FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

Patent Data: New Metrics and New Linkages

or

How can we be more clever in using our data?

May 2016

Lee Fleming

This work is supported by NSF SCISIP grant #1360228, the Kauffman Foundation, the US Patent and Trademark Office, the American Institutes for Research, and the Coleman Fung Institute for Engineering Leadership. I would like to thank co-authors Gustavo Manso, Benjamin Balsmeier, Olav Sorenson, Guan-Cheng Li, Sonja Lueck, Sandy Yu, and Tristan Fitzgerald

My earlier research on inventors and networks fits with this session…

…however, today I'm going to talk metrics and provide examples of how new computational approaches might squeeze new insight from old - and new - data

Underlying themes

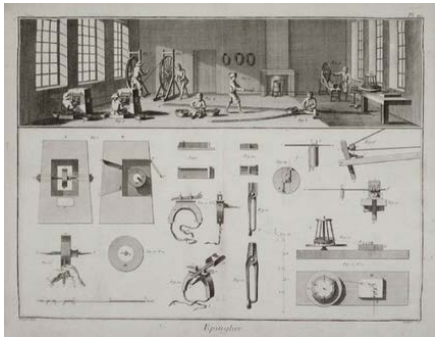
- Patent data are over-used and abused
- In particular, we need to stop relying solely on patent counts and citations to measure innovation
  - Richer measures are actually not that hard to calculate
- Advances in machine learning and natural language processing useful, though need thoughtful application
- Newly available data and tools provide opportunity

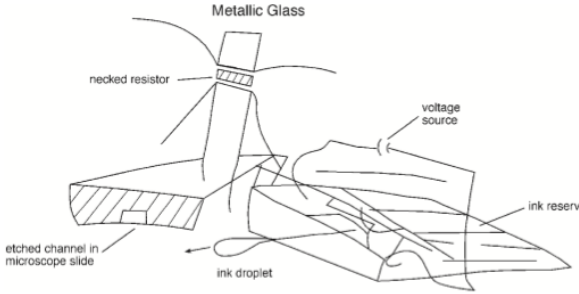
FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

RQ #1: What happens when you change fields?

"The man who employs either his labour or his stock in a grater variety of ways than his situation renders necessary…may hurt himself, and he generally does so…Jack of all trades will never be rich, says the proverb."

Adam Smith, Book IV, Chapter V, p. 563.





"Almost always the men who achieve these fundamental inventions of a new paradigm have been either very young or very new to the field whose paradigm they change."

Thomas Kuhn, The Structure of Scientific Revolutions, pp. 89-90.

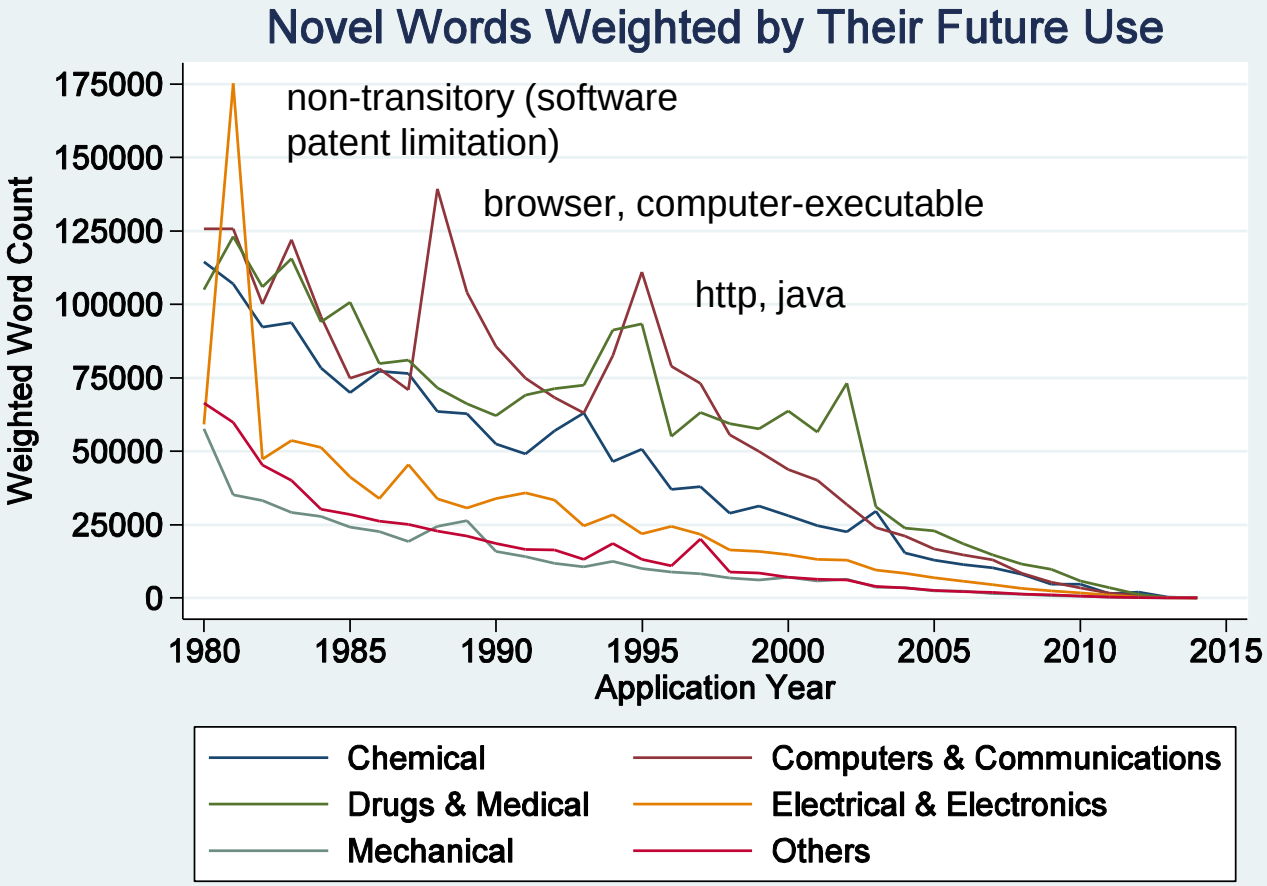
Intuitive answer to this question – if separate measures of novelty vs. value

