Using patent data as indicators

Prof. Bronwyn H. Hall
University of California at Berkeley, University of Maastricht; NBER, NIESR, and IFS

Outline

- Overview
- Sources for patent data
- Knowledge measurement
- Knowledge value
- Knowledge flows
- Knowledge types

Griliches (1990)

Patents and patent statistics have fascinated economists for a long time. Questions about sources of economic growth, the rate of technological change, the competitive position of different firms and countries, the dynamism of alternative industrial structures and arrangements all tend to revolve around notions of differential inventiveness: What has happened to the "underlying" rate of technical and scientific progress? How has it changed over time and across industries and national boundaries? We have, in fact, almost no good measures on any of this and are thus reduced to pure speculation or to the use of various, only distantly related, "residual" measures and other proxies.

In this desert of data, patent statistics loom up as a mirage of wonderful plenitude and objectivity. They are available, they are by definition related to inventiveness, and they are based on what appears to be an objective and only slowly changing standard.

(from his Introduction to Patent Statistics as Economic Indicators: A Survey, p. 1661)

Patents as indicators

- A patent is a property right to a knowledge asset => patent counts can be useful measures of innovative output
  - Counts at the firm, industry, country level over time
  - Counts weighted by the number of subsequent citations that the patents receive
- Citations from one patent to another
  - an imperfect but useful map of the links between these "bits" of output or knowledge

But.....

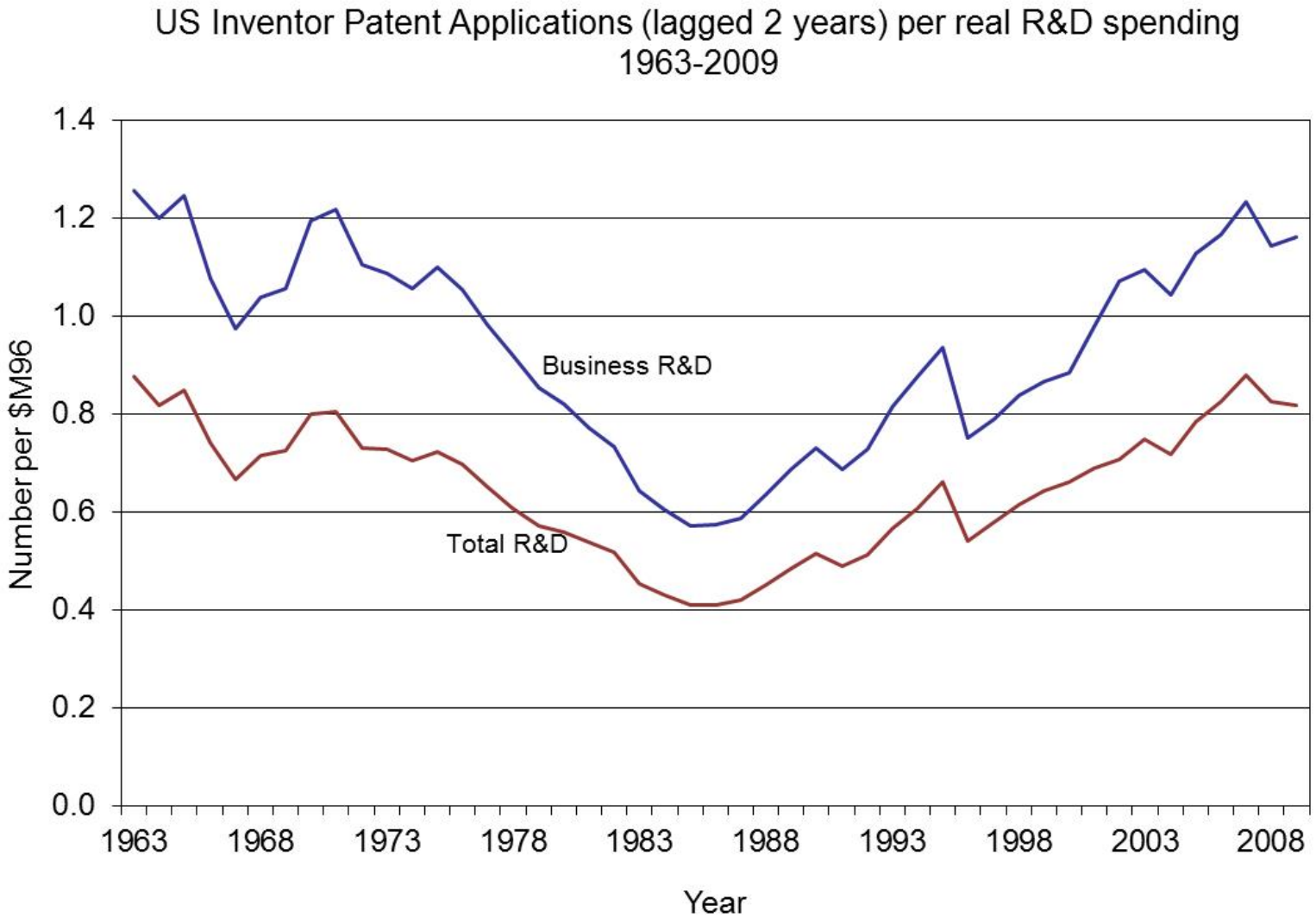
- Using patents as indicators requires some understanding of what they mean
  - how and why they are taken out
  - how they are administered
  - how they are enforced
  - how all this changes over time
- Simply assuming that patents are a stable measure of innovative output is not advisable

