Is External Research Assessment Associated with Convergence or Divergence of Research Quality Across Universities and Disciplines? Evidence from the PBRF Process in New Zealand

Robert A. Buckle, John Creedy and Norman Gemmell

WORKING PAPER 05/2019
April 2019

Working Papers in Public Finance

VICTORIA UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

The Working Papers in Public Finance series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
http://www.victoria.ac.nz/cpf/working-papers

Is External Research Assessment Associated with Convergence or Divergence of Research Quality Across Universities and Disciplines? Evidence from the PBRF Process in New Zealand

Robert A. Buckle, John Creedy and Norman Gemmell

Victoria University of Wellington

Abstract

Performance-based research quality measures have been adopted in many countries as a basis for allocating funding to universities. The question arises of whether this produces a divergence of research quality across universities and academic disciplines, or convergence whereby initially lower-quality institutions and disciplines catch-up? This paper examines whether the introduction of the New Zealand Performance-Based Research Fund process produced convergence or divergence in research quality scores of universities and disciplines between the 2003 and 2012 assessments. Anonymous individual researcher quality scores in 2003 and 2012 were used to derive average quality scores for disciplines and universities. Substantial convergence in average research quality is found over the period. With few exceptions, the hypothesis that rates of convergence have been uniform across almost all universities and disciplines is supported.

Key words: Education policy, New Zealand universities, Performance-Based Research Fund, research quality, convergence.

JEL classifications: I2; I28.

Acknowledgements: We are grateful to Amber Flynn for assisting with engagement with New Zealand Tertiary Education Commission (TEC), and the TEC for providing the data. We have benefited from discussions on the PBRF process and data with Sharon Beattie, Shelly Biswell, Morgan Healey and Paul Lawrence.

1. Introduction

Performance-based research quality measures have been adopted in many countries as a basis for allocating funding to universities. These schemes vary by coverage and assessment methods, which may be based on bibliometric data or peer review; see OECD (2010), Hicks (2012), Ministry of Education (2013), de Boer *et al.* (2015), Wilsdon *et al.* (2015). The New Zealand Performance-Based Research Fund (PBRF) scheme, introduced in the early 2000s, was designed to unbundle the research component of Government funding of New Zealand tertiary education organisations and allocate the research component based on research performance rather than the number of students. The Tertiary Education Commission explained the aims as follows: 'The purpose of the Performance-Based Research Fund (PBRF) is to ensure that excellent research in the tertiary education sector is encouraged and rewarded. This means assessing the research performance of tertiary education organisations (TEOs) and then funding them on the basis of their performance.'[1]

The introduction of the PBRF changed the incentives facing individuals, departments and universities. Each individual researcher was assigned, in a complex peer review process, to a 'quality category', and these were used to calculate, for each university and discipline, an 'Average Quality Score'.[2] There was considerable heterogeneity both within universities across disciplines, and within disciplines across universities, in these average scores. While the scheme was obviously designed to improve the overall quality ('excellence') of research, the question arises of whether it encouraged initially lower-performing disciplines and universities to improve such that they would catch-up on initially higher-performing ones? That is, would the PBRF process facilitate divergence or convergence of research quality across universities and disciplines, as measured by Average Quality Scores? There are two aspects of convergence. One relates to the relationship between changes in average quality and the initial quality level. The other relates to the dispersion in average quality levels over time. This paper examines both aspects.

It is important to distinguish between the concept of convergence, which refers to the nature of the appropriate joint distribution of the Average Quality Scores in two periods, and that of concentration. The latter refers to the distribution of the number of researchers across disciplines and universities. It is possible to have considerable convergence of research quality,

[1] See https://www.tec.govt.nz/funding/funding-and-performance/funding/fund-finder/performance-based-research-fund/. See also New Zealand Tertiary Education Commission (2002), Mahoney (2004), Ministry of Education (2012), and Smart and Engler (2013).

[2] The process, and these metrics, are described further in Section 2 below.

say across disciplines within a university, while at the same time increasing the proportion of researchers in the initially higher-quality disciplines. In that case, both convergence and increased concentration would act together to increase the average quality score for the university as a whole. However, the two different phenomena may not necessarily move in the same direction: increased concentration may occur alongside divergence among disciplines, with the initially stronger getting relatively stronger in average quality as well as larger in scale. Changes in the number of researchers in different discipline groups and universities, and associated concentration measures, have been examined in detail in Buckle and Creedy (2019c).

Concentration and specialisation has been a feature of some overseas schemes. For example, regarding the UK's PBRF-equivalent process (the Research Assessment Exercise, RAE), Hare (2003) identified concentration as both an RAE objective and outcome. He concluded that, 'the RAE system has resulted in an increasing concentration of public research funding, with barely 20 institutions receiving the lion's share of the funding, the weakest receiving very little' (2003, p.58). Similarly, for US universities Payne and Roberts (2010) found an increase in research specialisation following the introduction of performance measures by many States focussing on the evaluation of teaching. However, little attention seems to have been paid to the question of whether these schemes have led to convergence or divergence of research quality.

In the New Zealand case, it is necessary to recognise a caveat regarding the interpretation of results, due to an absence of suitable data on pre-PBRF research quality; that is, before 2003. Similarly, it is not possible to have a control group of the kind increasingly available to economists where it is possible to conduct 'quasi natural experiments', such as in the tax compliance literature.[3] Hence, it is not possible here formally to test whether it was the introduction of the PBRF that induced an increase in the overall research quality of New Zealand universities,

Nevertheless, there is some indirect evidence that PBRF has stimulated a significant improvement in research quality at New Zealand universities. Buckle and Creedy (2019a) examined the equilibrium quality category distributions that would arise from constant transition proportions, and entry and exit rates, for each university. These were found to involve highly unrealistic equilibrium frequencies in each category, suggesting that the changes since the introduction of PBRF are not sustainable, and supporting the view that the substantial changes observed over the PBRF period were to a large extent stimulated by the new incentive

[3] See, for example, Doidge and Dyke (2013), Gemmell *et al.* (2018) and Alm (2018). In the PBRF case, this was rolled out, and later repeated, uniformly across all universities and disciplines at the same time.

structure. Other analyses of the research performance of New Zealand universities are consistent with this conclusion. In the case of economics, see Anderson and Tressler (2014), and for all NZ universities, see Gemmell *et al*. (2017).[4]

Before examining the New Zealand case, Section 2 first clarifies the metrics used to define research excellence. It then summarises the data on average research quality of New Zealand's eight universities and academic disciplines (assigned to nine groups) since the introduction of the PBRF scheme in 2003. Section 3 discusses the factors likely to influence whether the introduction of the PBRF scheme results in convergence or divergence in the average quality of research among universities and disciplines. Section 4 defines different types of convergence. Section 5 provides an initial assessment of the patterns of change in research quality and evidence of convergence. Section 6 reports panel regression tests for convergence and for differences in rates of convergence among universities and discipline groups. Section 7 extends the analysis to include the effects of scale and age of researchers on the convergence process. Conclusions are provided in Section 8.

2. Average Quality Scores for universities and disciplines

The cornerstone of the New Zealand PBRF exercise is the assignment of each researcher to a quality category (QC). This is determined by a peer-review process. These QCs are used to allocate funding to universities, and are used to compute a quantitative performance score, referred to as an Average Quality Score, *AQS*, for each discipline area and university.[5] This section introduces the definition of the Average Quality Score, *AQS*, and provides some descriptive statistics by university and discipline in 2003 and 2012.

For each PBRF researcher portfolio submitted, a quality category, QC, is determined for each individual, h, by a panel assigned to a subject area or group of subject areas.[6] The relevant subject panel assesses the quality of each portfolio and assigns a score from 0 to 7 for each of three categories, r: these are 'research output'; 'peer esteem'; and 'contribution to research environment'. These three scores, s_r, are given weights, w_r, of 0.70, 0.15 and 0.15. The total

[4] Similarly, a significant change in research performance of UK universities, coinciding with RAE dates, was identified by Wang and Hicks (2013).

[5] For details of the PBRF funding formulae, see Buckle and Creedy (2019a).

[6] The assessment and scoring method used in the New Zealand PBRF system from 2003 to 2012 are described in more detail and critically evaluated in Buckle and Creedy (2019b).

score, S_h, for individual, h, is obtained by multiplying the weighted sum of the s_k values by 100. Hence:

$$S_h = 100 \sum_{r=1}^{3} w_r s_r \tag{1}$$

Thus, the maximum individual score is 700. A letter grade is then assigned depending on the assessed total 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.[7] A numerical score, G_h, is then assigned to each letter grade: 10 for an A; 6 for a B; 2 for a C; and 0 for R. A university's average quality score, *AQS*, is the employment-weighted arithmetic mean score, which can range from zero to 10.

Define the employment weight of person, h, as $e_h \leq 1$, and let n denote the relevant number of employees in a university. The average quality score is:

$$AQS = \frac{\sum_{h=1}^{n} e_h G_h}{\sum_{h=1}^{n} e_h} \tag{2}$$

Since the grade for R staff is equal to zero, their number only affects the denominator in (2).

The dataset used in this study includes anonymous PBRF data provided by the Tertiary Education Commission (TEC) following a confidentiality agreement; it is not publicly available. The data include, for each researcher, an anonymous identifier, age, research discipline, university of employment, and PBRF quality category for each PBRF round in which a researcher's evidence portfolio was submitted.

For present purposes it is useful to combine the various subject areas into nine discipline groups, also used by Buckle and Creedy (2019c). The composition of these nine groups is given in Appendix A. *AQS*s can also be derived for each of these discipline groups, by using the same weights or numerical scores, G_h, applied by the TEC, to the QCs achieved by each individual researcher, h, and by using the information indicating their subject area to assign their score to one of the nine discipline groups. The resulting *AQS* values are shown, by discipline and

[7] The recognition that new researchers may take time to establish their research, publications, and academic reputations led to the introduction in 2006 of the new categories, C(NE) and R(NE). These categories applied to new and emerging researchers who did not have the benefit of a full six-year period. The following analysis does not distinguish the NE categories, since neither numerical scores nor funding was affected.

university, in Table 1.[8] The dataset does not include information on part-time status, so in calculating the *AQS*s here, each value of e_h was set to 1.[9]