Simple prediction

- You change fields, you invent patents that are
  - more novel
  - less valuable
- Citations have been used to measure value *and* novelty
  - so that's not going to work…
- Methods aside:
  - use change in noncompete labor law as instrument
    - people under noncompetes must move further from old field
  - disambiguation is a classic machine learning problem

FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

Use patent renewal as measure of value – but how do you measure novelty?



Look for first time word appears in entire patent corpus: changing fields => +54% increase in new words and -35% decrease in renewals

RQ #2: How does governance influence innovation? (Forthcoming *JFE*)

- High-profile scandals transformed corporate governance
  "…the Enron Board of Directors failed to safeguard Enron shareholders and contributed to the collapse of the seventh largest public company in the United States…" *Forbes 2001*
- 2002: SOX requires majority of independent directors
- Consensus: board oversight has increased

…but what's the impact of monitoring on innovation search strategy and innovation?

- Stronger governance could increase innovation focus and effort

  …yet also inhibit creativity and risk taking
- Conflicting results from finance papers using patents and citations
- What does a patent really measure?
  - turns out, you can break it down surprisingly cleanly between axes of exploration and exploitation

FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

Very consistent signal: SOX => exploitation

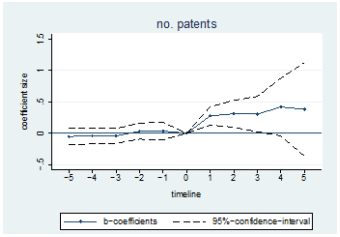
(and most of these measures available at time of issue)

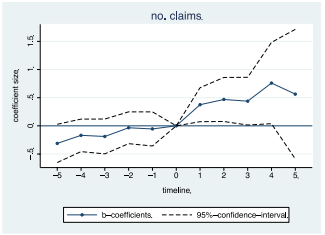
<u>These go up:</u>

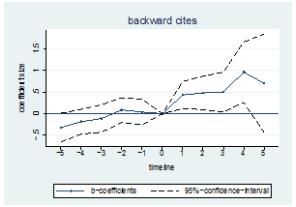
patents
patents in middle of cite distn
claims
backward cites
self cites
patents in old to firm classes

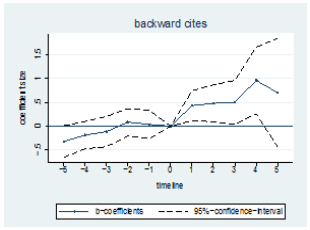
<u>These go down or don't change:</u>

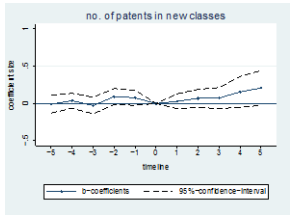
patents in bottom of cite distn
patents in top of cite distn
patents in new to firm classes
Jaffe distance measure











Reliance upon average counts and cites would have missed lack of exploration

- Results are very consistent across a variety of easily observed – and correlated - measures
- Can we develop canonical measures – and instruments - of exploration and exploitation?
  - Run a simple principle components analysis

FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

RQ #3: Where do these correlated measures fall out in a PCA? (Application still under construction)

