

AGENDA

A JOURNAL OF POLICY ANALYSIS & REFORM

Volume 25, Number 1, 2018

Editor and Editorial Committee 1

ANALYSIS

Has Sub-centre Policy Produced Sub-centres? An Evaluation
of Melbourne's Urban Spatial Planning since 1996 5
Jennifer Day, Weiqing Han, Amy Boxi Wu and Jiarui Zheng

The Impact on Research Quality of Performance-Based Funding:
The Case of New Zealand's PBRF Scheme 25
Robert A. Buckle and John Creedy

How Much Have Chinese Investors Invested in Australia? 49
Kerry Liu

ARGUMENT

Ethics in Economics 61
Gigi Foster

REVIEWS

Six Problems in the Biography of Alfred Deakin 79
William Coleman

Alex Millmow's *A History of Australasian Economic Thought* 93
Reviewed by Selwyn Cornish

Editor

Dr William Coleman

College of Business and Economics, ANU

Editorial Committee

Professor Harry Clarke

University of Melbourne

Professor Max Corden

University of Melbourne

Professor Anne Daly

University of Canberra

Professor Kevin Davis

University of Melbourne

Professor Henry Ergas

University of Wollongong

Dr Stephen Grenville

Lowy Institute, Sydney

Dr Arthur Grimes

Motu Research, Wellington

Professor Kim Hawtrey

Nimble Economics

Professor Doug Irwin

Dartmouth College

Dr Gary Johns

Australian Institute for Progress

Professor Tony Makin

Griffith University

Dr Deborah Peterson

Sapere Research Group

Professor David Vines

Balliol College, Oxford

Dr Ruth Williams

La Trobe University

ANALYSIS

Has Sub-centre Policy Produced Sub-centres? An Evaluation of Melbourne's Urban Spatial Planning since 1996

Jennifer Day, Weiqing Han, Amy Boxi Wu and Jiarui Zheng¹

Abstract

This study evaluates Melbourne's longstanding 'activity centres' (AC) policies—the first study to do so. It strongly suggests that, across the Melbourne metropolitan area, AC policies have had no effect on the propensity of people to work near their homes. The findings are robust to a number of validity hazards. The study does not warrant a wholesale abandonment of AC planning, but does signal that we may wish to question how we are currently going about transforming 'places' into 'centres'. For AC policies to be successful, designation as a 'centre' may be necessary, but is not sufficient.

Introduction

This study tests whether metropolitan activity centre (AC) planning actually produces activity centre-like behaviours among commuters. State planning authorities in Victoria, Australia, have been attempting to facilitate the development of urban sub-centres located outside Melbourne's central business district (CBD) since at least 1981 (Melbourne and Metropolitan Board of Works, 1981, pp. 57–9, 63–4). Since then, five successive metropolitan plans have been developed and implemented, with the latest released in 2016. One significant feature of all of these plans has been significant attention paid to developing ACs—concentrated

¹ University of Melbourne, jday@unimelb.edu.au. Sophie Sturup and Jennifer Day conceived the first iterations of this study in 2012. This study could never have happened without the sustained work and support of the Australian Urban Research Infrastructure Network (AURIN) and its staff, including Serryn Eagleson, Chris Pettit, Jack Barton and Bob Stimson. AURIN also funded the data assembly. Thanks also to Pietro Calogero and Stefan Al, along with Dr Sturup, for many useful comments on the writing.

nodes of activity that include jobs, entertainment and residential uses. The logic of this spatial planning strategy is that non-CBD ACs will house people and jobs to promote polycentric urban forms, reducing driving distances and trips to the urban core. In the 2014 strategy document, *Plan Melbourne*, this was described as the ‘20-minute neighbourhood’ concept (Victorian Department of Transport, 2014, p. 11).

A problem with these spatial policies is that they have never been evaluated for their long-term efficacy. This implies that Victorian state governments since at least 1981 have been making spatial policy in metropolitan Melbourne with little empirical evidence that those policies are creating the urban fabric that is the stated intention of the plans.

We do not contest that spatial planning is important for metropolitan regions. Our worry is that the particular type of planning that occurs in Melbourne may not be producing the outcomes that planners envision. Rather than a coherent vision that has persisted over time, Melbourne’s metropolitan planning shifts frequently. The statutory agency responsible for the implementation of metropolitan plans (which are developed by state government planning ministries) has changed names and responsibilities twice since 2008, when it was originally the Growth Areas Authority, charged with developing a strategic approach for new suburbs outside the urban core. In 2013, the agency’s name was changed to the Metropolitan Planning Authority, and it was tasked with implementation of metropolitan planning for the whole metropolitan region including the urban core. In 2016, the agency’s name and scope were again changed: now called the Victorian Planning Authority, it is responsible for state-wide planning implementation, including in the Melbourne metropolitan area.

Melbourne’s AC policies provide a good natural experimental framework from which to test the viability of the policies. The metropolitan spatial plans have seen frequent changes in language and location. For instance, the 1995 *Living Suburbs* strategy (Victorian Department of Planning and Development, 1995) designates 12 ‘Major Mixed Use Activity Centres’, while the 2002 *Melbourne 2030* (Victorian Department of Infrastructure, 2002a, 2002b) designates 25 ‘Principal Activity Centres’ and the 2014 *Plan Melbourne* designates 9 ‘Metropolitan Activity Centres’. The plans vary in the levels of ACs designated, from the CBD to ‘National Employment Clusters’ to ‘Major Car-Based Shopping Centres’ to ‘Specialised Activity Centres’—all which vary with each plan. This shift in spatial designation provides the opportunity to observe places that have changed designations, acquired designation later and lost designation in later plans.

Background

In this literature review, we start with the different attitudes towards AC policies in Melbourne from government and academic perspectives, the disconnect between which highlights a need for long-term evaluative study of metropolitan policies in Melbourne. We then turn to international discourses that reveal a disconnect between policy outcomes and academia, illuminating a need for robust methodologies in studying AC impacts. Lastly, we justify jobs–housing balance (JHB), as used in this study, as an effective measurement to examine the impact of AC policies.

The need for analysis of AC policy

Urban planning researchers and practitioners have long been concerned with the relationship between metropolitan spatial planning and economic outcomes associated with where people work, live and travel. Much of the discussion revolves around whether the decentralising of urban activities into ACs, or ‘sub-centres’, can influence the clustering of employment, housing or commercial activities (see Day et al., 2015). In Victoria, the designation of ACs has been implemented as a component policy of metropolitan planning documents since 1954, although the objectives of these policies have fluctuated over time.

Despite the popularity of ACs in government policymaking, the impact of these policies are not well understood in the literature. Analysing Census-derived journey-to-work (JTW) data, Gordon et al. (1991) find that from 1980 to 1985, there was an association between sub-centre development and decreased commuting time, as residents moved closer to their places of work. However, these findings are contradicted by Cervero and Landis (1991), who find that the suburbanisation of jobs resulted in a threefold increase in commuting time, as measured in vehicle miles travelled. As our previous AC research has posited (Day et al., 2015), statistical associations between urban structures and commuting activities are limited by the inherent heterogeneity of specific urban centres. This is further exacerbated by lack of clear consensus about metrics to study AC outcomes and the subsequent difficulty of aligning findings to applied policy. This study attempts to address this barrier, by proposing a quantitative methodology based on an existing metric: the jobs–housing ratio (JHR). This ratio is adopted and modified in our study, which will be discussed in detail in the following sections.

A further source of disconnect between academic discourse and policy comes from the difficulty in comparing the process of sub-centre development across geographies. While urban planning has traditionally been dominated by North American and European discourses, Melbourne’s distinctly top-down and ‘modernist’ (Mees, 2003, p. 297) approaches to metropolitan planning may in fact draw lessons from the increasing body of literature concerning AC models in Chinese cities.

For example, Monaghan et al. (2015) draw links between heavily regulated AC models in Shanghai and applications for ACs in Australia—the research posits that AC policy can successfully direct commercial activity into concentrated centres without negative impacts on retail competition and heterogeneity. This is in contrast to relatively neoliberal discourses in Australia to ‘remove planning and zoning regulation and encourage out-of-centre development’ (Productivity Commission, 2011, in Monaghan et al., 2015, p. 11).

Only two papers study temporal dynamics in Melbourne: our previous study (Day et al., 2015) which analysed two waves of data in 2006 and 2011 to evaluate the impact of AC policies in employment clustering, and a report (Goodman et al., 2010) looking at housing construction data from 1990 to 2007. There are obvious temporal gaps in these two studies: Day et al. (2015) fail to access data before the introduction of *Melbourne 2030*, and Goodman et al.’s report (2010) was limited to the 5 years following the plan’s release. Day et al. (2015) is the only Melbourne study to attempt to identify whether sub-centring has occurred outside of planned ACs, although these sorts of studies have been conducted in other metropolitan regions like Los Angeles: for example, Giuliano and Small (1991).

A proxy for evaluating AC outcomes: Jobs–housing balance (JHB)

Our literature review thus far suggests that AC policies in Victoria lack substantial directives about how to achieve their proposed objectives, and that there is a lack of adequate data for which outcomes can be measured. Recognising these deficits, we propose the JHB ratio as an appropriate metric for quantifying AC benefits and providing a way for policymakers to benchmark outcomes to objectives.

The JHB metric posits that balanced and sustainable urban sub-centres (or ACs) should have a comparable number of jobs to available housing units within geographic proximity (Schleith et al., 2016). This metric is commonly calculated as a ratio of jobs and housing (JHR), employing Census-derived travel data (origin–destination data, JTW, etc.). A key benefit being that such data is relatively accessible at the AC’s relevant municipal geography (Giuliano, 1991; Peng, 1997). The JHR, an operationalisation of JHB, has been widely adopted in a number of studies (such as Giuliano, 1991; Cervero, 1995; Peng, 1997; Lin et al., 2015). Cervero (1995), for instance, suggests that JHB is indicative of other characteristics of ACs, namely an area’s self-containment; i.e. the extent to which the built form ‘allows many to live, work, shop and recreate within a community or defined geographical area’ (Cervero, 1995, p. 1136).

This line of thinking aligns with that of the Victorian Government, which acknowledges the significance of JHB in its policymaking. *Plan Melbourne* contains strategies designed to promote employment, residential proximity and

shorter commute times. Consequently, this study is formed on the premise that JHB is an appropriate proxy for AC outcomes. We acknowledge that AC policies could have other impacts—as is stated in *Melbourne 2030* and *Plan Melbourne*, key objectives of ACs include stimulating business activities, boosting community activities, improving accessibility to ACs by public transport, etc. However, both plans have attached considerable importance to spatial distribution of activities. In *Melbourne 2030*, the primary objective of ACs is to reduce car dependency by concentrating activities in highly accessible places. Similarly, *Plan Melbourne* stressed the redistribution of jobs outside the central city under its Direction 1.2 (Victorian Department of Infrastructure, 2002a; Victorian Department of Transport, 2014). In the ‘Indicators’ section, we describe how we improve the JHR to provide a better indicator of JHB.

Hypotheses

If metropolitan AC policies are effective, then those ACs should be inducing people to live and commute locally—more so than people who do not live in the influence area of an AC. Compared with places that are not influenced by an AC, then, a place in the *influence area* of an AC should have:

Hypothesis 1: A higher proportion of trips originating in the area, ending in the same area or nearby.

Hypothesis 2: Stronger growth in the proportion of trips originating in the area, ending in the same area or nearby.

We are left, then, to operationalise the ideas of ‘nearby’ and ‘influence area’, as well as JHB. We take up these ideas in the Indicators section.

Data and quasi-experimental design

A significant contribution of this study is that it assembles datasets that have never been assembled and analysed in a peer-reviewed analysis. These are:

1. Panel data on JHB in the Melbourne metropolitan area, assembled from the Census of Population and Housing;
2. Historical panel data on AC locations, derived from examination of individual spatial plans.

To enable a historical look at our problem, we use data from the Australian Census of Population and Housing, which reports place of work data on origin and destination of commute. The Australia Census has occurred every five years since

1981, and metropolitan plans were released in 1981, 1995, 2002, 2014 and 2016. However, the 2016 Census does not contain geographies that concord with prior Censuses, so it is not possible yet to study the impacts of the 2014 and 2016 spatial plans with the Census. We have been unable to match 1981, 1986 and 1991 origin–destination-formatted commute data with commensurate spatial geographies due to missing unique identifiers. This leaves us with the quasi-experimental design shown in Figure 1 to test the hypotheses we outline above:

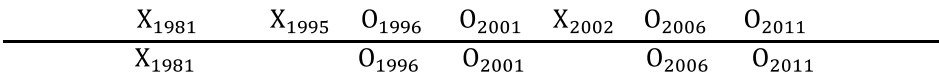


Figure 1. Timeline of quasi-experimental design

Source: Authors.

where O_{yyy} is an observation of JHB in a given year, from the Census data, and X_{yyy} represents the policy intervention in a given year.

Using the concept of ‘influence area’ described below, we operationalise the control group as being those places with no or less area influenced by ACs. We use the Statistical Local Area (SLA) as our unit of analysis. We do this for two major reasons. First, the next-smallest area at which origin–destination data are reported in the Australian Census is the Destination Zone (DZN). DZNs are problematic from a longitudinal perspective because of the lack of concordance we mention above. DZNs are also much smaller than SLAs: in 1996, there were 76 SLAs in the area that is currently designated as the Melbourne metropolitan area according to *Plan Melbourne*. There were 3,157 DZNs. This brings us to the second reason why we use SLAs. Because of its larger geography, the SLA represents a meaningful geography within which a person might commute. We concord all data back to a common set of SLAs produced in 1996 so that data across common areas could be compared. Some of these SLAs were added to the metropolitan area after 1996, but urban designation in 1996 does not matter for our analysis.

Indicators

This section defines the indicators developed for the analysis, including rationales for their selection.

JHB measures

The simplest framings of JHB simply use a ratio of trip origins to trip destinations, with trip origins a proxy for the number of workers living in an area, and trip destinations a proxy for the number of jobs in an area. However, such framings do not account for conditions where there is both a large inflow and a large outflow

of commuters to and from a place. It is possible that most workers who live in an SLA leave the SLA for work, and that other workers arrive from other places. Another problem with this framing is that any commute out of the SLA creates a jobs–housing imbalance—even if the worker commutes a few metres over an SLA boundary. This is a common problem with area-based studies that we mitigate by operationalising the idea of ‘nearby’.

Regarding nearby-ness, because we have origin and destination data, we can do better than a simple metric. We can allow places that are close to a given SLA to contribute to that SLA’s JHB, and for that effect to attenuate with distance. Equation 1 shows the equation we could use to estimate the JHB of a given SLA.

$$JHBO_i = \sum_{j=1}^n \frac{OD_{ji}}{O_i * DIST_{ij}} \quad (1)$$

where OD_{ji} is the number of trips originating in j and ending in i ; for $j \neq i$, $DIST_{ij}$ is the linear distance (in kilometres) between the centroids of SLA_i and SLA_j ; and for $j = i$, $DIST_{ij} = 1$. This is an SLA’s JHBO.

Essentially, Equation 1 measures the propensity of people who live in a given SLA or nearby to also work in that SLA. However, there is a converse measurement of AC behaviour that is also important: that people who work in a given SLA also live in that SLA. We therefore compute another JHB measurement and use it in the analysis, which reflects this converse relationship. This relationship is shown in Equation 2:

$$JHBD_j = \sum_{i=1}^n \frac{OD_{ij}}{D_j * DIST_{ij}} \quad (2)$$

where OD_{ij} is the number of trips originating in i and ending in j ; for $i \neq j$, $DIST_{ij}$ is the linear distance (in kilometres) between the centroids of SLA_i and SLA_j ; for $i = j$, $DIST_{ij} = 1$. This is an SLA’s JHBD.

The measurements contained in Equations 1 and 2 signal a higher JHB if the measure is larger.

ACs and influence areas

The AC locations are derived from the examination of the relevant metropolitan plans. We use the same dataset developed and used by Day et al. (2015) to locate the ACs in the metropolitan space. Our hypotheses describe the ‘influence area’ of an AC. We reproduce their logic in Table 1, which shows the alignment of class types.

Table 1. Classifications of AC types across plans

	Class 1	Class 2	Class 3	Class 4
<i>Plan Melbourne</i> (2014)	Central city (CBD)	Metropolitan activity centre	Activity centre	National employment cluster
<i>Melbourne 2030</i> (2002)	Central activities district (CBD)	Principal activity centres	Major activity centres	Specialised activity centres
<i>Living Suburbs</i> (1995)	Central city (CBD)	Major mixed use activity centre	Major growth area centre	Major education and employment centre
		Major car-based shopping centre		Manufacturing and production concentrations
<i>Metropolitan Strategy Implementation</i> (1981)	Central Melbourne	District centre	Potential district centre	Industrial concentrations
Scenario 1 influence area (km radius)	2.0	0.8	0.4	1.0
Scenario 2 influence area (km radius)	3.2	2.4	1.6	2.4

Source: Authors' summary of Day et al. (2015).

The next step was to operationalise each AC's influence area, which are also shown in Table 1. This is the area around an AC within which we can expect to see AC-like behaviour such as increased JHB. We use the above-mentioned AC centroids and draw buffers at these linear distances to identify the influence area of an AC. We describe how we deal with complexities like overlapping influence areas in the section below, 'AC influence areas'. For robustness, we conduct a sensitivity analysis using two scenarios for impact area:

- Scenario 1: small impact areas
- Scenario 2: large impact areas.

For Scenario 1, we follow Day et al. (2015), identifying influence areas according the AC classes. They cite a government publication that provides guidance on planning ACs, *Structure Planning for Activity Centres, Practice Notes* (Victorian Department of Planning and Community Development, 2010), which notes that neighbourhood ACs should be supported by improved walkability within 400 to 800 metres from the AC's core. With 400 metres as a starting point, Day et al. (2015) identify influence areas for each AC class as per the Scenario 1 influence area in Table 1.

For Scenario 2, we enlarge the impact areas established by Day et al. (2015). Our review of the planning documents concurs with Day et al. (2015), indicating that they provide very little guidance for constructing impact areas. We use *Melbourne 2030*'s language about neighbourhood ACs as a starting point (Victorian Department of Infrastructure, 2002b):

Design the centre and residential areas so that at least 60 per cent of dwellings and activities are within 400 metres safe walking distance from a bus stop, or 800 metres from a railway station (p. 45).

'Neighbourhood activity centre' is an AC type with smaller scale than Class 3 ACs. In *Melbourne 2030*, there are 900 neighbourhood ACs whose impact we do not include in this study due to their numerousness, their very small scale and their presence only in the 2002 plan—not in the 1995 plan or the 2014 plan.

Outside of neighbourhood walkability, the plans provide no guidance on impact areas around ACs. We start with this 800-metre buffer for neighbourhood ACs as a benchmark, and apply proportionally larger buffers to AC designations that would presumably be more impactful. Thus, in Table 1, Class 3 ACs are given an impact radius of twice the 800-metre radius. Class 2 are ACs are given an impact radius of 1.5 times the neighbourhood ACs, and Class 1 (the CBD) has quadruple the impact radius compared with neighbourhood ACs. National employment clusters and their equivalents also receive a buffer of 2.4 kilometres in radius. We do note the arbitrariness of these impact areas. However, as we describe in the Findings section, the size of the impact areas appears to matter very little in the actual JHB computations. Regardless of whether we use large or small impact areas (Scenarios 1 or 2), there is very little link between the metropolitan policies and JHB.

AC influence indicators

This section describes the SLA-based indicators of 'influence'. Because the unit of analysis (SLA) is a spatial polygon, there are a number of ways in which we can operationalise the impact of ACs on that polygon. For instance, we could conceive indicators that reflect whether the centroid of an AC is located in a given polygon. This ignores the possibility that an AC could have its centroid in one polygon but significantly affect JHB in a neighbouring polygon. We could measure whether an AC touches a polygon, assigning a dummy value of one if the AC is touched by the buffer around an AC and a value of zero if it is not. However, this process would count SLAs that are 90 per cent covered by an AC the same as SLAs that are 1 per cent covered—though they are very likely not influenced similarly. Also, because of the large number of ACs outlined in the policy, there is often overlap between the influence area of one AC and another.

Our final indicator is the proportion of total land area of an SLA covered by AC influence areas, with overlapping areas double-counted as necessary. This means that the indicators for influence can range from zero to more than 100 per cent. We do not apply weights to the different AC classes, since their influence is reflected in their spatial scope (400–2,000 metres) and applying a weight would amount to double-counting.

Analysis methods

We produce a variety of summary statistics and statistical tests, and we also estimate a series of simple panel models to test the hypothesis that there is a relationship between AC designation and JHB. We use a suite of models in order to rule out the possibility that an inappropriate (biased or inefficient) estimator is the reason we see no link between AC designation and AC-like behaviour among commuters. We start with a simple ordinary least squares (OLS) model with errors clustered on the SLA—a process that partially accounts for the persistence of JHB in an SLA over time. We next estimate the fixed effects estimator, which allows us to control for time-invariant effects such as distance from the urban core. We estimate a second model, an OLS model with a lagged *Coverage* variable, again with clustered error terms. We also estimate two fixed effects models with lagged coverage variables: a standard fixed effects model with bias correction, which Kiviet (1995) argues is a suitable way to handle dynamic panel bias (correlation between the dependent variable and lagged dependent variables used as regressors) and a generalised method of moments model. None of these models contains time-variant control variables such as public transportation accessibility or average age of housing stock, which could have some explanatory power. We leave out further controls because the purpose of this paper is to explore these simple relationships between AC designation and JHB—not to uncover the forces driving the creation of JHB. For readability, we do not fully describe those models here, nor do we report the detailed findings. The models are more fully specified and justified in a working paper that is currently under review (Day et al., 2017).

Findings

This section describes the data we have compiled, presents a descriptive narrative of the story it tells and reports the results of the regression analysis. We conclude that AC policy has had no observed effect on JHB.

AC profiles

The total number of AC designations has increased over the four policies, from 20 in 1981 to 45 in 1995. In 2002, there was a boom in the number of ACs, with 115 centres designated that are of comparable size to those in the other plans. In 2002, *Melbourne 2030* additionally designated 900 ‘neighbourhood ACs’, which we do not include in our analysis because their scale is too small. There were 133 ACs designated in 2014.

Coverage

This section describes the *Coverage* data. Figures 2 to 4 illustrate the influence area in the map for the 1981, 1995 and 2002 policies. As more ACs were designated in the 1995 and 2002 plans, the total influence area has increased over time. Thirteen SLAs were excluded as they were at the edge of the study area and not officially designated as part of the metropolitan area in 1981. Another five SLAs were excluded for missing data of JHB indicators in some years. This leaves us with 69 SLAs of the 76 SLAs in metropolitan Melbourne in 1996.

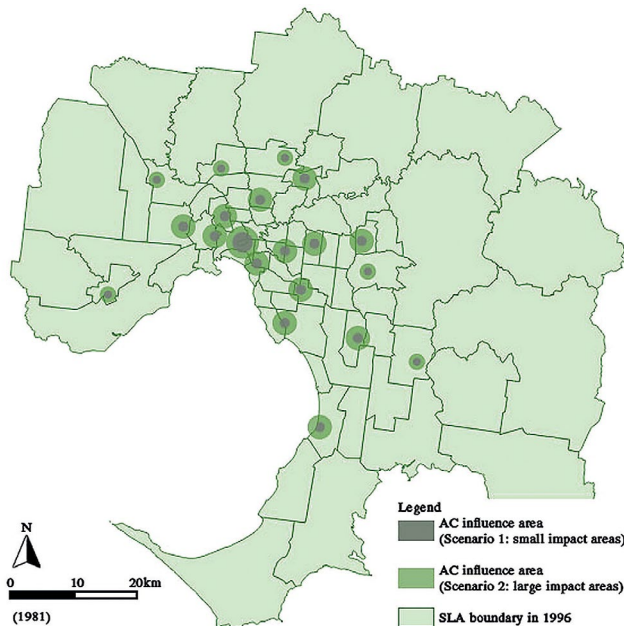


Figure 2. AC influence area for the 1981 policy

Source: Authors' summary of Melbourne and Metropolitan Board of Works (1981, pp. 57–9, 63–4).

