

# 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**

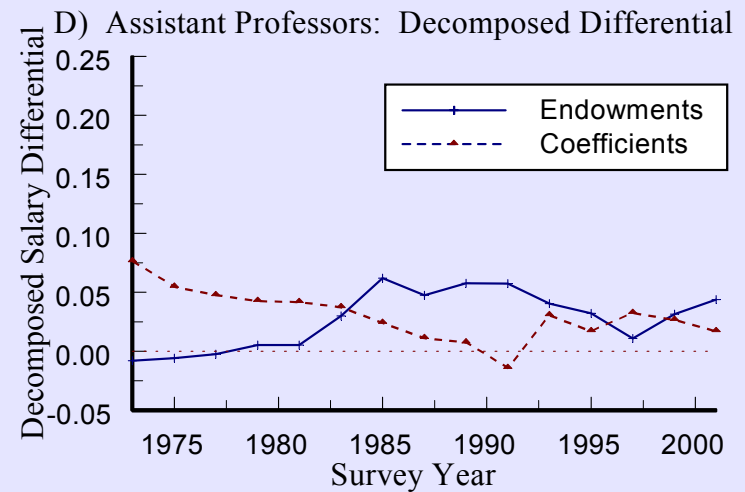
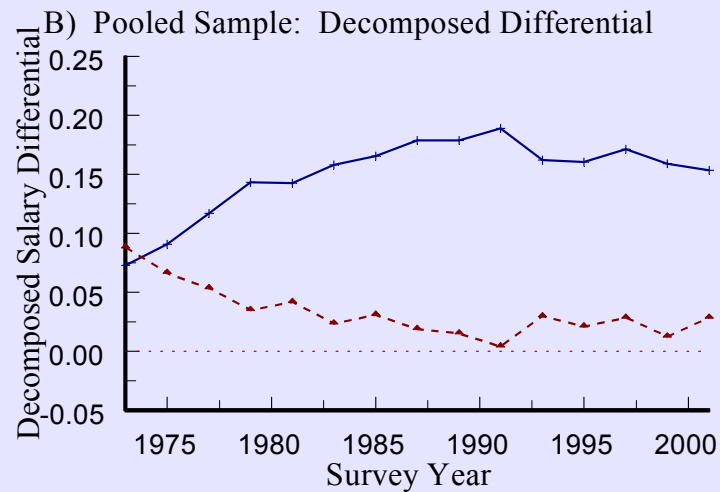
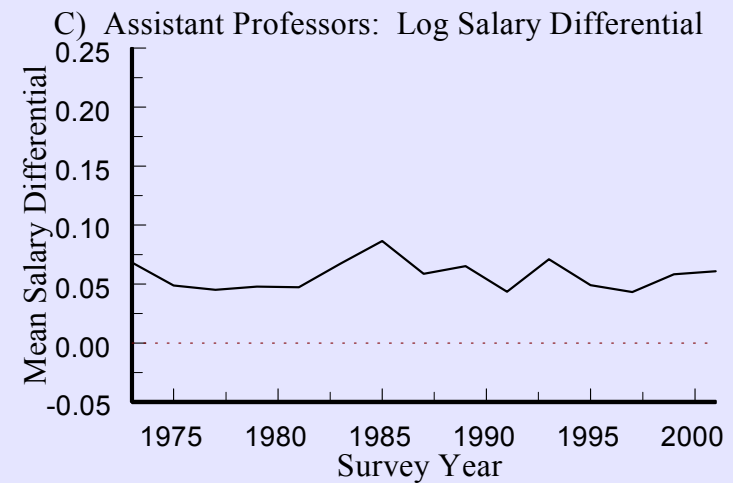
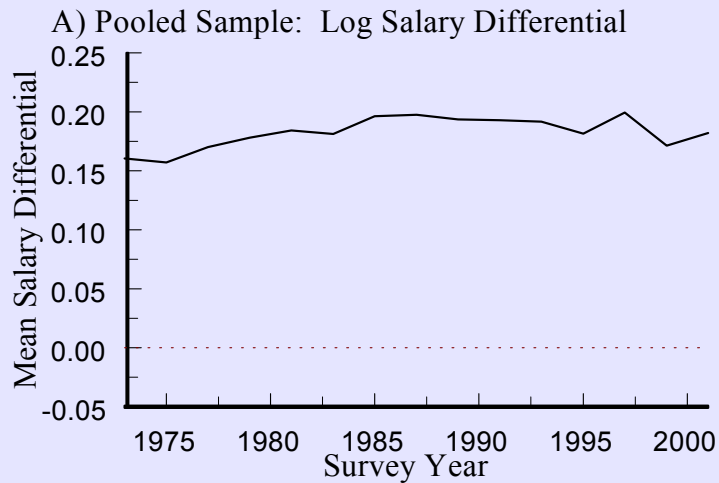
# Research Question

