SCIENCE & TECHNOLOGY POLICY INSTITUTE

Increasing Competitiveness of Investigators and Concentration Modeling: Quantitative Approaches

Thomas W. Jones
Brian Q. Rieksts
Brian L. Zuckerman

October 2014

Approved for public release;
distribution is unlimited.

IDA Document NS D-5320

Log: H 14-001123

IDA SCIENCE & TECHNOLOGY
POLICY INSTITUTE
1899 Pennsylvania Ave., Suite 520
Washington, DC 20006-3602

IDA | SCIENCE AND TECHNOLOGY POLICY INSTITUTE

Increasing Competitiveness of Investigators and Concentration Modeling: Quantitative Methods

American Evaluation Association Session 616
October 17, 2014

Thomas Jones

Brian Rieksts

Brian Zuckerman

Objective:

Perform an in-depth life-of-program assessment for the National Science Foundation (NSF) of its Experimental Program to Stimulate Competitive Research (EPSCoR) activities and these activities' outputs and outcomes

Fundamental Study Questions

1. What are the goals of the program as legislatively mandated?
2. What progress has there been in research competitiveness within EPSCoR jurisdictions over the period of their participation in the EPSCoR program? What evidence is there that this progress is attributable to EPSCoR?
3. How have EPSCoR funds been used to increase competitiveness of research universities, and what have been the outputs and outcomes of theses activities over time?
4. What are the program's eligibility criteria, how have they changed over time, and how have changes in eligibility criteria affected the attainment of NSF EPSCoR programmatic goals?
5. What role has been played by the EPSCoR jurisdictions' State Committees with respect to the EPSCoR program itself and to the relationships with State governments, the private sector, and universities in the jurisdiction?

Fundamental Study Questions

1. What are the goals of the program as legislatively mandated?
2. What progress has there been in research competitiveness within EPSCoR jurisdictions over the period of their participation in the EPSCoR program? What evidence is there that this progress is attributable to EPSCoR?
3. How have EPSCoR funds been used to increase competitiveness of research universities, and what have been the outputs and outcomes of theses activities over time?
4. What are the program's eligibility criteria, how have they changed over time, and how have changes in eligibility criteria affected the attainment of NSF EPSCoR programmatic goals?
5. What role has been played by the EPSCoR jurisdictions' State Committees with respect to the EPSCoR program itself and to the relationships with State governments, the private sector, and universities in the jurisdiction?

Key Language in EPSCoR Authorizing Statutes

- NSF Authorization Act of 1988, Section 113:

  “Assist those States that…historically have *received relatively little* Federal research and development funding” [emphasis added]

- NSF Organic Act:

  “Avoid *undue concentration*” of research and education [emphasis added]

- COMPETES Act of 2010:

  “National Science Foundation funding remains *highly concentrated”* [emphasis added]

Ambiguities in EPSCoR Legislation

- “Relatively little”
  - Relative to what?
  - How little is little?
- “Highly concentrated” and “undue concentration”
  - How concentrated?
  - What is undue?

Quantitative Analyses

1. Time-series regression modeling

   Has there been progress? What are the mechanisms?

2. Per-investigator analyses

   Have investigator-level mechanisms been effective?

3. Concentration analysis

   How does concentration of NSF funding compare to other agencies' funding?

Agenda

- Previous Research on the EPSCoR Program
- NSF Awards Database
- Time-Series Regression Model
- Per-Investigator Analyses
- Concentration Analysis

Agenda

- **Previous Research on the EPSCoR Program**
- NSF Awards Database
- Time-Series Regression Model
- Per-Investigator Analyses
- Concentration Analysis

Julia Melkers and Yonghong Wu 2009

- Compared EPSCoR to non-EPSCoR Jurisdictions
- Model-free analyses
- “Effect” =
  - Percentage change in Federal science and engineering (S&E) obligations for research and development (R&D) to academic Institutions
  - Researcher self-reported satisfaction
  - Size of researchers’ networks