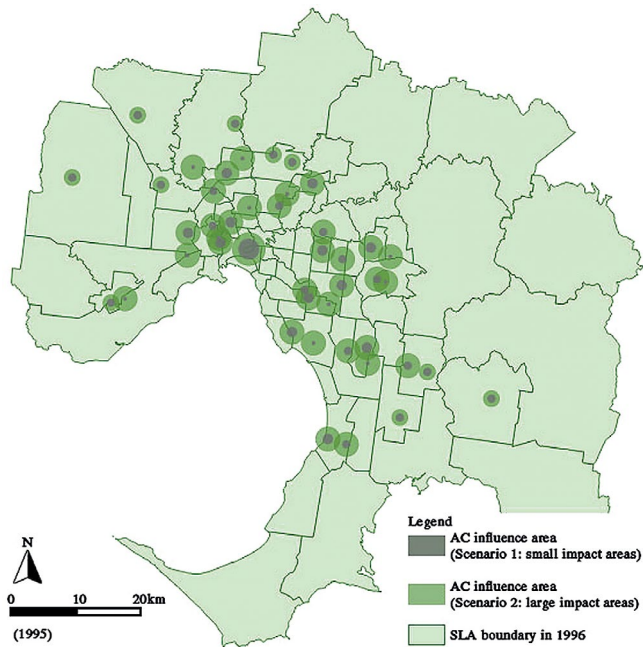


Figure 3. AC influence area for the 1995 policy

Source: Authors' summary of Victorian Department of Planning and Development (1995, p. 68).

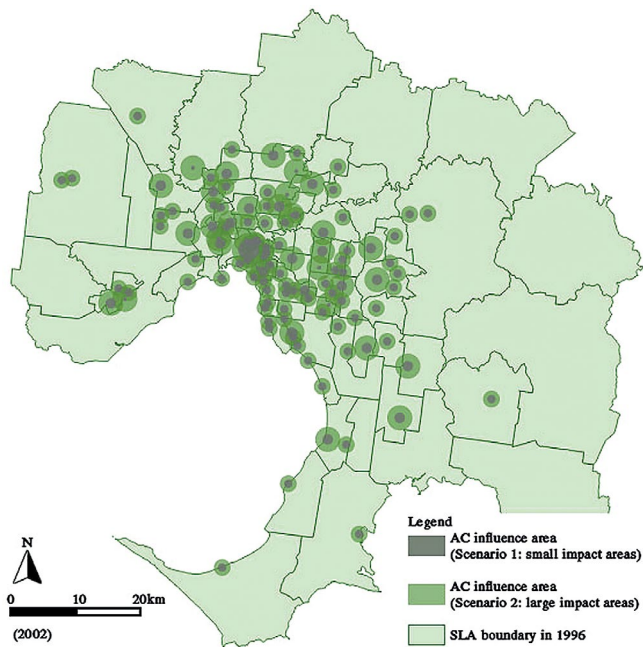


Figure 4. AC influence area for the 2002 policy

Source: Author's summary of Victorian Department of Infrastructure (2002a, pp. 46–51).

The total land area of metropolitan Melbourne was 7,453,381,888 m² in 1996. For constancy of comparison, we use the 1996 SLA boundaries throughout this analysis, collapsing all data to these geographic units. In 1981, under the impact scale of Scenario 2, 4.09 per cent of the total land was within an AC influence area. That proportion rose to 7.78 per cent in 1995 and then to 11.40 per cent in 2002. Impacts were smaller for Scenario 1, ranging from 0.87 per cent in 1995 to 3.2 per cent in 2002.

In aggregate, the *Coverage* ratio of SLAs increased dramatically between 1981 and 2002. Because of the growth in the total AC numbers, there was overlap in the AC areas. We dealt with this by double-counting the overlapping areas of two or more ACs.

JHB profiles

This section describes the change in JHB in the Census data years. Table 2 shows that the mean of JHBO (see Equation 1) declined marginally in 2006 and 2011 from 1996 and 2001 levels. The range and standard deviation, growing in 2011 and decreasing afterwards, suggests an overall decline in JHB in the metropolitan area. For JHBD (see Equation 2), the overall profile is fluctuating but relatively steady.

Table 2. Summary of statistics of JHBO/JHBD, 1996–2011

Indicator	Year	N	Mean	Std. Dev.	Median	Range
JHBO	1996	69	0.320	0.0993	0.302	0.113–0.642
	2001	69	0.321	0.107	0.301	0.0482–0.710
	2006	69	0.313	0.0933	0.299	0.145–0.660
	2011	69	0.309	0.0946	0.292	0.159–0.691
JHBD	1996	69	0.416	0.168	0.363	0.114–0.798
	2001	69	0.424	0.162	0.404	0.118–0.788
	2006	69	0.415	0.157	0.377	0.144–0.775
	2011	69	0.427	0.163	0.402	0.152–0.845

Source: Authors' computations on Australian Census data.

Figure 5 shows a selection of profiles using the JHBD metric and provides a few examples of SLAs that take on this pattern. (We do not provide JHBO as the trends are similar.) We note that we have not systematically identified profile types, as this study is about large overall trends and their relationship to AC influence. Our purpose here is to illustrate the features of the data we use for the analysis. The statistical tests we report later in this section seek to understand whether JHB variations systematically follow changes in AC designation. Among the profile types we identify are (with sample SLAs named in parentheses):

- Early increase—stronger increase in the earlier years of the study period followed by weaker change in later years (Hume).
- Early decrease—stronger decrease in the earlier years of the study period followed by weaker change in later years (Melton).
- Late change—a change in the trend near the end of the study period (Brimbank–Sunshine).
- Steady—varies over a narrow band of JHBD values (Cardinia South).
- Steady rise—steadily increasing JHBD values (Melbourne non-CBD, non-Southbank/Docklands).
- Flatline—JHBD values are virtually unchanged over the study period (Kingston North, Nillumbik South).
- Temperamental—JHBD values vary seemingly without pattern and over a wide band (Whittlesea North, Nillumbik).

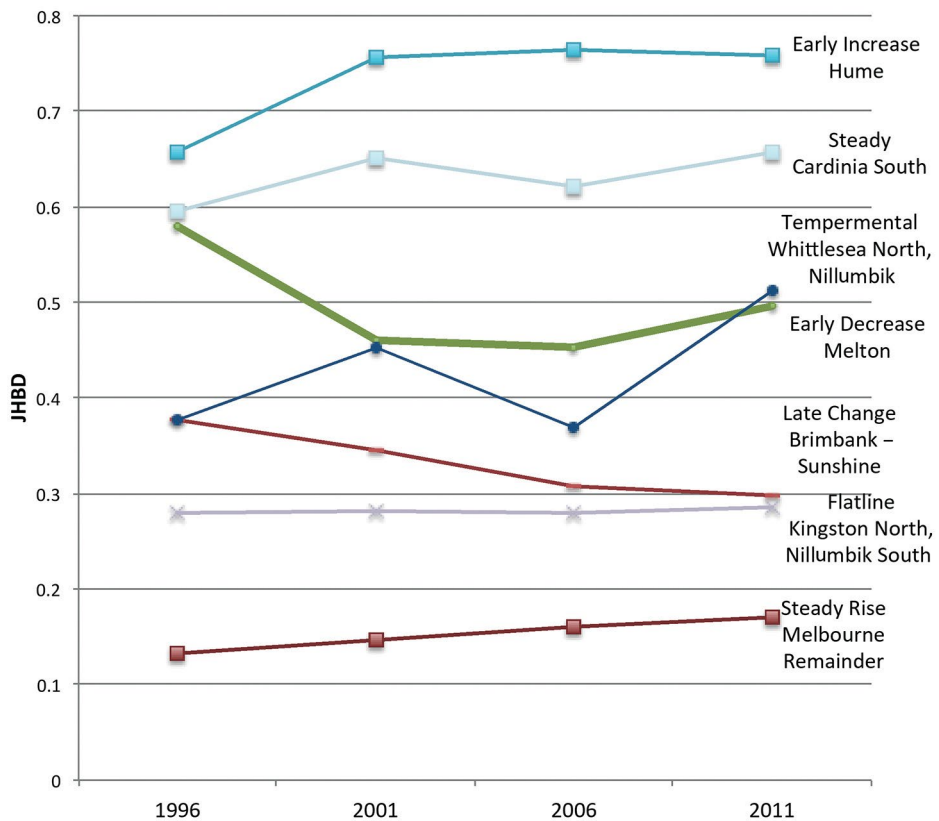


Figure 5. JHBD profiles, 1996–2011

Source: Author's computations on Australian Census data.

Association between AC influence and JHB

This section tests for the effect of any *Coverage* at all on JHB measures. We divide the 69 SLAs into two groups according to whether they were touched by AC influence area—that is, whether any portion of the SLA is within an AC influence area. Before 2002, there were 52 SLAs influenced by an AC. After 2002 and *Melbourne 2030*, 10 more SLAs were touched by an AC influence area. This section suggests that AC policies have produced at best no effect on JHB and at worst an effect that is counterproductive. Either way, the observed statistical direction of the relationship does not suggest policy efficacy.

We use a two-sample t-test to test whether there is significant difference between the mean JHBO and JHBD of AC influence and non-AC influence groups in each policy year. Table 3 shows the mean JHBD and JHBO values for SLAs in each category. In every test year, the mean JHBO of the AC influence group is marginally lower than non-AC influence group—suggesting that SLAs with no AC influence have higher jobs-housing balance. This relationship is only statistically significant in the year 2001, and the p-value is slightly higher than 0.05. These results do not suggest a strong and sustained difference between influenced and non-influenced SLAs in any of the data years.

Table 3 also shows the same relationships for JHBD, and the results are even more troubling. In each data year, the JHB of the non-influenced group is higher than the JHB of the influenced group. This effect is exactly the opposite of the effect we would expect if the AC policies were increasing JHB. These differences in the two metrics also appear in the regression results in the next section, suggesting that precision in the measurement of JHB is important for evaluating AC policies. There, we also address the possibility that ACs may have been created in order to improve JHB in low-JHB places.

Table 3. Two-sample t-test of JHBO/JHBD

Year	JHBO				JHBD			
	1996	2001	2006	2011	1996	2001	2006	2011
N of SLAs (influenced/non-influenced)	52/17	52/17	62/7	62/7	52/17	52/17	62/7	62/7
AC influence	0.311 (0.0125)	0.307 (0.0125)	0.308 (0.0119)	0.306 (0.0121)	0.370 (0.0190)	0.392 (0.0186)	0.391 (0.0179)	0.402 (0.0181)
Non-AC influence	0.347 (0.0299)	0.365 (0.0342)	0.358 (0.0315)	0.341 (0.0345)	0.558 (0.0433)	0.521 (0.0487)	0.621 (0.0556)	0.652 (0.0617)
Difference	0.0360 (0.0276)	0.0575 (0.0292)	0.0504 (0.0370)	0.0350 (0.0378)	0.188*** (0.0413)	0.129*** (0.0426)	0.229*** (0.0564)	0.250*** (0.0577)
p-value	0.1969	0.0531	0.1772	0.3576	0.0000	0.0034	0.0001	0.0001

(Std. Err.); *** p<0.01

Source: Authors' computations on Australian Census data.

Regression analyses

This section describes the regression results, wherein we test the associations between *Coverage* and JHB. We first ran the regressions on the full set of SLAs. To test the possibility that low-JHB places may have been designated as places to receive AC designation in order to encourage JHB there, we ran a second set of regressions to look at only those places that had the most to gain from AC policies.

We do not present the statistical findings in this section. To keep the paper interesting to a wide audience, we have placed the regressions in a working paper (Day et al., 2017).

The full-sample regression results reveal no supportable relationship between *Coverage* and JHB. The results further suggest that AC policies have not produced either form of JHB (JHBO or JHBD) in those areas where JHB was not present initially. This finding is robust across model specifications and estimators used, so the effect is likely to be a true effect and not a result of statistical misspecification or poorly constructed indicators. We test a variety of model specifications, including models using the Scenario 1 and Scenario 2 impact areas and models for both Scenarios with fixed and lagged effects. The estimated models for Scenario 1 (small impact areas) and Scenario 2 (large impact areas) are qualitatively identical. That is, they are identical across all models estimated in sign and are very similar in the magnitude of the significances and model fit statistics. This suggests that the results are robust to changes—even significant changes—in the *Coverage* indicator.

The results for the initially underperforming (low-JHB) places are even more discouraging for AC policies. For these SLAs, the results suggest that AC policies have at best produced no effect, and at worst have somehow been counterproductive to JHB.

Conclusions: Can AC policies be reinvigorated?

Our findings unfortunately suggest that ACs policies in Melbourne are not promoting JHB. The JHBO tests are very clear: AC policies appear to have no effect on the propensity of people to work nearby their homes. The JHBD tests are harder to interpret, but they also suggest that AC policies have at best had no effect on encouraging people to live nearby their workplaces. At worst, AC-influenced places are worse off when it comes to JHB. Moreover, those places that did not start with AC-like behaviour have not grown to be more AC-like. The State of Victoria dedicates a government agency to develop metropolitan activity spatial plans, including AC plans, and AC planning has been going on since 1981. This is a significant resource base dedicated to planning that may not be achieving the desired outcomes.

Our results are robust to a number of threats to validity. Given the quasi-experimental design and the sensitivity analysis, the results appear to be robust to both unspecified temporal effects and variations in the *Coverage* measurement. Our results also do not suffer from bias associated with the spatial definitions or spatial autocorrelation, as we have incorporated nearby-ness into the indicators for JHB. Our results observe the system back to 1996, so it is not likely that AC effects have yet to materialise to a sufficient degree to be observable. In short, these findings are ones that policymakers should take seriously, especially given the resources that are devoted to metropolitan planning in Victoria.

What this research demonstrates is there is probably more to creating vibrant ACs than just designating places as ACs. We also know that the state government provides varying levels of consultation and support in structure planning for localities that seek to cooperate with the metropolitan plans and develop ACs. Our study does not reflect significant issues like local capacity and uptake of offered assistance—again due to a lack of available data on these processes. We propose extensive follow-up research to confirm our findings and expand the analysis to include queries on what it is that drives AC success. Some next steps may be to conduct comparisons of places that have been given support, like Dandenong, to places that have received less support. Unfortunately, changes to the 2016 Census formats mean that further follow-up will not be straightforward using the same indicators. The upcoming 2016 Census data does not have the SLA as a proffered spatial unit of analysis. Australia's public agencies should consider a continuity of data that allows for ongoing, refreshed examination of the policy outcomes, but at present appears to have no plans to do so.

Our study does not imply that a wholesale abandonment of AC planning is warranted. It does, however, signal that we may wish to question how we are currently going about making places into centres. We hope that future studies will demonstrate that some AC planning has indeed been fruitful under some conditions. This future research will reveal what needs to happen in places where AC behaviour is not yet occurring. For AC policies to be successful, designation is perhaps necessary, but it is not sufficient.

References

- Cervero, R. (1995). Planned communities, self-containment and commuting: A cross-national perspective. *Urban Studies*, 32(7), 1135–61. doi.org/10.1080/00420989550012618
- Cervero, R. & Landis, J. (1991). *Suburbanization of jobs and the journey to work*. UC Berkeley: University of California Transportation Center.
- Day, J., Han, W., Wu, A.B. & Zheng, J.R. (2017). *Does policy invoke sub-centres? Econometric evaluation of Melbourne's urban spatial planning since 1996*. (Working Paper).

- Day, J., Sturup, S., Chen, Y., Budahazy, M., Wu, A. & Fan, L. (2015). *Activity centre policy effects on employment clustering: A spatial study of job density in Melbourne, Australia*. Paper presented at the State of Australian Cities National Conference, Gold Coast, Queensland, Australia.
- Giuliano, G. (1991). *Is Jobs-Housing Balance a Transportation Issue?* UC Berkeley: University of California Transportation Center.
- Giuliano, G. & Small, K.A. (1991). Subcenters in the Los Angeles Region. *Regional Science and Urban Economics*, 21(2), 163–82. doi.org/10.1016/0166-0462(91)90032-I
- Goodman, R., Buxton, M., Chhetri, P.R., Taylor, E.R. & Wood, G.R. (2010). *Planning and the characteristics of housing supply in Melbourne*. Melbourne: Australian Housing and Urban Research Institute (AHURI).
- Gordon, P., Richardson, H.W. & Jun, M.J. (1991). The commuting paradox evidence from the top twenty. *Journal of the American Planning Association*, 57(4), 416–20. doi.org/10.1080/01944369108975516
- Kiviet, J.F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, 68, 53–78. doi.org/10.1016/0304-4076(94)01643-E
- Lin, D., Allan, A. & Cui, J. (2015). The impact of polycentric urban development on commuting behaviour in urban China: Evidence from four sub-centres of Beijing. *Habitat International*, 50, 195–205. doi.org/10.1016/j.habitatint.2015.08.018
- Mees, P. (2003). Paterson's Curse: the attempt to revive metropolitan planning in Melbourne. *Urban Policy and Research*, 21(3), 287–99. doi.org/10.1080/0811114032000113671
- Melbourne and Metropolitan Board of Works (1981). *Metropolitan Strategy Implementation*. Melbourne.
- Monaghan, A., Burke, M. & Pan, H. (2015). *The Shanghai Model for Activity Centres and its Potential in the Australian Context*. Paper presented at the State of Australian Cities National Conference, Gold Coast, Queensland, Australia.
- Peng, Z. R. (1997). The jobs-housing balance and urban commuting. *Urban Studies*, 34(8), 1215–35. doi.org/10.1080/0042098975600
- Productivity Commission (2011). *Economic structure and performance of the Australian retail industry*. Canberra: Commonwealth of Australia.
- Schleith, D., Widener, M. & Kim, C. (2016). An examination of the jobs-housing balance of different categories of workers across 26 metropolitan regions. *Journal of Transport Geography*, 57, 145–60. doi.org/10.1016/j.jtrangeo.2016.10.008
- Victorian Department of Infrastructure (2002a). *Melbourne 2030: Planning for Sustainable Growth*. Melbourne.

HAS SUB-CENTRE POLICY PRODUCED SUB-CENTRES?

Victorian Department of Infrastructure (2002b). *Melbourne 2030: Planning for sustainable growth—Draft Implementation Plan 4 Activity Centres*. Melbourne.

Victorian Department of Planning and Community Development (2010). *Structure Planning for Activity Centres, Practice Notes*. Melbourne.

Victorian Department of Planning and Development (1995). *Living Suburbs: A Policy for Metropolitan Melbourne into the 21st Century*. Melbourne.

Victorian Department of Transport (2014). *Plan Melbourne: Metropolitan Planning Strategy*. Melbourne.

The Impact on Research Quality of Performance-Based Funding: The Case of New Zealand's PBRF Scheme

Robert A. Buckle and John Creedy¹

Abstract

This paper appraises the impact on the research quality of New Zealand universities of the Performance-Based Research Fund (PBRF), a peer-review process that assesses individual researchers. The paper identifies the contribution to improvement in research quality arising from transitions among research quality categories, and entrants and exits of individuals. It traces a substantial component of change from 2003 to 2012 to the removal of non-research active staff. It also finds a large reduction in the number of younger researchers, and population ageing due to retention of older and higher-quality researchers. Significant differences among universities are evident in the patterns of transformation. The paper also critically considers the PBRF assessment process and characteristics of the metrics used, suggesting scope for improvement in the assessment of researchers and the way in which universities are ranked.

1. Introduction

The Performance-Based Research Fund (PBRF) was introduced in New Zealand in 2003 as a method of allocating research funding to tertiary education organisations on the basis of research performance, rather than the number of students.² This is

1 Victoria University of Wellington, bob.buckle@vuw.ac.nz and john.creedy@vuw.ac.nz. We are grateful to Amber Flynn, Morgan Healey and Sharon Beattie of the New Zealand Tertiary Education Commission (TEC) for helpful discussions, and the TEC for providing the anonymous PBRF data requested by the authors. We should also like to thank William Coleman and a referee for their helpful comments on an earlier draft of this paper.

2 For background and detailed discussion of the PBRF, see New Zealand Tertiary Education Commission (2002, 2013) and Ministry of Education (2012).

part of a worldwide emergence of performance-based evaluation schemes designed to underpin funding of tertiary institutions (see Coryn, 2007; OECD, 2010; Jones & Cleere, 2014; de Boer et al., 2015; Wilsdon et al., 2015). Similar schemes were introduced earlier in the United Kingdom, Australia and Hong Kong, and subsequently in Denmark, Norway and Sweden and in other countries.³ These schemes vary by coverage and assessment method, which may be bibliometric or peer review-based. The New Zealand scheme uses a peer-review assessment method and assesses individuals rather than groups.⁴

The Tertiary Education Commission (TEC) explained the aims of the PBRF process as follows:

The primary purpose of the Performance-Based Research Fund (PBRF) is to ensure that excellent research in the tertiary education sector is encouraged and rewarded. This entails assessing the research performance of tertiary education organisations (TEOs) and then funding them on the basis of their performance.⁵

The purpose of this paper is to contribute to understanding the potential effects of these schemes by summarising the results of a research program analysing the impact on universities of the New Zealand PBRF scheme. The results can contribute towards the wider debate on performance-based schemes for which, as de Boer et al. (2015, p. 5) conclude, ‘there still is not sufficient evidence on the effects of the systems and ... our understanding of the proper design and implementation of performance agreements is still incomplete’.

The PBRF has changed the incentives facing universities and individual researchers. It is important to examine both the nature of those incentives and their potential effects. Improvement in a university’s research quality—assessed as some form of average of individual researcher qualities—can come about for three reasons. First, changes arise from recruitment and retention of high-quality staff; second, from encouraging the exit of lower-quality staff; and third, from converting low-quality researchers to high-quality researchers. Each of these paths has different ‘effective prices’ and faces various constraints. The speed of transformation is affected by the supply of high-quality researchers and the form of university contracts, including the flexibility in pay scales. The effective prices and constraints are likely to differ among universities. It is neither feasible nor desirable to have a university consisting

3 See also Jonkers and Zacharewicz (2016) for a comparison of a wide range of schemes, and Roberts (2005) on the Australian scheme.

4 In their summary of the literature regarding research evaluation and metrics, Wilsdon et al. (2015, p. viii) argue, ‘Peer review is not perfect, but it is the least worst form of academic governance we have’. An example of how the use of bibliographic metrics can affect research evaluation results in the context of the PBRF is provided by Tressler and Anderson (2012, p. 17), who conclude ‘that the time-lags between publication and citation make it difficult to rely on citation counts to produce a meaningful measure of output in a PBRF-like research-evaluation framework, especially one based on the assessment of individual academics’. See Davidson (2013) for a different perspective in the context of the Australian ERA system.

5 See www.tec.govt.nz/funding/funding-and-performance/funding/fund-finder/performance-based-research-fund/.

only of researchers judged by the PBRF process to be of the highest quality at a given date: the balance between high-quality and low-quality researchers is also likely to vary between universities. This paper explores the possible implications of PBRF by examining the nature of the relevant flows and their contribution to changes in New Zealand universities' average research quality.

Others have endeavoured to assess the impact of PBRF by comparing publication measures before and after the introduction of PBRF.⁶ However, no previous work has utilised the data on individual researcher performance produced by the PBRF process itself. The insights summarised in this paper are derived using a database consisting of an anonymous 'quality category' for each individual assessed in each of the three PBRF assessment rounds in 2003, 2006 and 2012. The 2006 round was a partial round and hence the results discussed in this paper refer to the changes between 2003 and 2012.⁷ This database, which was provided by the TEC under a confidentiality agreement and is not publicly available, includes the quality category, age, discipline and university of each individual assessed from 2003 to 2012.