- What are the explanations for the Gender Gap in
  - Salary
  - Promotion
- Evidence is inconclusive given limited information on productivity

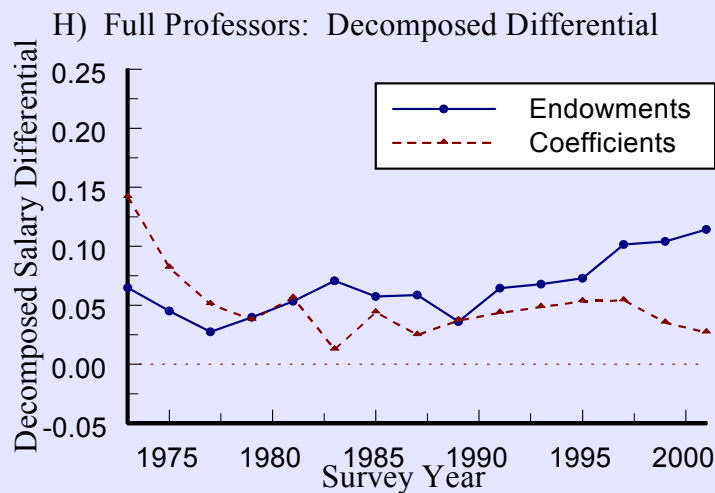
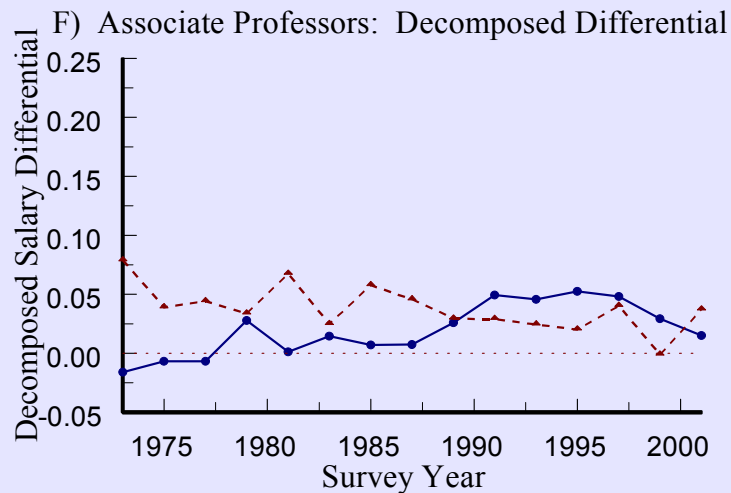
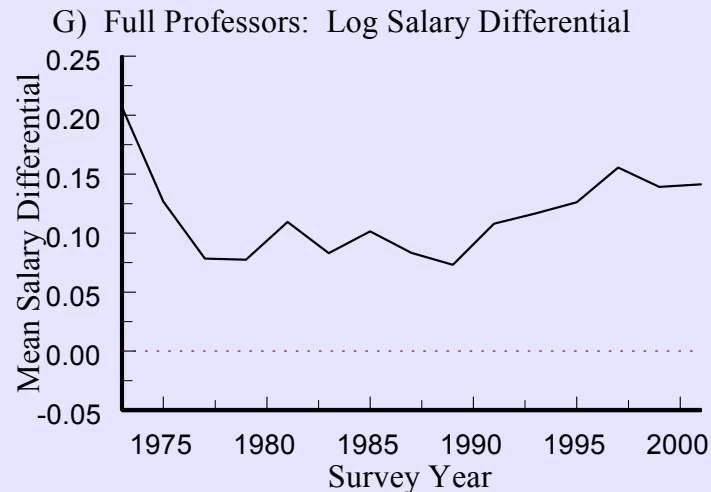
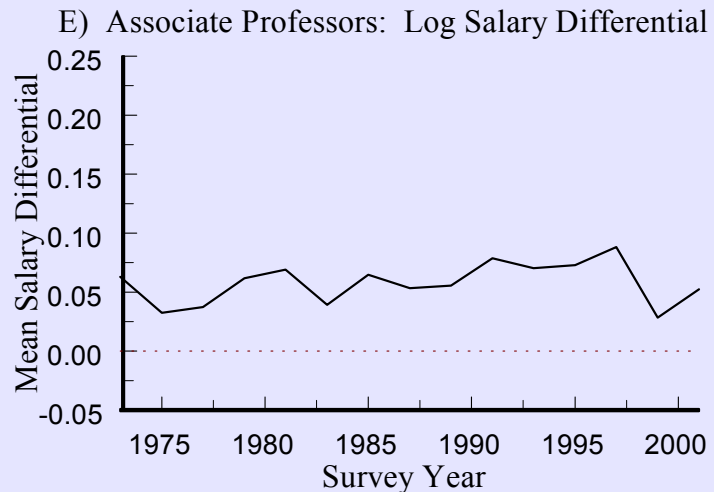
# Does Science Discriminate Against Women? (2001)

- Used 1973-1997 SDR to examine gender differences in pay and promotion in Science.
- Find gender salary gap at full professor level.