Pavitt (1988)

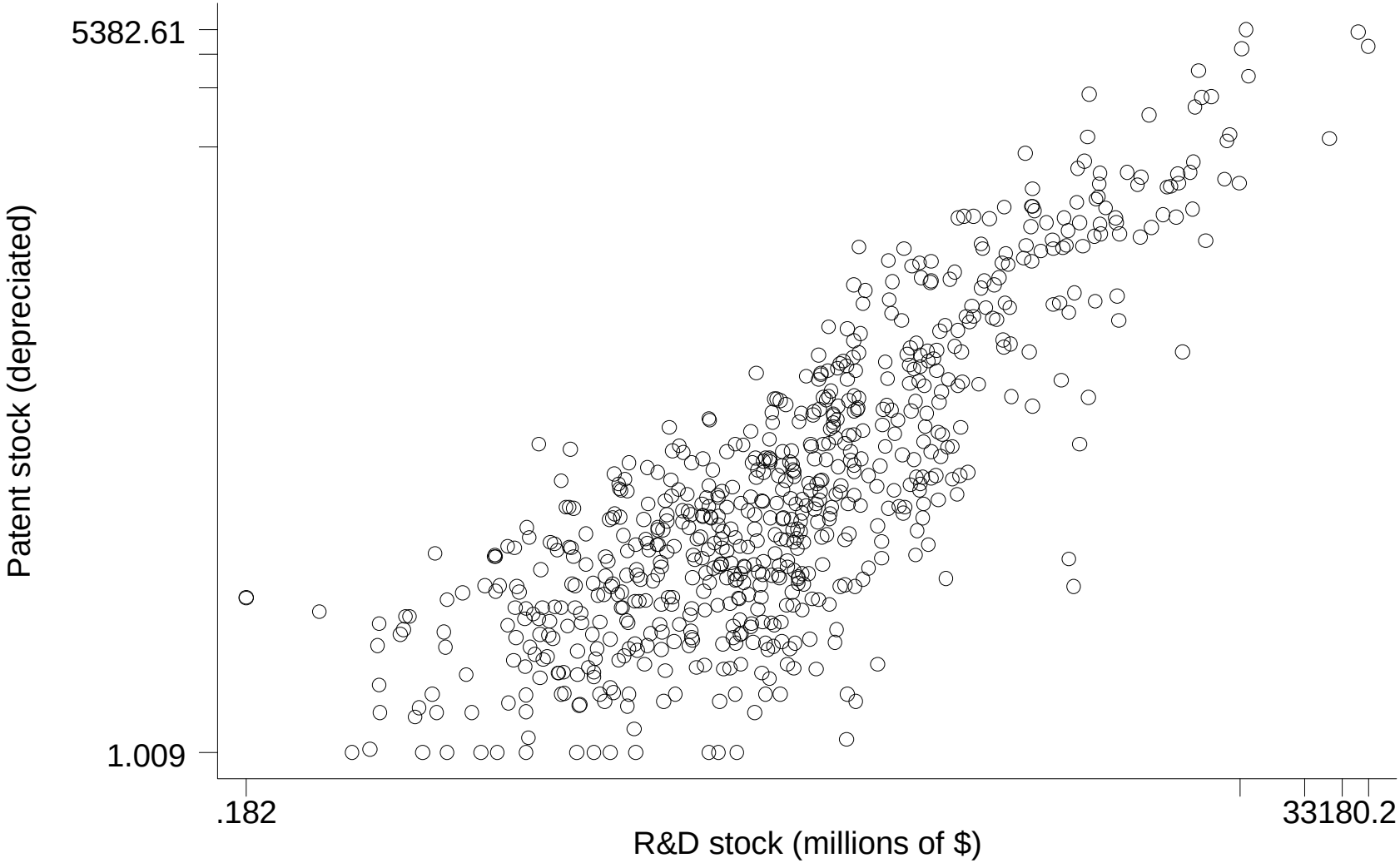
Three sources of bias in patent counts:

