

AGENDA

A Journal of Policy Analysis and Reform

Volume 20, Number 2, 2013

Contents

Editor and Editorial Committee	1
--------------------------------	---

ANALYSIS

Excellence in Research for Australia: An Audit of the Applied Economics Rankings Sinclair Davidson	5
Fabricating Invention: The Patent Malfunction of Australian Patent Law Hazel V. J. Moir	21
Stimulating Savings: An Analysis of Cash Handouts in Australia and the United States Sinclair Davidson and Ashton de Silva	39

ARGUMENT

Is Policy Too Important to be Left to Empiricists? Lessons of the 2012 Nobel Prize in Economics Richard Cornes and José A. Rodrigues-Neto	61
A Critique of the Productivity Commission's Cost-Benefit Analysis in the 'Disability Care and Support' Report Mark Harrison	77
<i>The Power to Tax</i> , 33 Years Later Jonathan Pincus	89

DEBATE

"Why Johnny Can't Regulate": A Reply to Ergas Daryl Biggar	105
A Rejoinder to Biggar Henry Ergas	111
A Final Word Daryl Biggar	113

Editor

Dr William Coleman College of Business and Economics, ANU

Editorial Committee

Professor Harry Clarke	La Trobe University
Professor Max Corden	University of Melbourne
Dr Anne Daly	University of Canberra
Professor Kevin Davis	University of Melbourne
Professor Henry Ergas	University of Wollongong
Professor John Gibson	University of Waikato
Dr Stephen Grenville	Lowy Institute, Sydney
Dr Arthur Grimes	Motu Research, Wellington
Professor Kim Hawtrey	BIS Shrapnel
Professor Doug Irwin	Dartmouth College
Dr Gary Johns	
Professor Tony Makin	Griffith University
Dr Deborah Peterson	Department of Primary Industry, Melbourne
Professor Jeff Pope	Curtin University of Technology
Professor David Vines	Balliol College, Oxford
Dr Ruth Williams	La Trobe University

ANALYSIS

Excellence in Research for Australia: An Audit of the Applied Economics Rankings

Sinclair Davidson¹

Abstract

The Excellence in Research for Australia (ERA) Report attempts to quantify the quality of research undertaken in Australian universities in the field of 'Applied Economics'. The paper shows it is difficult to reconcile the ERA rankings with the underlying data drawn from the Scopus database. Since the ERA rankings cannot be replicated, and since the ERA process is non-transparent, its rankings should be treated with some caution.

Introduction

The second Excellence in Research for Australia (ERA) National Report was released in December 2012. The importance of this Report cannot be understated. It will be used to allocate government funding to universities and also to separate out 'top performers' from 'also-rans'. As Stephen Matchett (2012b) of *The Australian* wrote:

Anybody contemplating an MBA can use ERA 2012 to assess every university claims that its bized (sic) staff are world-class researchers; in most cases they aren't. Have a look at the commerce and management list, there is a majority of low and no scores.

And the elite institutions can make the case they deserve all public research money, given they do so much of the world-class work. Of course everybody else can argue that the presence of Wollongong and Macquarie among the top performers makes the case for Canberra spreading the cash about. But that is an argument which applies to those two, not the 20 or so also-rans.

¹ RMIT University, sinclair.davidson@rmit.edu.au. I thank William Coleman, Ashton de Silva, and two anonymous referees for valuable comments on an earlier draft of this paper.

Some have gone further; Peter Roberts (2012) writing for the *Australian Financial Review* has used the 2010 ERA results to discredit policy advice given by Australian economists.

The average ranking of all fields of Australian research ERA measured was 3.4 on a scale of five, a score not achieved by any of the four measured strands of local economics.

Applied economics and other economics rated 2.1 and 2 respectively with any rating in the twos indicating a field which is performing below world average quality. Economic theory rated 3.1 and econometrics 3.3 which, while considered world average, were nevertheless below the Australian academic average.

Finally, of the 35 universities teaching economics here, ERA rated only seven institutions as above or well above world average quality; 14 received a score of 1, very much below world average quality.

At the same time, the ERA exercise has quickly gained broad acceptance. As *The Australian's* Matchett (2012a) explains: 'ERA doesn't just quantify research output it measures productivity. Which will upset academics who never quite get around to writing anything.'

In plain language: anyone speaking out against the ERA must either be bone idle or simply making excuses for underperformance.

The reality is that the ERA report results rely on some courageous assumptions. First, that government (or its agencies) can *define* quality. Second, that government (or its agencies) can *measure* quality. Third, that quality can be sufficiently represented in a single number between 1 and 5.

In this critique I do not challenge these fundamental assumptions, important as they are. Nor do I attempt to critique the ERA ranking process relative to the existing literature that attempts to provide similar rankings (for example, see Rodgers and Valadkhanim 2006 for a recent Australian example of this literature). I simply investigate whether ERA rankings are consistent with objective information.

To that end, I calculate the average citations for articles published between 2005 and 2010 for the 30 Australian universities that have a 2012 ERA ranking in the area of 'Applied Economics'. I employ the same global database that the ERA uses. I investigate that data along with an alternative measure of quality (the H-Index). Finally, I relate the ERA rank to the objective data I have gathered and discuss the anomalous results.

In short, it is difficult to reconcile the published ERA rankings with the objective data that the ERA claims as being an input into its ranking process. Some universities are ranked well above the level we might otherwise expect, while others well below the level we might expect.

The ERA process

The ERA 2012 Evaluation Handbook claims several objectives, with the first being:²

- Establish an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australia's higher education institutions.

To that end, the ERA defines a five-point scale:

5. '... well above world standard ...'
4. '... above world standard ...'
3. '... at world standard ...'
2. '... below world standard ...'
1. '... well below world standard ...'

It is quite remarkable that the ERA does not define what 'world standard' is, but rather tells us what it is not.³ We are told, "'World Standard" refers to a quality standard. It does not refer to the nature or geographical scope of particular subjects, nor to the locus of research nor its place of dissemination.' As reassuring as that is, there is still no definition of the 'quality standard' that 'world standard' entails. Further, 'The ERA evaluation measures research quality, not scale or productivity.' Indeed, 'The methodology and rating scale allow for [units of evaluation] with different volumes of output to achieve the same rating.'

Universities submit data to the Australian Research Council (ARC) that administers the ERA. That data relate to so-called Units of Evaluation — that is, a research discipline within a university. That Unit of Evaluation need not be a single department or school within a university. The Australian Research Council recruits individuals for Research Evaluation Committees and individuals

² Handbook: 3.

³ Ibid: 22–3.

who undertake peer review. Evaluation takes place, outcomes are determined, and results are published. The peer reviewers and members of the Research Evaluation Committees are bound by very strict confidentiality clauses.⁴

[Research Evaluation Committee] members and Peer Reviewers are required to sign a confidentiality agreement with the ARC prior to their participation in ERA. The agreement covers all aspects of their work with ERA, and the agreement survives the conclusion of their engagement for the purposes of ERA.

[Research Evaluation Committee] members and Peer Reviewers may not contact researchers and/or institutions under any circumstances in relation to material that has been submitted for evaluation in ERA, or seek additional information from any sources. [Research Evaluation Committee] members and Peer Reviewers may not reveal details about any evaluation at any time.⁵

The most important consideration, however, is this (emphasis added):

[Research Evaluation Committee] members exercise their *knowledge, judgement and expertise* to reach a single rating for each [Unit of Evaluation]. ... The rating for each [Unit of Evaluation] reflects the [Research Evaluation Committee] members' *expert and informed view* of the characteristics of the [Unit of Evaluation] as a whole.

In other words, the 'evaluation framework' that the ERA provides is not a transparent and replicable process; but, rather, an exercise where unknown individuals, acting in secret, selected by the government (or its agencies), express an opinion as to the quality of research relative to an undefined benchmark.

The ERA claims to have calculated 'Indicator principles' that are quantitative and 'internationally recognised'.⁶ To that end the ERA undertakes a citation analysis using the world-recognised Scopus database. According to its blurb, this database is 'the largest abstract and citation database of peer-reviewed research literature with more than 20 500 titles from more than 5000 international publishers. Scopus offers researchers a quick, easy and comprehensive resource to support their research needs in the scientific, technical, medical and social sciences fields and arts and humanities.'⁷

4 Ibid: 12.

5 I had been a Peer Reviewer in 2010 and was approached to be a Peer Reviewer in 2012. Upon reflection I refused to agree to the confidentiality clause.

6 Handbook: 21.

7 <http://www.info.sciverse.com/scopus/scopus-in-detail/facts>.

ERA interrogated Scopus as at 1 March 2012 to calculate the following citation count:

(1) World benchmark = [Sum of cites for all eligible articles in world data set] / [Total sum of eligible articles in world data set].

The ERA process is quite complicated but is more or less a citation count. It is possible to undertake a back-of-the-envelope replication of that process. In the next section I perform that back-of-the-envelope calculation for the Field of Research (FoR) code 1402 'Applied Economics'.⁸

Benchmarking the 2012 ERA results for Applied Economics

Table 1 below shows the overall ERA results for 'Economics' as a discipline, together with the various sub-disciplines that make up 'Economics'. While I concentrate on the 'Applied Economics' results (third column), I quickly wish to discuss some of the (other) anomalous results that are immediately apparent in the table. For example, both Flinders University and the University of Newcastle have overall scores, but no individual score in any of the sub-disciplines. There is no obvious explanation as to how that could occur. On the other hand, there are several universities that have multiple scores in the sub-categories but there is no discussion as to how those scores were aggregated to an overall score. This is problematic; consider The Australian National University and the University of New South Wales. Both those institutions have identical scores in the subcategories but overall The Australian National University scores '5', while the University of New South Wales scores '4'. Finally, some universities have lower overall scores than the subcategory scores would suggest (James Cook University, University of Canberra and University of Wollongong). Each of those universities has a lower overall score than its subcategory scores would have suggested.

⁸ The Economics discipline has the Field of Research code '14'. Within that Field is the sub-field 'Applied Economics' with a Field of Research code '1402'. The other fields are 'Economic Theory' code '1401', 'Econometrics' code '1403' and 'Other Economics' code '1499'. I have chosen Field of Research code '1402' for my analysis.

Table 1: ERA gradings of economic research by university and field of research

	14 Economics	1401 Economic Theory	1402 Applied Economics	1403 Econometrics	1499 Other Economics
Australian National University	5	4	4	5	n/a
Bond University	2	n/a	2	n/a	n/a
Central Queensland University	2	n/a	2	n/a	n/a
Charles Darwin University	3	n/a	3	n/a	n/a
Charles Sturt University	2	n/a	2	n/a	n/a
Curtin University of Technology	2	n/a	2	2	n/a
Deakin University	3	n/a	3	n/a	n/a
Edith Cowan University	2	n/a	2	n/a	n/a
Flinders University	3	n/a	n/a	n/a	n/a
Griffith University	2	n/a	2	n/a	n/a
James Cook University	1	n/a	2	n/a	n/a
La Trobe University	3	n/a	3	n/a	n/a
Macquarie University	3	n/a	3	n/a	n/a
Monash University	5	n/a	4	5	n/a
Queensland University of Technology	3	n/a	3	n/a	n/a
RMIT University	2	n/a	2	n/a	n/a
University of Adelaide	4	n/a	4	n/a	n/a
University of Ballarat	2	2	n/a	n/a	n/a
University of Canberra	1	n/a	2	n/a	n/a
University of Melbourne	5	5	4	5	n/a
University of New England	2	n/a	2	n/a	n/a
University of New South Wales	4	4	4	5	n/a
University of Newcastle	1	n/a	n/a	n/a	n/a
University of Queensland	4	5	4	4	n/a
University of South Australia	3	n/a	3	n/a	n/a
University of Southern Queensland	2	n/a	2	n/a	n/a
University of Sydney	4	3	4	4	n/a
University of Tasmania (inc. Australian Maritime College)	2	n/a	2	n/a	n/a
University of Technology, Sydney	5	n/a	5	5	n/a
University of Western Australia	4	n/a	4	n/a	n/a
University of Western Sydney	2	1	2	n/a	n/a
University of Wollongong	2	n/a	3	n/a	n/a
Victoria University	2	n/a	2	n/a	n/a
Total UoEs evaluated	33	7	30	8	0

According to the ERA, there were 3765.3 journal articles submitted by Australian universities in the category 1402 'Applied Economics'. There were also 112.1 books, 926.5 book chapters, 511.9 conference papers and 41.9 non-traditional research outputs submitted for evaluation. (The fractions are explained by universities submitting the same research output in more than one category.)

I accessed the Scopus database and undertook a search for all articles that can be described as being 'Applied Economics' over the period 2006–10.⁹ I collected for each university (i) the number of articles recorded by Scopus (henceforth 'Documents'), (ii) the number of citations reported by Scopus (henceforth 'Citations'), and (iii) the H-Index calculated by Scopus (henceforth 'H-Index').¹⁰ Finally, I have calculated the average number of citations by dividing the sum of citations by the sum of articles (henceforth 'Average-Cites'). I calculated these four metrics for each of the 30 Australian universities that have an ERA score for '1402 Applied Economics' (see Table 1) and report them in Table 2.¹¹

The first thing to note is that there are far fewer documents recorded in Scopus than the number of items Australian universities have submitted for evaluation. That could be explained by Scopus' limited coverage relative to the ERA list of journals. That suggests that Scopus may not be as reliable in this analysis as may have been hoped. It also suggests that any articles not published in a Scopus journal and/or not referenced in Scopus journals are likely to be undervalued by this process.

The raw data of Table 2 also suggests some interesting anomalies. For example, 'Applied Economics' research at Charles Darwin University is ranked by the ERA as being at the world standard. Yet over a six-year period up to December 2012 it has just one article in the Scopus database; an article that has received just four citations. It may be the case that Charles Darwin University has a (large) number of articles not published in Scopus-listed journals, and that these papers, in turn, were heavily cited by those very journals.¹² Similarly, the University of Technology, Sydney with a mere 20 articles, 241 citations, and an

9 While the ERA analysis is at 1 March 2012, my analysis is at 8 December 2012. I have a slightly longer timeframe, but given that most journals appear quarterly (and sometimes less frequently) I have less additional information than the nine-month difference indicates.

10 The H-Index is described at <http://en.wikipedia.org/wiki/H-index>. An academic with an index of h will have published h papers each of which has been cited at least h times.

11 Over that six-year period the Scopus database has 15 848 articles that, in some way, can be described as being 'Applied Economics'. Those 15 848 articles have been cited 64 362 times. Using equation (1) above, that implies an average citation rate of 4.06. That rate, however, is biased — articles published in 2006, for example, have been cited on average 10.1 times, while articles published in 2010 had only been published 0.91 times. Quite clearly there are publication and citation lags. That is to be expected and the ERA acknowledges that lag. Results of this exercise are shown in Table 2. ERA also calculated equation (1) on a year-by-year basis while I have only calculated it for the entire period 2005–10. The year-by-year analysis would sum up to the score I have calculated.

12 There is a low-volume threshold that must be reached before any analysis can occur. According to the Handbook (p.5), that threshold is 50 research outputs.

H-Index of 8 (barely at the Australian average) has earned an ERA ranking of 5 — ‘well above the world standard’. They too may have a number of articles not published in Scopus-linked journals that have in turn been cited by Scopus-linked journals. That sort of argument, however, cannot explain why Griffith University has 114 articles, 1157 citations, and an H-Index of 18 yet is scored ‘below world standard’ on its research by ERA.

Table 2: Measures of research performance in Applied Economics from Scopus

Institution	1402 Applied Economics (ERA Rank)	Documents in Scopus 2005–10	Citations in Scopus	Average Cites	H-Index
Australian National University	4	89	599	6.73	13
Bond University	2	14	109	7.79	6
Central Queensland University	2	8	50	6.25	8
Charles Darwin University	3	1	4	4.00	1
Charles Sturt University	2	16	234	14.63	6
Curtin University of Technology	2	44	95	2.16	5
Deakin University	3	74	589	7.96	13
Edith Cowan University	2	15	55	3.67	4
Griffith University	2	114	1157	10.15	18
James Cook University	2	11	273	24.82	6
La Trobe University	3	31	188	6.06	8
Macquarie University	3	29	85	2.93	6
Monash University	4	189	1708	9.04	23
Queensland University of Technology	3	25	250	10.00	10
RMIT University	2	40	257	6.43	8
University of Adelaide	4	30	137	4.57	6
University of Canberra	2	7	11	1.57	2
University of Melbourne	4	91	481	5.29	11
University of New England	2	20	259	12.95	7
University of New South Wales	4	44	224	5.09	8
University of Queensland	4	83	597	7.19	11
University of South Australia	3	25	145	5.80	7
University of Southern Queensland	2	4	17	4.25	3
University of Sydney	4	42	367	8.74	9
University of Tasmania (inc. Australian Maritime College)	2	13	191	14.69	3
University of Technology, Sydney	5	20	241	12.05	8

Institution	1402 Applied Economics (ERA Rank)	Documents in Scopus 2005–10	Citations in Scopus	Average Cites	H-Index
University of Western Australia	4	72	518	7.19	11
University of Western Sydney	2	33	88	2.67	5
University of Wollongong	3	34	129	3.79	6
Victoria University	2	58	325	5.60	10
Total UoEs evaluated	30	1276	9383	7.35	8.07

Source: ERA, Scopus, Author's calculations.

I then calculate a correlation matrix to investigate the relationships between the data. The result of that exercise is shown in Table 3.

Table 3: Correlations of ERA rankings with Scopus Performance Metrics (p statistics in parentheses)

	Documents	Citations	Ave_Cites	H-Index
ERA Rank	0.4147 (0.0230)	0.3329 (0.0720)	-0.0756 (0.6910)	0.3966 (0.0300)
Documents		0.9411 (0.0000)	-0.0263 (0.8900)	0.9218 (0.0000)
Citations			0.2261 (0.2300)	0.9380 (0.0000)
Ave_Cites				0.1540 (0.4160)

Source: ERA, Scopus, Author's calculations.

The single largest correlation with the ERA Rank is the number of documents contained in the Scopus dataset. While the ERA has emphasised that it intends to measure quality and not quantity, this result is not problematic. It is quite likely that more articles published in journals tracked by the Scopus database is itself an indicator of quality. At the very least it is likely to be a proxy for quality (Silva 2012). That does, however, raise questions about the ERA's rankings of institutions such as Charles Darwin University and the University of Technology, Sydney.

The second-largest correlation with the ERA Rank is H-Index. This is entirely expected — the H-Index is a measure of quality. If anything, it is somewhat surprising that the correlation is as low as it is and is not higher. Similarly, the third-highest correlation with the ERA Rank — Citations — is expected. What is somewhat surprising is the small, *negative*, and not statistically significant correlation between ERA Rank and average citations. That measure is derived

from equation (1) — the very measure the ERA employed as its ‘World citations per paper benchmark’.¹³ The low correlation between the ERA benchmark and the well-known H-Index is also somewhat troubling.

Next Table 4 shows some summary statistics for the data I have extracted from Scopus relative to the ERA Ranks. The first thing to note is that there is no monotonic increase in any of the measures across the four ERA Ranks. The single university awarded a rank of 5 (University of Technology, Sydney) underperforms the eight universities with a rank of 4 in three of the four measures. It is only for Average Citations that the University of Technology, Sydney outperforms the eight universities in the 4 ranking. In that metric, however, the 14 universities ranked 2 outperform the seven universities ranked 3. Overall there is no clear relationship between the ERA Ranks and the data extracted from Scopus.

Table 4: Scopus performance metrics by ERA rank

	ERA Rank	2	3	4	5	ALL
Documents	Mean	28.4	31.3	80.0	20.0	42.5
	Median	15.5	29.0	77.5	20.0	30.5
	Maximum	114.0	74.0	189.0	20.0	189.0
	Minimum	4.0	1.0	30.0	20.0	1.0
	Std. Dev.	29.4	21.7	49.9		40.2
	N	14	7	8	1	30
Citations	Mean	222.9	198.6	578.9	241.0	312.8
	Median	150.0	145.0	499.5	241.0	229.0
	Maximum	1157.0	589.0	1708.0	241.0	1708.0
	Minimum	11.0	4.0	137.0	241.0	4.0
	Std. Dev.	288.5	188.7	486.2		358.5
	N	14	7	8	1	30
Ave-Cites	Mean	8.4	5.8	6.7	12.1	7.5
	Median	6.3	5.8	7.0	12.1	6.3
	Maximum	24.8	10.0	9.0	12.1	24.8
	Minimum	1.6	2.9	4.6	12.1	1.6
	Std. Dev.	6.5	2.5	1.7		4.8
	N	14	7	8	1	30
H-Index	Mean	6.5	7.3	11.5	8.0	8.1
	Median	6.0	7.0	11.0	8.0	7.5
	Maximum	18.0	13.0	23.0	8.0	23.0
	Minimum	2.0	1.0	6.0	8.0	1.0
	Std. Dev.	4.0	3.7	5.1		4.6
	N	14	7	8	1	30

Source: ERA, Scopus, Author’s calculations.

¹³ Handbook: 95.

It is possible to display the data visually.¹⁴ The box and whisker plots (see Figure 1) demonstrate the similarities between the various measures as opposed to the differences. In terms of Documents, Citations and H-Index, it is unlikely that the data statistically significantly differentiate Rank 2, Rank 3, and Rank 5. Placing that in context, it appears there is no difference between those universities ranked by the ERA as being below world standard and well above world standard. On the other hand, those eight universities merely 'above the world standard' (Rank 4) perform *better* than University of Technology, Sydney, which is ranked 'well above the world standard'.

By contrast, on Average Citations there is unlikely to be a statistically significant difference between those universities with Rank 2, Rank 3, and Rank 4. But 2 is 'below world average', while 4 is 'above world average'. There is little to separate the 30 institutions from each other, yet the ERA has been able to rank them from 'below world standard' to 'well above world standard'.

I then use the data collected from the Scopus database to estimate the simple Ordinary Least Squares regressions shown in Table 5.¹⁵ The regressions are ranked in terms of their Adjusted R² measures. The values in parenthesis are p-values.

Table 5: Scopus performance metrics as a predictor of ERA rank

Model 1	ERA Rank = 2.455630 + 0.009664Documents (0.0000) (0.0227)	Adj-R ² = 0.142378
Model 2	ERA Rank = 2.210717 + 0.081316H-Index (0.0000) (0.0300)	Adj-R ² = 0.127214
Model 3	ERA Rank = 2.594483 + 0.000870Citations (0.0000) (0.0722)	Adj-R ² = 0.079066
Model 4	ERA Rank = 2.977474 - 0.014837Ave_Cites (0.0000) (0.6914)	Adj-R ² = -0.029799

Source: ERA, Scopus, Author's calculations.

Table 5 reveals that Average Citations has zero predictive power with respect to ERA rankings. The coefficient on Average Citations is negative, but not statistically significantly different from zero. This confirms the previous analysis suggesting that there is no relationship between the ERA ranks and the citation measure that the ERA employs in its analysis. In addition the Citations variable is only statistically significantly different from zero at the 10 per cent level.

¹⁴ I thank Ashton de Silva for providing the box and whisker plots.

¹⁵ Strictly speaking OLS is not entirely appropriate for this purpose — the dependent variable is discreet and bound between 1 and 5. An ordered choice model would be more appropriate; however, when I perform that exercise the coefficients are not statistically significantly different from zero.

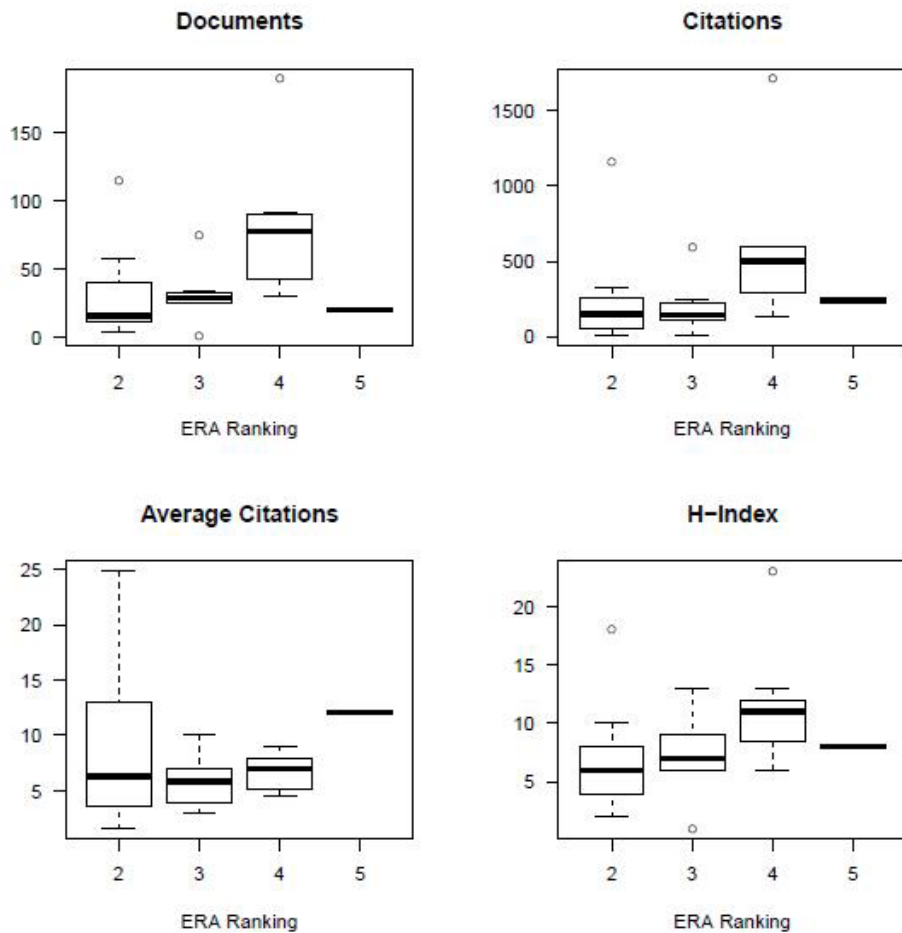


Figure 1: Box and whisker plots

Source: Ashton de Silva (see p15, footnote 14).

But the Documents variable and the H-Index variable are both statistically significantly different from zero at the 5 per cent level. Thus Model 1 (Documents variable) and Model 2 (H-Index variable) are my preferred model specifications in the analysis that follows. Model 3 (Citations) and Model 4 (Average Citations) are shown for completeness.

