

# Policy Screening by Structural Change Detection: Can Policies Effectively Boost Research System Performance?

Jian Wang\* and Diana Hicks\*\*

\* [jianwang@gatech.edu](mailto:jianwang@gatech.edu)

Institute for Research Information and Quality Assurance (iFQ) Schützenstraße 6a, D-10117 Berlin, Germany

\*\* [dhicks@gatech.edu](mailto:dhicks@gatech.edu)

School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, Atlanta, GA 30332-0345, USA

## Abstract

The university research environment has been undergoing profound change in recent decades. Aiming at international competitiveness and excellence, performance based university research funding systems have been implemented in many countries. However, evidence-based analysis of policy effects is scarce. This paper develops methods for evaluating the effect of university research policy on university system research output. We assume stable dynamics between inputs and outputs, and that effective policy change introduces external interventions and therefore structural changes into the system. We examine trends in input—HERD—and output—publications—at the national level over three decades and look for breaks signaling structural change.

## Introduction

In many countries where universities are governed as a national system, the university research environment has undergone profound change in recent decades. Aiming to enhance international competitiveness and research excellence, a variety of policies have been implemented to introduce research performance incentives. In particular more than a dozen countries have implemented national university research evaluation systems (Hicks, 2012), others have introduced center of excellence competitions (China, Germany, Japan), while others developed national individual level evaluations, i.e. Spain, South Africa. Other countries have not implemented explicit research performance incentive systems. The universally stated goal of governments that do implement such systems has been to increase international research excellence, not in one university, but rather in their university system as a whole.

These policies tend to be controversial and unpopular—often accused of damaging the systems they seek to enhance. Yet rigorous assessment of these systems is in its infancy. Academics dislike the introduction of evaluation systems on principle and have therefore concentrated more on commentary than on impartial evaluation. The academic literature tends to report anecdotes or be based on surveys gathering complaints from effected scholars (Butler, 2010). If evaluations have been quantitative, they have not narrowed down to university output, and/or have not been internationally comparative (Butler, 2010). So, for example, nobody has actually compared countries with and without policies to explore whether having a policy makes any difference. There is a need to view the policies from a distance and attempt to connect policy shifts aimed at increasing research excellence with changes in international research output.

Existing literature is largely national. International studies are challenging because they require some level of understanding of multiple, complex shifting policy landscapes. It is difficult to get this right. For example, one of the only comparative policy analyses misdated some policy

introductions; misidentified administrative shifts in policy implementation with first introduction of a research incentive system, and excluded the most developed system because a limited time period was analyzed (Franzoni & *al.*, 2011). In a review for the OECD, Hicks found fourteen performance based research evaluation systems that had been implemented as of 2010. The systems were found to be complex and dynamic, balancing peer review and metrics, accommodating differences between fields, and involving lengthy consultation with the academic community and transparency in data and results. Although the importance of the systems seems to be based on their allocation of universities' research funding, this is something of an illusion, and the literature agrees that it is the competition for prestige created by public, university level performance metrics that creates powerful incentives within university systems. The study identified center-of-excellence policies as an alternative method of achieving research excellence in a university system. Here we carry out the research recommended in the OECD study, namely, comparison of the effectiveness of different methods in achieving the primary, stated goal of increasing the excellence of a nation's research.

The motivation for performance based funding of university research is simple:

The rationale of performance funding is that funds should flow to institutions where performance is manifest: 'performing' institutions should receive more income than lesser performing institutions, which would provide performers with a competitive edge and would stimulate less performing institutions to perform. Output should be rewarded, not input. (Herbst, 2007, p. 90)

Thus many governments have shifted away from a system that distributed research funding based on input (number of faculty, students etc.) to one based at least in part on output (publications). To understand why this was necessary, it is necessary to understand the context of foreign university research.

Universities in most countries can be viewed as national systems, governed by a Ministry of Education, with professors often having civil service status. Like much of the public sector, such systems are subject to a principal-agent problem (Miller, 2005). The agents (university scientists) take actions (conduct research) that determine the payoff to the principal (the government) of its investment in research. The principal can readily observe the outcome (papers, citations, standing of universities in world rankings etc.) but not the day-to-day activities of the agents. The agent has control over the daily operations but less control over the ultimate impact of the work and is therefore inclined to report performance in terms of daily activities regardless of impact. The principal's interest is in maximum international impact for its research investment but cannot afford day to day control over the agent's activities. Therefore, the principal has an incentive to set up the terms on which university funding is awarded using independent measures of impact as a criterion.