Table 1: Average Quality Scores for universities and disciplines

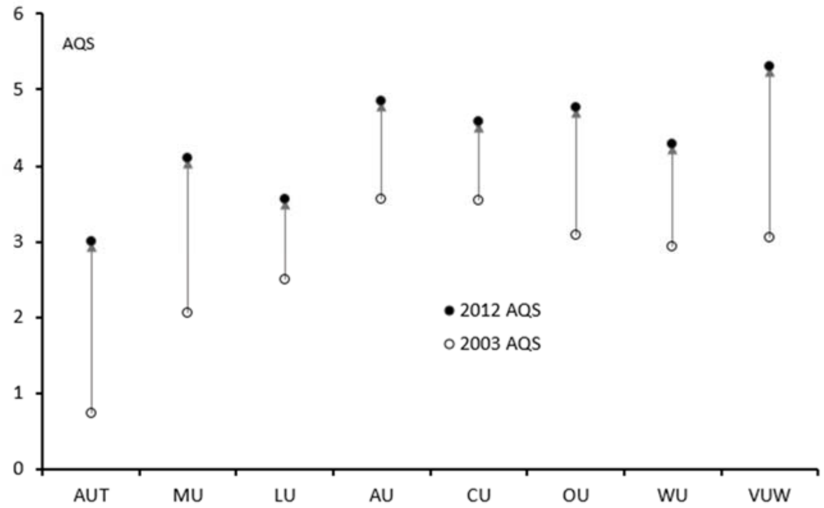
	University *AQSs*								
	AUT	**MU**	**LU**	**AU**	**CU**	**OU**	**WU**	**VUW**	
2003	0.730	2.060	2.490	3.560	3.540	3.080	2.930	3.060	
2012	3.010	4.110	3.560	4.850	4.570	4.770	4.290	5.310	
% change	312.33	99.52	42.97	36.24	29.10	54.87	46.42	73.53	
	Discipline *AQSs*								
	Med	**Eng**	**CS**	**Man**	**AFE**	**Hum**	**Ag**	**Law**	**Edu**
2003	2.79	2.94	3.77	2.17	2.42	3.13	3.61	3.00	1.33
2012	4.43	4.51	5.34	3.93	4.01	4.69	4.95	4.94	3.79
% change	58.58	53.38	41.69	81.16	65.60	49.62	37.16	64.81	185.12
	Discipline *AQSs* within universities								
	Med	**Eng**	**CS**	**Man**	**AFE**	**Hum**	**Ag**	**Law**	**Edu**
AUT: 2003	0.51	1.00	0.67	0.86	0.49	0.80	1.82	0.73	0.55
2012	2.88	3.19	3.5	3.19	3.03	3.10	3.67	2.31	2.13
% change	468.61	219.19	425.00	270.63	521.94	285.90	101.67	217.31	291.11
MU: 2003	2.12	1.48	3.00	1.53	2.08	2.34	2.87	0.66	1.34
2012	3.63	4.35	5.30	3.51	3.50	4.23	4.55	1.67	4.00
% change	71.28	193.95	76.67	129.08	68.44	81.01	58.56	150.00	198.00
LU: 2003	2.14	2.33	5.00	1.80	1.92	2.63	2.89	0.00	2.00
2012	3.77	3.31	3.33	2.72	2.31	3.86	4.13	na	na
% change	76.30	41.96	-33.33	51.11	20.40	46.94	42.58	na	na
AU: 2003	3.42	4.00	4.11	3.03	3.42	4.28	4.75	3.62	1.58
2012	4.59	5.02	5.41	4.42	4.94	5.18	5.31	4.98	4.05
% change	33.98	25.45	31.61	46.15	44.51	21.19	11.99	37.72	156.29
CU: 2003	4.22	4.62	3.91	2.58	2.42	3.45	4.66	3.71	0.90
2012	4.67	5.01	4.91	4.32	4.07	4.51	5.19	5.37	3.06
% change	10.63	8.45	25.40	67.23	67.75	30.61	11.44	44.53	240.02
OU: 2003	2.98	2.61	3.61	2.45	2.76	3.43	4.42	4.07	1.59
2012	4.66	4.33	5.57	4.07	4.97	4.89	5.44	6.29	3.38
% change	56.51	65.57	54.29	66.18	80.00	42.65	22.98	54.48	112.52
WU: 2003	3.39	4.24	4.71	2.98	3.08	2.70	4.50	2.38	1.80
2012	4.00	4.45	4.77	4.36	3.90	4.03	5.02	3.85	4.39
% change	17.81	4.91	1.45	46.24	26.73	49.22	11.65	61.94	144.60
VUW:2003	2.82	3.35	3.71	2.94	2.61	3.71	3.70	3.13	0.76
2012	5.77	4.59	6.43	4.57	4.83	5.60	6.21	5.86	3.89
% change	104.62	36.96	73.22	55.56	85.00	50.88	67.95	88.15	414.29

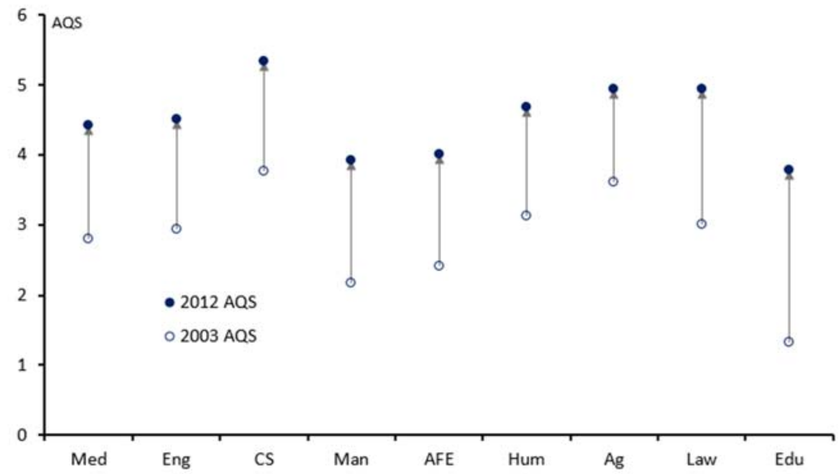
Notes: Med = Medicine, Eng = Engineering; CS = Core Science; Man = Management; AFE = Accounting, Finance and Economics; Hum = Humanities; Ag = Agriculture; Law = Law; Edu = Education. AUT = Auckland University of Technology; MU = Massey University; LU = Lincoln University: AU = Auckland University; CU = Canterbury University; OU = Otago University; WU = Waikato University; VUW = Victoria University of Wellington.

[8] The number of staff who did not submit a portfolio (NP) do not enter into the calculation of *AQS*s derived in this paper because, while NPs can be identified for each university, they cannot be identified by discipline. Buckle and Creedy (2018, 2019a) derive *AQS*s for universities that include NP-staff in the denominator and show that all universities reduced the proportion of NP-staff. The conclusion that all universities substantially increased their *AQS* and, apart from raising the 2012 *AQS* for CU above AU and OU, the rankings in 2012 *AQS*s are unchanged.
[9] The values do not differ substantially from those reported by TEC.

Figure 1 summarises the 2003 and 2012 Average Quality Scores by university and discipline. This identifies Auckland (AU) as (marginally) the leading university in 2003, closely followed by Canterbury. However, by 2012 VUW had become the clear leading university in terms of overall *AQS*. Among universities, AUT, had a particularly low *AQS*, with MU and LU to a lesser extent. Across disciplines, Core Science (CS) was the leading discipline in 2003 and remained the leading discipline in 2012. Education had the lowest *AQS* in 2003, and retained this position in 2012. The figure clearly demonstrates substantial improvements in the *AQS* values for all universities and disciplines, although the rates of improvement vary.[10]

Figure 1 *AQS* values in 2003 and 2012





[10] Some of the gains may have arisen from a learning process relating to the preparation of evidence portfolios.

3. The PBRF and sources of growth in research quality

The 2003 PBRF exercise was innovative for New Zealand universities in various respects. It provided the first comprehensive audit of research outputs by university research staff; it was the first university-wide research evaluation based on quality metrics, and much of the quality information collected was available across the university system for the first time. In addition, the research funding formula used to allocate TEC research funding across universities after the PBRF exercise explicitly linked research quality scores to financial allocations. This directly linked a dollar amount of research funding to each university per 'point' in the numerical score, G_h, described above. Thus, for example, the research funds received by a university for an A-rated researcher (for whom $G_h = 10$) was five times the amount received for a C-rated researcher ($G_h = 2$).[11] Following the 2003 PBRF there was, therefore, a positive proportional relationship between future research funding to each university and its aggregate G_h score: $\sum_{h=1}^{n} G_h$. These post-2003 PBRF changes therefore represented a new knowledge environment in which universities were operating compared to pre-2003, and a new financial incentive structure for research quality improvements.

In considering whether the responses by universities to these new incentives led to convergence or divergence in research quality, it is useful to borrow from analyses of efficiency or productivity of the university (or broader tertiary) sector in Australia and New Zealand adopted by, for example, Abbott and Doucouliagos (2000, 2003, 2010), and Worthington and Lee (2008). Using either Data Envelope Analysis or Malmquist Indices, they sought to decompose university efficiency or productivity gains into those attributable to technical change (a shift of the technology frontier) and technical efficiency improvements (movement towards the frontier).[12] Alternatively these may be thought of as, respectively, gains from innovation by productivity leaders and gains from imitation by followers via technology spillovers, facilitating a process of catch-up (convergence) on the leaders.

As the parallel literature on international technology spillovers makes clear, these spillovers across firms or countries need not necessarily generate convergence; see, for example,

[11] The amounts received also depended on the discipline to which a researcher belongs, with three separate categories given financial weightings of 1, 2 and 2.5. Roa *et al.* (2009) estimate that, following the 2006 PBRF, universities received $34,166 per year per A-researcher (in a discipline with a weighting of 1) and $6,832 per year for each C-researcher in the same discipline. The disciplines in each of the three TEC financial categories were as follows: Māori knowledge and development, law, humanities, business studies (weight = 1); Sciences, IT, nursing, sport and visual arts, theatre, media (2); Engineering, applied sciences, clinical medicine, veterinary science (2.5); see Roa *et al.* (2009), Tertiary Education Commission (2007). These funding ratios and financial weightings remained the same throughout the period.

[12] The latter may be further decomposed into 'pure' technical efficiency gains and 'scale-related' efficiency gains.

Abramovitz (1986), Baumol (1986), Dowrick and Gemmell (1991) and Quah (1993). In particular, while there may be some ‘advantages of backwardness’ where public good qualities of newly created knowledge makes imitation easier than innovation, there may also be ‘disadvantages of backwardness’ to be overcome, such as constraints on complimentary inputs; see Dowrick and Gemmell (1991, p.263).

Unfortunately, *AQS* data available for the current analysis do not allow productivity or efficiency to be measured, since PBRF-assigned quality scores are based on performance as measured by research outputs. Nevertheless, the research quality improvements that the PBRF targeted can be considered as simultaneously encouraging improvements or innovations in research and knowledge within the university system as a whole, and the spread of that knowledge across universities via learning or imitation. Thus the 2003 PBRF provided all universities with improved information on what was regarded as high-quality research, how this was currently distributed across universities and disciplines, and a financial incentive towards raising their quality scores.

This new information also enabled initially-lagging universities and disciplines to identify research leaders and the mechanisms by which they might achieve higher-quality research in future and thereby catch up with those research leaders. Of course, several conditions may inhibit this catch-up process in lower-ranking institutions, such as the mixture of historical factors (for example, research reputation) and geographical conditions (for example, international connections) which might be expected to enhance or inhibit a university’s or discipline’s ability to take advantage of the newly created opportunities and financial incentives. The fact that outside labour market opportunities differ according to the discipline group means that the relative cost of raising research quality is likely to vary; see Boyle (2008). Differences in research processes, availability of contestable research funding, access to higher-ranking journals, can also influence the ability to increase average quality: see, for example, Wanner *et al.* (1981), Shin and Cummings (2010), Jung (2012) and Subharwal (2013).

In practice, research quality, and the *AQS* metric in particular, can be increased by several means. These include the development of existing staff, recruitment of researchers with strong track records or demonstrated potential, and redundancy or retirement of lower-quality academic staff. The ability of each university to implement these improvements can be expected to vary by institutional conditions. Buckle and Creedy (2019a, 2019c) show that universities undertook somewhat different combinations of turnover and improved performance of those who remained in the same university. These differences varied by university and discipline and were found to depend on the initial quality score.