Table 6: Scopus metrics prediction of ERA rank

Institution	Actual ERA Rank	Fitted 1	Fitted 2	Fitted 3	Fitted 4	Excess Score 1	Excess Score 2	Excess Score 3	Excess Score 4
Australian National University	4	3	3	3	3	1	1	1	1
Bond University	2	3	3	3	3	-1	-1	-1	-1
Central Queensland University	2	3	3	3	3	-1	-1	-1	-1
Charles Darwin University	3	2	2	3	3	1	1	0	0
Charles Sturt University	2	3	3	3	3	-1	-1	-1	-1
Curtin University of Technology	2	3	3	3	3	-1	-1	-1	-1
Deakin University	3	3	3	3	3	0	0	0	0
Edith Cowan University	2	3	3	3	3	-1	-1	-1	-1
Griffith University	2	4	4	4	3	-2	-2	-2	-1
James Cook University	2	3	3	3	3	-1	-1	-1	-1
La Trobe University	3	3	3	3	3	0	0	0	0
Macquarie University	3	3	3	3	3	0	0	0	0
Monash University	4	4	4	4	3	0	0	0	1
Queensland University of Technology	3	3	3	3	3	0	0	0	0
RMIT University	2	3	3	3	3	-1	-1	-1	-1
University of Adelaide	4	3	3	3	3	1	1	1	1
University of Canberra	2	3	2	3	3	-1	0	-1	-1
University of Melbourne	4	3	3	3	3	1	1	1	1
University of New England	2	3	3	3	3	-1	-1	-1	-1
University of New South Wales	4	3	3	3	3	1	1	1	1
University of Queensland	4	3	3	3	3	1	1	1	1
University of South Australia	3	3	3	3	3	0	0	0	0
University of Southern Queensland	2	2	2	3	3	0	0	-1	-1
University of Sydney	4	3	3	3	3	1	1	1	1
University of Tasmania (inc. Australian Maritime College)	2	3	2	3	3	-1	0	-1	-1
University of Technology, Sydney	5	3	3	3	3	2	2	2	2
University of Western Australia	4	3	3	3	3	1	1	1	1
University of Western Sydney	2	3	3	3	3	-1	-1	-1	-1
University of Wollongong	3	3	3	3	3	0	0	0	0
Victoria University	2	3	3	3	3	-1	-1	-1	-1

Source: ERA, Scopus, Author's calculations.

Table 6 reports the ERA rank as predicted by the four models of Table 5, and the difference between the actual rank and the predicted rank.

For 24 of the 30 universities all four models provide similar results. In particular, all four models would have predicted that the University of Technology, Sydney should have earned an ERA rank of 3, rather than 5. The data suggest that the University of Technology, Sydney is at 'the world standard' (however that is defined), and not 'well above the world standard'.

The Group of Eight universities are said to be research-intensive and generally hold the reputation of being Australia's best universities. All eight have an ERA rank of 4 — 'above world standard'. But only Monash has a fitted value of 4 (for three of the four models). Every other one of the other Group of Eight has fitted values of 3 — 'at world standard' — for all four models. In other words, there is evidence that the ERA process overestimated the ERA rank for, at least, seven of the eight Group of Eight universities. Random variation of the ERA process would suggest that only three or four of the Group of Eight ERA ranks would be overestimated, while three or four might be underestimated. It is extremely unlikely that seven would be overestimated if ranking errors were unbiased.

Griffith University has been the big loser from the ERA process. Three of the four models show a fitted value of 4 — 'above world standard' — while only one model (Model 4 — the least-preferred model) shows a fitted value of 3 — 'at world standard'. The ERA process actually allocated an ERA rank of 2 — below world average.

The evidence on Charles Darwin University is mixed. The two better models (Model 1 and Model 2) suggest that it should be ranked 'below world standard', while the less-preferred models (Model 3 and Model 4) suggest that its existing rank at the world standard is appropriate.

Discussion

The results shown in this paper are not mere quibbles. Given the time, effort and money invested in the ERA process, it is important to get the rankings and orderings correct (Trounson 2012a). It is not good enough to suggest that the difference between, say, an ERA rank of 2 and an ERA rank of 3 is a matter of subjective opinion. As the quotes from Matchett and Roberts in the introduction make clear, these numbers are important.

Furthermore, the Australian federal government has indicated that federal funding decisions will be decided on the basis of these rankings. To the extent that there is a deliberate and systematic upward bias for some universities,

an attempt to bias federal funding has occurred. It is easy to imagine that misrepresentation for financial gain may be viewed by some as being unethical if not illegal.

At the same time, Trounson (2012b) reports that hiring decisions are being made on the basis of earned ERA ranks.¹⁶ Yet the analysis shown here indicates that some ERA ranks might be substantially overstated — on the objective evidence the University of Technology, Sydney is at world standard, not well above world standard. Similarly, some ERA ranks are substantially understated. Griffith University has an ERA rank of 2 — below world standard — while the objective evidence suggests that it should be ranked as being above the world standard. An uncritical application of the ERA data would see a world-standard research environment being funded and promoted while an above-world-standard research environment being unfunded and not promoted. Given the stated aims of the ERA this would be a perverse result.

There may well be additional explanations for the ERA ranks that have been awarded. For example, the ERA included so-called esteem factors when forming their opinions as the world standing of research. The ERA Evaluation Handbook, however, states that esteem facts may be used to increase ERA ranks, but not reduce them.¹⁷ It could be that case that the seven of the eight Group of Eight universities that received higher ERA ranks than can otherwise be justified had very high esteem factors. The ERA Handbook provides examples of esteem as being editors of prestigious reference works, or various category 1 fellowships, or members of learned academies.

It might be possible that a small number of superstars are driving the results. University X may employ person Y and so generate a high ERA rank. This may happen, but would constitute a violation of the ERA rules. For example, under esteem measures, the ERA Evaluation Handbook states, ‘Individual researchers cannot be identified through the esteem measures’.¹⁸ Later we read, ‘[Research Evaluation Committee] members do not make comment about the contributions of individual researchers.’¹⁹ Later again: ‘Drill down menus are generally not available where information would allow the viewer to identify and track individual researchers.’²⁰

Ultimately, it is not at all clear why there is such a large discrepancy between the actual ERA ranks and objective information that can be derived from Scopus — the ERA’s chosen dataset. It is not surprising that most Australian universities should score at the world standard (however defined). What is surprising is that

16 <http://www.theaustralian.com.au/higher-education/competition-hots-up-for-research-big-fish/story-e6frgjcjx-1226534758787>.

17 Handbook: 22.

18 Ibid: 20.

19 Ibid: 23.

20 Ibid.

the Australian government and its research agency would claim that 14 out of 30 Australian universities were below the world standard without bothering to define what that standard might be. It is even more surprising that several universities could be ranked at above the world standard (or even well above the world standard) without any reliable objective evidence to support that ranking except for the 'expert and informed view' of the individuals making the ranking.

Until such time as the government or the ARC is capable of producing transparent and objective rankings, these rankings and this whole exercise should be viewed with some caution, if not scepticism and contempt.

References

- Australian Research Council 2012, 'ERA 2012 Evaluation Handbook', Commonwealth of Australia; http://www.arc.gov.au/pdf/era12/ERA%202012%20Evaluation%20Handbook_final%20for%20web_protected.pdf
- Matchett, S. 2012a, 'Big ERA is watching you', *The Australian Higher Education*, 7 December 2012; http://blogs.theaustralian.news.com.au/thecommonroom/index.php/theaustralian/comments/appalled_by_accountability/
- 2012b, 'Research race is on', *The Australian Higher Education*, 12 December 2012; http://blogs.theaustralian.news.com.au/thecommonroom/index.php/theaustralian/comments/research_race_is_on/
- Roberts, P. 2012, 'Dismal science's drive to mediocrity', *The Australian Financial Review*, 16 January 2012; http://www.afr.com/p/opinion/dismal_science_drive_to_mediocrity_OU3Wv88THQH0c1o3QeHKbN
- Rodgers, J. and Valadkhanim, A. 2006, 'A Multi-Dimensional Ranking of Australian Economics Departments', *Economic Record* 82(256): 30–43.
- Silva, P. 2012, 'More papers, better papers? The curious correlation of quality and quantity in academic publishing', *The Australian Higher Education*, 13 December 2012; <http://www.theaustralian.com.au/higher-education/more-papers-better-papers-the-curious-correlation-of-quality-and-quantity-in-academic-publishing/story-e6frgcjx-1226535254934#sthash.GvbqGhEJ.dpuf>
- Trounson, A. 2012a, 'ERA attacked for being too expensive, too detailed and too frequent', *The Australian Higher Education*, 8 December 2013; <http://www.theaustralian.com.au/higher-education/era-attacked-for-being-too-expensive-too-detailed-and-too-often/story-e6frgcjx-1226532299099#sthash.Usgx3nR8.dpuf>
- 2012b, 'Competition to hot up for research big fish in new ERA', *The Australian Higher Education*, 13 December 2012; <http://www.theaustralian.com.au/higher-education/competition-hots-up-for-research-big-fish/story-e6frgcjx-1226534758787#sthash.lMy4opGt.dpuf>

Fabricating Invention: The Patent Malfunction of Australian Patent Law

Hazel V. J. Moir¹

Abstract

Despite advice to parliament that patents are granted only for ‘a significant advance over what was known and what was available to the public’² the evidence shows this is not the standard used. The actual standard is a scintilla — a marginal difference from what is known. The consequence of such a low standard is that thousands of patents are granted for things that contribute no public benefit. Such trivial patents can impede genuinely inventive companies.

‘The production of the knowledge of how to do in a somewhat different way what we have already learned to do in a satisfactory way would hardly be given highest priority in a rational allocation of resources.’

(Machlup 1958: 51)

The reform process

In 2009 the Australian government — through IP Australia — began a process to reform the standard patent system. This covered a range of issues but the focus of this article is the reform of the ‘inventiveness standard’. IP Australia commented that there were ‘a number of areas in which Australia’s requirements for inventive step are set at a lower level than those of other jurisdictions’ (IP Australia 2009a: 9).³ Particular concerns were ‘allowable existing knowledge’ and the ‘test for inventiveness’, as each can lead to the grant of patents for trivial inventions.

1 The Australian National University; hazel.moir@anu.edu.au. I would like to thank an anonymous referee and William Coleman for extremely helpful comments on the initial draft. Needless to say, the views presented in this paper are my own and do not represent the views of any other person or organisation.

2 Explanatory Memorandum, Intellectual Property Laws Amendment (Raising the Bar) Bill 2011: 42.

3 Lawson, reviewing a number of legal cases, also concluded that legal doctrines had driven Australia’s inventive step to a very low level. He concludes that now ‘obviousness will almost never be relevant in assessing patentability’ (Lawson 2008: 44).

The reform proposals were that:

- global common general knowledge replace Australian common general knowledge;
- statutory restrictions on allowable existing knowledge ('prior art') be removed;⁴ and
- one particularly low inventiveness test be replaced (see below).

The subsequent legislative amendments — the so-called Raising the Bar amendment bill — implemented the first two of these measures, but not the third. Removing the restrictions on allowable existing knowledge should reduce the volume of trivial patents granted in low-technology fields. But these changes will not affect trivial patenting in high-technology fields — there are no cases where restrictions to prior art have been used to uphold the validity of trivial patents in high-technology fields.

In 2009 IP Australia had proposed increasing the inventiveness threshold test from whether a skilled person 'would be *led directly as a matter of course* to try a particular approach with a reasonable expectation of success' (*Lockwood v. Doric* [2007] HCA 21, emphasis added) to 'obvious for the skilled person to *try a suggested approach*, alternative or method with a reasonable expectation of success' (IP Australia 2009a: 12–13, emphasis added). Such a change would have raised the inventive step from a scintilla to a small amount — a small step in the right direction. IP Australia later changed its position, advising that concerns about the low height of the inventive step would be addressed 'through restating the guidelines for inventive step in the Examiners' Manual and through more rigorous application of the inventive step tests during examination, rather than through changes to the law' (IP Australia 2009b: 12). On its face this reliance on the Examiners Manual seems dubious. Are Federal and High Court judges likely to accept that doctrines (policies) they have established are over-ridden by changes to a manual?⁵

This article proceeds by reviewing why the inventiveness proxy is so central to economic balance in a patent system (Section 2). A consideration of how inventiveness is administered shows that the threshold for grant of a patent is set at a very low level. Section 3 presents empirical evidence on the quantum

4 Specifically, wording that allowed patent information to be excluded from allowable knowledge in low-technology fields (*Commissioner of Patents v Emperor Sports* [2006] FCAFC 26).

5 The changes are to the obvious to try section (2.5.3.3.5, modified August 2010) and to sections on mirror image chemicals (enantiomers) (section 2.5.3.4.1 and section 2.5.4.1.3, both modified in September 2012) (communication from IP Australia 13 February 2013). The Examiners' Manual is at http://www.ipaustralia.gov.au/pdfs/patentsmanual/WebHelp/Patent_Examiners_Manual.htm (accessed 23 September 2013). The amended 'obvious to try' section clarifies that such objections are relevant only if 'there is a reasonable expectation that the solution might well solve the problem'. Examiners are advised they must give reasons why they consider the solution obvious to try; for example, that the technology field is highly predictable. On its face, this does not appear to deal with the 'led directly as a matter of course' doctrine.

of inventiveness required for a standard, demonstrating that this is indeed a scintilla not 'a significant advance in what is known'. With such a low threshold a very large proportion of granted patents will be for inventions with no spillover benefits. In the absence of systematic data on the cost of trivial patents, section 4 considers some limited evidence from legal decisions about the costs of trivial patents. Section 5 considers the options for reform, pointing particularly to the proposed Intellectual Property chapter in the Trans Pacific Partnership Agreement as a potential, but critical, impediment.

Inventiveness

Monopolies are generally anathema within the economics profession, and for good reason. But in patent policy the argument that the static efficiency losses from the monopoly grant will be offset by dynamic efficiency gains from new inventions is often accepted. This positive view of the patent intervention is based on a widespread belief that new technology provides significant spillover benefits. If such spillovers did not exist, the rationale for the patent intervention would disappear.

The patent eligibility criterion of the inventive step can be seen as a very rough proxy for the existence of spillover benefits. If an invention embodies significant new knowledge then spillover benefits are more likely. But if an invention contributes only trivial differences then there is no new knowledge, no spillovers and no rational basis for patent grant. Indeed grant of such patents will be welfare-reducing.

The legal and patent communities see the patent *quid pro quo* as the public disclosure of the invention. This is radically different to the economic perspective focusing on spillover benefits. Where the inventive step is high, the two perspectives can come into alignment with the public disclosure, providing a path through which the new knowledge can be disseminated.⁶ But if the inventive step is low, there is no useful knowledge content. In such cases disclosure simply adds noise to the system, hiding truly inventive patents.

Ideally patents would only be granted for inventions which would not otherwise take place *and* which provide spillover benefits. Both theory and data suggest that those inventions which are highly codified can be copied more quickly and cheaply (Mandeville 1996; Levin *et al.* 1987). In such cases, first-mover advantages could be insufficient to fully recoup expenses, particularly for high-cost inventions. Unfortunately TRIPS (Article 27(1)) precludes limiting patents

⁶ Survey evidence suggests that very few industrial (or indeed academic) researchers look to patent databases for new knowledge (Mandeville, Lamberton and Bishop 1982; Oppenheim 2000; Macdonald 2003).

to high-cost codified inventions. Firstly, TRIPS mandates no discrimination by technology field, so a limitation to inventions with highly codified technologies is not allowed. Secondly, TRIPS appears to mandate the current utility, novelty and inventiveness concepts as the sole screening devices for patent eligibility, implying that any other eligibility criterion, such as cost of R&D, would contravene the treaty.

Neither TRIPS nor the Australia–US Free Trade Agreement (AUSFTA) preclude setting the inventiveness requirement at a level which maximises the likelihood of spillover benefits and minimises the likelihood of granting patents where there are no dynamic efficiency gains. Proposals to prevent such policy reform have been tabled during negotiations for the Trans Pacific Partnership Agreement (TPPA).⁷

Economic reviews have (reluctantly, as it is a blunt instrument) endorsed the patent system based on the *assumption* of a reasonable inventiveness standard (BIE 1994; IPCRC 2000). Given the clear evidence of a very low standard for grant of a patent,⁸ the assumption needs testing. It is, after all, critical to whether the dynamic efficiency gains are likely to offset the static efficiency losses.

Turning to how the inventiveness criterion is administered, it is surprising to an economist to discover that the first step is a series of rules which eliminate substantial existing knowledge from consideration. These rules determine the subset of existing knowledge ('prior art') that forms the baseline against which inventiveness is assessed. This is followed by rules determining how existing knowledge can be combined. Only after this is the question of the quantum of inventiveness addressed.

One might imagine this would be done by asking 'How much new knowledge or know-how is in this invention?' Not so. A reverse approach is taken — what the patent office and the courts ask is 'Is the invention obvious?' If it is not, a patent is granted. There are complex and detailed rules about determining obviousness. For example, the 'obviousness' judge must be a person skilled in the relevant art (PSA) and must not be imaginative (van Caenegem 2007: 85).⁹ In this context 'art' means technology field, and this has become very narrowly defined. For example, the Federal Court determined that well-known dynamic

7 <http://keepthewebopen.com/assets/pdfs/TPP%20IP%20Chapter%20Proposal.pdf> Article 8(1), footnote 15.

8 The large literature on the low inventiveness standard is mostly based on US data (for example, Jaffe and Lerner 2004; Lunney 2004; Quillen Jr. 2006). Australia's inventiveness standard is considered to be lower (IP Australia 2009a: 6).

9 In the USA the Supreme Court has recently ruled that this skilled person has normal imagination (*KSR v. Teleflex* 127 S.Ct. 1727 (2007)).

storage technology was inventive when used in a consumer-loyalty smart chip.¹⁰ Dynamic storage was well-known (obvious) to IT experts, but not to consumer-loyalty experts. This seems to reward ignorance rather than inventiveness.

There is a large gulf between inventions embodying significant advances in knowledge and inventions which pass the 'Is it obvious?' test. A possible distribution of patent applications by inventiveness, where very many inventions have only a small degree of inventiveness, and only a small number are radically inventive, is shown in Figure 1. This allows a focus on the different outcomes achieved by asking 'Is it obvious?' compared to 'Is it sufficiently inventive?' Given the narrow and constrained approach to the patent test for obvious, many applications which pass this test will not be reach the 'significant advance' standard.

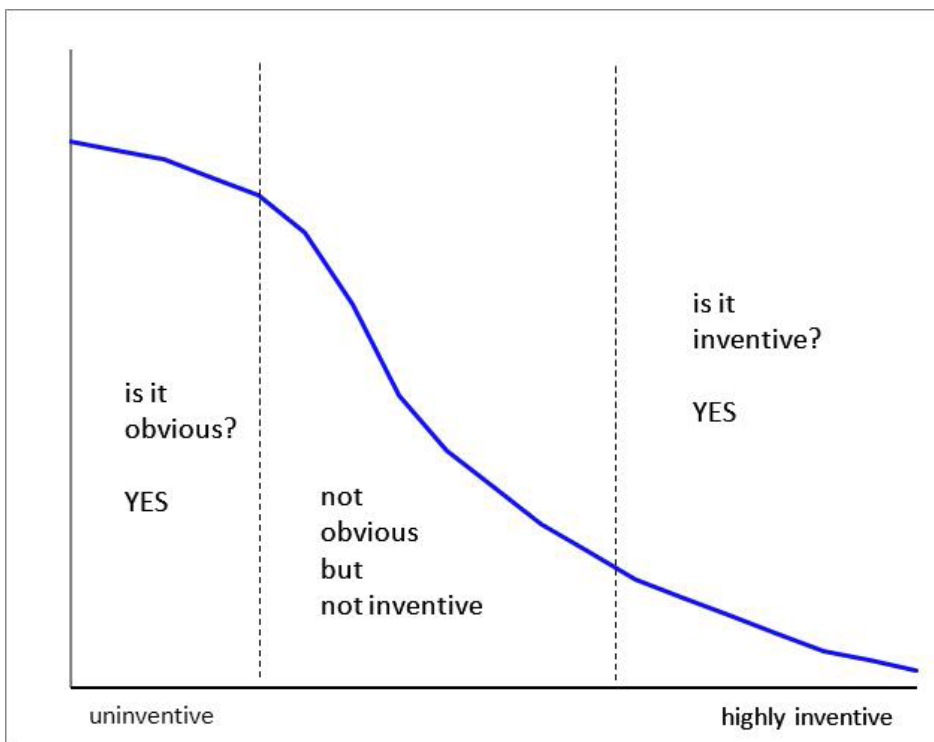


Figure 1: Hypothetical inventiveness continuum and determining patentability

Source: Author's representation.

¹⁰ *Welcome Real-Time SA v. Catuity Inc*, [2001] FCA 445 (17 May 2001). This issue, breadth of the technology field, is more fully considered by Bagley (2001), though with respect to US patent law.

The two issues that set the inventiveness standard are the rules determining the 'prior art' and the rules determining inventiveness. Problems with the prior art rules should be fixed by the Raising the Bar amendments, though there is a substantial time lag before these take full effect (21 years from Royal Assent).¹¹ But these rules have not been used to uphold trivial patent validity in high-technology fields so will have no impact on, for example, the grant of the secondary (evergreening) pharmaceutical patents. This is where next step — the rules determining inventiveness — are important.

Inventiveness: Evidence from business method patents

Inventiveness is qualitative and thus difficult to measure. While some studies use variables such as prior art citations, there is a lack of independent verification of these as good proxies for inventiveness (Gittelman 2008). Dahlin and Behrins (2005) have shown that the number of technology classes can provide a good measure of radical inventiveness, but the reverse does not work to identify obvious inventions. The one empirical study that does provide genuine insights into the quantum of inventiveness is of a set of 50 of the 'best' US software patents granted.¹² This concluded that all were only incrementally inventive (Campbell-Kelly and Valduriez 2005).

My empirical study of business method patents was designed to provide an assessment of the inventiveness of granted Australian standard patents from an economic perspective. Instead of asking whether inventions were obvious, the research question was 'What new knowledge or know-how is contributed?' The dataset is a point-in-time universe of business method patents. All applications classified as business methods and filed in the years 2003 to 2006 were selected provided they had been sealed or accepted by 1 July 2007. The full specification was read for each of the 72 cases, and the claims were 'construed' to identify the essence of the invention. This was compared to knowledge at the time to identify what new knowledge was disclosed.¹³ None of the 72 cases disclosed any new knowledge though there were two possibly new ideas (which did not contribute

11 These changes apply only where examination was requested on or after 15 April 2013. The rationale for this lengthy implementation delay is unclear, given that all existing applications were grandfathered to the previous low standards. There was an avalanche of applications just before the change — the month to 14 April 2013 saw 6300 applications compared to around 2000 normally (<http://blog.patentology.com.au/2013/04/raising-bar-raises-patent-filings.html>).

12 'Best' was defined on the basis of forward citations (subsequent patents citing that patent), a variable which is associated with private value but has not been shown to be related to spillover benefits.

13 This step was completed by the December 2007. The relevant date for existing knowledge for each application (the 'priority date') ranges from 1998 to 2005.

any new knowledge when implemented).¹⁴ In this article several cases from this study are discussed briefly to demonstrate the lack of new knowledge required for an Australian patent.

Many inventions take known elements and combine these in different ways. Such new combinations of well-known elements or processes ('combination' patents) raise substantial policy challenges. Early patent law did not face this difficulty as patents were available only for new inventions, not for improvements (Walterscheid 1995). Once patents for improvements were allowed (from 1776), courts had to face the challenge of dealing with combination patents. The synergy doctrine, still used in Europe, requires that to be patentable a combination of known elements must produce either a new and unexpected result or a result that is greater than the sum of its parts. That is, a combination of old elements can only be granted a patent if some new knowledge or know-how is produced.

Unfortunately this doctrine was overturned in Australia in 1980 by the High Court.¹⁵ The US Court of Appeals for the Federal Circuit made a similar ruling a few years later (Harris 1986; Lunney 2004). Now, in both the USA and Australia, examiners may not reject applications which combine well-known elements or processes *unless they can find written evidence that it is obvious to combine them*. This can be hard to find. Examiners are thus forced to grant patents for such 'inventions' as teaching children about finance by having them work for their pocket money (more below). Although the 'suggestion' doctrine clearly leads to a very low inventiveness standard there have been no moves to reform it in the over three decades since its introduction. As the data in this section demonstrate, the consequence of this doctrine (policy), and other doctrines determining obviousness, is a standard that falls far short of a significant advance in what is known or used.

All 72 cases in my study involve at best only trivial variations from other known processes. Of these, 41 can be classified as 'combinations', of which 21 simply involve combining a known process with a computer. A further three combine known processes with modern electronic networks. Examples of computerisation cases include printing documents near the delivery point; automating financial analyses for pharmacies; an expert system to manage dialysis treatment; lodging patent applications; assessing organisational business risks; predicting when work elements become due; and valuing assets. The last of these (2005203023)

14 The original study provides a full discussion of the inventiveness and knowledge contribution of each case (<https://digitalcollections.anu.edu.au/bitstream/1885/49313/4/02whole.pdf>). An updated version looking at what rules lead to granting trivial patents (in Australia, Europe and the USA) is provided in Moir (2013).

15 *Minnesota Mining and Manufacturing v Beiersdorf* (1980) 144 CLR 253 at 117.

simply computerises what every property valuer has done for decades.¹⁶ The examiner objected that the process was just common general knowledge, but the attorney argued that valuers often cut corners and failed to keep records. A patent was granted.

One of the network patents simply transfers the processes for ordering ID cards to the internet (2004201620). The process starts by verifying the client's computer's identity. The supplier then obtains appropriate data (over the internet) to populate the fields in the ID template. An image of the product is shown to the client for approval or amendment. That's it. Clearly this 'invention' would fail any ordinary meaning of the word 'inventive'.

Combination 'inventions' that were not simple computerisations include eight initially rejected by examiners. Claims were then amended and a patent granted. The correspondence between applicant and examiner shows how trivial are the distinguishing features which create what the patent system calls inventiveness.