Yonghon Wu 2009

- Compared EPSCoR to non-EPSCoR jurisdictions *and* changes within EPSCoR jurisdictions over time
- Regression model
  - Indicator for EPSCoR status
- “Effect” =
  - Per-capita *State* S&E funding for R&D from State budgets to academic institutions

Yonghong Wu 2010

- Compared EPSCoR to non-EPSCoR jurisdictions *and* changes within EPSCoR jurisdictions over time
- Regression model
  - Indicator for EPSCoR status
  - Indicator for years in EPSCoR
- “Effect” =
  - States’ shares of Federal S&E support for R&D to academic institutions

“Effect” Variables from Wu et al.

- *Percentage Change* in Federal S&E obligations for R&D
- *Proportion* of Federal S&E obligations for R&D
- Researchers’ satisfaction
- Researchers’ networks
- Per-capita *State* S&E funding for R&D at universities

National Academy of Sciences 2013

- Initial look at EPSCoR jurisdictions' performance at per-investigator level
- Considered change in number of proposals, success rates at jurisdiction level in aggregate
- Did not analyze on per-investigator basis
- Did not compare with non-EPSCoR jurisdictions

Agenda

- Previous Research on the EPSCoR Program
- **NSF Awards Database**
- Time-Series Regression Model
- Per-Investigator Analyses
- Concentration Analysis

Why Awards-Level Data?

- EPSCoR-related activities are numerous and heterogeneous
- Correlation is insufficient! Need attribution mechanisms
- Single, self-contained data source that is internally consistent

What Data Is Readily Available in the Awards Database?

- ~300,000 NSF awards for FY 1980–2010
- Highly granular:
  - Investigator-level, institutional-level
  - NSF program and activity codes

What Data Is More Challenging?

- Identifies unique and relevant education institutions
- Per award, not per fiscal year
- Principal investigators (PIs) are named, but not uniquely identified

Address Data Challenges: Awards

- Create flags for special programs:

  REU
  RET
  LSAMP
  AGEP
  I-3
  S-STEM
  ADVANCE
  IGERT
  MRI
  CREST
  HBCI-RISE

- Remove travel awards and non-research contracts

Address Data Challenges: Institutions

- Hand-curated dictionary to disambiguate institution names & merge with other sources
- Re-distribute administrative funding from system offices to schools within the system
- Assign jurisdictions to satellite campuses
- Identify HBCUs and Tribal Colleges
- Drop non-educational institutions and non-baccalaureate institutions

Address Data Challenges: PIs

- MUCH harder problem

- Email addresses for PIs, not Co-PIs

- People move, names change

- Some people have common names

Agenda

- Previous Research on the EPSCoR Program
- NSF Awards Database
- **Time-Series Regression Model**
- Per-Investigator Analyses
- Concentration Analysis

Competitiveness for NSF Funding: Time-Series Analyses

Structure:

(NSF Funding) = f(policy variables,
control variables)

NSF Funding = Annual Percentage Change of Non-EPSCoR NSF Funding

1. Directly modeling the proportion of NSF funding to jurisdictions is problematic
   - Limits other analyses
   - Constraint makes model more complicated
2. Directly modeling dollars may not make statistical sense
3. Percentage change solves both problems

Policy-Relevant Explanatory Variables

1. Percentage change of the *number* of non EPSCoR Awards
2. Tenure in EPSCoR
3. Change in *number* of EPSCoR co-funded awards
4. Percentage change in funding for *large awards*
5. Cohort fixed effects (or flag for non-EPSCoR)

Control Variables

- % change in non-NSF Federal S&E obligations to R&D (NCSES)
- Flagged recessions from 1980 to 2009
- Flagged Idaho 1998 (one HUGE award)

Final Time-Series Model

(Percentage change in non-EPSCoR $) =

f (Percentage change in number of non-EPSCoR awards

Tenure in EPSCoR

Change in number of co-funded awards

Percentage change in large award funding

EPSCoR cohort (or non-EPSCoR flag)

Percentage change in other Federal funding

Recessions

Idaho in 1998)

