May, 2008

The U.S. Research Enterprise in a Changing Global Science System

Diana M Hicks, *Georgia Institute of Technology - Main Campus*

The U.S. Research Enterprise in a Changing Global Science System

Diana Hicks

School of Public Policy

Georgia Institute of Technology

Atlanta GA, 30332

Wednesday, August 13, 2008

Abstract

Scientific communities outside Europe and the U.S. have become more competitive over recent decades due to greater investment, restructuring and increased evaluation. Most scientists probably have a vague sense of this, aware that there are many more competent scientists and interesting results coming from abroad than in the past. This paper argues that as a result, it is possible for the U.S. scientific community to perform as well as it has ever done, yet for it to appear to be in decline. This paper will explore the trajectory of the U.S. research enterprise in an international context to highlight some of the surprising consequences that strengthening of foreign R&D systems can have for trends in U.S. research output.

Introduction

The global scientific landscape has changed. Over the past decade many governments became convinced that their economic futures lay with knowledge-based economies in which research is central. Governments sought to strengthen national research, swiftly building capability and fostering a sharper competitive culture. As a result, foreign scientific communities have become more competitive. These changes are easy to underestimate because the size of the US scientific enterprise still dwarfs that of any other country (though the scientific output of the EU is now larger). Nevertheless, in aggregate these shifts are beginning to have an impact on the perceived strength of US research.

Around the world governments' science policy trajectories and goals are clear and so we can imagine the changed world we might inhabit in say, 2015. It will be a place in which science and technology advance even faster than they do at present. More diverse approaches will be brought to bear on tough problems and will accelerate their solution. At the personal level, the implications are less positive. US scientists will compete for the best students and for corporate research support not just with other US universities but with Asian and European universities. And US scientists will have to compete for limited space in top journals not just with other US scientists, but with more and better scientists from around the world. Races to patent discoveries will involve more players. Inevitably, this will reduce the perceived achievements of younger generations of US scientists, who though they will work far harder than any previous generation of scholars will somehow not command the same dominating position in world science as did their predecessors. It is quite possible that in 2015 measured scientific output from the United States will be less than that from the EU-15 countries or Asian countries as a group. This paper explores how a situation might develop in which US scientific output is smaller than that of countries we now consider to be less proficient in research than ourselves.

Measured rates of growth in scientific output

To examine growth rates in scientific publishing, we use data from NSF's *Science and Engineering Indicators 2006 (S&EI)*. S&EI appendix table 5-41 reports the number of scientific papers published by various countries including the U.S., EU-15 and Asian countries. In these counts countries are given fractional credit for papers based on how many addresses on each paper list the country. The papers are counted in the full Science and Social Science Citation Indices (S/SCI) database so changes in journal coverage will affect the trends. Table 1 reports the growth in output between 1988 and 2002 for South Korea, the EU-15, the database as a whole and the United States.

Table 1 – Growth in publications between 1988 and 2002

Entity	Growth	% share of database
South Korea	15.2	2
Asian countries[1]	2.5	18
EU-15	1.5	32
Database	1.4	100
United States	1.1	31

In comparison to the US, whose paper output as indexed in this database was flat during this period, we see between four and fifteen fold growth in the output of up and coming Asian science systems such as China, Singapore, Taiwan, and South Korea. In aggregate the Asian countries including Japan and India more than doubled their output. Of course, the strongest growth was from a low base, and the US still publishes far more papers in the journals indexed in the S/SCI, accounting for 31% of the database. Indeed, the database has always been criticized for being biased towards Anglo-Saxon researchers. However, the relatively static position of the U.S. in these numbers is a worry because it contrasts with the healthy growth of prior decades and thus may signal a transition to an actual decline in measured scientific output.

Open publication in peer-reviewed research journals traditionally characterizes scientific communication, and counting a country's journal articles is a basic indicator of a nation's scientific output. Increases in published output are routine, expected and taken as indicators of a healthy scientific research system. A declining publication count would be worrying, perhaps signaling weakness or decay in the research system, which in turn might threaten future economic growth in our science-driven, high tech economy. Could a country where almost everything measurable increases, most especially things associated with high technology, accept a decline in scientific output with equanimity? Perhaps not.[2] Certainly, the issue deserves scrutiny.

Does the database grow?

It is of course possible that U.S. scientists are indeed publishing more and more articles in scientific journals as they always have, but *S&EI* counts are somehow missing that and getting it wrong. After all, these publication counts are indicators, and indicators can go astray.

Bibliographic databases index part of the scientific literature. No representation of the scientific literature is more complete than these databases, yet there are published scientific papers that are not indexed. Thus, publication indicators can go astray if the database is not a faithful representation of the scientific literature. Of particular relevance here, if the database grows more slowly than the literature, U.S. scientific output could appear to decline when measured in the database yet could still be expanding.