1. Differences across countries in economic costs and benefits of patents – rigor of exam; size of market; subject matter coverage
2. Differences among technologies and sectors in the importance of patents as protection against imitation
3. Differences among firms in propensity to patent, especially unimportant innovations; filing under different names

Changes over time



Across firms - patent stock versus R&D stock (log scale) corr = 0.79



US Manufacturing Sector 1992

Some sources of patent data

- NBER patent citations data file for US
- Patstat for worldwide – OECD/EPO/….
- Japanese patent data at IIP
- Chinese patent data – early days
- (Free) online searching:
  - USPTO (detailed status and assignee info in the PAIRS System)
  - EPO Espacenet (for families, equivalents, all docs)
  - Google patents for USPTO, including older pats

NBER Patent Citations Data File

- Available at http://www.nber.org/patents
- ~3 million U.S. patents granted between January 1963 and December 1999 (now updated to 2006)
  - Patent number, application and grant dates
  - Country and state of first inventor (up to 2002)
  - Main US patent class; IPC classes; number of claims
  - Number of citations, forward and backward; generality and originality measures based on citations
- All citations made to these patents between 1976 and 2006 (over 16 million).
- Match of patenting organizations to Compustat (the data set of all firms traded in the U.S. stock market).
  - enables ownership assignment for part of the dataset

PATSTAT

- Worldwide statistical patent database, developed by the EPO in 2005, updated semi-annually.
- Data from the EPO's master bibliographic database, DocDB.
- Bibliographic details on patents filed at 70+ patent offices worldwide, covering 50 million+ documents.
  - claimed priorities, application and publication nos & dates
  - technology classes
  - Inventor and applicant names & addresses
  - title and abstract
  - patent citations and non-patent literature text
- Coverage may be partial/delayed (e.g., US nonpublished apps).