Attribution: Model Counterfactual

- “EPSCoR effect” =
  (Model fitted values) – (Fitted values without EPSCoR variables)
- Average EPSCoR effect by cohort
- Causality is still tricky

Agenda

- Previous Research on the EPSCoR Program
- NSF Awards Database
- Time-Series Regression Model
- **Per-Investigator Analyses**
- Concentration Analysis

Two Per-Investigator Analyses

1. Awards to EPSCoR-funded faculty

   Source: NSF Awards database

2. Comparison of award success rates

   Sources: NSF Awards database, NSF-provided data on award proposal and success rates

Per-Investigator Analysis 1: Rationale

- Jurisdictions get awards because researchers apply for them!
- EPSCoR has two individual-PI mechanisms
  1. Fund direct hires of faculty
  2. Provide co-funding for awards

Per-Investigator Analysis 1: Identifying Hired Faculty

- EPSCoR-provided list of hired faculty per jurisdiction
- Subset of faculty was identified in the NSF Awards database
- STPI tracked their funding histories
  - Did they have sustained ability to get awards or only one-off funding from EPSCoR?

Per-Investigator Analysis 1: Identifying Co-Funded Faculty

- Start with PIs on Co-funded awards
- Target PIs: First award is EPSCoR co-funded
- STPI tracked their funding histories
  - Did they have sustained ability to get awards or only one-off funding from EPSCoR?

Per-Investigator Analysis 2: Rationale

- From the logic model, three primary mechanisms by which EPSCoR jurisdictions can increase their relative share of funding were considered:
- Increase number of proposals
- Increase success rate of proposals; and
- Increase relative size of successful awards

Per-Investigator Analysis 2: Data

- Data on award size from NSF awards database
- Data on number of proposals made, success rate provided by NSF
  - Per jurisdiction level
  - Since 1990
- Needed to transform data from per jurisdiction to per investigator to allow for comparisons across jurisdictions and years
  - Normalized based on National Science Foundation, Science and Engineering Indicators 2012, Table 8-48; Appendix Table 5-14

Per Investigator Analysis 2: Execution

- Comparison: 1980–1992 cohorts versus 2000+ cohorts
- All indicators as percentage of average non-EPSCoR levels
- Indicators:
  - Average award size (Awards database)
  - Average number of awards per-investigator (Awards database)
  - Proposal rate (NSF provided)
  - Proposal success rate (NSF provided)

Agenda

- Previous Research on the EPSCoR Program
- NSF Awards Database
- Time-Series Regression Model
- Per-Investigator Analyses
- **Concentration Analysis**

Measures of Concentration Background

- Economists have studied the concentration of a resource across members of a group
  - The resource is less concentrated if it is more evenly distributed across members in the group
  - Concentration is often measured for market shares across firms or the income distribution across populations
- Herfindahl-Hirshman Index (HHI) is used to measure market shares across firms
  - HHI places more weight on larger members of the group
  - 10,000 is the maximum value when one member of the group has all of the resources and approaches zero for a perfectly even distribution
- Gini coefficient is used to measure income distribution
  - Gini coefficient places less weight on larger members of the group than HHI
  - Ranges from 0 (least concentrated) to 1 (most concentrated)

Relevant Literature

- Some researchers have used the Gini coefficient to evaluate the concentration of research publications across universities
  - Halffman and Leydesdorff (2010) predominately found university rankings globally and within nations were becoming more homogenous
  - Ville et al. (2006) found measures of research output were becoming less concentrated across Australian universities
  - López-Illescas et al. (2011) examined the concentrations within Spanish universities across disciplines
  - Xie (2014) found research across universities was becoming more concentrated due to the increase in the number of universities participating in research

Measures of concentration in research capacity could be evaluated across EPSCoR jurisdictions