- In the case of an on-line employment register (2005234625) the applicant successfully argued that the *absence* of features included in earlier systems created an inventive difference in this combination — a kind of 'less is more' argument.
- The inventive feature in a process for allocating patients to emergency facilities (2003248001) is a satellite communications gateway — a feature added to avoid rejection for lack of inventiveness. By adding one extra (known) feature the invention now became sufficiently inventive for a patent.
- In a process for combining date-stamped photos and written reports (2003246060) the report is generated on a server, rather at the site. This change in physical location was the key feature making this sufficiently inventive for a patent.

The dataset contained 18 cases where the 'inventions' were only trivial variations.

- An 'invention' to teach children about finance by working for their pocket money (2003203582) has multiple features — learning about credit by getting an advance is probably an 'invention' that every child has independently worked out. Less obvious elements are sickness insurance and buying shares or real estate. The combination of these many well-known elements in the environment of pocket money meant that the examiner could not find documented evidence of obviousness. If the synergy doctrine was still used

¹⁶ Application 2005203023. Numbers given for cases are application numbers and the specification can be found by entering this number into the AusPat database (<http://pericles.ipaustralia.gov.au/ols/auspat/>) then following the link for the B2 document.

or if the inventiveness standard were a 'significant advance over what was known' this patent would not have been granted.

- A combined loan/investment 'invention' (2005265435) uses the loan both for the investment and to pay off part of the loan. It has five simple steps: receive loan parameters; determine rule-based payments; determine allocations for repayments between the loan and acquiring the investment; receive payments; and use payments based on allocation. A standard patent was immediately granted. There is nothing new about any element of this process — at the priority date (November 2004) a wide variety of geared investment products were on the market.
- An 'invention' predicting recurring events (2006200104, priority May 2005) is very similar to Microsoft's well-known Outlook product, available globally since 1997. The only difference is the option of the Julian calendar, a small additional feature. After examiner objections the claims were amended, moving some words from claim 2 into claim 1. These words were a highly generalised specification of converting dates to numbers and back — required for any computerised calendar system as computers can only work with binary data (ASCII). The patent was granted.

Of the remaining cases in the dataset, three involved analogous use of known processes, four simply link databases and three are just ideas, written at such a high degree of generality that no new knowledge is disclosed.

The analogous-use doctrine was developed to set a floor to the patent system by ensuring that patents were not granted for every trifling device. The English courts firmly established the principle that analogous uses (uses of a well-known thing for a different purpose for which its properties were well-suited) were not patentable as 'there would be no end to the interference with trade, and with the liberty of any mechanical contrivance being adapted, if every slight difference in the application of a well-known thing were held to constitute a patent ...'.¹⁷

There are three cases in this dataset that appear to be simple uses of well-known processes in different contexts where their properties make them ideally suitable:

- Benchmarking was widely used during the micro-economic reform period of the late 1980s and early 1990s.¹⁸ The 'invention' here (2004200942, priority March 2003) is a classic application of benchmarking, using sustainability indicators for real estate projects and comparing their values with benchmarks. That's it.

¹⁷ Lord Chancellor in *Hamilton* (1865) at 1499. The principle was originally determined in 1838 when considering a patent claim for using a wheel on railway carriages (Brennan and Christie 1997: 239, citing *Losh*).

¹⁸ Australian government reports using this technique and dating from 1995 are at www.pc.gov.au.

- Reward systems for consumer loyalty have a very long history. Loyalty ‘stamps’ were widely used in UK retailing in the 1960s. Modern versions include frequent-flyer reward schemes and Coles’ Fly-Buys process. Patent 2003280529, priority November 2002, simply applies such an award system to tickets. A clear analogous use of a well-known process.
- The surprising thing about a patent for audit processes on the quantum of chemicals in foodstuffs (2004233489, priority November 2004) is that many similar patents had already been granted. As audit is a process designed to be used in a wide range of environments, none should have been granted. This one was in fact granted because it *covered a slightly longer part of the supply chain*. This marginal difference in the breadth of application was sufficient to achieve the alleged ‘significant advance’ standard for patent grant.

In case after case this research project demonstrates that the quantum of inventiveness required for grant of a standard patent in Australia is close to zero. None of the patents in the dataset contributed any new knowledge or know-how. The minimum inventiveness standard for grant is clearly substantially below the claimed ‘significant advance over what was known or used’.

How far can these results be extended to other technology fields? TRIPS mandates no discrimination by field of technology, so the patentability requirements of utility, novelty and inventiveness are applied in the same way whatever the technology field. The fact that none of the patents embodied any new knowledge (let alone a significant contribution) suggests that a very high proportion of granted patents generate no spillover benefits.

But some granted patents are inventive and deliver significant new knowledge. Work by Dahlin and Behrens (2005) suggests about 1 per cent are truly radical with up to 4 per cent being quite radical. Their dataset implies that in total about 20 per cent have some degree of real advance. This aligns well with informed (but never written) views from the patent community that about 5 per cent of patents are truly inventive, about another 10–15 per cent are quite inventive and the 80–85 per cent are ‘rubbish’.¹⁹

In 2012, IP Australia granted 17 724 patents.²⁰ There are likely to be at least some genuine inventions among these — possibly as many as 3550. But some 14 000 to 15 000 may contribute few, if any, spillover benefits. Even if only 50 per cent of granted patents have no spillovers, this substantially increases the volume of patenting, imposing higher search costs on all innovating firms. It also allows for the substantial growth of a body of intermediaries who have a vested interest in larger volumes and hence lower standards. Allowing a very low standard also

19 Personal discussion with a (now retired) senior member of the Australian patent community, 7 October 2004.

20 Table 1, IP Australia, Australian Intellectual Property Report 2013 (<http://www.ipaustralia.gov.au/about-us/corporate/reports/?doc=139839&view=Detail>).

encourages a rent-seeking attitude and brings the law into disrepute. These all imply substantial resource misallocation — a resource misallocation that is a direct consequence of the patent system.

What are the costs? Case law concerning trivial inventions

Patents are a powerful exclusive right — the right to exclude others from operating in a particular technology space. As such, the only parties against whom patents can be enforced are innovating firms.²¹ Such costs can include changes to the firm's technology mix, additional R&D to work round the patented space, licensing fees for technology which is very similar to what the firm invented itself, diversion of management time from running the business, fines and penalties as well as not insubstantial legal costs. Although the major policy concern is with the social costs of patent systems, most of the limited available data are on private costs (Moir 2009).

When it comes to the key area of concern — the possibly chilling effect on the innovation activities of other firms through exclusion or high license fees — few data are available. Some argue, *without any empirical evidence*, that few patents are used, so trivial patents are unlikely to inflict any economic damage (Lemley 2001; Blonder 2005). Other commentators (for example, Gans, Williams and Briggs 2004) consider that in most circumstances the effective monopoly power conferred by a patent is weak, and such costs will therefore be low to non-existent. If this is so, it is hard to see how patents create an incentive for R&D investment.

The policy interest is in how often innovating firms are held up by trivial patents owned by others and how significant such hold-up is. The National Innovation Survey does not collect data on whether firms have incurred costs due to patents held by other parties. Nor does the patent office collect any data on the use of the over 17 000 patent monopolies it currently hands out each year.²² In any other policy area such lack of outcome data would be severely criticised. But patents are off-budget so are not subject to normal audit and evaluation processes. Reviewing the historical evidence, Cohen finds that technological hold-ups can and do occur (Cohen 2005).

In the absence of systemic data on this important issue, one can consider the evidence from court cases. Even here, however, data are limited as costs and

21 Bessen and Meurer (2007). Individuals and micro-firms are usually not worth pursuing.

22 Despite IPAC's recommendation that such data be collected at the time patents are renewed (IPAC 1984). No patent office collects data on patent use.

damages are determined separately from substantive outcomes. As a result, databases on litigation outcomes rarely include cost data and where they do the data are limited to legal costs.²³ But the greatest policy concern is not legal costs but the range of other costs an innovating firm can experience if it is held up or diverted because of trivial patents.

Australian judgements where trivial patents are upheld and found to be infringed each point to firms who have been fined and forced to work around their own. Doric fought for seven years against a patent owned by Lockwood. But the High Court found that it was not obvious to take an element from a storeroom lock to fix a problem with a rim-mounted lock.²⁴ Doric had to pay costs, damages and stop using technology it invented. When the Federal Court upheld the validity of a dynamic storage card for a consumer loyalty use — even though all parties agreed dynamic storage was well-known in the IT area — Catuity had to stop using its own similar technology.²⁵ In every case where a court upholds a trivial patent as valid there is a firm which pays the price for these low standards.

When these trivial patents involve pharmaceuticals, some elements of their cost can be estimated. Clear policies for price falls for drugs listed on the Pharmaceutical Benefits Scheme (PBS) after generic entry allow estimation of the costs to taxpayers of delayed generic entry. Submissions to the recent pharmaceutical patents review provided data on a number of such cases.²⁶ The cost to taxpayers is different to the cost to generic firms of delayed entry, but it does show the magnitudes that can be involved:

- Efexor (venlafaxine) came off patent in late 2008. In the market it had effectively been replaced by the extended release version Efexor-XR. Generic entry of Efexor-XR was delayed for 2½ years by a patent dispute. With this high volume drug the estimated cost to the taxpayer from the delayed price fall is A\$209 million.
- During this time Pfizer heavily marketed a closely similar drug — desvenlafaxine — which has the identical therapeutic effect as venlafaxine. This achieved a 37 per cent market share for the replacement drug at an estimated annual cost to the taxpayer of A\$6.7 to A\$7.6 million a year. As venlafaxine is metabolised to desvenlafaxine in the human body, it is surprising that a patent was granted for such a small variation.

23 For example, Dent and Weatherall (2006) have no cost data. Helmers and McDonagh (2012) include limited data on private legal costs. A very small survey undertaken for the UK Strategy Advisory Board for Intellectual Property Policy (SABIP) asked firms involved in patent disputes about legal costs and management time taken in disputes. The survey was supplemented by in-depth interviews which show that the management time issue had a major negative effect on many businesses (Greenhalgh *et al.* 2010).

24 *Lockwood v. Doric* [2007] HCA 21.

25 *Welcome Real-Time SA v. Catuity Inc.*, [2001] FCA 445 (17 May 2001).

26 The website for the pharmaceutical patents review and the submissions to this has been taken down. Copies of submissions, the issues paper and the draft report can be obtained on request to the author.

- It is well known that one isomer of a racemic drug will be more effective than the other, and often more effective than the racemic mixture. The Patent Examiner's Manual was revised in September 2012 to state that patents on isomers would rarely be granted. Before this, Sanofi was granted a patent over the racemic mixture of clopidogrel then a further patent over its isomer (marketed as Plavix). A term extension was granted on the isomer patent. The isomer patent was eventually challenged and revoked, but it kept generic versions off the market for some 2½ years at an estimated cost to the taxpayer of A\$60 million during the injunction period.
- Clopidogrel is an anti-platelet agent that inhibits blood clotting. Sanofi also obtained a patent on the combination of clopidogrel and aspirin. Like clopidogrel, aspirin is used as a blood thinning agent — and has been known since 1899. Although both drugs were known, and were known for their use for the same purpose, the patent for the combination delayed generic entry. Eventually the combination patent was challenged. Sanofi resisted the challenge until the directions hearing on 29 May 2012 when it chose not to defend the patent — a clear sign of a very 'weak' patent. But this patent gave Sanofi 2½ years of higher prices at an estimated cost to taxpayers of A\$11 million.²⁷

Patent policy reform: Opportunities and threats

In 1984 the Fraser government asked IPAC to review the economic impact of the patent system. The review committee commissioned a range of empirical work and held a number of excellent seminars (Mandeville *et al.* 1982b, 1982a). But the one economist on the committee lodged a dissenting report claiming that there was little economic analysis in the actual report (IPAC 1984: 79–80). Instead, he said, the proposals pandered to sectional interests and missed an important opportunity for reform. Similarly the review of the patent system under the Competition Principles Agreement (the Ergas Report) simply *assumed* that patents were necessary and that the inventive step was at least reasonably high. Another opportunity for effective reform was missed. The draft report of the 2012–13 Pharmaceutical Patents Review accepted the proposition that a change had been made to inventiveness standards for pharmaceuticals despite the lack of any evidence to support this (Australian Government 2013: 121–4).

As noted in the introduction, IP Australia put forward proposals for a marginal increase in the inventiveness test in 2009, but then backed away this. The reasons given were that a number of submissions considered this

²⁷ Ibid: 26.

unnecessary (IP Australia 2009b: 12). None of the submissions to this reform process were ever made public, so it is unclear which parties argued to maintain the current very low standards.

It is clear, however, that a low standard means many more applications and thus directly affects the income for intermediaries. Patent Attorneys were established as a protected profession in the 1952 Patent Act — the patent equivalent of tariff consultants — and there are now 860.²⁸ Patent applicants are officially advised to ‘seek professional advice from a patent attorney’.²⁹ The cost per application is \$8000 or more (IP Australia 2013: 32). Another major beneficiary of low standards is the brand pharmaceutical industry which uses these low standards to obtain follow-on patents to extend their period of market exclusivity (Kapczynski *et al.* 2012). Burdon and Sloper (2003: 238) advise that while very uninventive patents may incur litigation, the returns to the patent owner will easily outweigh litigation costs, even when the patent is found invalid.

The imperviousness of the patent system to genuine reform is explainable by focusing on the interest groups involved and regulatory capture theory. The risks faced by innovating firms are widely distributed and there are no systems in place for collecting data on these risks and their cost.³⁰ Associated with this lack of any direct data on costs and benefits, there have been no economic evaluations of Australia’s patent system.³¹ The Productivity Commission, for example, has never been tasked with assessing whether the patent system contributes a net welfare benefit to Australia.

A further reason for the lack of any reform effort is that the patent system is off-budget. This was one reason why it appealed to notoriously poverty-stricken European monarchs. Now it reduces the imperative for reform. Combined with substantial lobbying resources — which successfully promote unsubstantiated slogans such as ‘patents are essential for innovation’³² — the lack of action is readily explained.

It becomes, however, a very risky strategy now that large global companies have successfully persuaded governments to insert these restraints on competition

28 These data were provided by IP Australia in response to my enquiry. Data on the number of Patent Attorneys registered each year do not appear to be publicly available — the regulatory board publishes no summary data or annual reports (<http://www.psb.gov.au>).

29 <http://www.ipaustralia.gov.au/get-the-right-ip/patents/patent-application-process/> as at 21 April 2013.

30 There is evidence that in the USA data collection and analysis has been successfully opposed by patent attorneys and major patenting firms (Kahin 2003). The Federal Trade Commission recommendation (FTC 2003) most prominently rejected by the Intellectual Property Owners Association (dominated by patent lawyers from large firms) was recommendation 10 ‘expand consideration of economic learning and competition policy concerns in patent law decisionmaking’ (Bessen and Meurer 2008: 293–4).

31 Indeed, I have not been able to locate a sound economic evaluation of any patent system in any country.

32 One of the global experts on the relationship between patents and innovation has concluded that the patent system ‘has been neither necessary nor sufficient for technical and/or economic progress at country and company level historically’ (Granstrand 1999: 44).

into the global trading system. If a country wants to participate in GATT it must provide patents, trademarks, copyright and so on, as mandated in TRIPS. The US government is pursuing TRIPS-Plus standards in bilateral and regional trade agreements, and the AUSFTA is an example. The Productivity Commission has warned strongly against including such non-trade issues in trade agreements (Productivity Commission 2010).

A leaked draft of the Intellectual Property chapter of the Trans Pacific Partnership Agreement provides detailed prescription on how patents, trademarks and copyright should be administered, including determining inventiveness by the 'Is it obvious?' approach. While this draft has been withdrawn and replaced, the secrecy surrounding the discussions continues to raise concerns about inappropriate intellectual property privileges being traded for small gains in agricultural exports. This would be a poor bargain as it is domestic market reforms, not enhanced export access, which lead to the major gains from global trade (Armstrong 2012: 1641). Further, it is information-intensive industries which are expected to show most future growth. To tie the hands of current *and future* governments with respect to reform in this important area of innovation policy would be an extremely expensive outcome.

New material about the cost of the low-inventiveness standard was placed on the table during the previous government's pharmaceutical patent review panel. This showed that the cost to taxpayers of secondary pharmaceutical patents can be substantial. In the current budgetary climate, proper reform of the patent inventiveness standard would flow directly to substantially reduced outlays through the PBS as exclusive marketing periods would no longer be extended through evergreening patents. Perhaps this can be an incentive to undertake genuine reform of the patent system?

While Australia is not alone in having a very low inventive step for grant of a patent monopoly, the fact that other countries have dysfunctional patent systems is no reason for Australia to follow suit. When it came to dismantling tariff barriers Australia had no hesitation in identifying that it was in Australia's economic interests to do this, whether others did so or not. We also showed leadership on reforming agricultural subsidies. Using the principles set forth in the Competition Principles Agreement, we could now show leadership in reforming patent systems to remove their worst excesses. This could be of benefit to many of our trade partners as well as to ourselves.

References

- Armstrong, S. 2012, 'Australian Trade Policy Strategy Contradictions', *The World Economy*: 1633–44.
- Australian Government 2013, 'Pharmaceutical Patents Review: Draft Report', Canberra; available at: <http://www.ipaustralia.gov.au/about-us/ip-legislation-changes/review-pharmaceutical-patents/> (viewed 23 September 2013).
- Bagley, M. 2001, 'Internet business model patents: obvious by analogy', *Michigan Telecommunications Technology Law Review* 7: 253–88.
- Bessen, J. E. and Meurer, M. J. 2008, *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*, Princeton and Oxford: Princeton University Press.
- BIE 1994, 'The Economics of Patents', Canberra.
- Boldrin, M. and Levine, D. K. 2008, *Against Intellectual Monopoly*, Cambridge: Cambridge University Press.
- Burdon, M. and Sloper, K. 2003, 'The Art of Using Secondary Patents to Improve Protection', *Journal of Medical Marketing* 3(3): 226–38.
- Campbell-Kelly, M. and Valduriez, P. A. 2005, 'A technical critique of fifty software patents', *Marquette Intellectual Property Law Review* 9(2): 249–305.
- Cohen, W. M. 2005, 'Patents and appropriation: concerns and evidence', *Journal of Technology Transfer* 30(1–2): 57–71.
- Dahlin, K. B. and Behrens, D. M. 2005, 'When is an invention really radical? Defining and measuring technological radicalness', *Research Policy* 34(5): 717–37.
- FTC 2003, *To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy*, Washington, D.C.: Federal Trade Commission (US).
- Gans, J. S., Williams, P. L. and Briggs, D. 2004, 'Intellectual property rights: a grant of monopoly or an aid to competition?' *The Australian Economic Review* 37(4): 436–45.
- Gittelman, M. 2008, 'A note on the value of patents as indicators of innovation: implications for management research', *Academy of Management Perspectives* 22(3): 21–7.

- Granstrand, O. 1999, *The Economics and Management of Intellectual Property — Towards Intellectual Capitalism*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
- Greenhalgh, C., Philips, J., Pitkethly, R., Rogers, M. and Tomalin, J. 2010, 'Intellectual Property Enforcement in Smaller UK Firms', Report for the Strategy Advisory Board for Intellectual Property Policy (SABIP).
- Harris, R. W. 1986, 'Prospects for Supreme Court review of the Federal Circuit standards for obviousness of inventions combining old elements', *Journal of the Patent and Trademark Office Society* 68(2): 66–82.
- Helmets, C. and McDonagh, L. 2012, 'Patent Litigation in the UK', London.
- Hemphill, C. S. and Sampat, B. N. 2012, 'Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals', *Journal of Health Economics* 31: 327–39.
- IP Australia 2009a, 'Getting the Balance Right: Toward a Stronger and More Efficient IP Rights System', Canberra: IP Australia.
- 2009b, 'Toward a Stronger and More Efficient IP Rights System: Proposed Reforms', Canberra: IP Australia.
- 2013, 'Australian Intellectual Property Report 2013', Canberra.
- IPAC 1984, *Patents, Innovation and Competition in Australia*, Canberra: Industrial Property Advisory Committee; available at: http://www.acip.gov.au/reviews_other.html.
- IPCRC 2000, 'Review of Intellectual Property Legislation Under the Competition Principles Agreement: Final Report', Canberra: Attorney-General's Department (Intellectual Property and Competition Review Committee).
- Jaffe, A. B. and Lerner, J. 2004. *Innovation and its Discontents: How Our Broken Patents System is Endangering Innovation and Progress, and What to Do About It*, Princeton: Princeton University Press.
- Kahin, B. 2003, 'Information process patents in the US and Europe: policy avoidance and policy divergence', *First Monday* 8(3).
- Levin, R. C., Klevorick, A. K., Nelson, R. R. and Winter, S. G. 1987, 'Appropriating the returns from industrial research and development', *Brookings Papers on Economic Activity, Special Issue on Microeconomics* 1987(3): 783–831.
- Lunney Jr., G. S., 2004, 'Patent law, the Federal Circuit, and the Supreme Court: a quiet revolution', *Supreme Court Economic Review* 11: 1–79.

- Macdonald, S. 2003, 'Bearing the burden: small firms and the patent system', *The Journal of Information, Law and Technology (JILT)* 2003(1) at: http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2003_1/macdonald/.
- Machlup, F. 1958, 'An Economic Review of the Patent System (Study No.15 of the U.S. Senate Subcommittee on Patents, Trademarks and Copyrights)', Washington, D.C.: US Government Printing Office.
- Mandeville, T. D. 1996, *Understanding Novelty: Information, Technological Change, and the Patent System*, Norwood, N.J.: Ablex.
- Mandeville, T. D., Lamberton, D. M. and Bishop, E. J. 1982, *Supporting Papers for the Economic Effects of the Australian Patent System*, Canberra: Australian Government Publishing Service.
- Moir, H. V. J. 2009, 'What are the costs and benefits of patent systems?' in *Intellectual Property Policy Reform*: 29–54, W. Van Caenegem and C. Arup (eds), Cheltenham, UK and Northampton, MA, US: Edward Elgar.
- Oppenheim, C. 2000, 'Patent information usage in British small and medium sized enterprises'.
- Productivity Commission 2010, 'Bilateral and Regional Trade Agreements', Canberra: Productivity Commission Research Report.
- Quillen Jr., C. D. 2006, 'Innovation and the U.S. patent system', *Virginia Law and Business Review* 1: 210–25.
- Steensma, D. P. and Loprinzi, C. L. 2006, 'Epoetin Alfa and Darbepoetin Alfa Go Head to Head', *Journal of Clinical Oncology* 25(15): 2233–6.
- Walterscheid, E. C. 1995, 'The Early Evolution of the United States Patent Law: Antecedents (Part 3 continued)', *Journal of the Patent and Trademark Office Society* 77(11):847–57.

Stimulating Savings: An Analysis of Cash Handouts in Australia and the United States

Sinclair Davidson and Ashton de Silva¹

Abstract

At the onset of the Global Financial Crisis governments around the world implemented fiscal stimulus packages. A key component of many of these packages was aimed at stimulating consumer spending. In Australia and the United States, for example, households received one-off cash payments. We assess the changes in the macroeconomic levels of consumption and savings of both countries coinciding with the timing of the household bonuses using an econometric time series method known as seemingly unrelated time-series equations. The results suggest that the one-off cash bonuses did not stimulate consumption. On the contrary, the evidence suggests savings was stimulated.

Introduction

The Global Financial Crisis (GFC) began in mid-2007, peaking in 2009. In that time a massive economic dislocation occurred around the world, but largely centred on North America and Western Europe. The causes and consequences of the GFC will be debated for decades. In this paper we examine the fiscal response to the GFC in two economies; Australia and the United States. In contrast to typical approaches we provide a unique comparison; namely the effectiveness of fiscal policy under different prevailing economic conditions.

It is important to appreciate that although these economies are based on the same economic principles, their size and economic character prior to the GFC differ considerably. Australia is a small open economy whereas the US is the largest open economy in the world. Further, their economic experience during the GFC differs considerably. Specifically, the United States was severely impacted by the crisis, millions of individuals lost their jobs and old and venerated financial

¹ RMIT University; ashton.desilva@rmit.edu.au.

institutions were swept away into insolvency. By contrast, Australia weathered the storm reasonably well; not even experiencing two consecutive quarters of negative GDP growth.

The fiscal responses in each economy were similar. Indeed, the Australian government response was predicated on the US experience during the GFC and previous US recessions. Both the Australian government and the US government undertook a series of cash handouts to households (often described as tax refunds), various tax cuts, and spending programs in the hope of maintaining economic activity. The objective, in the language of policymakers, was to substitute public demand for collapsing private demand. In this paper we focus on the cash handouts that occurred in Australia and the US during the GFC. Australian policymakers, for example, had the explicit objective of getting 'the money into the pockets of people who would take it straight to the shops' (Taylor and Uren 2010: 77). This policy, however, appears to be inconsistent with the permanent income hypothesis where a temporary increase in disposable income should have no impact on consumption (Hall 1978).

The Australian experience would suggest that the fiscal interventions were successful. The *Economic Security Strategy* paid a total of A\$8.7 billion to households in December 2008, while the *Nation Building and Jobs Plan* delivered an additional A\$12 billion to households over March, April and May 2009. The Australian economy avoided two consecutive quarters of negative GDP growth (the local, if somewhat unscientific, definition of a recession). Unemployment did not rise anywhere near the forecast levels (with or without the stimulus packages), whereas in the US unemployment rose beyond forecast levels even with the stimulus packages.

By contrast, the US experience would seem to suggest that the fiscal intervention was unsuccessful. The 2008 *Economic Stimulus Act*² saw US\$168 billion in tax refunds and the 2009 *American Recovery and Reinvestment Act* saw an additional US\$288 billion in tax cuts to business and families (households). Given the deep and prolonged recession in the US and the high levels of unemployment, it seems fair to say the US fiscal policy response to the GFC failed. Perhaps the packages were too small, or poorly implemented, or poorly targeted; all manner of argument can be made and has been made.

We examine aggregate consumption in Australia and the US during the GFC using an econometric time-series technique known as seemingly unrelated time-series equations. By using this method we are able to construct a counterfactual and so estimate the impact of government measures. Importantly, our analysis does not rely on examining survey data.

2 <https://www.govtrack.us/congress/bills/110/hr5140/text>.

In our analysis we consider three important questions: Is there any evidence that household spending increased significantly as a result of the handouts? Conversely, were the handouts saved? Why is it that handouts appeared to ‘work’ in Australia but not the US? To answer these questions we model consumption, income and savings together with gross domestic product, consumer sentiment, and stock market indices. By using these variables we are effectively examining macroeconomic data for evidence of the permanent income hypothesis (Hall 1978) while taking into account economic conditions. Overall our results are consistent with the permanent income hypothesis.