JPO Data

- IIP Patent Database
  - developed in 2006 by the Institute of Intellectual Property of Japan (IIP) and the University of Tokyo. See Goto and Motohashi (2006)
  - http://www.iip.or.jp/e/e_patentdb/
  - Contains information on
    - Applications
    - Grants
    - Applicants
    - Rights holders
    - Citations
    - Inventors

SIPO patent data

- Zhen Lei, Zhen Sun, Brian Wright at ARE Berkeley – comprehensive SIPO data. See

  http://is.jrc.ec.europa.eu/pages/ISG/patents/documents/LeiChinaPatentSystem.pdf

  http://faculty.haas.berkeley.edu/neil_thompson/Innovation_Seminar/papers/patent_subsidy_Zhen.pdf

- Eberhardt & Helmers (2011) have a match to Oriana (Chinese state-owned and private firms) See

  http://www.csae.ox.ac.uk/workingpapers/pdfs/csae-wps-2011-15.pdf

- Unfortunately, no citation data as far as I know

Patents vs patent families

- A patent is a single document with coverage over a specific region (US, EPO designated countries, etc.)
- A patent family is a collection of docs from different patent offices with coverage of the same invention
- BUT….
  - The precise definition and scope of a patent may vary in different regions, so that there can be 2 equivalents in one country to a single patent in another, and more complex possibilities.
  - Leads to multiple definitions of patent families

Some definitions

- Priority patent – the patent application which establishes the date before which the examiner searches for prior art
- Equivalent – a patent in another jurisdiction that names a particular application as the priority application
- Note that
  - Priority patents may have more than one equivalent, even in the same jurisdiction
  - Later patents may have more than one priority

Patent families

- Patent families – collections of equivalent patents
- Patent documents example:
  - Conservative: only D2 and D3 are equivalents
  - Families: D1,D2,D3; D2,D3,D4; D4,D5
  - Extended family: all 5

Application D1	Priority P1		
Application D2	Priority P1	Priority P2	
Application D3	Priority P1	Priority P2	
Application D4		Priority P2	Priority P3
Application D5			Priority P3

What should you use?

- To analyze application, grant, opposition or litigation behavior, the appropriate unit of observation is an individual patent (or patent application)
- To analyze invention, the appropriate unit of observation is a patent family
  - For citations from one patent to another, - requires also consolidation of citations
  - E.g., US patent citing German patent, and the German equivalent citing the same patent is one citation from the family, not two

Measuring innovation using patents

- Schmookler (1960 book) – pioneer in the use of patent statistics
- Scherer's (1960s) work in oil, chemicals, steel
- Griliches et al (1980s) – first large sample work using computerized USPTO data. Conclusions:
  - Patents strongly related to R&D across firms, elasticity close to one
  - Controlling for unobserved differences across firms, elasticity lower (about 0.3)
  - Difficult to determine lag structure – R&D very smooth over time within firm
  - Poisson-type models – patents exhibit overdispersion
  - In the presence of R&D, patents add little explanatory power for sales, profits, market value, etc. Why?

