Inventor Mobility and Knowledge Transmission in Nanotechnology

Jinyoung Kim
Korea University

Sangjoon John Lee
Alfred University

Gerald Marschke
Harvard University, NBER and IZA

Prepared for "Nanotechnology and Society Workshop," October 3, 2008, Amherst, MA. This is a work in progress. Please do not quote without the authors' permission.

- Scanning electron microscope image shows rows of horizontal zinc-oxide nanowires grown on a sapphire surface. Gold nanoparticles are visible on the ends of each row.

Objectives

- Using U.S. patent records, we study how nanotechnology is transmitted from firm to firm.
- While knowledge can be disseminated via published patents, papers, and textbooks, at conferences where research is presented and where industry and academic research personnel comingle, and via informal social networks…
- This paper focuses on the role of innovator mobility as a pathway for the diffusion of ideas through industry.
- When innovators patent, they leave a paper trail.
- We use this trail to follow inventors from firm to firm and measure whether an employer's innovation reflects the employment history of its inventors.

Main Findings

1. Our principle finding is consistent with a story that, in at least one important nanotechnology subfield, when inventors move among firms they spread knowledge.

2. In particular, we find if we consider any two patents in the "Chemicals, misc." subclass, A and B, where A and B are assigned to different firms and where A is granted after B, patent A is more likely to cite patent B if the patent A firm employs an inventor who earlier worked for the patent B firm.

Outline

- Literature Review
- Data Sources
- Empirical Findings
- Conclusion

Literature Review (1)

- S&T scholars have long suspected that inter-firm mobility of scientists transmits technological know-how, but econometric evidence scarce

  Arrow (1962), Stephan, 1996

- Why is mobility necessary for transmission? Tacit knowledge?

  Polanyi, 1958, 1966

- Geographically limited technical diffusion…

  Jaffe, 1989; Jaffe, Trajtenberg, and Henderson, 1993; Audretsch and Feldman, 1996; Zucker, Darby, and Brewer, 1998; and Mowery and Ziedonis, 2001

- …often interpreted as evidence of tacit knowledge

  Feldman (1994)

Literature Review (2)
"Star" University Scientists→Industry

- Universities importance source of knowledge for industry?
  - Evidence for geographically localized spillovers occurring in areas around major universities

    Jaffe, 1986, 1989; Audretsch and Feldman, 1996; Henderson et al, 1998
- University knowledge tacit?
  - Best predictor that academic idea succeeds in industry is involvement of academic originator (scientist) in development

    Jensen and Thursby (2001), Agrawal and Henderson (2002), and Thursby and Thursby (2002)

Literature Review (3)
"Star" University Scientists→Industry

- University star scientists are important source of knowledge, which is probably tacit
  - Biotech industry—entry, measures of firm success correlated with firm collaboration with star scientists

  Zucker, Darby, and Armstrong, 1998, 2001; Zucker, Darby, and Brewer, 1998; and Zucker and Darby, 2001
  - Nanotech industry—nanotech agglomeration around universities, knowledge transfer takes place via collaboration (publishing between university and industry) and this is conducive to nanotechnology progress

  Zucker and Darby (2007), Darby and Zucker (2003), Zucker, Darby, Furner, Liu, and Ma (2007)

Literature Review (4)
Industry→Industry?

- We know interfirm mobility very high in some areas
- Econometric evidence that job to job mobility facilitates transmission of knowledge is indirect and circumstantial
  - Semiconductor firms are more likely to cite patents of other firms in their region if inventor mobility rates are high

    Almeida and Kogut (1999)
  - Patenting rates higher where firms face higher mobility

    Kim and Marschke (2005)
  - Scientists accept wage cuts early in career for prospect of higher wages later

    Moen (2005)

In This Work

- We use patent citations to track the diffusion of university innovations to nanotechnology industry.
- We use patent data to track the movement of inventors among firms
- We ask in this paper: Does the movement of inventors correlate with citations, or

Is patent A more likely to cite patent B if the patent A firm employs an inventor who earlier worked for the patent B firm?