A review of the literature

Macroeconomic theory suggests that most of the cash handouts would have been saved rather than being spent. In a Ricardian world, the beneficiaries of the cash handout would realise that their future tax liabilities had increased; they would then increase savings to meet that future cost (Barro 1974). Ricardian equivalence implies that temporary fiscal measures to stimulate the economy should fail.

At the time the Australian policy was formulated, a small literature examining the 2001 US cash handout policy suggested that handouts would be spent, and not saved.

Johnson, Parker and Souleles (2006) had used survey data to examine the effects of the 2001 US cash handout and reported that 20 to 40 per cent of the rebate had been spent over three months and up to two-thirds of the rebate was spent over six months. They interpret these results as being inconsistent with the permanent income hypothesis. By contrast Shapiro and Slemrod (2003), also using survey data, find that the 2001 cash handout was saved. In particular that spending was low relative to economists’ expectations.

Agarwal, Liu and Souleles (2007) employ credit card data to analyse the response to the 2001 US tax rebates. They report that consumers initially saved some of the rebate and then increased expenditure: ‘For consumers whose most intensively used credit card account is in the sample, spending on that account rose by over \$200 cumulatively over the nine months after rebate receipt, which represents over 40 per cent of the average household rebate.’

It is interesting to reflect on that statement, *40 per cent of the rebate was spent over nine months*. Presumably the other 60 per cent was either saved or consumed via other means. Nonetheless, they too interpret their results as being inconsistent with the permanent income hypothesis.

The Australian Treasury *explicitly* relied on an unpublished study by Broda and Parker (2008) as support for its argument that cash handouts would be spent (Uren 2009).³ This study investigated the 2008 US\$950 tax rebate by exploiting timing differentials and comparing spending in those households that had received the rebate to spending in households that were eligible but had not yet received it. Those households that had received the rebate consumed more than those that had not. Broda and Parker had concluded 'the stimulus payments are initially being spent at significant rates'. Yet, on average, only US\$448 was spent in additional purchases, the other 52 per cent of the handout was (at least temporarily) saved. In a later, more complete, paper Broda and Parker (2012) argue that households increased their spending by 10 per cent in the week the cash handout arrived — while that sounds quite large, the US-dollar amount was just \$14 in that week. Over the next seven weeks US consumers in receipt of the cash handout spent an additional US\$30–50.

In an Australian replication of the Broda and Parker methodology, Aisbett *et al.* (2012) report that the change in household consumption was less than one per cent, which they describe as being insignificant and quantitatively small.

By contrast Chakrabarti *et al.* (2011) employ several data resources, including credit card records and household surveys, and show that average consumption decreased while savings increased during and after the 2007 recession.

A reading of the international literature would have suggested to policymakers that there was some scope for demand management policies in countering an economic downturn. Within Australia, for example, policymakers were of the opinion that the permanent income hypothesis was not empirically valid (Taylor and Uren 2010: 74). Olekalns (1997) does report that the permanent income hypothesis is inconsistent with the evidence prior to financial deregulation in the early 1980s, but since that time Australian household behaviour is consistent with the permanent income hypothesis. After the second stimulus package was completed, a survey to determine what had happened to the cash component of the stimulus package was conducted (Leigh 2009). Overall it found that 60 per cent of the cash had been saved or used to pay off debt. Comparing his results to Shapiro and Slemrod (2003), Leigh argued that this was at twice the rate that Americans had spent similar tax rebates in 2001.

The difficulty with most of this literature is that it relies on survey instruments and requires respondents to accurately recall, and then accurately disclose

3 Interestingly, an investigation by Harris *et al.* (2002), published in Australia's leading economic journal, seems to have been ignored. Important findings of the investigation include: 'Based on these estimates, a negative shock to the economy that increases consumer pessimism will increase the number of people saving by a considerable proportion....' (pp.219–20) and 'The results suggest that individuals who are pessimistic in their outlook are likely to save more to account for an uncertain future' (p.221), which raises serious doubt as to whether households would actually spend the bonus (or a large portion thereof).

their consumption behaviour. Also problematic is that researchers can never know the counterfactual. What would otherwise have happened? Alternative empirical strategies do exist, but this usually involves estimating the size of various Keynesian-type multipliers. This approach is controversial as it relies heavily on *a priori* theorising on the size and magnitude of the multipliers (see Auerbach *et al.* 2010 for a detailed discussion).⁴ In the next section we set out our empirical strategy that does not rely on survey results, or Keynesian multipliers, and does generate a plausible counterfactual.

The econometric method

The econometric technique we employ is known as Seemingly Unrelated Time Series Equations (SUTSE, Harvey 1989).⁵ The SUTSE approach falls under the umbrella of state space modelling. State space modelling has been used successfully in many areas of economic research. Key examples include Engle and Watson (1987), Aoki (1987), Harvey and Jaeger (1993), Harvey and Koopman (1997), Prioretti *et al.* (2007), Panher (2007) and de Silva *et al.* (2009).

If we let y_t denote a collection of variables at time t , the multivariate state space formulation can be represented as follows:

$$y_t = Hx_t + \varepsilon_t, \quad (1)$$

$$x_t = Fx_{t-1} + v_t, \quad (2)$$

where (1) captures how the observations evolve over time according to the latent trend and seasonal components of each series, specified in x_t . How these components evolve over time and combine to measure the observations (y_t) is (pre)determined by the H and F . In the base form, inter-series associations are modelled through the variance-covariance matrices of the error terms ε_t and v_t Σ_ε and Σ_v respectively. Further details of the model specification are contained in Appendix 1.

The motive for using the SUTSE approach in this investigation is three-fold. One, it provides the means of modelling the series consistent with the Sims (1980) method; that is, we do not need to impose any prior theoretical restrictions on the model, thus letting the data speak. Two, we can model more than one variable simultaneously. Further, these variables can be modelled ‘as is’; that is, we do not need to engage in transformations such as differencing that

4 Recent evidence suggests that multipliers with respect to the ARRA are ‘modestly’ positive (0.52) in the short run and ‘modestly’ negative (-0.42) in the long run (Drautzburg and Uhlig 2011). Similarly, Makin and Narayan (2011) suggest that multipliers are close to zero for the Australian context.

5 The model is fitted using the computer package STAMP 8.2 (Koopman *et al.* 2009).

will necessarily result in changes in interpretation. Finally, the model assumes stochastic time-series characteristics, thus it is flexible enough to account for any variation in the components, such as seasonal factors, over time. This last point is particularly important as it distinguishes it from the typical form of modelling approaches such as ordinary least squares.

The ability to estimate the (latent) trend and seasonal component of a given set of time series is particularly important as it provides the means to be able to assess whether structural changes have occurred in the different components of the economic variables under consideration. Specifically, the employment of the SUTSE formulation allows us to determine whether the time path of the economic variables has significantly altered from its historically defined underlying trend. We do this by testing whether there has been a change in the trajectory of the underlying trend by testing the series for level breaks. A positive (negative) significant level break indicates the level trend has moved up (down). In addition we also identify one-off departures from trend, referred to as impulse breaks.

The US experience

Figure 1 presents the three time series of interest for the US; consumption, income and savings.⁶ All three variables are available from the Bureau of Economic Analysis (BEA <http://www.bea.gov/>) and are supplied in seasonally adjusted form only.⁷

Three instances of cash stimulus have occurred in the US since 1985: 2001 Q3 and Q4; 2008 Q2 and Q3; and 2009 Q2. Each of these periods is identified in Figure 1 by the grey shading.⁸

Consumption and income appear to have grown fairly steadily over the period since 2000. Close inspection of the chart reveals that income peaked slightly in 2001 and 2008 corresponding to the periods in which cash handouts occurred. Conversely, consumption does not appear to exhibit any local peaks; however, in recent times it seems to have deviated negatively from its long-run path.

6 Savings is a residual term and includes down payment of debt etc.

7 The BEA provides the annualised data at quarterly intervals; thus the data are divided by four to get the approximate seasonally adjusted quarter values. The seasonally adjusted nature of the data does somewhat modify our empirical strategy. The difference being that a seasonal component was not modelled.

8 During 2001 quarters three and four, US households received rebates as part of the Economic Growth and Tax Relief Reconciliation Act of 2001. Rebates were administered on a randomised basis during this period ranging from \$300 to \$600 per eligible household. A summary of this package is presented in Johnson, Parker and Souleles (2006).

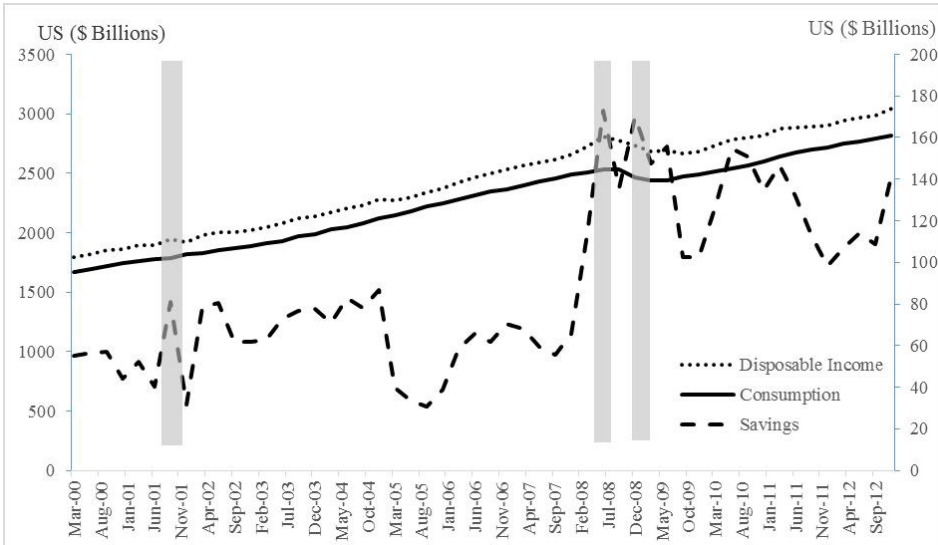


Figure 1: US macroeconomic data

Source: Authors' calculations; Bureau of Economic Analysis.

Unlike consumption and income, savings is highly volatile and therefore possesses a significant modelling challenge. It appears a number of breaks in the series have occurred. To capture the behaviour of aggregate savings during the period of interest these breaks have to be modelled. An example of a break is the period from 2005 Q1 to 2005 Q4 where the series exhibits a downward shift in the underlying average. Interestingly, periods corresponding to various cash handouts demonstrate positive spikes in saving. Importantly, the SUTSE approach allows us to capture these structural breaks in a straightforward manner.

In addition to the three variables of interest, Gross Domestic Product (GDP), consumer sentiment and the Dow Jones Industrial index were modelled.⁹ The motivation to include these additional control variable stems from the argument that to model the three variables appropriately general macroeconomic conditions also need to be taken into consideration.

To test the effectiveness of the various cash handouts impulse dummies are fitted to the data. An impulse dummy is defined as taking a value of one in the period of interest and zero in all other periods. In addition to these dummies a level break was also tested for in 2009 Q4. By including a level break, we are aiming to determine whether the time path has permanently altered. In summary, the

⁹ The Dow Jones Industrial Average was sourced from DX time series. The consumer sentiment index from the Surveys of Consumers, University of Michigan.

purpose of fitting these dummies is to determine whether there has been any statistically significant departure from what can be considered as ‘normal’ or ‘baseline’ according to the historical properties of the series as determined by the SUTSE formulation and fitting technique.

The key results of our analysis are presented in Table 1 (further results are available upon request in the form of an appendix). These results are based on fitting two models to the natural log of the data spanning 1985 Q1 to 2012 Q4 (T=110).¹⁰ Table 1 is divided into two panels; focus and control variables. The focus variables are those that are of primary interest. The table shows whether deviations from the trend are positive or negative and whether the deviation is statistically significantly different from the trend.

Table 1: Significant departures from trend coinciding with household payments (US)

Date	Panel A: Focus Variables		Panel B: Control Variables			
Year (Quarter)	Consumption	Savings	Sentiment	GDP	Stock Index	Disposable Income
2001(3)		Positive***				
2001(4)		Negative***	Negative**			Negative**
2008(1)	Negative*	Positive***		Negative**	Negative**	
2008(2)	Negative**	Positive***	Negative***	Negative**	Negative**	Positive***
2008(3)	Negative***	Positive***	Negative**	Negative***	Negative***	
2008(4)	Negative***	Positive***	Negative***	Negative***	Negative***	Negative**
2009(1)	Negative***	Positive***	Negative***	Negative***	Negative***	Negative***
2009(2)	Negative***	Positive***		Negative***	Negative***	Negative***
2009(3)	Negative***			Negative***	Negative***	Negative***
2009(4) [Level Break]	Negative***			Negative***	Negative***	Negative***

Source: Authors’ calculations.

Note: All departures are ‘impulse’ (one-off) departures except in 2009 quarter 4 where the departure represents a shift in the underlying mean. Blank cells indicate no significant departure. The results relate to two models, each model has one focus variable and the entire set of control variables. Results of control variables are consistent across models. Note, two models were required to avoid problems of multicollinearity between income, consumption and savings. ***, ** & * denote significance at the 1 per cent, 5 per cent and 10 per cent levels respectively.

After the 2001 cash bonus, Consumption did not deviate from the trend; however, savings did. In 2001 Q3 savings was significantly more Positive than trend, but appears to have been offset immediately in 2001 Q4.

¹⁰ Model diagnostics indicate the model fits reasonably well. There does appear to be a hint of autocorrelation in for some of the results — however, we believe our conclusions to be robust with respect to its presence.

Turning our attention to the most recent episode of household bonuses (together with other Government/Federal Reserve stimuli) we observe a similar but more compelling picture of the ineffectiveness of cash bonuses (and other related macroeconomic measures). In particular, the results in Table 1 indicate that from the period of 2008–09 aggregate Consumption was statistically significantly below trend. Furthermore this ‘Consumption gap’ — the deviation from trend — has carried through to the end of 2012, as seen in Figure 2.

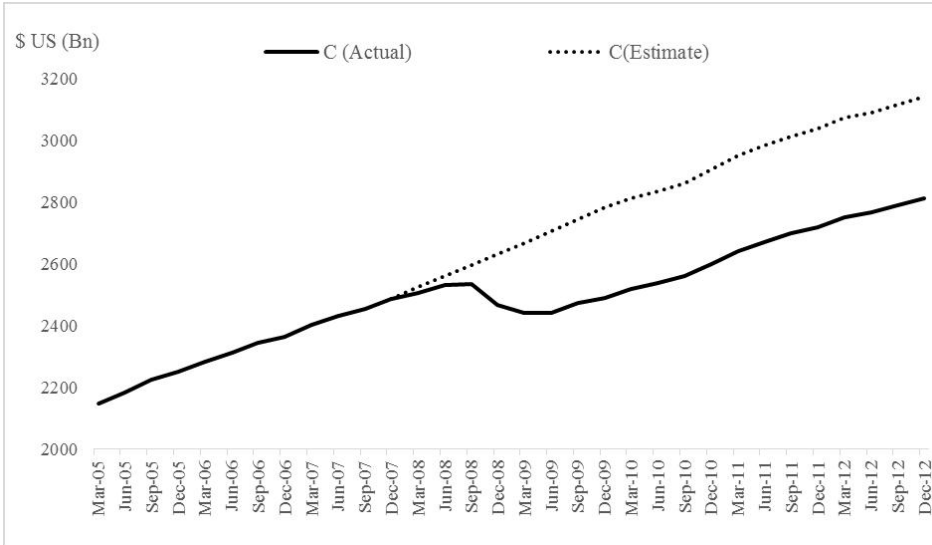


Figure 2: Estimated US consumption gap

Source: Authors' calculations.

In Figure 3, we present the estimated dollar value of the Consumption gap together with the increase in Savings. The increase in Savings corresponds to the cash handouts. The level of Savings, unlike Consumption, does not appear to permanently change in the aftermath of the GFC. Given the significant decreases in GDP, Dow Jones Industrial Index and disposable income, however, it is likely that the return in trend in savings is a result of a deteriorating economy where consumers, in the interests of maintaining a satisfactory lifestyle, are choosing (whether voluntarily or involuntarily) not to save over and above the trend. Our results indicate that households increased their savings in the beginning of 2008, thus suggesting that their precautionary behaviour preceded the more notable events such as the Lehman collapse.

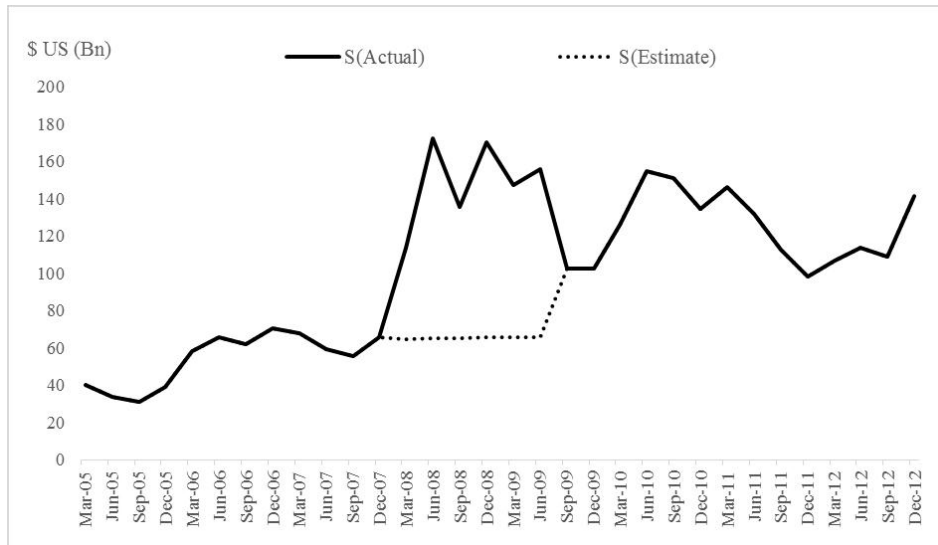


Figure 3: Estimated US consumption gap and savings gap

Source: Authors' calculations. This gap estimate is calculated from non-annualised rates less the statistically significant deviations.

The scale of the changes in Consumption and Saving are economically significant. The Savings gap, in aggregate is estimated to be \$500 billion. Using 2006 as the base year, this is equivalent two years' worth of savings. The Consumption gap is estimated to be approximately \$5 trillion, representing six months' worth of Consumption in 2006 terms.

As a means of verification we test the robustness of these estimates using an ordinary least squares regression. Specifically, we model consumption and savings (separately) as a function of a trend structural break in the intercept at 2008 Q1. It is important to appreciate that this 'benchmark' model is more rigid than the SUTSE specification as the time series components are now necessarily deterministic.¹¹ The gap is found to be statistically significant estimated to be seven trillion for consumption and 1.4 trillion for savings.¹² These estimates are consistent in character; however, we note that OLS model does not represent the data as well and thus conclude that the SUTSE estimates, the more conservative of the two, is more representative of what the US economy has experienced in recent times.

¹¹ The rigidity also means that the OLS estimates should be indicative only as there are patterns in the residuals, thus implying OLS assumptions do not hold.

¹² In the savings regression, a shift in the underlying mean (intercept) was found to be significant and permanent. This is in contrast to the SUTSE approach. Over the corresponding six quarters, 2008(1) to 2009(2), OLS and SUTSE estimates are approximately equal at \$500 billion.

In summary, the results in Table 1 and Figure 2 suggest that during both rounds of GFC stimulus US households chose to increase Savings rather than increase Consumption. This observation is consistent with the permanent income hypothesis.¹³

The Australian experience

Once again we focus on the cash handouts administered, of which there were two; one in 2008 Q4 and the second in 2009 Q2.

The analysis begins with a brief discussion of plots of the (household) Gross Disposable Income and Consumption series presented in Figure 4. These data are available on a quarterly basis from the Australian Bureau of Statistics (ABS).

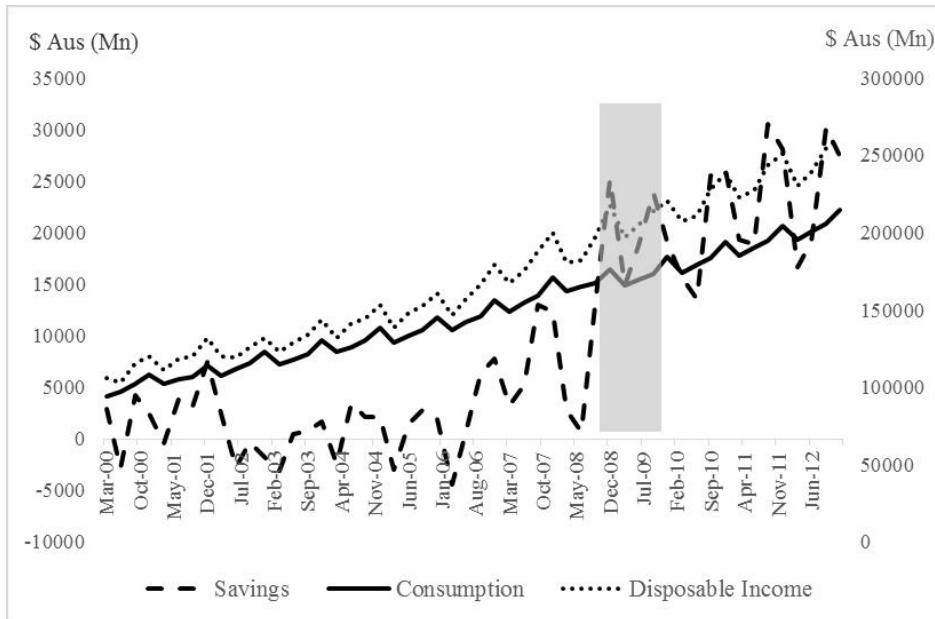


Figure 4: Australian macroeconomic data

Source: ABS National Accounts Cat 5206; left axis corresponds to savings only.

¹³ This phenomenon is consistent with the economic concept of precautionary savings. Lusardi (1998) and more recently Benito (2006) indicate that precautionary savings increase in times of employment uncertainty. This has been a notable characteristic of the US economy since 2008 Q1 (Chakrabarti *et al.* 2011).

In addition to displaying the original data series for Consumption, Income, and Savings,¹⁴ the chart also identifies periods in which two rounds of cash handouts were administered (2008 Q3 and Q4 and 2009 Q1 and Q2). The Consumption and Income plots do not seem to have deviated from underlying trend at the time the cash handouts were administered.

The apparent insensitivity of Consumption is an interesting phenomenon. In order to further investigate this we ran a third model, employing Food expenditure and Hotel expenditure to proxy non-discretionary and discretionary spending.

The data in this investigation spans 1985 Q1 to 2012 Q4.¹⁵ Table 2 presents the statistically significant deviations from trend.

Following Table 1, there are two panels in Table 2. The first panel identifies the main variables of interest. The results from three different models are presented in this table, these being Consumption and Savings, modelled separately with the entire set of control variables, and Hotel and Food expenditure modelled together with the control variables.

The first and fourth columns together with Figure 4 tell a similar story to the US experience; namely, that Savings increased at the same time that household cash bonuses were administered. The scale of the increase eclipses the US experience, in aggregate exceeding three times the level of Savings in 2006.

Australian households during and following the onset of the financial crisis have consumed less according to the model estimates. The scale of difference is on par with the US experience, representing 40 per cent of total consumption in 2006 terms. An interesting extension of this finding is to consider whether this has been uniform across consumption categories. The middle two columns of Table 2 panel 1 show that this is not the case. Noticeably, non-discretionary spending only departed from the trend once during the financial crisis, in aggregate representing an economically negligible 1 per cent in 2006 terms. Discretionary spending, however, seems to have permanently underperformed according to model estimates. In 2006 terms it represents a decrease of approximately 16 per cent.

Panel 2 of Tables 1 and 2 look remarkably similar, suggesting that both economies began experiencing a noticeable economic downturn prior to the administration of the cash bonuses and the stimulus packages. Importantly, the results in both tables also demonstrate that this downward course has not been reversed.

¹⁴ Savings is a residual term and includes down payment of debt etc.

¹⁵ Unlike the US data the Australian data are not seasonally adjusted. Savings is mean adjusted to address the mathematical problem of taking logs of negative values. Further there does appear to be a hint of autocorrelation for some of the results — however, we believe our conclusions to be robust with the respect to its presence.

Table 2: Significant departures from trend coinciding with household payments, Australia¹⁶

Date	Panel A: Focus Variables				Panel B: Control Variables			
	Consumption	Food	Hotel	Savings	Sentiment	GDP	Stock Index	Income
2008(1)					Negative **		Negative ***	Negative **
2008(2)	Negative **	Negative ***	Negative **	Negative **	Negative ***		Negative **	Negative ***
2008(3)	Negative ***		Negative **		Negative **		Negative ***	Negative ***
2008(4)	Negative ***		Negative ***	Positive *	Negative *		Negative ***	
2009(1)	Negative ***		Negative ***	Positive ***			Negative ***	
2009(2)	Negative **		Negative **	Positive *		Negative **	Negative ***	
2009(3)	Negative ***		Negative ***			Negative **	Negative **	Negative *
2009(4) [LB]	Negative ***		Negative **			Negative *	Negative **	Negative **

Source: Authors' calculations.

Note: All departures are 'impulse' (one-off) departures except in 2009 Q4 where the departure represents a shift in the underlying trend. Blank cells indicate no significant departure. The results relate to three models, each with one focus variable and the entire set of control variables. Results of control variables are consistent across models. Note, three models were required to avoid problems of multicollinearity between income, consumption and savings. ***, ** & * denote significance at the 1 per cent, 5 per cent and 10 per cent levels respectively.

¹⁶ Sentiment is the Westpac Consumer Sentiment index. The Stock index is the All Ords sourced from Yahoo finance. The remaining variables were sourced from the ABS.