The above framework therefore enables the relative contributions of university-wide improvements in research quality and convergence towards best practice or frontier research quality to be separately identified following the initial PBRF. The empirical analyses in Sections 5 to 7 seek to identify how large was the common growth in *AQS* shared by all universities and disciplines, and how large was any convergence towards or divergence from this common growth.

4. Types of convergence

Before proceeding it is helpful to consider the precise meaning attached to alternative types of convergence in the present context. Consider first the discipline area. Each discipline, aiming to improve its research quality, faces particular advantages and constraints which are likely to be influenced by the university, its location, and the characteristics of the discipline.

Let $q_{i,j}$ denote the *AQS* for discipline *i* in university *j*, determined at the time of a PBRF round. Between two PBRF rounds, these change by absolute amounts $\Delta q_{i,j}$. On the simplified assumption that all quality changes are positive, convergence among any two disciplines, *i* and *k*, within a university, *j,* can be said to exist when:

$$\frac{\Delta q_{i,j}}{q_{i,j}} > \frac{\Delta q_{k,j}}{q_{k,j}} \tag{3}$$

for $q_{i,j} < q_{k,j}$. That is, the proportional growth in measured average discipline quality is larger for those starting from a relatively lower initial average quality level.

Typically, a university has a variety of discipline groups, so for convergence among disciplines within a university, it is required only that there is a significant tendency for the relatively weaker disciplines to improve proportionately more than the stronger disciplines. Thus, consider the following specification (for variations in *i* for given *j)*, where an additional subscript relating to the time period is needed:

$$\log q_{i,j,t} - \log q_{i,j,t-1} = \alpha_j + \beta_j \log q_{i,j,t-1} + u_{i,j,t} \tag{4}$$

Here, *u* is a Normally distributed random error term, and α_j is equal to $\mu_{j,t} - (1+\beta_j)\mu_{j,t-1}$, where $\mu_{j,t}$ is the (unweighted) logarithm of geometric mean quality in university *j* at time, *t*.

This arises from a process in which proportional changes in quality, relative to the geometric mean, are related to the initial relative quality.[13]

The coefficient β_j reflects the nature of convergence. In the cross-country growth and convergence literature it has become known as conditional or β-convergence; see, for example, Barro and Sala-i-Martin (1991) and Quah (1993). This specification of the nature of quality changes forms the basis of the empirical analysis of β-convergence in Sections 6 and 7. A value of β_j equal to zero implies that there is no tendency for convergence: all proportional changes, $\log q_{i,j,t} - \log q_{i,j,t-1}$, are equal to the proportional change in geometric mean quality, except for the stochastic variation. The extent to which β_j is less than zero measures the degree of systematic convergence towards the (unweighted) geometric mean quality of disciplines in the university, with $\beta = -1$ reflecting complete convergence. In this case, $\log q_{i,j,t} = \mu_{j,t} + u_{i,j,t}$ and all disciplines converge, in the absence of the stochastic term, on the geometric mean from *t-1* to *t*. Conversely, a value greater than zero implies divergence, or systematic movement away from the geometric mean.[14]

Clearly, if there is β-divergence in this sense, then the combination of systematic disequalising changes with the random component, *u*, ensures that the dispersion of *q* must increase from period *t*-1 to t. However, the existence of a negative β, or β-convergence, does not itself indicate whether the distribution of *q* is becoming more or less dispersed. Hence, it is possible to define another type of divergence or convergence, according to whether the variance of $\log q_{i,j,t}$ falls or rises: this is referred to in the growth literature as σ-convergence and σ-divergence respectively. The relationship between the two types can be seen as follows.

---

13 Writing $z = q/\mu$, if only random proportional changes in the ratio, z, occur, then $(dz/z)dt = u$ and in discrete time this converts to $\log z_t - \log z_{t-1} = u_t$. To allow for relative size to have a systematic contribution, add a term, $\beta \log z_{t-1}$ to the right-hand side, Hence, when $q_{t-1}/\mu_{t-1} < 1$, so that q is initially below the geometric mean, the logarithm of the relative value, *z*, is negative. Hence if $\beta < 1$, the proportional change is positive and hence greater than if *q* is initially above the geometric mean.

14 In the absence of a random component, $\beta_j < 0$ involves no changes in the rank order of disciplines, but a gradual compression of the distribution, since those at the extremes move relatively faster to the geometric mean (while there are positive absolute changes in all cases). In the case where there are no random variations, it is therefore also possible to think of convergence in terms of movement towards the top of the distribution, such that eventually all disciplines, including the top, take the same log $q_{i,j}$ value.

Denoting the variance of $u_{i,j,t}$ in (4) by σ_u^2, the variance of $\log q_{i,j,t}$ is given by (dropping *i,j* subscripts):

$$\sigma_{q,t}^2 = (1+\beta)^2 \sigma_{q,t-1}^2 + \sigma_u^2 \quad (5)$$

hence,

$$\sigma_{q,t}^2 - \sigma_{q,t-1}^2 = \beta(2+\beta)\sigma_{q,t-1}^2 + \sigma_u^2 \quad (6)$$

The term on the left-hand side captures σ-convergence if it is negative, or σ-divergence if it is positive. Clearly σ-convergence cannot arise with β-divergence ($\beta > 0$). However, given β-convergence, σ-convergence requires $(\sigma_u^2 / \sigma_{q,t-1}^2) < -\beta(2+\beta)$. For example, with $\beta = -0.5$, σ-convergence requires the ratio of variances to be less than 0.75.

Further, from (5) it can be seen that *ceteris paribus*, the dispersion – reflected in the variance of logarithms – stabilises asymptotically to $\sigma_q^2 = -\sigma_u^2 / \beta(2+\beta)$. The value of β, when it is negative, thus determines not only the extent of systematic equalising movements from period *t-1* to *t*, but also the eventual dispersion, as measured by the variance of logarithms. This is highly sensitive to β. For example, if $\beta = -0.7$ the stable variance, σ_q^2, is about 10 per cent higher than σ_u^2, but if $\beta = -0.5$ the difference rises to 33 per cent, and when $\beta = -0.2$ the difference becomes 178 per cent higher (becoming 426 per cent higher when the extent of β-convergence is reduced to –0.1). In the stable situation, inequality-reducing convergence is just balanced by the inequality-increasing random changes. Of course, such stable values assume that the parameters remain constant over time.

Clearly a number of discipline characteristics, in addition to any systematic convergence or divergence tendencies, can be expected to affect research quality improvements including, as suggested above, the particular university in which it is located. Hence, convergence among universities involves a tendency for relative changes to favour those universities below average, and a similar relationship would exist to that in (4) above. That is, for variations in *j* for given *i*:

$$\log q_{i,j,t} - \log q_{i,j,t-1} = \alpha_i + \beta_i \log q_{i,j,t-1} + v_{i,j,t} \quad (7)$$

Again, *v* is a random term, and the coefficient, β_i, reflects the nature of convergence among universities for discipline, *i*.

The discussion has so far been at the discipline level, involving either relative changes within each university, or relative changes for each discipline across all universities. However, particular interest typically attaches to average quality changes at the university level, which influences the ranking of universities by *AQS*. Similarly, it is also of interest to consider the possibility of convergence for disciplines, over all universities combined.

In the context of convergence, or otherwise, among universities in their *AQS* (measured over all disciplines combined), the discipline composition of each university is crucial. It is here that the concept of concentration plays a part in influencing the nature of convergence. That is, a university can improve its overall *AQS*, and rank position among universities, by expanding the size of high-performing discipline areas relative to low-performing areas. Of course, the ability to do this depends on a range of factors, and is likely to vary substantially among universities. It is therefore possible to have convergence across disciplines within each university, along with an entirely different distribution of growth rates of *AQS*s among universities, for all disciplines combined.

To consider aggregate relationships, and to highlight the importance of composition effects, let $n_{i,j}$ denote the number of researchers in discipline i in university j. The total number of researchers in university j, and the total number in discipline i, are given respectively by $n_j = \sum_i n_{i,j}$ and $n_i = \sum_j n_{i,j}$. The *AQS* of university j over all disciplines, is thus:

$$Q_j = \frac{1}{n_j} \sum_i n_{i,j} q_{i,j} \tag{8}$$

The *AQS* of discipline i over all universities, is:

$$d_i = \frac{1}{n_i} \sum_j n_{i,j} q_{i,j} \tag{9}$$

Hence, convergence among universities in *AQS*s is expressed in terms of a regression equation involving Qs. For convergence among disciplines, for all universities combined, the regression equation involves changes in the ds.

It was mentioned above that concentration among disciplines can influence the convergence or divergence of university Average Quality Scores. Consider changes in Q_j and rewrite (8) as:

$$Q_j = \sum_i n'_{i,j} \, q_{i,j} \tag{10}$$

where $n'_{i,j} = \frac{n_{i,j}}{n_j}$. Totally differentiating, and letting, for example, $\dot{Q}_j = dQ_j / Q_j$, it can be shown that:

$$\dot{Q}_j = \sum_i \alpha_{i,j} \left(\dot{q}_{i,j} + \dot{n}'_{i,j} \right) \tag{11}$$

where $\alpha_{i,j} = \frac{n'_{i,j} q_{i,j}}{\sum_i n'_{i,j} q_{i,j}}$. These weights are the (staff weighted) shares of each discipline in the overall *AQS* of the university. Equation (11) shows how the proportional change in a university's *AQS* depends on changes in the shares of each discipline within the university and the changes in individual discipline qualities. Hence it is possible to have, for example, convergence of disciplines within universities and divergence across universities.

5. *AQS* growth and σ-convergence

The previous section identified both *σ*-convergence and *β*-convergence as possible responses following the introduction of the PBRF regime. This section presents some evidence on σ-convergence and considers how observed *AQS* growth between 2003 and 2012 compares with a counterfactual of maximum possible convergence across universities and disciplines. Sections 6 and 7 then examine evidence on *β*-convergence.

Table 2 Variance of log*AQS* by university and discipline

	2003	2012	Obs.
Across disciplines:	0.0997	0.0137	9
Across universities: All universities	0.2722	0.0339	8
Across universities: excluding AUT	0.0381	0.0169	7
Across universities and disciplines	0.3785	0.0687	70

Data on σ-convergence are shown in Table 2 which reports the variance of log*AQS* across universities and disciplines separately and combined, in 2003 and 2012.[15] Across all universities and disciplines, the bottom row shows that variances are larger, as expected given the greater level of disaggregation. It is clear that there are substantial reductions in all variances. These fall by around eighty per cent over the period, suggesting a substantial process of *σ*-

[15] As shown in Table 1, Lincoln university has no observations for Education and Law, so that the total number of universities and disciplines shown in Table 2 is 70 rather than 72.

convergence. However, when AUT is excluded the decline in the variance (0.0381 to 0.0169) is much less than when AUT is included.

Given this evidence, a natural question is: how important is this σ-convergence process in comparison to the overall *AQS* growth observed across the universities and disciplines? Based on the $AQS_{i,j}$ data, Figure 2 addresses this by considering the hypothetical question of what *AQS* growth would have been across universities if each university had fully converged on the 2003 leader, Auckland University. The same question is then applied to disciplines, relative to the 2003 leader, Core Science. In each case, the figure shows actual *AQS* growth by university or discipline alongside the relevant hypothetical full-convergence *AQS* growth.