The paper is structured as follows. Key features of the process used to assess researcher quality are explained in Section 2. Section 3 reports changes in the number of researchers between 2003 and 2012. A summary of the change in the quality of researchers at New Zealand universities since the introduction of PBRF is provided in Section 4. This section also discusses the relative performance of each university. Section 5 examines the contribution of exits and entrants, along with transitions among quality categories at the aggregate level. Section 6 evaluates these transitions at the level of each university, revealing how they vary. Although information about management practices within universities is not available, a number of reasons for these different responses are discussed. The associated change in the age distribution of researchers is discussed in Section 7. The metrics used in the PBRF are critically evaluated in Section 8. Conclusions and policy issues are discussed in Section 9.

6 Examples of this type of work as applied to the economics discipline include Gibson et al. (2008), Anderson et al. (2013), and Anderson and Tressler (2014). For examinations of a wider range of disciplines, see Smart (2009), who concluded that research output and quality increased for most New Zealand universities after the introduction of PBRF; Hodder and Hodder (2010) for New Zealand, and Butler (2003) for Australia, who concluded that research quantity increased but quality declined. A critique of Butler is provided by van den Besselaar et al. (2017), who find that both quantity and quality of research has increased, and refer to other studies showing a positive relationship between research quantity and impact.

7 In 2006, universities could choose to submit a new portfolio for researchers covering the previous six years of research, or retain the quality category of those researchers who submitted portfolios in 2003 covering the six years prior to 2003.

2. Key features of the PBRF process

The PBRF uses three measures to allocate government funding to support research at universities and other New Zealand TEOs; these are ‘Quality Evaluation’, ‘Research Degree Completions’ and ‘External Research Income’. The quality categories form the basis of the Quality Evaluation component, which comprises 60 per cent of PBRF funding.⁸ Furthermore, the funding per quality category varies across disciplines.

The allocation of the Quality Evaluation component of PBRF funding is based on information submitted by each university researcher, which is referred to as an ‘evidence portfolio’. The information provided by the evidence portfolios is used to assess the research quality of each researcher. The evidence portfolio is evaluated by relevant subject expert panels to eventually assign each evidence portfolio to a quality category.

The subject panel assigns a score from 0 to 7 for each of three components of the evidence portfolio: ‘research output’, ‘peer esteem’ and ‘contribution to research environment’.⁹ These three components are given scores, s_j , which are then given weights, q_j , of 0.70, 0.15 and 0.15. The total weighted score, s , for an individual is obtained by multiplying the weighted sum of the s_j values by 100. Hence:

$$S = 100 \sum_{j=1}^3 q_j s_j \quad (1)$$

The maximum individual score is therefore 700. A letter grade, indicating the quality category, is then assigned depending on the assessed total. These are as follows: R for scores 0 to 199; C for scores between 200 and 399; B for scores from 400 to 599; and A for scores from 600 to 700.¹⁰ A numerical score, G , is then assigned to each letter grade: 10 for an A; 6 for a B; 2 for a C; and 0 for R.¹¹ These categories are described further in Ministry of Education (2012, p. 21). When individuals move between institutions, all their quality score goes to the new institution, irrespective of how long they have been at the new university and how much has been invested in them by the previous employer.¹²

⁸ Since 2016, the percentage allocations have changed to 55:25:20 respectively.

⁹ The TEC guidelines refer to these scores as ‘unweighted individual component scores’.

¹⁰ Only research output from the previous six years qualifies for assessment.

¹¹ The recognition that new researchers may take time to establish their research, publications and academic reputations led to the introduction in 2006 of new categories, C(NE) and R(NE), although having the same numerical score as C and R respectively.

¹² However, if they move within a year of the census date, the funding is divided between the two relevant institutions.

3. Changes in the number of researchers assessed by PBRF

A number of factors have contributed to changes in the number of portfolios submitted by universities between 2003 and 2012. The period since 2003 coincides with changes in policy with respect to former Colleges of Education. Following the 2003 PBRF round, the Auckland, Christchurch, Dunedin and Wellington Colleges of Education were subsumed within the University of Auckland, University of Canterbury, University of Otago and Victoria University of Wellington (VUW) respectively.¹³ The Colleges of Education at Hamilton and Palmerston North had been amalgamated with University of Waikato and Massey University respectively prior to the 2003 PBRF. These amalgamations in turn triggered reforms of the former colleges and significant reductions in staff, particularly at the Auckland and Wellington colleges. Another factor influencing the change in the number of researchers assessed was the change in 2012 to the PBRF process. In the 2003 and 2006 PBRF rounds, all eligible staff were assigned a quality category and if an evidence portfolio was not submitted, the staff member was assigned a quality category of R.¹⁴ At a late stage in the 2012 process, the TEC changed the rules such that only submitted portfolios were assigned a quality category. Hence universities could choose not to submit portfolios for those who were expected to be assigned a quality category of R. This affected the calculation of average quality scores (AQS), as discussed further below.

The TEC publishes aggregate university data on an employment-weighted basis, shown in Table 1.¹⁵ The published aggregates show an increase in employment-weighted evidence portfolios between 2003 and 2012. In 2003, in particular, New Zealand universities employed a considerable number of people on a part-time basis, the extent of which varied among universities.

Table 1 also shows the number of employment-weighted evidence portfolios and total non-administration staff for each university in 2003 and 2012. The corresponding ratios are shown in Figure 1. The ratio for the entire university system increased from 0.58 in 2003 to 0.73 in 2012. The ratios varied substantially in 2003. Auckland University of Technology's (AUT's) ratio in 2003 was only 0.15 while VUW's ratio was 0.76. All universities increased this ratio by 2012, and particularly AUT, for which the increase was threefold. The rise in the ratio of employment-

13 The staff numbers for the Universities of Auckland, Canterbury and Otago and VUW used in this paper include the staff from these respective Colleges of Education.

14 However, in 2006, those who submitted a portfolio in 2003 did not need to submit a new one if they took the view that their quality category had not changed. In those cases, they were assigned their 2003 category.

15 The aggregates therefore give the numbers of 'full-time equivalent persons'.

weighted portfolios to total employment-weighted non-administration staff implies that universities reduced the proportion of non-administration staff who were not research active.¹⁶

Table 1. Number of FTE (full-time equivalent) evidence portfolios (N_p) and number of non-administration staff (N_T)

University	2003		2012	
	N_p	N_T	N_p	N_T
AUT ^a	135.3	892.7	429.5	952.1
Lincoln	139.1	221.1	174.1	250.9
Massey	689.3	1326.8	918.6	1316.2
Auckland	1152.5	1669.7	1556.1	2023.8
Canterbury	497.7	695.1	617.3	661.2
Otago	845	1297.8	1168.2	1567.5
Waikato	369.8	685.7	440.6	601.3
VUW ^b	459.9	606.4	641.5	779.9
Total	4288.6	7395.3	5945.9	8152.9

Note: (a) Auckland University of Technology; (b) Victoria University of Wellington.

Source: Data from New Zealand Tertiary Education Commission (2013).

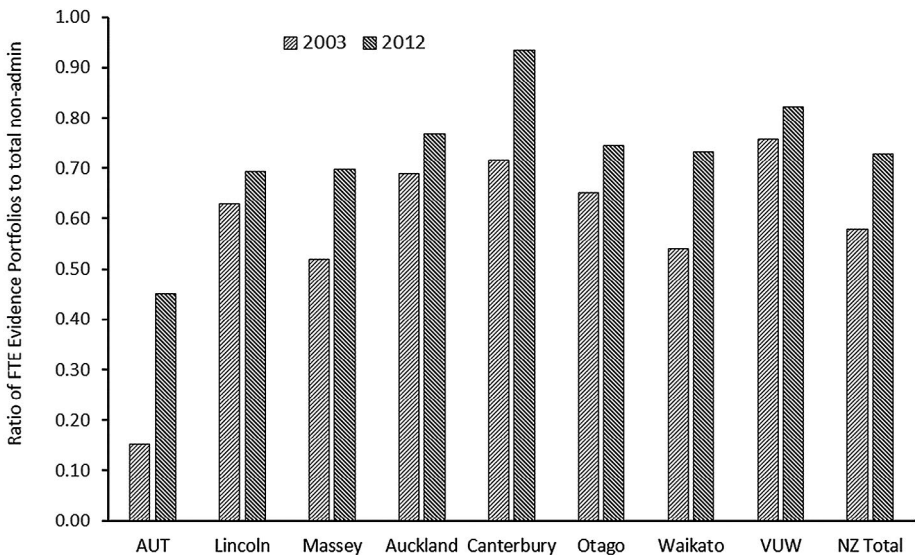


Figure 1. Ratio of FTE evidence portfolios to total number of staff

Source: Authors' calculations using the data in Table 1.

16 Using as a measure of 'research active' those individuals publishing at least one article, in whole or in part, during a six-year period in an EconLit-listed journal, Anderson and Tressler (2014, p. 7) concluded that the proportion of research active academic staff increased from 71 per cent in 1994–99 to 79 per cent in 2000–05 and 83 per cent in the period 2006–11.

4. Changes in AQSs for New Zealand universities

This section reports changes between 2003 to 2012 in the average quality scores, AQSs, of the eight universities, derived as follows. Define the employment weight of person i as $e_i \leq 1$ and let n denote the relevant number of employees in a university.¹⁷ The choice of n has an important effect and is discussed in more detail below. The AQS is:

$$AQS = \sum_{i=1}^n e_i G_i / \sum_{i=1}^n e_i \cdot \quad (2)$$

As explained in Section 2, the value for each researcher, G , ranges from 0 to 10, depending on the individual's quality category. Hence, a university's AQS can range from 0 to 10.¹⁸

The TEC produced a range of AQS measures, depending on the choice of denominator. One measure uses the employment-weighted total of all researchers for whom PBRF portfolios were submitted, while another measure excludes R-type researchers. Since the grade for R-type staff is equal to zero, their inclusion affects only the denominator in equation (2). The change in the rules in 2012, which allowed universities to withhold submission of the portfolios of potential R-type researchers, means that the AQSs published by the TEC for 2012 are not necessarily comparable with those for 2003.¹⁹

However, an alternative preferred measure can be derived by using the total number of non-administration staff as the denominator in equation (2). The use of this denominator avoids the problems of comparability mentioned in the previous paragraph. An AQS based on all non-administrative staff can be obtained by multiplying the AQS, which includes all portfolios, including R-researchers and based on the unweighted TEC data, by the ratio of the employment-weighted number of portfolios to the total number of non-administrative staff. The latter ratio is obtained from the published data in Table 1. The details of this calculation are explained in Buckle and Creedy (2018).²⁰

17 As mentioned above, when using the anonymous TEC dataset, the employment weights are all set to unity.

18 There is an inconsistency in the terminology used by the TEC. The term 'Total Weighted Score' refers, for each individual, to the weighted sum (multiplied by 100) of the scores given for the three components of research quality. The term 'Average Quality Score' is not in fact an average of *scores*, but is a weighted average of the *values* attached to individual quality categories, where the latter are determined by the individual's position in the distribution of Total Weighted Scores.

19 Two additional measures include as the denominator either the sum of effective full-time students or the sum of postgraduate students. These are affected by the discipline mixture and in particular by student-staff ratios, and hence do not necessarily reflect average research quality of staff.

20 The AQS values published by the TEC use employment-weighted data for both numerator and denominator, but they do not report AQSs based on all non-administrative staff.

Table 2 shows AQSs derived on this basis for each university and all universities combined, and indicates that all universities have improved.²¹ However, there are substantial differences in the rates of improvement across the eight universities. The proportional growth rates for each university are shown in Table 3. The final row of this table shows the extent to which the growth in each AQS exceeds that which would arise in a scoring system under which all non-R researchers are given the same score. It can be seen that, for all universities combined, the change in the distribution of C, B and A researchers has contributed only an additional 9 per cent to the growth of the AQS. This appears to suggest that there has been little overall improvement over and above the removal of a large number of non-research active staff and, to a lesser extent, the transition of former Rs to higher-quality categories.

Table 2. AQSs based on all non-administration staff: 2003–12

University	2003	2012
AUT	0.110	1.364
Lincoln	1.579	2.458
Massey	1.071	2.878
Auckland	2.457	3.736
Canterbury	2.532	4.262
Otago	2.002	3.555
Waikato	1.579	3.145
VUW	2.310	4.374
All	1.733	3.307

Source: Authors' calculations using the TEC dataset.

However, there has been considerable variation among universities. For example, in VUW the change in the quality composition of researchers (arising both from recruitments and transformations) has produced a growth in its AQS that is 83 per cent higher than it would be under a scoring system that gives all non-R researchers the same weight. The equivalent contribution for Lincoln was 52 per cent, and yet this university had the second-lowest growth rate in its AQS. The percentage additional contribution arising from the composition of researchers is lowest for AUT, Waikato and Canterbury, yet they are in the group having the highest growth rates in their AQSs. Hence, no clear pattern of change emerges.

21 This increase complements the finding of Gemmell et al. (2017) that research productivity in New Zealand universities has increased markedly since the early 2000s.

Table 3. Growth of AQSs: 2003–12

	AUT	Mass.	Linc.	Wai.	Otago	VUW	Auck.	Cant.	NZ
AQS 2003	0.11	1.07	1.58	1.58	2.00	2.30	2.46	2.53	1.73
AQS 2012	1.36	2.88	2.46	3.15	3.55	4.37	3.74	4.26	3.31
% growth	1143	169	56	99	78	89	52	68	91
%	Percentage improvement of AQS growth rate above that which would result from a common score (G) for A, B and C quality categories								
Improvement	10	29	52	14	29	83	25	17	9

Source: Authors' calculations using the TEC dataset.

The question arises of whether the changes exhibit convergence, implying an inverse relationship between the proportional improvement of AQSs and initial AQSs. Table 3 arranges universities from left to right from the lowest to highest AQS in 2003. Convergence is consistent with growth rates decreasing from left to right. AUT has the lowest AQS in 2003 and has by far the highest growth rate of over 1,100 per cent. Massey has the second-lowest AQS in 2003 and the second-highest growth rate at nearly 170 per cent. However, the remaining observations do not follow the pattern predicted by convergence: there is no systematic tendency towards uniformity of AQSs across universities.²²

A related issue concerns the policy objectives of the PBRF scheme. While one stated purpose is to raise the quality of research in New Zealand universities, this can create a tension between 'equity' and 'efficiency' in research funding (see, for example, Hicks & Katz, 2011). The incentives may have the effect of generating a concentration of higher-quality research in a small number of universities. However, objections may be made on equity grounds involving, for example, the geographical distribution of researchers.

5. Patterns of change in research quality of New Zealand universities

This section examines the characteristics of research quality change within New Zealand universities. This is achieved by considering the range of flows within universities that contributed to changes in their AQSs. First, transitions among the quality categories, along with entrants and exits from 2003 to 2012, for all universities combined are shown in Table 4. The flows are from rows to columns; for example, of those who were As in 2003, 54.8 per cent were As in 2012 in the same university.

22 More formal testing of convergence of AQSs is provided in Buckle and Creedy (2017a).

These flow percentages show, for example, that a very small proportion, just under 6 per cent, of those who enter (that is, submit a portfolio for their employing university in 2012 for the first time) do so with a grade R in 2012. The largest exit rate, of just over 70 per cent, is for those who were classed as R in 2003.²³ Hence, low recruitment and high exit rates of Rs is a strong feature of transitions over the PBRF period.

A low proportion, 7.5 per cent, of entrants between 2003 and 2012 were classified as A-researchers in 2012. The majority of entrants are classed as B and C in 2012, at 35 and 52 per cent respectively. Just under 20 per cent of the 2003 Rs moved upwards to become Cs in the same institution by 2012, and 6 per cent moved upward to B. Upward movements within the same institution came mainly from C and B researchers, where for each category about a quarter of individuals moved one step upwards from 2003 to 2012. Hence for Bs and Cs, it could be said that about one third of entrants (between 2003 and 2012) fell into each category, and about a quarter of those who were classed as either B or C in 2003 progressed to a higher grade in the same institution by 2012.²⁴

Table 4. Matrix of flows (percentages): All universities combined, 2003–12

Quality category in 2003	Quality category in 2012				Exits	Total
	A	B	C	R		
A	54.8	16.1	0.5	0	28.7	442
B	24.7	52.6	12.8	0	9.8	1294
C	2.9	26.3	23.3	1.2	46.4	2374
R	0	6.0	19.2	4.6	70.2	2453
Entrants	7.5	34.7	51.8	5.9		3165
Total	868	2622	2833	329	3076	9728

Source: Authors' calculations using the TEC dataset.

It also seems that of those classed as C-researchers in 2003, a high proportion, 46 per cent, had exited by 2012. It is likely that many of these moves were to another university, and may have involved some kind of promotion, but this information is not available from the data. The B-researchers experienced much less outward mobility, but 29 per cent of those who were A in 2003 had exited. This is consistent with a small proportion of entrants over the period being classed as A-researchers

²³ However, some Rs may have entered another university at the same or a higher level. And, as discussed above, some could have been given new contracts, which meant that they avoided the need to submit a portfolio.

²⁴ It has been noted that, in 2006, a distinction was made for 'new and emerging' (NE) C and R staff. The effect of the NE category on upward movement was examined using matrices of transitions between 2006 and 2012. The distinction in the case of Cs appears to have had little value. But in the case of R researchers, those who were NE in 2006 experienced much more upward movement than other Rs. This suggests that the PBRF encouraged more careful selection of entry-level researchers.

in 2012 because there can be a large proportion of As who left between 2003 and 2012, since the denominator (initial number of As) is much smaller than the total number of entrants in all categories over the period.

6. Individual university responses

This section examines the extent of heterogeneity among universities in their recruitment and transitions. The results reported in Section 5 suggest that the conversion of Bs and Cs to higher grades is less ‘costly’ than that of Rs. The relative ability to recruit higher-quality researchers and retain high-quality researchers may vary across universities. Universities were incentivised to recruit at the higher levels and appoint very few researchers with low research outputs. However, the incentives vary among universities depending on their initial AQS. For example, it can be shown that the proportional increase in the AQS of a university with n researchers (assuming these are all full-time employees), resulting from hiring one additional type-A researcher, is given by:

$$\text{Proportional change in AQS} = \frac{\left(\frac{w_A}{\text{AQS}} - 1\right)}{n+1} \quad (3)$$

Where w_A is the value assigned to a quality category of A. By contrast, the change in AQS from eliminating n_R type-R people is given by:

$$\text{Proportional change in AQS} = \frac{n_R}{n - n_R} \quad (4)$$

Figure 2 illustrates some variations in the change in AQS for the case where $n = 700$. Clearly, if a university has an AQS greater than 2 its AQS would in fact fall as a result of hiring one more C-type researcher. It would need to recruit at least at level B, or reduce the number of R-type people. There are sharply decreasing returns, in terms of the impact on AQS, from hiring A and B researchers. These properties suggest that university responses to the introduction of PBRF are likely to vary depending on their initial situation.

To illustrate the varying responses, Table 5 provides information about the flows of researchers from 2003 to 2012 for each university. The actual flows are shown on the left-hand side. It is also possible to construct hypothetical flows, on the assumption that each university has the same transition proportions as all universities combined, but start from their actual numbers in 2003. The right-hand side of Table 5 shows the differences between the actual flows and the hypothetical flows. If all entries in this right-hand block were zero, all universities could be said to display identical transition propensities. A formal test of whether a university differs from ‘All NZ

universities' can be obtained by computing a standard chi-square statistic, using the hypothetical flows as 'expected' values. The resulting chi-square statistics are reported in each case on the same line as the university name in Table 5.²⁵

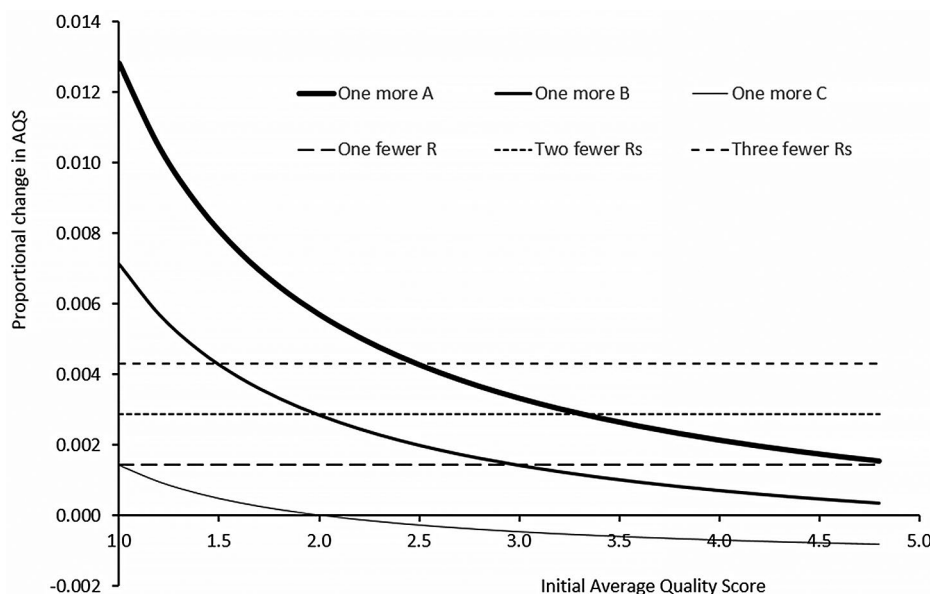


Figure 2. Effects on AQS of varying quality categories

Source: Authors' calculations.

As anticipated, the transition proportions clearly differ among universities. The only universities whose transitions do not differ significantly from those of all universities combined are Massey and Auckland. The university that differs most from the average is VUW, which had the highest AQS in 2012. The next-highest chi-square value is for AUT, which had the highest percentage change in its AQS over the period. While AUT and VUW differ most from the overall pattern, they also differ from each other considerably. For example, if VUW is compared with AUT (that is, by computing expected frequencies by combining VUW transition and exit rates with the initial AUT stocks) a chi-square value of 323.6 is obtained.²⁶

25 The appropriate chi-square values, for type I errors of 0.05 and 0.10, and for 16 degrees of freedom, are 26.296 and 23.542 respectively.

26 This is the highest value in such pairwise comparisons; for example, comparing VUW with Auckland and Lincoln gives chi-square values of 250.7 and 245.6 respectively. The lower-ranked universities also differ significantly from each other; hence, comparing Lincoln with AUT and Massey gives chi-square values of 54.51 and 174.5 respectively.