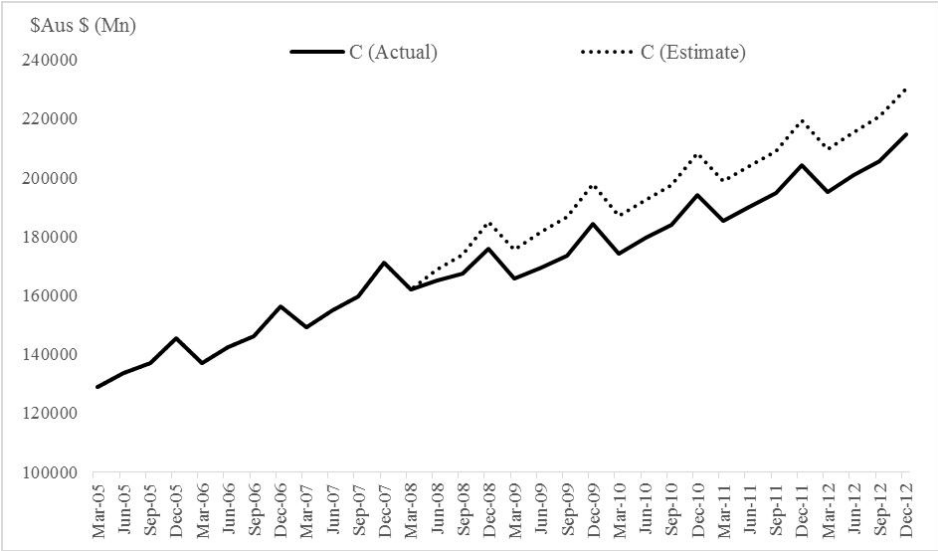


Figure 5: Estimated consumption gap (Australia)

Source: Authors' calculations. This gap estimate is calculated as the actual less the statistically significant deviations.

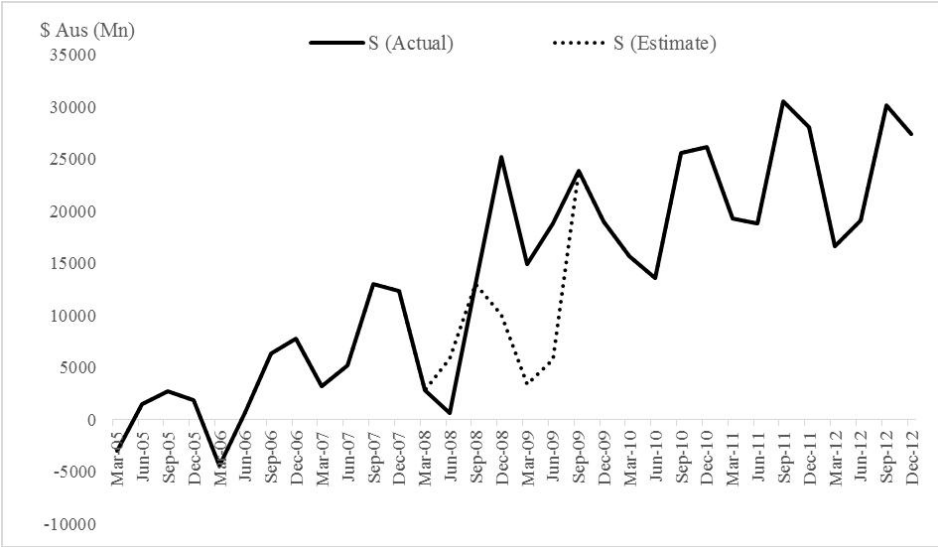


Figure 6: Estimated savings gap (Australia)

Source: Authors' calculations. This gap estimate is calculated as the actual less the statistically significant deviations.

Further, the results in Table 2 and Figures 5 and 6 indicate that Australian households have also behaved consistently with the permanent income hypothesis; that is, they did not increase consumption upon receiving the cash bonuses.

As before, we examine the veracity of these estimates by comparing them to the results of a comparable standard linear regression approximation.¹⁷ In the case of the consumption, savings and hotel expenditure the estimates are consistent in direction.¹⁸ We note that in the case of hotel expenditure the OLS estimated gap is considerably larger by a ratio of approximately five (7 billion versus 34 billion). The consumption (240 billion and 165 billion) and savings (35 billion and 57 billion) gaps are broadly consistent for SUTSE and OLS over the corresponding periods respectively. Once again, we consider the gap estimates of SUTSE to be a better representation of the underlying consumption and savings phenomena.

Conclusions

At the start of this paper three research questions were stated.

Is there any evidence that household spending significantly increased as a result of the cash handouts? Our results suggest there is no evidence that aggregate spending increased. In the US case an economic downturn was not avoided. Our modelling suggests a growing consumption gap consistent with the policy of cash handouts being ineffective.

While Australia did not experience a downturn, the response to the cash handout is very similar. Many advocates of the Australian stimulus package claim this proves that the stimulus package was a success. Our findings, however, suggest otherwise; specifically, Consumption did not increase.

Is there any evidence that the cash handouts (or proportion thereof) was saved? In the case of both economies there is clear evidence that Savings increased. The scale of the 'additional' savings for each economy is significantly large, suggesting that consumers behave cautiously in periods of uncertainty. This is consistent with documented international experience (Nielsen 2010 and IMF 2010).

Why has the stimulus (cash handouts) appeared to have worked in Australia but not in the US? In short, the appearance is a deception. The evidence suggests that in neither economy did the cash handouts increase spending (even after controlling for economic conditions). Further, savings in both economies dramatically increased. Therefore, there must be another reason for the resilience of the Australian economy.

¹⁷ OLS regression is fitted using a trend and quarterly dummies. A level break is fitted started in 2008(2) for all series. For all series this was found to be significant at the 5 per cent level. We do note that the OLS model diagnostics do indicate the model is not a good representation of the dynamics of the variables.

¹⁸ In the case of Food, a significantly positive shift in expenditure was found to occur.

In short, our findings suggest both Australian and US consumers acted consistently with the permanent income hypothesis during the time of the crisis. That is, they did not change (increase) their spending behaviour in response to the one-off government handouts.

This is an important conclusion as the results are consistent despite the different economic structure and conditions prior to the onset of the crisis. This suggests that cash bonuses, at least, are ineffective at stimulating the economy.

Limitations to our findings include the fact that the models fitted do not perfectly fit the data. Further, given the sequential nature of break testing, the power of the tests may be compromised. In addition, the a-theoretic approach taken means that we have not explicitly tested the permanent income hypothesis. Despite these limitations we believe our results are robust. Importantly, they are consistent with the conclusions reached in other investigations, including Taylor (2011).

References

- Agarwal, S., Liu, C. and Souleles, N. 2007, 'The Reaction of Consumer Spending and Debt to Tax Rebates — Evidence from Consumer Credit Data', *Journal of Political Economy* 115: 986–1019.
- Aisbett, E., Brueckner, M., Steinhäuser, R. and Wilcox, R. 2012, 'Fiscal Stimulus and Household Consumption: Evidence from the 2009 Australian Nation Building and Jobs Plan' (mimeo).
- Aoki, M. 1987, *State Space Modeling of Time Series*, New York: Springer Verlag.
- Auerbach, A., Gale, W. and Harris, B. 2010, 'Activist Fiscal Policy', *Journal of Economic Perspectives* 24(4): 141–164.
- Barro, R. 1987, 'Are Government Bonds Net Wealth?', *Journal of Political Economy* 82: 1095–117.
- Benito, A. 2006, 'Does Job insecurity affect household consumption?', *Oxford Economic Papers* 58: 157–181.
- Broda, C. and Parker, J. 2008, 'The Impact of the 2008 Tax Rebate on Consumer Spending: Preliminary Evidence', mimeo University of Chicago, GSB, July.
- 2012, 'The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption' (mimeo, October).

- Chakrabarti, R., Lee, D., Klaauw, W. and Zafar, B. 2011, 'Household Debt and saving during the 2007 recession', National Bureau of Economic Research Working Paper 16999 at: <http://www.nber.org/papers/16999>.
- de Silva, A., Hyndman, R. and Snyder, R. 2009, 'A multivariate innovations state space Beveridge Nelson decomposition', *Economic Modelling* 26: 1067–74.
- Drautzburg, T. and Uhlig, H. 2011, 'Fiscal Stimulus and Distortionary Taxation', Centre for European Economic Research Discussion Paper No. 11-037; at: <http://ftp.zew.de/pub/zew-docs/dp/dp11037.pdf>.
- Durbin, J. and Koopman, S. 2001, *State Space Modelling*, Oxford University Press.
- Engle, R. F. and Watson, M. W. 1987, 'The Kalman Filter: Applications to Forecasting and Rational-Expectations Models' in Bewley, T. F. (ed.), *Advances in Econometrics*, Fifth World Congress.
- Hall, R. 1978, 'Stochastic implications of the life-cycle permanent income hypothesis', *Journal of Political Economy* 86(6): 971–87.
- Harris, M. N., Loundes, J. and Webster, E. 2002, 'Determinants of Household Saving in Australia', *Economic Record* 78: 207–23.
- Harvey, A. 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- Harvey, A. and Jaeger, A. 1993, 'Detrending, stylised facts and the business cycle', *Journal of Applied Econometrics* 8: 231–47.
- Harvey, A. and Koopman, S. 1997, 'Multivariate Structural Time Series Model' in Heij, C., Schumacher, H. and Hanzon, B. (eds.), *System Dynamics in Economic and Financial Models*, John Wiley and Sons: 269–96.
- IMF 2010, 'Crisis Alters Pattern of U.S. Consumption' at: <http://www.imf.org/external/pubs/ft/survey/so/2010/res020110a.htm>.
- Johnson, D., Parker, J. and Souleles, N. 2006, 'Household Expenditure and the Income Tax Rebates of 2001', *American Economic Review* 96: 1589–610.
- Koopman, S., Harvey, A., Doornik, J. and Shephard, N. 2009, *Structural Time Series Analyser, Modeller and Predictor (STAMP 8.2)*, Timberlake Consultants.
- Leigh, A. 2009, 'How Much Did the 2009 Fiscal Stimulus Boost Spending? Evidence from a Household Survey', unpublished paper, Research School of Social Sciences Australian National University, at: <http://econrssh.anu.edu.au/~aleigh/>.

- Lusardi, A. 1998, 'On the Importance of the Precautionary Saving Motive', *The American Economic Review* 88(2): 449–53, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association.
- Makin, A. and Narayan, P. 2011, 'How Potent is Fiscal Policy in Australia', *Economic Papers* 30(3): 377–85.
- Nielson Global 2010, *Online Survey 2007–2009*, Consumer Confidence.
- Olekalns, N. 1997, 'Has Financial Deregulation Revived the Permanent Income/ Life Cycle Hypothesis?', *Australian Economic Review* 30(2): 155–66.
- Panher, G. S. 2007, 'Modelling and controlling monetary and economic identities with constrained state space models', *International Statistical Review* 75: 150–69.
- Proietti, T., Musso, A. and Westermann, T. 2007, 'Estimating potential output and the output gap for the Euro area: a model-based production function approach', *Empirical Economics* 33: 85–113.
- Shapiro, M. and Slemrod, J. 2003, 'Consumer Response to Tax Rebates', *American Economic Review* 93(1): 381–96.
- Sims, Christopher A. 1980, 'Macroeconomics and Reality', *Econometrica* 48: 1–49.
- Taylor, J. 2011, 'An Empirical Analysis of the Revival of Fiscal Activism in the 2000s', *Journal of Economic Literature* 49(3): 686–702 at: <http://www.aeaweb.org/articles.php?doi=10.1257/jel.49.3.686>.
- Taylor, L. and Uren, D. 2010, *Shitstorm: Inside Labor's Darkest Days*, Melbourne University Press, Melbourne.
- Uren, D. 2009, 'Cash handouts "likely to be spent, not saved"', *The Australian*, 10 February.

Appendix 1: Seemingly unrelated time series equations

The SUTSE specification used in this investigation has the form:

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \Phi D + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t$$

$$\gamma_{j,t} = \cos \lambda_j \gamma_{j,t-1} + \sin \lambda_j \gamma_{j,t-1}^* + \omega_{j,t-1}$$

$$\gamma_{j,t}^* = -\sin \lambda_j \gamma_{j,t-1} + \cos \lambda_j \gamma_{j,t-1}^* + \omega_{j,t-1}^*$$

The bolded characters denote N -vectors, where N denotes the number of series. In particular, the term \mathbf{y}_t denotes a vector of observations at time t . The terms, $\boldsymbol{\mu}_t$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\beta}_t$ denote vectors of trends, seasonals and growth rates at time t . The association between the series is captured by the off-diagonal elements of the various $(N \times N)$ $\boldsymbol{\Sigma}$ matrices.

The first equation is often referred to as the observation equation. All the other equations are typically denoted as component equations. The formulation presented above is often referred to as the Basic Structural Model (Harvey 1989). It depicts the trend as being equivalent to a Vector Auto-Regressive Integrated Moving Average (0,2,2). The formulation is easily extended to include a cyclical component. For more information refer to Harvey (1989).

ARGUMENT

Is Policy Too Important to be Left to Empiricists? Lessons of the 2012 Nobel Prize in Economics

Richard Cornes and José A. Rodrigues-Neto¹

Abstract

Fifty years ago, a paper entitled ‘College Admissions and the Stability of Marriage’ was published in a somewhat obscure journal, the American Mathematical Monthly (currently a ‘B’ journal, according to the Australian Business Deans Council). The research program and policy developments that have flowed from that abstract and apparently slight seven-page paper recently led to the award of the 2012 Nobel Prize for Economics to one of its authors, Lloyd Shapley. (Shapley’s co-author, David Gale, died in 2008.) Shapley shared the Nobel Prize ‘for the theory of stable allocations and the practice of market design’ with US economist Alvin Roth, who has been responsible for much of the applied work that has built on Gale and Shapley’s insights. The history of the path leading from the abstract Gale/Shapley insights to the design of resource allocation mechanisms in 2012 is a fascinating and instructive one for many reasons. This article tries to give the reader an idea of what this literature is about, and of the many ways in which Matching Theory has led to real improvements in the design of operational resource-allocation mechanisms.

Historical background

Gale and Shapley were interested in situations in which a group of agents want to be matched with one another. The agents may be prospective marriage partners, firms and workers, schools and schoolchildren, universities and students, hospitals and junior doctors, and so on. Of course, institutions and individuals faced the need to be matched to one another long before the Gale/Shapley contribution. Decentralised procedures for college admission evolved in which colleges would independently solicit applications from students, and make offers to those considered most acceptable. This arrangement has several

¹ The Research School of Economics, The Australian National University; Richard.Cornes@anu.edu.au. We would like to thank the editor and referees for detailed comments. All remaining errors are our own responsibility.

drawbacks, of which one in particular regularly arose. If one institution wants to 'steal a march' on its competitors and attract the best students, it may be tempted to do so by getting in with an earlier offer, and trying to obtain binding acceptances from students. Thus, over time, one might expect to see institutions gradually making earlier and earlier offers. This is indeed what happened — by 1945, some hospitals in the US were making offers to medical students up to two years before they had graduated from medical school! Such unravelling of the market, as this process has come to be called, is a common form of market failure, and generates substantial inefficiency. Although each institution may correctly perceive it to be in its interest to move earlier than its competitors, the outcome is that each thereby forgoes any useful information about that student's quality that may emerge during those subsequent two years — information which, had it been available, might have led to an outcome in which some students and/or institutions obtain more desired matches. Ignoring such information may well leave all agents — both students and hospitals — worse-off.

In response to such inefficiencies, in time, clearinghouses evolved. Under such an arrangement, each institution may give up the right to make independent offers, instead committing itself to working through a central agency — the clearinghouse — which operates under prescribed rules and procedures concerning the submission of preferences over institutions by prospective students and the identification of acceptable students by institutions. Eventually a matching of students and colleges is achieved. Central clearinghouses evolved in various settings in a somewhat ad hoc way — one of the authors of the present discussion once applied for an undergraduate place in the UK through precisely such a clearing house, then known as UCCA (Universities Central Council on Admissions). Some seemed to work quite well, others less well, and some were quite short-lived. But at that time there was no clear understanding of what, precisely, might constitute a satisfactory clearinghouse, why some seemed to work well while others did not, and how one might go about analysing, let alone designing, effective clearinghouses.

Gale and Shapley changed that. They established an approach that has provided the accepted framework for the analysis of clearinghouses that involve no explicit money transfers — in this sense they are quite distinct from the typical commodity market — but which seek simply to attain a satisfactory matching of institutions and individuals. First, they stated clearly what one might mean by a 'satisfactory' matching. To this end, they introduced the idea of a 'stable matching'. Second, they showed that under certain conditions a satisfactory matching would certainly exist. Finally, and perhaps most importantly, their demonstration of the existence of such satisfactory mechanisms took the form of a simple algorithm that actually found such an outcome. Their proposed algorithm has many attractive properties.

A statement of the college admission problem

Suppose there are m students applying for college admission and n colleges available, each one of them offering some positions/places for new students. (Typically, m and n are different positive integer numbers.) Each student has a complete ranking over the colleges, based on her individual views and goals. Similarly, each college has a complete ranking over the new students. This may be based on information provided by students about academic attainments, but may well be influenced by other considerations — different colleges usually come up with different rankings over the students. What matters is that this ‘market’ has two sides made of indivisible units: students and positions. Each member, on each side of the market, has her individual complete ranking of agents on the other side. Ties are not allowed.

A matching is a list of which college each student will attend. This list may say that a particular student does not have a college; or it may say that a college does not have all (or any) of its positions filled. In other words, a matching is a list of matches between students and colleges. We permit each college’s list of students to exclude some students. That is, each college is allowed to deem some students unacceptable, meaning that the college prefers to stay with some unfilled posts rather than admitting what it believes to be unacceptable students. Similarly, each student may find some colleges unacceptable. We say that an ‘agent’ (a student or a college) *blocks* a proposed allocation if in the proposed matching the agent is to be matched to an unacceptable pair, according to her preferences.

In this situation, is there a natural matching/allocation? That is, what would be natural or efficient or fair outcomes? How should we match/assign students to academic places? Some outcomes could be very favourable to some or even to all students, while other outcomes would be better for colleges. How should we combine the preferences of all agents in both sides of the market? Depending on how a mechanism does this, there could be incentives for misrepresentations of preferences; that is, for strategic manipulations. How could we prevent agents from ‘gaming the system’?

An interesting consideration is that, on each side of the market, agents have idiosyncratic preferences. There may be some correlation between preferences or it may be the case that each agent ranks the other side of the market in a completely different way. There is not a unique ‘correct’ ranking.

Gale and Shapley argued that a desirable property for an allocation is that it would be not possible to find a student i and a college j such that i prefers j to her prescribed college and, at the same time, college j prefers i to at least one of its prescribed students. In other words, no pair of agents blocks the proposed/

prescribed matching. They defined a matching to be ‘stable’ if it is not blocked by any pair of agents and, at the same time, the individual rationality condition of every agent is satisfied; that is, no single agent blocks the proposed allocation.

A special case: The Stable Marriage Problem

The easiest of the original problems posed by Gale and Shapley is the ‘Stable Marriage Problem’. This is a particular case of the College Admissions Problem where each college can admit only a unique student. In other words, the matching allocation is one-to-one. Each agent on each side of the market matches with at most one agent on the opposite side of the market.

The original paper refers to one side of the market as ‘females’ while the other side would be composed of ‘males’. As before, each female ranks all males up to a point, considering the remaining males as unacceptable (she would rather be alone rather than matched to any unacceptable male), and, likewise, each male identifies and ranks the females that he considers acceptable.

Gale and Shapley’s contribution

We have already drawn attention to the first of Gale and Shapley’s contributions — their statement of stability as a desirable characteristic of a matching outcome. In addition they suggested an algorithm that allows us to find stable outcomes in the Stable Marriage Problem. This is the famous Deferred Acceptance Algorithm. A variation of this algorithm generates a stable matching in the more complex Stable College Admissions Problem.

They proposed the use of a centralised mechanism for discrete resource allocations. This might explain the lack of interest by US economists in the 1960s and 70s. The efficiency of free markets was a cornerstone of economics and the Arrow-Debreu Theorem had just made great academic impact.

The set of all stable allocations — that is, allocations such that no subset of agents can collude to make all agents in this subset better off — corresponds with the notion of the core, a concept that was systematically developed during the 1950s and 60s within the research program on General Equilibrium. In any coalitional game, an allocation is in the core if there is no blocking coalition of any size. In the matching context, for a matching to be stable, there can be no unilateral deviation or any blocking coalition of two players. If an allocation is in the core, then, trivially, it is stable. For the Stable Marriage Problem, the reciprocal statement is also true. Why is this true? Well, if there is an improving

coalition with more than two agents, there must be a subset of agents in this coalition with one or two players that would prefer to deviate from the original allocation. If an allocation is stable, there is no such subset. Hence, this allocation must be in the core.

The Deferred Acceptance Algorithm (DAA) is remarkable in many ways. First, it proves the existence of at least one stable matching.

Second, the algorithm is reasonably simple and intuitive. While current economists may find it a very desirable and convenient feature, this is somewhat surprising for a paper written by two mathematicians 50 years ago. Most papers in mathematics at that time, and even nowadays, propose very complex structures and algorithms. By contrast, Gale and Shapley's rigorous demonstration of the existence of a stable outcome — to which we will return later — contains, as they point out (p.15), '...no obscure or technical terms. Knowledge of calculus is not presupposed. In fact, one hardly needs to know how to count.'

Third, the DAA runs in finite time and is relatively easy to a program in a computer. Gale and Shapley proved that their procedure has at most $n^2 - 2n + 2$ stages, where n is the number of agents on each side of the market.

Fourth, usually there are *multiple* stable allocations for any given instance of the Stable Marriage Problem. Depending on how we set the algorithm to run, it can generate the allocation that is best, among all stable ones, for males or the best allocation for all females. In other words, if we are restricted to the collection of stable matching allocations only, then all males would never disagree in their ranking of any two proposals. The same is true for females. Moreover, in the set of all stable matching allocations, the 'male best' stable matching is exactly the 'female worst' matching and vice-versa. According to mathematicians, the set of all stable allocations form what they call a 'distributive lattice', a very useful but somewhat unknown mathematical concept. Analogous considerations hold for the Stable College Admissions Problem. Topkis (1978) has more on lattices and optimisation on such domains.

Fifth, even though the algorithm itself may involve arbitrary choices, the set of all stable matching allocations does not depend on chance. Although chance does play a role sometimes, nobody can complain too much of 'bad luck'. If an agent is not matched by the Deferred Acceptance Algorithm to some other agent ranking relatively high in her preference list, it is likely that she is simply not ranked highly by any one of her most favoured matches. This property may be an important feature in persuading the public to accept the implementation of a central mechanism.

The Deferred Acceptance Algorithm

Consider an instance of the Stable Marriage Problem. We will describe a version of the Deferred Acceptance Algorithm (DAA) in which men propose to women. The algorithm has a sequence of such proposals. All agents start 'free'. Men can alternate between being free or 'engaged' several times during the process, but women cannot become free after they are engaged. Women who are free always accept an engagement proposal of any acceptable man, but will turn this man down if she ever receives a proposal from a more desirable agent. Engaged women reject proposals from men that are less desirable than her current engaged partner. Men who are free keep making proposals, but engaged men do not.

The version of the DAA in which men are the proposers works as follows:

- All agents start free.
- The following procedure is repeated as long as any man is free; when no free men remain, the algorithm terminates.
- The men in a randomly chosen order make proposals.
- The current man makes a proposal to the most desirable woman in his list who has not already rejected him.
- If this woman is free, then she and her proposer become engaged.
- Otherwise, if she prefers the proposer to her current partner, she becomes disengaged from her existing partner — who thereby becomes free — and becomes engaged to the current proposer; otherwise, she rejects the proposer.
- This procedure is repeated for the next man.
- As already stated, the algorithm ends when no free men remain.

When the algorithm ends, the couples who are engaged at that stage become matched.

It turns out that the order in which men make their proposals does not affect the outcome of the algorithm. In the process, each man starts with highly desirable partners, but as the algorithm proceeds, he can only go down in his preference list. However, each woman can only move up in her preference list, as she receives proposals from a range of different men, but switches partners only if it is in her advantage to do so (she rejects all proposals from men who are less desirable than her current engagement). Because both the set of men and the set of women are finite, the algorithm always ends in a finite number of steps.

We urge curious readers to indulge themselves by reading the original proof of the existence of a stable matching in Gale and Shapley (1962: 11–12). The following argument captures its main insight. Suppose that man X_1 is matched to woman Y_1 , but prefers woman Y_2 to Y_1 . During the DAA, man X_1 proposes

to woman Y_2 before he proposes to woman Y_1 . The reason he is not matched to Y_2 is *either* because woman Y_2 has rejected X_1 's proposal (since she had a better partner) *or* because woman Y_2 accepted the proposal by man X_1 , but afterwards obtained a proposal from a more desirable man. In either case, Y_2 ended up with a man that she considers better than man X_1 .

Consider the following example of the men-proposing version of DAA: there are three men (X_1, X_2, X_3), and three women (Y_1, Y_2, Y_3). Preferences are as follows: the last option of each agent is to end up alone. Among members of the opposite sex, preferences are:

$X_1: Y_1, Y_2, Y_3$ (meaning that man X_1 prefers woman Y_1 to Y_2 , and prefers Y_2 to Y_3);

$X_2: Y_2, Y_3, Y_1$;

$X_3: Y_1, Y_2, Y_3$;

$Y_1: X_2, X_3, X_1$;

$Y_2: X_1, X_2, X_3$;

$Y_3: X_3, X_1, X_2$.

Suppose men make proposals in the order: X_1 first, then X_2 , and finally X_3 . In this case, man X_1 proposes to woman Y_1 , who accepts. Man X_2 proposes to woman Y_2 , who accepts. Man X_3 proposes to woman Y_1 , who accepts (because woman Y_1 prefers X_3 to X_1), making man X_1 free again. Man X_1 proposes to woman Y_2 who accepts (because woman Y_2 prefers X_1 to X_2), making man X_2 to become free again. Man X_2 proposes to woman Y_3 who accepts. The resulting matching is: (X_1, Y_2) , (X_2, Y_3) , and (X_3, Y_1) . The theory tells us that this matching is a stable matching in this particular example.

But there are multiple stable matching outcomes. The two versions — distinguished by whether men or women are the proposers — lead to different stable outcomes. For instance, if women make proposals in the order Y_1 first, then Y_2 , and finally Y_3 , the DAA proceeds as follows: woman Y_1 proposes to man X_2 , who accepts. Woman Y_2 proposes to man X_1 , who accepts. Woman Y_3 proposes to man X_3 , who accepts. The resulting matching is: (X_1, Y_2) , (X_2, Y_1) , and (X_3, Y_3) . This is also a stable matching.

