Measuring Knowledge Flows from Public Research

The State of the Art and Future Directions

Michael Roach – University of North Carolina

National Academies Workshop:

Measuring the Impacts of Federal Investments in Research

April 18, 2011

Overview

- What do we know?
  - Virtues and limits of existing measures of knowledge flows from public research
- New insights on the validity of patent citations
  - Findings from a comparison of patent citations and survey data (with Wes Cohen)
- Where do we go from here?
  - Suggestions for current and future measures

Measuring Knowledge Flows

- The challenge
  - To understand the impact of public research, we need to measure which firms use it, how much, in what ways, and the subsequent outputs… but data are limited and imperfect
- Past and current measures
  - Aggregate university research expenditures (Jaffe, 1989) or scientific publication output (Adams, 1990), but we would like to observe direct flows to firms
  - Valuable insights from survey data (Mansfield, 1991; Klevorick et al., 1995; Cohen et al., 2002); cross-sectional, small samples, etc.
  - Most prominent measures are patent citations to university patents (Jaffe et al., 1993; Narin et al. 1997; Henderson et al., 1998) and, more recently, patent citations to scientific publications

Measuring Knowledge Flows

- Patent citations to public research
  - Widely available across industries, firms, and over time
  - Descriptive tool for understanding the social welfare impact of public research (NSF, OECD, Narin et al. 1997, etc.)
  - Measure of source and intensity of knowledge flows (Jaffe, et al. 1993; Henderson et al. 1998; Mowery et al., 2001, 2002) that are often used as a predictor of firm innovative performance
- Known limitations
  - Not all inventions are patentable or patented
  - Not all knowledge flows are citable or cited
  - As a result, citations likely mismeasure knowledge flows
    - Are citations simply noisy measures that understate knowledge flows?
    - Or is error systematic, and is it related with other variables of interest?

Assessing the Validity of Citations

"Patent Citations as Measures of Knowledge Flows from Public Research" (M. Roach & W. Cohen, 2011)

- Objective
  - Empirically explore for possible sources of error in patent citations as measures of knowledge flows
- Approach
  - Measurement model to identify sources of error
    - Factors that citations should reflect, but do not *(errors of omission)*
    - Factors that citations should not reflect, but do *(errors of commission)*
  - Compare citations to survey measure; assume that both are independent measures of latent of knowledge flows
  - Residual analysis to estimate impact of "unobserved" component of knowledge flows on firm innovative performance