Figure 2 shows that actual *AQS* growth rates across universities and disciplines are similar to those which would be expected if each university or discipline were able to converge on the 2003 leader. The top panel shows that, while Auckland's growth is equal to the full-convergence case by assumption, the other universities display *AQS* growth rates that, except for VUW, are all somewhat lower than their observed *AQS* growth. VUW's counterfactual full-convergence growth exceeds the actual because VUW overtook AU during the period. Those universities with the largest gaps between actual growth rates and those growth rates required for full catch-up appear to be AUT, LU and MU. Considering differences across disciplines in the lower panel, all disciplines grew more slowly than would be required to catch up fully on Core Science, with the largest gaps appearing to be for Education, Management, and AFE (Accounting, Finance and Economics).

As suggested above, these observed increases in *AQS* scores are associated with changes in concentration; that is, the proportion of staff in the higher research categories. The extent of changes in the proportions of staff in different categories can be seen in Figure 3. This reveals that all universities substantially reduced the share of R-staff. All universities increased the proportions in all three scoring categories (A, B and C), with the exception of VUW. This university had the highest proportional increase in A and B-rated staff and was the only university to reduce its proportion of C-rated staff. This enabled it to become the *AQS* leader in 2012: for further details, see Buckle and Creedy (2019a).

The largest increase in shares occurred with B-rated staff except at AUT and AU. At AUT (initially lowest ranked), the largest increase was in C-rated staff, whereas at AU (initially highest ranked) the largest increase was in A-rated staff. It is also evident that, in general, proportional increases in the share of A-rated researchers was often small, perhaps reflecting the difficulty of recruiting A researchers or transforming existing Bs or Cs into As. These transitions are examined in greater detail in Buckle and Creedy (2019a).

Figure 2 *AQS* growth and convergence

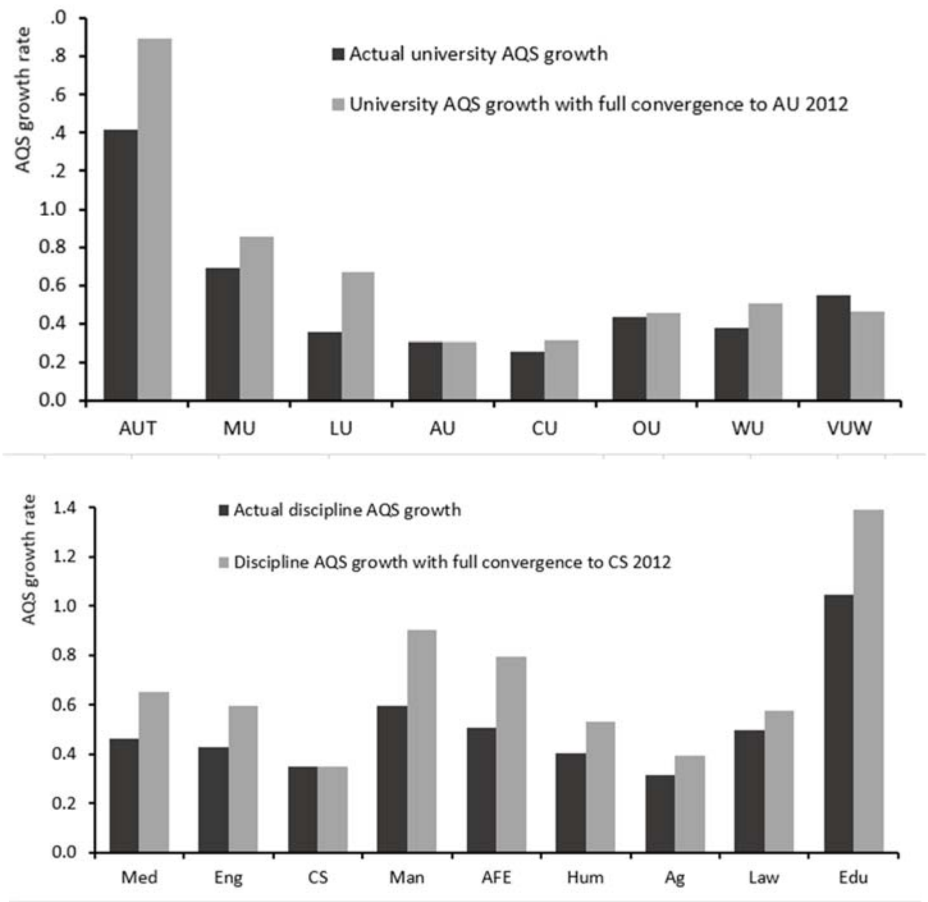
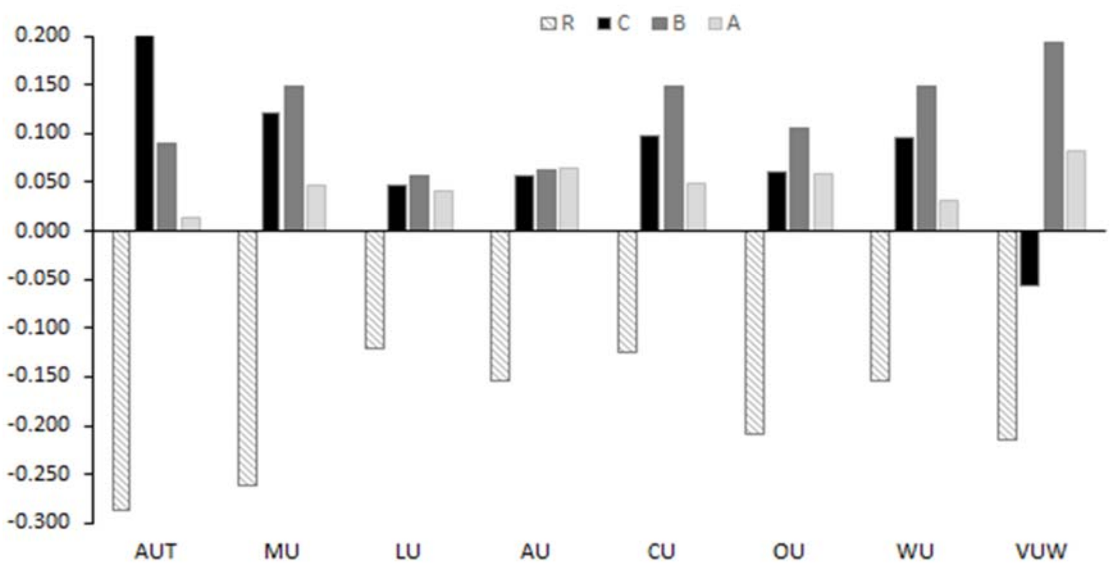


Figure 3 Change in shares of researchers in quality categories: 2003 to 2012



6. Estimating β-convergence among universities and disciplines

In Section 3, it was suggested that a conditional process of β-convergence can be tested by considering the parameters of equations like (4) and (7), enabling this systematic component to be separated from other influences, including university-specific fixed effects and random shocks.[16] Hence, using the $AQS_{i,j}$ data, the present section estimates regression equations of the form:

$$\log AQS_{ijt} - \log AQS_{ijt-1} = \alpha + \beta \log AQS_{ijt-1} + \varepsilon_{ijt} \qquad (12)$$

where $\log AQS_{ijt} - \log AQS_{ijt-1}$ measures (approximate) proportional AQS growth for university j, and discipline i, and t and t-1 refer to 2012 and 2003.

Two approaches are used to test for differences across universities and disciplines in the growth and convergence or divergence of their AQS values, involving the inclusion or exclusion in regressions of shift dummy variables, D_i, D_j, where $D = 1$ for each university or discipline in question and zero otherwise, and similar slope dummies, $(D_i \times \log AQS_{it-1})$, and $(D_j \times \log AQS_{jt-1})$. The parameters α and β in (12) and their university- and discipline-specific equivalents therefore respectively capture autonomous AQS growth and the rate of convergence ($\beta < 0$) or divergence ($\beta > 0$) of AQSs for each university and discipline. The parameter, β, may be interpreted as the rate of convergence conditional on autonomous sources of AQS growth, including any university or discipline fixed effects. The next section considers the effect of adding other conditioning variables to (12).

The first approach initially includes all dummy variables: 2×8 university shift and slope dummies and 2×9 discipline equivalents. These are progressively eliminated using the 'general-to-specific' (*Gets*) approach proposed by Castle *et al*. (2011) and Hendry and Doornik (2014) until the most parsimonious specification of the data-generating process is obtained: Appendix B provides details.[17] This approach effectively treats the null hypothesis as $\alpha_i \neq \alpha_j \neq \alpha$ and $\beta_i \neq \beta_j \neq \beta$, against the alternative of common values of α and β across universities and disciplines.

[16] Such random shocks in this case might include, for example, the 2010-11 Canterbury earthquakes which affected Canterbury University's ability to recruit staff after 2011. In New Zealand's small university system this may have indirectly affected recruitment practices, including recruiting higher-quality staff from CU, in other universities.
[17] For an application of the method in the context of NZ public expenditure growth, see Gemmell *et al.* (2018).

As a check on the robustness of the above approach, a second, 'specific-to-general' approach was adopted. This begins instead with the null hypothesis that all universities and disciplines share the same parameter values, α and β, against the alternative that each university or discipline, tested in turn, deviates from the average across all universities or disciplines. Thus university and discipline shift and slope dummy variables are added sequentially to (12). These are omitted from subsequent regressions where they fail variable-addition *F*- or *t*-tests; see Appendix Table B1. Applying both approaches yielded the same preferred regression specification that includes shift dummies only for Lincoln University, D_{LU}, and Victoria University of Wellington, D_{VUW}, and shift and slope dummies for the Law discipline. Results are summarised in Table 3.[18]

The table reveals strong β-convergence properties across all universities and disciplines at a common rate of convergence. The convergence parameter of –0.722 is relatively large (compared to the maximum convergence parameter of –1), robustly identified, and this rate is confirmed across all universities. That is, in all cases the hypothesis that any individual university β–converges faster or slower is rejected: see relevant *F*- and *t*-tests in Appendix B.

The regression results indicate that *AQS* growth (over 2003-12) would have been around 1.2 (or 120 per cent over the nine years) in the absence of any convergence. But a systematic convergence process resulted in growth being lower, on average, by around 62 per cent (given by –0.722×0.871). Thus, β-convergence effectively reduced *AQS* growth to around half what it would have been in the absence of any convergent tendencies.

The data and regression line are displayed in Figure 4, where the strong β-convergence tendencies and close fit of the regression are clear. This simple model of university-specific and shared autonomous growth and convergence is nevertheless capable of explaining a large fraction of the observed *AQS* growth. That is, for almost all universities and disciplines, 2003 *AQS*s are good predictors of subsequent *AQS* growth to 2012.

The results in Table 3 also indicate that although β-convergence rates are generally constant across all universities, there are three university- or discipline-specific differences. Firstly, LU

[18] Regressions in Table 3 use 87 observations: 70 for the 8 universities and 9 disciplines within universities, less two missing observations for Law and Education at Lincoln, plus 17 AQS_{ij} values, averaged across all universities and all disciplines. This enables parameters for each university's growth and convergence to be compared directly with the average across all universities or disciplines rather than adopting one university and discipline as the omitted variable. However, adjusted-R^2s must be interpreted cautiously since they are somewhat inflated by the inclusion of individual observations and their cross-university or cross-discipline averages. For example, running the regression specification in Table 3 on the 70 AQS_{ij} observations yields a slightly lower adjusted-R^2 = 0.940.