Man X_1 and woman Y_2 are matched in both the man-proposing and woman-proposing versions. However, the man-proposing version of the DAA is strictly preferable to the woman-proposing version for men X_2 and X_3 . Indeed, man X_2 prefers to be matched with woman Y_3 (his second-best option) than with Y_1 , which is his third-best option. Man X_3 prefers Y_1 (his first option) to woman Y_3 , which is only his third-best option. On the other hand, women Y_1 and Y_3 strictly

prefer their partners under the woman-proposing version to those under the man-proposing version of the DAA. For instance, woman Y_1 prefers man X_2 rather than man X_3 .

Later developments by Roth and others

The National Resident Matching Program (NRMP) is a North American NGO that matches medical school students with residency positions that are available in any given year. The NRMP started to operate in 1952 and is sponsored by several medical associations. Before that, there was keen competition among hospitals for the best talent. Hospitals would make job offers independently of one another. Around 1944 individuals received offers up to two full years before their graduation. Agreement on the need for some kind of clearinghouse, or centralized market, was reached around 1950.²

Matches occurring two years before students complete their training tend to suggest that students lack incentive to study really hard during their last two years. As a consequence, they probably become less competent doctors. This is a clear source of inefficiency. Because many of the best positions are taken two years before they graduate, unmatched medical students also find fewer reasons to study hard in their last two years. Hence, matching a group of medical students two years before graduation may severely diminish the incentives of all students — both earlier matched and unmatched — to study hard.

Of course, two full years before graduation the information available regarding future doctors' skills is less precise than what we would obtain if matches were made at graduation time. This was another significant source of inefficiency.

Up until 1995, the version of the Deferred Acceptance Algorithm used by the NRMP always generated the matching assignment that was the best for hospitals, among the subset of all stable allocations, and worst for doctors. Since 1998 the revised NRMP algorithm finds the doctor-optimal stable matching. Interestingly, only a small proportion of the assigned doctor-hospital pairs would change, whether we run the previous or the current algorithm. This is a positive feature of the procedure. Later, the NRMP started to allow for the possibility that couples would submit joint preference lists. This makes the assignment problem significantly harder.

The unravelling of the market has been particularly acute in the process by which US federal court judges hire federal judicial law clerks. Avery *et al.* (2001)

2 Roth and Peranson (1999: footnote 5) claim that the '...initial NRMP algorithm, dating from 1951, [was] the first we know of'. For more details on the history of the DAA and NRMP, see Roth (1984), Gusfield and Irving (1989) or Roth and Sotomayor (1992).

note that, at the time when many prominent law schools in 1999 abandoned their attempts to regulate the time at which letters of recommendation could be sent, the clerk-hiring process got under way two years before the clerkship positions themselves were due to begin. The behaviour of judges in their attempts to jump the gun on their rivals in making early offers and seeking binding acceptances in order to attract the best candidates, led the *New York Times* to refer to the process as a judicial ‘free-for-all’ in which judges ‘behave[ed] like 6-year-olds’. One judge likened the process to a ‘calf scramble ... the low point of many western rodeos’ — not an image that most of us have of the judiciary!

Some significant characteristics of this piece of intellectual history

In the original formulation, the Stable Marriage Problem assumes that both sides of the market have n agents. It turns out that this assumption is not terribly important. The Deferred Acceptance Algorithm runs unchanged when the numbers of agents on each side of the market are distinct/different from each other.

In reality, a large part of the work done by Alvin Roth and his co-authors deals with institutional particularities. Yet the central ideas of Matching Theory are always present and the researchers continue to create and develop new abstract ideas, concepts, algorithms and mathematical techniques that will, in the future, become the basic knowledge upon which concrete, applicable real-world solutions will be developed.

The lesson is two-fold: first we learn from this history that one must be willing to adapt abstract frameworks to tackle real-life problems — what Roth refers to as ‘social engineering’. Second, the search for pure, abstract algorithms and formal mathematical results in idealised environments is crucially important. Abstract models have provided critically important changes in paradigm. They can shed light on difficult problems and indicate directions where possible solutions lie and where there is need for more knowledge and more pure or applied research. Too frequently governments have a myopic focus. However, the power of very abstract ideas is enormous.

Investment in abstract modelling may save public funding because the same central principles can typically be used in a wide variety of applications. In short, a single investment may generate many fruits, if we are prepared to adapt it to the idiosyncratic features of each institutional framework.

The award of this particular Nobel Prize provides a good opportunity for reflection. Where are our tax dollars best spent — in very concrete yet somewhat limited research of clear immediate practical use, or in more abstract research with no immediately apparent connection to reality, but deep scientific significance? For instance, would it be better to invest one dollar with the certainty of collecting two, or to invest one dollar with a one per cent probability of collecting a thousand?

We would argue that research in economic theory has, on average, more of a ‘public-good nature’ than a good deal of applied research. Very applied research is designed — by its very nature — to deal with a very particular case. Lessons may sometimes be transported to some other situations, but typically on a rather small scale. By contrast, the ideas in Gale and Shapley’s analysis of matching are non-rival — their use in one setting does not limit their use in others. Cornes and Sandler (1996) provide an extended discussion of this non-rivalry property in the context of public-good theory, and Romer (1990) discusses its crucial importance in the context of technological change and welfare. A matching algorithm can be used time and time again — to match medical interns with hospitals, students with schools, law clerks with law courts or legal firms. Further, once it is developed, its extension and adaptation for other uses is a relatively low-cost matter, as is evidenced in the developments from Gale and Shapley’s original insights up to the more complicated algorithms being developed today by Roth and many others.

The example of combinatorial analysis and auctions

A new important research trend inside auction theory is the study of combinatorial auctions; that is, procedures to auction bundles of heterogeneous goods, instead of a single product. The idea is that these different products may complement or substitute each other and these degrees of complementarities or substitutability may be idiosyncratic. For instance, suppose you want to sell a table and a set of chairs that fit the table. Is it better to hold two auctions, one for the table and another for the chairs, or to sell the entire bundle in a single auction?

Matching Theory and Auction Theory are the two faces of what current researchers call “Market Design”. Rassenti *et al.* (1982) use experimental analysis to study the design for airport time-slot auctions. They propose the use of a combinatorial auction mechanism. De Vries and Vohra (2003) provide a classic survey of combinatorial auctions. Their survey reveals that over the last 10 years there have been substantial theoretical advances on this front and

that many new applications have emerged in a wide variety of contexts — sales of many products and services such as furniture, airport time slots for airlines, delivery routes, and network routing are among the huge variety of situations in which combinatorial analysis is proving useful. More recently, Vohra (2011) has shown how combinatorial reasoning, in the form of linear programming, can be used to develop and understand modern mechanism design.

Other significant areas within economics that use the ideas of combinatorial analysis are Voting Theory and Matching Theory, as well as all problems related to the allocation of indivisible goods. However, for some traditional economists educated to work only with calculus, optimisation and differential methods with continuous variables, the language of combinatorial analysis represents a huge challenge. It is so because, unfortunately, discrete mathematics, even its elementary ideas and concepts, is no longer part of the typical university training in the field of economics.

The basic ideas of permutations and combinations are extremely helpful in the education of university students. Few concepts are as efficient to help students learn how to think analytically, in an organised way, as the basic counting and ordering principles of combinatorial analysis. Their presence in the syllabus helps to make students learn how to think and how to organise complex structures and ideas in their minds. Ironically, these virtues may partly explain their gradual disappearance from university economics syllabuses since the 1960s — under ever-increasing time and resource pressures, it may have been easier and quicker to teach alternative approaches that rely more heavily on a relatively uncritical application of ‘cookbook recipes’.

Current applications of Matching Theory

In the last two decades ever more applications for Gale and Shapley’s Matching Theory have been found. Although they rarely make front-page news in the manner of the latest macroeconomic or financial ‘crisis’, these applications are beginning to have a significant beneficial impact on the lives of millions of people — as resident doctors and patients with transplanted kidneys in the USA or parents of children benefiting from a better high school allocation in New York or Boston public schools system would tell you — and to change the world we live in.

Matching Theory is used today to allocate secondary students to public high schools in New York and Boston, two very large educational systems. Before entering as incoming freshmen, all students submit a preferred high school list,

and high schools set criteria that effectively rank all new students in the city system. This work was done by Alvin Roth together with Atila Abdulkadiroglu, Parag A. Pathak and Tayfun Sonmez (Abdulkadiroglu *et al.* 2005).

Roth (2002) brings quantitative evidence about observed benefits of stability in matching mechanisms used in practice, focusing on different regional markets for new physicians and surgeons in the United Kingdom. Roth's paper describes quantitative evidence found in previous papers and displays (Table 1: 1354) the average costs of markets. Among markets with a clearinghouse, one group indicates similar costs between the DAA and the Newcastle Priority Algorithm — an alternative algorithm that does not take the Gale and Shapley concept of stability into account. However, in two other cases, the costs associated with the DAA are significantly lower than the cost associated to the Newcastle Priority Algorithm.

An unexpected benefit of using Matching Theory in practice was to decrease corruption in some cases. As the centralised mechanism was implemented in school districts, local authorities lost their power to accept or reject students to high schools. To the surprise of some, extra positions became available in many schools.

Matching Theory has had a dramatic effect on the allocation of high school students. Every year, in New York City School District alone, there are around 100 000 new high school students entering the system. All of their families provide a list of preferences and the schools provide their criteria to rank the entrant students. Obtaining a better allocation of school leaves thousands of families better off and their kids will probably learn more and become more productive workers. As economists know well, enhancing human capital leads to a permanent increment in productivity.

Organ transplantation is another major beneficiary of Matching Theory. Typically, the waiting list for deceased donors is very long. Many patients die before a compatible organ is available. A partial solution to this problem is to increase the incentives for live donors. Often a patient can find a living donor, usually a spouse or sibling. However, rarely is the donor organ compatible with the patient's biology. In this context, Matching Theory and Market Design have played a crucial role in the design of the institutional rules that generate the best incentives and the largest possible number of successful organ donations. Indeed, it seems to have been precisely this application that drew the Nobel committee's attention to the work of Alvin Roth.

A feature of the market for organ donation is that it is illegal to buy or sell human organs in most countries, including all developed countries. Even if both parties agree on the terms, the law does not allow for such transactions.

(For example, in the US, the National Organ Transplant Act forbids contracts for organ transplantations.) Moreover, organs are indivisible. This is exactly the framework of Matching Theory, and indeed much has been done recently.

Alvin Roth, together with Itai Ashlagi, David Gamarnik, Utku Unver and Tayfun Sonmez, and with the support from the US National Science Foundation, developed the ideas, procedures and software needed to match living kidney donors with recipients. Their research and ideas from idealised abstract models are achieving what had earlier seemed impossible: that a single extra donation can help to spark a long chain of donations, saving many lives directly and decreasing the waiting times of patients in the waiting lists, which in turn, greatly increases the probability of success in organ transplantations. Also, reduced waiting lists improve dramatically the welfare of thousands of patients. It is no longer the case that one donor helps a single sick person.

The literature on two-sided matching uses elements of economic theory as well as empirical and experimental analysis, as explained by Roth (2002).

Conclusions

The history of Matching Theory teaches us several valuable lessons.

First, it could be dangerous to dismiss basic research as necessarily less important because of its lack of obvious immediate impact on pressing problems of the day. This is a crucial observation these days as policy circles discuss the introduction of policies that try to measure research impact.

Second, some of the most applicable and useful ideas often come along first in the most abstract forms. They require time to mature and to prove their true value; in this case a few decades. This does not fit with the politics of the democratic electoral cycle. Hence, support for deep research must be a national priority above and beyond the politics of the day. Funding for basic research should be allocated in a way that it becomes untouchable to political actors, both within and outside universities. Matching Theory is a great example — perhaps the best — of a line of research in economics which, initially, may have seemed like a piece of academic self-indulgence, yet ultimately having — as its authors surmised — real significance to real allocation issues. It is not unfair to ask how the Gale/Shapley piece would have been shoehorned into the straitjacket of national priorities.

Third, the power of abstract mathematical ideas should never be underestimated. Looking to social sciences or policy studies through the lens of what seems to be of immediate national significance will probably lead to myopic incentive

schemes that are detrimental to mathematics and to the creation of powerful abstract new ideas and concepts. At the end of the day, such short-sightedness may generate an academic sector producing less significant outcomes, most of which use, at best, simplistic and conventional ideas. This would be a poor path for our society to follow.

Fourth, Matching Theory is a great example of the breadth of current economic theory. It explores the ideas behind allocation of resources, and their implications for human wellbeing. Of course, topics related to the allocation of goods and agents are always interesting to economists, whatever the particular circumstances.

As John Maynard Keynes famously wrote in *Essays in Persuasion*: ‘If economists could manage to get themselves thought of as humble, competent people on a level with dentists that would be splendid.’

In this piece, we have drawn attention to economists who, through their association not with dentists but with school administrators, renal transplant surgeons and others, have greatly enhanced the wellbeing of many individuals. We suspect that many contemporary critics of economics, and indeed some of our colleagues within the discipline, are simply unaware of its involvement in the design of resource allocation mechanisms of the kind that we have described, and of their potential benefits.

Finally, Matching Theory is an area with proven positive results in terms of public policy. However, in Australia, high school graduates are not allocated to universities with the use of a centralised mechanism that uses the DAA. The implementation of such a scheme would improve access of students to universities and would also lead to better matches between the two sides of the market. There are plenty of other areas that could benefit if Australian policymakers decided to use Matching Theory to formulate and implement better public policies.

We leave the last word to the writer of the piece in *The Economist*³ which announced the award of the 2012 Nobel Prize: ‘The recognition of Mr Shapley’s and Mr Roth’s work is also a reminder: that for all the bad press economics has received since the [financial] crisis, the discipline still brims over with insights that can solve real-life problems.’

3 20 October 2012.

References

- Abdulkadiroglu, A., Pathak, P., Roth, A. E. and Sonmez, T. 2005, 'The Boston Public School Match', *The American Economic Review Papers and Proceedings* 95: 368–71.
- Avery, C., Jolls, C., Posner, R. A. and Roth, A. E. 2001, 'The market for federal Judicial law clerks', *University of Chicago Law Review* 68: 793–902.
- Cornes, R. C. and Sandler, T. 1996, *The Theory of Externalities, Public Goods and Club Goods* (2nd edition), Cambridge University Press.
- de Vries, S. and Vohra, R. 2003, 'Combinatorial auctions: A survey', *INFORMS Journal on Computing* 15(3): 284–309.
- Gale, D. and Shapley, L. S. 1962, 'College admissions and the stability of marriage', *American Mathematical Monthly* 69: 9–15.
- Gusfield, D. and Irving, R. W. 1989, *The Stable Marriage Problem: Structure and Algorithms*, Foundations of Computing Series, The MIT Press.
- Hajduková, J. 2004, *On Coalition formation games*, IM Preprints series A, No. 5/2004.
- Knuth, D. E. 1976, *Mariages Stables*. Le Presses de l'Université de Montréal, Montreal.
- O'Brien, M. and Kellan, A. 2012, *New Software Matches More Kidney Donations, Faster*, National Science Foundation. Available at: http://www.nsf.gov/news/special_reports/science_nation/kidneyexchange.jsp
- National Resident Matching Program. Information available at: <http://www.nrmp.org> and http://en.wikipedia.org/wiki/National_Resident_Matching_Program
- Rassenti, S. J., Smith, V. L. and Bulfin, R. L. 1982, 'A combinatorial auction mechanism for airport time slot allocation', *The Bell Journal of Economics* 13: 402–17.
- Romer, P. M. 1990, 'Endogenous Technical Change', *Journal of Political Economy* 98: S71–102.
- Roth, A. E. 1984, 'The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory', *Journal of Political Economy* 92: 991–1016.

- Roth, A. E. 2002, 'The economist as engineer: Game theory, experimentation, and computation as tools for design economics', *Econometrica* 70: 1341–78.
- Roth, A. E. and Sotomayor, M. A. O. 1992, *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis*, Cambridge University Press.
- Roth, A. E. and Peranson, E. 1999, 'The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design', *The American Economic Review*: 748–80.
- Topkis, D. M. 1978, 'Minimizing a submodular function on a lattice', *Operations Research* 26: 305–21.
- Vohra, R. V. 2011, *Mechanism Design: A Linear Programming Approach*, Cambridge University Press.

A Critique of the Productivity Commission's Cost-Benefit Analysis in the 'Disability Care and Support' Report

Mark Harrison¹

Abstract

In its 2011 NDIS report, the Productivity Commission rationalises its policy recommendation by means of a cost-benefit analysis, claiming that 'the benefits of the [National Disability Insurance] scheme would significantly outweigh the costs'. But methodology the PC adopts departs from conventional cost-benefit analysis in ways that understates costs, presumes the benefits, muddies policy comparisons, and jumbles equity and efficiency issues. These problems are traceable to the Commission's use of a 'distributional weights approach' to equity benefits. The 'basic needs approach' is an alternative way of dealing with equity considerations that better captures the underlying preferences of citizens and the rationale for disability care and support policies.

Introduction

In its 2011 'Disability Care and Support' report, the Productivity Commission (PC) recommended a major overhaul of the disability support system. The PC found the current State-based disability support system is 'underfunded, unfair, fragmented and inefficient' (PC 2011a: 2) and recommended replacing the State-based schemes with a new National Disability Insurance Scheme (NDIS). In March 2013, both major political parties supported legislation to establish the scheme and a 0.5 percentage point increase in the Medicare levy to help fund it. The scheme would provide support for Australians who are born with, or acquire, a severe or profound disability. The PC estimates 410 000 people would be on the scheme (ibid.).

The PC bolstered its policy recommendation with what it describes as a cost-benefit analysis. The PC acknowledges its cost-benefit analysis is not conventional,

¹ University of Wollongong; harrison@uow.edu.au.

for its analysis does not attempt the difficult task of a welfare *efficiency* analysis of the benefits from expenditure on the disabled, but focuses instead on equity or distributional effects. Benefits are measured using a distributional weights approach — dollars to people with disabilities have more social value than dollars to taxpayers. The methodology the PC adopts understates costs, presumes the benefits, muddies policy comparisons and jumbles equity and efficiency issues.

In cost-benefit analysis we should always ask: what is the essential benefit the project confers? Clearly any assessment of disability care and support policy options must evaluate the equity benefits. But in conventional cost-benefit analysis, equity and efficiency are strictly separated. Efficiency is all about the willingness to pay, and can, in principle, be estimated from observed behaviour using well-established economic techniques. In contrast, there is a lack of agreement on how to judge equity effects, which involve ethical or value judgments. In particular, there is no general consensus about the weights that should be attached to the welfare of different groups under the distributional weights approach. There is not even agreement on whether the distributional weights approach the PC adopts is the appropriate way to assess equity issues. I argue that although it is an approach favoured by economists, it is peculiar and does not represent the values of the general public. The basic-needs approach to evaluating equity effects is superior, especially for disability care and support policy.

The cost-benefit analysis

Chapter 20 of *Disability Care and Support* presents a cost-benefit analysis of the NDIS (PC 2011b). Unfortunately, it appeared for the first time in the final report and did not go through the scrutiny of the draft report hearings and consultation process. It constitutes an *ex post* rationalisation of the PC's recommendations, rather than being used to shape and inform policy formulation and compare options.

The PC presents the result of its cost-benefit analysis as: 'The benefits of the scheme would significantly outweigh the costs. ... The NDIS would only have to produce an annual gain of \$3800 per participant to meet a cost benefit test. Given the scope of the benefits, that test would be passed easily.' (PC 2011a: 2.)

This presentation is mysterious, and to the extent it can be fathomed, highly misleading. Their analysis assumes 410 000 participants and puts total expenditure on the NDIS at \$13.5 billion; an increase of \$6.5 billion over

current spending (PC 2011a: 3).² But benefits of \$3800 per participant come to \$1.56 billion, an amount less than 30 per cent of the increase in spending. How can \$1.56 billion 'pass' \$6.5 billion? The answer is that the PC puts the marginal excess burden (MEB)³ at a (conservative) 24 per cent, and so claims: 'That suggests that the approximate economic cost of funding the NDIS would be around $0.24 \times \$6.5$ billion or about \$1.56 billion.' (PC 2011b: 955.) Thus the only 'cost' of the PC's NDIS policy is the Marginal Excess Burden of the taxes needed to finance it. Thus if the MEB of taxation was zero, the PC would presumably claim the scheme had no economic cost at all, and was justified no matter how small the benefits.

Let's get it straight: an MEB of 24 per cent means an extra dollar of revenue to the government costs taxpayers \$1.24. That is, the proposed reform increases government expenditure by \$6.5 billion, which imposes a cost on taxpayers of \$8.06 billion. For the project to be justified, the benefit it produces must exceed the budgetary costs multiplied by one plus the marginal excess burden of taxation (that is, the marginal cost of funds).

Done properly, the NDIS must produce incremental benefits of \$19 658 per recipient ($=\8.06 billion/ $410\ 000$), not \$3800. Moreover, that is the increase in benefits compared with *current* arrangements, which may themselves be quite inadequate on a cost-benefit criterion. In the most comprehensive application of cost-benefit logic, not only must the incremental benefits of the reform exceed incremental costs, the *total* benefits of the NDIS must exceed the total costs for a project to pass a cost-benefit test.⁴ It must produce gross benefits of more than \$40 830 ($= \$13.5\text{b} \times 1.24 / 410\ 000 = \$16.74\text{b} / 410\ 000$) per recipient to justify the \$13.5 billion spent on it. That is, the NDIS must produce at least \$40 830 per recipient to pass a cost-benefit test, not \$3800.

The PC does recognise that 'the \$6.5 billion reduces the income (and therefore consumption) of one group of people — taxpayers in the general community and raises the consumption of people with disabilities and informal carers by the same dollar amount. In conventional cost-benefit analysis, taxpayers lose from such an exchange.' (PC 2011b: 955.)

2 The Australian Government Actuary (AGA) pointed out that the PC's cost estimates of \$13.5 billion per annum were for a full scheme implemented in 2009. Under the PC's plan, the scheme would not be fully operational until 2018. Accounting for inflation, wage rises (especially the Fair Work Australia decision to increase award wages for a large number of social and community sector workers from 1 December 2012) and population increases through to 2018, the AGA puts the number of recipients at 441 000 and the gross cost of the NDIS at \$22 billion in 2018–19, around a \$7.5-billion increase in spending in today's dollars (AGA 2013: 26).

3 The PC recognises that taxes distort behaviour and impose a marginal excess burden. When the government raises an extra dollar of revenue from taxpayers, the cost to taxpayers (marginal cost of funds) is greater than \$1. It is $1 +$ the marginal excess burden (MEB). Accounting for the cost of raising funds to finance a project means applying an extra charge or benefit equal to the marginal cost of funds to each and every cash outflow or cash inflow from and to the public treasury, over the life of the project (Harberger 2007: 6).

4 Ideally, the chosen scheme is that which maximises the excess of total benefits over total costs.

But the PC (2011B: 941, 971) assumes that:

The net economic cost of the NDIS is not the budgetary cost of around \$6.5 billion (which is a transfer of resources from one group to another). ... It is assumed that the dollar value of transfers to the NDIS participants are equivalent to an income transfer of the same amount.

So the PC recognises the budgetary cost is a cost to taxpayers, but gives an offsetting benefit to people with disabilities, which, in its mind, leaves its economic cost equal to the excess burden of the taxes needed to finance it.

Is the dollar paid to Paul worth the dollar taken from Peter?

What is wrong with the PC's approach?

To begin, the NDIS is not a simple cash transfer. It is spending \$13.5 billion on services for people with disabilities. Even when recipients are given more control over how services were planned and delivered — so called self-directed funding — the transfer must still be spent on disability services. The PC (2011b: 971) admits that 'in conventional economic analysis, the benefits to people of hypothecated payments are less than their equivalent income value' but lists reasons why the assumption it is valued at cost is reasonable for this particular policy. At best the methodology is a special case, not a general approach.

The issue is how much benefit people with disabilities get from that spending — the value they place on those services — which needs to be established, not assumed. It needs to be allowed that a portion of the spending goes on a bureaucracy to administer it, so not all of the money makes it to the disabled. People with disabilities do not benefit directly from that spending — the question is how they value what the bureaucracy produces.

Although costs and benefits are presented in dollars, cost-benefit analysis is conducted in real terms — using either the consumer price index or the GDP deflator. Thus, real economic magnitudes are either expressed in 'consumer baskets' or in 'producer baskets'. Good cost-benefit analysis should also account for possible changes in relevant relative prices. The NDIS almost doubles disability spending, which is likely to bid up the price of disability services. If the relative price of disability services increases, then people with disabilities will receive fewer services per dollar of expenditure. Doubling expenditure doesn't automatically double the amount of real services transferred to the disabled. Some of the expenditure goes to scarce factors in the industry (such as monopoly equipment manufacturers). The PC has considered this possibility

(2011b: ch.15) and concludes that 'There is a danger that a rapid increase in demand for support staff will result in staff shortages as well as wage inflation (PC 2011b: 693).⁵

In summary, the PC assumes an extra billion dollars of expenditure is always valued by the recipients at a billion dollars. But to evaluate the net benefit of extra spending, we need to measure the benefits and costs of that spending, not assume the recipients value it at cost. This assumption does not allow the PC to compare different policies. If each dollar of spending on the NDIS is valued by the recipient at a dollar, how do we know what to spend it on within the scheme? What is the optimal amount to spend on bureaucracy? The PC has asserted that current state programs are poor, but what if we doubled spending on them? Say we doubled the wheelchair subsidy? The PC presumes that a dollar spent on its scheme is valued more than a dollar spent on existing schemes, but that is what needs to be established.

Willingness to pay or marginal utility?

The PC (2011b: 957) sets out three incremental benefits from the transfer.

Accordingly, the NDIS and associated reforms produce three broad economic benefits:

- the excess value of consumption from resources transfers to people with disabilities (V or C2-C1)
- efficiency benefits (E), such as more efficient service provision
- the economic benefits of fiscal savings, such as those arising from reduced DSP payments (F).

The PC focuses on 'V'; that is, the gain in total utility from transferring from taxpayers to people with disabilities — which comes from the recipients having a higher marginal utility of income than the average taxpayer. But putting costs and benefits in terms of utility departs from the efficiency criterion and conventional cost-benefit analysis, which measures costs and benefit in terms of *willingness to pay*, not *utility*. The PC does not clearly separate equity and efficiency — offsetting utility gains from redistribution against dollar efficiency costs from taxation and presenting the result as a net benefit.