In recent decades, public sector reform has moved in the direction of governments as principals imposing output based incentives on its agencies and programs across policy domains (Heinrich & Marschke, 2010). This movement, called new public management, was broadly international and affected university governance and funding too (Kettl, 2005). New public management was an international movement to change how governments run with an eye to enhancing productivity, relying more on private markets, instilling a stronger public service orientation and enhancing accountability. However, the manifestation of new public management and the extent

of its influence varied with the context of public administration in each country. Pollitt and Bouckaert identified five dimensions along which countries vary and which jointly explain differences public management reform. These dimensions are:

- Degree to which the state is centralized or decentralized/federal
- Whether the governing convention is consensual or majoritarian
- Whether the political appointees and civil servants are integrated or separate and whether relations are politicized or not
- Type of administrative culture: Anglo-Saxon public interest motivation or legal tradition based on the centrality of the state to society.
- Policy advisors—whether restricted to civil service, or sourced more broadly.

The variety of administrative and governance cultures meant that new public management manifested not as a monolithic standard but as a series of interventions differing in timing and character. These interventions are amenable to causal analysis since they constitute a series of natural experiments. New Zealand and the UK saw particularly strong and early implementations including early introduction university research evaluation upon which university funding depended (Hicks, 2012). Other countries lay somewhere in between the strongest systems and no systems introduced.

Though scholars trace the origins of performance based university research funding to the new public management movement, the stated goals of governments introducing these systems frame things differently. Excellence appeared frequently as a stated goal of national governments for these policies. In Australia the stated goal of the new research evaluation system was to identify and promote excellence across the full spectrum of research activity, including discovery and applied research, in Australia's higher education institutions (ARC, 2009, p. 11). In New Zealand, the primary purpose was to ensure that excellent research in the tertiary education sector is encouraged and rewarded (New Zealand, Tertiary Education Commission, 2010). In Norway the goal was to increase research activities and allocate resources to centers performing excellent research (European Commission, 2010, p. 120). Sivertsen reports the goal of the Norwegian publication indicator was “to measure and stimulate the research activity at the level of institutions and to enhance the focus and priority they give to research as organizations” (Sivertsen, 2009, p. 6). Governments’ stated rationales for performance based university research funding systems thus concentrate on the general pursuit of excellence. In its concern for goals other than efficiency, movement toward performance based university research funding is perhaps more reminiscent of the newer “public values” movement (Stoker, 2006), than of new public management. The rhetoric of excellence became so dominant in government statements about universities that something of a backlash has developed in the form of critique of the concept of “excellence” as it is applied to universities (Rosten & Vaira, 2011). In sum, the governmental emphasis on excellence as the articulation of the policy objectives in this context justifies an investigation of trends in international research output.

## **Approach**

We assume that a national university research system has stable dynamics between research inputs and outputs, and effective policy change is an external intervention that introduces structural change into the system. Different from classical policy evaluation approaches which focus on a specific policy and aim to assess impacts of this focal policy, our approach starts with evidence-based empirical data analysis to determine whether or not a country’s science system

exhibits a structural change in the relationship between input and output over the past three decades. We then compare the empirical results with narratives of the development of university science policy over the same time period to identify possible relationships between policies and the presence or absence of structural change. In this approach we seek to connect detection of structural change with mapping of policy change.

In addition, we are interested in national university research policies. However, more than other fields, research in clinical medicine involves non-university institutions, i.e. hospitals and firms, and has considerable funding sources outside HERD. This blunts the impact of the factors we investigate here, HERD and policy changes to national university systems. Therefore, we examine all fields of research, and where available, data excluding the field of clinical medicine. We analyze almost three decades of publication and funding data between 1981 and 2007. Although our intent is to eventually analyze all OECD countries, in this initial round we began with the large countries whose policy histories are most familiar to us: United States, Germany, United Kingdom, Canada and Australia.

