2 US RESEARCH INSTITUTIONS

In collaboration with researchers from the Andrew Young School of Policy Studies at the Georgia State University (Atlanta, GA), American addresses in the header of publications indexed in the Web Of Sciences from 1995 onwards have been cleaned and assigned to about 7,600 US institutions. The same procedure as described in the second chapter of the first section and developed for the cleaning of Brazilian and European addresses is used for this project. More than 15 million addresses were processed to yield a list of about 375,000 unique words in the institution field. 19,282 of these words/names occurred at least 20 times or more in the total list of addresses. 15,504 of them were assigned to 7,604 US institutions. Using this procedure about 94.3% of all publications with at least one American address could be assigned to at least one American institution.

Finally, a set of 5,998 institutions publishing between 2001 and 2003 was identified and classified into the 8 different groups. Table 46 presents the number of institutions in each group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Institutions</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO</td>
<td>495</td>
<td>8.3%</td>
</tr>
<tr>
<td>AGR</td>
<td>248</td>
<td>4.1%</td>
</tr>
<tr>
<td>MDS</td>
<td>2167</td>
<td>36.1%</td>
</tr>
<tr>
<td>GSS</td>
<td>205</td>
<td>3.4%</td>
</tr>
<tr>
<td>TNS</td>
<td>558</td>
<td>9.3%</td>
</tr>
<tr>
<td>CHE</td>
<td>270</td>
<td>4.5%</td>
</tr>
<tr>
<td>GRM</td>
<td>569</td>
<td>9.5%</td>
</tr>
<tr>
<td>SPM</td>
<td>1485</td>
<td>24.8%</td>
</tr>
</tbody>
</table>

Table 46. Number of US institutions per group

Next to the replication analysis described earlier two other applications are scheduled with this set of American institutions. First a comparison between the US and the European Union will be done. For the second application bibliometric data will be enriched with data on enrolment and staff and on R&D expenditures. These two applications are in line with the two main applications of bibliometric tools and methodologies as mentioned in the first part of this thesis. First, the main focus will be at measuring and describing and comparing the research performance of institutions at both sides of the Atlantic. The latter application will help us with the understanding of the influence of factors like staff or expenditure on the structure and creation of science.
2.1 Comparing US with Europe

With this first application we want to extend the comparisons between the US and Europe to the institutional level. For instance, the Third European Report on Science & Technology Indicators (ERSTI/REIST, 2003) presents such comparisons between European countries and the US, however, based on a large set of bibliometric or technometric indicators at the national level. None of these indicators are used for a comparison of institutional performance.

The unique combination of cleaned datasets of both European and US research institutes creates a vast range of opportunities for bibliometric research. First, the research landscapes can be compared although we assume that the “European” landscape is still somewhat heterogeneous. Concerning the distribution of institutions over the eight different groups, we expect a bit larger share of multidisciplinary institutions larger in the US than in Europe. It can, of course, be tested whether this difference is large enough to be statistically significant.

In a second step, the average performance within a group can be compared between the US and the EU. Here the wide range of indicators previously already used in this dissertation will be applied.

Last, within the individual groups, institutions can be merged over both continents. This will reveal the relative position of institutions in the joint US-EU dataset.

2.2 Enriched data

The bibliometric data on the institutions can be enriched by two additional sources provided by researchers at Andrew Young School of Policy Studies. First there is the enrolment and faculty data from the Integrated Postsecondary Education Data System (IPEDS, see http://nces.ed.gov/ipeds) available through the National Center for Education Statistics. The second source is the WebCASPAR data on academic institution research and development expenditures derived from the National Science Foundation’s Integrated Science and Engineering Resources Data. Main research questions here will focus on the influence of institutional characteristics like number of students, number of staff, level of education offered, research expenditures on the classification of the institution into our system and on research performance.

2.2.1 IPEDS

IPEDS is a system of interrelated annual surveys of the US Department’s National Center for Education Statistics (NCES, see http://nces.ed.gov/). Since the Higher Education Act of 1965 each institution that participates in a federal student aid program is required to report data on various issues like enrolments, graduation rates, faculty and staff. According to the IPEDS website more than 6700 institutions complete the surveys annually.
In our research we are only interested in those institutions that are active both in education and as research institutions which is a subset of both sets of institutions.

Relevant data in the IPEDS:

• Fall Enrolment data with total number of students and full time equivalent. Also the share of female students is available.

• Fall Staff Data: This survey provides data on full- and part-time staff by primary occupational activity.

2.2.2 Carnegie classification

The IPEDS data also contains the Carnegie classification. This classification system developed by the Carnegie Foundation in the 1970’s is based on academic mission without any indication of quality (Carnegie Commission, 1973 or Carnegie Foundation, 2001). In IPEDS the 1994 classification codes are recorded. This system of 10 categories classifies institutions based on their highest degree offered, on the number or share of other degrees and in some cases on selectivity of the institution’s admission. The list of categories is included in the appendix.

2.2.3 WebCASPAR

The institutional research and development expenditure data are derived from the National Science Foundation’s Integrated Science and Engineering Resources Data System (see http://webcaspar.nsf.gov). The expenditures are given in thousands of constant dollars with the base being year 2000 and using the ‘Fiscal year GDP implicit price deflator’ available within the webcaspar environment.