It is quite likely that the scientific literature grows differently from the literature databases because different mechanisms underlie the expansion of the literature and the expansion of databases. The literature expands as journals grow and split and new journals are founded to serve new specialist interests. This reflects the continuous expansion of research as scientists become more specialized and new specialties emerge. Growth in literature may be slowed if subscriptions are harder to come by, for example if library budgets are static or journal subscription costs rise. Databases must expand or risk looking old fashioned and losing subscribers, yet expansion per se probably does not increase subscriptions, though it increases costs. Also, databases are affected by changes in company policy, for example a decision to include more foreign literature, or to add all health sciences journals (as happened in the SSCI in 1996). Perhaps most crucially, the literature grows as the result of decisions of innumerable highly motivated individuals and publishers looking for expansion. Database growth is controlled by a management skilled in assessing the costs and benefits of expansion in business terms. Thus, the literature is biased towards exuberant and uncontrolled growth, while databases are biased towards staying the same size, or if they must, growing slowly, steadily and predictably.

This disparity is not as worrying as it might seem at first because of what might be termed "the quality factor" in science. Simply put, not all literature that claims to be scientific is, and not all scientific literature is equal. In the first category would be, for example, journals on astrology or homeopathy. In the second would be house journals, locally oriented journals, and often new journals, because researchers may be wary of submitting good work to new journals with an uncertain future and limited circulation. Clearly, the boundary between the best scientific literature and the rest is subjective and shifts over time as approaches once considered obsessions of the fringe gain acceptance. However, from the policy perspective, quality counts. If U.S. output in *Science* and *Nature* declined, it would not be comforting to know that increased publishing in the "Vegetable Journal" and "Astrology Today" more than made up for it.

To generalize this principle, in bibliometrics we rely on databases to draw a line somewhere and to incorporate the best scientific literature. We hope, and in general this is the case, that the database indexes the best literature, and that at the bottom end of its quality spectrum questions may arise and coverage may change, but that overall what is missed is much less important than what is indexed. This works because the impact of scientific research is not arrayed in a normal distribution (like height or intelligence). Measured by citations, impact follows a power law distribution, meaning that a very few papers earn very high citation counts and a large number earn no citations at all. Normal distributions are well described by a mean and standard deviation; power law distributions are not. Thus, database providers work not to cover a spread about a quality mean, but rather to identify the top of the distribution which is very visible, given the nature of the distribution. And the nature of the distribution means that however far down the distribution databases draw their line, the literature excluded will be much less significant than the literature included.

The implications for assessing growth of U.S. scientific output are these. We would like to count papers in peer reviewed, internationally oriented journals because this is the yardstick by which a nation's science should be measured against other nations. However we count – in a few top journals, in a fixed set of journals that avoids database coverage changes (as was done for *S&EI* counts until the 2006 edition), or in the full database (S&EI 2006) – what we count is relatively fixed in size when compared to the full scientific literature. Nevertheless, growth in number of U.S. papers is possible using each of these three counting methods because the total number of papers in each set grows. Between the five year periods ending 1994 and 1999, the number of papers in the journal *Science* increased by 9%; the number of papers in a fixed set of journals grew by 5%;[3] and the number of papers in the *Science and Social Science Citation Index* grew by 13%. Growth rates measured in these sets of journals are probably lower than those that would be obtained from a count of the full scientific literature. But this does not mean that our indicators have gone astray. If U.S. output in the world's peer reviewed, internationally oriented journals begins to decline, that in itself may signify a problem. Our indicators may be more complex than we initially thought, but they raise policy relevant questions.

Prediction

Might we be able to predict the trend in U.S. authored publications whole counted in the full *S/SCI*? Perhaps this is possible if we can identify fundamental factors that underpin trends in the number of U.S. publications. There would seem to be two such factors. The first is growth in the *S/SCI*, which sets an upper bound on growth in number of U.S. authored publications indexed in the *S/SCI*. Between 1988 and 2002 the number of articles and reviews indexed in the *S/SCI* increased from about 466,000 to 640,000, growing in a linear fashion by about 12,500 papers per year on average or about 2.3% per year.

The second factor shaping the trend in number of U.S. papers in the *S/SCI* is the U.S. share of articles. The U.S. share of articles is the ultimate outcome of innumerable peer-review and editorial decisions allocating space in *S/SCI* indexed journals. For our purposes we can imagine a contest between U.S. and foreign authors for limited space in high quality, internationally oriented, scientific journals.[4] To perpetually increase measured output, U.S. authors must succeed in the peer review contest at the same rate they always have. Unfortunately, the American share of world scientific output, measured in the *S/SCI*, decreases every year. The decreasing share has not been a concern, because after World War II, the U.S. so dominated world scientific output that it was only natural that other nations would recover, new entrants would build scientific capability, and the U.S. share would decline.