Our first variable is publication output, **PUB**, the number of Thomson Reuters Web of Science (WoS) indexed journal publications (articles, letters, notes, and reviews), with at least one university affiliation. The United States National Science Foundation (NSF) journal field classification scheme developed by the Patent Board is used for classifying journals into research fields. It is a two-level system classifying journals into one unique research field and subfield. However, we keep journals with WoS subject category 'multidisciplinary sciences' as 'multidisciplinary sciences.' Furthermore, the NSF scheme does not cover the arts and leaves some social sciences and humanities journals as 'unassigned,' so we manually code the remaining journals (which are not classified by NSF scheme or classified as 'unassigned'). Most of them are about literature and arts. After assigning each paper into a unique research field, we analyze all fields as well as exclude the field of 'clinical medicine' and analyze publications in all other research fields.

Our second variable is research funding, **HERD**, the annual higher education research and development (HERD) expenditure in constant 2005 dollars and discounted for purchasing parity power collected from OECD. We use both total HERD and HERD excluding the field of medical sciences. The field breakdown of HERD is not available for all countries and not over the entire time span, limiting the scope of the non-medical analysis. Missing values are interpolated by Loess regression. To ensure reliability, we avoid extrapolation, that is, we do not fill missing data before the earliest available year or after the latest available year.

In order to assess whether growth in publication has departed from the expected path, we need to specify our expectations of growth. Publications and research funding grow not by a set amount every year (1000 papers or \$1 million for example) but rather they grow by a percentage (2% per year on average for example). If publications and research funding grew by a set amount, growth would be linear over time. However, growing by a set percentage every year is an exponential growth pattern. Because the time series of **PUB** and **HERD** grow exponentially we take natural logarithm of the variable, and then take the first order differentiation. This transformation indicates the growth rate of the original variable and is widely used in econometric analysis. It gives two new variables for time series modeling:  $d\ln\text{PUB}$ ,  $d\ln\text{HERD}$ .

## Detecting Structural Change

We look for significant structural change in the time trends for **PUB** and **HERD**. Two methods are used: (1) the cumulative sums of standardized residuals method (CUSUM) proposed by Brown et al. (1975); and (2) the structural change model (SCM) methods based on F statistics initially proposed by Chow (1960). The tests provide a general picture of whether there is any structural change and around what time did the structural change(s) take place. Subsequently, we use the SCM method to estimate the exact year and 95% confidence interval of the year of the structural change(s). The idea is to partition the time series into several segments to minimize the residual sum of squares (RSS) of the linear regression system. However, RSS always decreases as the number of segments increases; therefore, we put a penalty on number of segments and instead choose to minimize BIC. In other words, the partitioning solution that minimizes BIC is selected. Both testing and estimation are implemented by the R package “strucchange” available at: <http://cran.r-project.org/web/packages/strucchange/index.html>.

**PUB:** Detecting structural change in the growth of **PUB**, regardless of any causes, is to ask the question: Does  $dlnPUB$  have a constant mean, namely is  $dlnPUB$  invariant over time? We do not limit ourselves to any specific causes at any specific time, therefore, we test each possible time point during the whole time period of 1982-2008 (we lose the 1981 data point when taking differentiations). We test whether there is a significant structural change at each possible time point. In other words, at each time point, we partition the time series into two segments and test if both segments have the same mean. We have the time series between 1982 and 2008 available, the CUSUM method can test all years, while the Chow’s test can test years between 1983 and 2006, because it has to reserve certain number of initial (and final) points. Testing results are reported in Table 1. The testing finds a variety of breaks in the PUB trends, some of which appear to be associated with structural changes in the growth of the underlying database. We interpret these results below.

**HERD:** we use the same procedure to analyze  $dlnHERD$ . The only country to show breaks in growth of **HERD** is the US. This suggests that in most countries funding growth has been stable, and so structural change in **PUB** is not due to changes in funding.

**Table 1.** Structural changes detected.

Country	Structural break	Year	95% confidence interval	Confirmed in non-medical data	Interpretation
Australia	<i>PUB</i> up	1991	[1988, 1996]	year becomes 1989	Dawkins reform/quality assurance scheme
Germany	<i>PUB</i> up	1991	[1990, 1994]	*	reunification
	<i>PUB</i> down	1998	[1997, 1999]	*	end of reunification effect
	<i>PUB</i> up	2003	[2001, 2004]	*	pay law
Canada	pub down	<b>1994</b>	<b>[1991, 2001]</b>		database effect
	pub up	<b>2002</b>	<b>[2001, 2003]</b>		database effect
United Kingdom	<i>PUB</i> up	1989	[1988, 1991]		RAE 2
United Kingdom	pub down	<b>1995</b>	<b>[1991, 1996]</b>		database effect
	pub up	<b>2003</b>	<b>[2001, 2009]</b>		database effect
United States	<i>HERD</i> down	1989	[1988, 1998]		
	<i>HERD</i> up	1998	[1988, 1998]		NIH doubling begins
	<i>HERD</i> down	2003	[2002, 2004]		NIH doubling ends
	pub down	<b>1995</b>	<b>[1991, 2002]</b>		database effect
	pub up	<b>2002</b>	<b>[2001, 2003]</b>		database effect
database	pub down	1984			
	pub up	1988			
	pub down	<b>1995</b>			
	pub up	<b>2002</b>			