Data Description

- The data used in this paper are a part of the inventor-firm panel database that we created.

- Data sources

  (1) Patent Bibliographic data (Patents BIB) by USPTO that contain bibliographic information for U.S. utility patents issued from 1975 to 2002.

  (2) NBER Patent-Citations data collected by Hall, Jaffe and Trajtenberg (2001) which contain all citations made by patents granted in 1975-1999.

  (3) Nanobank database collected by Zucker and Darby (2007) that identifies patents in nanotechnology.

Inventor Name Matching in Patent BIB data

- The inventors listed is comprehensive and include only those inventors that make a creative contribution to the innovation underlying the patent.

- We treat each entry in inventor name field as a unique inventor.

- With N names, generate $N(N\text{-}1)/2$ pairs (N = 5.1 million, $N(N\text{-}1)/2$ = 13 trillion).

- For each pair, decide if the two names belong to the same inventor, based on the following criteria:

Inventor Name Matching in Patent BIB data

- Two names in a pair are matched if the SOUNDEX codes of their last names and their full first names are the same, and at least one of the following 3 conditions is met:

  (1) the full addresses are the same;

  (2) one name is an inventor of a patent that is cited by another patent whose inventors include the other name; or

  (3) the two names share the same co-inventor.

- Two names in a pair are matched if the two names have the same full last and first names, and at least one of the following 3 conditions is met:

  (1) they have the same zip code;

  (2) they have the same full middle name; or

  (3) they reside in the same MSA area.

- A pair is considered a '**mismatch**' if middle name initials are different.

Inventor Name Matching in Patent BIB data

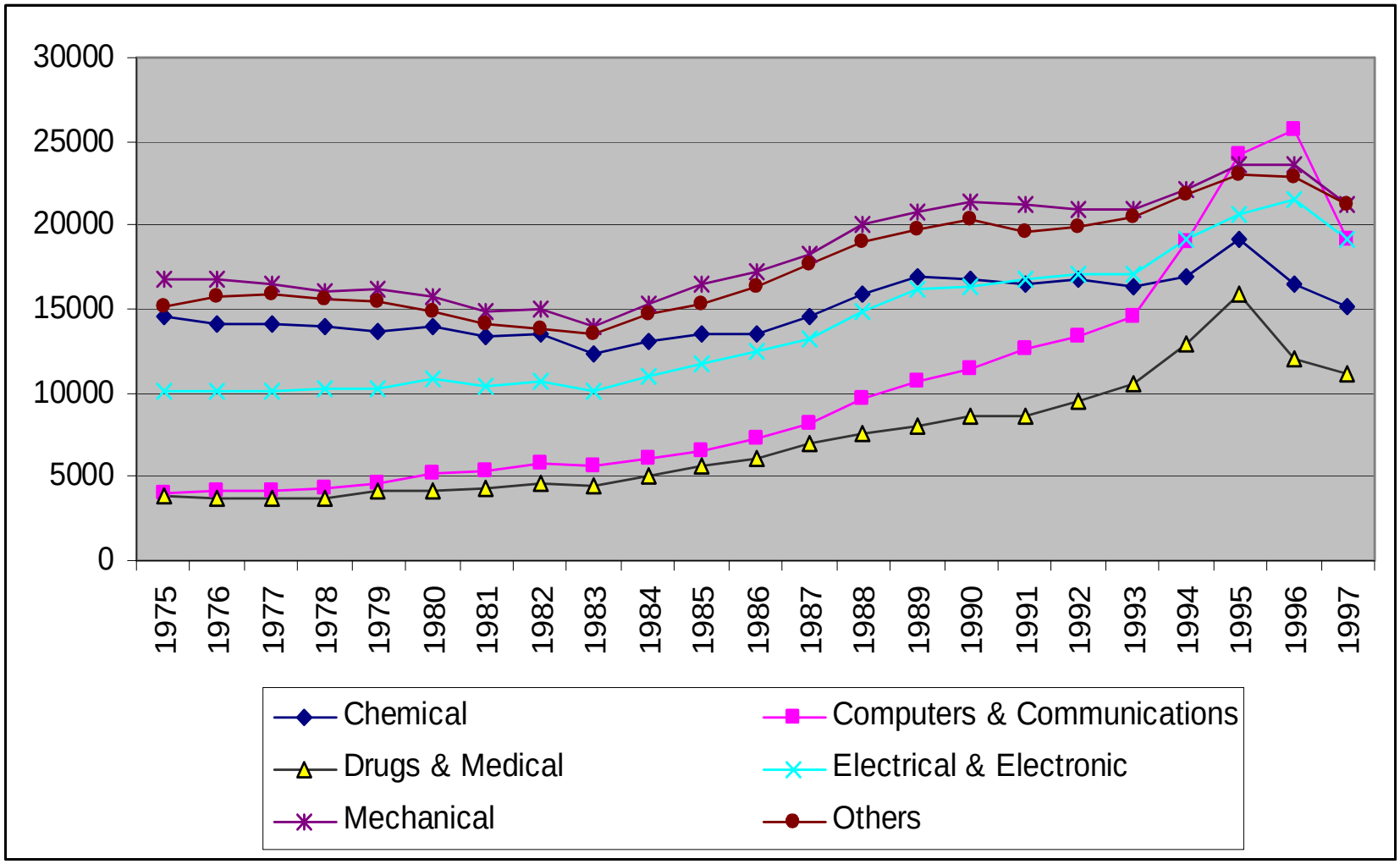
- We impose transitivity: If A⇒B and B⇒C, then A⇒C.
- Using this method, we identify 1.72 million unique inventors (34%) out of 5.1 million names in the entire patent data.
- After name matching, we add information on all citations from the NBER Patent-Citations data where each citing patent that was granted between 1975 and 1999 is matched to all patents cited by the patent.
- As the final step, we select only those patents in nanotechnology that are identified in the Nanobank database.