In isolation, neither the rate of *S/SCI* expansion nor the declining U.S. share of *S/SCI* papers is cause for concern. However, when number of *S/SCI* papers is multiplied by U.S. share, we obtain number of U.S. papers. Since both *S/SCI* expansion and decline in U.S. share are linear, we can use the trends in both to predict the future growth in U.S. output. The results do provide cause for concern. Here are the equations:

S/SCI= Total *S/SCI* papers

US% = U.S. share of *S/SCI* papers

US = U.S. papers = *S/SCI*·US%

With linear growth in the *S/SCI* and linear decline in U.S. share, we have:

(1) S/SCI = n·year + a

(2) $US\% = m \cdot year + b$

Thus:

(3) $US = S/SCI \cdot US\% = (n \cdot year + a)(m \cdot year + b) = m \cdot n \cdot year^2 + year\,(n \cdot b + m \cdot a) + a \cdot b$

With the *S/SCI* expanding and U.S. share of papers declining, a parabola describes the trend in number of U.S. papers. Even with share falling, we can expect number of U.S. papers to rise for a number of years, though in ever decreasing increments until the top of the curve is reached at which point number of papers will begin to decline. From the policy viewpoint, the important question is: when will we reach this inflection point? If it comes in 300 years, it is unimportant now; but if arrives soon there is a policy problem.

To find the inflection point we take the first derivative of the equation and solve for $dUS/dt = 0$:

(4) $dUS/dyear = 2 \cdot m \cdot n \cdot year + n \cdot b + m \cdot a = 0$

(5) $year = -(n \cdot b + m \cdot a)/\ 2 \cdot m \cdot n = -½(b/m + a/n)$

Linear regression on the trend in *S/SCI* papers and U.S. share gives:

(6) $S/SCI = 12{,}756 \cdot year - 24{,}876{,}247$

(7) $US\% = -0.006 \cdot year + 12.6$

Using these parameters in equation 5, we calculate the year at which decline in U.S. output began to be 2001. However, the trend in U.S. share appears to change in the early 1990s and enter a much steeper decline. The linear regression line does not really fit the data very well, its slope being too steep at the beginning and not steep enough at the end. If things had remained as they were in the late 1980s, U.S. share would decline so slowly that number of U.S. papers would not decline until 2106–making it a non-issue. However, if there has been a change, and the regression based on years 1992-2002 is more appropriate, decline began even earlier, in 1999. Examining year-to-year changes in number of U.S. papers as reported in S&EI reveals that between 1999 and 2002 the number of U.S. papers increased by 2% between 2000 and 2001, and the other years were static or declining.[5]

Some conjectures and stylized facts

Suppose the article database were fixed in size, and furthermore suppose the U.S. R&D enterprise was the best in the world, working at maximum efficiency. And suppose that the database had a U.S. bias such that more minor U.S. journals were indexed than minor journals from other countries (we might call this the "Idaho Potato Journal" effect). In contrast, the R&D systems of other countries contain slack resources and slack R&D capacity. When the other countries decide to emphasize R&D more they quickly (because of slack capacity) improve their R&D enterprise and produce more and higher quality work. What is the effect on the US in the fixed database? Things get more competitive, the bar is raised a bit, some lower quality U.S. work can't find a place anymore and so the US share declines.

In reality the database isn't fixed in size; it is almost fixed. The database grew by about 13,000 articles per year, or on average by 2.0% between 1990 and 2002. But there is a limit to growth, Thomson Scientific's costs scale with articles indexed. To keep costs constant, they would have needed to improve productivity by 2.0% per year.

The U.S. R&D system is not perfect, but as a stylized fact it is very plausible that U.S. scientists have worked harder than anyone else for longer than anyone else and that the incentives, particularly in academia, over the long term have extracted maximum efficiency and that this contrasts with the situation in other countries which did have slack R&D capacity. For example, the UK RAE evaluation of universities was implemented and output rose. The 1990s saw aggressive expansion of government R&D support in Asia – esp. South Korea and China – as well as Latin America and Pacific countries. Substantial government expenditure on training and infrastructure is not captured in R&D statistics. Doctoral programs have expanded across Asia and in Latin America as a point of government policy. Many US-trained PhDs now return to home countries that can offer them good professional opportunities. Even in Europe and Australia, which like the US suffered a relative decline in government R&D funding, new systems of evaluation have played a significant role. Governments imposed aggressive accountability and evaluation for research into systems in which stringent evaluation was not used at any level.