Table 5. Actual flows and differences between actual and hypothetical flows for each university: 2003–12

	Actual					Differences from actual				
	A	B	C	R	Exits	A	B	C	R	Exits
AUT	<i>Chi-square =75.94</i>									
A	3	2	0	0	0	0	+1	0	0	-1
B	3	18	1	0	0	-2	+6	-2	0	-2
C	2	29	24	2	48	-1	+1	0	+1	-1
R	0	11	84	22	357	0	-17	-7	0	+24
Entrants	9	52	171	33		-11	-40	+34	+17	
Lincoln	<i>Chi-square =51.05</i>									
A	5	1	0	0	1	+1	0	0	0	-1
B	7	17	6	0	1	-1	+1	+2	0	-2
C	2	22	30	2	39	-1	-3	+8	+1	-5
R	0	0	10	9	41	0	-4	-2	+6	-1
Entrants	5	18	64	15		-3	-17	+11	+9	
Massey	<i>Chi-square =18.02</i>									
A	25	6	0	0	13	+1	-1	0	0	0
B	34	83	27	0	13	-5	0	+7	0	-2
C	14	101	117	6	212	+1	-17	+12	+1	+3
R	0	34	93	23	430	0	-1	-19	-4	-23
Entrants	27	157	273	27		-9	-11	+22	-2	
Auckland	<i>Chi-square =18.77</i>									
A	93	17	1	0	48	+6	-9	0	0	+2
B	115	192	45	0	48	+16	-19	-6	0	+6
C	20	153	115	8	253	+4	+9	-13	+1	-2
R	0	29	86	20	279	0	+4	+6	+1	-11
Entrants	80	303	444	55		+14	-3	-13	+3	0
Canterbury	<i>Chi-square =53.59</i>									
A	24	9	1	0	21	-6	0	+1	0	+5
B	31	78	9	0	21	-3	+5	-9	0	+7
C	5	59	49	0	140	-2	-8	-10	-3	+23
R	0	12	47	10	70	0	+4	+20	+4	-28
Entrants	20	124	173	17		-5	+8	0	-3	
Otago	<i>Chi-square =28.06</i>									
A	56	16	0	0	24	+3	+1	0	0	-4
B	78	161	43	0	24	+2	0	+4	0	-6
C	14	127	130	1	207	0	+1	+19	-5	-15
R	0	26	67	14	305	0	+1	-12	-5	+16
Entrants	43	211	313	14		-1	+9	+12	-20	

	Actual					Differences from actual				
	A	B	C	R	Exits	A	B	C	R	Exits
Waikato		<i>Chi-square =57.84</i>								
A	12	13	0	0	9	-7	+8	0	0	-1
B	18	59	25	0	9	-9	+1	+11	0	-2
C	5	61	54	6	69	-1	+10	+9	+4	-21
R	0	9	51	13	109	0	-2	+16	+5	-19
Entrants	12	76	84	16		-2	+11	-13	+5	
VUW		<i>Chi-square =91.18</i>								
A	24	7	0	0	11	+1	0	0	0	-1
B	34	73	10	0	11	+2	+6	-6	0	-2
C	6	72	33	4	133	-1	+7	-25	+1	+18
R	0	26	34	2	130	0	+14	-3	-7	-5
Entrants	42	158	119	10		+17	+44	-52	-9	

Source: Authors' calculations using the TEC dataset.

Inspection of Table 5 reveals that the universities differ from each other in different ways.²⁷ As suggested by Figure 2, changes evident in the universities with a lower initial AQS differ from those with a higher initial AQS. Thus, AUT, which had the lowest initial AQS, recruited relatively few A and B researchers and converted few people to higher grades. It was relatively strong in recruiting Cs and achieving a high number of exits from Rs, but it also recruited more Rs than average. Lincoln also recruited relatively more Cs and Rs, but was about average in its ability to convert researchers to higher grades and achieving exits of Rs. Massey also recruited relatively more Cs, was quite successful at converting Rs to Cs and in achieving exits of Rs.

Thus, the experience differed among the three universities that were consistently ranked at the low end of the AQSs. The common features for this group are the low recruitment of As and Bs, and their reliance on recruiting Rs, with improvements coming from recruiting relatively more Cs and getting high exits among Rs. Conversion rates were also low, with the exception of Massey, regarding Rs. It could be argued that these universities face relatively high effective costs of recruiting As and Bs. Nevertheless, these universities managed to retain relatively more of their higher-scoring researchers.

Auckland was strong at recruiting As, but was in other respects similar to the pattern revealed by all universities combined. Canterbury recruited relatively more Bs, but at the same time suffered more than average exits of A and B researchers. It was also successful at converting Rs to Cs. Otago recruited relatively few Rs and had higher

²⁷ In comparing performance relating to the exits of R-researchers, it has to be kept in mind that this figure is distorted for reasons discussed in the introduction and in Section 4.

exits of Rs, while being strong at recruiting Bs. Nevertheless, it was about average at converting researchers to higher-scoring academics. Waikato had a relatively strong conversion of Rs to Cs and was stronger at appointing Bs, while achieving fewer exits of Rs.

At the top of the AQS ranking in 2012, VUW achieved success at appointing As and Bs, and keeping its higher-scoring staff. In addition, it appointed relatively fewer Cs and Rs. Unusually, it had success in converting Rs to B researchers. Among the top-five universities, Canterbury stands out in terms of its ability to appoint people who were rated B researchers in 2012. Canterbury's higher losses of As and Bs may be explained to some extent by the earthquake experience, since those researchers are expected to be more mobile. The higher-ranked universities were also relatively better at converting researchers to higher grades, with the possible exception of Otago.

It might be argued that these higher-ranked universities have the kind of environment—including stronger academic leadership from A and B researchers—and perhaps also more resources devoted to internal research grants, which stimulates higher outputs. Also, in recruiting Rs, they may be selecting from those who are at the higher end of the scale in terms of their potential. The greater ease of attracting As and Bs to the higher AQS universities, combined with the effect of the funding formula, may be said to involve an effective reduction in their 'price'.

One difficulty in interpreting the findings is that the analysis of transitions during which PBRF has operated does not have a control group with which to compare performance.²⁸ Hence, it is not obvious that all the changes can be attributed directly to the PBRF process. Nevertheless, there is no doubt that New Zealand universities in 2003 had a large number of Rs, and this was inconsistent with the incentive structure created by the PBRF exercise. Indeed, the high exit rate of those categorised as R in 2003, combined with the low recruitment rate over the 2003 to 2012 period of those categorised R in 2012, combined with the relatively high recruitment of B and C researchers, is most unlikely to represent an equilibrium situation.

This question can be formally assessed by considering the implied equilibrium distribution generated by the transitions and entries observed over the PBRF period, if they are held constant. After a sufficient time, the system would settle into an equilibrium distribution of academics across the quality categories: in this situation, outward movements from each category would be balanced by inward movements

28 Ideally, in order to isolate the effect of PBRF it would be necessary to compare the post-PBRF transitions with those that would have occurred in the absence of PBRF. This would require information about previous transitions, including the likely behaviour of those who may not have left the system in the absence of PBRF. In addition, there may have been other external influences, such as the effect of the global financial crisis on the ability of New Zealand universities to recruit academics from other countries, the effects of which cannot be isolated.

each period. An equilibrium distribution derived by using the transition matrix and vector of entrants in Table 4 was found to give equilibrium stocks of each quality category that are not feasible and very different from the totals observed for 2012.²⁹ It is therefore not unreasonable to suggest that the evolution of New Zealand universities since 2003 represents to some extent a structural shift in response to the introduction of PBRF, and this path cannot be expected to continue indefinitely. These results help to provide an indication of the extent to which the changes since 2003 are in fact large, despite the many characteristics of universities that make structural change difficult.

7. Changes in the age distribution of researchers

The desire to retain A and B researchers, who are typically relatively older, along with considerable caution in recruiting entry-level researchers, so that only those who have already displayed some publishing success are likely to be appointed to lectureships, suggests a tendency for population ageing in the New Zealand university system. The age of some researchers was not reported in the TEC dataset, and hence the sample size used to evaluate changes in the age distribution of researchers is smaller than that used in the other Sections. Nevertheless, Table 6 shows a marked change in the age distribution of researchers. Between 2003 and 2012, there were large declines in the number of researchers in each age group, except for the 60–89 group, with a 77.7 per cent growth rate. The largest absolute decline was in the 40–49 age group, which fell by 415 researchers (a decline of 18.2 per cent). But the largest percentage decline was in the 20–29 years age cohort which fell by 64.3 per cent.

Table 6. Change in the age distribution of PBRF researchers from 2003 to 2012

	Age					Totals
	20–29	30–39	40–49	50–59	60–89	
Absolute change in number in age group	–153	–244	–415	–231	+602	–441
Change as per cent of 2003 age group	–64.3	–16.1	–18.2	–10.9	+77.7	–6.4

Source: Authors' calculations using the TEC dataset.

The changes shown in Table 6 may be associated with changes in the age distribution by grade. To compare the distributions, the observed proportions for 2003 were used to derive 'expected frequencies' for the 2012 age distribution by grade and 2012 grade distribution by age (using the absolute totals in 2012). The differences between the observed and expected frequencies for the 2012 age distribution by grade are shown in Table 7; these may be referred to as 'unexpected changes'.

29 See Buckle and Creedy (2018) for further details.

The sum of each column in Table 7 shows the differences for each age group between observed and expected frequencies; the sum of each row is zero. These totals reveal that the large increase in the frequency of researchers in the 60–89 age group is much higher (by 599.71) than could be expected from the 2003 marginal frequencies. Conversely, the decline in the frequency of researchers in each of the other age groups is larger than would be expected from the 2003 marginal frequencies. The largest differences are the decline of approximately 306 in the 40–49 age cohort, and the rise of approximately 600 older age cohort researchers.

Table 7. Differences between observed and expected 2012 age distribution by quality category

QC	Age				
	20–29	30–39	40–49	50–59	60–89
A	0.00	–13.10	1.77	–79.90	91.23
B	3.22	–82.09	–130.50	–77.55	286.92
C	–29.32	2.27	–157.85	–3.43	188.33
R	–15.34	–13.75	–19.25	15.11	33.23
Totals	–41.44	–106.67	–305.83	–145.77	599.71

Source: Authors' calculations using the TEC dataset.

However, there were higher than expected numbers of R-category people in the older age groups. Conversely, there are large unexpected declines in the number of people in C and R categories aged 20–29, of Bs in the 30–39 age cohort, of C and B researchers aged 40–49, and large unexpected declines in A and B researchers aged 50–59.

The unexpected high number of older people of R-quality may be because it is more difficult and expensive to remove people in the older age cohorts.³⁰ The 'excess' decline in lower-category researchers in the youngest age cohort may reflect the possibility that people in that stage of their careers are more career mobile or are easier to remove.³¹

8. Assessment of PBRF appraisal method and metrics

Having examined some of the characteristics of changes in AQSs, it is useful to return to the choice of metric itself. A fundamental initial judgement required in any research quality appraisal involves the question of whether quality is perceived

³⁰ In some cases, they may be valued for administration or teaching qualities.

³¹ Formal chi-square tests based on Table 7 showed that the differences are statistically significant at all reasonable significance levels.

to be a continuous (or quasi-continuous) variable, or whether it is discrete, whereby individuals are placed into a small number of well-defined categories. The basic view of 'quality' has not been articulated by the Ministry of Education and the PBRF process involves what might be described as a hybrid (or perhaps ambivalent) method.

As explained in Section 2, each research portfolio is assigned an initial total weighted score by discipline experts, which is the sum of the weighted individual component scores for research output, peer esteem and contribution to research environment. These initial scores are subject to a further round of assessment, where adjustments may be made. After this second round, the portfolio is then placed into one of four quality categories, R, C, B and A.

The method used to derive total weighted scores, involving equation (1), has some properties which may not be immediately apparent. With three components s_j , where $j = 1, 2, 3$, each is given an integer score from 0 to 7, the total weighted score, S , can range from 0 to 700, but many values in the range are not possible. The number of possible values actually depends on the weights, q_j , used to form the aggregate, S . The number of permutations is 512. With PBRF weights of 0.7, 0.15 and 0.15, the possible values start from 0 and increase in increments of 15 until reaching 60. The next two numbers in the sequence are 70 and 75, after which they increase in increments of 5 up to 570. The next number is 580 and then the values increase in increments of 5 up to the maximum of 700. Furthermore, the number of ways of achieving a given value of S varies with S itself. To give a few examples, there are two ways of obtaining a score of 15, but 5 ways of getting a score of 60, 6 ways of obtaining 345, 7 ways of getting a score of 90, and 8 ways of getting 315. This means that, if everyone has an equal independent probability of obtaining a score from 0 to 700 for each of the quality types, there is not an equal probability of falling into equal ranges of S .

Buckle and Creedy (2017b) show that the scoring system can itself have a substantial effect on the measured quality distribution of researchers. Examination of the distribution of initial weighted scores in the 2003 and 2012 PBRF rounds, and movements in those scores during the expert panel assessments stages, indicates that the use of quality categories and associated thresholds appear to have influenced the performance scores assigned to research portfolios and the final allocation of portfolios to quality categories. In particular, the process generated large spikes in the distribution of total weighted scores at the lower thresholds for each quality category. Furthermore, the distribution of scores is roughly symmetric. These features contrast strongly with the distribution of metrics used by other performance evaluation methods. Without a clear rationale for the use of quality categories and the selection of threshold values, the choice of weights, q_j , assigned to each component also appears arbitrary. Yet, these weights and the distribution of total weighted scores can affect the research quality ranking of universities and disciplines.

This raises questions as to whether the presence of threshold levels created by the use of quality categories, and the weights attached to each component, influence the decisions of peer reviewers and discipline expert panels in assigning preliminary scores and moving people to the next threshold. It also raises the possibility that the final distribution of total weighted scores has properties that differ from those that would emerge in a system using only an overall measure based on the distribution of weighted scores, S , for disciplines and universities, rather than the AQSs based on the weighted quality categories.

This system, used by the TEC to summarise research quality, discards information about the relative quality of researcher portfolios that could be used to derive alternative summary measures of average university quality.³² Furthermore, it is not possible to know what distribution of scores would have arisen without the reviewers' prior knowledge of the use of quality category thresholds. An alternative method, using quality categories, would involve the independent evaluation of scores, followed by the determination of thresholds. This is still subject to the criticism that the use of quality categories compresses all scores within a given range to a single value.

Considerable attention is given to the AQSs and resulting rankings, in view of the kind of reputation effects stressed by the Organisation for Economic Co-operation and Development (OECD, 2010). Nevertheless, as mentioned in the introduction, funding is based on a formula which uses individual quality categories, rather than AQSs, with funding weights that vary by discipline.

9. Conclusions

This paper has summarised the changes in New Zealand universities' measured research quality since the introduction of the PBRF scheme in 2003, which was designed to promote and reward research excellence in the tertiary education sector. The New Zealand scheme is particularly interesting because it uses a peer-review system of assessment and assesses all eligible university staff.

The 'Quality Evaluation' component of the PBRF funding formula takes account of the number of FTE researchers and their quality. Furthermore, all universities increased their submitted evidence portfolios as a ratio of their total number of non-administrative staff.

Based on the data for individual quality categories, the average quality of researchers has nearly doubled between 2003 and 2012. This inference is based on a measure of the AQS for all universities combined, where the AQS is the ratio of the sum

³² The variation could perhaps be regarded as 'noise' rather than a clear quality signal. If this view is taken it is not clear why the first stage is used, as individuals could be assigned directly to discrete categories.

of all quality categories expressed as a proportion of the sum of all university non-administration staff. Although the TEC has published other AQS measures, they either do not capture the average quality of researchers, or they are compromised by the decision in 2012 by the TEC to remove R-quality researchers from the calculations.³³

There are considerable differences in the rates of improvement amongst the eight universities. The fastest rates of improvement were achieved by AUT and Massey universities, two of the lower-ranked universities in terms of AQSs in 2003. The slowest rate of improvement was experienced by Lincoln, also one of the lowest-ranked in 2003. Rates of improvement also varied markedly amongst the top-ranked universities. Consequently, there has not been a clear tendency for convergence to the mean level of research quality for the eight universities. There has been persistence in the relative quality of the group of top-ranked and the group of bottom-ranked universities. Rankings have changed within the groups, with Auckland declining amongst the group of higher-ranked universities and Lincoln declining in the lower group, but there has been a persistence in membership of the two quality groups.

A strong feature contributing to improvements in research quality was a large reduction in R-quality staff. In addition to low recruitment and high exit rates of R-quality researchers, it also seems that a high proportion of those classed as C-researchers in 2003 had left their 2003 university by 2012. The majority of entrants were B and C quality researchers. Upward movements within the same university came mainly from C and B researchers. But the experiences varied significantly across universities, and this variation tended to be associated with the research ranking of the university in 2003. In particular, of the new entrants between 2003 and 2012, the higher-ranked universities tended to be more successful at attracting and retaining A and B quality researchers. This was a standout feature of VUW, which experienced the highest proportional improvement in its AQS compared to the other higher 2003 ranked universities of Auckland, Canterbury and Otago. The lower-ranked universities in 2003, consisting of AUT, Massey and Lincoln, were more successful at recruiting C-quality researchers.

It is not clear whether the PBRF process was intended to encourage a more concentrated allocation of funding to the more highly-ranked universities. While there is not strong evidence for convergence or divergence in quality growth rates, the responses have been found to depend on each university's initial research quality. Thus, the characteristics of growth for AUT and Massey were very different from, for example, VUW and Canterbury. Furthermore, despite the high growth rates of AUT in particular and, to a lesser extent, Massey University, the margin in the AQSs

33 Researchers using bibliometric measures to evaluate the impact of PBRF on research quality have drawn varying conclusions. For example, using data from the Web of Science, Smart (2009) concludes that research output and quality increased at New Zealand universities since the introduction of PBRF. However, Hodder and Hodder (2010) conclude that quantity increased while quality decreased.

between the top- and bottom-ranked universities has persisted. It would be helpful if the Ministry of Education were to make an explicit statement about whether it believes that a more efficient use of funding would involve a greater concentration of research quality or, on the other hand, it prefers a more dispersed allocation. The preferred goal should influence the design of the performance-based funding schemes and their incentives.

Care needs to be taken in judging the AQSs. For all universities combined, the AQS is below 4, out of a maximum of 10. Because of its unique features, it is not possible to compare New Zealand universities with those elsewhere. Whether this value is judged to be poor or satisfactory really depends on a clear view of what is being measured and what optimal balance of staff is appropriate, since a situation where most non-administrative members of a university are A-quality may not be desirable. The Ministry of Education has not expressed a view on what it regards as a 'satisfactory' AQS for the university system. This is very important when considering the merit or otherwise of retaining the scheme and its associated metrics.

A strong feature associated with the improvements in researcher quality has been a significant change in the age distribution of researchers. There was substantial population ageing over the period. This arose from a combination of an increase in the average age of entry and reduced exits from older age groups. This led to changes in the age distribution within quality category grades. The cost of recruiting younger researchers who do not progress to higher-quality categories has increased with the introduction of PBRF. This incentive has indeed influenced recruitment, with a tendency to recruit mainly early career academics with clearer evidence of research capability. However, there has not been a relatively high rate of exit from older Rs, and this may be because of employment-related constraints associated with tenure.

Without similar pre-PBRF data, it is not obvious that the observed changes can all be attributed to the introduction of the PBRF process. Nevertheless, it can be argued that the transformation that has taken place is not sustainable, in the sense that if the transition proportions remained constant over time, the equilibrium distributions of researchers would be unrealistic. It is therefore reasonable to suggest that the large changes over the period 2003 to 2012 have, to a substantial degree, been stimulated by the introduction of the PBRF and that such large changes are unlikely to be repeated.

The PBRF scoring system involves several stages of evaluation and weighting to derive each individual's quality category. This combines the use of an idiosyncratic quasi-continuous scoring system with the use of discrete quality categories. It reflects an ambivalence concerning the view taken of how research quality should be measured. The system could be significantly simplified given a clear judgement on this crucial issue. For example, if it is believed that research quality can only be classified into

discrete broad categories, there would seem to be no role for a first stage derivation of a total weighted score for each person. On the other hand, if continuous variation over a wide quality range, reflected in many metrics of research quality, is believed to be appropriate, the PBRF scoring system could be considerably simplified and improved to reflect the kind of heterogeneity revealed in other metrics.

Where the PBRF process is used to rank universities in terms of research quality, there are no good reasons not to include, as denominator in the calculation of AQSs, the total number of non-administrative staff.³⁴ This is the denominator used in the present paper, although it has not been used by the TEC (which excludes R-quality researchers) when publishing PBRF outcomes. In addition, the use of student numbers as one denominator has no clear rationale and the publication of AQSs on this basis has been subject to misinterpretation.

It is clear that an incentive structure involving some kind of assessment is needed to allocate public funds and to stimulate the adoption of appropriate policies by universities. A challenge is to avoid unintended negative consequences. This paper has shown that the research quality of New Zealand universities has improved substantially, and it can be argued that this is a direct result of the introduction of the PBRF system. Nevertheless, it cannot be concluded unambiguously that universities have been appropriately rewarded for their improvements or that the procedures and metrics used are ideal. Given the large compliance costs, there is clearly scope to improve the system in respect of the demands on universities and academics in producing their evidence portfolios, the way in which these are evaluated, and the way funds are allocated. The debate on reforms to PBRF would also need to address issues raised by other commentators regarding, for example, the high administrative and compliance costs, the types of research supported and incentives to work on contemporary New Zealand issues.

References

- Anderson, D.L. & Tressler, J. (2014). The New Zealand performance-based research fund and its impact on publication activity in economics. *Research Evaluation*, 23(1), 1–11. doi.org/10.1093/reseval/rvt017
- Anderson, D.L., Smart, W. & Tressler, J. (2013). Evaluating research—peer review team assessment and journal based bibliographic measures: New Zealand PBRF research output scores in 2006. *New Zealand Economic Papers*, 47(2), 140–57. doi.org/10.1080/00779954.2013.772879

³⁴ Indeed, a case can be made for including all staff.

- Buckle, R.A. & Creedy, J. (2017a). *The evolution of research quality in New Zealand Universities as measured by the Performance-Based Research Fund process* (Victoria University of Wellington Working Papers in Public Finance, no. WP11/2017). Wellington: Victoria University of Wellington.
- Buckle, R.A. & Creedy, J. (2017b). *An evaluation of metrics used by the Performance-Based Research Fund Process in New Zealand* (Victoria University of Wellington Working Papers in Public Finance, no. WP16/2017). Wellington: Victoria University of Wellington.
- Buckle R.A. & Creedy, J. (2018). *The Evolution of Research Quality in New Zealand Universities as Measured by the Performance-Based Research Fund Process*. New Zealand Economic Papers. Retrieved from: www.tandfonline.com/doi/full/10.1080/00779954.2018.1429486
- Butler, L. (2003). Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143–55. doi.org/10.1016/S0048-7333(02)00007-0
- Coryn, C.L.S. (2007). *Evaluation of Researchers and Their Research: Toward Making the Implicit Explicit* (PhD thesis, Western Michigan University, USA).
- Davidson, S. (2013). Excellence in research for Australia: an audit of the Applied Economics Rankings. *Agenda*, 20(2), 5–20. doi.org/10.22459/AG.20.02.2013.01
- de Boer, H., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K. & Vossenstey, H. (2015). *Performance-based funding and performance agreements in fourteen higher education systems*, Report for the Ministry of Education, Culture and Science, Centre for Higher Education Policy Studies, C15HdB014, Universiteit Twente, Enschede. Retrieved from: research.utwente.nl/en/publications/performance-based-funding-and-performance-agreements-in-fourteen-
- Gemmell, N., Nolan, P. & Scobie, G. (2017). *Estimating quality-adjusted productivity in tertiary education: methods and evidence for New Zealand* (Victoria University of Wellington Working Papers in Public Finance, no. WP17/2017). Wellington: Victoria University of Wellington.
- Gibson, J., Anderson, D.L. & Tressler, J. (2008). *Do research assessment exercises raise the returns to publication quality? Evidence from the New Zealand market for academic economists* (University of Waikato Department of Economics Working Paper, 08\11). Hamilton: University of Waikato.
- Hicks, D. & Katz, J.S. (2011). Equity and excellence in research funding. *Minerva*, 49(2), 137–51. doi.org/10.1007/s11024-011-9170-6
- Hodder, A.P.W. & Hodder, C. (2010). Research culture and New Zealand's Performance-Based Research Fund: some insights from bibliographic compilations of research outputs. *Scientometrics*, 84(3), 887–901. doi.org/10.1007/s11192-010-0201-0

- Jones, A. & Cleere, L. (2014). *Furthering the research impact of UCD: report of the Beyond Publications Committee*. Dublin: University College Dublin. Retrieved from: hdl.handle.net/10197/7292
- Jonkers, K. & Zacharewicz, T. (2016). *Research Performance Based Funding Schemes: A Comparative Assessment*. European Commission Joint Research Centre Science for Policy Report. EUR 27837 EN.
- Ministry of Education (2012). *A History and Overview of the PBRF*. Wellington: Ministry of Education. Retrieved from: education.govt.nz/assets/Documents/Further-education/Policies-and-strategies/Performance-based-research-fund/PBRFHistoryAndOverview.pdf
- New Zealand Tertiary Education Commission (2002). *Investing in Excellence: The Report of the Performance-Based Research Fund Working Group*. Wellington: Ministry of Education and Transition Tertiary Education Commission.
- New Zealand Tertiary Education Commission (2013). *Performance-Based Research Fund: Evaluating Research Excellence—the 2012 Assessment*. Wellington: Tertiary Education Commission.
- OECD (2010). *Performance-Based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. OECD Publishing. Retrieved from: www.oecd.org/science/sci-tech/performance-basedfundingforpublicresearchintertiaryeducationinstitutionsworkshopproceedings.htm
- Roberts, G. (Chair) (2005). *Research quality framework: Assessing the quality and impact of research in Australia*. Department of Education, Science and Training Expert Advisory Group for a Research Quality Framework: Issues Paper.
- Smart, W. (2009). *The Impact of the Performance-Based Research Fund on the Research Productivity of New Zealand Universities*. Wellington: Ministry of Education. Retrieved from: www.educationcounts.govt.nz/publications/tertiary_education/education-research/49897
- Tressler, J. & Anderson, D. L. (2012). Citations as a measure of the research outputs of New Zealand's economics departments: the problem of 'long and variable lags'. *Agenda*, 19(1), 17–34. doi.org/10.22459/AG.19.01.2012.02
- van den Besselaar, P., Heyman, U. & Sandström, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*, 11(3), 905–18. doi.org/10.1016/j.joi.2017.05.016
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Higher Education Funding Council for England. doi.org/10.4135/9781473978782

How Much Have Chinese Investors Invested in Australia?