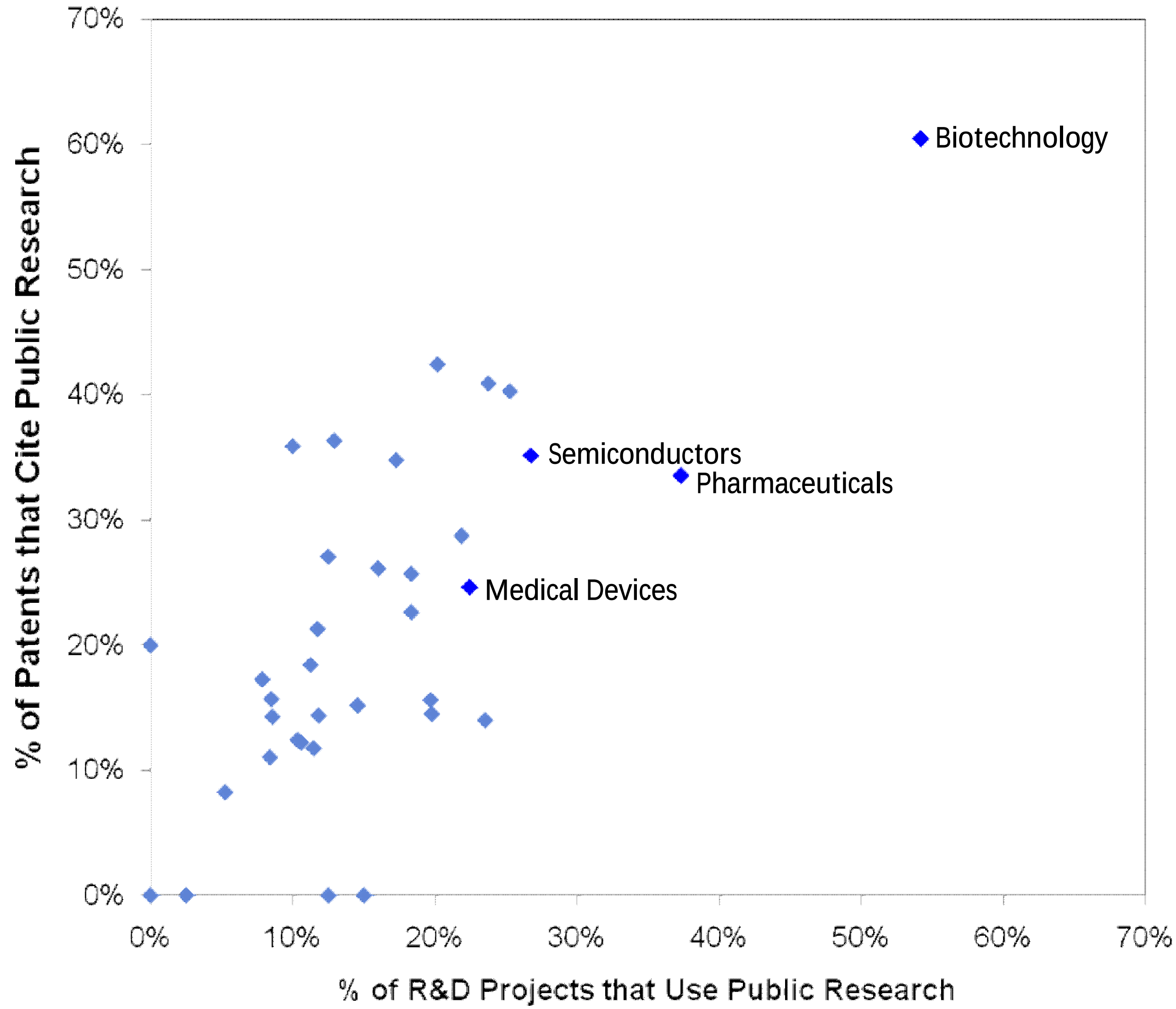
Two Measures of Knowledge Flows

- Patent-based measures
  - Citations to patent references (PR) assigned to university, federal lab, or research institute
  - Citations to non-patent references (NPR) w/author affiliated with university, federal lab, or research institute
    - NPRs are primary output of public research and less influenced by patent examiners (Sampat 2005; Lemley & Sampat 2010); likely more accurate measure that citations to patent references
- Survey-based measure (Carnegie Mellon Survey)
  - Reported fraction of R&D projects that use public research (5-point scale; percentage recoded at midpoints)

Mean Comparison by Industry

Industry	Obs	% R&D Projects that Use Public Research (survey)	% Patents that Cite Public Research (patent)
Food	35	26.1	30.9
Agriculture, mining, etc.	24	17.3	15.9
Chemicals	92	16.4	24.3
Steel, metal, etc.	54	11.5	19.2
Drugs & biotech	38	42.0	56.9
Medical devices	51	22.4	32.1
Information Technology	60	23.8	28.9
Aerospace	66	19.8	36.7
Machinery	90	13.8	15.5
Electrical equipment	32	8.0	18.8
Automobiles	20	7.5	9.8
General manufacturing	36	8.9	11.6
Total	598	18.2	25.6

Mean Comparison by Industry



Correlations with Survey Measure

	Obs.	(1) % Patents Cite Public Research	(2) # Patent Refs. to Public Research	(3) # Nonpatent Refs. to Public Research
Correlations				
Industry-level Correlation[a]	34	0.74*	0.51*	0.87*
Firm-level Correlation	598	0.26*	0.13*	0.21*
Firm-level Partial Correlation[b]	598	0.16*	0.12*	0.18*
Chronbach's alpha[c]				
Industry-level alpha[a]	34	0.85	0.72	0.94
Firm-level alpha	598	0.41	0.24	0.38

[a] Industry averages; 34 ISIC industries represented; [b] Controlling for 34 ISIC industries; [c] Scale reliability coefficient

- Although there is a high correlation across aggregate industry measures, correlations across firms within is very weak; unclear whether this is due to noise or systematic variation

Summary of Citations as Measures

- What NPR citations capture
  - Knowledge flows through channels of open science (e.g. publications)
  - Direct use of technological opportunities in new R&D projects
  - Knowledge flows to firm's applied research (patentable output)
- What NPR citations *don't capture*, but should (errors of omission)
  - Knowledge flows through contract-based relationships
  - Intermediate use in existing projects (foundational knowledge)
  - Knowledge flows to firm basic research activity (less patentable)
- What NPR citations capture, but *shouldn't* (errors of commission)
  - Patent effectiveness and citing propensity increase citations
  - Secrecy decreases citations; likely through lower patent propensity
  - Industrial PhDs cite NPRs *above and beyond* that explained by use of public research alone; perhaps due to academic norms to extensively cite prior research

