Lag between Staffing and Publication in Research Entities: The Case of a University Faculty

Gregory Lee

gregory.lee@wits.ac.za
University of the Witwatersrand

Abstract

Entity-level research is often a primary focus for the management of research institutions such as universities. When measuring entity-level research output, such as departmental outputs, a crucial input is the size and quality of the staff body of that entity. It is unfortunately rare for measurement even to take this into account (for instance by calculating research output per standardised staff member). What practical scientometrics has yet to include on any systematic basis is the acknowledgement that there exists a natural lag in time between staffing and research, given the lag between hiring, writing, submission, acceptance and citation of manuscripts. This paper explores the initial implementation of a simple exponential smoothing system to deal with this lag by examining five years of staffing and publication data among 5 entities of a South African university faculty.

Introduction

Higher education institutions and researchers are often interested in assessments of publication at the level of research units (RUs), i.e. those constituent parts of institutions such as departments, centres and the like that are seen as discrete research-producing bodies (e.g. Alewell 1990; Baird 1991; Colman et al. 1995; Crewe 1988; Krampen 2008; Nederhof et al. 1993; Okrasa 1987; Schloegl et al. 2003; Tan 1986).

When measuring RU-level research, such as departmental publications, a fundamental input is the size and quality of the staff body of that entity. It is unfortunately rare for measurement even to take this into account (for instance by calculating research output per 'standardised' staff member, i.e. staff member measures adjusted for level). What practical scientometrics has yet to include on any systematic basis is the acknowledgement that there exists a natural lag in time between staffing and research, given the lag between hiring a researcher, writing, submission, acceptance and citation of manuscripts. In an examination of articles with relevant RU-level research measurement keywords, Lee (2010) shows that only one (Schloegl et al. 2003) includes a crude form of lag in staffing (i.e. average of staff numbers over a number of years prior to citation).

Lee (2010) asserts that including lag in the staff input measures is a superior practical methodology for ongoing use. However, little or no research appears to exist on the issue. This paper accordingly attempts an initial exploration of this lag by examining five years of staffing and publication data among six entities of a South African university faculty.

Adjustments for Lag

Before beginning the discussion of lag, I establish and denote the basic measures. First, let R be a measure indicating the research staff strength of a given RU. R can be a simple headcount, or a much more complex and realistic measure adjusted for contractual lengths, seniority, experience, and the like. Second, let P be some score of an RUs publication output, be that score quantity, quality, or the like. See Lee (2010) for considerations of these base measures.

Probably the simplest lag method is to link a given period's publication scores to a prior period's staffing complement for each RU (e.g. a one-year lag). Let time be $t \geq 0$ where $t = 0$ is the most recent period and $t > 1$ indicates number of time periods prior to the most recent period. Then P_t and R_t represent an RU's publication and staff values at time t respectively. One can then simply calculate staff for a specific $t > 0$ (e.g. for R_1) but use P_0 (research measures at t_0). Therefore, a measure of per-staff publications that employs a lag of staffing is:

$$\frac{P_0}{R_{t>0}} \tag{1}$$

This prior-period approach assumes that research at $t = 0$ is completely due to the prior period's staffing complement. It is more likely that research at $t = 0$ (P_0) is a cumulative effect of not only of the current staffing complement at $t = 0$ (R_0), but of a combination of both current staffing and potentially also of *several* prior periods rather than one. In other words, research might be a combined effect of several years of interactions between staff members, planning, acquisition of skills, loss of skills, phases of grants, and so-forth. Therefore, optimally, a staff measure should take into account staffing going back several periods. The simplest of these is an average of staffing over several past periods, however this is perhaps overly simplistic.

Generally, the strategy in multiple-period lagging is to take some weighted average of several periods, where periods closer to the present are often, although not always, given a higher weight. Lee (2010) suggests exponential smoothing as the preferred practical method.