About the information on a patent…

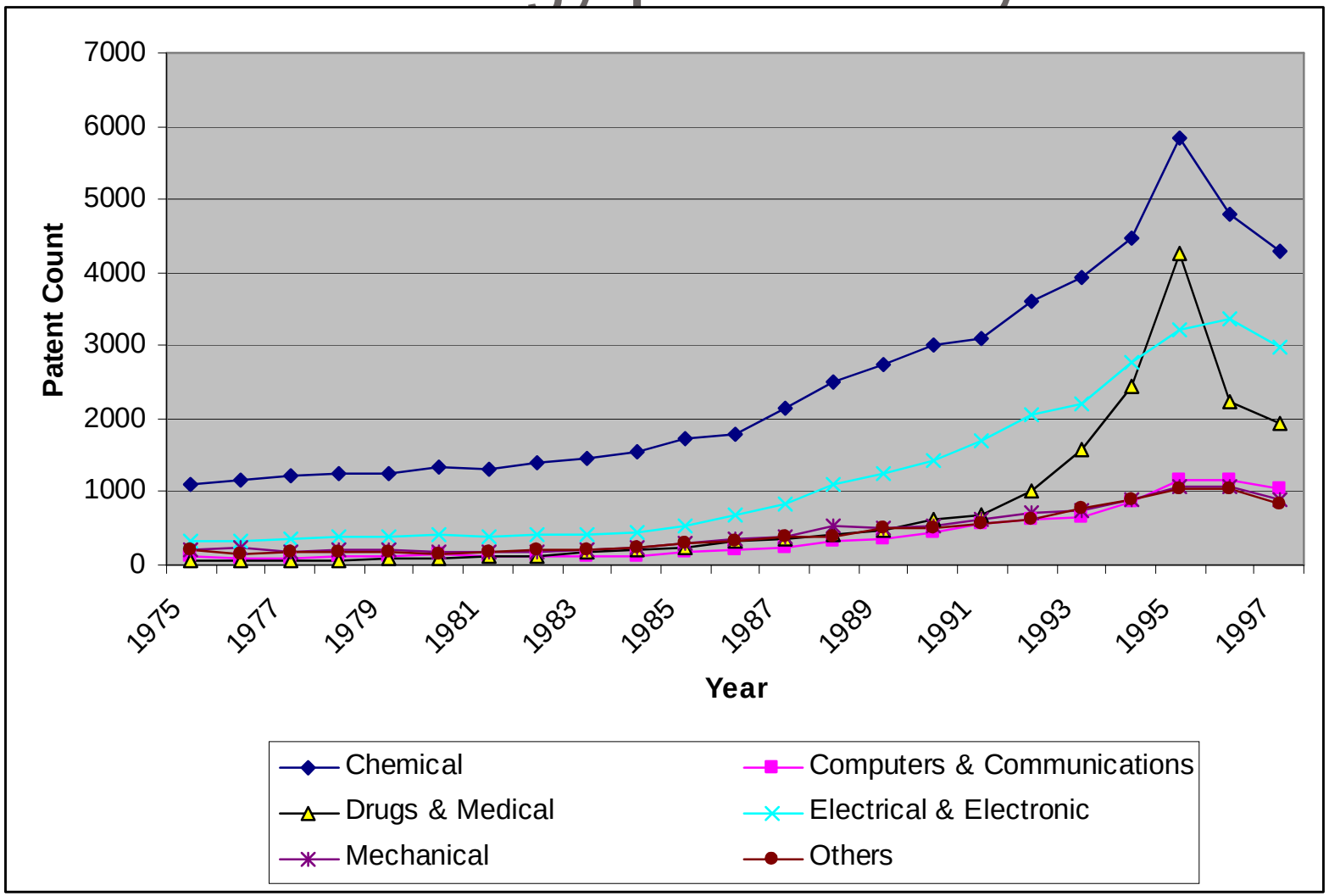
- *Assignee*—usually the inventors' employer, sometimes a collaborating firm
- *Citation*—The patent citations documents the "prior art" upon which the new innovation builds.
  - Evidence that citations proxy for knowledge flows

  Jaffe, Fogarty, and Banks, 1998; and Duguet and MacGarvie, 2005
- *Technological class*

All patents by field



Nanotechnology patents by field



Empirical Methodology (1)

- We examine only those nanotechnology patents in patent technological category “Chemicals, miscellaneous” (USPTO subcategory 19)
- We will estimate the determinants of one Nano patent in subcategory 19 citing another patent in subcategory 19
  - Challenge: for each citing patent there are thousands of potential or cite-able patents

Empirical Methodology (2)

- Solution: estimate a weighted logit (Manski and Lerman, 1977; used by Singh, 2006; Stolpe, 2003; and Jaffe et al, 1993)
  - Use all patents that cite another patent in subcategory 19
  - For comparison group, use a random sample of the potential patent pairs in subcategory 19 that do not cite one another
  - When forming likelihood, randomly sampled patent pairs in the comparison group are weighted up by the inverse of their probability of being sampled so that their contribution to the likelihood is proportional to their numbers in the population

Empirical Methodology (3)
Key Variables

- MOBILITY: Indicator variable, equal to one if citing patent and cite-able patent share an inventor.
  - Suppose for example, patent B is filed on June 1, 1980 by assignee b and A is filed on June 1, 1990 by assignee a. If an inventor on A was an inventor on any patent assigned to b between June 1, 1975 (5 years before B was filed) and June 1, 1990 then the dummy is coded 1, else it is coded zero.