Computing Research Concentration

- STPI selected the Gini coefficient as the most appropriate measure of concentration to use across EPSCoR jurisdictions
  - The Gini coefficient does not place additional weight on larger jurisdictions
  - The HHI and other metrics of concentration could also be computed to determine whether results are robust
- Several different indicators of research capacity could be selected
  - E.g., Federal R&D funding, NSF Funding, Publications, Patents
  - Indicators of research capacity could also be normalized across jurisdictions by population

Methodology

- Concentration of research capacity can be measured across EPSCoR jurisdictions over the history of the program
  - A correlation between a decrease in concentration and the existence of EPSCoR program indicates a possible positive effect from EPSCoR
  - Such a correlation does not necessarily imply that the decrease in concentration was caused by EPSCoR
  - Other factors could have led to a decrease in concentration
- Concentration could be measured among only non-EPSCoR jurisdictions to see whether factors exogenous to EPSCoR are increasing or decreasing concentration

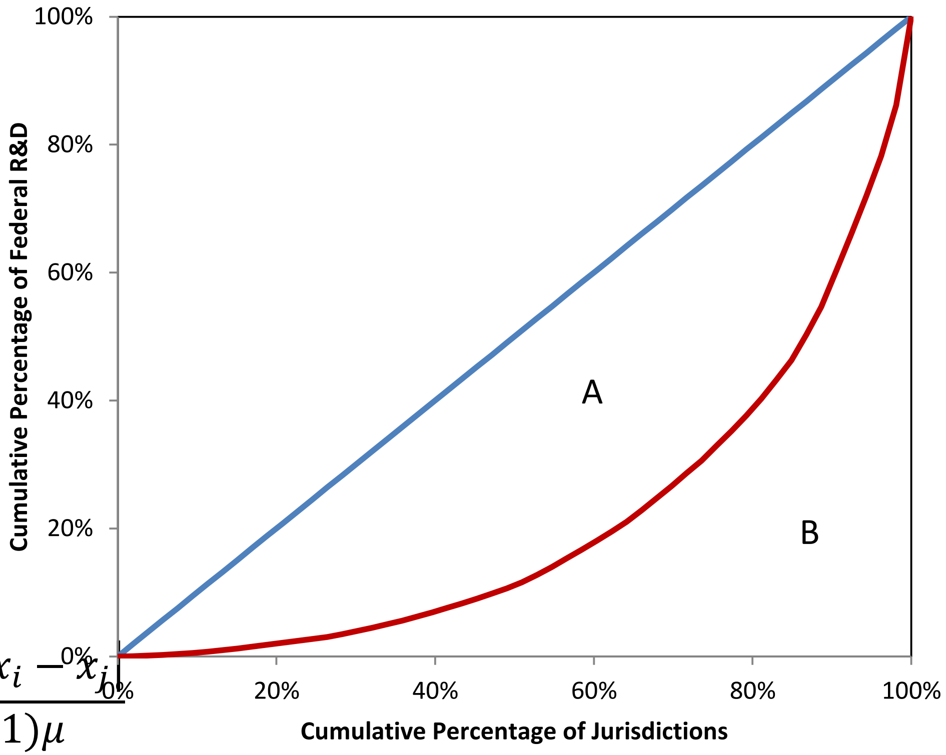
Definition of Measures of Concentration

- Let x_i be the proportion of the resource of interest for group member *i*
- Herfindahl Hirshman Index (HHI):

$$\sum_i {x_i}^2$$

- Gini coefficient:
  - Based on the Lorenz curve (right)
  - The area of region A divided by the area of regions A and B
  - Sample Gini coefficient are multiplied by n/(n-1) to become unbiased estimators of the population coefficients
  - This normalized formula is

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|}{2n(n-1)\mu}$$



Other Concentration Analyses

- Measures of research concentration could also be calculated across various Federal agencies
  - The concentration of research funding for an agency could be benchmarked against other agencies
  - Agencies with low levels of concentration in research capacity could be studied for lessons learned
- Concentration can also be used as a metric to evaluate the effect of policy changes
  - E.g., formula funding

Quantitative Analyses Performed

1. Time series regression modeling

   Has there been progress? What are the mechanisms?