Exponential smoothing (Brown et al., 1961) is a specific weighted average method with no particular cut-off for number of prior periods – the measure continually updates every new period, down-weighting periods further back and adding a portion (α) of the new period's value to the last forecast. If there is trend, Holt's double exponential smoothing adds a trend factor dictated by the second smoothing constant β. Specifically, if the current time period is t, then the forecast with trend for the next time period $t+1$ ($\hat{Y}_{t+1}$) is:

$$\hat{Y}_{t+1} = E_t + T_t \text{ where} \tag{2}$$
$$E_t = \hat{Y}_t + \alpha(X_t - \hat{Y}_t) \text{ and} \tag{3}$$
$$T_t = T_{t-1} + \beta(\hat{Y}_t - \hat{Y}_{t-1} - T_{t-1}) \tag{4}$$

Here, E_t = the basic horizontal exponentially-smoothed estimate, α is the smoothing constant dictating how the most much recent datum (X_t) is built into the most recent forecast ($\hat{Y}_t$) to produce the next forecast, T_i = the extra trend estimate at time i, and β is the trend smoothing constant dictating how much recent changes in forecasts are used to update trend estimates.

Exponential smoothing has advantages in that it has no particular limit on which prior periods the researcher builds into the current estimate, and after the researcher decides on the update factor it updates automatically.

The major consideration is how high one sets the values of the smoothing constants α and β. The higher the update factors (towards 1) the more that the index includes recent changes at the expense of past trends, and vice versa.

Aims of this Paper

This paper does *not* seek to prove lag between staffing and publication per se. As pointed out by a reviewer, such an analysis is far better served at the individual level rather than that of units, where data is obscured by other considerations. However, it is precisely this problem that this method seeks to address: at the institutional and practical levels of analysis, institutions simply cannot afford to do micro-level analysis on an annual basis of all staff and research – these data are almost inevitably too hard to collate and clean. Instead, a simpler system using unit-level data, which is typically all that Research Offices and like have, is required.

This paper therefore seeks simply to assess the implementation of an exponential smoothing system in a few RUs to make initial evaluation of the extent to which (a) the smoothing algorithm optimises β using recent movements versus prior forecasts, and (b) the extent to which the smoothed staffing figures more or less accurately track the actual research outputs in a year compared to the raw staffing data. The research will assist in drawing initial lessons in implementing the techniques.

Method

Population and Sample

The sample used for the analysis consists of six entities of a Faculty of Commerce, Law and Management at the author's institution, the University of the Witwatersrand, Johannesburg. The six RUs will not be identified by name in the findings, and include Schools of Law, Economic & Business Sciences, Accounting, a Business School, a Graduate School of Public & Development Management, and a non-teaching Centre of Applied Legal Studies.

There is no particular reason to wish to extend the sample to a given population, since this study focuses on the implementation of specific techniques rather than general inferences. However, the findings may apply to similar school elsewhere.

Measures

The key measures are staffing strength and publication output. The task of choosing measurements for RU-level publication productivity and staffing is made easier by the fact that the regulatory framework promotes a given publications output system, and the university under scrutiny has a specific method of measuring staffing intensity.

Publication outputs. First, the South African Department of Higher Education funds and judges higher education institutions on a specific publication output system that informs this paper. The department awards publication 'points' to institutions for (a) every journal article published in the ISI citation indices, the IBSS list, or a list of 'accredited' local journals, (b) academic books that have been subjected to a defined peer review system prior to publication, and chapters in such books, (c) papers published in conference proceedings with certain qualities such as peer review and a reputable external publisher, (d) certain others. Each publication point garners the institution a significant subsidy (over R100,000 or approximately \$15,000 and 11,000 euro). This system is deeply ingrained, measures are made available, and, although having its shortcomings, strongly dictates national and institutional policy. Accordingly, this paper uses these points as the measure of RU research output (P_t).