## Results

Table 1 reports the structural changes we detected in the **PUB** and **HERD** time trends for five countries: Australia, Canada, Germany, United Kingdom and United States as well as breaks in the growth rate of the database. Our interpretations of each break follow.

**Australia:** We find a significant structural change in *PUB* 1991(all fields) or in 1989 (data excluding clinical medicine). The 1991 dating is visible in the data of Butler (2003), in particular in the graphs plotting Australian share of ISI publications. There are two major policy changes in the Australian system that stand as possible causes.

In 1988, a Unified National System of universities was introduced, trebling the number of universities eligible for research funding (Guená & Martin, 2003, 293). Between 1989 and 1992, 15 Colleges of Advanced Education, 11 Institutes of Technology, and 6 Institutes of Advanced/Higher/Tertiary Education were merged with universities or converted into universities. These institutions transitioned from non-degree awarding, no research expectation to degree awarding with expectation of research.<sup>1</sup> This was not the end of the changes imposed on the Australian university system.

<sup>1</sup> Source: Wikipedia page entitled "College of Advanced Education," viewed May 30, 2012

The government also imposed research output evaluation upon which part of university research funding depended. The dating of this policy implementation varies somewhat:

- In October 1991 the Education Minister, Peter Baldwin, announced that the government had decided to establish an academic audit scheme. Both teaching and research would be evaluated and money would depend on the outcomes (Harman, 1998, 337). The resulting quality assurance scheme operated from 1993-1995.
- “Since 1992, all universities have been required to supply details of their publication output” to the Education Department (Butler, 2003).
- In her 2010 OECD piece, Butler dates the introduction of publication collection to 1993 (Butler, 2010, figure 4.1).
- “In 1993, the Minister for Education announced that, as from 1995, the Research Quantum would be allocated on the basis of [a formula incorporating publication counts].” (Guena & Martin, 2003, 293).
- 2006 is noted as the year the Research Quality Framework was introduced by Franzoni et al. This however marked a redesign in a pre-existing system, not the first introduction of research evaluation. As such, we would not expect a major shift in behaviour in response.

Butler (2003) attributed the shifts she saw to the introduction of evaluation rather than the creation of new universities. She isolated the upturn to university publishing, separating it from other sectors, and the other sectors did not show the same trend. She pointed out that the new universities were not especially research active and published in journals not indexed in ISI. Also, she examined output from two universities not involved in the university reorganization and found increase in publication per faculty member in response to the introduction of more strategic university management attempting to raise the universities' research profile.

However, the breakpoint date is rather early to be attributed to an evaluation scheme that really only became relevant to funding in 1995. It would seem that some credit must be given to the creation of the Unified National System to explain the dating of the breakpoint. On the other hand, starting with the October 1991 announcement, a culture of evaluation and audit was developed and imposed on Australian universities with the intent of building more strategic management capacity. Butler's case studies illustrated the responses of two universities and the results they achieved in increased publication output. It seems likely that the enhanced trajectory initiated with the system unification was maintained over the long term because the system evolved to respond to the new system wide incentives strongly focused on ISI publications.

**Database/Canada:** One of the challenges in looking for changes in scientific output due to policy implementation is that scientific output is measured in a database, Web of Science (WoS), and Thomson-Reuters can choose to enlarge the database or not for commercial reasons. So the database is subject to change for reasons entirely unconnected with national policy shifts. To control for this effect, we used the same techniques to find breakpoints in the growth of number of papers indexed in WoS. We found four breaks. The database grew slower between 1984 and 1988, and between 1995 and 2002 than it did before or immediately after. The acceleration in growth was especially dramatic in 2002, likely because in 2003 Elsevier announced the launch of Scopus, the first direct competitor to WoS, whose point of differentiation is its broader coverage.