Kerry Liu¹

Abstract

Chinese outward direct investment (ODI) in Australia has been debated for many years. Different data sources provide quite different indications of how much Chinese investors have actually invested in Australia. This study analyses each data source's application and limitations, and provides some guidelines on how to interpret and use these data. The findings include: first, although the Australian Bureau of Statistics and Chinese official data measure direct Chinese ODI in Australia, the real value of Chinese capital flow into Australia is greater than these measures. Second, KPMG/University of Sydney and American Enterprise Institute/Heritage Fund data measure contracted value, which may be expected to be higher than the true value of capital flow given the uncertainty of the business environment. Third, as Foreign Investment Review Board data is the value of proposed investment, using this data to measure Chinese ODI in Australia is misleading.

Introduction

Chinese outward direct investment (ODI) in Australia has been a hot topic for debate for many years. For example, during 2008–09, the scale and speed of the surge of Chinese state-owned enterprises' (SOEs') investment in Australia raised the question of whether investments by SOEs require special scrutiny (Drysedale, 2011). Since then, many concerns including job issues, food security issues and so on have been raised about Chinese ODI in Australian public debate (Drysedale et al., 2016). The most recent concern is the possibility that Chinese Government could strategically leverage Chinese investments in critical infrastructure to impede

¹ Kerry Liu is an independent scholar of the Chinese economy, kerry.luke@gmail.com. The author would like to thank one anonymous referee and the editor for their helpful comments.

Australia's national security. The climax of the Australian Government's response to this debate is the establishment of the Critical Infrastructure Centre, announced on 23 January 2017. The main reason given for this is '[given] the shift in our international investment profile, Australia's national critical infrastructure is more exposed than ever to sabotage, espionage and coercion' (Australian Government, 2017). This shift in Australia's investment profile mainly means the rise of Chinese foreign direct investment (FDI) in Australia, evidenced with a string of recent events including the sale of the Port of Darwin to the Chinese-owned Landbridge for AU\$506 million in October 2015, the Treasurer's decision to refuse the sale of the AU\$370-million Kidman cattle property in April 2016 to Chinese investors, and the Treasurer's further refusal of Chinese State Grid's AU\$10-billion bid of AusGrid in August 2016.

Chinese ODI in Australia has attracted extensive attention from a variety of stakeholders including the media. Yet the issue of how much Chinese investors have invested in Australia is a mystery. Media reports often provide conflicting figures, which, sometimes if not always, are puzzling. For example, on 27 February 2017, *The Australian* announced, 'Chinese investment in Australia soars to \$4.8bn' (Callick, 2017). Around two months later, on 1 May 2017, *The Sydney Morning Herald* reported, 'Chinese investment in Australia surged 11.7 per cent to \$15.4b last year: report' (Scott, 2017). Just eight days later, on 9 May 2017, the *Financial Review* offered, 'Chinese investment in real estate grows to \$32b: FIRB' (Tan, 2017). So, what is really going on?

With regard to the exact figure of Chinese ODI in Australia, there are three types of data sources. The first is the official statistics from Australian Bureau of Statistics (ABS) and the Department of Commerce, China (DCC). The second is research institutions, including the KPMG/University of Sydney (UniSyd) 'Demystifying Chinese Investment in Australia' series, and the American Enterprise Institute (AEI) and the Heritage Foundation (HF) data. The third is the data released by the Foreign Investment Review Board (FIRB). Each of the above data sources show a different picture of Chinese ODI in Australia. While not trying to give a fourth type of Chinese ODI in Australia data, this study analyses the applications and limitations of each data source, and provides some guidelines on how to interpret and use these data.

Official data: ABS vs DCC data

In terms of ABS data, the conceptual framework used in compiling Australia's FDI statistics is based on the *Balance of Payments and International Investment Position Manual* (sixth edition), produced by the International Monetary Fund (IMF, 2009). Accordingly, a foreign direct investor is an entity resident in one economy that has acquired, either directly or indirectly, at least 10 per cent of the voting power of

a corporation, or equivalent for an unincorporated enterprise, resident in another economy. Furthermore, FDI statistics are presented according to the directional principle rather than asset/liability principle. Under the directional presentation, the direct investment flows and positions are organised according to the direction of the investment for the reporting economy—either outward or inward. So, for a particular country, all flows and positions of direct investors resident in that economy are shown under outward investment and all flows and positions for direct investment enterprises resident in that economy are shown under inward investment. The ABS also mentions how to treat reverse investment, which is when an affiliate invests in its parent. Under the directional presentation, reverse investment is subtracted to derive the amount of total outward or inward investment of the reporting country. So, if a resident parent borrows money from one of its foreign affiliates, this is subtracted in calculating the reporting country's outward investment because it reduces the amount of money that the resident country's parent entity has invested in its foreign affiliates on a net basis. Similarly, if a resident affiliate lends money to its foreign parent, this is subtracted when calculating inward investment because it reduces the amount of money that the foreign parent has invested in the resident country on a net basis. The ABS also notes that the country of residence of the ultimate benefit owner/recipient is not identified, so the country allocation of financial transactions and levels is based on the immediate country of residence (ABS, 2016), suggesting that ABS data may potentially underestimate Chinese flows in Australia. The ABS also includes reinvested earnings and excludes Hong Kong (HK, and Taiwan) from its calculation of Chinese FDI in Australia.

As to the DCC data, the DCC does not explicitly state that it follows the IMF guidelines. However, the DCC does use the same definition of FDI as ABS, including adopting the 10 per cent voting rights criterion. The DCC also adopts the reverse investment concept, which is defined as an affiliate's not more than 10 per cent investment in its parent or resident enterprise. Also, the DCC states that the country allocation of financial transactions and levels is based on the immediate country of residence (DCC, 2010). The DCC also includes reinvested earnings in its calculation of Chinese ODI. So, we can conclude that while there are some differences, such as the DCC's 10 per cent criterion for reverse investment, the methodologies of FDI/ODI statistics of the ABS and the DCC are similar.

Table 1 shows Chinese ODI in Australia based on ABS data and DCC data.

Table 1. Chinese ODI in Australia: ABS data vs DCC data (AU\$ billion)

	2008	2009	2010	2011	2012	2013	2014	2015	2016
DCC	2.18	3.13	1.87	3.09	2.10	3.54	4.48	4.50	–
ABS	3.19	4.86	2.63	3.27	3.44	6.15	9.86	3.38	3.86

Source: Author's summary of ABS and DCC data.

First, since the ABS excludes HK (and Taiwan) from its calculation of Chinese ODI in Australia, and only considers immediate country of residence, it is highly likely that some Chinese investments in the disguise of HK/Cayman Island/British Virgin Island (BVI) identities enter Australia. Since HK, Cayman Island and BVI are the top three destinations for Chinese ODI (see Table 2), the real amount of Chinese ODI in Australia would thus be higher than the ABS data shows.

Table 2. Top 10 destinations of Chinese ODI by end of 2015 (US\$ billion)

HK	Cayman Island	British Virgin Island	USA	Singapore	Australia	Holland	UK	Russia	Canada
656.86	62.40	51.67	40.80	31.98	28.37	20.07	16.63	14.02	8.52

Source: DCC.

Second, comparing ABS data and DCC data shows that ABS data are generally higher than DCC data before 2014. While Chinese private capital flows are dominated by FDI (Hatzvi et al., 2015), this difference may reflect the systematic inaccuracy (underestimation) of DCC data as a result of Chinese investors' long-term intention of trying to escape from capital account control regime.

Third, the year 2015 shows a different picture. In this year, Chinese ODI in Australia significantly decreased from AU\$9.9 billion in 2014 to just AU\$3.4 billion in 2015, according to ABS data. There are possibly two explanations.

- First, the Australian dollar depreciated a lot against USD and RMB in 2015, which means that investment in Australia would suffer exchange rate losses.
- Second, Chinese capital may have mainly flowed to HK in 2015. The People's Bank of China suddenly depreciated its currency against USD by 1.9 per cent on 11 August 2015. This unexpected policy change led Chinese investors (and international investors as well) to think China would join the 'currency war' and further depreciate the RMB to boost its exports (Liu, 2016). As Chinese capital flow seems to be driven by expected changes in the exchange rate (Hatzvi et al., 2015), Chinese investors traditionally use HK as the primary stop for capital outflow when encountering turmoils such as during 2008–09 and 2015 (Table 3; Prasad & Wei, 2007).

Table 3. Fraction of Chinese ODI in HK (per cent)

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
% of Chinese ODI invested in HK	32.8	51.8	69.1	63.0	56.0	47.8	58.4	58.3	57.6	61.6

Source: DCC.

Research institution data: KPMG/UniSyd data vs AEI/HF data

KPMG/UniSyd started publishing its ‘Demystifying Chinese Investment’ report series in 2011. However, KPMG/UniSyd reports did not disclose the methodology employed to collect and analyse data until 2013. Starting in 2013, KPMG/UniSyd reports stated that they cover direct investments into Australia made by entities from the People’s Republic of China through mergers and acquisitions (M&A), joint ventures (JVs) and greenfield projects. However, ‘greenfield projects’ were not mentioned in the 2016 report.² Furthermore, it is not clear how the yearly greenfield investment data are obtained and analysed, but rather ‘The University of Sydney and KPMG team obtains raw data on China’s ODI from a wide variety of public information sources which are verified, analysed and presented’ (KPMG/UniSyd, 2013–2017). KPMG/UniSyd reports also state that their datasets track Chinese investment by subsidiaries or special purpose vehicles based in HK, Singapore and other locations. Moreover, completed deals that are valued below US\$5 million are not included.

The AEI/HF also publishes Chinese ODI data since 2005 (AEI/HF, 2017). In terms of original data source, ‘we predate Dealogic for this and use the companies themselves—press releases, Shanghai or Hong Kong exchange disclosures, and the like’ (Derek Scissors, AEI/HF data author, personal communication, May 2017). The AEI/HF also publishes Chinese construction contracts data. Since construction contracts may be classified as services exports, construction contracts are excluded from the calculation of Chinese ODI in this study. However, the inclusion or exclusion of construction contracts does not make a significant difference to results.³ Moreover, the AEI/HF also includes greenfield investments, but only transactions valued at or above US\$95 million.

Table 4 shows Chinese ODI in Australia based on KPMG/UniSyd data and AEI/HF data.

Table 4. Chinese ODI in Australia: KPMG/UniSyd data and AEI/HF data (US\$ billion)

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Data source: KPMG/UniSyd	16.00	8.50	3.70	9.30	11.40	9.19	8.35	11.10	11.49
Data source: AEI/HF	15.78	9.13	3.34	9.19	9.24	8.38	10.03	10.59	5.26

Source: Author’s summary of KPMG/UniSyd and AEI/HF data.

² Greenfield FDI projects require the establishment of new entities overseas, including offices, buildings and factories. Greenfield FDIs involve capital flow. Cross-border M&As involve taking over or merging with the overseas enterprise’s cash, assets and liabilities (UNCTAD, 2009).

³ During 2008–16, there are only (in total) US\$6.5 billion construction contracts in Australia from China, while the total non-construction Chinese ODI is worth US\$80.9 billion (AEI/HF, 2017).

First, the comparison of Table 4 and Table 1 shows that the Chinese ODI in Australia, as indicated by data compiled by research institutions, is generally much higher than what the official data may suggest (also note that Table 4 is in US\$ while the Table 1 is in AU\$). There are reasons for this difference. Generally speaking, the transaction values announced either through media or in the form of company announcements issued through the exchanges are contracted values. They are not necessarily equal to the real value of capital flow. It is highly likely that the investment that actually happens is lower than the contracted value considering the uncertainty of business environment. Moreover, both KPMG/UniSyd and AEI/HF data do not include reverse investment, meaning that KPMG/UniSyd and AEI/HF may potentially inflate the real value of capital flow. At the same time, both KPMG/UniSyd and AEI/HF data do not include earning reinvestment, meaning that KPMG/UniSyd and AEI/HF may potentially deflate the real value of capital flow. However, the net effect, which is difficult to evaluate, should be marginal compared with the difference between contracted value and real value.

Second, the results from KPMG/UniSyd and AEI/HF during 2008–2015 are in fact very similar. However, since KPMG/UniSyd datasets have a lower criterion of inclusion (only US\$5 million) than AEI/HF, KPMG/UniSyd results should be higher than AEI/HF results. Since KPMG/UniSyd did not disclose detailed methodology of data collection, including whether construction contracts were included, the question of why KPMG/UniSyd and AEI/HF results are so similar will remain a mystery.

In fact, a reasonable approach of collecting M&A data is to retrieve original data from data vendors including Dealogic, ThomsonOne and MergerMarket. Since these three data vendors often provide repetitive data items, the combined results will first need to be cleaned. After that, the cleaned datasets will become the foundation of Chinese ODI data (excluding greenfield investment) for further updating or upgrading.

FIRB data

FIRB data reveals a stunning picture of Chinese ODI in Australia. In 2015–16, China was the largest source of foreign investment approvals by both value and number. The value of approvals for Chinese investors was AU\$47.3 billion in 2015–16, representing 26 per cent of the total value of approvals (FIRB, 2017). However, these figures are too inflated as a measure of Chinese ODI in Australia.

First, the statistics contained in the FIRB *Annual Report* reflects investors' intentions (not actual purchases) to acquire Australian assets. There are substantial differences between these statistics on proposed investments and actual investment

flows. FIRB itself also recommends the audience use ABS data, which more comprehensively reflect investment transactions between residents of Australia and non-residents (FIRB, 2017).

Second, FRIB data are based on the assumption that investment funds will be sourced from overseas. The extent to which approved investment proposals will actually be funded from outside of Australia and result in foreign capital inflows depends not only upon whether they are implemented, but also upon the proportion that is financed from foreign sources. The proposed funds to be invested may be contributed by Australians.

Third, FIRB data may include some transactions that do not actually proceed. They include proposals that are approved in a given year but that are not actually implemented in any year; approvals for multiple potential acquirers of the same target; approvals for shares or units where only a portion of the intended shares or units may be acquired.

Fourth, FIRB also points out that policy and legislative change can have a considerable impact on the continuity of data. Changes in administrative practice and foreign investment application requirements have also impacted year-to-year data comparability.

Fifth, there are some factors that especially skew Chinese ODI measurement in FIRB data:

- First, while the acquisitions of interests from foreign government investors in a few sectors in Australia need special approval from FIRB,⁴ SOEs account for a large fraction of Chinese ODI in Australia (for example, over half of Chinese project value was from SOEs in 2016, and this fraction is even higher before 2016 (KPMG/UniSyd, 2013–2017: issue 2017)). This especially increases the significance of Chinese investment in FIRB data.
- Second, Chinese investors have a special interest in the real estate sector. Under Australia's foreign investment framework, foreign persons generally need to apply for foreign investment approval before purchasing residential real estate in Australia. This also improves the significance of Chinese investment in FIRB data.

⁴ See the *Foreign Acquisitions and Takeovers Act 1975* (Cwth) (available here: www.legislation.gov.au/Details/C2016C01144) and the *Foreign Acquisitions and Takeovers Regulation 2015* (Cwth) (available here: www.legislation.gov.au/Details/F2015L01854).

Conclusions

Many people are keen to know ‘how much Chinese investors have invested in Australia’ just as asked in this paper’s title. A number of institutions try to provide answers. Unfortunately, these answers do not completely resolve the question, but, to some extent, create more confusion. This study provides a brief review on the methodologies employed by these institutions to conduct data collection. Table 5 presents a summary of main findings.

Table 5. A summary of Chinese ODI data

Data source	Scope of coverage	Nature of data	Size
ABS	M&A and JV deals with at least 10% voting rights + greenfield projects	real value of direct capital flow	FIRB > KPMG/UniSyd > AEI/ HF > ABS > DCC
DCC	M&A and JV deals with at least 10% voting rights + greenfield projects		
KPMG/UniSyd	M&A and JV deals valued not less than US\$5 million + greenfield projects(?)	contracted value of deals	
AEI/HF	M&A and JV deals valued not less than US\$95 million + greenfield projects		
FIRB	Foreign proposed investment that needs approval	value of proposed transactions	

Source: Author’s summary of datasets.

First, ABS and DCC data only measure Chinese ODI in Australia that comes directly from China. The real value of Chinese capital flow in Australia should be greater than ABS/DCC data. Second, researchers should be cautious when using KPMG/UniSyd data, since no detailed description on data collection and analysis is disclosed. Furthermore, KPMG/UniSyd and AEI/HF data only measure contracted value, which is may be higher than the real value of capital flow. Third, it can be misleading to use FIRB data to measure Chinese ODI in Australia, as it offers only the value of proposed investment, and there are a number of other factors that affect the measurement.

References

ABS (2016). *International Investment Position, Australia: Supplementary Statistics, 2016* (Cat. No. 5352.0). ABS. Retrieved from: www.abs.gov.au/ausstats/abs@.nsf/exnote/5352.0

AEI/HF (2017). *China Global Investment Tracker* [Data set and online tool]. Retrieved from: www.aei.org/china-global-investment-tracker/

- Australian Government (2017). *A foreign investment regime that facilitates investment and protects the national interest* (2017–18 Budget Papers—Fact Sheet). Retrieved from: www.budget.gov.au/2017-18/content/glossies/factsheets/download/FS_ForeignInvestment.pdf
- Callick, R. (2017). Chinese investment in Australia soars to \$4.8bn. *The Australian*, 27 February. Retrieved from: www.theaustralian.com.au/business/economics/chinese-investment-in-australia-soars-to-48bn/news-story/1e3503d4940f69366e583448bf5fe0b7
- DCC (2010). 对外直接投资统计制度 [The Statistics Rules of China's Outward Direct Investment]. December. Retrieved from: images.mofcom.gov.cn/hzs/accessory/201102/1297930121981.pdf
- Drysdale, P. (2011). A new look at Chinese FDI in Australia. *China & World Economy*, 19(4): 54–73. doi.org/10.1111/j.1749-124X.2011.01250.x
- Drysdale, P., Armstrong, S. & Thomas, N. (2016). *Chinese ODI and the Need to Reform Australia's Foreign Investment Regime* (EABER Working Paper No. 25459). 2016. Retrieved from: www.eaber.org/system/tdf/documents/EABER%20Working%20Paper%20117%20Drysdale_0.pdf?file=1&type=node&id=25459&force=
- FIRB (2017). *Annual Report 2015–16*. Commonwealth of Australia: cdn.tspace.gov.au/uploads/sites/79/2017/04/1516-FIRB-Annual-Report.pdf
- Hatzvi, E., Meredith, J. & Nixon, W. (2015). Chinese Capital Flows and Capital Account Liberalisation. *RBA Bulletin*, December, 39–48. Retrieved from: www.rba.gov.au/publications/bulletin/2015/dec/pdf/bu-1215-5.pdf
- IMF (2009). *Balance of Payments and International Investment Position Manual* (Sixth Edition). IMF. Retrieved from: www.imf.org/external/pubs/ft/bop/2007/bopman6.htm
- KPMG/UniSyd (2013–2017). *Demystifying Chinese Investment in Australia* [Report series, issues 2013, 2014, 2015, 2016 and 2017]. Demystifying Chinese Investment in Australia: demystifyingchina.com.au/
- Liu, K. (2016). Chinese Renminbi after 11 August 2015. *JASSA: The Finsia Journal of Applied Finance*, 4, 11–20.
- Prasad, E. & Wei, S.-J. (2007). The Chinese approach to capital inflows: Patterns and possible explanations. In S. Edwards (ed.), *Capital Controls and Capital Flows in Emerging Economies: Policies, Practices and Consequences* (pp. 421–80). University of Chicago Press. doi.org/10.7208/chicago/9780226184999.003.0010
- Scott, J. (2017). Chinese investment in Australia surged 11.7 per cent to \$15.4b last year: report, *The Sydney Morning Herald*, 1 May. Retrieved from: www.smh.com.au/business/chinese-investment-in-australia-surged-117-per-cent-to-154b-last-year-report-20170501-gvw8pc.html

Tan, S.-L. (2017). Chinese investment in real estate grows to \$32b: FIRB. *Financial Review*, 9 May. Retrieved from: www.afr.com/real-estate/chinese-investment-in-real-estate-grows-to-32bn-firb-20170509-gw0sla#ixzz4hPUa2uPw

UNCTAD (2009). *UNCTAD Training Manual on Statistics for FDI and the Operations of TNCs. Volume I: FDI Flows and Stocks*. United Nations Conference on Trade and Development Division on Investment and Enterprise. Retrieved from: unctad.org/en/Docs/diaeia20091_en.pdf

ARGUMENT

Ethics in Economics

Gigi Foster¹

Abstract

This paper presents a slightly modified version of a speech given to the Economic Society of Australia's ACT branch in Canberra in November 2017, as a keynote address at the organisation's annual general meeting. It considers the relationship between economics as a science, and ethical principles both as they guide the actions of practising economists and as they arise in the surrounding social and political context in which economists ply their trade.

Introduction

Thank you very much for inviting me to give tonight's address. In considering what to speak about that would be of most interest and use to the Economic Society of Australia membership here in Canberra, I was inspired by my understanding that my friend Dr Cameron Murray's recent talk to you was very well-received, and by a confluence of recent events and media stories that centre on the theme of ethics. These events and stories focus on behaviour spanning different economic sectors—from academia to the government sector to the private sector—and exemplify Game of Mates-style collusive behaviour along networks of favouritism (Murray & Frijters, 2017), as well as single individuals getting up to various supposed misdeeds all on their own. So tonight, I want to talk about some examples of unethical behaviour, to offer some conceptual musings about ethics from the perspective of economics, and (hopefully of most use to you) to share my views about how to combat behaviour that could be legitimately deemed unethical from the point of view of an economist.

¹ School of Economics, University of New South Wales, gigi.foster@unsw.edu.au.

The ethics underpinning economic science

A sad and surprising lesson I've learned over my 14 years in Australia is that the reputation of economics in the Australian public discourse is not great. The public does not associate our profession with ethical principles (and we could talk for hours about why that is). The reason I was surprised about this is partly that it's not as true in my home culture of America, and partly that my undergraduate major at Yale was Ethics, Politics and Economics, which is Yale's version of the Oxford PPE (Philosophy, Politics and Ethics) program. This program's very *raison d'être* rests on the existence of mutual relevance across the three branches of the degree. I was indoctrinated with this mutual relevance so early in my life that it is second nature to me that economic decisions are inevitably bound up with decisions in the moral and political spheres. To my mind, attempting to deny these profound interdependencies is not only foolish, but dangerous.