Staffing. Second, staffing strength at the focal institution is measured via the product of a "Full-Time Equivalent" score (i.e. a proportion where 1 is full-time and lower-proportion contracts garner a commensurately lower proportion) and "Research Senior Lecturer Equivalent" score (which standardises the level of academic staff, so that an associate lecturer = .6, a lecturer = .8, senior lecturer = 1, associate professor = 1.1, full professor = 1.3). This leads to a measure of staffing strength adjusted for research capacity, similar to Lee's (2010) Base Researcher Equivalent score, and is used for the RU staffing score (R_t). It is a useful measure for this analysis since in policy it coincides with expected average publications per seniority level, therefore, at least nominally, providing a score in the same metric as P. I note that this staffing score can be altered for a unit by both changes in employee flows (the entry/exit of more or less senior staff) as well as changes in employee stocks (an increase in seniority, experience and expertise of current staff, most commonly seen through promotions so that the R_t score increases proportionally as per the system above).

Results

Analysis of the Raw Data

Tables 1 and 2 give the raw data for the investigation.

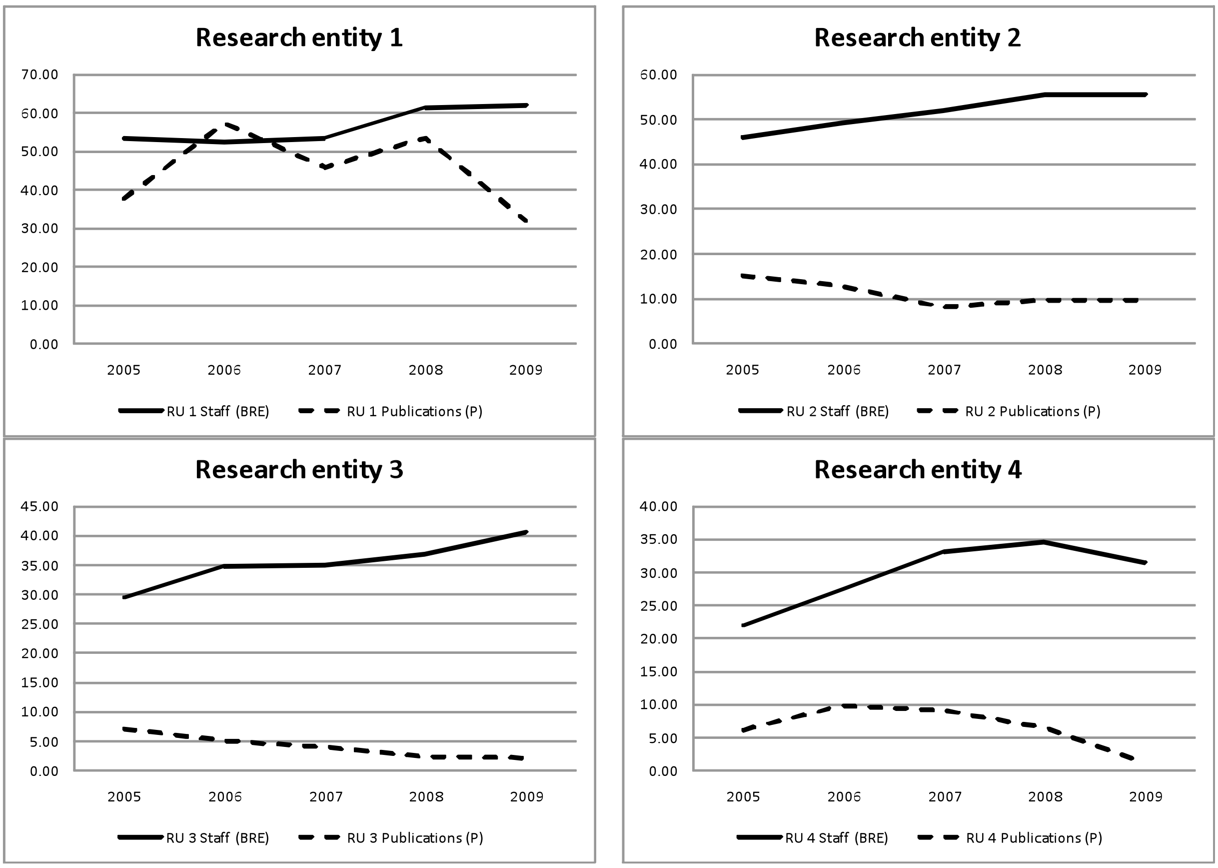
Table 1. Staffing intensity score (*R*) for each RU by year