The last two break points, in 1995 and 2002 appear also in the publication trends of the core English language countries whose publications are favored in WoS coverage—Canada, the US and UK. Like the database, each of these countries showed a slowdown in publication growth in

1995 (1994 in Canada). Like the database, each countries' publication growth accelerated in 2002 (2003 in UK). Other countries do not exhibit this 1995/down-2002/up breakpoint pair.

**Germany:** German publication output grew faster between 1991 and 1998 than before or after. Germany reunified on October 3, 1990.<sup>2</sup> Thus 1991 was the first full year of unity. Publications from former East Germany are included in the publication count from the beginning. The same is not true of HERD, which incorporated East German data for the first time in 1991. Interestingly, there is no breakpoint in HERD. This is because HERD took a large step up in 1991, but thereafter continued to grow at the same rate as before reunification. The reunification did however seem to unleash dynamics of higher publication growth, at least for seven years. Presumably this reflects Eastern academics benefiting from more resources to realize their latent potential. The process may have just reached a natural conclusion in 1998. Though it is notable that 1998 marks the end of the 16 rule of Chancellor Helmut Kohl with the loss of an election to Gerherd Schröder who took office on October 27, 1998.

Faster growth in publication resumed in 2003. Since Germany is not a core English language country, we look for an explanation other than the growth in the database discussed above. Causal attribution is not easy. Franzoni et al. identify 2004 as the year in which an individual level performance incentive was introduced with the implementation of the professor salary reform law passed in 2002. "It enables institutions to more freely negotiate professors' salaries and to link salary to performance." (Franzoni & al., 2011, Table S1). However, the impression gathered from the German literature is not so clear cut. In 2002 a supplement to the Fifth Amendment of the Federal Framework Act governing universities was passed that addressed professors' salary. It is known as *ProfBesReformG*. This did allow for salary bonuses as part of appointment or tenure negotiations, but these were not new incentives as they were allowed under the previous law. *ProfBesReformG* introduced performance bonuses for engaging in university administration as well as for outstanding achievement in research, teaching, art, education and youth development (Kräkel, 2006). Thus the incentives were not focused on publishing but accrued to all types of achievement.

The extent of implementation is also unclear. The new bonuses had to be implemented without increasing the overall salary budget. Authors emphasize the resulting need to decrease salaries, presumably for newly hired junior faculty, to make room for the bonuses (Pritchard, 2006). Pritchard notes that implementation of the new salary scales depends on a complicated interaction between the *Bund* (federal government), *Länder* and universities and so is inconsistent across the country. Procedures had not been finalized as of 2006 when Pritchard's article was published (Pritchard, 2006, 108).

This is typical of the German system. Universities are under the control of the *Länder*; the Federal government provides the overall structure with the Framework Act. Within universities, professors have a notoriously high degree of autonomy since chairs control a research budget, staff and junior faculty. As civil servants, these cannot be taken away. Academic salaries are paid by the Ministries in the *Länder* and so are not in the universities' budgets. Reform thus proceeds slowly and terms can only be changed for new hires. Lange's discussion of attempts to reform the German system by introducing research evaluation emphasizes the resistance in the system to change and the inability of higher levels to force change upon professors. Lange's discussion of changing governance of academic research in Germany makes no mention of *ProfBesReformG*.

---

<sup>2</sup> Source: Wikipedia, German Reunification page.



In conclusion, the German situation is complex. The literature would lead one to believe that no reform has been implemented on the national scale that would be strong enough to shift the dynamics of the system. Yet, Franzoni et al. point to the new ability to reward performance in salary. Though the extent to which this has been implemented is unclear, the data suggest a shift in the system timed to coincide with the new law. Other possible explanations for this breakpoint include the accelerated growth in the database around this time, though Germany would be the only non-Anglo Saxon country affected, and the accumulations of signals at all levels of the German system *Bund, Länder* and university that times are changing and more is expected—especially in the area of English language publication.