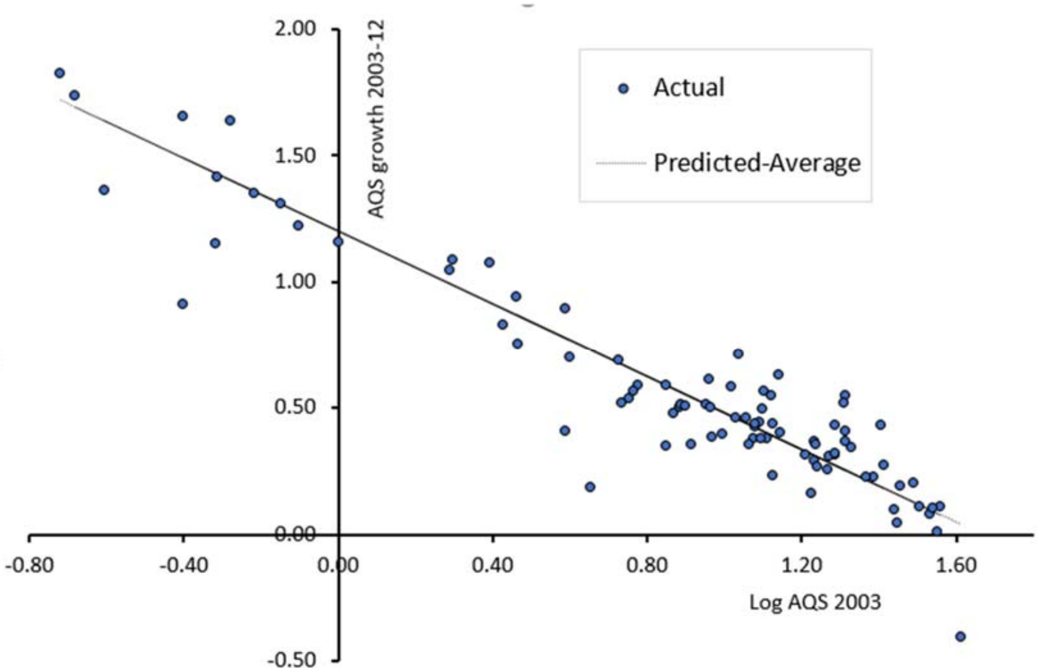
and VUW have respectively slower and faster growth, *ceteris paribus*, in their *AQS*s over 2003 to 2012. Secondly, the growth of *AQS*s in Law appear to have involved both lower growth across all universities (parameter = –0.317) and a lower rate of convergence across universities (parameter = –0.722 + 0.340 = –0.382).

Table 3 Final regression results

	Coefficient	Standard Error	*t*-value	95% Confidence Interval	
logAQS_{2003}	-0.722	0.020	-36.14	-0.761	-0.682
D_{LU}	-0.258	0.038	-6.88	-0.333	-0.183
D_{VUW}	0.166	0.034	4.89	0.098	0.233
D_{Law}	-0.317	0.058	-5.46	-0.432	-0.201
D_{Law}×logAQS_{2003}	0.340	0.055	6.18	0.231	0.450
Constant	1.201	0.021	56.56	1.159	1.243
R^2 = 0.947	Adj-R^2 = 0.943		F (7, 79) = 268.78		Obs. = 87

Notes: Dependent variable: *AQS* growth, 2003-2012. D_{LU}, D_{VUW} and D_{Law} are shift dummy variables for Lincoln University, Victoria University of Wellington and the Law discipline respectively. D_{Law}×logAQS_{2003} is a slope dummy variable for the Law discipline.

Figure 4 Cross-plot of 2003 log*AQS* and *AQS* growth with regression line



The results reported here for *β*-convergence may be compared with the σ-convergence results concerning the dispersion of average research quality, discussed in Section 5. As Figure 1 and Table 1 showed, AUT is a large initial outlier with an especially low *AQS* in 2003. Nevertheless, in terms of the systematic process of *β*-convergence observed across all $AQS_{i,j}$, AUT is no different from the other seven universities. Likewise, Lincoln and Massey Universities (the other two low-*AQS* universities in 2003) appear to share the same average rate of *β*-convergence with all other universities, even though Lincoln also appears to have experienced relatively low *AQS* growth. Similarly, across disciplines, Education appeared to be a low outlier in Figure 1 and has a large effect on inter-discipline *σ*-convergence, yet it is Law, not Education, that appears to have experienced a somewhat different rate of *β*-convergence compared with the average of all others.

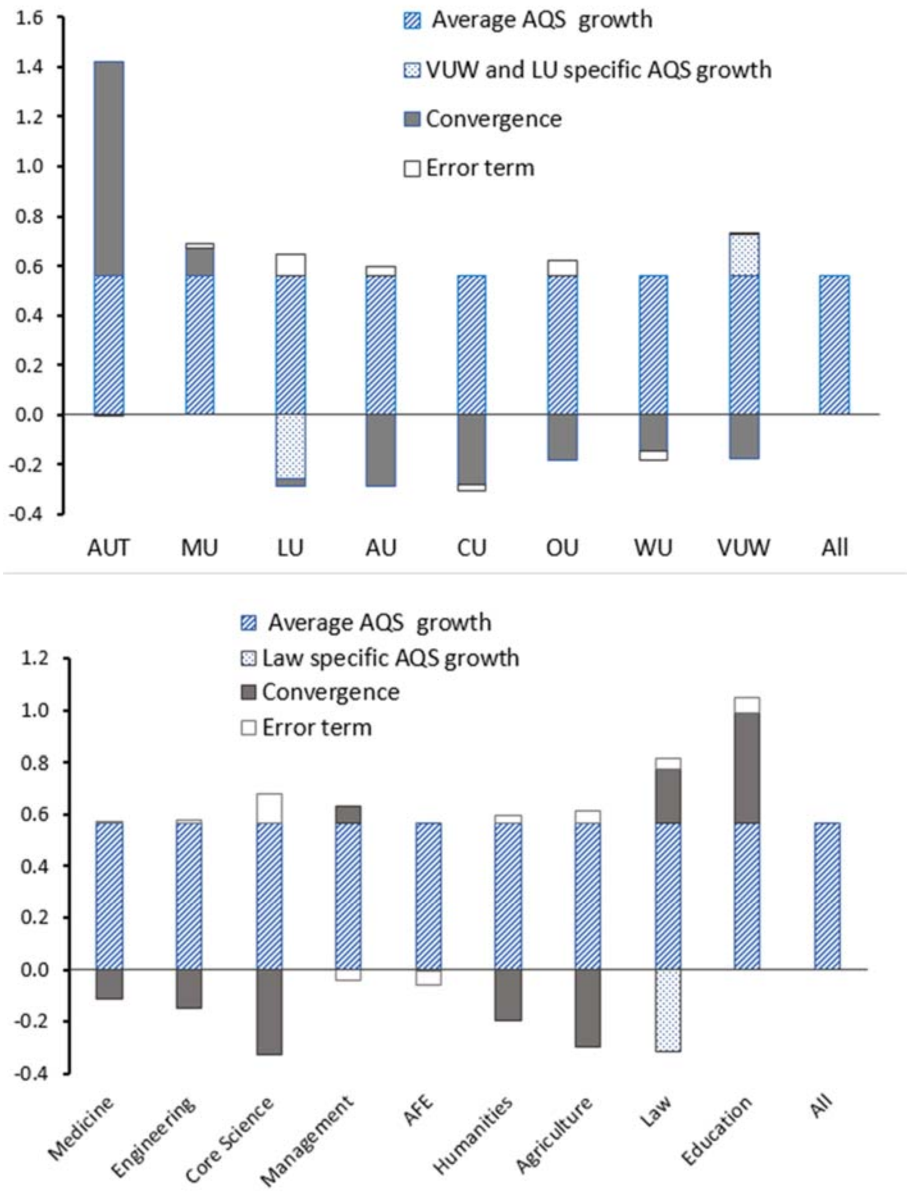
Although Table 3 suggests a common rate of *β*-convergence across all universities and disciplines, except Law, the relative contribution of *β*-convergence to each university's *AQS* growth outcome differs, depending on their initial *AQS* value. The decompositions are shown in Figure 5: the top panel shows differences across universities and the lower panel shows differences across disciplines. Using the parameter estimates from Table 3, together with mean values for each variable, allows the decomposition of *AQS* growth to be identified, in terms of average *ASQ* growth and the sources of deviations from that average by individual universities or disciplines.

The average $AQS_{i,j}$ growth rate is 0.56, shown in both panels of the figure. Each university (upper panel) or discipline (lower panel) deviates from that average due to a combination of the convergence effect and university-specific factors. These average and convergence components can be interpreted, respectively, as the effect of an increase in research quality by all units over the 2003 to 2012 PBRF period as the research frontier shifted, maintaining all relative positions constant, and relative movements towards that frontier by those units initially inside.

The top panel shows that *β*-convergence generated greater than average growth in two of the three initially lowest research quality universities: Massey (MU) and AUT. The third, Lincoln (LU), experienced a slightly negative convergence component despite starting with an initially low *AQS*. In addition, Lincoln experienced a further negative fixed effect that resulted in substantially lower than average *AQS* growth.

Thus, the five remaining universities (AU, CU, WU, OU and VUW) experienced a negative relative growth contribution from the convergence process: that is, they were converged upon by MU and AUT, though to differing degrees as shown in Figure 5. VUW's higher autonomous growth effectively compensated for the negative impact of the inter-university β-convergence component, such that VUW became the new leading university by 2012.

Figure 5 Contributions to *AQS* growth across universities and disciplines



The lower panel of Figure 5 provides a similar breakdown across disciplines, indicating different contributions from inter-university β-convergence for different disciplines. Thus, convergence contributed positively to *AQS* growth in only three disciplines: Law, Education and (slightly) in Management. For the remaining six disciplines, β-convergence was a source of lower-than-average *AQS* growth, although it was close to zero in Accounting, Finance and

Economics (AFE). Unsurprisingly, it was most negative in Core Sciences (CS) the discipline leader in *AQS* values in 2003, with relatively large negative values also for Agriculture and Humanities which, like CS, had high initial *AQS*s, as shown in Figure 3.

It was acknowledged in the introduction that, with the available data, the β-convergence measured over the period cannot be directly attributed to the PBRF process. Indeed, it is likely that some convergence existed before 2003. Yet, the extent reported here (of around -0.7) is large, bearing in mind, as mentioned in Section 4, that as β approaches -1 the equilibrium variance of logarithms of the *AQS* approaches the residual variance (on the assumption that the parameters governing the dynamic process remain constant). This would imply a considerably lower value than the observed variance in 2012, shown in Table 2, which demonstrated nevertheless substantial σ-convergence over the period. Such convergence over a longer period, producing a relative uniformity of research quality among NZ universities, seems unlikely. It is plausible therefore that, as with the average growth in measured research quality, a non-trivial proportion of convergence can be attributed indirectly to the incentives created by the PBRF.

7. Conditioning convergence on other variables

The results in the previous section estimated β-convergence effects conditional only on the shift and slope effects associated with individual universities and disciplines. A paucity of data on suitable variables available by university and discipline limits further testing of the sensitivity of these results to other conditioning factors. However, two relevant variables are available: the total number of research portfolios submitted by each university and discipline to the PBRF exercises, $n_{i,j}$, and the ages of the individuals concerned, $age_{i,j}$. Subsection 7.1 considers those effects on estimated rates of convergence. Subsection 7.2 examines the potential impact on estimated convergence properties of the fixed maximum achievable *AQS* of 10.