Skewness of the distribution of patent value or importance

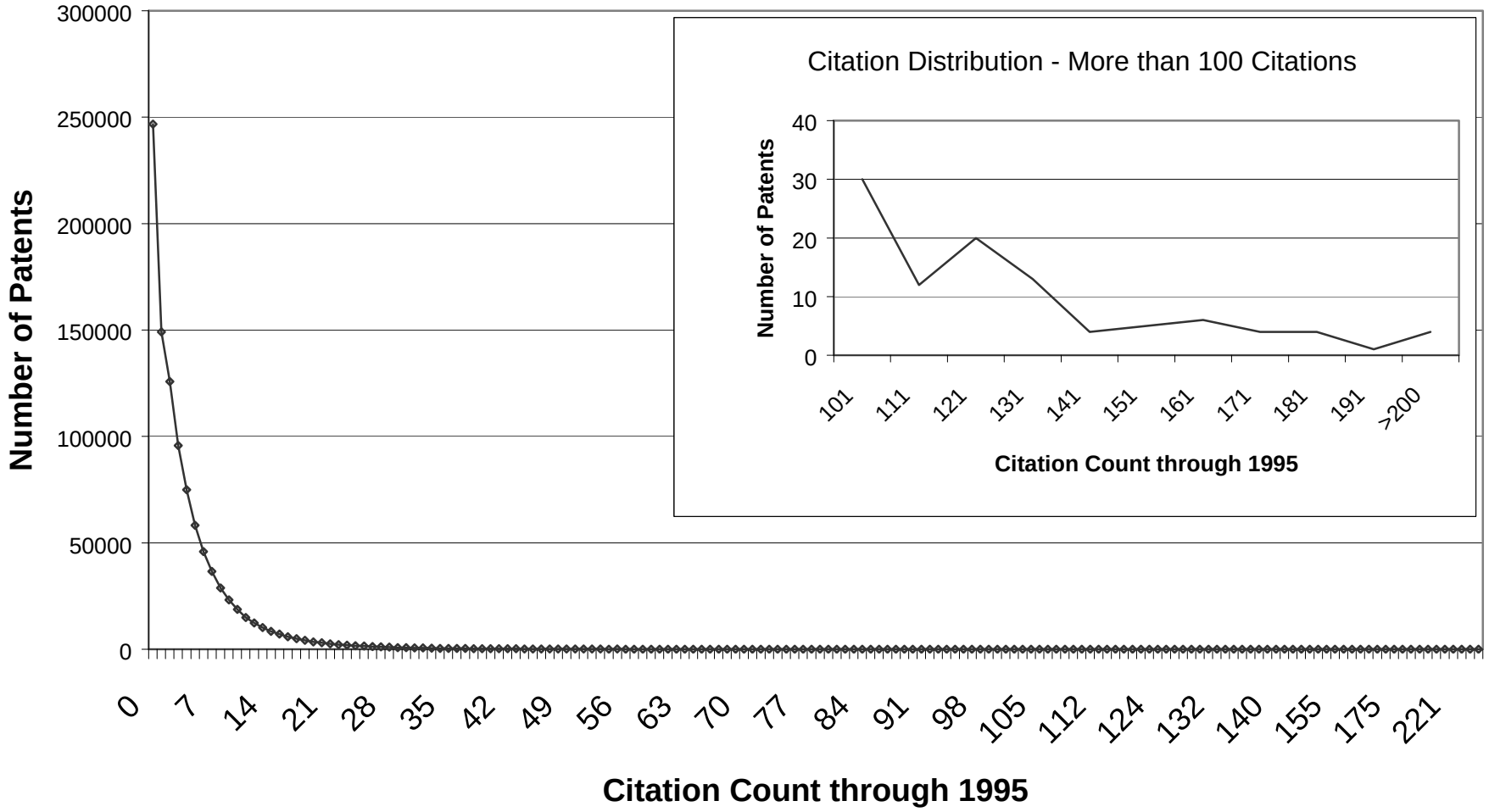
What are patent citations?

- Somewhat like citations in a research paper:
  - References to prior technology, either patents or other scientific literature on which the current patent builds or which it uses
  - Some added by the examiner (the "referee")
  - Some added after the fact (not used by inventor)
  - Some added to avoid infringement (limit scope, defense against suits)
  - Some added for "teaching" (like survey articles)
- EPO differs from the USPTO in citation practice
  - Examiner minimzes the number of cites
  - Most added by examiner
  - Cites are tagged with an indicator of why they are useful
    - Most important are X,Y references
  - Average number is 3 rather than 6-7

Some facts about U. S. citations

- More valuable patents are cited more often
- One quarter of patents receive no citations
- 0.01% receive more than one hundred citations
- Lag distribution is skew to the left with a mode at about 3.5 years. Most cites happen by 10 years, but there can be long lags (30 years)
- Number per patent has increased recently with the advent of computerized search