**United Kingdom:** As described above the U.K. exhibits the two database breaks in publication growth—a slowing in 1995 and an acceleration in 2003. There is one additional break to be explained, an acceleration in publication growth after 1989. This break coincides with the second round of the Research Selectivity Exercise (RSE, as the Research Assessment Exercise was initially called). The first university research assessment in the UK was conducted in 1986, the second in 1989, the third in 1992. That the break seems related to the second and not first round of the evaluation is interesting. It makes quite a bit of sense, given changes between the first and second rounds. Martin and Whitley (2010, 54-57) discuss the evolution of the RSE and the differences between the 1986 and 1989 versions. In the 1986 round, departments submitted details of their five best publications from the previous five years. The effect on funding was rather limited and “some in the more established universities paid relatively little attention (hoping, no doubt, that the RSE would 'go away'), others took it much more seriously” (Martin, & Whitley 2010, 55). This suggests a limited effect on university behaviour, which is consonant with the lack of a break in the publication trend data. In 1989, departments submitted details on up to two publications per faculty member (raised to 4 in 1996) as well as the total number of publications in relation to full-time staff. Further, the results of the evaluation were more explicitly linked to a larger amount of funding—half of the research portion of the block grant was allocated on the basis of the 1989 ratings (Martin & Whitley, 2010, 56-57). The increased importance of each individual's productivity in the ranking, as well as the greater financial stakes, and no doubt the sense that this was not “going away,” all suggest a more substantial impact on faculty behaviour, aligned with the shift in UK university publications to a faster growth trajectory beginning in 1989. In 1992 departments were allowed to submit information only for research active staff. Thereafter, tweaks were made, but the method had settled down.

The breakpoint analysis suggests not only that the RAE policy was strong enough to shift the university system in the desired direction, but that the design of the system mattered and that the big shift was seen when individual level productivity began to matter and when significant money began to move. Unfortunately, it is not possible to disentangle the effects of these two factors in this instance. The conclusion that individual level attention is needed to create a shift is aligned with the findings of Franzoni & al. (2011).

**United States:** The United States shows two breakpoints in publication count. As described above, we attributed these to shifts in database coverage. The only breakpoints in HERD were found in the US data, and were not associated with breaks in publication trend. U.S. HERD grew faster between 1998 and 2003. This corresponds to the period during which NIH research funding doubled. U.S. HERD also grew more slowly after 1989 than before. This is more difficult to explain. We consulted *Science & Engineering Indicators 2012* Appendix table 5-2 to see academic R&D expenditure going back further than the OECD data, and also broken down by source. Visual inspection of the data suggests that academic R&D funding grew faster in the

late 1980s than before or after. The drop is seen across all funding sources: federal government, state and local government, industry and academic institution own funding (though with a delay). This coincided with the second, and last, term of President Reagan, and the buildup in defense expenditure designed to bring down the Soviet Union. President Reagan's term in office ended in January 1989 when President Bush was inaugurated.

## **Conclusion**

In this preliminary analysis we looked for structural change in growth of HERD and publications from universities in five large countries: US, Germany, UK, Canada and Australia. We found a series of breakpoints in the data, some more surprising than others. First we note that growth in the database in which publications are measured is uneven and showed four breakpoints. The last two of these seemed to be echoed in the data of the US, UK and Canada, which is not a surprise because these core English speaking countries are favored in the database coverage. We also found that US HERD grew faster between 1998 and 2003 than it did before or after. This is to be expected as this was the period during which NIH funding doubled. Unfortunately this did not seem to be echoed in a corresponding upshift in publication growth.

Also not surprising, at least in retrospect, is the accelerated growth of German publication after reunification and continuing for seven years thereafter. Adding a highly skilled, yet underresourced set of faculty and institutions to your system, and giving them more resources, seems guaranteed to generate faster growth in publication output.

Breaks were detected in publication growth in Australia and UK. Both shifted to a higher growth trajectory, the UK in 1989 and Australia in 1991. In a way this study was framed to look for these two breaks which appear to be the result of policy interventions. Both Australia and the UK early implemented strong evaluation systems that produced public rankings of universities' research strengths upon which non-trivial amounts of money depended. Both countries shifted to a higher growth trajectory in publication output. However, the Australian shift started before the evaluation began in earnest, perhaps relating also to the conversion of a number of institutions to university status in the late 1980s.

Most surprising was the downshift in growth in HERD in the US in 1989 and the up shift in German publication growth in 2003. Like the 1998 end of faster publication growth in Germany, the 1989 drop in growth of HERD corresponds to a change in administration. The upturn in German publication growth in 2003 coincides with a change in law allowing bonuses to be paid to professors for exceptional performance. However, it is not clear how far this has been implemented or what effect it has had.

