Linking Academic Productivity Data to the SDR: An Attainable Goal

for presentation at:
National Science Foundation
February 29, 2008

Donna K. Ginther
University of Kansas

Funded by NSF grant SES-0353703
Introduction

- Welcome
- Research Question
- Current Match Proposal
- SDR Productivity Data
- Research Benefits of Linked Data
- Goals of Conference
Welcome

Throughout my research, I have to say:

- The use of NSF data does not imply NSF endorsement of the research, research methods, or conclusions contained in this report.

But we’re here today because of NSF financial and institutional support for the creation of the SDR Productivity Data Set.

THANK YOU
Research Question

- What are the explanations for the Gender Gap in Salary and Promotion?
- Evidence is inconclusive given limited information on productivity.
Does Science Discriminate Against Women? (2001)

- Find gender salary gap at full professor level.
Figure 1A--Salary Decompositions: 1973-2001 SDR Science Doctorates

A) Pooled Sample: Log Salary Differential

B) Pooled Sample: Decomposed Differential

C) Assistant Professors: Log Salary Differential

D) Assistant Professors: Decomposed Differential
Figure 2B--Salary Decompositions: 1973-2001 SDR Science Doctorates

E) Associate Professors: Log Salary Differential

G) Full Professors: Log Salary Differential

F) Associate Professors: Decomposed Differential

H) Full Professors: Decomposed Differential

- Endowments
- Coefficients
Does Science Discriminate Against Women? (2001)

- Unpublished because paper cannot rule out the fact that productivity differences explain the gender gap.

- Gender Promotion gap in Economics.
  - Gathered our own data and publication data.
  - Productivity does not explain the gender gap in economics.
Figure 4--Predicted Survival without Tenure Functions, by Gender and Discipline

A. Economics

B. Statistics

C. Political Science

D. Life Sciences

E. Physical Sciences

F. Engineering

Ginther and Kahn (2006)

- Gender Promotion gap in Social Science
  - Use 1973-2001 SDR
  - Find large, unexplained Gender Promotion gap
  - We cannot rule out the fact that productivity can explain much of these differences
<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Social Science x Economics</th>
<th>Economics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Promotion To Tenure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Probit Coefficient</td>
<td>-0.090</td>
<td>-0.054</td>
<td>-0.520</td>
</tr>
<tr>
<td>Promoted 11 Years Past Ph.D.</td>
<td>(0.00)</td>
<td>(0.09)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>(Covariates ex. Productivity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Ratio Estimate:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Risk Ratio</td>
<td>0.873</td>
<td>0.886</td>
<td>0.796</td>
</tr>
<tr>
<td>(No Covariates)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Model 1 Female Risk Ratio</td>
<td>0.831</td>
<td>0.841</td>
<td>0.694</td>
</tr>
<tr>
<td>(Covariates ex. Productivity)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Model 2 Female Risk Ratio</td>
<td>0.846</td>
<td>0.857</td>
<td>0.686</td>
</tr>
<tr>
<td>(Including Productivity Covariates)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>Promotion To Full</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Probit Coefficient</td>
<td>-0.083</td>
<td>-0.072</td>
<td>-0.209</td>
</tr>
<tr>
<td>Promoted 15 Years Past Ph.D.</td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>(Covariates ex. Productivity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Ratio Estimate:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Risk Ratio</td>
<td>0.784</td>
<td>0.794</td>
<td>0.729</td>
</tr>
<tr>
<td>(No Covariates)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Model 1 Female Risk Ratio</td>
<td>0.814</td>
<td>0.819</td>
<td>0.670</td>
</tr>
<tr>
<td>(Covariates ex. Productivity)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.16)</td>
</tr>
</tbody>
</table>
Why Measure Productivity?

- For academics, especially, productivity is quantifiable and can be quality adjusted.
- Workers are paid and receive promotions based on productivity.
- NSF grant rejected because the SDR data insufficient to answer the research question I posed.
Why Measure Productivity?

- Other researchers have done so:
  - Stephan and Levin (1992). *Striking the Mother Lode in Science*
  - Georgine Pion. (2001). *The Early Career Progress of NRSA Predoctoral Fellows and Trainees*
SDR Productivity Dataset

- Proposal to match productivity data onto the 1993 – 2003 SDR:
  - Publications and Citations from Thomson-ISI for 40,000 tenure track academics in SDR
    - Measure year of publication and citations
SDR Productivity Dataset

- **USPTO Inventor File**

- **USPTO Patent Citations / NBER Patent Database**
  - Citations to and from patents
SDR Productivity Dataset

- NBER Patent Citation Database
    - Variables include number of patent applications, granted, citations, inventor sequence number.
- Research Doctorate Programs in the United States
  - Ranking of Ph.D. programs 1992
Matching Issues

- Both publications and patents need to be matched by name.
- Develop algorithms based on name and address.
- Patent data needs to be reshaped into an Inventor database before matching.
Matching Issues

Matching will be imprecise:

- Develop verification methods to ensure match quality
- Contractor pulling 150 CVs from web to hand-check publication matches
- Can cross reference using patent and publication counts in SDR
Matching Issues

Resource-intensive process:
- This process started in August 2003
- Costs approximately $70,000 to perform the various matches
- Personal sunk costs in hours of lost productivity due to complex logistics of facilitating match.
Goal is to create a comprehensive data set where academic productivity is measured by individual and across time.

- Publications
- Patents
- Citations
- Other related variables
Research Benefits

- Matched Data allow researchers and policy-makers to pose and answer new questions.
- Less-expensive than modifying and fielding surveys
- More-responsive to immediate policy questions
Research Benefits

- These data will allow me to further my current research:
  - Explain the ‘unexplained’ gender gap in salary and promotion
  - Facilitate new research questions:
    - Are Foreign students more likely to cite works from home countries?
Research Benefits

- Examine factors associated with research productivity
  - Age
  - Funding
  - Training
  - Networks
Research Benefits

- Patent and Publication citations facilitate study of knowledge diffusion
- What factors have contributed to publication slowdown in US Science?
  - Other than publication lags at Economics Journals...
Other Potential Data to Match

- NSF and NIH have Funding Databases that could be matched to SDR:
  - Evaluate the impact of Federal Funding on Research Productivity
  - Causal impact of NSF / NIH graduate fellowships on career outcomes
Conclusion

- Conference designed to bring together stake holders:
  - NSF & NIH invested in SDR
  - Research Community
  - Data Linking Experts
- How do we create high-quality, enhanced version of SDR?
Goals of Conference

- Seek input from research community on value of data to be created.
- Foster discussion about best practices for linking NSF data to other sources.
- Discuss Data Dissemination