The PC (2011b: 955–6) admits 'controversy about the measurement of the MUI and the extent to which cost-benefit analysis should take account at all of the redistributive effects of government interventions ... Nevertheless, it is common

⁵ The PC implicitly assumed that direct wage costs for those providing the care and support accounted for roughly half the costs of this element of the scheme (AGA 2013: 12).

practice for cost-benefit analysis to provide higher weights from benefits or costs to people with lower incomes.’⁶ The PC justifies its approach by quoting from the Department of Finance and Administration Handbook of Cost Benefit Analysis:

As a general practice, it is recommended that analysts refrain from attaching distributional weights to cost and benefit streams in the interest of avoiding subjective bias. The exception is where an unambiguous government policy objective can be identified to assist the specific group at which the project or programme is aimed; and where the priority of assistance to this group relative to other groups is also clearly established. These are stringent and restrictive conditions. (Commonwealth of Australia 2006: 83.)

The PC, reasonably, points out that the conditions are satisfied here, but they fail to quote the next lines:

Moreover, even in these instances, it is important to include in the report the estimate of the *unweighted* net present value, so that the absolute cost of the distributional judgement can be measured. It is also highly desirable to develop an estimate of the efficiency cost of the alternative means of achieving a similar income redistribution, for example, through a direct transfer payment. (Ibid.)

It is true that the underlying rationale for disability support is an equity one. Yet the efficiency effects are worth knowing, and something that can be objectively measured. An efficiency-based cost-benefit analysis can help to inform the decision and clarify the trade-offs when comparing alternative policy proposals, such as how much income may need to be sacrificed to achieve other objectives. By abandoning efficiency analysis, the PC forgoes these benefits.

Separating equity does not mean that distributional concerns are unimportant or should be neglected. It means that they should be brought into account as a separate part of an overall analysis of the problem in question — which may be more important than the resource allocation part, but should not be mixed up with it.

Doing equity properly

Equity is about what is fair and is based on ethical judgements, and so is inevitably subjective. People will have different notions of equity and its importance.

⁶ A referee commented that using distributional weights was ‘common practice’: ‘Not in this writer’s knowledge’. The PC does not cite any examples of studies that do so.

There are two major views of equity: the distributional-weights and the basic-needs approaches.

The PC adopts the distributional-weights approach. It is based on a social welfare function, which represents some ethical judgement about the appropriate distribution of welfare across people affected by a policy change. Just as a utility function shows how a person ranks different combinations of consumption goods, the social-welfare function represents a value judgement of how society should rank different distributions of utility across people. The standard social-welfare function values equality — dollars to the poor increase social welfare more than dollars to the rich and receive a higher distributional weight.

The social-welfare function approach is one particular view of social choice that may not capture how most people think about social welfare or account for equity. For example, the standard form of social-welfare function focuses on equality. But most people would be unconcerned about a transfer of income from a very rich person to a comfortably rich person, yet a standard social-welfare function would say it raises social welfare.

Harberger (1978) points out that the distributional-weights approach has radical implications. It provides a justification for government intervention in any market where buyers and sellers have different incomes. Any policy that redistributed from the richer to the poorer group (such as interventions to drive the market price up or down) would give distributional gains.

Balancing distributional gains against efficiency losses requires acceptance of large efficiency losses for the distributional weights chosen in practice. For example, the PC uses distribution weights of $w_i = (y_a/y_i)^{1.24}$ where y_a is median income, y_i is the income of individual i and 1.24 is the elasticity of the marginal utility of income (PC 2011b: 971). If we use the example from the PC's Box 20.1 'A Tale of Two People' (PC 2011b: 957), about a redistribution from Mike on \$150 000 to Mary, a disabled person on \$25 000, then using the PC's distributional weights, a dollar to Mary is worth $(150\,000/25\,000)^{1.24} = 9.2$ times more to society than a dollar to Mike.

So taking \$100 off Mike, wasting \$89 of it (say on administration costs) and giving \$11 to Mary would increase social welfare according to the PC's distributional weights.

Harberger argues such outcomes are unacceptable. He concludes that the distributional-weights approach does not capture how most people think about distributional issues. It does not represent the value system of most citizens and risks economists' peculiar opinions on distributional issues swamping all other considerations, something that is beyond the economist's professional role (Harberger 1978: S118–9).

An important policy issue is the size and scope of the NDIS. The PC recommended limiting it to 410 000 people with a significant 'tier 3' disability. But in 2009 there were four million people with disabilities in Australia (PC 2011b:15; ABS 2011). There is much scope, and political pressure, for the scheme to expand to cover more people and more disabilities. The only limit to expansion under the PC approach is the MEB; each dollar spent on the scheme costs taxpayers \$1.24. But this is dwarfed by the PC's assumed equity benefits, where redistribution can easily lead to a \$9 benefit for each \$1 cost to taxpayers. The PC's ethical judgement about the benefits of redistribution justifies massive expansion of this and many other redistributions.

An alternative way to bring equity benefits into the cost-benefit framework is what Harberger has labelled the basic-needs approach, (Harberger 1984; Jenkins *et al.* 2011). Rather than rely on the differential weighting of the welfare of different individuals, this approach imputes external benefits connected with the improvement in the circumstances of others. But the basic-needs approach assumes it is not the recipient's utility or income that enters the donor's utility function but the consumption of particular goods and services (food, education, medical care, housing, and so on) or the attainment of certain states (better nourished, better housed ...) that are closely correlated with the adequate consumption of certain goods and services — the extent to which the basic needs of certain segments of society are met. It seems in practice that the altruism we observe is more closely linked to the basic needs of individuals rather than to their utility or entire consumption bundle. Citizens want welfare payments to be spent on food and clothing, not beer and drugs.

The basic-needs approach implies gifts in kind rather than cash transfers. If donors get benefits from seeing the poor better fed, then give them food. Most people would think a food-stamps program has failed if the recipients convert it to cash and spend it on drugs or gambling.

Harberger (1984) suggests that, judging by people's charitable giving, redistributions within their family and gambling behaviour, most seem to care about alleviating poverty. Genuine deprivation — an inability to afford the necessities of life — is what motivates most charity. Most people genuinely believe it is good for the sick to be healed, the homeless sheltered and so on.

Families step in to help fellow members meet basic needs, but seldom redistribute to equalise income (for example, bequests are usually divided equally rather than to offset income differences between children).

Further, the fact that a large proportion of the public participates willingly in lotteries suggests that many people do not value equality. Lotteries increase inequality — many people purchase tickets and make themselves poorer in order to make a few winners rich.

The basic-needs approach follows standard cost-benefit analysis techniques and assumes external benefits come from alleviating poverty and fulfilling basic needs. External benefits from giving can rationalise a role for government in redistributing income.

People may wish to transfer income to the disabled — and do so through charities — for a number of reasons. In these cases, government involvement in redistribution may be efficient. Charity may be a public good that is under-provided because of the free-rider problem. We all benefit when people with disabilities are helped.

That is, if one person gives to a disabled person, this benefits everyone who cares about the disabled person (giving can involve dollars or goods). It is a public good (non-rival and non-excludable). If givers do not take account of these positive externalities, they may under-provide private transfers.

It can be efficient for the government to redistribute. For example, if A is altruistic to B, taking a dollar off A and giving it to B is efficient (but not a potential Pareto improvement) — B gains a dollar, A loses less than a dollar (as A benefits somewhat if B gets a dollar).

It is possible for everyone, or most people, to be better off if the government forces extra giving to those in poverty. Although individuals lose by being forced to give more than they want, they benefit from the increased poverty alleviation from other people. The high degree of consensus about the need to adopt the PC's recommendations indicates it is a realistic possibility in the disability sector.

Although redistribution can be given an efficiency justification, equity and efficiency should still be separated because the externalities from redistribution are controversial and difficult to measure.

The basic-needs approach, then, says that 'society' is willing to pay a premium in order to meet more fully the basic needs of people with disabilities. This premium reflects a willingness to put up with certain amounts of extra cost, or of economic inefficiency, if this makes possible the fulfilment of some unmet basic needs of the disabled. The size of the premium assigned for a given basic need, and the definition of the base to which that premium applies, defines the precise trade-off involved — that is, how much society is willing to pay for what specific sign of improvement.

Although economists tend to mock ‘needs’ — which tend to expand when the user doesn’t pay — the basic-needs approach does better capture the underlying rationale for disability policy. Indeed the PC Disability report focuses on providing support to meet basic needs, such as self-care, mobility and communication, and increasing funding to ‘address the high level of unmet need’ (p.156). The PC recommends the program be limited to third-tier ‘people with support needs that would otherwise not be reasonably met without taxpayer funding’ (p.157).

That is why the transfers are in the form of services. In contrast, the distributional-weights approach is about maximising the utility of the recipients — which implies cash transfers.

A focus on basic needs rather than distributional weights is a better way to analyse the equity effects of disability policy than the PC’s distributional-weights approach.

The basic-needs approach focuses attention on the effectiveness of policies: to what extent they fulfil basic needs. In contrast the distributional-weights approach employed by the PC implies a large willingness to accept inefficiency and invites waste.

The basic-needs approach has the largest externalities from helping the worst-off groups. But the distributional-weights approach may consider the effects of a larger subsidy on less-poor groups to bring more social benefits, as it focuses on the income transfer implicit in the subsidy.

In the basic-needs approach, the social benefit of each incremental increase in subsidy is smaller, but this may not be true in the distributional-weights approach, where the benefits may increase with the size of the subsidy. With distributional weights, the greater the expenditure, the greater the benefits.

Conclusions

Although the PC uses the language of conventional cost-benefit analysis and concludes that the benefits of its proposed disability care reforms exceed their costs, the PC adopts an unconventional approach which focuses on benefits from ethical presumptions.

The PC’s cost-benefit analysis is not efficiency based and so cannot reap the benefits that such an evaluation allows: such as evaluating different options and helping develop policy. The PC considers only its recommended policy, which means it can only establish that its proposals are better than doing nothing, not

that they are the best amongst alternative approaches, or cannot be improved, or should be scaled up or down. Cost-benefit guidelines usually require a number of options be analysed.

The efficiency effects of a policy are worth knowing to inform decision-making. The process of trying to describe and measure costs and benefits is valuable in itself and encourages policymakers to come up with better ways to achieve their objectives. By examining what determines the costs and benefits and how they are likely to vary, policymakers are encouraged to consider different approaches and determine the best way to achieve objectives. Identifying and measuring costs and benefits encourages close examination of the factors that influence them and assists in minimising costs and maximising benefit, helping decision-makers increase net benefits to society. Is the policy the best way of producing them — or could a better outcome be produced by some alternative? An efficiency approach encourages analysts to think about where the costs and benefits come from and how to increase the benefits and decrease the costs.

Further, it is not even clear whether the PC's methodology for assessing the equity effects — a distributional-weights approach — is the best way to so. A basic-needs approach better captures the underlying preferences of citizens and the rationale for disability care and support policies.

References

- Australian Bureau of Statistics (ABS) 2011, *Disability Australia, 2009* cat no, 4460.
- Australian Government Actuary (AGA) 2013, 'NDIS Costings — Review by the Australian Government Actuary'.
- Commonwealth of Australia 2006, *Handbook of Cost Benefit Analysis* (January).
- Harberger, A. 1978, 'On the use of distributional weights in social cost-benefit analysis', *Journal of Political Economy* 86: S87–120.
- 1984, 'Basic Needs versus Distributional Weights in Social Cost-Benefit Analysis', *Economic Development and Cultural Change* 32(3): 455–74.
- 2007, 'Suggested New Steps Toward the Practical Implementation of Cost-Benefit Analysis', Paper presented at a Conference on Cost-Benefit Analysis, University of Washington, Seattle.
- Jenkins G., Kuo, C. and Harberger, A. 2011, 'Shadow Price of Government Funds, Distributional Weights, and Basic Needs Externalities', chapter 14 in *Cost-Benefit Analysis for Investment Decisions* (manuscript).

Parliament of the Commonwealth of Australia 2013, 'Medicare Levy Amendment (Disability Care Australia) Bill 2013 - C2013B00105, Explanatory Memorandum at: <http://www.comlaw.gov.au/Details/C2013B00105/Explanatory%20Memorandum/Text>.

Productivity Commission 2011a, *Disability Care and Support: Overview and Recommendations* Report no. 54, Canberra.

——— 2011b, *Disability Care and Support*, Report no. 54, Canberra.

The Power to Tax, 33 Years Later

Jonathan Pincus¹

Abstract

The basic puzzle about the power to tax is how to limit the capacity of government to exploit taxpayers, while at the same time not overly hampering the government in going about its useful activities. Standard economics fondly believes that it is giving advice to benevolent despots as to how to collect a given target of tax revenue at the least possible harm to the size of the economic pie. The Constitutional Political Economy approach of Geoff Brennan and Jim Buchanan showed that that very same advice is exactly what the non-benevolent government wants to hear in its efforts to maximise tax revenue. Brennan and Buchanan were concerned about excessive exploitation of taxpayers in the large; standard economics is concerned with second-order small triangles of economic inefficiency; government is concerned about the size of first-order revenue rectangles: and so should we be.

‘Good intentions will always be pleaded, for every assumption of authority. It is hardly too strong to say that the Constitution was made to guard the people against the dangers of good intentions. There are men in all ages who mean to govern *well*, but they mean to govern. They promise to be *good* masters, but they mean to be masters.’

Daniel Webster, American politician and diplomat (1782–1852)

The standard economic approach to taxation is to give advice to a benevolent despot. Courtesy of a referee, here is what motivates a benevolent government:

A benevolent politician seeks revenue (or a transfer of resources from the private sector) to establish and monitor property rights for the market sector, fund public goods and correct other market failures, and to achieve society’s equity goals via direct income transfers, progressive income taxation and the provision of some goods and services at below

¹ The University of Adelaide; jppincus@gmail.com. Thanks, but no blame, go to Geoff Brennan, Henry Ergas and the referees. An earlier version was given at a conference — ‘Freedom to Choose 2013: the economics of government failure’, at the University of Notre Dame, Fremantle, sponsored by the Mannkal Foundation. The opinions are my own.

cost. The benevolent politician/economist, in assessing the scale of taxation and expenditure, seeks to balance the marginal social cost of any tax used with the marginal social benefits of any item of expenditure. The total of taxation is the result of this balancing.

In the standard theory of public finance, the predominant form of benevolence is utilitarian, which involves a trade-off between redistribution and the aggregate size of the economic pie. Less commonly, the standard theory is cast in Paretian terms: the benevolent ruler is only concerned with efficiency with which the revenue target is reached (for example, Battaglini and Coate 2008). When it comes to taxation, in both cases the benevolent politician's objective is to meet a target of tax revenue while doing the minimum harm to the economy. The standard approach worries about the efficiency of the tax system; that is, it is concerned with excess burden triangles.

In 1980, Geoffrey Brennan and James Buchanan published *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Their book was directed at the intelligent citizen and not at a benevolent despot. Brennan and Buchanan assumed that the state acted selfishly, and spent on things that the rulers wanted, and not necessarily what the citizen wanted. A monopolist wants to gather in as much revenue as it can for its own purposes. Brennan and Buchanan applied the name 'Leviathan' to a state that did the same — that is, the Leviathan state engrosses as much revenue as it can for its own purposes.

Brennan and Buchanan acknowledged that their Leviathan is not a fully realistic description of government in the developed world — in part, because many countries, including Australia, have constitutional provisions and conventions that constrain the Leviathan-like tendencies of the state. In mainstream economics, these constitutional provisions and conventions are regarded as unfortunate, because they prevent benevolent governments from achieving the best of all possible outcomes for their citizens. In contrast, from the Brennan and Buchanan viewpoint, these constitutional provisions and conventions prevent non-benevolent governments from exploiting their citizenry to the full. Brennan and Buchanan's objectives in writing *The Power to Tax* were to understand the political economy of these kinds of constitutional provisions and conventions, to assess their effectiveness and to canvass additions and modifications.²

Brennan and Buchanan's book was analytical and did not provide a short description of what motivates a Leviathan government/politician. However, if it had, maybe it would run something like this:

2 The recently launched campaign of the Centre for Independent Studies, to limit tax revenues to 30 per cent of GDP, has a similar concern — to restrain the state within reasonable macro-economic limits so as to prevent government from excessive exploitation of taxpayers generally.

In a democracy, governments gain power via complex majoritarian processes. Because the probability is trivial that your vote will affect the electoral outcome, voting is mainly an expressive act: expressing support of those who support your interests and values. Governments provide some public goods, but mostly provide benefits to particular segments of the population through laws and regulations, and through government transfers and final expenditures; and distribute burdens, including by taxation. Decisions about government spending and taxing are made on many grounds, including serving the interests of sections of the population who reward government in return, and who may engage in expenditure-seeking and tax-diverting activity. Government decisions are made within a polyarchy, and not by a single mind applying calculus to maximise a social welfare function. One government cannot commit the next. The Leviathan politician seeks to balance the marginal political and personal cost of taxation with the marginal political and personal benefits of a range of specific expenditure items or categories. These costs and benefits may be extrinsic or intrinsic, ideological or material. Total taxation very likely exceeds what a benevolent ruler would choose.

The benevolent despot liveth

In 1986, six years after the publication of *The Power to Tax*, Buchanan won the Nobel Prize for helping establish and nurture the intellectual fields of ‘public choice’ theory and ‘constitutional political economy’. Therefore, you may well expect to see, in any serious study of tax theory, a reference to *The Power to Tax* and to its arguments.

Do not hold your breath.

In 2008, the superb tax economist Louis Kaplow published *The Theory of Taxation and Public Economics*. Kaplow never referred to *The Power to Tax*; moreover, he did not even acknowledge that there was any need to discuss what Buchanan called ‘the structure within which political decisions are made’. Instead, Kaplow proffered his policy advice to a benevolent despot.

More recently Robin Boadway, a top theorist in the field of public economics, published a book entitled *From Optimal Tax Theory to Tax Policy*. It is meant to offer theory-based but practical advice to government. Boadway explicitly based his recommendations on the assumption that tax policy is run by a beneficent

social planner with an unchallenged ability to commit future governments when it came to tax policy: to assume otherwise would be to stifle the usefulness of his analysis (Boadway 2012: 239).³

Mention should also be made of a major recent study of the UK tax system, commissioned by the Institute for Fiscal Studies in London. It was conducted under the direction of the Nobel Prize winner James Mirrlees (2010; 2011). In their reports, Mirrlees and colleagues decided not to take any account of the possible failure of government to act benevolently. This led the prominent Harvard economist Martin Feldstein (2012: 782) to remark that the Mirrlees Review contains no discussion of 'the political legitimacy of unlimited redistribution'. 'What kind of nation', Feldstein asked, 'places no value on the welfare of those with income in the top tax bracket, treating them only as the revenue producing property of the state?'

But to regard individual people as merely 'the revenue producing property of the state' is not specifically a 'British disease'. Front and centre of the Australian Treasury's (2012) presentation of self, in its 'Strategic Framework', is a statement entitled *The Wellbeing of the Australian People*. It is redolent with the presumption — seemingly pervasive at Treasury — that all incomes and assets are 'the community's', so that, if something is left in private hands, then it is only by the grace and favour of government or through government's incapacity to devise mechanisms that better achieve its goals while leaving less as the private income of the taxpayer.

In a speech about the 'wellbeing framework', senior Treasury official David Gruen (2011) remarked: 'We [in Treasury] of course see rewarding effort and risk as important from an efficiency perspective, but for that reason only.' Gruen's reference to 'efficiency' may comfort readers used to thinking in Paretian terms. However, that comfort is unjustified: the words 'economic efficiency' or close synonyms do not appear in the 'wellbeing framework'. Rather, one of the main purposes of the framework clearly seems to be to attempt to protect Treasury from the claim that it continues to place too much emphasis on economic efficiency (Pincus, forthcoming). Certainly, there is no hint in the framework (or, from what I have seen, in senior Treasury speeches) of concern that taxation is too high. Rather, Treasury Secretary Martin Parkinson has expressed anxiety that taxation is too low to fund government spending.

Treasury has travelled a long way from the time when, it is claimed, the then Secretary opposed the introduction of a goods and services tax on the

3 'Optimal tax research in the spirit of Mirrlees (1971) has generally avoided situations in which the Revelation Principle does not apply, such as if the social planner cannot commit to a future policy plan' (Mankiw, Weinzierl and Yagan 2009: n.3).

grounds that it was so much more efficient than the taxes it was to displace that, inevitably, it would lead to an increase in government revenue and, therefore, even more wasteful spending.

Tax theory in orthodox public finance

Standard economic theory of taxation says if it cannot move, then tax it hard; but if it can easily move (or change) to avoid the tax, then tax it lightly. Thus, the state should place a heavier tax on entities that find it hard to avoid some of the tax by fleeing the jurisdiction; on activities that do not have close, untaxed or lightly taxed substitutes; on characteristics of the taxpayer that are costly or slow to alter. (I leave it to the reader to construct the contra-positive statements.)

Let me illustrate. Gordon Tullock, then a famous economist, stayed with me in Canberra for a few days in 1974. At one of his talks, Tullock said that it was only when he noticed that Jonathan Pincus was doing his own brick paving, that he came to appreciate just how high were the marginal tax rates of the Australian income tax schedule. Jonathan had to earn an extra 4000 dollars in order to retain 1000 post-tax dollars with which to pay someone to do his paving: the marginal tax rate was 75 per cent. Therefore, said Tullock, Pincus found it worthwhile to do his own paving, even though his comparative advantage lay in economics, not in laying paving.

The high marginal tax rates encouraged people into do-it-yourself activities, which were tax-free; to this extent, the high marginal tax rates on those with high earning ability reduced the size of the monetised or 'market' economy and, therefore, reduced the taxable capacity of Australia. So, if you were advising benevolent government how to collect the required amount of tax revenue, your advice would be to tax income less heavily. Something like Milton Friedman's 'negative income tax' schedule seems to be approximately optimal, with a large 'demogrant' and a flat marginal rate schedule (see Mankiw, Weinzierl and Yagan 2009).

But what should be put in its place? Here standard economics advocates heavier taxes on natural resources and land, neither of which can move in response to taxes. High resource and land taxes would harm the economy less than would high marginal income tax rates.

But say you were advising a Leviathan government, keen to maximise its tax take. Again, you would advise against very high tax rates on incomes — avoid the backward-bending part of the Laffer curve. Rather better to cut those rates and increase and broaden the taxes on land and other natural resources.

Colbert, in the seventeenth century, saw the problem from both sides, when he allegedly said: 'The art of taxation consists in so plucking the goose as to obtain the largest possible amount of feathers with the smallest possible amount of hissing.'

Standard economics aims to minimise the hissing for a desired amount of feather; but the same tax rules would produce the maximum amount of feathers for a given amount of hissing. Brennan and Buchanan argued that their analysis leads to the conclusion that there would be more feathers (tax) and more hissing (burden) under Leviathan than under a benevolent ruler; yet both would follow the same kind of rules of tax design.

Mainstream economists like Mirrlees, Boadway and Kaplow fondly think that they are giving advice to a benevolent despot. The brilliant insight that Brennan and Buchanan brought to bear in *The Power to Tax* was that Mirrlees and Co are unknowingly giving exactly the advice that the Leviathan seeks: how to gain the maximum in tax revenue.

To show this requires two steps.

Firstly, the tax rates that Mirrlees and Co derive for a system of 'optimal taxation' are the very same tax rates to achieve largest tax revenue compatible with given a pre-specified amount of harm being caused to the economy by the taxation. In mathematical terms, the two optimising problems are duals.⁴ Even worse, the Leviathan is not worried about maximising revenue for a given, pre-specified amount of harm to the economy. No, the Leviathan wants the maximum flow of revenue: he is a despot, and in for the long term, even if the political complexion of government changes periodically.

At this point comes the second step: the general tax rules that standard theory derive for a system of 'optimal taxation' are the very same rules that the Leviathan would apply: high tax rates on land and resources and other things that find it hard to escape the tax; lighter rates on fugitive tax bases. The rates would be higher and the quantity of tax revenue greater than if the marginal social-welfare calculus set a constraint on the amount of harm caused to the economy. But the general pattern of tax rates would be just as standard theory prescribes.

4 The same equimarginal conditions arise from the following two problems: $\min B(\tau_1, \tau_2, \tau_3, \dots)$ s.t. $R(\tau_1, \tau_2, \tau_3, \dots) = R^*$, and $\max R(\tau_1, \tau_2, \tau_3, \dots)$ s.t. $B(\tau_1, \tau_2, \tau_3, \dots) = B^*$, where B is burden, B^* is any given burden target, R is revenue, R^* is a revenue target, and τ_i are the tax rates. The Benevolent ruler, using a social welfare function, maximises welfare by altering the budget size and composition until the shadow price (or cost) of additional revenue from any source equals the shadow price (or benefit) of additional spending of any type. The Leviathan rulers select that quantum of taxation at which the Leviathan's own marginal benefit equals the Leviathan's own marginal cost.

By proffering tax policy advice to benevolent rulers, unknowingly Mirrlees and Co were giving advice to the revenue-seeking Leviathan. For many decades, Mirrlees used his magnificent intellectual power to assist government to pluck more feathers with less hissing. He devoted no effort at all to advising citizens about how to stay the hand of the pluckers.

Even if, on moral or merely practical grounds, it were the case that all income and wealth ‘belongs to the community’, then it would still be sensible for ‘the community’ to take precautions against the possibility that governments tax and spend too much, and leave too little for private disposition.

Taxation of mining⁵

In my opinion the standard theory of public finance was offered in support of a tax on pure rent, which would have garnered modest revenues from the mining industry; and that the theory was used by government to justify a tax that would, in fact, have fallen mostly on *quasi-rents*, and which *would* have gathered huge revenues.

This effort drew support from a range of mainstream economists.

Dr Ken Henry, then Secretary of Treasury, chaired the last major review of the Australian tax system (Treasury 2010). In keeping with standard public finance economics, the review recommended lower taxes on personal and company incomes, the abolition of mining royalties, and the introduction of a new tax on mining rents.

By definition, a pure rent is any payment made to a factor of production that is not needed to bring it into the existence and is not needed to cause it to be supplied as a factor of production. A quasi-rent is a payment made to an already-existing factor, and in excess of what would be needed to elicit its supply of productive services. Therefore, a tax on pure rents does not alter the quantity supplied of the taxed objects or services; however, a tax on quasi-rents can alter the future supply of the taxed factors of production.