Some people seem to get confused about the relationship between economics and morality due to the apparent juxtaposition between 'positive' economics on the one hand, and 'normative' economics on the other. Some of this confusion relates to the idea that scientific inquiry should be objective and unbiased. Let me address this latter idea first.

There is no doubt that the ideals of science, both in Australia and internationally, include the notion that a scientist should do everything in his power to maintain objectivity when searching for truth about the world. Whether or not it is justified to draw conclusions on a given subject that are influenced by personal biases, belief or pre-conceptions has been the question that divides 'scientists-as-a-group' from other groups throughout the centuries, from the condemnation of Galileo by the Catholic Church to the Scopes trial.

Economists are fundamentally scientists, to the extent that we seek ever more accurate theories to explain human behaviour. This then implies that there is a moral role for economists-as-a-group in upholding scientific ideals like objectivity, which we promote when we call for transparency in research, for example. Objectivity itself exists as an aspirational ideal, guiding individual economists like an internal lighthouse, constantly keeping us in check to more or less of a degree. We develop that internal lighthouse through years of indoctrination: it's part of what the slog of graduate school is for.

Ultimately, the useful application of economists' pursuit of truth is that it helps us craft better and better policy advice for the stewards of our economies. Hence we are a special variety of social scientist, identified not only by our adherence to scientific ideals like objectivity, but by our deep commitment to the notion of promoting total social welfare. As expressed succinctly by Henry Hazlitt in his classic primer, *Economics in One Lesson*:

The art of economics consists in looking not merely at the immediate but at the longer effects of any act or policy; it consists in tracing the consequences of that policy not merely for one group but for all groups (Hazlitt, 1962, p. 17; italics in original).

We draw on a raft of conceptual and analytical tools to advise governments on how to lead society towards a situation in which resources are used best, as defined in this way. Our concept of Pareto efficiency captures neatly the ideal world in which ‘the most’ has been produced ‘for all’ with the limited resources available to society.

To say that you wish to see social welfare promoted—or to make any other similar claim about your maximand, such as that you wish to see more social surplus, to get the most welfare bang for a given expended buck, or to move society in the direction of the production possibilities frontier for every productive enterprise—is inherently a moral statement: it is a statement about what matters (and, by implication, what does not matter). As such, a statement like this might be branded as the product of ‘normative economics’ by those who proudly advertise that they instead promote ‘positive economics’.

In my wild youth, I used to proclaim myself a ‘positive economist’, thinking of the term as denoting my intent to engage in an objective and clear analysis of the facts in front of me, rather than having my analytical efforts muddled by trying to incorporate into my analysis things I couldn’t see. The problem with ‘positive economics’, despite the fact that it sounds more similar to the broader scientific ideal of objectivity, is that if it alone is applied to the analysis of an issue, it cannot deliver a recommendation for any policymaker. Like it or not (and most of us do like it), a primary role for economists in the modern world is to advise policymakers on how to spend their scarce resources. If we are not prepared to make any normative judgement at all—if we content ourselves with merely reporting facts (‘inflation has declined’, ‘we spend \$170 billion per year on welfare’, ‘the life expectancy of Indigenous Australians is a decade less than that of non-Indigenous Australians’)—then we fail in our social role of providing advice. Worse, we open the door for others who *are* prepared to make normative judgements to interpret our reported facts in whatever way serves the story they would like to tell. Lobby groups can cherry-pick the facts we produce to argue for whatever policy they favour. If we content ourselves with being ‘positive economists’ in this way, are we truly doing a service for our society, or are we merely offering fuel to be added to the flames of whatever lobby group can afford to pay the cleverest storyteller?

Ethics and the interface of economics and politics

This is where the interdependence between economics and politics comes to the fore. Every politician perceived to have power to allocate resources is lobbied by competing interests, all telling stories about why allocating more to their preferred cause is the right thing to do—the ethical thing to do. These arguments, in Australia and in many other countries, do contain a whiff of the social welfare maximisation goal that forms the centrepiece of the economist's moral code, but they are made up of other ingredients as well.

Amongst other influences, a nation's culture significantly determines what sorts of arguments are put to its politicians by lobbyists or, in less euphemistic words, how the requests for resources of the various groups that plead for them are packaged. Here in Australia, arguments about protecting 'Aussie battlers' and giving everyone 'a fair go', for example, have been powerful players in our historical political discourse, and not always in the service of maximising the good of all. The trade-offs inherent in making any resource allocation decision—following the principle of opportunity cost, beautifully articulated in Bastiat's parable of the broken window—mean that any allocation represents a compromise amongst the interests of various constituents. While an economist may march into Parliament valiantly with a firm commitment to maximising aggregate social welfare, a few days in the thick of politics will make clear to him that things are not so straightforward. Compromise is everywhere, muddying the ethical principles of the economist. Whose welfare should count the most? How valuable is long-run welfare relative to short-run welfare? Which members of society are the most versus the least deserving of an increase in social surplus flowing to them? The politician's predicament—and that of his economist advisers—is like that of the mother bird staring into the open mouths of her many chicks. There is little clear and economically defensible guidance within economics on the question of optimal distribution: it is aggregate welfare of all that matters, not to whom it flows. So, faced with such open-mouth chicks, the economist must step above the fray and ask: which distribution best promotes the good of all?

Of course, this situation should come as no surprise to economists. One of the fundamental tenets of our discipline, if not the bit of doctrine that we are most famous (or infamous) for, is that people are self-interested and greedy. When resources are on the table, when a pie is to be split, everyone wants a slice. They cannot help themselves. We as economists are prepared to admit this fact of human nature and to move forward in spite of it, rather than—as many other disciplines and professions do—trying to pretend that greed is not of prime importance in motivating people, or even condemning humanity for its greed. We pass no negative judgement upon humanity's greed: in this dimension there is no normativity flowing

from our discipline. In fact, our entire discipline is based on the notion that given the reality of greed, certain types of structures are best for maximising social welfare. This pursuit of an ultimately normative, moral objective in the face of hard realities that we do not judge *per se* is what we are engaged in when we argue for the free market. It is an activity that Adam Smith implicitly sanctions in his comment that:

It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities, but of their advantages (Smith, 1776: pp. 9–10).

Not all so-called ‘freshwater economists’ are the demons they are sometimes made out to be: freshwater economics is merely an extremely purist application of the invisible-hand principle to the objective of maximising social welfare.

Even a freshwater economist, with no interest in promoting any distribution of resources for its own sake, would take seriously the recent evidence that more unequal societies grow more slowly (Stiglitz, 2016). This relationship, seemingly driven by mechanisms like weakened aggregate demand and decreased stability when inequality is very high, is the start of a truly pure economic argument in favour of policies to reduce inequality. Rather than trumpeting the notion that inequality is bad *per se*, the economist would argue that if too much inequality holds all of us back in the long run, then on that moral basis alone we should pursue policies that constrain it.

Of course, not every professed freshwater (or saltwater, for that matter) economist is what one might call a true believer in the moral code of economics. Some economists are seduced into highly partisan camps, where they use the technical and critical thinking skills that the discipline imparts to its apprentices to promote the good not of society as a whole, but of whichever subgroup of society employs them. The demand for our services outside of academia and the public service is high not only because the skills taught in economic programs are applicable to such a wide variety of jobs, but also because economists have a reputation as objective scientists: as students of human behaviour, who will speak the truth independently and without bias. That reputation can be free-ridden upon by those who seek to gain personally by offering information or advice under the banner of economics without themselves truly believing in the moral code of the economist. Rather than economists aiming to promote aggregate social welfare, such people can become economics-trained apologists for their employers, trading in the reputation of the economics profession as if it were a currency, lowering themselves into the basket of those pushing more for their preferred distribution of the pie than for what will maximise its size. But this too is to be expected; after all, there is a lot of money and status that flows from economic consulting and, as we all know, people are greedy—they respond to material incentives.

A final note on the interface of economics and politics concerns laws. It can be seductive for an economist to think of the existing laws in his society as the exemplification of what is ethical, but this is incorrect. Laws are merely the codification of what a society has decided should be deemed 'right' and 'wrong'. There is nothing absolutist about them: they evolve as social norms evolve, typically slowly, along with the rest of our social institutions. At any given instant, the activities in an economy are supported by a network of current laws (e.g. contract laws) that make it difficult for the parties to a transaction to violate what the society deems to be 'correct behaviour'. A clear and enforced set of laws that support free, cooperative and market-competitive interactions greatly assists societies in maintaining economic efficiency and stability, and some laws at some moments will be strongly in alignment with the promotion of Pareto efficiency. By no means, however, does any of this imply that economists should view any society's current laws as the almighty and absolute judgement about what is and isn't ethical—by the ethical standards of economics, or by anyone else's ethical standards except those of 'this society today, as a whole'. Ultimately, a law, like any other social institution (or even culture), is subject to the economist's moral calculus about whether in the long run it is good for all.

Examples of economically relevant ethics in practice

I want to share now some examples of ethically laden phenomena that will inform my recommendations about economists' role in identifying and combating unethical behaviour in the course of their practice.

The replicability crisis

One modern incarnation of the longstanding tension between scientists (including economists) and other members of society can be seen in the increased worldwide focus on the motives of social scientists that has been sparked by the so-called 'replicability crisis'. For those who haven't heard of this 'crisis'—and it has been felt more strongly in psychology than in economics, but it is trickling down into our discipline as well—over the past 10 to 15 years there has been a huge increase in the scrutiny of published results in social science, accompanied by public humiliation (sometimes career-ending) for the scientists responsible for producing results that are found not to be replicable by other scientists, or in other ways found not to have resulted from what are held to be the proper standards of scientific inquiry (e.g. Bartlett, 2017). Fuelled by papers with titles such as 'Why Most Published

Research Findings are False' (Ioannidis, 2005), this 'crisis' has given birth to all manner of fantastical new jargon, from specification curves to p-hacking to replicability tests and Caliper tests.

This is tricky territory for a well-meaning scientist to navigate because nowhere is it written down exactly what the 'proper standards' of science are. Who determines the content of scientific ideals? Who is the steward of scientific morality? Ultimately, it is 'scientists-as-a-group' who are the custodians of scientific ideals and, for this reason, the replicability crisis contains the essential characteristic of a purge. The ideals that justify the violence of a purge do not come from outside of the group being targeted, as in the case of your stock-standard war, but from within the group itself—akin to the case of the Spanish inquisition or the Salem witch hunt. The violence that accompanies the modern replicability crisis is perpetrated upon individual scientists who are believed to have failed to uphold the ideals of scientists-as-a-group, and it is justified by those same ideals. The vehemence and fanaticism on display are evidence of how crucial those ideals are to the self-image of scientists: if we cannot claim to the rest of society to uphold ideals like objectivity and transparency, then we as a group lose our identity, our sense of common purpose and our coherence as a profession.

The most important ethical dimension of this problem from the economist's perspective, in which aggregate social welfare is the ultimate maximand, is that if science fails in this way then society becomes the larger loser—space is created for people selling 'alternative facts' that have no less credibility than anyone else's (as highlighted in Laine and Taichman, 2017). With no trusted source of independent analysis about a social scientific issue, society loses its common understanding about what problems actually exist, how big they are and what we might do to solve them. We descend into a chaos of conflicting, baseless opinion, and resources are inevitably wasted. No economist would want such a world. Through this lens then, in spite of the suffering of individual scientists (including some who are doubtless fervent adherents to scientific ideals), there is something about the efforts being spawned from the modern replicability crisis that is morally justifiable to an economist.

Norm evolution in groups

A paper I recently published (Bose et al., forthcoming) tests the malleability of group norms in the presence of what my co-authors and I refer to as 'true believers'. We wanted to know whether the norms of a whole group could be influenced by placing within the group a subset of players who were totally committed to a particular norm—i.e. true believers.

In an experimental lab setting, we asked participants to play simultaneously as giver and receiver in 50 rounds of the Ultimatum Bargaining Game, a standard game used in experimental economics. In this two-player game, the giver is presented

with an initial ‘pie’ (say, \$100) and asked to offer the receiver some fraction of the pie (say, \$20), with the remainder to be retained by the giver. The receiver can then choose whether to accept the giver’s offer, in which case the proposed division is implemented and the round ends, or to reject the offer, in which case both parties receive nothing and the round ends.

Into this environment we introduced computerised participants, indistinguishable to the human participants from the perspective of any individual player, who were programmed to make and reject offers in accordance with very strict rules—i.e. to adhere strictly to a particular norm about which division of the pie was ‘acceptable’.

We found strong evidence that the human players adjusted their offers in the presence of a sufficient number of true believers, whether the norm adhered to was higher or lower than the ‘natural’ offer rate by human givers, in the absence of computerised participants, of approximately 30 to 40 per cent of the initial endowment. However, the humans’ behavioural adjustments only arose in treatments where there was punishment—i.e. in which the computerised true believers rejected offers that were too low, even though that rejection ran against their own short-term interest. Merely offering a certain amount each round, but accepting any offer made, resulted in the true believers making lower profits than the human participants and failing to influence humans’ offers.

Our findings highlight both the malleability of group norms and the potency of group members who are prepared to enforce norms at a cost to themselves. While in our experiment the different possible norms had no natural ethics-based ordering, we expect a similar dynamic to arise with respect to norms that carry an ethical dimension.

International students and the monetary incentives of universities

In another paper, this time to do with the university sector, I looked for academic consequences of universities’ financial incentives to admit full fee-paying international students (Foster, 2012).

Based on a multi-year set of panel microdata on undergraduate students studying in the business faculties at two Australian universities, I documented lower average performance of international and non-English language speaking students, along with negative spillover effects on both learning and academic standards in their classes. The reason to investigate performance differentials between international and domestic students in the first place was derived from economic logic: with little to no academic oversight of the international student admissions process at most universities, and a strong financial incentive on the part of university leadership to admit international

students, there was reason to suspect that the ideals academics adhere to—including academic quality and preparation by commencing students—would not be enforced in the admissions process for international students.

This work was widely covered in the media, but also led to the data streams that had supported this research agenda being cut off by the theretofore-participating universities. I attempted to get the Commonwealth Department of Education to sign on to a program of data organisation and access for researchers, so that more such studies could be undertaken for a wider sample of Australian universities, but such an effort was not deemed politically expedient by departmental officials at the time. This is still the sole paper of its ilk to have been published using Australian data.

The formation and stability of exclusive coalitions

In another working paper, my colleague Ben Greiner and I are using laboratory experiments to explore the ways in which exploitative collusion is supported or suppressed by different communication institutions (Foster & Greiner, 2017). In our laboratory experiment, groups of five participants play repeated rounds of a game in which they elect a ‘dictator’ to split the endowment of the whole group. We are interested in what institutional settings catalyse or deter the formation and stability of a ‘minimum-winning coalition’—a subset of three of the five players, led by the dictator—to whom the vast majority of the entire group’s endowment is allocated. Such exploitative coalitions exclude the other two group members, and in that sense are socially damaging. We test the impact of adding to this election game a ‘proposal’ stage in which all participants can propose a division of the endowment before the election, although these proposals are not binding (similar to the real-world phenomenon of election campaigns); and, in a separate treatment, the impact of having a binding proposal stage.

We find that the non-binding proposal treatment is best for the establishment and stability of exclusive subgroups. The mechanisms behind this effect are still unclear. However, given the laboratory-based design, these mechanisms clearly relate to communication protocols, thus underscoring the importance of communication norms in generating optimal social outcomes.

How to be an ethical economist

Armed with these research-founded ideas about norms, politics and communication, I now want to address the following core question for the remainder of my time tonight: what, to an economist, is unethical behaviour, and when and how should we try to stop it? Specifically, in what cases do we as representatives of the profession

have a moral mandate to call out behaviour as unethical, and what interventions can be effective in stopping such behaviour? The answers I give to these questions here are intended to demonstrate the artistry required to practise economics: notwithstanding the pretty purity of many of our abstractions, making important economic judgements in the messy real world inevitably involves weighing up many different bits of information, based on one's prior knowledge and experience, to work out what is probably best for all in the long run. Such judgements are rarely accompanied by certainty.

Let's begin by considering the example of the replicability crisis. Because of the status one can gain by being seen to be a successful social scientist, there are plenty of examples out there of so-called scientists who are actually not committed to the ideals of objectivity and transparency. In one of my favourite subfields, there are people calling themselves behavioural economists who charge lots of money for advice on how to 'nudge' people to buy goods or services that they do not really need, and others who produce TED talks in which they wave around bottles of allegedly magic 'love potion' (Paul Zak, of the team responsible for Kosfeld et al., 2005). I've already put to you the social welfare argument in favour of maintaining the reputation of social scientists as independent, objective, transparent truth-seekers. If you as an economist buy that argument, and if a purge of the ranks is unavoidable in the quest to maintain the reputation of social science, then you will hold your nose and countenance at least some purging efforts, seeing them as part of the periodic self-cleansing of scientists-as-a-group that serves the interest of public welfare. The hope is that, thanks to these efforts, some of the truly bad apples will find themselves no longer part of the group. While that may hurt those bad apples, this would arguably be a worthy sacrifice because it would be outweighed by the benefits thereby accruing to so many others.

However, there is also a danger that the zealotry of these efforts will bring consequences that overshoot the 'Goldilocks point' of adjustment: consequences like heightened compliance checking for all scientists and attempts to very closely monitor exactly what it is that scientists do in their labs and their offices. We have seen a flavour of this attempt already in the increasing administrative burden—for example, in the area of Human Subject Research Ethics compliance protocols—across Australian universities. The operation of groups in Australia like university ethics boards that have tried to define and take control of 'objectivity'—rather than letting it remain the internal compass for scientists that it is—has resulted in some outcomes that themselves are welfare-reducing (e.g. Paul Frijters' battle with the University of Queensland over his publication of the results of a natural experiment to measure racism on public buses (Mujcic & Frijters, 2013), which ultimately led to his departure from Australia). In that sense, these groups can be argued to have behaved unethically, from the standpoint of an economist. Because I see this descent

from purported promotion of ethics to actual unethical behaviour as inevitable given the current human subjects ethics review structure in place at many universities, I have publicly proposed an alternative system (Foster, 2015).

The underlying problem of how to ensure society produces quality scientific research is twofold: first, no monitoring is perfect; and second, monitoring is extremely expensive for society. Scientists are at the end of the day experts, with the knowledge and skills to circumvent any monitoring system if they really want to. Making life difficult for ‘good apple’ scientists through the establishment of complicated ethical review systems will only hamper scientific progress by frustrating those scientists and taking their time away from actually doing science, following the classic logic of opportunity cost. The ideal state to aim for is one where only those people who are true believers in scientific ideals become scientists, and other people do not, such that only a very low level of external monitoring is necessary.

Naturally, we can never expect to arrive at this ideal state—there will always be a few bad apples in any group—but as with our pursuit of scientific objectivity, we can and should strive for it. Low external monitoring and common ideals to which all group members truly subscribe (which are often, though not always, codified into a system of laws) are characteristics of a healthy group, as they signal stability and the potential for efficient economic exchange on terms that are well-understood and agreed to by all.

In the presence of shared ideals such as tax-paying morale that create efficient internal monitoring of economic players, specific definitions advance neither our understanding of the economic situation nor the outcomes that can be achieved, because of the abstract rather than practical nature of what is actually important. Written rules, templates, forms and other control-oriented mechanisms are merely a by-product: what people hold in their minds to be ‘right’ is far more important in determining many economic behaviours that we observe in modern society. Through our interactions within groups, a sense of ‘right’ gradually endogenously forms and then drives the creation of explicit rules—for example, in our legal system.

Taking a final example closer to home, let’s consider the welfare fraud and robo-debt debacle that Centrelink recently found itself in the middle of (McIlroy, 2017). Taking a big-picture view of this situation, we should expect that some people will indeed abuse the welfare system—we cannot construct a perfect world—and the degree of our outlays of scarce taxpayer money on monitoring and enforcement should be roughly proportional to the degree to which those taxpayers are being ripped off. The total amount of money that welfare fraud is costing the Australian people is a few billion dollars at most, across all individuals committing some form of welfare fraud. An ethical economist considering the right area of monitoring and enforcement in which to spend taxpayer money might note that this figure is very close to the estimated amount of taxation revenue foregone through a much

smaller number of very rich companies who manage through various mechanisms to avoid paying Australian tax on their operations (Khadem, 2017). I would never claim that welfare fraud is a good thing, but tax fraud by large corporations is just as large a drain on the public purse, with its benefits flowing to a significantly smaller number of richer entities.

The economic problem here is not that corporate tax avoidance or welfare fraud contravene politically correct notions of ‘a fair go’ for everyone. The problem is that when we as a society allow people or groups to free-ride upon what we produce and thereby get more pie for themselves—in this case, when we allow companies to use our infrastructure and institutions for a lower price per use than the price paid by ordinary Australians, or when we allow people to take more from our social welfare system than what our laws say is due to them—this can destabilise our society and hence, potentially, our longer-run productive capacity. Perceiving our collective wealth and will to be blatantly trampled upon by a particular subgroup pushes us apart from one another, and can ultimately lead to an abandonment of individuals’ internal commitment to ideals of tax morale and other internal monitoring that supports a thriving economy, as seen for example in Greece today.

Yet which problem, welfare fraud or corporate tax avoidance, is worse? Any economist knows that the marginal utility of income falls as income rises, so if the choice the economist faced was between spending money to recover \$100 million in tax avoided by the rich or \$100 million in tax avoided by the poor, then the most welfare-preserving choice would likely be to spend the resources to recover the tax money from the rich (not to mention that fewer rich people would have to be collected from in order to recover that \$100 million, cheapening the collection effort). This argument would be only strengthened by the progressive effect of such a policy in a country that already has high inequality, such as the US, as per the findings of Stiglitz (2016).

That we are speaking of corporate rather than personal tax means we must also make a judgement on the longer-term economic effects on everyone of increasing the effective (not actual) tax rate on companies, through changes to their activities in the marketplace provoked by having to pay that tax. Many considerations must be weighed here, including how the tax is to be paid back (a one-off recovery effort? A closing of a loophole in the tax code?), how big a change it would really be for each affected company, and the economic importance of the addition to aggregate demand of poor people who take more money from the welfare system than is their due, compared to the economic importance of the provoked decisions of companies to relocate, invest less, hire fewer workers, and so on. My own personal weighing of the relevant evidence leads me to believe that on net, for most developed economies today and for small to medium changes in tax payable per company, pursuing national-tax-avoiding companies is a more economically ethical choice than pursuing national-welfare-system-defrauding poor people.

Finally, what can we as economists do about unethical behaviour that we observe in others? Here are a few ideas, once again grounded in my experience and weighing up of the evidence.

First, whenever we are designing schemes to implement regulation, oversight or any sort of monitoring by the ‘whole group’ (i.e. Australia), we should remember that no monitoring mechanism is perfect. If the stakes are high enough, the target of the monitoring will work out a way to elude it (misreporting, regulatory capture, and so on). For this reason and to minimise the total cost of the monitoring system, limit the costs of compliance for most of those you are targeting. At the same time, to catch the truly bad apples, try to find levers that allow you to reduce the size of the personal gains available to them of doing the wrong thing. Salary caps are one such lever, in cases where observed salaries in a particular level by occupation seem high and politically rather than economically mediated, unable to be clearly linked to performance (which appears to be likely in the case of universities’ vice-chancellors, for example (Lyons & Hil, 2018)); others include instituting legal barriers to operating a business in Australia without paying the associated company tax, or making it more difficult for lucrative but total-welfare-draining networks of favouritism to emerge by periodically rotating the personnel doing the regulating.

Second, try to push for adherence to a common set of clearly articulated, pro-social ideals, in order to reduce the need for external monitoring in the first place. This can mean organising and supporting whole-group efforts, like celebrations on Anzac Day or Australia Day that bind us all together as Australians, or the reporting that appears on our tax statements listing where the money is going that is proof positive that we are all interconnected. It can take the form of setting school curriculums that indoctrinate children with a commitment to the critical moral duty of citizens to make choices that support the health of our nation and the welfare of our society.