A recent NSF study of publication counts concludes that the key factors driving academic publication rates are money, (research funding, especially federal); researchers (number of faculty) and students (number of PhD students). The hypothesis that an international perspective is needed to understand the decline in U.S. output is supported by quantitative data on these research resources: total U.S. R&D expenditure, numbers of scientists and engineers and numbers of graduate students and doctoral degrees awarded increased in the 1990s. This is the domestic perspective. In the international perspective, U.S. resources are expressed as a share of a group of countries.

In the domestic perspective, there is a general pattern of growth with perhaps some leveling and a bit of decline, but nothing especially worrying. Long periods of growth are interrupted in the early 1990s (except in doctoral degrees awarded which continued to grow). There was a slight decline in R&D expenditures, a leveling of growth in scientists and engineers and a drop off in graduate enrollment.[6] The most worrying decline is the drop in number of graduate students beginning in 1993 which followed four decades of increase.[7] In the domestic perspective there are hints of problems, but things really do not look too bad.

The international perspective is truly worrying. The U.S. share of G7 R&D expenditure declined sharply from 1987 to 1990; share of G7 scientists and engineers declined sharply from 1991 to 1993; and share of U.S.-Asian doctoral degrees began declining sharply in 1993. The international perspective reveals patterns very similar to the drop in U.S. share of publications—a long-term gentle decline in share is rudely interrupted by a slide down a much steeper slope. The declines in resources begin a couple of years before the declines in U.S. output, which makes perfect sense. Share of R&D expenditure and scientists and engineers has since turned up, offering hope that we are simply seeing a transition to a new regime which will soon stabilize.

Summary

In the *Economist* on September 16, 2006 it was noted that:

Last year the combined output of emerging economies reached an important milestone: it accounted for more than half of total world GDP (measured at purchasing-power parity). This means that the rich countries no longer dominate the global economy. The developing countries also have a far greater influence on the performance of the rich economies than is generally realized. Emerging economies are driving global growth and having a big impact on developed countries' inflation, interest rates, wages and profits.

This paper argues that other countries are having a big impact on the rate at which our scientific output grows as well. We seem to be entering a new era in science policy. The U.S. has long accepted that its share of world scientific output will decline as scientific communities in other countries strengthen. This process seems to have accelerated sharply, as other governments have become convinced that their economic futures lie with knowledge-based economies in which research plays a central role. Foreign scientific communities have become much more competitive.

The institutions of modern science have in many ways been a gift from the United States to the rest of the world. The U.S. has demonstrated that the best quality scientific research is fostered when funding is awarded competitively, plentiful rigorously trained PhD students and postdocs are available cheaply, substantial amounts of money are spent, modern equipment is used, and transfer of research to technological application is encouraged. In many ways, other countries have sought over the past decade to incorporate more of these elements into their systems. Furthermore, the U.S. has probably trained or at some point employed many of the scientists now doing so well back home. As a result, American universities no longer stand alone at the scientific frontier.

This paper has explored the trajectory of the U.S. research enterprise in an international context to highlight some of the surprising consequences that strengthening of foreign R&D systems can have for trends in U.S. research output.

---

[1] China, India, Japan, South Korea, Singapore, Taiwan

[2] The worrying trend has not gone unnoticed - "Less Ink for U.S. Economists, Journals print more foreign work", *Business Week*, July 30, 2001, p. 26.

[3] The fixed set comprising in this case all journals indexed in the S/SCI in 1985, taking into account journal administrative changes.

[4] Presumably in the top journals, article space is the scarce quantity and at the bottom of the journal quality spectrum, good papers are scarcer than publication places.

[5] The year 2003 is excluded from this analysis because in that one year the database grew by more than it had in the previous six years combined, and by more than in any other single year in the 15 years recorded in S&EI 2006. Between 2002 and 2003 the database increased in size by an unprecedented 9%, and so measured U.S. output increased by 8%. Perhaps the worries about declining U.S. output convinced Thomson-Scientific to solve the

problem. Unfortunately, analysis of the 2003 changes in database coverage is not possible using the information released in S&EI 2006.

[6] Note that development expenditures and personnel are included, and as military expenditure declined during this period substantially, an unknown proportion of the drop must be due to military development which would not affect the trend in paper output much.

[7] Historically, growth in graduate enrolment has generally echoed shifting patterns of Federal R&D support, with an influx of foreign students complicating matters. Some of this decline may trace to the favourable U.S. job market after 1992. The decline in graduate students may well be connected to the decline in university output, though the decline in graduate enrolment seems to affect the biological sciences as well as the physical sciences, which does not echo the pattern in paper output.

References

National Science Board. 2006. *Science and Engineering Indicators 2006.* Two volumes. Arlington, VA: National Science Foundation (volume 1, NSB 06-01; volume 2, NSB 06-01A).