Figure 3
Citation Distribution



Hall, Jaffe, Trajtenberg

- *Rand Journal of Economics 2005*
- Large firm level study which relates market-book value ratio to
  - Stock of R&D spending
  - Average patent yield per R&D
  - Average cite yield per patent
- Findings
  - Cites per patent are more important than patent yield itself
  - Increase of one cite per patent => increase of 3% in market value
  - Below the median, cites per patent has no effect, but
    - 10% increase in value if cites per patent average 7-10
    - 35% increase in value if cites per patent average 11-20
    - 54% increase in value if cites per patent average above 20
  - Self-cites worth twice as much as other cites (appropriability)
- Timing – do citations received before value is measured matter more or less than those received after?
  - Less, although they are useful for forecasting future cites
  - Predictable and unpredictable citations approximately equal

Other value correlates

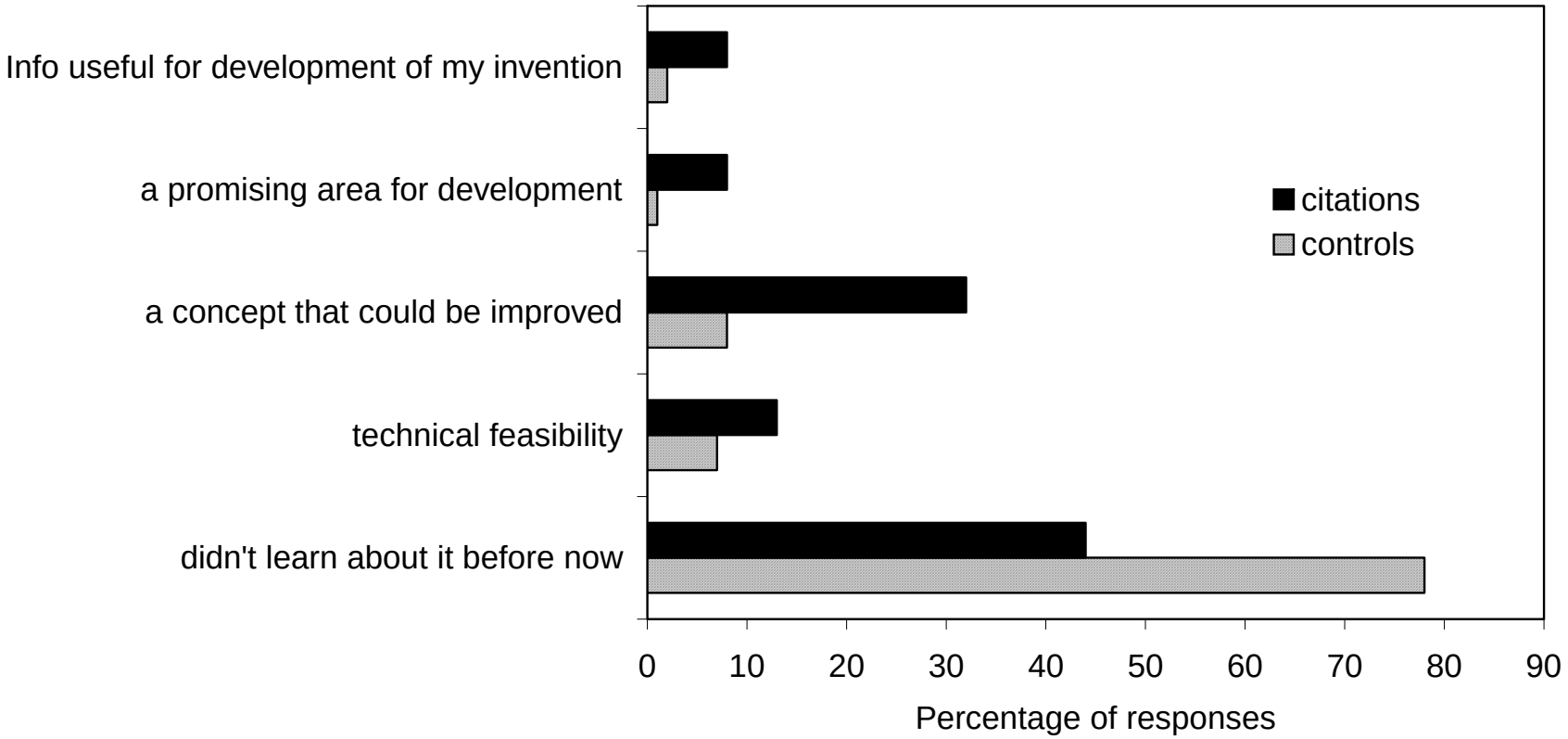
- Opposition or litigation
- Family size
- Backward citations as well as forward
- Claims, in some cases
  - independent claims if available
  - Cites per claim
- Type of citation
  - X and Y more valuable than others (EPO)