Tables 4 and 5 – Principal Component Analysis

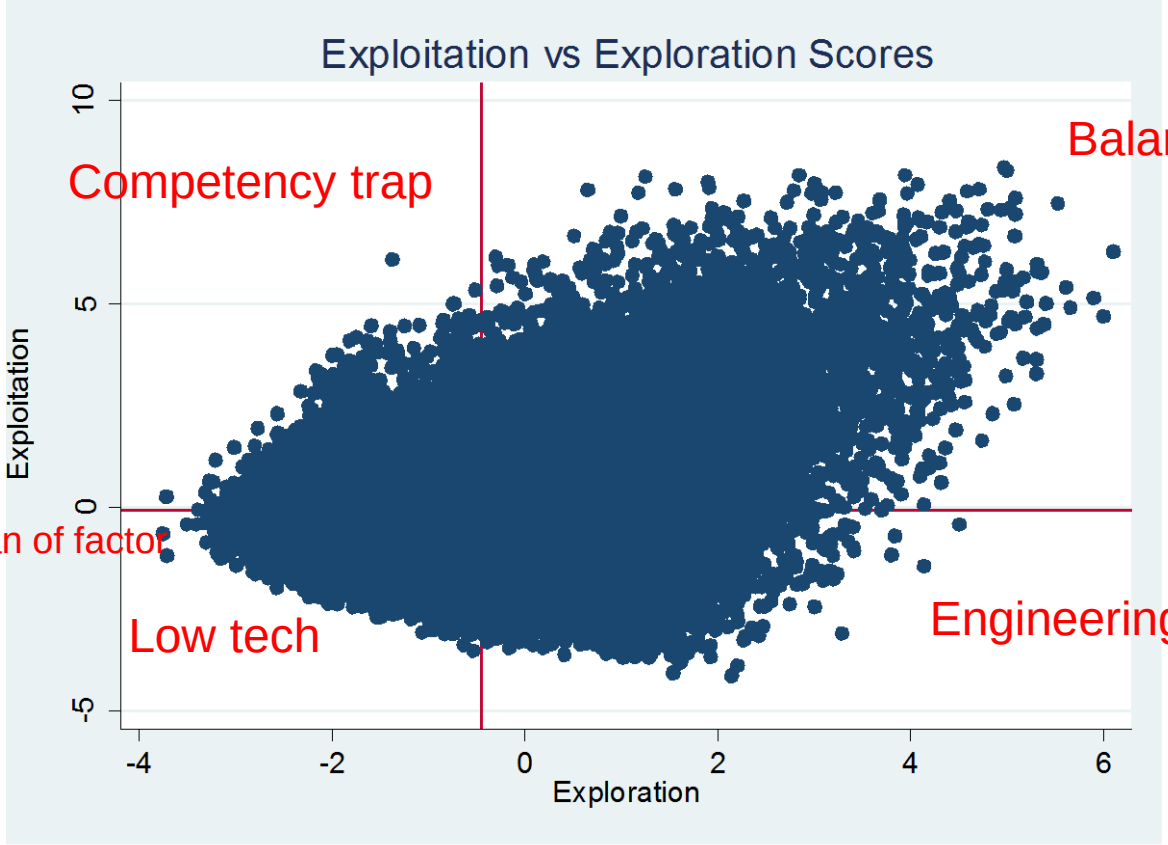
Component	Variance	Difference	Proportion	Cumulative
Comp1	4.02	1.72	0.50	0.50
Comp2	2.30		0.29	0.79

Notes: This table reports the results of a Principal Component Analysis after Varimax Rotation. Only components with Eigenvalues above one are extracted. The 8 variables that entered the PCA are: new classes entered, patents in new/known classes, technological proximity, av. age of inventors, backward citations, self-citations, and claims; all variables log-transformed.

Variable	Comp1	Comp2	Unexplained
New tech classes entered		0.58	0.08
Patents in new classes		0.58	0.08
Patents in known classes	0.45		0.09
Technological proximity	0.39	-0.38	0.47
Backward citations	0.41		0.10
Self-citations	0.45		0.16
Claims	0.41		0.09
Av. age of inventors	0.31	-0.37	0.60

Notes: This table reports the results of a Principal Component Analysis after Varimax Rotation. Only components with Eigenvalues above one are extracted. All variables log-transformed. Variable definitions provided above.

Figure 1 – Scatter Plot of PCA scores

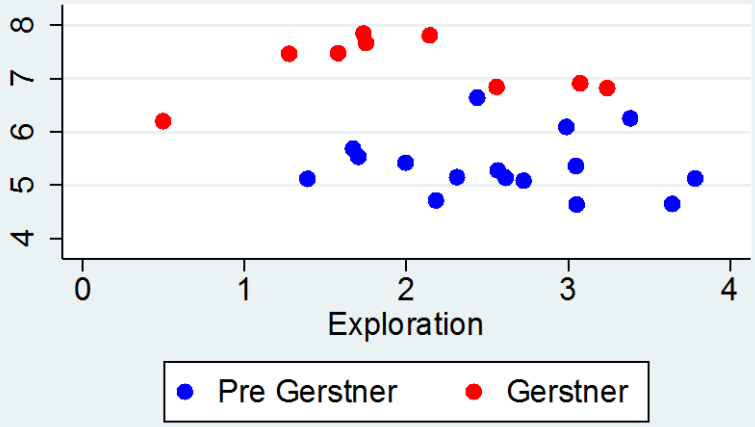


Notes: This graph plots the component scores of 'Exploration' and 'Exploitation' extracted from the Principal Component Analysis shown above. Red lines mark the median values of each factor. 19% of the observations are each in the upper left and lower right quadrants, 31% in each of the other quadrants.

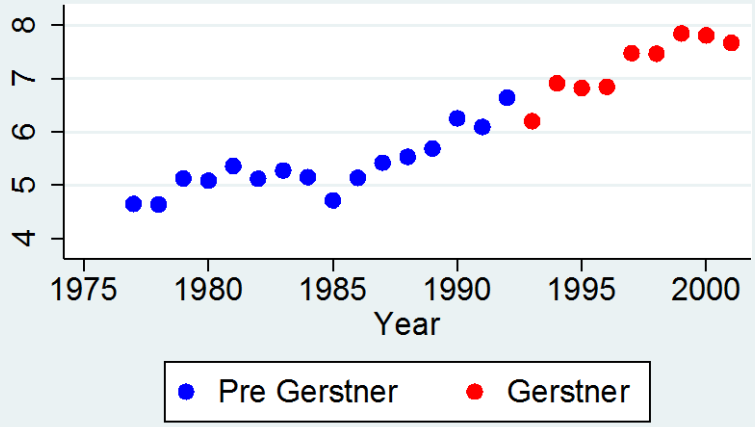
Even better! Firms respond to exogenous shocks along these dimensions (SOX, MARA, Anti-trust, banking deregulation)