Year	*RU 1*	*RU 2*	*RU 3*	*RU 4*	*RU 5*	*RU 6*
2005	53.38	46.04	29.42	21.90	5.60	20.27
2006	52.40	49.24	34.86	27.37	5.60	23.52
2007	53.52	51.92	34.97	33.11	5.60	29.14
2008	61.26	55.64	36.95	34.57	4.25	30.70
2009	62.00	55.57	40.61	31.47	4.84	33.40

Table 2. Research output (P) for each RU by year

Year	*RU 1*	*RU 2*	*RU 3*	*RU 4*	*RU 5*	*RU 6*
2005	37.57	15.17	7.00	6.01	3.50	12.18
2006	57.20	12.58	5.00	9.68	4.46	7.02
2007	45.59	8.02	3.93	9.03	3.95	3.05
2008	53.22	9.58	2.33	6.50	2.50	12.39
2009	31.68	9.55	2.03	1.05	2.17	17.11

Figure 1 shows graphical analyses of RU publications and staffing for each of the five years under investigation based on the data in Tables 1 and 2.



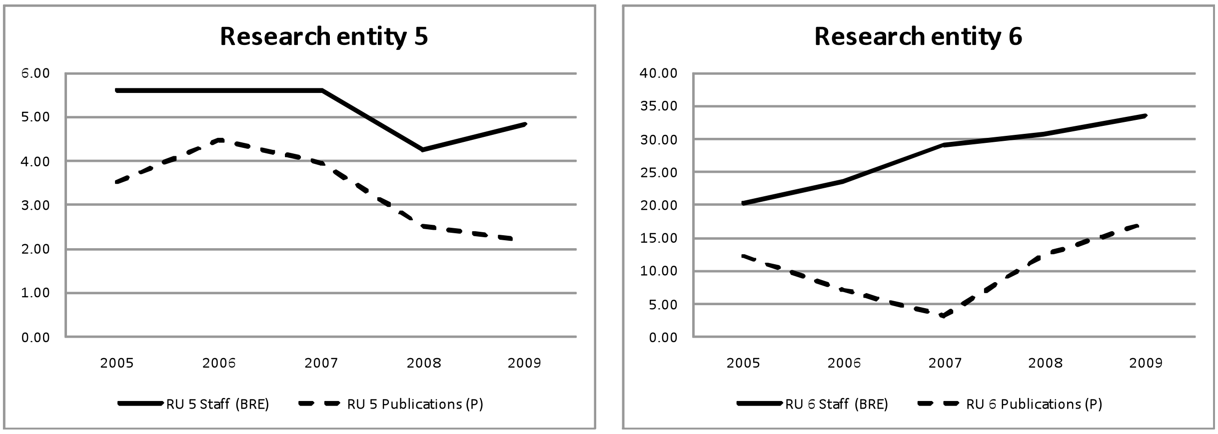


Figure 1. Graphs of research unit staffing intensity (*R*) and research output (*P*) by year

An examination of the graphs in Figure 1 allows some initial albeit tentative conclusions. Evidence for lag in this sample and time period is not in fact clear, with varied conclusions. RUs 2, 4, 5 and 6 mostly have similar-shaped staffing and publications graphs with the exception of an initial 2-year research plummet in RUs 2 and 6. RU 1 has mostly steady staffing but sporadic (high levels of) research. One unit (RU 3) increases slowly in staffing but steadily decreases in publications. It is clear that such data cannot be used to illustrate lag itself. However, the general principle of possibly using exponential smoothing to ameliorate recent significant movements in either research or staffing remains.