7.1 Effects of median age and number of researchers in each discipline

If average research quality benefits from greater concentrations of researchers in the same unit (university or discipline), it might be expected that the number of researchers, $n_{i,j}$, in initially higher-quality units would positively affect the Average Quality Scores for that unit, due to a higher spillover effect within higher-quality research units. This may be tested by adding an interaction term, $\log N_{2003} \times \log AQS_{2003}$, to the previous regression specification. If larger numbers of portfolios in initially higher quality units encourages higher *AQS* growth, the parameter on this term is expected to be positive, while the parameter on $\log AQS_{2003}$ remains negative. Alternatively, if there are ‘pure scale’ effects, whereby greater numbers of researchers *per se* raise research quality in any unit regardless of initial *AQS* (perhaps due to general research culture factors), this can be tested by instead adding $\log N_{2003}$ to the original specification.

Secondly, research quality may be improved by appointing new high-quality staff or transforming existing staff. In the case of the former, this has been shown to be associated with older staff who arrive with a proven record of experience in research performance.[19] On the other hand, improving research quality over the longer-term may be more associated with replacing non-performing (often older) staff with new young staff more capable of building a high quality research portfolio. Thirdly, transforming existing staff involves no age effect (other than the effects of time that affect all staff). Table 4 shows results from testing for the above effects, and may be compared with Table 3.

First, in column (1), ‘age’ and ‘pure scale’ variables are added: the logarithm of the median age of portfolio submitters in 2003, $\log Med\text{-}age_{2003}$, and the number of portfolios submitted in 2003, $\log N_{2003}$.[20] Initial year values are used because interest is in how far initial conditions (other than initial *AQS* levels) affected subsequent research quality growth, and to minimise the risk of endogeneity effects on parameter estimates.[21]

Table 4, column (1) shows that neither the scale nor age variables have statistically significant effects on *AQS* growth and have minimal effect on the estimated parameter sizes for other included variables.[22] Adjusted-R^2s are similar for both the regressions in Tables 3 and 4 but the regression *F*-statistic is lower in Table 4.

Nevertheless, the estimated positive parameter on the scale variable is consistent with initially larger research units being positively correlated, *ceteris paribus*, with higher subsequent *AQS* growth (significant at around the 15 per cent level). Conversely, initial age appears to be negatively associated with subsequent *AQS* growth (significant at around the 13 per cent level), perhaps indicating that replacing initially older lower-quality staff with higher-quality staff (young or old) after 2003 was the dominant age-composition effect.

---

[19] See Buckle and Creedy (2019a).

[20] These data are reported in Appendix A, Tables A1 and A2.

[21] Results were also obtained for non-log versions of these included variables and when mean, rather than median, age was used. Results were uniformly inferior to those reported in Table 4, with no variables even close to statistical significance at the 10 per cent level.

[22] Testing whether the shift and slope dummy variables included in Table 3 are rendered redundant when the new variables are added to regressions confirms that this is not the case. A nested model always supports the inclusion of all dummy variables over the age and/or scale variables.

Table 4 Regression results: adding scale and age effects

	Coefficients (standard errors)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log AQS_{2003}$	-0.737 (0.022)**	-0.802 (0.044)**	-0.820 (0.083)**	-0.796 (0.044)**	-0.764 (0.089)**	-0.764 (0.097)**	-0.758 (0.108)**	-0.728 (0.035)**
D_{LU}	-0.227 (0.041)**	-0.225 (0.040)**	-0.232 (0.040)**	-0.230 (0.040)**	-0.196 (0.072)**	-0.254 (0.079)**	-0.240 (0.058)**	-0.247 (0.051)**
D_{VUW}	0.172 (0.034)**	0.173 (0.034)**	0.173 (0.034)**	0.174 (0.034)**	0.163 (0.039)**	0.146 (0.043)**	0.213 (0.070)**	0.209 (0.068)**
D_{Law}	-0.281 (0.060)**	-0.307 (0.057)**	-0.329 (0.067)**	-0.317 (0.057)**	-0.431 (0.455)	-	-0.325 (0.071)**	-0.325 (0.070)**
$D_{Law} \times \log AQS_{2003}$	0.336 (0.055)**	0.358 (0.055)**	0.367 (0.062)**	0.357 (0.055)**	0.463 (0.363)	-	0.298 (0.123)*	0.289 (0.117)*
$\log Med\text{-}age_{2003}$	-0.206 (0.134)	-0.182 (0.134)	-	-	-	-	-	-
$\log N_{2003}$	0.014 (0.009)	-	-0.007 (0.020)	-	-	-	-	-
$\log N_{2003} \times \log AQS_{2003}$	-	0.015 (0.009)†	0.022 (0.018)	0.016 (0.009)†	0.023 (0.011)*	0.010 (0.011)*	0.006 (0.022)	-
Constant	1.919 (0.511)**	1.189 (0.510)**	1.226 (0.091)**	1.200 (0.021)**	1.112 (0.137)**	1.216 (0.147)**	1.196 (0.025)**	1.197 (0.025)**
Adj-R^2	0.945	0.946	0.944	0.945	0.773	0.713	0.929	0.931
Regression F	211.8	214.1	209.3	246.9	25.4	27.8	93.1	114.5
Obs.	87	87	87	87	44 (upper)	44 (upper)	43 (lower)	43 (lower)
Implied convergence parameter:				-0.722	-0.653	-0.718	-0.730	-0.728

Notes: Dependent variable: *AQS* growth, 2003-2012. D_{LU}, D_{VUW} and D_{Law} are shift dummy variables for Lincoln University, Victoria University of Wellington and the Law discipline respectively. $D_{Law} \times \log AQS_{2003}$ is a slope dummy variable for the Law discipline. Standard errors in parentheses; **, (*), (†) = significant at 1%, (5%), (10%). 'Convergence parameter' = $d(AQS \text{ growth})/d(\log AQS_{2003})$, calculated at the mean value of $\log N_{2003}$.

When the possibility of effects on *AQS* growth from larger numbers in higher quality units is allowed for in columns (2) and (3), this clearly outperforms both the age and general scale variables. The parameter on the interaction term, $\log N_{2003} \times \log AQS_{2003}$, in (4) is significantly positive at the 6.6% confidence level at 0.016 (s.e. = 0.009; t = 1.86), while the parameter on $\log AQS_{2003}$, is relatively unaffected: –0.796 (s.e. = 0.044; t = –17.9). Of course, the overall impact of $\log AQS_{2003}$ on *AQS* growth now has to be estimated from the two parameters; these are shown in the final row of Table 4, based on mean $\log N_{2003}$.

At –0.722, this overall effect is identical to that reported in Table 3, though the regression *F*-statistic marginally favours the Table 3 result. Nevertheless, there is some evidence here that, although faster research quality growth is associated with lower initial values, this is counteracted somewhat if initially higher quality units have larger numbers of researchers. For example, decomposing the convergence parameter in Table 4, column (4) shows that the direct negative convergence effect is –0.796, while the indirect effect is +0.074; that is the scale component has only a modest compensating effect, at least at the average $\log N_{2003}$ value.[23]

7.2 Does the maximum achievable *AQS* value affect convergence?

Since the maximum quality score that an individual researcher can achieve, by being awarded an A rating, is 10 points, this also fixes the maximum *AQS* for a research unit. If there are increasing costs associated with raising *AQS*s as the maximum is approached, as seems likely, this creates a specific 'mechanical' source of convergence. This makes the *AQS* convergence case rather different from the convergence analysed in the tertiary productivity and cross-country GDP growth literatures noted earlier. In these cases there were no fixed maximum values achievable (university productivity, or GDP levels). This raises the question of whether the observed *AQS* β-convergence is merely the result of the increasing costs (diminishing marginal returns) associated with initial *AQS*s closer to the maximum?

One approach to address this question is to consider the long-run value of *AQS* to which universities and disciplines are converging based on the earlier estimated convergence parameter, β, and in the absence of the 'random' effects, $u_{i,j,t}$. Calculations of the time profile of $\log AQS$ based on equation (12) and the parameters in Table 3 ($\alpha = 1.201$; $\beta = -0.722$), reveal that, starting from 2003 values, and based purely on the systematic convergence component, all $\log AQS$ values would be expected to β-converge towards 1.665, or $AQS \rightarrow 5.284$.[24] This compares with the maximum achievable $AQS = 10$ or $\log AQS = 2.303$. Hence, at roughly half the value of the maximum achievable *AQS*, this result suggests that the β-convergence properties of the *AQS* data are not influenced by this hypothetical long-run maximum.[25]

---

[23] At the maximum value of $\log N_{2003}$ this component becomes 0.12; that is it remains quite small.

[24] Similar results are obtained using parameters from Table 4 ($\log AQS \rightarrow 1.657$.), where convergence properties are captured using parameters on the two variables which include $\log AQS_{2003}$.

[25] The mean of observed values of $\log AQS$ increased from 0.871 to 1.435, 2003 to 2012. Maximum values observed in the dataset increased from 1.609 in 2003 to 1.861 in 2012 highlighting that, for some units, observed *AQS* growth was substantially impacted by other factors, consistent with evidence from regression results and Figure 5.

An alternative approach is to split the sample into above- and below-median *AQS* sub-samples.[26] This allows testing of the hypothesis that the rate of β-convergence is equal across the two samples. If the maximum achievable *AQS* was indeed constraining the ability of initially higher *AQS* units to grow, it would be expected that the estimated value of β for the below-median sub-sample would be greater in absolute value than the equivalent for the above-median *AQS* sub-sample.

Results for the two sub-samples are reported in the four right-hand columns of Table 4 with the regressions highlighted in bold representing the preferred specifications based on parameter and regression F-tests. As noted above, with the inclusion of $\log N_{2003} \times \log AQS_{2003}$ in all regressions the rate of convergence is no longer estimated simply from the parameter on $\log AQS_{2003}$. Nevertheless, it is clear from both sets of parameters that there is no support for the hypothesis that the rate of convergence differs across the two sub-samples. Indeed the convergence parameter in the final row of the table confirms that the estimates for the two separate sub-samples and for the full sample are close to, and statistically indistinguishable from, each other.

Thus, although in principle a fixed maximum long-run value of *AQS* in 2003 and 2012 could potentially generate some convergence tendencies in the data generating process, these results suggest that this has not been an important determinant of the convergence estimates obtained earlier.

8. Conclusions

This paper has examined changes in research Average Quality Scores across universities and academic disciplines in New Zealand from the first research evaluation (PBRF) exercise in 2003 to 2012, the latest year for which data are currently available. Particular interest focused on whether, and how far, growth in Average Quality Scores between 2003 and 2012 can be attributed to a catch-up process whereby initially lower-quality disciplines and universities converge on those with higher-quality. Two aspects of convergence were examined: these are the relationship between growth rates and initial quality (β-convergence), and the changing dispersion of quality levels (σ-convergence).

It was suggested that changes in universities' Average Quality Scores after the introduction of the PBRF could be expected to follow a growth and convergence process. That is, the PBRF encouraged a shift in the research 'technology frontier' such that all universities could

---

[26] In principle, several sub-samples could be created but limited numbers of observations here would limit the reliability of further sample divisions.

potentially increase their *AQS*s via quality improvements, as well as 'movement towards the frontier' whereby institutions inside the frontier could catch-up on higher-quality institutions, depending on the convergence constraints they face.