2.2.4 Research questions

This multidimensional data (different institutions, years, indicators) allow us to answer several questions. If we consider scientific output - as all other products – being the result of input of labor and capital we can investigate the relation between these input variables and the output at institutional level in the US. Several authors already described a paradox in the relation between research investment and publication output. Adams and Griliches (1996) were among the firsts to describe a diminishing return in terms of papers produced per R&D dollar. They conducted an analysis at the institutional level and identified the effect of the rising cost of achieving new results within specific scientific fields and of the rising competition caused by the expanding overall size of the scientific enterprise. For citations they showed a brighter picture although they question this effect by pointing out the growing number of journals included in the database and the changing citation practices. Adams and Griliches used merely pure citation counts instead of relative citation indicators that could control for the two mentioned factors.
Shelton (2008) studied at the national level the decline of the American share of publications despite its leading investments in R&D. By using regression analysis he tried to find the causes for this decline and to determine the effect of changes in input factors on changes in output factors. He concluded that the decline in US publication share is in line with the declining share of investment in science of the US which is due to the huge increase of world R&D investment mainly by Asian countries. The observation of the decline of US publication share despite the increase in absolute numbers of the US investment was called the American Paradox by Shelton referring to the European Paradox first described in the European Commission’s “Green paper on Innovation”. This European Paradox relates the strong European scientific performance to a weaker R&D outcome as it expresses “Europe’s inferiority in terms of transforming the results of technological research and skills into innovations and competitive advantages” (EC, 1995).

The research questions I would like to address are:

1. Does this American Paradox as described by Shelton also hold for individual US institutions when looking at relative citation indicators.
2. Is there a difference in the relation between input and output indicators across the eight groups of institutions.

This project thus combines the approaches by Adams and Giliches and by Shelton with our own institutional classification and relative citation indicators in order to obtain a better understanding of the role of changing input variables on production and impact of scientific papers. By including also the collaboration with EU institutions we can also control for the growing internationalization in several fields. Regression analysis will be used as statistical tool to enable us to answer these questions.
References


Appendix

These are the included categories as described in the IPEDS data dictionary:

11 - RESEARCH UNIVERSITIES I: These institutions offer a full range of baccalaureate programs, are committed to graduate education through the doctorate, and give high priority to research. They award 50 or more doctoral degrees each year. In addition, they receive annually $40 million or more in federal support.

12 - RESEARCH UNIVERSITIES II: These institutions offer a full range of baccalaureate programs, are committed to graduate education through the doctorate, and give high priority to research. They award 50 or more doctoral degrees each year. In addition, they receive annually between $15.5 million and $40 million in federal support.

13 - DOCTORAL UNIVERSITIES I: These institutions offer a full range of baccalaureate programs and are committed to graduate education through the doctorate. They award at least 40 doctoral degrees annually in five or more disciplines.

14 - DOCTORAL UNIVERSITIES II: These institutions offer a full range of baccalaureate programs and are committed to graduate education through the doctorate. They award annually at least 10 doctoral degrees (in three or more disciplines), or 20 or more doctoral degrees in one or more disciplines.

21 - MASTER’S (COMPREHENSIVE) UNIVERSITIES AND COLLEGES I: These institutions offer a full range of baccalaureate programs and are committed to graduate education through the master’s degree. They award 40 or more master’s degrees annually in three or more disciplines.

22 - MASTER’S (COMPREHENSIVE) UNIVERSITIES AND COLLEGES II: These institutions offer a full range of baccalaureate programs and are committed to graduate education through the master’s degree. They award 20 or more master’s degrees annually in one or more disciplines.

31 - BACCALAUREATE (LIBERAL ARTS) COLLEGES I: These institutions are primarily undergraduate colleges with major emphasis on baccalaureate degree programs. They award 40 percent or more of their baccalaureate degrees in liberal arts fields and are restrictive in admissions.

32 - BACCALAUREATE COLLEGES II: These institutions are primarily undergraduate colleges with major emphasis on baccalaureate degree programs. They award less than 40 percent of their baccalaureate degrees in liberal arts fields or are less restrictive in admissions.

40 - ASSOCIATE OF ARTS COLLEGES: These institutions offer associate of arts certificate or degree programs and, with few exceptions, offer no baccalaureate degrees.
SPECIALIZED INSTITUTIONS: These institutions offer degrees ranging from the baccalaureate to the doctorate. At least 50 percent of the degrees awarded by these institutions are in a single discipline.

51 - Theological seminaries, Bible colleges and other institutions offering degrees in religion: This category includes institutions at which the primary purpose is to offer religious instruction or train members of the clergy.

52 - Medical schools and medical centers: These 23 institutions award most of their professional degrees in medicine. In some instances, their programs include other health professional schools, such as dentistry, pharmacy, or nursing.

53 - Other separate health profession schools: Institutions in this category award most of their degrees in such fields as chiropractic, nursing, pharmacy, or podiatry.

54 - Schools of engineering and technology: The institutions in this category award at least a bachelor’s degree in programs limited almost exclusively to technical fields of study.

55 - Schools of business and management: The schools in this category award most of their bachelor’s or graduate degrees in business or business-related programs.

56 - Schools of art, music, and design: Institutions in this category award most of their bachelor’s or graduate degrees in art, music, design, architecture, or some combination of such fields.

57 - Schools of law: The schools in this category award most of their degrees in law. The list includes only institutions that are separate entities or campuses.

58 - Teachers colleges: Institutions in this category award most of their bachelor’s or graduate degrees in education or education-related fields.

59 - Other specialized institutions: Institutions in this category include graduate centers, maritime academies, military institutes, and institutions that do not fit any other classification category.

60 - Tribal colleges: These colleges are, with few exceptions, tribally controlled and located on reservations. They are all members of the American Indian Higher Education Consortium.