This first round of analysis illustrates some of the difficulties of connecting system wide policy shifts with changes in research output. First shifts in the publication database must be recognized and accounted for (which is why many analysts including Butler (2003) look at share of world publications). Second, policies other than science policies may have the largest effects—German reunification, changes in national leadership. Third, the timing of policy events and response is difficult. Perhaps people respond to the announcement, or the consultation process before the announcement. Or perhaps they ignore the first implementation hoping it will not be repeated. Perhaps response is staggered, with early adopters jumping in first and later years seeing increasing participation. Although further work will be required to develop a convincing interpretation of the German publication shift in 2003 and the US HERD shift in 1989, our approach seems to have potential in that we identified shifts in the Australian and British systems that seem to connect to policy as well as the US shift in HERD due to the doubling of NIH's budget.

## References

- Australian Research Council (ARC). (2009). ERA 2010 Submission Guidelines. Retrieved on June 9, 2011 from [www.arc.gov.au/era/key\\_docs10.htm](http://www.arc.gov.au/era/key_docs10.htm)
- Brown, R.L., Durbin, J. & Evans, J.M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society, B* 37, p. 149-163.
- Butler, L. (2003). Explaining Australia's Increased Share of ISI Publications—the Effects of a Funding Formula based on Publication Counts. *Research Policy* 32 (1), 143-155.
- Butler, L. (2010). Impacts of Performance-based Research Funding Systems: A Review of the Concerns and the Evidence. In OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions*. Workshop Proceedings: OECD Publishing.
- Chow, G.C. (1960). Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28 (3), 591-605.
- European Commission (2010). *Assessing Europe's University-Based Research*. EUR 24187 EN, Science in Society 2008 Capacities, 1.4.1, European Commission, Brussels.
- Franzoni, C., Scellato, G. & Stephan, P. (2011). Changing Incentives to Publish Science, *Science*, 333 (6043), 702-703.
- Geuna, A. & Martin, B.R. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, 41 (4), 277-304.
- Harman, G. (1998). Quality Assurance Mechanisms and Their Use as Policy Instruments: Major International Approaches and the Australian Experience Since 1993. *European Journal of Education*, 33 (3), 331-348.
- Heinrich, C.J. & Marschke, G. (2010). Incentives and their Dynamics in Public Sector Performance Management Systems. *Journal of Policy Analysis and Management*, 29 (1), 183-208.
- Herbst, M. (2007). Financing Public Universities. *Higher Education Dynamics*, 18. Dordrecht: Springer.
- Hicks, D. (2012). Performance-based University Research Funding Systems. *Research Policy*, 41 (2), 251-261.
- Kettl, D.F. (2005). *The Global Public Management Revolution*. Washington: Brookings Institution Press.
- Kräkel, M. (2006). Zur Reform der Professorenbesoldung in Deutschland. *Perspektiven der Wirtschaftspolitik*, 7 (1), 105-126.
- Martin, B.R. & Whitley, R. (2010). The UK Research Assessment Exercise: A Case of Regulatory Capture? In R. Whitley, J. Glaser and L. Engwall (Eds), *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation* (pp. 51-80). Oxford: Oxford University Press.
- Miller, G.J. (2005). The Political Evolution of Principal-Agent Models. *Annual Review of Political Science*, 8, p. 203-25.
- New Zealand, Tertiary Education Commission. (2010). Performance-Based Research Fund—Tertiary Education Commission. Retrieved on June 9, 2011 from: [www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF/](http://www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF/).
- Pritchard, R. (2006). Trends in the Restructuring of German Universities, *Comparative Education Review*, 50 (1), 90.
- Rostan, M. & Vaira M. (Eds). (2011). *Questioning Excellence in Higher Education: Policies, Experiences and Challenges in National and Comparative Perspective*. Rotterdam: Sense Publishers.
- Sivertsen, G. (2009). A Bibliometric Funding Model Based on a National Research Information System. Presented at the ISSI 2009, Rio de Janeiro, Brazil. Retrieved June 9, 2011 from: [www.issi2009.org/agendas/issiprogram/activity.php?lang=en&id=108](http://www.issi2009.org/agendas/issiprogram/activity.php?lang=en&id=108)
- Stoker, G. (2006). Public Value Management: A New Narrative for Networked Governance? *The American Review of Public Administration*. 36 (1), 41-57.