The Journal of Academic Librarianship 42 (2016) 300–312

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

ARL Libraries and Research: Correlates of Grant Funding

Ryan P. Womack

Rutgers University–New Brunswick, New Jersey, USA

ARTICLE INFO

Article history:
Received 21 January 2016
Received in revised form 19 May 2016
Accepted 13 June 2016
Available online 24 June 2016

Keywords:
Academic libraries
Research
Universities

ABSTRACT

While providing the resources and tools that make advanced research possible is a primary mission of academic libraries at large research universities, many other elements also contribute to the success of the research enterprise, such as institutional funding, staffing, labs, and equipment. This study focuses on members of the Association for Research Libraries (ARL) in the United States. Research success is measured by the total grant funding received by the University, creating an ordered set of categories. Combining data from the NSF's National Center for Science and Engineering Statistics, ARL Statistics, and IPEDS, the primary explanatory factors for research success are examined. Using linear regression, logistic regression, and the cumulative logit model, the best-fitting models generated by ARL data, NSF data, and the combined data set for both nominal and per capita funding are compared. These models produce the most relevant explanatory variables for research funding, which do not include library-related variables in most cases.

BACKGROUND AND LITERATURE REVIEW

Academic libraries are under increasing pressure to demonstrate their relevance to the scholarly enterprise via concrete metrics. The literature of professional librarianship is replete with discussions the importance of libraries, but thorough quantitative studies are somewhat rarer. Several quantitative approaches to evaluating the impact of academic libraries have been used, as discussed in the literature review below.

Some studies demonstrate the importance of the library to student outcomes. Whitmire (2002) found that gains in critical thinking skills among undergraduates, as measured by the College Student Experiences Questionnaire, were linked to library measures taken from the Integrated Postsecondary Education Data System (IPEDS). Mezick (2007) also used IPEDS data along with data from the Association of Research Libraries (ARL) and the Association of College and Research Libraries (ACRL) to show a correlation between library expenditures and professional staff and student retention. Researchers at the University of Minnesota (Soria, Fransen, & Nackerud, 2013) used detailed student records to demonstrate a positive relationship between academic performance and library use.

A second approach has been to look for the impact of library resources on faculty publications. For example, Budd (2006) studies faculty productivity and uses rank-order correlations to show a moderate association between the quantity of faculty publishing at ACRL institutions and library expenditure and volumes held. The number of PhD's awarded also shows similar levels of correlation. Surveys of faculty attitudes towards academic libraries, such as Mikitish and Radford (2013), are another way to establish value.

Hendrix (2010) used principal components analysis to study the relationship between faculty citations and library variables from the ARL Statistics. While strong associations were present in the initial dataset, no associations with faculty citations were found when using size-independent measures of library activity. In an earlier article, Hendrix (2008) also conducted a bibliometric study on medical schools using principal components methodology.

Another way of addressing the pressure to demonstrate the continuing relevance of libraries is to adopt business paradigms, such as return on investment (ROI) (Coyle, 2006). In contrast to a business environment with clearly defined profit and loss, the inputs and outputs in the library context are harder to pin down, and are imperfectly addressed by existing data sources. Tenopir (2010) described the evalution of ROI by working with administrators to understand their attitudes towards library support and its impact on grant funding. At several institutions, the article citations used in grant proposals were studied and combined with qualitative information from surveys of faculty submitting grant proposals which testified to the value of the library.

Turning to studies that use larger data sets and more extensive quantitative methods, Allen and Dickie (2007) built a regression model that relates library expenditure as the response variable to various institutional measures such as the size of programs, enrollments, and faculty.

Weiner (2009) built a dataset that combined IPEDS, ARL, and *US News and World Report* peer assessment scores, along with several other sources to determine factors influencing institutional reputation.

E-mail address: rwomack@rutgers.edu.

http://dx.doi.org/10.1016/j.acalib.2016.06.006
0099-1333/

She then used stepwise linear regression to build explanatory models. Library expenditure was influential in all models, and grants and instructional expenditures were also influential.

GOALS OF THIS STUDY

A major characteristic and limitation of most of the studies mentioned above (with the notable exception of Weiner) is that they use only library data to explain the outcome of interest. But in the context of a university, a well-performing library may be correlated with many other factors that more directly influence student success, since the best libraries are typically at the best schools with the best funding, best support services, best faculty, and so on. Allen and Dickie's work shows how library funding can be predicted from these factors. Working only with library variables to demonstrate the library's relevance does not allow for alternative explanations and is a weak form of proof.

However, Weiner's study did include other institutional reputational factors in order to select a model that combined variables from different spheres. The present study takes a similar approach to modeling both library and other academic factors, but with a wider range of statistical methods and a larger selection of variables. This will provide one method of determining whether library characteristics are the primary explanatory factors for the outcome, or whether they are only secondary factors that have some explanatory power due to their correlation with other primary factors.