2. Per-Investigator Analyses

   Have investigator-level mechanisms been effective?

3. Concentration analysis

   How does concentration of NSF funding compare to other agencies' funding?

BACKUP

Use of Database: Conduct Descriptive Analyses Based on EPSCoR Outputs

- Approach
  - Identify NSF awards associated with EPSCoR-related outputs
    - Awards to EPSCoR hired faculty
    - Subsequent awards to faculty whose first awards were EPSCoR co-funded
    - Large awards attributed to EPSCoR
    - Awards making use of research centers that leverage EPSCoR funds
  - Calculate the percentage of NSF funding associated with these EPSCoR-associated awards for each cohort and year
- Caveats
  - Assumes that these outputs would not have happened without EPSCoR
  - Does not include other categories of EPSCoR-related outputs (e.g., awards to EPSCoR-trained students)

Competitiveness for NSF Funding: Time Series Analyses

- Used to test significance of EPSCoR participation/EPSCoR influence, controlling for other factors
- Used percentage change in NSF funding (data source: NSF awards database) as dependent variable modeled
- Modeled effect of both EPSCoR-related (e.g., years in EPSCoR program) and non-EPSCoR-related (e.g., percentage change in non-NSF Federal R&D funding) independent variables

Attribution Methods: Time Series Model

- Approach
  - Set the underlying rate of growth for EPSCoR jurisdictions to non-EPSCoR level
  - Estimate percentage of NSF funding to EPSCoR jurisdictions using this non-EPSCoR underlying rate of growth
  - Compare estimated result to actual result
- Caveat
  - Assumes the difference in underlying growth rates is associated with the EPSCoR program

Competitiveness for NSF Funding: Per-Investigator Analyses

- Analyzed as ratio of EPSCoR/non-EPSCoR
- Proposal rates, success rates, and awards per investigator
  - Data from NSF BFA (number of proposals, number of awards by jurisdiction and year)
  - Normalized based on number of S&E faculty/jurisdiction (from NCSES survey data)
- Award size
  - Data from NSF awards database

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
XX-10-2014

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
Oct 2013 - Oct 2014

4. TITLE AND SUBTITLE
Increasing Competitiveness of Investigators and Concentration Modeling: Quantitative Approaches

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jones, Thomas W.
Rieksts, Brian Q.
Zuckerman, Brian L.

5d. PROJECT NUMBER

5e. TASK NUMBER
AE-20-S243

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
IDA Science and Technology Policy Institute
1899 Pennsylvania Avenue, NW, Suite 520
Washington, DC 20006-3602

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Document NS D-5320

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
IDA Science and Technology Policy Institute
1899 Pennsylvania Avenue, NW, Suite 520
Washington, DC 20006-3602

10. SPONSOR/MONITOR'S ACRONYM(S)
STPI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This presentation was prepared for a meeting of the American Evaluation Association in October 2014. It describes two quantitative analyses conducted for the purpose of identifying the effect of the National Science Foundation (NSF) Experimental Program to Stimulate Competitive Research (EPSCoR) on capacity development defined with respect to funding competitiveness. The first analysis used NSF awards and proposal data to compare the number of proposals submitted per investigator, success rates, award sizes, and awards per funded investigator for EPSCoR and non-EPSCoR jurisdictions. This comparison, by disaggregating the determinants of funding levels, identified the particular drivers of funding levels where EPSCoR and non-EPSCoR jurisdictions differ, and suggested hypotheses regarding EPSCoR impact. The second analysis used NSF awards data to conduct a time-series analysis of changes in funding levels in EPSCoR and non-EPSCoR jurisdictions, assessing the statistical significance of EPSCoR-related variables (years in program and number of EPSCoR-funded awards received).

15. SUBJECT TERMS
econometrics, quantitative methods, time-series analysis, quasi-experimental designs, descriptive statistics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lewis, Mark J.
Unclassified	Unclassified	Unclassified	Same as Report	52	19b. TELEPHONE NUMBER *(Include area code)* 202-419-5491