Illustration of the Dangers of Not Applying Lag

Given the data from Tables 1 and 2, if an institution judges the research from its units on the basis of publications per standardised staffing member *in the same year* (which is precisely what occurs at the author's institution), i.e. P/R_t, then the following will occur:

- In 2009, RU6 with 17 publications and 33 senior lecturer equivalents in 2009 will come out as the best performer with .52 points per researcher. This is possibly equitable for this year since this unit bounced back from a steady decline until 2009 with commensurate increases in both staffing and research;
- RU1 would be rated a close second with .51 publications per researcher. However, this ratio is strongly dictated by a 2008-2009 increase in staffing. If the weaker staff bodies prior to 2008 had been included in the ratio, and it is quite likely that a good proportion of these publications were really written and submitted in 2007 and even before, this unit would be rated first;
- RU5 with .45 would come in third. Again, this ratio is harmed by the 2009 increase in staffing, whereas the staff body prior to this is lower;

Therefore, the rank ordering could be quite different if prior staff bodies are taken into account as part of the input into current publications.

Examination of Mean Absolute Deviation

Table 3 examines mean absolute deviation (MAD, i.e. the absolute value of the difference between two scores, in this case between *R* and *P*) for each RU and for various single-period lags. Average MAD is lowest at a three-year lag, e.g. when staffing is calculated for 2005 and research output for 2008, although this is calculated for only two research periods.

Table 3. Mean absolute deviation between staffing and research output with lag x

Lag	*RU1*	*RU2*	*RU3*	*RU4*	*RU5*	*RU6*	*Ave*
Zero	13.38	40.70	31.30	23.23	1.86	17.06	21.26

One year	10.13	40.78	30.73	22.67	1.99	16.02	20.39
Two years	10.15	40.02	30.32	21.93	2.73	13.46	19.77
Three years	10.44	38.08	29.96	20.86	3.27	7.15	18.29
Four years	21.70	36.49	27.39	20.85	3.43	3.17	18.84

Table 3 does seem to indicate that, for this sample, staffing levels predict future research with at least some lag, although evidence is weak at best.

Initial Exponential Smoothing Solution

If it holds true that exponential smoothing could help describe the average historical staffing, then the task of the analyst is to solve for α and (if using trend) β in Equations 3-4. This section explores this approach as well as the implications for predicting research.

Here to allow for relative comparisons I took the approach of also including Mean Absolute Percentage Error (MAPE) as an output measure in conjunction with MAD and adopted the following procedure. First, I applied the exponential smoothing formula for the staffing of each school, using NCSS's exponential smoothing modules which is set to optimise α and β for lowest MAPE.

The two approaches were to calculate smoothing constants for the entire sample, therefore minimising net MAD, or to allow smoothing constants and MAD to vary between RUs. For the horizontal forecast, minimising a single overall MAD results in an optimised $\alpha = .81$ and a minimised MAD of 22.46 for 2007-2009. Setting a different α per RU results in vastly varying βs, ranging from 0 to 1 (i.e. no updates for new values versus single period lagging) and βs in-between (i.e. allowing for an iterative staff value). Net average MAD with this method is 19.96 for 2006 to 2009, which is lower than an net average MAD for single-period lagging of 21.49 for the same periods when that was available.

With trend, disaggregation by unit remains a better prospect. The smoothing constants again differ from zero in one case each upwards to a high of .62 for α and .56 for β. Overall, the application of trend appeared to add more to the concordance between staffing and research, reducing overall MAPE by 7.46%.

Discussion

This paper provides an initial, simple exploration of the use of lagging when linking staffing in RUs to research output. The exploration does rely heavily on idiosyncratic data from a handful of RUs. However, the preliminary evidence presented here appears to indicate several things of interest to the lag hypothesis.

First, at least some evidence of lag appeared to exist, although the evidence is not entirely conclusive. In at least a few of the RUs presented, patterns in staffing do appear to parallel patterns in research. Furthermore, the seeming predictive ability of staffing appears to improve, albeit not particularly strongly, if lag is used in staff measures, especially around the 2-3 year mark in single lagging. However, this paper does not seek to conclusively prove lag: as stated earlier, a micro-level analysis would be a far better approach to this. Because institutions cannot generally hope to do extensive annual-level data collection on this basis, however, their limited data may require an approach such as that taken here.