IBM

Exploration vs Exploitation

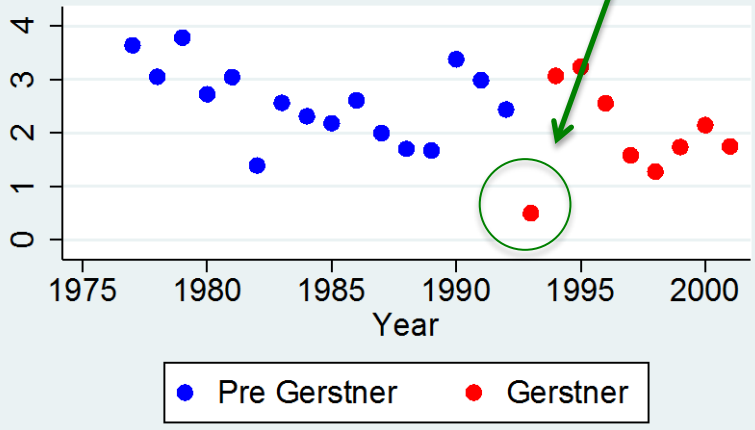


Exploitation

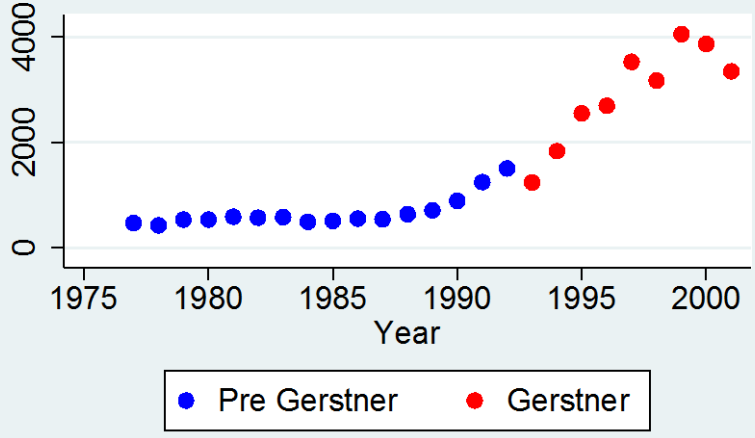


IBM's near death experience

Exploration

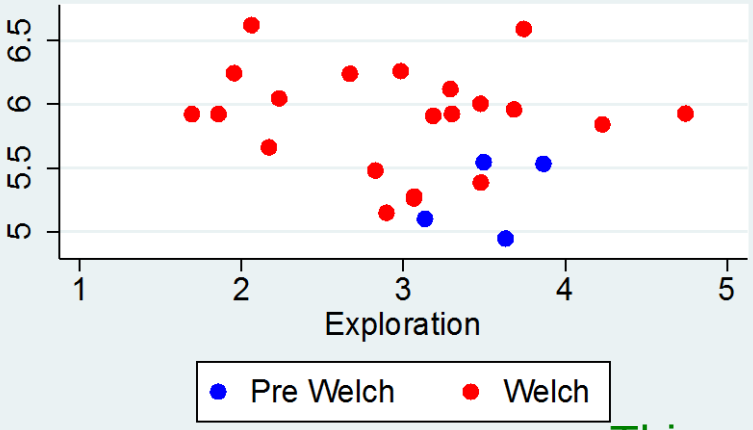


Patents

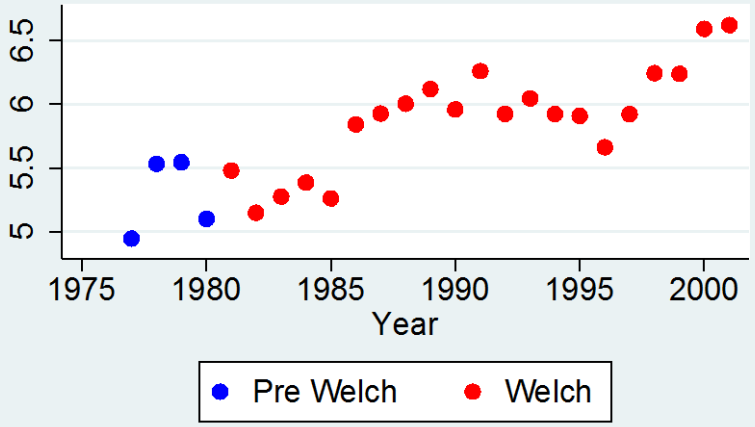


GE

Exploration vs Exploitation

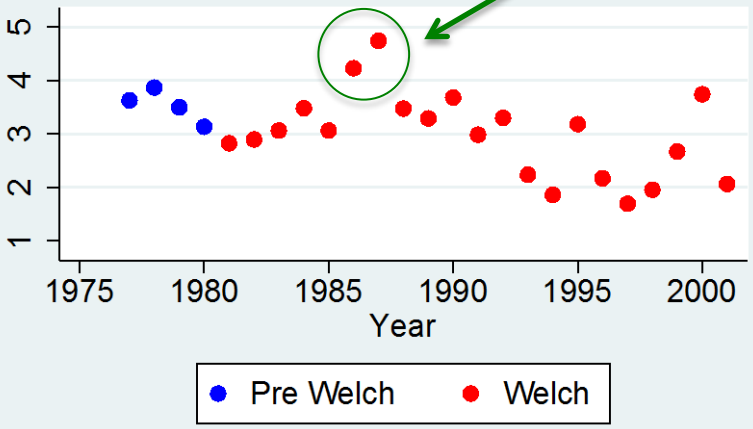