Strong convergence properties of *AQS*s both across universities and academic disciplines were found. Importantly, rates of β-convergence appeared to be uniform across universities and disciplines even though the extent of σ-convergence depended on the inclusion or exclusion of initial *AQS* outliers. The null hypothesis of all universities sharing a common rate of β-convergence was strongly supported.

The robustness of results to the introduction of several additional control variables was tested. The variables were the initial scale (number of portfolios), the interaction between scale and average quality, and the median age of researchers. The interaction term tested whether research quality improvements were greater where higher-quality research staff were concentrated in larger research units, while an age variable allowed testing of whether initially older or younger research units experienced faster *AQS* growth. The age and simple scale variable were not statistically significant. However, there was some evidence that greater numbers of higher-quality research staff in a unit was associated with faster *AQS* growth. That is, *ceteris paribus*, β-convergence properties whereby initially lower quality units grew faster, were counteracted somewhat where larger numbers of higher-quality researchers were concentrated. Importantly, the robustness of the previous parameter estimates was unaffected.

A concern with the above convergence tests is that, since the definition of the TEC quality metric imposes a fixed maximum value, the observed growth and convergence process may simply have resulted from the nature of the metric used. However testing the regression estimates against the alternative hypothesis that observed *AQS* growth and convergence were dictated by a maximum *AQS* value was clearly rejected.

Therefore, regarding the question posed in the title of this paper, of whether the PBRF process in New Zealand was associated with convergence or divergence of average research quality across universities and disciplines, the evidence from 2003 to 2012 suggests strong average growth and convergence of higher-quality research across New Zealand universities and disciplines. As a result, the process of convergence observed in the data suggests a reduction in the dispersion, and substantial catch-up in average research quality by initially lower-ranked universities and disciplines, even though in the two specific cases of Lincoln University and the Law discipline, other factors tended to counteract that catch-up process.

Appendix A Composition of Discipline Groups

The PBRF data for each anonymous researcher, provided by TEC, include a research subject area, of which there are 42 (see New Zealand Tertiary Education Commission, 2013). This information was used to allocate each researcher to one of nine discipline categories. The discipline categories comprise the following research subject areas.

Medicine: Biomedical Science; Clinical Medicine; Dentistry; Molecular, Cellular and Whole Organism Biology; Nursing; Other Health Studies (including Rehabilitation Therapies); Pharmacy; Psychology; Public Health; Sport and Exercise Science.

Engineering: Architecture, Design, Planning, and Surveying; Computer Science, Information Technology, Information Sciences; Design; Engineering and Technology.

Core Science: Chemistry; Physics; Pure and Applied Mathematics; Statistics.

Management: Management; Human Resources; Industrial Relations and Other Business; Marketing and Tourism.

AFE: Accounting and Finance; Economics.

Humanities: Anthropology and Archaeology; Communications, Journalism and Media Studies; English Language and Literature; Foreign Languages and Linguistics; History; History of Art; Classics and Curatorial Studies; Human Geography; Māori Knowledge and Development; Music; Literary Arts and Other Arts; Philosophy; Political Science; International Relations and Public Policy; Religious Studies and Theology; Sociology; Social Policy; Social Work; Criminology and Gender Studies; Theatre and Dance; Film, Television and Multimedia; Visual Arts and Crafts.

Agriculture: Agriculture and Other Applied Biological Sciences; Earth Sciences; Ecology; Evolution and Behaviour; Veterinary Studies and Large Animal Science.

Education: Education.

Law: Law.

The number of portfolios in each of these nine discipline areas and in each university are shown in Table A1.

Table A1 Discipline groups and PBRF portfolios, 2003 and 2012

Discipline category:	Total NZ	University: AUT	MU	LU	AU	CU	OU	WU	VUW
Medicine:									
2003	1804	178	203	14	483	55	762	43	66
2012	1808	112	199	18	585	78	720	36	60
Change	4	-66	-4	4	102	23	-42	-7	-6
Engineering:									
2003	905	94	207	36	240	110	88	50	80
2012	864	99	139	32	277	119	55	58	85
Change	-41	5	-68	-4	37	9	-33	8	5
Core Science:									
2003	534	15	92	4	166	94	82	34	47
2012	496	12	60	3	172	88	79	31	51
Change	-38	-3	-32	-1	6	-6	-3	-3	4
Management:									
2003	473	72	115	20	76	24	62	55	49
2012	398	47	85	25	52	25	54	61	49
Change	-75	-25	-30	5	-24	1	-8	6	0
AFE:									
2003	358	37	77	24	62	33	42	37	46
2012	337	39	68	26	51	30	35	42	46
Change	-21	2	-9	2	-11	-3	-7	5	0
Humanities:									
2003	1458	142	250	16	320	187	190	145	208
2012	1354	102	219	14	315	168	219	117	200
Change	-104	-40	-31	-2	-5	-19	29	-28	-8
Agriculture:									
2003	548	11	186	85	59	58	71	32	46
2012	652	12	192	95	108	64	82	41	58
Change	104	1	6	10	49	6	11	9	12
Education:									
2003	738	55	149	1	191	71	44	137	90
2012	553	30	79	0	153	77	42	102	70
Change	-185	-25	-70	-1	-38	6	-2	-35	-20
Law:									
2003	223	11	18	3	63	28	29	32	39
2012	190	13	6	0	63	19	28	26	35
Change	-33	2	-3	-3	0	-9	-1	-6	-4
Total:									
2003	7041	615	1297	203	1660	660	1370	565	671
2012	6652	466	1047	213	1776	668	1314	514	654
Change	-389	-149	-250	10	116	8	-56	-51	-17

Source: Buckle and Creedy (2019c, Appendix A, pp 29-30).

Table A2 Median age in each university and discipline group

	AUT	MU	LU	AU	CU	OU	WU	VUW	All unis
Medicine	43	42	40	39	39	41	41	42	41
Engineering	42	42	44	40	38	40	37	43	40
Core Science	52	40	52	37	38	37	50	45	39
Management	46	45	45	42	45	39	47	41	44
AFE	37	45	42	44	41	42	46	39	42
Humanities	46	43	44	45	44	43	46	45	44
Agriculture	38	39	44	40	37	38	43	36	38
Education	52	51	43	50	46	46	50	48	50
Law	46	52	62	42	48	48	45	41	45
All disciplines	45	43	44	41	41	41	46	44	42

Appendix B Derivation of Convergence Results

This appendix provides more detail on the general-to-specific and specific-to-general (*Gets*) approaches used to obtain the regression results in Table 3. The process to identify the data generating process is identified in Appendix Table B1. Column 1 shows the university or discipline unit, while col. 2 indicates whether *t*-ratios shown in relevant columns relate to a shift (D) or slope (S) dummy variable.

Initially following the *Gets* approach (see, for example, Campos, Ericsson and Hendry, 2005a, b), regressions were first run with all possible variables, including all dummies shown.[27] Dummy variable were eliminated one at a time based on a parameter *t*-test, in each case eliminating the variable/parameter with the lowest *t*-ratio. (The convergence variable, log*AQS*2003, always passed relevant *t*-tests).

For example, in Table B1 the first variable eliminated was the shift dummy for Management (Man) with a *t*-value of -0.01. The regression was re-run omitting this variable. This led to the slope dummy for medicine (Med) being identified as the lowest *t*-ratio and eliminated ($t = 0.30$). This process was repeated until only variables with *t*-ratios > $|3|$ were retained, yielding the regression specification in column 5 (and shown in Table 3).

A critical *t*-ratio = 3 was chosen following Castle *et al.* (2011) and Castle and Hendry (2014) who argue that substantial pre-testing and variable selection from many possible models increases the risk of retaining irrelevant variables. For example, with a *t*-test significance level

[27] Campos *et al*. (2005b, p. 2), for example, summarise the *Gets* approach as follows: '1. *Ascertain that the general statistical model is congruent. 2. Eliminate a variable (or variables) that satisfies the selection (i.e., simplification) criteria. 3. Check that the simplified model remains congruent. 4. Continue steps 2 and 3 until none of the remaining variables can be eliminated*'.

of $c_\alpha = 0.05$, there is a 1 in 20 chance of an irrelevant variable being retained (with a threshold $t > 2$) in the model on average. However, for $c_\alpha = 0.01$ ($t_\alpha \approx 2.6$) this becomes 1 in 100, and $c_\alpha = 0.001$ ($t_\alpha \approx 3.35$) implies 1 in 1000. Hence t_α between 2.6 and 3.35 substantially reduce the risk of a false positive. Castle and Hendry (2014) recommend setting α = min.(1/N, 1/T, 1%

Table B1 Identifying the data generating process

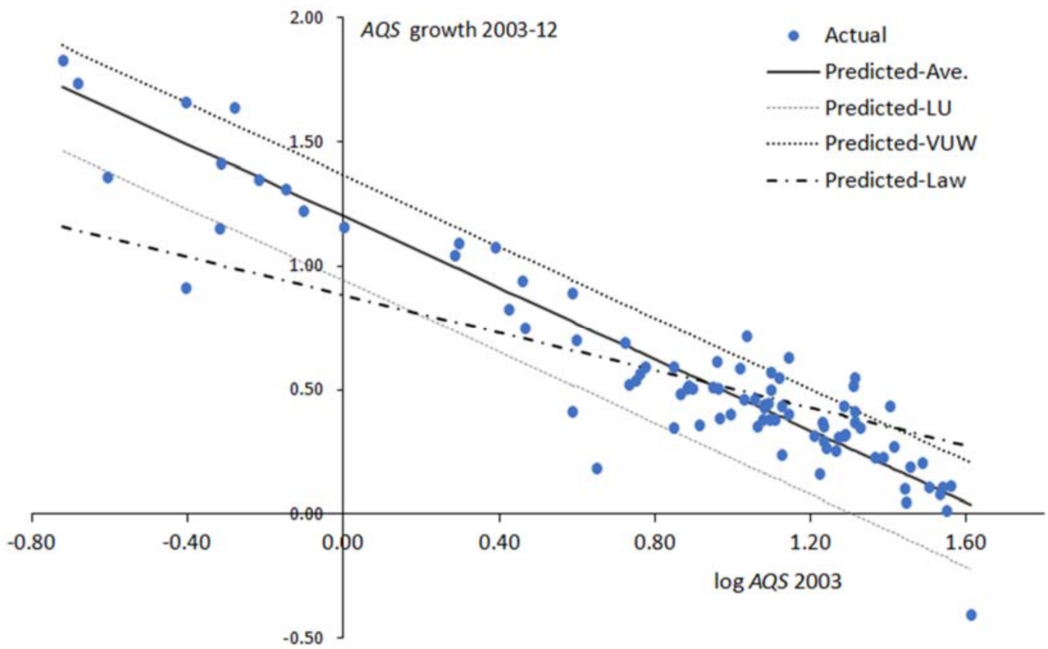
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		GENERAL TO SPECIFIC			SPECIFIC TO GENERAL									
		Order of	Lowest	Retained	t-test	F-test	t-test	t-test	t-test	F-test	t-test	t-test	t-test	t-test
Unit		elimination	t-ratio	t-ratio	(Unis)	(Unis)	(Unis)	(Unis)	(Disc)	(Disc)	(Disc)	(Disc)	(Unis & Disc)	
AUT	D	22	0.94		0.64	0.91								
	S	17	-1.36		-0.96									
MU	D	24	-0.55		**-2.40**	3.46		-3.32					-0.97	
	S	23	1.38		**2.58**			3.49					1.13	
LU	D			**-6.88**	-1.35	16.63	-5.69	-6.09					-6.77	**-6.88**
	S	11	-0.73		-0.48									
AU	D	8	0.74		0.47	0.22								
	S	12	1.03		-0.34									
CU	D	6	0.51		0.03	0.07								
	S	9	-0.38		-0.18									
OU	D	28	-2.50		-1.00	1.66								
	S	29	2.00		1.37									
WU	D	27	-2.35		0.60	1.69								
	S	4	0.17		-1.03									
VUW	D			**4.89**	2.37	8.93	4.18	4.37					4.85	**4.89**
	S	18	-1.16		-0.70									
Med	D	10	-0.80						0.77	0.46				
	S	2	0.30						-0.96					
Eng	D	14	1.13						0.54	0.46				
	S	16	-0.41						-0.84					
CS	D	26	2.23						**2.36**	2.78	2.18	1.37		
	S	15	-1.17						**-2.14**		-1.73			
Man	D	1	-0.01						-0.42	0.48				
	S	19	-1.91						-0.04					
AFE	D	20	-1.91						0.16	0.91				
	S	3	0.18						-0.91					
Hum	D	5	0.50						0.09	0.01				
	S	7	-0.14						-0.5					
Ag	D	25	1.80						-0.33	0.22				
	S	13	-1.35						0.46					
Law	D			**-5.46**					**-3.69**	10.52	**-3.52**	**-3.69**	-4.05	**-5.46**
	S			**6.18**					**4.58**		**4.47**	**4.69**	4.90	**6.18**
Edu	D	21	-1.46						1.01	1.05				
	S	30	2.17						0.59					
Adj-R2:				**0.943**										**0.943**
Reg F:				**286.8**									203.2	**286.8**