Residual Analysis

- Observed and unobserved knowledge flows
  - When using citations as a measure of k_i certain dimensions are observed (k_1) while others are unobserved (k_2)
    - Citations = k_1; $\text{cov}(k_2, v_1) \neq 0$
    - Survey = $k_1 + k_2$; where k_1 is shared variance between measures
  - Regress survey measure onto citations to predict unobserved component ($\hat{k}_2$), which can then be included in performance equation with citations
- Measure of innovative performance
  - *Citation-weighted patent counts*: widely used, but suffer from own limitations and potential sources of error
  - Not all inventions are patentable or patented
  - Citations to patents are influenced by attorneys, examiners, and others (Sampat 2005; Alcacer & Gittelman 2008; Alcacer et al. 2009)

Residual Analysis:

Impact on firm innovative performance

- Dependent variable: citation-weighted patent counts
- Method: Poisson QML; marginal effects reported

	(1)	(2)	(3)	(4)	(5)
%NPR (k_1)	17.18**		16.17**	15.53**	17.14**
%Survey ($k_1 + k_2$)		22.74*		15.60*	
%Survey_residual ($\hat{k}_2$)					15.80*
Patent Effectiveness (p)			7.51	4.55	4.50
Secrecy (p)			-10.49	-10.02	-9.99
Citing Propensity (p)			-0.06	-0.09	-0.09
ln(RD)	12.16**	14.28**	11.83**	12.14**	12.13**
Industry dummies	Incl.	Incl.	Incl.	Incl.	Incl.

N = 598; ** ≤1%, * ≤5%, + ≤10%

- *Suggests that patent citations underestimate impact of public research*
- *Unobserved contribution is comparable to what is observed*

Summary

- NPRs useful—but limited—measure of knowledge flows
  - Significant relationship between survey and patent citations
  - Citations to NPRs are most strongly associated with
    - Knowledge through channels of open science (e.g., publications)
    - Technological opportunities stimulating new R&D projects
- Limitations of citations
  - Fail to reflect tacit knowledge, private interactions, use in firm basic research activity
  - Industrial scientists make citations over explained by use
- Non-classical measurement error bias
  - Citations are downward biased; understate impact of public research on firm innovative performance
  - PRs appear more biased than NPRs
  - Other regressors of innovative performance potentially biased

Future Directions

- Citations to nonpatent references
  - Although limited and difficult to obtain, they are best we have
  - *Agencies*: make nonpatent data more readily available with greater uniformity and detail
  - *Citation origins*: not just examiner v. applicant, but also attorney and inventor
- Complementing with other data sources
  - NSF Business R&D and Innovation Survey (BRDIS)
    - *Breadth:* "What share of R&D projects use research findings from universities and government labs?"
    - *Intensity:* "For a recent important project, how much did the project utilize research findings from universities and government labs?"
  - Micro-level data on channels, mechanisms, and impact of knowledge flows
    - Surveys, archival firm data, qualitative field studies, etc.

Thank you

michael.roach@unc.edu

Comparing Citations and Survey Data

- Measures: % R&D projects (Survey), num. patent citations (PR, NPR)
- Marginal effects reported

Dependent Variable	(1) Survey	(2) PR	(3) NPR	(4) Survey	(5) PR	(6) NPR	(7) Survey	(8) PR	(9) NPR	(10) Survey	(11) PR	(12) NPR
Channels												
Open Science	0.11**	0.04	0.16*									
Private Interactions	0.04**	0.00	0.10									
Industrial Scientists	0.14**	0.39*	1.52**									
Uses of Public Research												
Suggesting New R&D Projects				0.10**	0.11	0.48*						
Completion of Existing R&D Projects				0.07**	0.10	0.12						
Composition of R&D Activity												
ln[Basic]							0.02*	-0.00	-0.01			
ln[Applied]							0.04*	0.03**	0.07**			
ln[Development]							0.00	0.02*	0.02			
Patenting Behavior												
Patent Effectiveness										0.04	-0.06	0.37**
Secrecy										0.02	-0.17	-0.43*
Citing Propensity										0.00	0.00*	0.03**

Survey models are fractional logistic, PR & NPR models are negative binomial; N = 598; ** ≤1%, * ≤5%, + ≤10%