Exploitation

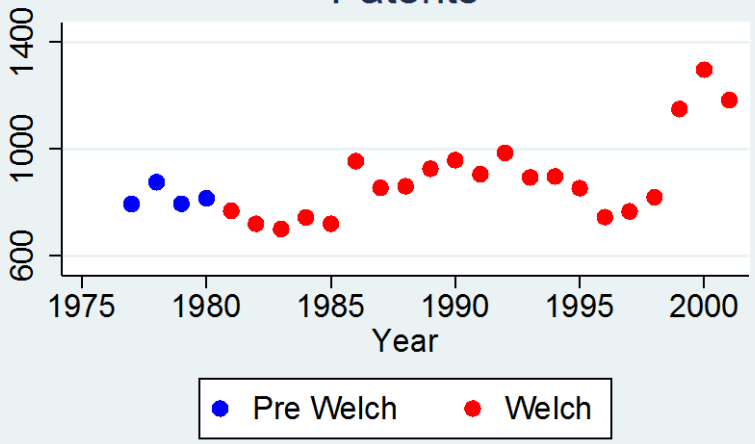


This appears to be pretty rare

Exploration

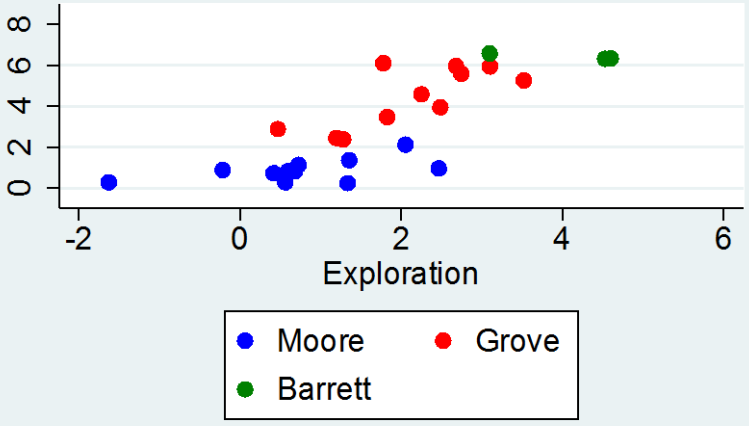


Patents

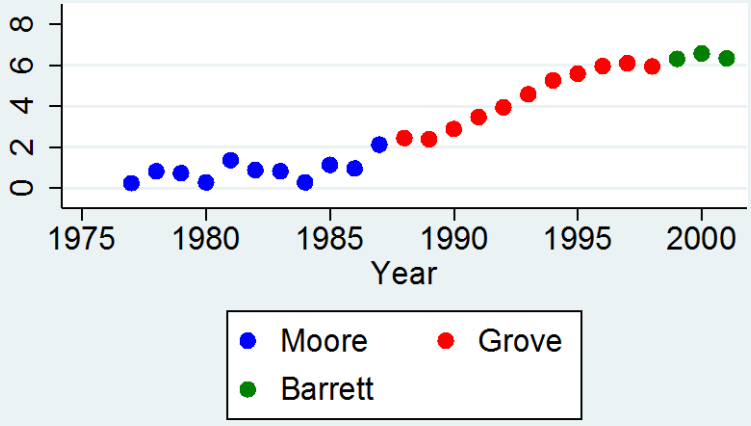


Intel

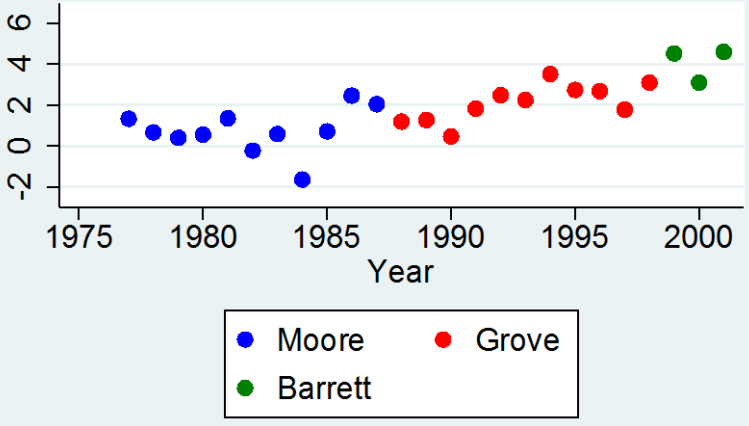
Exploration vs Exploitation



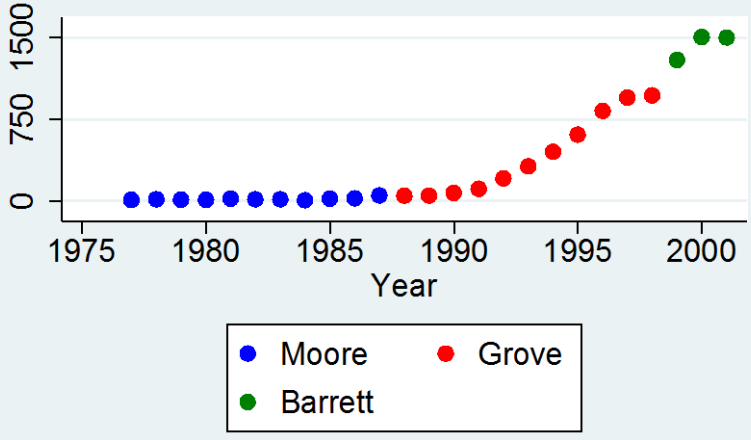
Exploitation