Citations as indicators of K flow

- Can they be used in this way?
- Jaffe, Trajtenberg, Fogarty surveyed 1300 inventors (37% response), find
  - About half correspond to some kind of knowledge flow
  - About one quarter to a very substantial flow
  - Remainder are primarily those added by others (not the inventor)

Jaffe, Trajtenberg, Fogarty (2002)

Distribution of answers to:
What did you learn from the previous invention?



Using citations to measure K flow

- "Self" measure in HJT for appropriability
- Geographic localization
  - Henderson, Jaffe, and Trajtenberg
  - Many successor papers
- Branstetter (2000); Macgarvie (2003)
  - Citations used to measure knowledge flow induced by exporting or importing
  - French firms begin exporting to Germany
    - Do they cite German patents more after than before?
- Spillover from alliances?
  - Ham (1997) – Sematech
  - Mowery and coworkers – universities and industry

Citations as measures of K types

- Henderson, Jaffe, Trajtenberg suggested the following measures:
  - Generality – One minus HHI of cites to the patent
  - Originality – One minus HHI of cites from the patent
  - where HHI is computed across technology classes
- Problems
  - Defining appropriate classes (they used US system, which is not ideal)
  - Not all classes are equidistant from each other
  - Small numbers - bias correction is easy (see Hall 2005), but it still means measurement is noisy

Newer work

- Jones & Uzzi – define a radical scientific paper as one that combines citations that are rarely seen together
  - Interesting to use this idea with patent data
- Gorodnichenko, Hall, Roland – work in progress using a refined measure of originality that constructs weights for technology distance
  - Idea is to relate indivdualistic culture to greater originality or radicalness in invention
  - First results are promising

Conclusions

- Patents as indicators
  - Can be useful, especially citation-weighted – correlated with value, R&D, litigation, profits, etc.
  - However, important, especially over time, to understand the impact of policy changes on these indicators.
- Citations
  - Defensible as a partial measure of knowledge transfer
  - Suggest spillover localization in region and country, or via contact
  - Work on richer citation measures continuing

Data needs

- Major patent offices have put an enormous amount of data online, *but*
  - more suited to search than statistical analysis
  - researchers need to download large blocks of data
  - ftp access desirable

Data needs

Two major problems for research:

1. Inconsistent assignee names, and no common register of assignees (even within POs)

   Name harmonization projects at KU Leuven, OECD, HBS, etc.

2. Classification by industry, which needs to be done by patent, not by tech class

   Lybbert-Zolas paper at WIPO – uses text analysis and keywords to allocate patents to industries with probability; data available online. See:

   http://www.wipo.int/export/sites/www/econ_stat/en/economics/pdf/Working_Paper_No._5_Lybbert.pdf

Some surveys available

- Basberg (1987), "Patents and the Measurement of Technological Change: A Survey of the Literature," *Research Policy.*
- Pavitt, Keith (1988), "Uses and Abuses of Patent Statistics," A. F. J. van Raan (ed). *Handbook of Quantitative Studies of Science and Technology.* Amsterdam: Elsevier Science Publishers.
- Griliches (1990), "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature.*
- Nagaoka, Motohashi, and Goto (2010), Patent Statistics as an innovation indicator, in Hall and Rosenberg (eds.), *Handbook of Economics of Innovation*