There are also the tried-and-true methods of promoting social welfare that most people in this room will know. For example, aim for consumer choice in all markets that can support it, including access to a high-quality government provider; design institutions that require people to pay the true value of a public good or a good with externalities; and, in other ways, work towards reducing the personal gains from free-ridership upon public goods.

On a personal note, do not be afraid to speak up when you see what you feel is a contravention of the principle of social welfare maximisation. Reaffirming the morals of the group, being a punishing role model yourself, helps to keep other people in line. There is no moral in the country more defensible than the desire to make all Australians better off in the long run, so you will always find support for that objective, as long as Australia remains strong. Even if your initial belief about the likely welfare effects of particular policy might be wrong, give voice to it anyway, and make space for more people to speak out. The more open debates

and whistleblowers there are, the fewer costly and unwieldy institutionalised protections there will need to be for them. We need to have more frank discussions about our policy options, and to get deep into the reeds with the pro and con arguments, using objective, scientific arguments and without being side-tracked by defensiveness, bullying or self-righteousness. Is it on balance good for Australia to have the quarantine restrictions that it does? Who are the winners and who are the losers, in the short run and the long run? Is it on balance good or bad for Australia to pursue an industrial policy that supports certain domestic industries? These are economic *and* political decisions, so there are always going to be trade-offs and compromises, but we should push whenever we can for arguments that are clearly tied to the maximisation of the welfare not just of one segment of Australia, but of Australians-as-a-group. That is a goal that unites all Australians, and as economists it is our moral duty to pursue that goal.

References

- Bartlett, T. (2017). Spoiled Science: How a seemingly innocent blog post led to serious doubts about Cornell's famous food laboratory. *Chronicle of Higher Education: Chronicle Review*, 17 March. Retrieved from: www.chronicle.com/article/Spoiled-Science/239529
- Bose, G., Dechter, E. & Foster, G. (forthcoming). Behavioral coordination as an individual best-response to punishing role models. *Journal of Economic Behavior & Organization*. doi.org/10.1016/j.jebo.2017.03.016
- Foster, G. (2012). The impact of international students on measured learning and standards in Australian higher education. *Economics of Education Review*, 31(5), 587–600. doi.org/10.1016/j.econedurev.2012.03.003
- Foster, G. (2015). A new model for research ethics reviews. *The Conversation*, 19 March. Retrieved from: theconversation.com/a-new-model-for-research-ethics-reviews-38296
- Foster, G. & Greiner, B. (2017). *The Anatomy of Exclusion: How selfish subgroups form and are maintained in a repeated dictator election game* (Working Paper).
- Hazlitt, H. (1962). *Economics in One Lesson*. New York, NY: Manor Books.
- Ioannidis, J.P.A. (2005). Why Most Published Research Findings are False. *PLOS Medicine* 2(8): e124. doi.org/10.1371/journal.pmed.0020124
- Khadem, N. (2017). ATO's tax gap figures revealed: \$2.5 billion missing from corporates, multinationals. *Sydney Morning Herald*, 10 October. Retrieved from: www.smh.com.au/business/the-economy/atos-tax-gap-figures-revealed-25-billion-missing-from-corporates-multinationals-20171010-gyxpk1.html
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435(7042), 673. doi.org/10.1038/nature03701

- Laine, C. & Taichman, D.B. (2017). Alternative Facts Have No Place in Science [Editorial]. *Annals of Internal Medicine*, 20 June. doi.org/10.7326/M17-0878
- Lyons, K. & Hil, R. (2018). Vice-chancellors' Salaries are Just a Symptom of What's Wrong with Universities. *The Conversation*, 5 February. Retrieved from: theconversation.com/vice-chancellors-salaries-are-just-a-symptom-of-whats-wrong-with-universities-90999
- McIlroy, T. (2017). 20,000 people sent Centrelink 'robo-debt' notices found to owe less or nothing. *Canberra Times*, 13 September. Retrieved from: www.canberratimes.com.au/national/public-service/20000-people-sent-centrelink-robodebt-notices-found-to-owe-less-or-nothing-20170912-gyg8mm.html
- Mujcic, R. & Frijters, P. (2013). *Still Not Allowed on the Bus: It Matters If You're Black or White!* (IZA Working Paper No. 7300).
- Murray, C. & Frijters, P. (2017). *Game of Mates: How favours bleed the nation*. Australia: Cameron Murray and Paul Frijters.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (Vol. 1) [ebook version]. Hampshire: Harriman House Ltd. doi.org/10.1093/oseo/instance.00043218
- Stiglitz, J. (2016). Inequality and Economic Growth. In M. Jacobs and M. Mazzucato (eds), *Rethinking Capitalism*. Chichester, West Sussex, UK: Wiley-Blackwell in association with *The Political Quarterly*.

REVIEWS

Six Problems in the Biography of Alfred Deakin

William Coleman¹

Judith Brett's *The Enigmatic Mr Deakin* constitutes the fifth biographical study of Alfred Deakin. It succeeds the admiring salute of Murdoch (1999 [1923]), the inscriptional chronicle of La Nauze (1965), the revelatory exposé of Gabay (1992) and the waggish sallies of Rickard (1996). We now have Brett's full-length portrait of the man, in all his parts (Brett, 2017).

But the topic of Alfred Deakin will not be exhausted by even this attention. Let me nominate six problems that remain for future biographers to puzzle over.

1. Deakin's spiritualism

Deakin was a committed spiritualist.

Contemporary observers—you and me—are baffled by how Deakin could so readily soak up such twaddle. His biographers have sought to reconcile his spiritualism into our own understanding by means of two strategies. Both are 'apologetic' in some measure.

1. **Minimisation.** Murdoch and La Nauze present Deakin's spiritualism as a youthful *jeu d'esprit*. This contention, however, became untenable with the remarkable disclosures of Gabay. And Brett has reinforced Gabay with further details of how much Deakin's spiritualist activities extended into his maturity. As late as 1891—the same year he was a delegate at the Federation Convention—Deakin was seeking guidance from the séances of the dubious Mrs Cohen.
2. **Rationalisation in terms of rationality.** Gabay and Brett are inclined to present spiritualism as a wayward progeny of the conquering rationalism of Deakin's

1 College of Business and Economics, The Australian National University, william.coleman@anu.edu.au.

age; a spawn of the decomposition of dogmatic Christianity, and the quickening of 'free thought'. Spiritualism, it is held reinforcingly, is rooted in some of the infant, if erroneous, investigations of human mind; mesmerism and phrenology.

But any attempt to claim spiritualism as one rationalist precipitate of the dissolution of Christianity falls at the starting gate. It is no triumph of the critical intellect to refuse to believe that the dead may be restored to life, but firmly believe they may be talked to. Spiritualism—need it be stated?—was grossly antithetical to science. It was, specifically, devoid of the key impersonality of science, and beholden to what might be called a suite of self-selected miracle workers. Given this, there may be a temptation to reverse direction, and trace Deakin's spiritualism to deficiency in his critical faculty. Deakin's mind was essentially artistic, was it not? So we dispose of Deakin's spiritualism by invoking his inclination to gambol in the folds of fancy. But this is too pat: it is an unnerving truth that several distinguished natural scientists, and Cambridge philosophers, were firm adherents to spiritualism—or at least sympathisers.

Rather than diagnose spiritualism as a product of its believers' rationality or irrationality, it might be more useful to seek its success in how it served the values of the period in which it thrived; the mid-19th century to about 1930, a near-normal life span of someone born in Deakin's birth year, 1856.

Spiritualism gratified one key value of its age; an intense individualism. Historically, individualism had, in religion, been expressed through Protestantism. This had eliminated, first, the rule of the priest, then later the rule of scripture, and then, with 'modernism', even the rule of the Almighty. But one restraint remained; the fundamental and near unbridgeable chasm between This World and The Next. It was spiritualism that destroyed this last constraint on the individual will. It annihilated the dichotomy of the sacred and the profane; in its own language, it allowed you to sweep aside the 'flimsy film' between the two worlds. You were no longer confined in This World; you could summon your dead loved ones at chez Cohen, with almost as much ease as you could buy there a pound of nuts from her almond tree. Granted: it is this destruction of the distinction between the sacred and the profane of that made spiritualism so vulgar, sometimes hilariously so. But here was a doctrine that gratified the invincible will of a Syme and hundreds of others of his ilk. Including Alfred Deakin.

When conjoined to the equality that Deakin's age also aspired to, individualism additionally served spiritualism by a second, if indirect, route: by repudiating a necessary foe of spiritualism—intellectual authority. If individualism and equality are taken literally, then all authority is redundant. And not just religious authority, but also scientific authority. Thus the 19th century was not only the age of spiritualism, but also flat earth propagandising. And it is no great surprise that there was an

overlap in membership of these two movements (Gardwood, 2007). Individualism and equality, taken literally, freed one from religion; and from science, too. Thus spiritualism: bad science, and still worse religion.

The significance of spiritualism being so congenial to its age is that Deakin was so acutely of his own age. The man was rootless; quite without any tie with the past, such as grandparents or family history. He was wholly of his newborn, ‘instant society’ of Melbourne; a society that was not only, inevitably, without memory, but one that with an equal inevitability identified itself with the new.

So to formulate the thesis: Deakin, being so perfectly of his point in time, was vulnerable to the religious responses that were so characteristic of the time.

2. Frederick Deeming

Frederick Deeming was a serial killer who in 1892, amid great publicity, was put on trial for murder in Melbourne, and executed. His defence counsel was Alfred Deakin. Why would Deakin choose to take on the defence of a psychopath? Ordinary financial motives did not apply; Deakin was not remunerated. So: why would a politician in mid-career take on a case seemingly prejudicial to the future success of that career?

His biographers treat this incident briefly and without explanation. Deakin’s enemies did not forget it, but neither do they explain.

The authors of the one extended study of Deeming case suggest of Deakin: ‘Presumably he believed that the Deeming case would help restore his name as a barrister’ (Gurvich & Wray, 2000, p. 202).² So a spectacular victory would make Deakin’s reputation, just as Paddy Crick had won a reputation with the Buttner case 1889; and, in 1904, came within a Governor’s scruple of becoming Premier of New South Wales. Or as G.H. Reid acquired prestige in 1888 in securing, by means of his ‘brilliant cross-examination’, the acquittal of one of the accused in the Mt Rennie rape case (McMinn, 1989, p. 41).

There is another possibility: Deakin’s willingness to take the case arose from his own peculiar views on the mind, arising from his spiritualism. Deakin’s defence of Deeming was that he was ‘insane’. Insanity is a weak descriptor of Deeming: he was plainly a sociopath and not a psychotic. But mental illness was poorly understood; thus Deakin’s ridiculous suggestion that Deeming could have committed the murder in an epileptic fit. Might Deakin have had a particular interest in the ‘Sacred

2 The word ‘restore’ is unsuitable: Deakin had no reputation as a barrister in 1892.

Disease'? He believed, or at least claimed, that he had written a book—*A New Pilgrim's Progress*—in a frenzy of 'automatic writing', where he had lost his mind to some directing other.³

Neither of these candidate motivations, however, reduce the damage Deakin's career would presumably have suffered if he had actually won. Reid politically benefited from extricating his client from the dock only because the public believed that the Mt Rennie accused had not received justice. Similarly, Ernest Buttner was probably innocent of rape, and Crick was probably saving an innocent man.

Was it possible that in 1892 Deakin had abandoned hope of a further political career?

3. Sexual orientation

Put simply, 'was Deakin gay?' The answer, it seems, is 'fairly gay'.

We might begin with the fact that, as a boy, he had 'a certain girlishness of appearance' (Murdoch, 1999, p. 20), and was known by his peers at Melbourne Grammar as Miss Deakin, Polly and Pretty Polly. This impression of effeminacy did not, however, survive adolescence, and he disliked receiving the impression of it in others.⁴

As a young adult, Deakin evidently could hold considerable appeal to men much older than himself. Thus a (childless) squatter, Alexander Strachan, bequeathed to him the bulk of his estate. Another squatter, Sydney Grandison Watson, on the basis of slight acquaintance, invited the 23-year-old spiritualist to join him for a 'delightful trip' to Fiji (Deakin, 1957, p. 44). Over even the remote, granitic monolith of David Syme, Deakin appears to have held a distinct beguiling power. Despite the two being 'unlike each other in almost every way' (Sayers, 1976), a close tie grew between them; Deakin was granted the privilege of 'free access to Syme's private room at all hours' (Murdoch, 1999, p. 55).

There was, no doubt, an asymmetry in these attractions. His popularity with elderly men appears to have been of little more than an instrumental value to Deakin. But with men not so different in age, Deakin was certainly emotionally involved on some occasions. A pattern emerges of close attachments to men either a decade or so older than himself, or about a decade younger. In his youth, these include David Mickle, 12 years his senior, and, most importantly, Frederick Derham, businessman and politician to whom Deakin expressed what he betokens as 'that love which passeth the love of woman' (Brett, 2017, p. 109). On his departure overseas, Deakin wrote

3 This is revealed in the book's full title: *A New Pilgrim's Progress: Purported to be Given By John Bunyan, Through an Impressional Medium* (published W.H. Terry, 1877).

4 But A.F. Davies (1973) judged the adult Deakin to be 'covertly feminine'.

to Derham, 'Men must not seem women but believe me it is a pull on my heart strings to say good bye for you for so long' (Brett, 2017, p. 125).⁵ The Victorian age was, certainly, an age of agonised partings.⁶

As he moved into middle age, Deakin took the role of the senior member of these dyads; Walter Murdoch (18 years younger) and—above all others—Herbert Brookes (11 years younger) assumed the junior role. As Brett put it, 'Deakin had no male friend whom he loved, and he took Brookes immediately into his heart' (Brett, 2017, p. 234).

There is no evidence that these attachments were sexual in act. And it would appear he never conceived them as having a sexual dimension. But to simply categorise them as 'male friends' seems to banalise them in a way untrue to their significance. The problem is, how quite to conceive them? The relationship of Syme and Deakin puts in mind nothing more than the king and his male favourite. But how to categorise that? Part of the difficulty is in translating the words and symbols of the 19th century into our own. Thus, while in London in 1887, James Service records that Henry Holland, Secretary of State for the Colonies, 'has fallen in love' with Deakin (Murdoch, 1999, p. 116). We are left wondering how much is the past another country; or, how much another country is another country.

What may be said with confidence is that Deakin found recreation almost only in male company, and barely at all in female. His remarks on women's physical attractions rare and terse.⁷ By contrast, his two political memoirs reveal a distinct sensitivity to male beauty. His political allies are 'handsome' (Quick, described in Deakin, 1944, p. 55), 'extremely handsome' (Playford, described in Deakin, 1944, p. 29), and 'one of the handsomest young fellows I have ever seen' (a youthful supporter in Victoria, described in Deakin, 1957, p. 49). And Barton had 'eyes of remarkable beauty' (Deakin, 1944, p. 32). The significance of these observations is reinforced by Deakin's disdainful references to the (heterosexual) love shack that was NSW politics of the day: Reid is a 'squire of dames' (Deakin, 1944, p. 52), Jack Want is Reid's 'brother bachelor' (Deakin, 1957, p. 54). In venting his seething contempt of William Lyne, Deakin notes, with an air of triumph, that the Premier's 'vagaries' were known to him. I would guess this is a reference to Lyne's apparent mistress-keeping.⁸

Deakin seems to have been quite beyond the force of female eros.

5 Deakin's relationship with Derham is wholly ignored by La Nauze and Murdoch.

6 Thus, John Robertson 'wept like a girl' (his words) at the harbourside departure for Europe of W.B. Dalley.

7 In a peculiar inversion, he on one occasion refers to a woman as 'dashing' (Deakin, 1944, p. 61).

8 In 1911, Lyne was to secretly marry—in Montana—Sarah Olden, 25 years his junior. I am indebted to John Hawkins for this detail.

So what was to be the root of his relations with women? His play *Quentin Massys*—a technically brilliant accomplishment for an 18-year-old—prefigures in miniature his vision of the foundation of male–female relationships; it was a matter of the strong protecting the weak. Desexualised chivalry, in other words. This vision was realised in the domestic ménage Deakin established at the age of 25, with his marriage to Patty Brown. This produced three daughters, but was cold at the core. Both exacerbating this frigidity, and reflecting it, was the peculiarly significant relationship between himself and his sole sibling, Catherine (Kate). ‘I have worshipped the baby as I have the man all our lives’, Kate once declared (Brett, 2017 p. 17). For his part, Deakin seems to have elevated Kate above Patty. In Deakin’s chosen words towards the end of his life (quoted in La Nauze, 1965, p. 11), his father, mother and sister were ‘the Trinity’ under whose ‘shelter, guidance tenderness and forgiveness’ he had lived. His wife is not mentioned.

Thus resulted a disastrous confusion of roles; of wife, sister, daughter.

Kate’s bed was the bridal bed of Alfred and Patty. Alfred expressed a wish for Kate to accompany him on his honeymoon. On their father’s death, Deakin invited Kate to with live with himself and Patty; a mix-up of even oedipal categories. When—at Patty’s bitter protests—Alfred withdrew this invitation, he, in Brett’s words, ‘broke Catherine’s heart’ (2017, p. 191): a somewhat unexpected usage of Brett, but not inappropriate, given the role confusion the Deakin siblings laboured under. In this confusion, Alfred and Kate would make decisions about the three daughters in conferences from which Patty was absent; indeed, unaware of. Kate ‘acts as if she owns the family’ complained Patty, while she was reduced to a nanny; a ‘dearest child’; and even mistaken by strangers for Alfred’s sister.

Does Deakin’s sexual orientation matter? Not in the ‘Alfred the Lawgiver’-type biography of La Nauze. But the topic belongs in any full-length portrait. And the implicit assumption that Deakin was conventionally heterosexual has also prompted false solutions to certain puzzles. At the age of 32, Deakin composed a ‘prayer’ that has left a mystery for biographers:

After again a long silence something has been awakened in me by the burning iron of remorse. Aid me O God to atone for the past and if it be possible to undo even remove the evil done to others. After this enable me to conquer the evil which it had done to others. After this enable me to conquer the evil which it has done to myself and to kill the root of that evil within me. (Deakin, n.d.)

Some have speculated it was a piece of sexual infidelity for which Deakin felt the burning iron of remorse (Langmore, 1992). But there is no evidence for this hypothesis.

So what was ‘the root of that evil’ within him?

4. His final illness

While he still in his 50s, Deakin's epigrammatic, lucid and conjuring mind was overthrown, leaving him, in his own words, 'a shadow in a dream'. Beyond the complete collapse of memory, his symptoms include giddiness, dizziness, exhaustion, a 'wavering' heart, chronic insomnia, weight loss and gastric upsets.

Brett diagnoses this disaster as progressive vascular dementia, produced by hypertension. This hypothesis is simple, and well-hinged on the one objective measurement of his health known to biographers; at his retirement in January 1913 his systolic blood pressure stood at 164.

But can this hypothesis bear the burden of all particulars?

Vascular dementia is 'uncommon' in persons less than 65 years of age (Alzheimer's Society, 2014). Deakin was dead at 63.

A diagnosis of vascular dementia sits also uneasily with the fact that his illness barely reduced his impulse—if not compulsion—to take up the pen and cover the page, even when his condition was well-established. If it had any effect, it was to reduce, on occasion, the control he exercised in passing the pen over the page.

Third, vascular dementia has a mean duration well below a decade.⁹ Deakin's illness lasted well above a decade. At the age of 56, he retreated from public life in the face of his illness. From the age of 52, at end of 1908, he was forgetting things he had just read. At 51, in September 1907, he was, in Hughes' contemporary judgement, 'a very sick man. Nervous breakdown' (quoted in La Nauze, 1965, p. 424). In A.B. Piddington's retrospective judgement, Deakin was already losing his mental grip on the occasion of his Notice to Quit speech to G.H. Reid of June 1905, when he had not yet turned 49. In 1901, the year of his 45th birthday, Deakin was 'frequently unwell' (Brett, 2017, p. 255) and stooped.

Similar doubts cloud, without eliminating, the suggestion that his illness was 'almost certainly Alzheimer's disease' (Headon, 2018). Alzheimer's will not often kill by the age of 63. But it may. It will generally not last longer than a decade; but, again, it may. However, Alzheimer's typically does damage skill in longhand, and there is no sign of that in Deakin.

Is there a more likely diagnosis?

This paper advances two hypotheses.

⁹ 'Some data suggest that those who develop dementia following a stroke survive three years, on average': www.alz.org/dementia/vascular-dementia-symptoms.asp.

Over the past 40 years, the increased sensitivity of diagnostic markers have revealed that vegetarians are at severe risk of vitamin B12 deficiency. The consequences include ‘irreversible neurological damage’ cognitive difficulties, memory loss, fatigue and loss of balance (Herrmann & Obeid, 2008). As the body is endowed with reserves of B12, a deficiency can take years to develop in the face of a vegetarian diet.

The significance of vitamin B12 deficiency is that Deakin was a strict vegetarian in young adulthood (Gabay, 1992, p. 11). He relaxed this regimen somewhat in his maturity.¹⁰ However, it is known that once a deficiency is established it is not remedied by simply switching to a non-vegetarian diet.

The second hypothesis is more sensational. It invokes, not the Emperor of all maladies, but their Lucifer, syphilis.

The hypothesised identification rests on the mental collapse (‘neurosyphilis’) that is often occasioned by syphilis in its ‘tertiary’, or concluding, stage.

‘His mind grew weak; his utterance became hesitating, for want of the right word to express is meaning; his gait tottering ...’ (Beaney, 1872, p. 204). Thus speaks one 19th-century report of a sufferer of neurosyphilis.

Granted: many illnesses can produce such symptoms of mental disarray. But the hypothesis of syphilis sheds an apparent gratuitousness on account of its shocking prevalence, by our standards, before the antibiotic revolution. Thus, the British Royal Commission into Venereal Disease of 1913 estimated, on the basis of results of Wasserman tests, that 8.3 per cent of the male upper- and middle-class residents of London in their mid-30s were infected (Szreter, 2014).¹¹ A substantial minority of these, say around a third, would eventually develop neurosyphilis, say about 2.75 per cent. A calamity.¹² Epidemiological considerations suggest that 19th-century Australia would not be more fortunate than London; rather less, in fact. The epidemiological predisposing factors for syphilis are considered to be: a booming economy, a port town and a population with a high ratio of men to women.¹³ The Melbourne of Deakin’s youth would have been a syphilitic inferno.

10 Meat at lunch remained a rarity in Deakin’s regimen (Rickard, 1996, p. 102).

11 Other attempts at measurement of the prevalence of syphilis in developed countries before antibiotics also produce a number in the region of 10 per cent (Zajdowicz, n.d.)

12 By comparison, the prevalence of vascular dementia in the 65–69 age group is broadly around 0.1 per cent (Dubois & Rejean, 2001). Other candidate causes of mental collapse are still rarer. Thus CADASIL—an ‘inherited form of vascular dementia ... caused by a mutation on the Notch 3 gene’, and proposed as the cause of the madness of Nietzsche and Ruskin—is ‘very rare’ (www.dementia.org.au/about-dementia/types-of-dementia/vascular-dementia).

13 The 1881 Victorian census records 123 men aged 21 and over for every 100 women so aged.

The toll of those stricken certainly includes J.F. Archibald.¹⁴ The exact coincidence of his years of his birth and death with Deakin's (1856–1919) bodes uneasily.¹⁵ The toll would also include Walter Richardson of 'Richard Mahony'; William Astley, Barton's press officer in the Federation campaign (Andrews, 1976, p. 166); and John Curtin's father (Day, 2006, p. 224). I strongly suggest the fatalities would extend to the spiritualist John Tyerman (1838–1880); the first registrar of Sydney University, Hugh Kennedy (1829–1882); and Melbourne law academic J.B. Gregory (1844–1910), the originator of the 'Gregory Fractional Transfer' in the Hare-Clarke system. One may easily extend the list of Australian noteworthies of the period whose death from 'insanity' or 'softening of the brain' before the age of 60 are grounds for a dismaying apprehension.