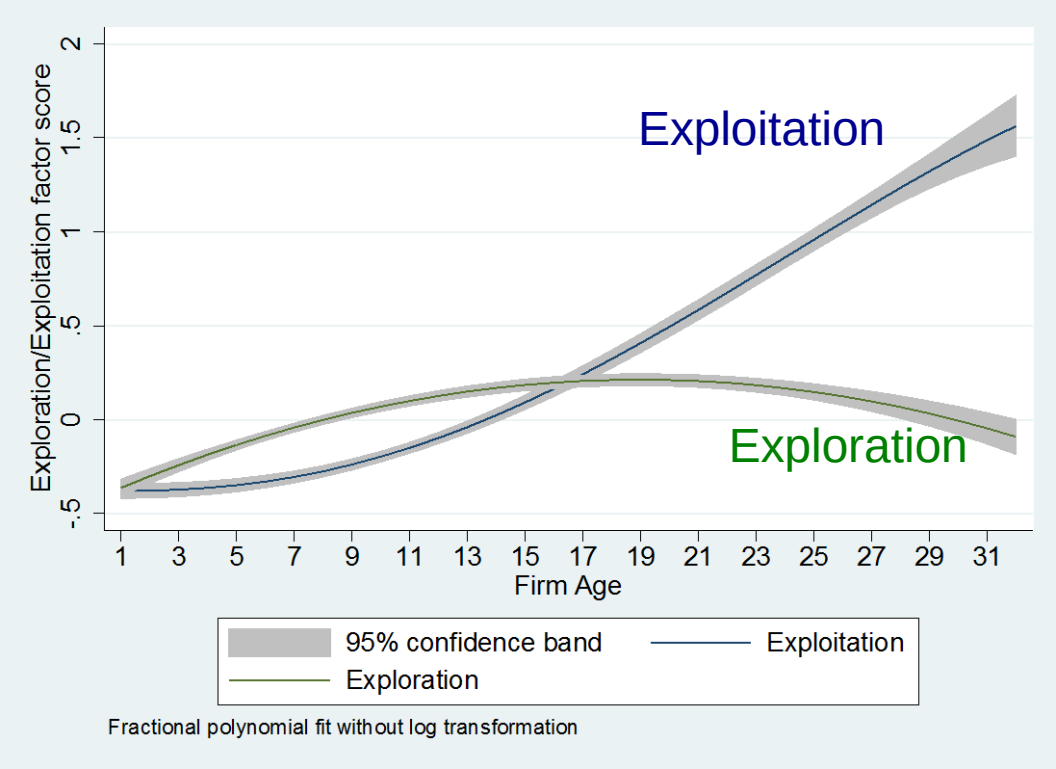
Exploration



Patents



An inexorable exploitation path dependence (except of course for Intel)?



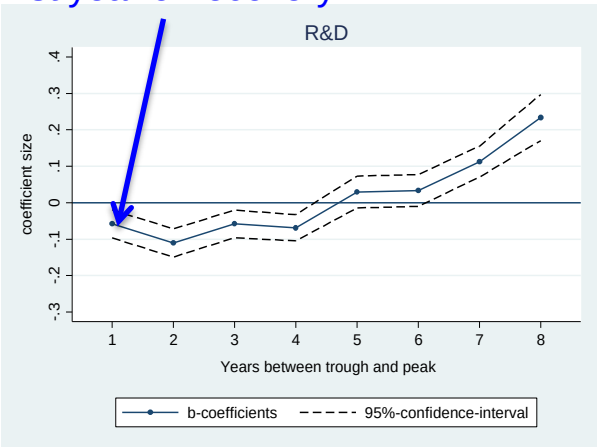
Also looks like exploration precedes new market entry and productivity improvements

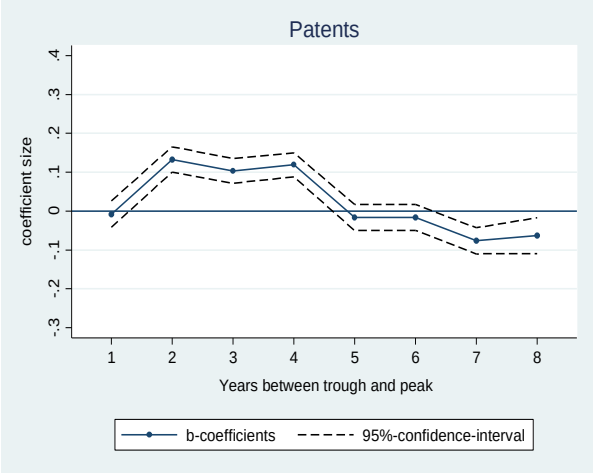
RQ #4: Can we see these components in the economy?