**Figure 1A--Salary Decompositions: 1973-2001 SDR Science Doctorates**



**Figure 2B--Salary Decompositions: 1973-2001 SDR Science Doctorates**



# Does Science Discriminate Against Women? (2001)

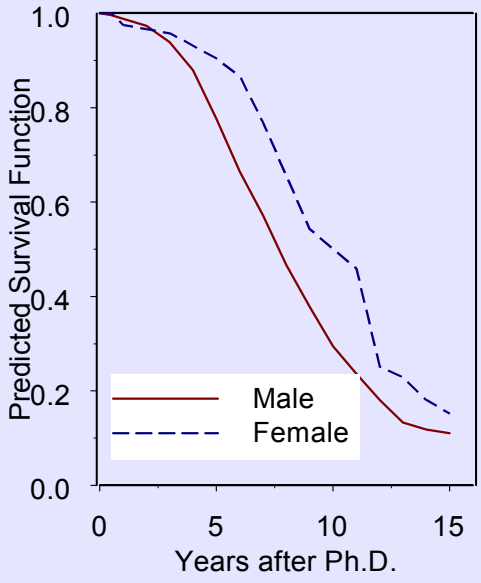
- Unpublished because paper cannot rule out the fact that productivity differences explain the gender gap.

# Ginther and Kahn (2004)

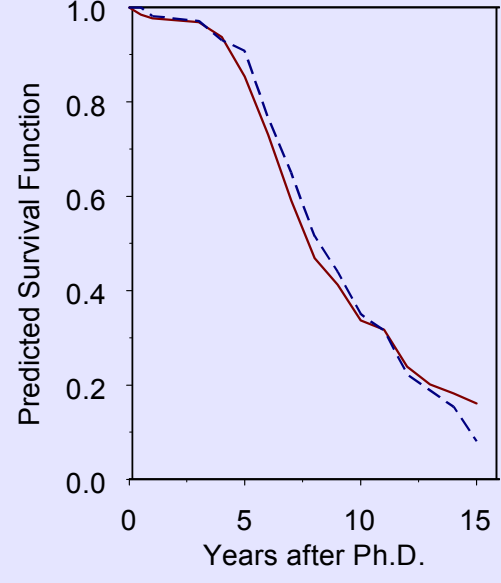
- Gender Promotion gap in Economics.
  - Gathered our own data and publication data.
  - Productivity does not explain the gender gap in economics.