Our primary response variable will be research productivity, as measured by grant funding. Grant funding for research is a central characteristic for the reputation and identity of major research universities. We will look at a representative group of research universities and assess whether library or other academic and institutional characteristics are related to grant funding. A secondary dimension of interest is the effect of fitting linear regression, logistic regression for binary outcomes, and cumulative logit models for multi-category ordered outcomes. Logistic and cumulative logit methods can help explain data that is categorical in nature, rather than continuous, and may provide a better fit than linear regression in many settings. By comparing different fitted models, we will begin to understand the variables that are most closely related to research funding. Most importantly, our model selection process will select the best explanatory variables from among all candidate variables. Whether or not the final fitted model includes library-related variables will be a strong indicator of library relevance to the university's research productivity.

METHODS

DATA COLLECTION

The Association of Research Libraries (ARL) is the leading grouping of large research libraries in North America. The ARL Statistics have been collected annually since 1908 (Association of Research Libraries, 2012). In 2012, there were 125 members, 17 of which were in Canada. The 108 members located in the United States consist of 99 university research libraries and 9 institutional libraries (e.g., the New York Public Library, National Library of Medicine, Library of Congress, and so on). This study uses data only from the 99 US university libraries, who compete for research funding under similar conditions. The Canadian research funding environment is not directly comparable.

Although the ARL membership contains most of the largest universities from a research funding standpoint, there are notable exceptions. Institutions that receive large research grants such as Stanford, the California Institute of Technology, Carnegie Mellon, and others are not ARL members. Other institutions in the ARL are ranked below the top 200 in research funding, such as Howard University (#208 in 2012) or Kent State University (#248), far below many non-ARL members. However, the ARL has the longest-running and most complete collection of library statistics, and this data sample has the most potential for detailed comparisons over time. With the exceptions noted, it remains a very representative grouping of the most active research universities. Data from the year 2012 was used for comparability with the most recently available data at the time of the study, collected from the other sources described below. Definitions of variables and data collected are as provided by ARL.

The National Center for Science and Engineering Statistics (NCSES) of the National Science Foundation (NSF) is the most systematic collection of data on research funding and inputs to research in the United States. The Higher Education Research and Development (HERD) survey, which is the most systematic collection of data on research funding and inputs to research in the United States (National Science Foundation. National Center for Science and Engineering Statistics, 2014). The HERD reports annually on levels of research funding from all sources: federal, state, local, nonprofit, business, internal institutional funding, and other sources. For the purposes of this study, the total research funding received in 2012 was the primary response variable of interest, although federal funding is the largest share of funding and closely tracks the total.

The NCSES Survey on Science and Engineering Facilities (National Science Foundation. National Center for Science and Engineering Statistics, 2013) reports on the total amount of existing square footage of research space, as well as newly constructed space in the last year, dedicated to science and engineering research at universities in the US, in laboratories, animal research facilities, computer labs, equipment rooms, and other such facilities. The latest available data, at the time of the study, from the fiscal year 2011 was used. Data is collected every two years, so there is no direct equivalent to 2012. Since these variables function as a likely input to future grant receipts, using the earlier year is reasonable. Planned construction and repair and renovation costs were excluded from the dataset since they are not likely to be directly related to grant success.

Finally, in order to add other measures of staffing and salary expenses in non-library categories, along with additional institutional characteristics, data for 2012 was extracted from the Integrated Postsecondary Education Data System (IPEDS) of the National Center for Education Statistics (National Center for Education Statistics, 2014). The IPEDS data reports the number of employees, the total salary expenses, and the number of Full-Time Equivalent (FTE) employees in several categories. All of these ways of measuring employment are included in the dataset.

Since medical research is a large component of overall research dollars, data from the Association of Academic Health Sciences Libraries was also considered (Association of Academic Health Sciences Libraries, 2014). However, the overall magnitude of medical library expenditures and staffing is not large compared to their general academic library counterparts. For example, at the University of Michigan, collections spending is $2 million in the medical library versus $24 million in the main library, and professional FTE employment is 15 versus 212. The medical data also has many missing values and introduces questions of comparability that would require investigation of each institution's library and institutional configuration of its medical research with regards to the rest of the campus. The IPEDS data contains indicator variables for medical degree-granting and presence of a hospital, so these can serve as a proxy for any distinctive medical effects. Based on these considerations, the AAHSL data was not included in the present study.

The ARL, NCSES, and IPEDS data described above were merged into a single dataset for the 99 US ARL institutions under study. At this stage there were 75 possible predictor variables representing inputs from library, research, infrastructure, and general staffing characteristics of the institutions. Details of the data cleaning process are described in the Appendix 0.

The data files used in this study, along with the R code used to conduct the analysis, are available from openICPSR at http://doi.org/10.3886/E45486V1. All statistical analysis was run with open source R software [available from http://r-project.org]. The R code provides more detail on the steps used in the modeling process described below. The abbreviated variable names used to report results in the paper correspond