- (Simplified) conundrum:
  - Innovation should concentrate in recessions
    - Lower opportunity costs due to weaker sales
  - R@D and patenting data not clear

- Gustavo's model:
  - *Explorative* innovation should concentrate in recessions
    - Lower opportunity costs because sales of extant technology and products are weaker

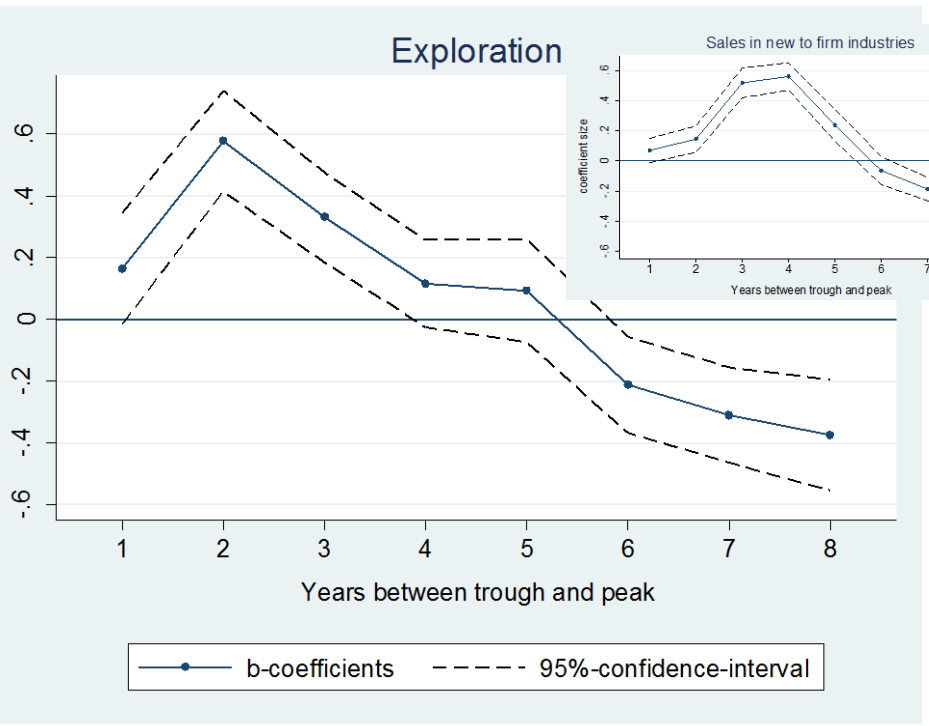
First year of recovery

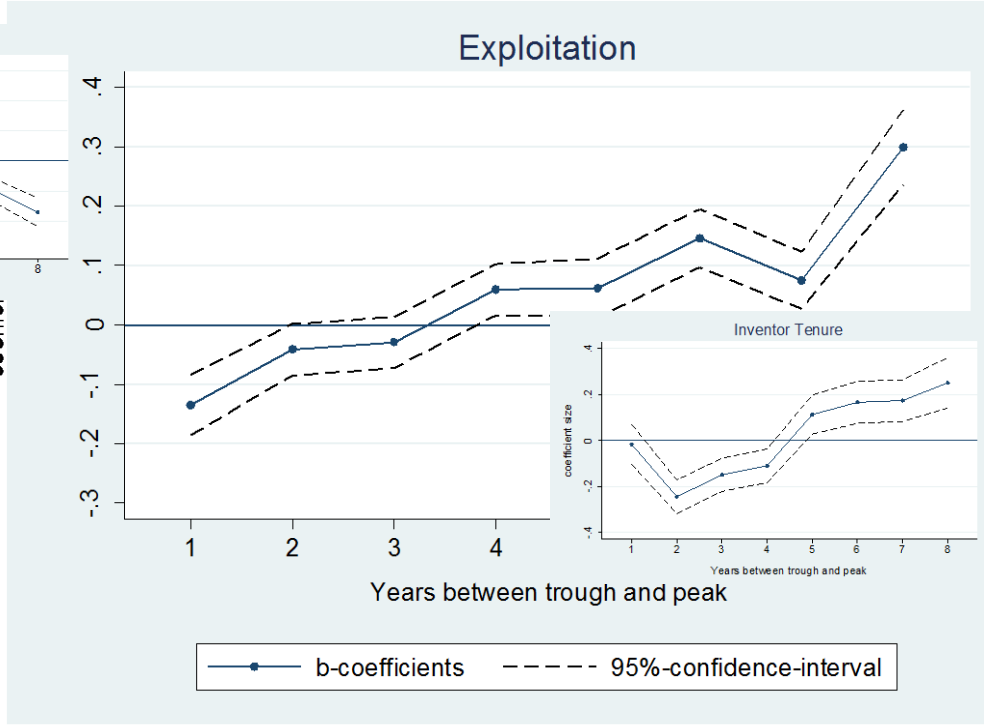




We find

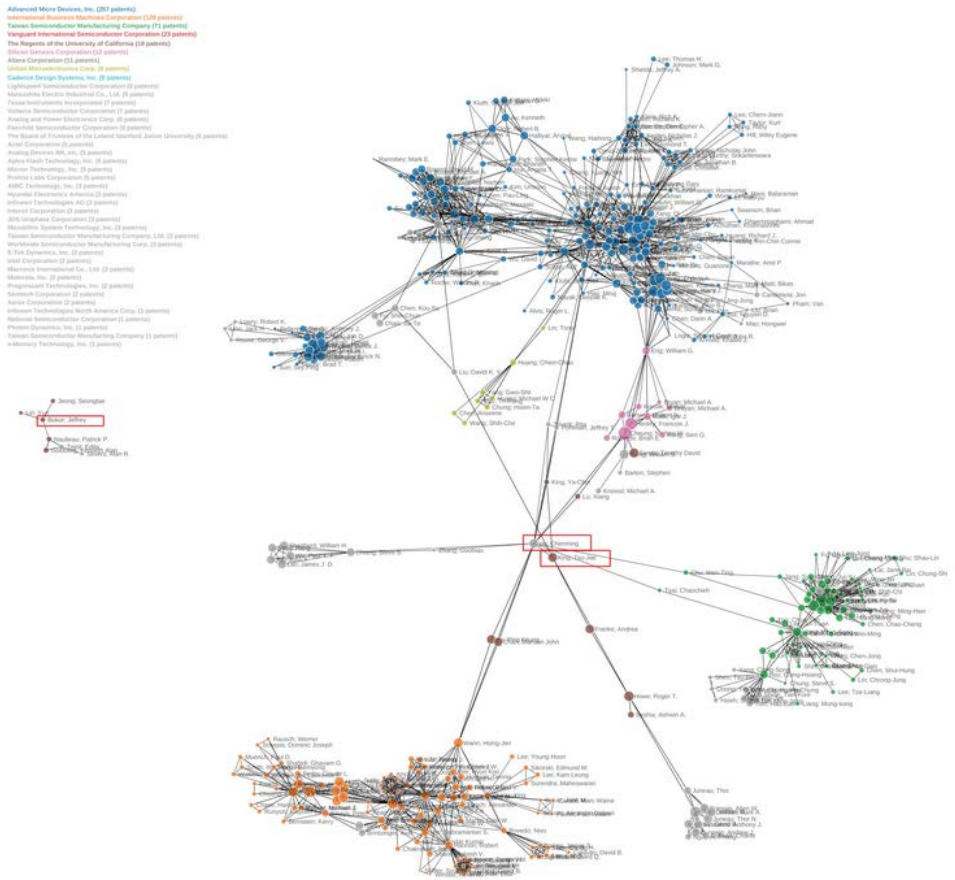
- Exploration is counter-cyclical
  - New market entry lags exploration
- Exploitation is pro-cyclical
  - Inventor tenure follows exploitation



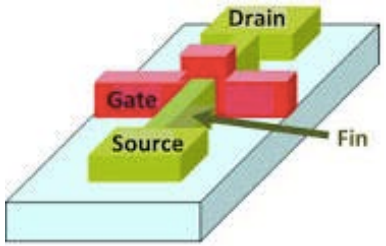


FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

New tools: real time inventor network rendering



The FinFET breakthrough



1999: First P-channel FinFETs

X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," *IEEE International Electron Devices Meeting Technical Digest*, pp. 67-70, 1999

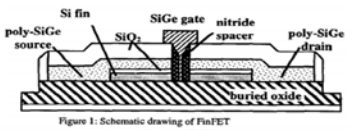
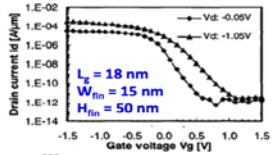
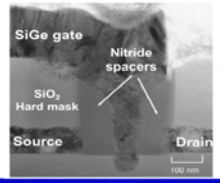
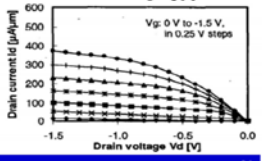


Figure 1: Schematic drawing of FinFET