A tax on pure rents is the economist’s version of the magic pudding or the perpetual motion machine: it is a tax that generates revenue but imposes no excess burden on the economy. However, its support is an incomplete theory, one that ignores information asymmetry of the kind emphasised by Mirrlees in his work on optimal income tax schedules (beginning with Mirrlees 1971), as well as Laffont and Tirole (1983) on regulation, and which bedevils actual

⁵ This section draws freely on Ergas, Harrison and Pincus (2010), where it is argued that no actual tax can produce revenue without any incentive effects on the parties liable to the tax.

attempts to tax pure rents;⁶ and it is a theory that ignores that there is a limited supply of the talent and knowhow, of the kind needed to access the full value of rents from minerals in the ground. Be that as it may, to recommend taxes on pure rents is in keeping with standard theory of public finance.

The new tax proposed by Dr Henry became the Resource Super Profits Tax, RSPT, projected to produce many billions: at that time, the government was desperately short of revenue. But the Resource Super Profits Tax elicited too much hissing, which helped to bring the first Rudd government to a sudden end. The Rudd #1 government was too unconstrained in its tax ambitions. Instead of the RSPT, the Gillard government imposed the Minerals Resource Rent Tax. Less hissing, but virtually no feathers — the MRRT was negotiated with the large established miners, and has produced almost nothing by way of revenue in its very short life so far.

The Resource Super Profits Tax, the Rudd #1 tax, was excessively exploitative — it was a clear case of government failure in the tax arena. The Minerals Resource Rent Tax, the Gillard tax, in contrast, represents a sensible compromise between the exploitation of property owners, and the need for public revenue. Although it has serious flaws and was oversold, the MRRT is not a clear case of government failure in the tax arena.

Dr Henry claimed that economic theory of the 'Brown' tax told him that a tax rate of 60 per cent, even 90 per cent, would not lead to any diminution of mining in Australia. He must have had in mind a tax on pure rents. Ministers and Treasury officials analogised the RSPT to a government taking a form of equity in mining ventures, and paying for it, as it were, with a promise of covering 40 per cent of any costs, including losses (the latter covered by the issuance of government promissory notes, bearing interest at the long-term bond rate).⁷

This could be a reasonable representation for new ventures (subject to how costs are defined and measured, and ignoring questions of asymmetrical information that bedevils attempts to tax rents, as well as questions of the supply of mining talent and knowhow). Via the RSPT, the government becomes an equity partner in an uncertain prospect. The history of these projects, for good or bad, lies in the future. When the RSPT was applied to mining ventures initiated after the imposition of the tax, it was reasonable to claim that the purpose of the RSPT was the taxing of pure rents, and pure rents only. Realistically, this source of RSPT revenue would have been small, delayed and uncertain. (Some prominent tax economists advocate a version of the super profits tax for all industries.)

6 Dasgupta and Heal (1979) derive the expression for a resource royalty that captures all of the pure rent. Its implementation requires that the taxing authority knows the supply or cost function of the miner.

7 Company tax is to be paid on profits net of RSPT payments.

However, the analogy between the RSPT and taking an equity punt does not work for existing ventures. Part of the history of an existing project has happened. Government would not have been taking its share of what came along, the good times and the bad. Rather, the first Rudd government wanted to impose the tax following and in response to a series of years of high and rising mineral prices; that is, when times were spectacularly good. For existing projects, this retrospectivity means that government avoided bearing the threat of losses that it would have borne in an actuarially neutral (that is, fair) bet. Most of the RSPT revenue would have come from quasi-rents, not from pure rents. Realistically, this source of RSPT revenue would have been large, soon and relatively certain.

I repeat one of the two rules of standard economics: tax heavily if the taxed activity or object cannot easily evade or avoid. When these two taxes were introduced, the mining companies had invested many tens of billions into mining in Australia. Their investments were sunk, literally in many instances — they could not be packed up and shipped cheaply to another country. They were just standing there, ready to be plucked, according to mainstream economics: tax revenue would have been large and, depending on the reaction of investors in mining and generally, the excess burden relatively small.

The MRRT, which applies to iron ore, coal, and coal seam gas, has a nominal 22.5 per cent tax rate applied to cash flow less an allowance for capital costs, with any negative cash flow carried forward at the long-term bond rate plus seven percentage points. Existing mines were able to set their initial capital at market valuation, and so the element of taxation of quasi-rents, such a feature of the RSPT, is largely absent.⁸ Rather, the Minerals Resource Rent Tax is primarily levied on the profits of *future* investments in mining, not past investments.

It is one thing to impose a rent tax prospectively, which is what the Minerals Resource Rent Tax does — then any project that pays the tax is a project that commenced after the tax was imposed. It is quite another thing to impose the RSPT retrospectively on existing projects, as the first Rudd government intended to do.

The Australian constitution requires the federal government to pay just or fair compensation for property taken by government.⁹ This contradicts standard economics — if it is natural or exogenous capital and it cannot flee, then tax it very hard, even at 100 per cent; if it is produced capital, then, having regard

⁸ Presumably, the market valuations capitalised expected profits, net of taxes and royalties and, therefore, capitalised residual resource rents. Ergas, Harrison and Pincus (2010) argued, however, that the MRRT discriminates against risky projects; and discussed sovereign risk.

⁹ Section 51 xxvi: The acquisition of property on just terms from any State or person for any purpose in respect of which Parliament has power to make laws.

for incentive effects, tax it heavily. The constitution's provision about fair compensation for property reflects a concern about exploitation by government and, moreover, about selective exploitation — only the miners were to pay the super-profits tax.

I am not arguing that no taxation is legitimate if it causes capital losses to anyone: that would rule out almost every tax, even a poll tax: people may leave the city or nation, if a severe enough poll tax were levied, but few would do so in response to the light poll tax that Margaret Thatcher imposed. But Mrs Thatcher learned that a poll tax also brings people out onto the street in protest, which helped to bring her down.¹⁰ The unpopularity of taxes seems to be in direct relationship to their approval rating from economists in the mainstream tradition.

My test is not that taxation should Do No Harm, but that taxation should make a reasonable compromise between exploiting or expropriating, on the one hand, and funding the useful activities of government, on the other hand. Taxes that take a sizeable share of the value of some long-lived assets, or the income or wealth of a small portion of the population, do not satisfy that vague test.

A height tax

In standard economics, the economics of the benevolent despot, the ideal tax base has two characteristics: it cannot be hidden; it is unchangeable. That is to say, the ideal tax base is easily observed and measured by the tax department, and it is difficult or impossible for the taxpayer to change its nature, size or location. Governments are keen to find and get access to tax bases that display these characteristics or come close.

In their search for revenue, governments have been well served by standard economics — many are the articles in leading economics journals advising governments along these lines. For example, James Mirrlees, mentioned earlier, won his Nobel Prize for his work on how governments should structure income tax scales in view of the fact that governments cannot directly observe the earning ability of a taxpayer, but only their actual earnings.

Recently, some economists have shown that taller people and better-looking people have higher earning capacities than shorter and less-beautiful people, all else being equal. This has led Mankiw and Weinzierl (2010) to discuss putting a

10 Mankiw, Weinzierl and Yagan (2009) remarked that 'As this [Thatcher's poll tax] episode suggests, the social planner has to come to grips with heterogeneity in taxpayers' ability to pay. If the planner could observe differences among taxpayers in inherent ability, the planner could again rely on lump-sum taxes, but now those lump-sum taxes would be contingent on ability.'

higher tax on taller people. Such a tax is not likely to cause many tall people to have an operation to make them shorter; nor would a modest tax on handsome people cause many to have an uglifying operation.

Standard economic theory says that income earners react to an income tax by legally shifting effort away from taxable activities, into home production or leisure; and some evade the tax illegally. However, people with higher earnings capacity will respond less than other taxpayers: it is optimal for such people to work harder and take less leisure, than those with lower ability.

Of course, governments may well decide not to take the economist's advice, and not put on a heavier tax on taller or more handsome people. Given my family name, I hope no economist can convince government to act on evidence that Jews have a higher earning capacity than non-Jews.

Any government that seriously went down the path of taxing according to personal characteristics — a path that is being strewn by mainstream economists with tall, beautiful poppies — any such government would be headed for failure of a serious kind: excessive and discriminatory exploitation of specific classes of taxpayers. As Mankiw and Weinzierl suggest (2010: 12), democratic societies may have an interest in avoiding the taxation of specific groups as a matter of course to counter the majority's temptation to tax minority groups. The benevolent ruler of standard public finance theory suffers no such temptation.

But there is another concern, one that does not often occur to these tax theorists: offering a new and efficient tax base to government will lead to more tax revenue and to more government spending of a kind not justified by the marginal analysis of standard public finance theory.

Conclusion

One of the most important functions of government is to impose taxes, chiefly in order to finance public outlays. Governments fail if they fail to make a reasonable compromise between exploiting people and limiting economic freedom, on the one hand, and filling the coffers of Treasury, on the other. The Resource Super Profits Tax was a government failure, being too exploitative; the Minerals Resource Rent Tax has faults, but was not an instance of government failure.

In this paper, I have contrasted the standard economics of taxation with the Constitutional Political Economy approach offered by Geoff Brennan, Jim Buchanan and colleagues. Standard economics fondly believes that it is giving advice to benevolent despots, advice as to how to collect a given target of tax revenue at the least possible harm to the size of the economic pie. Brennan and

Buchanan showed that in fact that very same advice, from standard economics, is exactly what the non-benevolent Leviathan government wants to hear in its efforts to maximise tax revenue.

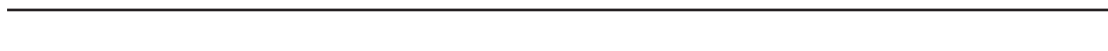
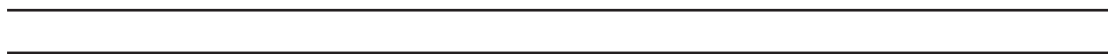
Brennan and Buchanan worry about excessive exploitation of taxpayers in the large. Standard economics is concerned with the second-order small, with the triangles of economic inefficiency; government is much more concerned about the size of first-order revenue rectangles, and so should we be.

References

- Battaglini, M. and Coate, S. 2008, 'Pareto efficient income taxation with stochastic abilities', *Journal of Public Economics* 92: 844–68.
- Boadway, R. 2012, *From optimal tax theory to tax policy: retrospective and prospective views*, Cambridge, Mass: MIT Press.
- Brennan, G. and Buchanan, James M. 1980, *The power to tax: analytical foundations of a fiscal constitution*, Cambridge U.K.: Cambridge University Press.
- Dasgupta, P. and Heal, G. M. 1979, *Economic theory and exhaustible resources*, Welwyn Eng.: J. Nisbet; Cambridge, Eng.: Cambridge University Press.
- Ergas, H., Harrison, M. and Pincus, J. 2010, 'Some economics of mining taxes', *Economic Papers* 29(4): 369–83.
- Feldstein, M. 2012, 'The Mirrlees Review', *Journal of Economic Literature* 50(3): 781–90.
- Gruen, D., Kelly, J., Gorecki, S. and Wong, T. 2011, 'Wellbeing, living standards, and their distribution'; at: http://archive.treasury.gov.au/documents/2153/PDF/speech_dg_nz.pdf
- Kaplow, L. 2008, *The theory of taxation and public economics*, Princeton, N.J. and Oxford: Princeton University Press.
- Laffont, J. J. and Tirole, J. 1993, *A theory of incentives in procurement and regulation*, Cambridge, MA: MIT Press.
- Mankiw, N. Gregory, Weinzierl, M. and Yagan, D. 2009, 'Optimal taxation in theory and practice', *Journal of Economic Perspectives* 23(4): 147–74.

- Mankiw, N. Gregory and Weinzierl, M. 2010, 'The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution', *American Economic Journal: Economic Policy* 2(1): 155–76.
- Mirrlees, James A. 1971, 'An Exploration in the Theory of Optimal Income Taxation,' *Review of Economic Studies* 38: 175–208.
- Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G. and Poterba, J. (eds) 2010, *Dimensions of Tax Design: the Mirrlees Review*, Oxford: Oxford University Press.
- , 2011, *Tax by Design: the Mirrlees Review*, Oxford: Oxford University Press.
- Pincus, J. (forthcoming 2013), 'The Wellbeing of the Australian People: Comments on the Treasury's Framework' in Podger, A. and Trewin, D. (eds), *Measuring and Promoting Wellbeing: How Important is Economic Growth? Essays in honour of Ian Castles AO and a selection of Castles papers*, Canberra: ANU E Press.
- Treasury 2010, 'Australia's Future Tax System. Report to the Treasurer', December 2009; at: http://taxreview.treasury.gov.au/content/Content.aspx?doc=html/pubs_reports.htm (accessed 29 August 2012).
- 2012, 'Australian Treasury Strategic Framework 2012–13'; at: <http://www.treasury.gov.au/About-Treasury/OurDepartment/Treasury-Strategic-Framework> (accessed 11 November 2013).

DEBATE



‘Why Johnny Can’t Regulate’: A Reply to Henry Ergas

Darryl Biggar¹

Henry Ergas is a talented, erudite and articulate economist, and a long-time critic of the ACCC/AER. In a recent paper in this journal² he raises the possibility that public utility regulation is doomed to end in failure. He draws on the vignette of primary-schooler Johnny, who wants to grow up to be a regulator. Is Johnny’s career destined to end up with nothing but disappointment?

Of course, this raises the question as to what public utility regulation is designed to achieve. Ergas addresses this issue at the outset, by allowing that, due to transactions costs, a Coasean bargain contract between the customers and a monopoly supplier would be subject to the risk of ex-post opportunism — both on the side of the supplier (who might threaten to shade quality or raise the price ex-post) and on the side of the customers (who might seek to expropriate the sunk investment of the supplier by collectively agreeing to lower the price to the supplier). Given these risks, either side may fail to make valuable sunk investments, or, as Ergas puts it, either side may make investments which are inefficient due to the threat of opportunism: ‘each side would invest in costly precautions so as to ward them off. But ... those investments, seen from a societal perspective, are merely a waste, reducing welfare.’³

Ergas is thereby led to ask: can the government directly step in to re-create the Coasean bargain on behalf of small customers? He rightly points out many problems in this, including problems of aggregating customer preferences, problems of monitoring and enforcing by customers of the government in its interactions with the supplier, and problems of time-inconsistency and opportunism.

Perhaps these problems can be resolved by delegating certain powers to an independent authority. Such delegation might make it easier for customers to monitor, might improve the information-collection ability, and might resolve the commitment problems. After all, delegating monetary policy to an independent central bank as a means of overcoming the time-inconsistency problem has been a success. Might not the same principle apply to an independent regulatory authority? Ergas goes to some lengths to distinguish the role of a central bank

¹ ACCC; Darryl.Biggar@accc.gov.au.

² Ergas 2013.

³ Ibid: 45.

from the role of an independent regulator. According to Ergas, an economic regulator is likely to have much wider range of instruments at its disposal and more vaguely specified objectives. Moreover, the incentives for time-inconsistency are much stronger for a regulator than for a central bank. He also suggests that newly established central banks seek to invest in a reputation for being tough on inflation, whereas newly established regulators seem to do the opposite:

This [need to establish a reputation] would suggest that the initial period following the transition to regulatory independence would be associated with regulator's credibly signalling a strong aversion to expropriating sunk investments. But in no major country has that been the case ... If anything, utility regulators seemed to preference delivering price reductions to consumers, as evidenced by tough price caps and steep falls in regulated revenues.⁴

As an aside, we may note that any tendency for new regulators to favour customers is still consistent with the regulator seeking to establish a reputation for protecting sunk investments — but with rather more focus on the sunk investments of the customers. If the primary reason why the regulator is established in the first place is a concern that — in the absence of the intervention of the regulator — the customers will be unwilling to make sunk investments for fear of expropriation by the monopoly supplier,⁵ it makes sense for a regulator to counter that fear, proving its mettle by being tough on the monopoly supplier. Such actions may improve the climate for investment by those customers who need to invest in reliance on the monopoly service.

'Last but not least', Ergas expresses concern that regulatory decisions are affected by the wider political context, citing as examples changes to the merits review process in electricity and ministerial powers of discretion in NBN. Ergas concludes:

In short, regulation is no magic wand that can, at no or low cost, replicate the outcomes of the idealised Coasean bargain. Moreover, conventional nostrums for the conundrums it raises — most obviously, the 'solution' of delegating powers to an independent regulator — create many problems of their own. To those problems there are no simple solutions, and most of the attempted solutions have yielded very mixed results.⁶

4 Ibid: 51.

5 This approach is consistent with the rationale for regulation Ergas articulates at the outset. See also, Biggar (2009).

6 Op. cit.: 52.

Johnny is destined for disappointment. Although he is assured of steady work as a regulator 'if he expects regulation to meet the lofty goals so often set for it, that is less simple and far less assured'.⁷

There is no doubt that Ergas has identified many real difficulties in regulation, including the problems of aggregating customer preferences and effective monitoring and oversight of the regulatory authority by customers. Some of these difficulties can be mitigated by separating the responsibility for aggregating and representing consumer views in regulatory processes to an agency separate from the regulator, leaving the regulator the task of arbitrating between competing views of the customers and the monopoly supplier. It is a common practice in institutional design to allocate separate tasks to separate agencies. Doing so clarifies the objectives of each agency — in this case clarifying that the task of the regulator is to act as a neutral objective decision-maker, while it is the task of the customer-representative organisation to negotiate and litigate for the interests of the customers in regulatory proceedings.⁸ Organisations responsible for representing customer interests in regulatory proceedings are common in the US.⁹ Australian governments are in the process of establishing a customer advocacy body for the energy sector.¹⁰

Some of the other problems seem to stem from expectations about what a regulator is established to achieve. One line of thinking suggests that the role of a regulator should be thought of as being akin to that of an arbitrator in a long-term contract.¹¹ In that sense the role of a regulator is something more akin to a dispute-resolution body — albeit a specialist, permanent dispute-resolution body: something closer to a tribunal or a court than a central bank.

The idea here is that faced with the need to protect sunk investments, firms face two basic choices: vertical integration or long-term contracts. Where the number of customers of the monopoly supplier is small we do find both vertical integration (for example, in the iron-ore railways in the Pilbara) and long-term contracts (such as the foundation contracts for some natural-gas pipelines). But when the number of customers is large (such as might be the case for, say, an electricity distribution business), vertical integration and long-term contracts are a bit more difficult. Vertical integration might take the form of a co-operative or customer-owned trust or even government ownership. But these ownership arrangements have their own governance issues.

⁷ Ibid: 53.

⁸ See Biggar (2011).

⁹ Nearly every state has such a body. The national umbrella body for such agencies is the National Association of Utility Consumer Advocates (www.nasuca.org).

¹⁰ See <http://www.scer.gov.au/workstreams/energy-market-reform/national-energy-consumer-advocacy-body/>.

¹¹ See Biggar (2009; 2012).

This leaves us with a long-term contract. I have argued elsewhere that public utility regulation should be viewed as a form of government-imposed long-term contract between the customers and the monopoly supplier.¹² The purpose of that long-term contract is to protect and promote the sunk investments of both parties (the supplier and the customers). But any long-term contract faces a problem: circumstances change over time. The terms and conditions of the contract must be continually adjusted to adapt to changing conditions, without disrupting the incentives for the parties to make sunk investments. This process of adjustment can be facilitated by allowing disputes to be resolved by an arbitrator. All substantial long-term commercial contracts include arbitration clauses. Public utility regulation is no different. The only difference is that the arbitration function is carried out by a specialist, permanent dispute-resolution body known as the regulator. This body resolves disputes in a way which allows for the long-term contract to adapt over time in a manner which is fair and which preserves the incentives for the parties to make sunk investments.¹³ In other words, the task of a regulator should be thought of as a form of dispute resolution, akin to that of an arbitrator or a judge.

If little Johnny had said he wanted to be a judge when he grew up, I doubt that Ergas would have had reason to comment, other than on a precocious interest in legal dispute resolution. Being a judge or arbitrator will not always win applause or accolades. But it can earn respect — respect for wisdom, fairness and consistency. Judges are not there to keep their constituents happy, but to resolve disputes in a manner which is wise, fair and which commands respect. The same is true for regulators. As a regulator, Johnny will not always be able to give the affected parties what they want — in fact Johnny will almost certainly have to get used to disappointing people. Mark Jamison (of the Public Utility Research Center of the University of Florida) is fond of saying that ‘the art of regulation is disappointing people at a rate they can endure’.¹⁴

But Johnny is not doomed to fail. My response to Ergas is that he seems to be expecting more from public utility regulation than it is designed to provide. Rather than the ‘lofty goals so often set for it’, we should expect no more of Johnny than to act with wisdom and fairness, to resolve disputes in a manner consistent with the long-term contract (the Coasean bargain) that the parties would themselves have entered into. In this, I can see no reason why Johnny cannot succeed. There is no need for the counsel of despair. I can see no reason why we could not (using Ergas’ words) ‘in good conscience commend that career choice, as one likely to ultimately lead to personal fulfilment and worthwhile community service’.¹⁵

12 Biggar (2011).

13 For a discussion of the role of fairness in public utility regulation, see Biggar (2010).

14 Jamison (2008).

15 Ergas 2013: 43.

References

- Biggar, D. 2009, 'Is Protecting Sunk Investments by Consumers a Key Rationale for Natural Monopoly Regulation?', *Review of Network Economics* 8(2): 128–53.
- 2010, 'Fairness in Public Utility Regulation: A Theory', *Agenda* 17(1): 5–29.
- 2011, 'Public utility regulation in Australia: Where have we got to? Where should we be going?', ACCC Working Paper No. 4, July 2011; available at: <http://www.accc.gov.au/regulated-infrastructure/about-regulated-infrastructure/regulatory-resources/working-discussion-papers>.
- 2012, 'Why Regulate Airports: A Re-examination of the Rationale for Airport Regulation', *Journal of Transport Economics and Policy* 46(3): 367–80.
- Ergas, H. 2013, 'Why Johnny Can't Regulate: The Case of Natural Monopoly', *Agenda* 20(1): 43–54.
- Jamison, M. 2008, 'Providing Stability While Leading Change', speech at the ACCC Regulatory Conference 2008, 16 July.

A Rejoinder to Biggar

Henry Ergas¹

There is much to agree with in Darryl Biggar's comments on my article. Nonetheless, it does demand the examination of a number of issues.

First of all, it is illusory to think preventing firms from recouping investments prudently incurred can protect the interests of consumers — the opposite is true. For example, if the regulator, confronted with an adverse cost shock, shifts the burden of that shock on to investors in the regulated business, thus stranding investments prudently made, the effect is simply to increase the risk premium those investors will demand in future. In the long run, that higher premium must be reflected in increased supply costs and so in prices or subsidies, affecting consumers, taxpayers or both.

Second, Biggar's suggestion that creating an entity that acts as a consumer representative solves any of the contracting problems I referred to seems inconsistent with both analysis and experience. After all, such an entity still has to identify preferences and aggregate them in a meaningful way; why merely having such a body should make those tasks more tractable is unclear. Moreover, there are obvious issues of governance associated with such entities. The US experience is that they are often captured by self-styled consumer advocates, whose focus ranges from promoting 'green' schemes (despite their higher costs) to a naïve form of populism. I would suggest that has also been the Australian experience with consumer advocates in telecommunications.

Third, just as regulators are not like central banks, so regulators are neither courts nor judges, for reasons explored in Ergas (2009). Precisely because the regulatory task is relatively loosely defined and relies on the collection and analysis of a wide range of information (and so requires a permanent bureaucracy), it cannot be shoehorned into the constraints we impose on courts.

Those constraints, which minimise the opportunities for courts to be 'captured' or to themselves engage in rent-seeking, include severe restrictions on the manner in which information is presented (embodied in the law of evidence and in the rules of proceeding), the nature of the interaction between the court and the parties and the complex structure of appeals. In contrast, economic regulators are generally subject only to the fairly permissive requirements of

¹ The University of Wollongong and Deloitte Access Economics; ergas.henry@gmail.com.

administrative law; especially when the legislation under which they operate is also relatively permissive (in terms of the goals it sets them and the instruments it provides), that accentuates the risks of time-inconsistency I stress in my article.

As a result, analogising regulators to courts is unhelpful. Indeed, the differences between courts and regulators — and the associated dangers of regulators acting in ways that are ultimately harmful — are precisely an argument for ensuring regulatory decisions are subject to adequate appeals mechanisms. In contrast, the ACCC has long advocated paring back such mechanisms, as I note in my article.

In short, when Johnny chooses to be a regulator, he is not choosing to be a judge, just as he is not choosing to be a policeman or a prosecutor. What he is choosing is to exercise a role that is of great importance to efficiency — and hence needs to be carefully designed. The essence of my article is that we are still struggling to do so.

Reference

Ergas, H. 2009, 'Error and Design: Economics in (and some Economics of) the Australian Competition Tribunal', *Agenda: A Journal of Policy Analysis and Reform* 16(3): 71–91.

A Final Word

Darryl Biggar¹

In his response, Ergas makes three points; I will briefly address the first and last.

I cannot be quite so categorical that firms should always be allowed to recover the costs of prudent investments. The Coasean bargain between the firm and its customers may well involve a promise by the regulated firm that it will bear the risks of speculative investments (as would occur in a competitive market). Where there is such a promise it seems unwise to later change the terms of the regulatory contract to raise prices to consumers to cover the costs of unprofitable investments. Doing so undermines the incentives for customers to invest in reliance on the monopoly service. In some cases, preventing firms from recovering prudent investments does protect the interests of consumers.

Ergas is correct to emphasise that a regulator is not exactly like a court. However, I emphasised that a public utility regulator can be and should be thought of as a kind of arbitrator. Like other arbitrators, regulators must develop rules of procedure which, while more flexible than those of the courts, still ensure fair and wise outcomes. Viewed as a form of dispute resolution, I suggest that Johnny is not doomed to fail. Johnny can assist in the ongoing elaboration of the regulatory contract through the resolution of disputes with wisdom and fairness. Like a judge, Johnny may occasionally disappoint and may err. But, like a judge, the dispute-resolution service role remains a valuable part of the long-term regulatory contract.

¹ ACCC; Darryl.Biggar@accc.gov.au.