The results are clouded by what may be some specific events or periods in some of the RUs. RU 3 experienced a consistent albeit shallow drop-off in staffing numbers, while research marginally improved over the period, while in both RU 2 and RU 6 research plummeted over the periods 2005-2007 and then picked up again. The research of RU was also quite erratic. This underscores the need in smaller samples to explore deeper-level qualitative data around the context. However, in bigger samples – and specifically if institutions use such measures over all RUs and years – average results would be less susceptible to idiosyncratic events.

With regard to the simple exponential smoothing exercise presented here, it does appear that measuring staff with a rolling, weighted update for multiple prior periods provides a smoother and more accurate staffing measure.

There are several considerations for future research. Obviously an institution should explore correct lag with a complex micro-analysis, and consider applying this lag within the exponential smoothing paradigm if required. Larger samples and timeframes would make for better estimations, as would more complex estimation techniques in this regard such as panel regression. A second consideration is the reverse time order may apply, in which research movements pre-figure staff changes. One example in this regard is senior researchers failing to publish in their last few years before retirement, while remaining on the books. This would lead to decreasing research but high staff scores. If such researchers were then replaced by productive but perhaps more junior members, research would increase but staffing decrease. Again individual-level research would perhaps show this.

Another example of such a reverse effect could be where generally poor research leads to the exodus of the few productive members to 'greener pastures', in which case research scores would pre-figure staffing changes. Although it is unlikely that practical, institutional measures at the RU level could account for this, the phenomemon may bear research and thought.

Having said this, an examination of the individual research records clearly shows a great number of cases in which publications are accredited to an RU in a certain year when the authors had left the RU in a previous year, sometimes many years prior. Clearly there is some scope for lag adjustments.

There exist substantial data problems in this sort of research. For instance, the researchers discovered an unexpected phenomenon: a considerable proportion of the research produced by and accredited to some RUs is authored by people not accounted for in the official institutional staff list: honorary and emeritus staff, students publishing dissertations and the like, even sessional lecturers. This is especially true of law and business, although it occurs throughout. Much effort had to go into dealing with these issues.

Conclusion

The staff complement of an RU is its most precious asset, and one that operates in complex ways. The fact that the publishing process is by no means immediate is well known. It seems clear that connecting current staffing in RUs to current publication will lead at best to some inaccuracies and at worst to considerable misunderstanding. For this reason, research should continue to explore how institutions can best include the lag between staffing and publishing in their measurements involving these variables.

Acknowledgments

This material is based upon work supported financially by the National Research Foundation. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author(s) and therefore the NRF does not accept in liability in regard thereto.

References

Alewell, K. (1990). Criteria for performance profiles of departments and universities. *Scientometrics*, 19(5–6), 337–347.

Baird, L. (1991). Publication productivity in doctoral research departments: Interdisciplinary and intradisciplinary factors. *Research in Higher Education*, 32, 303–318.

Brown, R. G., Meyer, R. F., & D'Esopo, D. A. (1961). The fundamental theorem of exponential smoothing. *Operations Research*, 9(5), 673–687.

Colman, A. M., Dhillion, D., & Coulthard, B. (1995). A bibliometric evaluation of the research performance of British university politics departments: Publications in leading journals. *Scientometrics*, 32(1), 49–66.
Crewe, I. (1988). Reputation, research, reality: The publication records of UK departments of politics 1978–1984. *Scientometrics*, 14(3–4), 235–250.
Krampen, G. (2008). The evaluation of university departments and their scientists. *Scientometrics*, 76(1), 3–21.
Lee, G.J. (2010). Assessing publication performance of research units: Extensions through operational research and economic techniques. *Scientometrics*, 84:717-734.
Nederhof, A. J., Meijer, R. F., Moed, H. F., & Van Raan, A. F. J. (1993). Research performance indicators for university departments: A study of an agricultural university. *Scientometrics*, 27(2), 157–178.
Okrasa, W. (1987). Differences in scientific productivity. *Scientometrics*, 12(3–4), 221–239.
Schloegl, C., Gorraiz, J., Bart, C., & Bargmann, M. (2003). Evaluating two Austrian university departments: Lessons learned. *Scientometrics*, 56(3), 287–299.
Tan, D. (1986). The assessment of quality in higher education: A critical review. *Research in Higher Education*, 24, 223–265.