Transmission Electron Micrograph



FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

Linking new data: blocking actions and crowdfunding

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

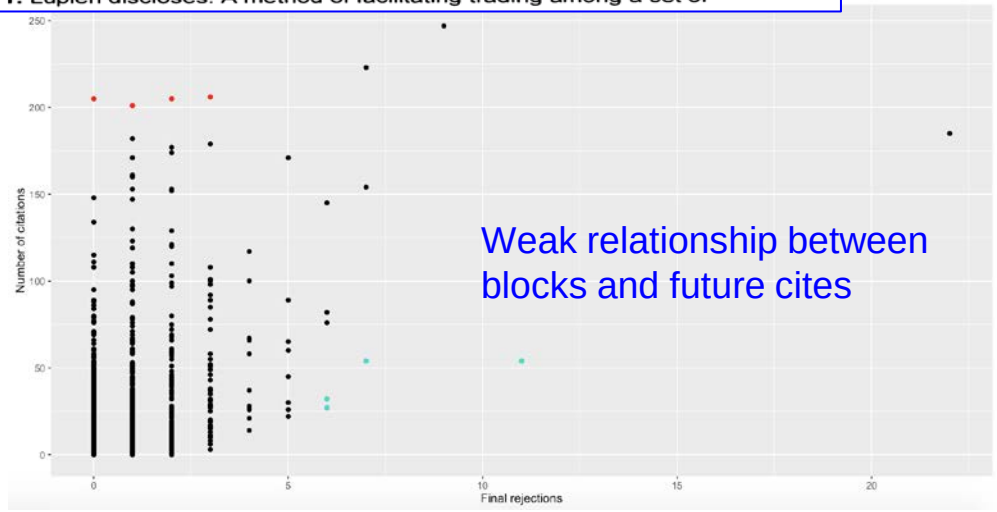
A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2. Claims 1-26 are rejected under 35 U.S.C. 102(b) as being anticipated by Lupien et al. (Hereafter Lupien, US PAT 5,101,353).

Re Claim 1: Lupien discloses: A method of facilitating trading among a set of

Campaign descriptions offer a wonderful window into non-patent innovation: ripe for NLP



Weak relationship between blocks and future cites

Is a block a measure of novelty or value – or just a well written patent?

Meta themes

- Patent data are (almost) dead. Long live patent data.
  - stop relying only on counts and citations to measure innovation
  - please help Alan Marco and friends at USPTO
- Advances in machine learning and natural language processing useful, though need thoughtful application
  - crowdfunding is fertile area
- Newly available data and tools provide opportunity
  - thank you to Kauffman, SCISIP, and others! please don't stop funding data and tools infrastructure!