Table 1
Descriptive Statistics
Nanobank Patents from Subclass 19 ("Chemicals, Miscellaneous")

Variable	Definition	Obs.	Mean	Std. Dev.
CITE	Indicator equal to one if patent cites another subclass 19 patent	469181	.335	.472
MOBILITY	Indicator equal to one if citing patent and assignee of cited/cite-able patent share an inventor	469181	.006	.076
LCLAIM_B	Log of the number claims made by cited/cite-able patent	436496	2.327	.794
CITELAG	Number of years between application dates of citing and cited/cite-able patent	469181	8.111	5.446
CRECEIVE_B	Number of citations received by cited/cite-able patent in 5 years following grant date	469181	10.335	17.216
CMADE_B	Number of citations made by cited/cite-able patent	439206	8.084	8.478
GENERAL_A	Generality, citing patent	297967	.326	.292
GENERAL_B	Generality, cited/cite-able patent	421936	.391	.279
ORIGINAL_A	Originality, citing patent	463655	.439	.278
ORIGINAL_B	Originality, cited/cite-able patent	429271	.380	.279

Table 2
Estimating Determinants of Citing Patent in Same Class
Weighted Logit
("Probability that Patent A cites Patent B")

Variable	I All	II All	III Ind.→Ind.	Marg. Effect
MOBILITY	7.010 (.448)	7.390 (.493)	7.331 (.496)	.080
CITELAG		-.154 (.009)	-.153 (.010)	-.00006†
CITELAG^2		.014 (.001)	.014 (.001)	.00002†
LCLAIMS_B		.113 (.014)	.108 (.014)	.000006
CMADE_A		.015 (.001)	.014 (.001)	.000009†
CRECEIVE_B		.027 (.002)	.028 (.002)	.00001†
GENERAL_B		.941 (.039)	.867 (.043)	.000017†
Obs. Wald Chisq.(*p.v.*) Log likelihood	469181 245.3 (.0000) -355.476	240599 5435.6 (.0000) -185.162	225454 5302.1 (.0000) -173.367	

Robust standard errors in parentheses. Marginal effects are from model III.
†Partial elasticity.

Discussion

- Finding: consider two patents in the “Chemicals, misc.” subclass, A and B, where A and B are assigned to different firms and where A is granted after B, patent A is more likely to cite patent B if the patent A firm employs an inventor who earlier worked for the patent B firm.
  - Having a shared inventor increases probability by .08

Discussion

- Other stories consistent with data:
  - We would expect that inventors when they move tend to remain in the same research areas and the fact that two firms share an inventor is an indication of, and not a cause of, similar research agendas
  - When inventors move they do not move far, hence mobility proxying for clustering by research area
  - Solutions in future work: Use more homogeneous technological subclass, dummies for clusters

Appendix SOUNDEX

- A SOUNDEX code for a surname is an upper case letter followed by 6 digits. For example the SOUNDEX code for Kim is K500000, while that for Marschke is M620000. The first letter is always the first letter of the surname.

- The rules for generating a SOUNDEX code are:

1. Take the first letter of the surname and capitalize it.
2. Go through each of the following letters giving them numerical values from 1 to 6 if they are found in the Scoring Letter table (1 for B, F, P, V; 2 for C, G, J, K, Q, S, X, Z; 3 for D, T; 4 for L; 5 for M, N; 6 for R; 0 for Vowels, punctuation, H, W, Y).
3. Ignore any letter if it is not a scoring character. This means that all vowels as well as the letters h, y and w are ignored.

Appendix SOUNDEX

4. If the value of a scoring character is the same as the previous letter then ignore it. Thus if two ‘t’s come together in the middle of a name they are treated exactly the same as a single ‘t’ or a single ‘d’. If they are separated by another non-scoring character then the same score can follow in the final code. The name PETTIT is P330000. The second ‘T’ is ignored but the third one is not since a nonscoring ‘I’ intervenes.
5. Add the number onto the end of the SOUNDEX code if it is not to be ignored.
6. Keep working through the name until you have created a code of 6 characters maximum.
7. If you come to the end of the name before you reach 6 characters, pad out the end of the code with zeros.
8. Optionally you can ignore a possessive prefix such as ‘Von’ or ‘Des’.