Note: An *F*-test on both MU dummies in col. 14 confirm that retention of both variables is rejected: *F* = 0.63.

Columns 6-15 in Table B1 show the results of a specific-to-general approach in which shift and slope dummy variables for each unit (university or discipline) were added to the basic regression specification in equation (4). The table shows results from *t*-tests and, where relevant, *F*-tests on their inclusion. These indicate (column 6) a clear case for retaining a shift dummies for LU and VUW, and possibly both dummy variables for Law using $t > 2$ as the critical value. When only those dummies are included in combination, results in col. 9 support inclusion of all four dummies. Similar test in columns 10-13 show results from an equivalent procedure based on discipline dummy variables, while columns 14-15 combine the parsimonious specifications produced by each of these approaches. This yielded the same regression specification (column 15) as that generated by the *Gets* approach (col.5).

The relevant regression lines of predicted values from the final regression in Table B1, and Table 3, are shown in Figure B1. This plots the observations for log*AQS*2003 against *AQS* growth, as well as regression lines for the average university or discipline, and those for LU, VUW and Law. As can be seen, both LU and VUW deviate from the average by fixed negative and positive amounts respectively, while Law demonstrates a significantly lower (less negative) rate of convergence than other disciplines.

Figure B1 Plot of 2003 log(*AQS*) and *AQS* growth with predicted values



References

Abbott, M. and Doucouliagos, C. (2000) Technical and scale efficiency of vocational education and training institutions: the case of the New Zealand polytechnics. *New Zealand Economic Papers*, 34, 1-23.

Abbott, M. and Doucouliagos, C. (2003) The efficiency of Australian universities, a data envelopment analysis. *Economics of Education Review*, 22, 89-97.

Abbott, M. and Doucouliagos, C. (2010) Research output of Australian universities. *Education Economics*, 12, 251-265.

Abramovitz, M. (1986) Catching up, forging ahead, and falling behind. *Journal of Economic History*, 46, 385-406.

Alm, J. (2018) What motivated tax compliance? *Journal of Economic Surveys*. doi: 10.1111/joes.12272.

Anderson, D.L. and Tressler, J. (2014) The New Zealand performance-based research fund and its impact on publication activity in economics. *Research Evaluation*, 23, 1-11.

Barro, R.J. and Sala-i-Martin, X. (1991) Convergence. *Journal of Political Economy*, 100 (2), 223– 251.

Baumol, W.J. (1986) Productivity growth, convergence, and welfare: what the long-run data show. *American Economic Review*, 75, 1072-1085.

Boyle, G. (2008) Pay peanuts and get monkeys? Evidence from academia. *The B.E. Journal of Economic Analysis and Policy*, 8, Article 21.

Buckle, R.A. and Creedy, J. (2017) *The Evolution of Research Quality in New Zealand Universities as Measured by the Performance-Based Research Fund Process*. Wellington: Victoria University of Wellington Working Papers in Public Finance, No. WP11/2017.

Buckle, R.A. and Creedy, J. (2018) The impact on research quality of performance-based funding: the case of New Zealand's PBRF scheme. *Agenda*, 25, 25-48.

Buckle, R.A. and Creedy, J. (2019a) The evolution of research quality in New Zealand universities as measured by the performance-based research fund process. *New Zealand Economic Papers*, DOI: 10.1080/00779954.2018.1429486.

Buckle, R.A. and Creedy, J. (2019b) An evaluation of metrics used by the performance-based research fund process in New Zealand. *New Zealand Economic Papers*, DOI:10.1080/00779954.2018.1480054.

Buckle, R.A. and Creedy, J. (2019c) *The 'Disciplinary Effect' of the Performance-Based Research Fund Process in New Zealand*. Wellington: Victoria University of Wellington Working Papers in Public Finance, No. WP03/2019.

Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), (2005a) *General-to-Specific Modelling*. Cheltenham: Edward Elgar Publishing, Cheltenham

Campos, J., Ericsson, N.R. and Hendry, D.F. (2005b) *General-to-Specific Modelling: An Overview and Selected Bibliography.* Board of Governors of the Federal Reserve System. International Finance Discussion Papers, No. 838 (August).

Castle, J.L., Doornik, J.A. and Hendry, D.F. (2011) Evaluating automatic model selection. *Journal of Time Series Econometrics*, 3, 1-31.

Castle, J.L. and Hendry, D.F (2014) 'Data mining' with more variables than observations. *VoxEU.org* (CEPR's Policy Portal), 13 August 2014. Available at: http://voxeu.org/article/data-mining-more-variables-observations.

de Boer H., Jongbloed B., Benneworth P., Cremonini L., Kolster R., Kottmann A., Lemmens-Krug K. and 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 http://www.utwente.nl/cheps.

Doidge, C. and Dyke, A. (2013) Taxes and corporate policy: evidence from a quasi natural experiment. *Journal of Finance*, 70 (1), 45-79.

Dowrick, S. and Gemmell, N. (1991) Industrialisation, catching up and economic growth: a comparative study across the world's capitalist economies. *Economic Journal*, 101, 263-275.

Gemmell, N., Nolan, P. and Scobie, G. (2017) *Estimating Quality-Adjusted Productivity in Tertiary Education: Methods and Evidence for New Zealand*. Wellington: Victoria University of Wellington Working Papers in Public Finance, No. WP17/2017.

Gemmell, N., Gill, D. and Nguyen, L. (2018) Modelling public expenditure growth in New Zealand, 1972–2015, *New Zealand Economic Papers*, DOI: 10.1080/00779954.2018.1458330.

Gemmell, N., Grimes, A. and Skidmore, M. (2018) Do local property taxes affect new building development? Results from a quasi-natural experiment in New Zealand. *Journal of Real Estate Finance and Economics*. Available online at https://doi.org/10.1007/s11146-017-9651-y.

Hare, P.G. (2003) The United Kingdom's Research Assessment Exercise: impact on institutions, departments, individuals. *Higher Education Management and Policy*, 15 (2), 43-62.

Hendry, D.F. and Doornik, J.A. (2014) *Empirical Model Discovery and Theory Evaluation: Automatic Model Selection Methods in Econometrics*. Cambridge, MA: MIT Press.

Hicks, D. (2012) Performance-based university research funding schemes, *Research Policy*, 41, 251-261.

Jung, J. (2012) Faculty research productivity in Hong Kong across academic discipline. *Studies in Higher Education,* 2, 1-13.

Mahoney, P. (2004) *Performance-Based Research Funding*. Wellington: Parliamentary Library Background Note, 2004/07.

Ministry of Education (2012) *A History and Overview of the PBRF*. Wellington: Ministry of Education. Retrieved from https://education.govt.nz/assets/Documents/Further-education/Policies-and-strategies/Performance-based-research-fund/PBRFHistoryAndOverview.pdf

Ministry of Education (2013) *An International Comparison of Performance-Based Research Funding Schemes*. Retrieved from: http://www.minedu.govt.nz/NZEducation/EducationPolicies/TertiaryEducation/PolicyandStrategy/~/media/MinEdu/Files/EducationSectors/TertairyEducation/PBRF/InternationalComparison.pdf

New Zealand Tertiary Education Commission (2002) *Investing in Excellence: Report of the Performance-Based Research Fund Working Group*. Wellington: Ministry of Education and Transition Tertiary Education Commission.

New Zealand Tertiary Education Commission (2007) *Performance-Based Research Fund. Evaluating Research Excellence – The 2006 Assessment.* Wellington: 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*. Paris: OECD. Retrieved from http://www.oecd-ilibrary.org/.

Payne, A.A. and Roberts, J. (2010) Government oversight of public universities: are centralized performance schemes related to increased quantity or quality? *The Review of Economics and Statistics,* 92, 207-212.

Quah, D.T. (1993) Galton's fallacy and the convergence hypothesis. *Scandinavian Journal of Economics*, 95, 427–443.

Roa, T., Beggs, J.R, Williams, J. and Moller, H. (2009) New Zealand's performance based research funding (PBRF) model undermines Maori research. *Journal of the Royal Society of New Zealand*, 39, 233-238.

Sabharwal, M. (2013) Comparing research productivity across disciplines and career stages. *Journal of Comparative Policy Analysis: Research and Practice*, 15, 141-163.

Shin, J.C. and Cummings, W.K. (2010) Multilevel analysis of academic publishing across disciplines: research preference, collaboration, and time on research. *Scientometrics*, 85, 581-594.

Smart, W. and Engler, R. (2013) *In Pursuit of Excellence: Analysing the Results of New Zealand's PBRF Quality Evaluations*. Wellington: Ministry of Education. Retrieved from http://www.educationcounts.govt.nz.

Wang, J. and Hicks, D. (2013) Detecting structural change in university research systems: A case study of British research policy, *Research Evaluation*, August. Retrieved from: http://works.bepress.com/diana_hicks/32/

Wanner, R.A., Lewis, L.S. and Gregorio, D.I. (1981) Research productivity in academia: a comparative study of the sciences, social sciences and humanities. *Sociology of Education*, 54, 238-253.

Wilsdon J., Allen L., Belfiore E., Campbell P., Curry S., Hill S., …, Johnson B. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. https://doi:10.13140/RG.2.1.4929.1363 .

Worthington, A.C. and Lee, B.L. (2008) Efficiency, technology and productivity change in Australian universities, 1998–2003. *Economics of Education Review*, 27, 285–298.

About the Authors

Robert Buckle is Professor Emeritus at Victoria Business School, Victoria University of Wellington
Email: bob.buckle@vuw.ac.nz

John Creedy is Professor of Public Finance at Victoria Business School, Victoria University of Wellington, New Zealand.
Email: john.creedy@vuw.ac.nz

Norman Gemmell is Professor of Public Finance at Victoria Business School, Victoria University of Wellington, New Zealand.
Email: norman.gemmell@vuw.ac.nz