**Figure 4--Predicted Survival without Tenure Functions, by Gender and Discipline**

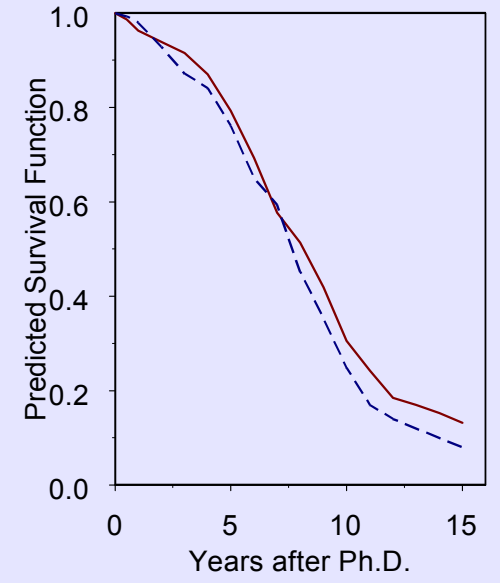
**A. Economics**



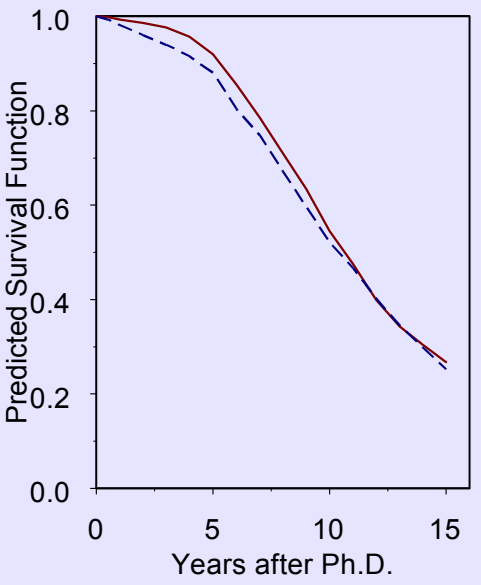
**B. Statistics**



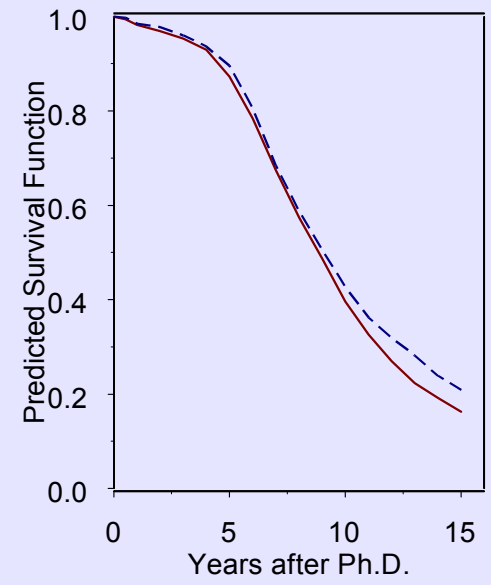
**C. Political Science**



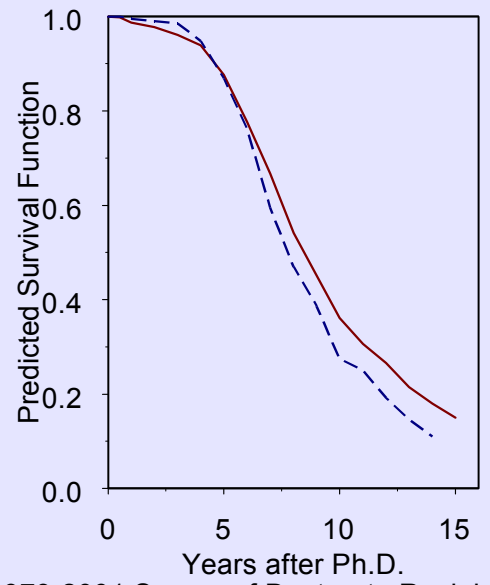
**D. Life Sciences**



**E. Physical Sciences**



**F. Engineering**



Source: 1973-2001 Survey of Doctorate Recipients

# 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 3: Gender Differences in the Probability and Hazard of Promotion  
1973-2001 Survey of Doctorate Recipients**

	<u>Full Sample</u>	<u>Social Science x Economics</u>	<u>Economics</u>
<b><u>Promotion To Tenure</u></b>			
<b>Female Probit Coefficient</b>	<b>-0.090</b>	<b>-0.054</b>	<b>-0.520</b>
Promoted 11 Years Past Ph.D. (Covariates ex. Productivity)	(0.00)	(0.09)	(0.00)
<b><u>Risk Ratio Estimate:</u></b>			
<b>Female Risk Ratio</b> (No Covariates)	<b>0.873</b> (0.01)	<b>0.886</b> (0.02)	0.796 (0.13)
<b>Model 1 Female Risk Ratio</b> (Covariates ex. Productivity)	<b>0.831</b> (0.00)	<b>0.841</b> (0.00)	<b>0.694</b> (0.03)
<b>Model 2 Female Risk Ratio</b> (Including Productivity Covariates)	<b>0.846</b> (0.00)	<b>0.857</b> (0.01)	<b>0.686</b> (0.03)
<b><u>Promotion To Full</u></b>			
<b>Female Probit Coefficient</b>	<b>-0.083</b>	<b>-0.072</b>	-0.209
Promoted 15 Years Past Ph.D. (Covariates ex. Productivity)	(0.02)	(0.04)	(0.12)
<b><u>Risk Ratio Estimate:</u></b>			
<b>Female Risk Ratio</b> (No Covariates)	<b>0.784</b> (0.00)	<b>0.794</b> (0.00)	0.729 (0.18)
<b>Model 1 Female Risk Ratio</b> (Covariates ex. Productivity)	<b>0.814</b> (0.01)	<b>0.819</b> (0.02)	0.670 (0.16)

# 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
  - 1975-2004 Patent Numbers, Inventor names.
- USPTO Patent Citations / NBER Patent Database
  - Citations to and from patents

# SDR Productivity Dataset

- NBER Patent Citation Database
  - 1963-99, currently being updated through 2004. Contains 4 million patents and related citations.
    - 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.

# SDR Productivity Dataset

- Goal is to create a comprehensive data set where academic productivity
  - Publications
  - Patents
  - Citations
  - Other related variables
- Are measured by individual and across time.

# 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