I have proposed; let medical expertise dispose.

5. Extraordinary seamen

Deakin's memoir of Federation (Deakin, 1944) includes his revelation of a remarkable escapade that he, the Victorian minister for defence and the Victorian Premier resolved upon in 1888 or 1889. The background was the friction between Britain and France over the New Hebrides; and Alfred Deakin's fixation on these islands. In November 1887, the two imperial powers had agreed to jointly assure the life and property of their respective subjects residing in the New Hebrides, but otherwise make no claim of sovereignty. And there the matter seemed to rest. But, says Deakin, rumours subsequently arose that French annexation was imminent:

It was therefore decided to forestall the French by dispatching a detachment of the Victorian permanent military forces in a swift steamer with orders to hoist the British flag and keep it flying (Deakin, 1944, p. 22).

To achieve requisite surprise, the forces were to covertly embark on a chartered boat while ostensibly on a training manoeuvre.

This foolhardy scheme appears to have remained completely uninvestigated by historians, and, apart from Murdoch, not even noticed.

¹⁴ I venture the account of Archibald's medical condition supplied by his biographer could not be interpreted otherwise. At times Archibald believed his condition was syphilitic (Lawson 1971, pp. 218–20).

¹⁵ As is the use of the phrase 'nervous breakdown' by Archibald's devotees to describe, and disguise, his collapse in 1906.

6. Who was the master?

Deakin's politics have been an object of contention of commentators, especially as to which contemporary political tendency may claim his legacy.

The defect with any concern to place Deakin on an ideological spectrum is that, for all his bibliomania, he was not a person of ideas; certainly not of systems of ideas. He was no theoriser.

Neither was Deakin a person of causes. Certainly not lost causes, as Brett notes. As Deakin conceded, scruples in doctrine soon departed him (Deakin, 1957). Winning causes, he favoured. Thus the stereotypical character of his various policy stances. It is hard to think of any that was not perfectly of his time and place.

Deakin, then, was one of the breed for whom policy is not an end, but a means. But a means to what? For Deakin it was the power, not the glory. Or to put the matter more precisely, mastery: mastery of séances, audiences, caucuses, chambers, cabinets. It was entirely characteristic to exalt in 1909, 'I have become more than ever the pivot of the whole situation' (Murdoch, 1999, p. 263).

Deakin's quest for mastery might in turn be resolved into that distinct sense of superiority which he clearly suns himself in his memoirs, and which could lead him to compare himself on occasion with Jesus Christ (Rickard, 1996, p. 112). In a word, narcissism. Not to vanity; Deakin was proud, rather than vain. The plaudits of others had little value. He had no interest in prizes at school; and he had no interest in the prizes of adult life. Thus his remarkable refusal of an offer of knighthood from the British Prime Minister at the age of 30. To rebuff a knighthood was not actually rare; those who rejected knighthoods included Kingston, Playford, O'Connor and Pember Reeves. For all that, his refusal remains an impressive, almost breathtaking, act. But it was not done out of republicanism—which Deakin repudiated. And certainly not humility. Quite the opposite. For Deakin, he would be the one to confer any honours. The only crown he wanted was that which he would place on his own head. The only cross he would care for is the one he would nail himself to.

This conviction of superiority fostered a distance and an inner coldness that belied the appearance of Affable Alfred. In his adolescence—the age of camaraderie—Deakin had no friends amongst his peers. The adult friendships he later formed were not insincere, but they were largely epistolary; they were, literally, 'composed'. Unlike so many of his political peers, he was a member of no club.

This emotional distance in Deakin accommodated a ruthlessness, which extended treachery.¹⁶

16 Deakin bears a comparison with Lloyd George, not only in his ruthlessness, but also in his oratorical power over audiences; his immunity to flattery; and the fact that he lived 'the life of the extrovert and yet had the strength of the introvert's inner life' (Salter, 1947). Dare one also mention Welsh provenance! Always attracted to the martial, one wonders what if Deakin, like George, had been his country's warrior champion in the First World War?

Deakin's parliamentary career began and ended with betrayals of trust.

As a freshly elected member of the Victorian parliament, Deakin announced to an 'incredulous' House his resignation (Gabay, 1992). He had won his seat only on account of the particular preferment that Berry had awarded him. He had not advised Berry of his decision to resign, or anyone else. His patrons had no intimation of this act of desertion.

And his parliamentary career closed with him voting for Joseph Cook as leader of the Commonwealth Liberal Party, rather than his deputy Forrest, to Forrest's perplexed distress.

Mileposts in the journey between include his repudiation in 1905 of the Protectionists' compact with Reid, which not only stranded certain rivals, but left Turner 'stabbed in the back by one of his oldest friends' (McMinn, 1989, p. 225). The frenzy of indignation that Deakin's abandonment of Fisher produced in Lyne, a stable personality if there ever was one, bespeaks an acute psychic wound.

Ruthless as he was in the deployment of power, the question still remains: how successful was Deakin in winning it? One locus for the investigation of that query would be Deakin's relationship with Barton. It seems that Deakin was successful in manipulating the glory-hunting Barton to secure his own ends. Thus, Deakin successfully forbore on Barton to appoint Kingston to his cabinet. And what of the greatest prize: the post of prime minister, which Barton so conveniently vacated in 1903? 'Take PM' was Syme's advice to Deakin in this period. Which brings us Deakin and Syme. In this game of king and the royal favourite, who was truly the master?

The Ariel Mr Deakin

The six questions above are unresolved. This lack of resolution, however, does not, I believe, lend itself to the moniker of Deakin the Enigmatical that has become popular. It was Deakin who referred to himself as 'enigmatical'. Do any contemporaries refer to him this way? True: his cabinet minister Patrick Glynn records Deakin as a 'puzzle'. And there are puzzles about Deakin's life—as there are, presumably, with every figure. But even the outright contradictions that Deakin does present—his arrogance in the company of others, and his humility in the company of himself—does not to my mind make for an enigma. An 'enigma' is closer to an uncertainty as to *what* is going on, rather than *why* it is going on. It is G.H Reid who receives the judgement of enigmatical from both his contemporaries (Piddington, 1929; Lang, 1956) and later historians (Martin, 1980), and with justice.

I would rather dub Deakin a marvel than a mystery. We might think of an Ariel Mr Deakin. Deakin's Ariel in service of Swedenborg's Prospero. A Shakespearean fairy; not inhuman, but unearthly; entrancing by words, rather than notes; manipulative and mischievous; hastening to put his girdle 'round the Earth'—and simple mortals—in 40 minutes.

References

- Alzheimer's Society (2014). *What is Vascular Dementia?* Factsheet 402P. Alzheimer's Society UK. Retrieved from: www.alzheimers.org.uk/download/downloads/id/2427/factsheet_what_is_vascular_dementia.pdf
- Andrews, B. (1976). *Price Warung (William Astley)*. Boston: Twayne.
- Beaney, J. (1872). *Constitutional Syphilis: Being a practical illustration of the Disease in its Secondary and Tertiary Phases*. Melbourne: F.F. Bailliere.
- Brett, J. (2017). *The Enigmatic Mr Deakin*. Melbourne: Text.
- Davies, A.F. (1973). *Politics as Work*. Melbourne: Melbourne University Politics Department.
- Day, D.A. (2006). *John Curtin: A life*. Sydney: HarperCollins.
- Deakin, A. (n.d.). *The Boke of Praer and Praes* [Prayer Diary]. Papers of Alfred Deakin (NLA MS 1540/5/1457. Typescript: NLA MS 1540/5/818). National Library of Australia, Canberra.
- Deakin, A. (1944). *The Federal Story: the Inner History of the Federal Cause* (Herbert Brookes ed.). Melbourne: Robertson and Mullers.
- Deakin, A. (1957). *The Crisis in Victorian Politics, 1879–1881: A Personal Retrospect* (J.A La Nauze and R.M. Crawford eds). Melbourne: Melbourne University Press.
- Dubois, M.-F. & Rejean, H. (2001). The incidence of vascular dementia in Canada: A comparison with Europe and East Asia. *Neuroepidemiology* 20(3): 179–87. doi.org/10.1159/000054785
- Gabay, A. (1992). *The Mystic Life of Alfred Deakin*. New York: Cambridge University Press. doi.org/10.1017/CBO9781139085007
- Garwood, C. (2007). *Flat Earth*. London: Pan Macmillan.
- Gurvich, M. & Wray, C. (2000). *The Scarlet Thread. Australia's Jack the Ripper*. Sydney: Fairfax.
- Headon, D. (2018). *An Enigma Revealed: The mystery and mysticism of Prime Minister Alfred Deakin*. 9 March. Speech given at Old Parliament House, Canberra.

- Herrmann, W. & Obeid, R. (2008). Causes and Early Diagnosis of Vitamin B12 Deficiency. *Deutsches Arzteblatt*, 105(40): 680–85.
- La Nauze, J.A. (1965). *Alfred Deakin: a Biography*. Melbourne: Melbourne University Press.
- Lang, J.T. (1956). *I remember: Autobiography*. Sydney: Invincible Press.
- Langmore, D. (1992). *Prime Minister's wives*. Ringwood, Victoria: Penguin Books.
- Lawson, S. (1971). *J.F. Archibald*. Melbourne: Oxford University Press.
- Martin, A.W. (1980). *Henry Parkes: A biography*. Carlton: Melbourne University Press.
- McMinn, W.G. (1989). *George Reid*. Melbourne: Melbourne University Press.
- Murdoch, W. (1999 [1923]). *Alfred Deakin: A Sketch*. Melbourne: Bookman.
- Piddington, A.B. (1929). *Worshipful masters*. Sydney: Angus & Robertson.
- Rickard, J. (1996). *A Family Romance: The Deakins at Home*. Melbourne: Melbourne University Press.
- Salter, A. (1947). *Personality in Politics: Studies in contemporary statesmen*. London: Faber and Faber.
- Sayers, C.E. (1976). Syme, David (1827–1908). In *Australian Dictionary of Biography*. Melbourne: Melbourne University Press.
- Szreter, S. (2014). The Prevalence of Syphilis in England and Wales on the Eve of the Great War: Re-visiting the Estimates of the Royal Commission on Venereal Diseases 1913–1916. *Social History Medicine*, 27(3): 508–29. doi.org/10.1093/shm/hkt123
- Zajdowicz, T. (n.d.). *Syphilis Epidemiology* [PowerPoint presentation]. Chicago Department of Public Health.

Alex Millmow's *A History of Australasian Economic Thought*

(Routledge, London, 2017)

Reviewed by Selwyn Cornish¹

Several surveys of Australian and Australasian economics have been produced over the last half century or so. Craufurd Goodwin's *Economic Enquiry in Australia*, published in 1966, covered the 19th century and ended in the 1920s. Peter Groenewegen and Bruce McFarlane's *A History of Australian Economic Thought*, published in 1990, covered both the 19th and 20th centuries, but it, like Goodwin, excluded New Zealand. The present reviewer's essay on 'Economics in Australasia', published in 2008 in the *Revised New Palgrave Dictionary of Economics*, included economic thought in Australia and New Zealand in the 19th and 20th centuries, as did William Coleman's chapter in the *Routledge Handbook of the History of Global Economic Thought*. The latest survey, by Alex Millmow, an Australian National University PhD graduate in economic history, now Associate Professor of Economics at Federation University, Ballarat, embraces both Australia and New Zealand. However, despite its title—*A History of Australasian Economic Thought*—the book is restricted to the 20th century, or more precisely from the mid-1920s, ignoring writing on economic subjects in the 19th and early 20th centuries. Even so, it will provide a useful guide for those wanting information about what was written, by whom, and in which year, between the 1920s and the turn of the century.

One of the difficulties of writing a book on the history of economic thought in two countries over nearly a century, with hundreds of writers and their many contributions, is finding a systematic and coherent method of organising the material. For the failure to devise a satisfactory framework could easily result in an amorphous mess. Millmow has avoided this danger for the most part by structuring the chapters of the book in chronological order and selecting work to be covered

¹ Research School of Economics, College of Business and Economics, The Australian National University, selwyn.cornish@anu.edu.au.

in each chapter according to pertinent themes. The book comprises 11 tightly and well-written chapters. The first three substantive chapters, following a general introduction, concentrate on the professionalisation of Australian economics—and to a lesser degree New Zealand economics—as departments and faculties of economics were established in universities, and governments began to seek out economists for advice as economic conditions deteriorated from the late 1920s. Economists played an even greater role in these countries during the Second World War, when economists were appointed to advisory committees or attached to government departments and agencies; many of them decided after the war to remain permanently in government service. Virtually everyone who has written on the history of economic thought in Australia agrees that these decades—the 1920s, 1930s, 1940s and into the 1950s—constitute the heyday of Australian economics, when the nation produced a remarkable group of economists, assisted by some who were New Zealand-born (the occupants of the first chairs in economics at the universities of Sydney, Tasmania and Melbourne, for instance, were born in New Zealand).

In the late 1920s, the so-called Brigden Report—*The Australian Tariff: An Economic Inquiry*—commissioned by Prime Minister Bruce was dominated by economists, including J.B. Brigden, D.B. Copland and L.F. Giblin. It was also at this time that Giblin formulated his (export) multiplier, a precursor to Richard Kahn's expenditure multiplier. Then there was the advice provided by economists during the depression of the early 1930s. About the Premiers' Plan (perhaps more accurately, the Copland Plan) prepared in 1931 by a group of economists, led by Copland and including Giblin, L.G. Melville and E.O.G. Shann, John Maynard Keynes in 1932 declared that he was 'sure' that it 'saved the economic structure of Australia' (Keynes, 1932, in the *Melbourne Herald*). Often neglected in histories of economic thought are the submissions and testimony by virtually all the leading Australian economists to the 1935–37 Royal Commission on the Australian Monetary and Banking Systems. Melville, the Commonwealth Bank's chief economist, in his submission, advocated the adoption of a monetary framework for Australia that was remarkably similar to what became known as the 'Bretton Woods system', based upon a fixed but adjustable exchange rate, with monetary policy guided by movements in price levels. On the eve of the Second World War, D.H. Robertson, the renowned Cambridge economist, could justly say that:

[t]here are said to be, in the far north [Sweden] and the far south [Australia], lands where economists all give the same advice, where the government listens to it, where the public understands why the government has listened. (Millmow, 2017, p. 49)

Again, the strategies adopted in Australia for financing the Second World War and the development of policies for post-war reconstruction, both domestically and internationally, owed a great deal to the work of local economists. Of Roland Wilson's performance at the first British Commonwealth conference in London in 1942 devoted to post-war reconstruction, Keynes said he:

took a prominent, indeed a leading part through all the discussions and played a major role in them with the greatest success. We had a great gathering of Whitehall officials, who came to feel great respect both for his wisdom and for his pertinacity. (Reserve Bank of Australia Archives, 1942, Keynes to Giblin, 13 November)

The principal economic policy that Australian governments formulated for the international economy in the post-war era—the 'Positive Approach' or 'Full Employment Approach'—was produced by three leading economists, Giblin, Melville, and H.C. Coombs. And it was Wilson again, who, as the leading economist attached to Australia's delegation at the 1945 San Francisco Conference that established the United Nations, was responsible for the adoption of Articles 55 and 56 of the UN Charter, which dealt with employment and unemployment.

Millmow is an authority on this period and the exceptional group of economists who began their careers in the 1920s and 1930s, and he covers the ground remarkably well. It is a story of a group of young economists—Giblin being the exception in terms of age—who worked assiduously and with considerable acumen and skill to moderate the effects of the international catastrophes of the age—the First and Second World Wars and the Great Depression. But he is not as successful in writing about developments in Australian economic thought from the 1960s. There are obvious reasons for this. One is that the number of economists working in Australia increased considerably, and the number of names that enter and depart the scene is more difficult to absorb and their work more difficult to digest. Another is that governments no longer sought outside economic advice to the extent they did before the 1960s—for several reasons. Large numbers of economists were now working in government departments and at the central bank; ministers began to appoint economists to their personal staff; economists became increasingly critical of government policy as the stability of the immediate post-war decades faded; and there was less uniformity in the advice provided by economists, often leading to confusion and bewilderment by politicians. In his discussion of New Zealand economic thought during the post-1960 decades, Millmow is more successful, perhaps because he has fewer economists to cope with, but also because there is a more dramatic story to tell—namely, the collapse of the heavily controlled economy and the sudden but systematic liberalisation of the New Zealand economy and the reformation of the monetary policy framework.

A key issue for authors of surveys of economic thought in Australia and New Zealand is whether there was anything original in what was produced in these countries or whether economic ideas were essentially derived from elsewhere and adapted where necessary to local conditions and circumstances. Goodwin, for example, concluded that 'Australians imported and adapted economic theory and analytical techniques as they sought to understand and control the functioning of their economy'; they 'selected, modified, and operated fundamentally within the structure taken from abroad' (Goodwin, 1966, p. xii). Groenewegen and McFarlane argued that:

the importance of ideas from abroad explains a great deal of Australian economics over the two centuries of its existence and that, on those few occasions when such borrowing from abroad was less important, what are significant *and* Australian contributions can be highlighted. (Groenewegen & McFarlane, 1990, pp. 4, 5)

The 'significant Australian contributions' they mention include the Brigden Report, the Giblin multiplier, work on fiscal equalisation (Giblin, Melville, R.L. Mathews), the problems of a small open economy (Wilson, T.W. Swan, R.G. Gregory), and economic growth (Swan). Yet they also refer to 'the derivative nature of much Australian economics', contending that 'from the 1950s, Australian economics has become dominated by the economics of the United States' (Groenewegen & McFarlane, 1990, p. 5). What position does Millmow take? He notes the tendency to put down Australasian contributions to economic ideas, agreeing that work undertaken in economics in the Antipodes has been essentially 'derivative', or borrowed 'with a lag' from overseas. And while he concedes that, with the creation of the Economic Society of Australia and New Zealand in 1925, 'the Australasian economics profession has made a respectable contribution to world economic literature', it 'lost not just its distinctiveness, but also its identity' after the Second World War (Millmow, 2017, p. 232).

There are a number of amusing jokes in the book. Some of them are illuminating, but occasionally they backfire. Take, for example, Lionel Robbins' joke about Walter Nash, the leader of New Zealand's delegation at the Bretton Woods Conference and later Prime Minister of New Zealand. According to Robbins, as quoted by Millmow, Nash was 'several bars behind the band' (Millmow, 2017, p. 97). Millmow might have added to the joke by quoting what Robbins—who was a member of the UK delegation at Bretton Woods—said about Melville, the leader of Australia's delegation. Robbins wrote in his diary—now published—that Melville was 'excruciatingly dingy-minded', referring to him as:

the unfortunate Professor Melville, who, rising as if to confess to matricide, announced that the Australian Government had instructed him to disassociate himself entirely from the final proceedings. Poor Melville! (Howson, 1990, p. 191)

Robbins wrote further that ‘the wretched Melville received authority from home to sign the Final Act, though only for certification purposes’. All this might seem hilarious to some. But what is not mentioned by Millmow is that Robbins, within a few months of making these comments about the leaders of the New Zealand and Australian delegations, wrote to a senior British Treasury official saying that he believed that there was something now to be said for what these countries’ delegations were proposing at Bretton Woods. The Employment Approach, he believed, had ‘considerable value’, and if a similar policy could be:

wrung out of the government of the United States and other leading economic centres ... [it] would be to our advantage. Most of our hopes for external reconstruction will be frustrated if intense depression is allowed to develop in any important part of the world. (National Archives of the United Kingdom, 1944, Robbins to Eady, 15 November)

That is exactly what New Zealand and Australia were saying at Bretton Woods, for which they were mercilessly pilloried, though not by Keynes. Robbins said that he could now understand why the governments of those countries ‘are considerably peeved’ by ‘our somewhat lukewarm reply’. In short, he now recognised that much of the criticism of the US proposals at Bretton Woods:

springs, in the end, from a belief, on the part of quite seriousminded people, that we run grave danger in entering into any agreement which involves running in harness, however loose, with a country like the United States which is uncommitted to the maintenance of internal outlay. (National Archives of the United Kingdom, 1944, Robbins to Eady, 15 November)

He had come to:

believe that it would make all the difference in the world ... if it were known that we were taking steps to promote a convention on the lines suggested by the Australians [and New Zealanders]. (National Archives of the United Kingdom, 1944, Robbins to Eady 15 November)

Rather than Australia and New Zealand being ‘several bars behind the band’, it would appear that the American delegation (led by the Soviet agent Harry White) was playing the wrong music! Surely W.M. Corden, a truly great Australian economist, was right when he wrote that the Australian economics profession ‘rates well and does not need to apologise for anything.’ (Millmow, 2017, p. 7).

While there is much in this book that will stimulate debate among economists in Australia and New Zealand, it is a pity that the copyediting leaves a lot to be desired. While some of the errors could be regarded as mere peccadilloes, others are more egregious. Here are some of them. Geoffrey Blainey’s book is not entitled *The Tyranny of Discipline* but rather *The Tyranny of Distance* (p. 12); exchange control was not ‘first used’ in Australia in 1952 but rather in 1939 (p. 15); the claim

that Australia ‘did not sign the Articles of Agreement arising from of [sic.] the Bretton Woods Conference’ (p. 18) is not strictly true, since it did sign the Final Act for ‘certification’ purposes; Syd Butlin is referred to as ‘Chair of Economics’ at the University of Sydney (p. 38); Swan constructed his forecasting model while he was employed at the Department of War Organisation of Industry, refining it later when he moved to the Department of Post-War Reconstruction (p. 104); Giblin was the chair of the Finance and Economics Advisory Committee throughout its existence, not Wilson (p. 105); Keynes never objected to the Australian’s ‘Full Employment Approach’ (sometimes referred to as the ‘Keynesian Crusade’), but rather to the tactics that the Australians (and New Zealanders) were employing—hence his witticism expressed in the quoted letter to T.S. Eliot about the Australians being ‘not only good but clever’ (p. 109); it was Wilson, the Treasury head, and not Coombs, as Governor of the central bank, who preferred economic growth over macro-economic stabilisation (p. 123); Bill Phillips held a chair in the Research School of Social Sciences (RSSH), not in ‘RSEs’ (sic) (p. 129); Solow’s ‘warm tribute’ to Swan did not come ‘much later’, but rather within months of Swan’s 1956 paper being published (p. 134); the National Economic Summit was held in 1983, not in 1984 (p. 209); it was not 1984, but 1983—the year of the decision to float the Australian dollar—that Australia ‘embarked upon an astonishing burst of economic reform’ (p. 209); John Rose was Prime Minister Fraser’s initial ‘principal economic adviser’, not Cliff Walsh (p. 201); Prime Minister Muldoon, the great advocate of direct controls, could hardly be called a ‘Keynesian’ (p. 212).

References

- Goodwin, C.D.W. (1966). *Economic Enquiry in Australia*. Durham, NC: Duke University Press.
- Groenewegen, P. & McFarlane, B. (1990). *A History of Australian Economic Thought*. London and New York: Routledge.
- Howson, S. (1990). *The Wartime Diaries of Lionel Robbins and James Meade, 1943–5*. London: Macmillan. doi.org/10.1007/978-1-349-10840-4
- Keynes, J.M. (1932). The Report of the Australian Experts. *Melbourne Herald*, Melbourne.
- Millmow, A. (2017). *A History of Australasian Economic Thought*. London and New York: Routledge.
- National Archives of the United Kingdom (1944). L. C. Robbins to Sir Wilfred Eady, 15 November, file CAB 123/229.
- Reserve Bank of Australia Archives (1942). J. M. Keynes to L. F. Giblin, 13 November, file c.3.7.6.33.