Exploration and Exploitation in US Technological Change

Vasco M. Carvalho[a] **Mirko Draca**[b c] Nikolas Kuhlen[a, d]

[a] *University of Cambridge*

[b] *University of Warwick*

[c] *CAGE Research Centre*

[d] *The Alan Turing Institute*

October 20, 2022

Measuring Technological Change?

Three main approaches (Griliches, 1990):

- **Outputs**: Patents, scientific papers. Modified with citation counts to measure 'quality'.
- **Inputs**: R&D spending, employment of scientists and engineers, investments in technological capital and 'intangible' capital.
- **Residuals**: TFP estimated for production function methods.

These methods face challenges when it comes to understanding qualitative changes in technology and rates of 'progress' ...

Contents lists available at ScienceDirect

Explorations in Economic History

journal homepage: www.elsevier.com/locate/eeh

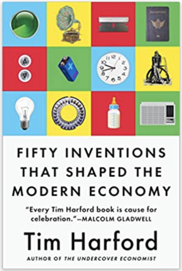
Declining American economic growth despite ongoing innovation☆

Robert J. Gordon[a,b,*]

[a] *Northwestern University, USA*
[b] *NBER, USA*

1. Introduction

A central result of economic history is that there was very little economic growth prior to the First Industrial Revolution of the late eighteenth century. For instance, Stephen Broadberry and his co-authors (2015) after extensive research have documented a growth rate of per-capita income in Britain from 1270 to 1700 of only 0.2% per annum. Beginning around 1750 the rate of economic growth accelerated gradually for a variety of reasons that are concisely explained in Joel Mokyr's (2018) companion paper in this volume. Because economic historians usually define their subject matter as economic behavior prior to the most recent forty or fifty years, they have understandably devoted less attention to the deceleration of economic growth in the developed world that began in the 1970s and has become more pronounced in the past decade.



Technological Change & ‘Innovation’.

- Another challenge here is linking technological change to the process of innovation.

- We know that innovation is about trial & error and experimentation. This is complemented by phases of consolidation where we double down on successful experiments.

- Further to this, there is qualitative change in what is studied. Ideas enter, rise and fall. There are ‘waves’ in the development of ideas.

Technological Change...and Text

- Text data offers the chance to try and capture what was previously thought of as qualitative change and how it relates to ‘progress’.

- A key recent contribution is Kelly, Papanikolaou, Seru, and Taddy (2020) who identify ‘breakthrough patents’. In short, these are defined as the patents that were the first to use sets of words that became much more common later on.

- Approach is based on a combination of ‘backward IDF’ and massive scale cosine similarity across patents. It makes progress in identifying ‘regime shifts’ in technological development.

Tesla’s Electric Motor Patent (1888).

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y., ASSIGNOR OF ONE-HALF TO CHARLES F. PECK, OF ENGLEWOOD, NEW JERSEY.

ELECTRO-MAGNETIC MOTOR.

SPECIFICATION forming part of Letters Patent No. 382,279, dated May 1, 1888.

Application filed November 30, 1887. Serial No. 256,561. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a subject of the Emperor of Austria, from Smiljan, Lika, border country of Austria-Hungary, now residing at New York, in the county and State of New York, have invented certain new and useful Improvements in Electro-Magnetic Motors, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

In a former application, filed October 12, 1887, No. 252,132, I have shown and described a mode or plan of operating electric motors by causing a progressive shifting of the poles of

sulated sections, so as to be susceptible to rapid variations of magnetism. This core is wound with four coils, C C C′ C′, the diametrically-opposite coils being connected in the same circuit, and the two free ends of each pair being brought to the terminals t and t', respectively, as shown. Within this annular field-magnet A is mounted a soft-iron cylinder or disk, D, on an axis, a, in bearings b b, properly supported by the frame-work of the machine. The disk carries two coils, E E′, of insulated wire, wound at right angles to one another, and having their respective ends joined, so that each coil forms a separate closed circuit.

What We Do.

- Innovation measure derived from the probabilistic ‘text information’ in patent documents.

- How different is patenting this year compared to previous patenting? This will be measured in terms of informational ‘bits’. That is, the change in the probability distribution of topics.

- Periods where there is a lot of change can be seen as phases of ‘exploration’, while steady periods are ‘exploitation’. Crucially, this means that lots of patenting $\neq$ lots of ‘innovation’ (necessarily).

A Sketch of the Approach.

- Get the USPTO data from the 1920s onwards. Run a separate topic model per firm over the whole time period. This gives you a distribution of topic shares over firm-year observations..

- Plug these topic shares into a KL 'Bayesian Surpise' measure. This will pick up the extent to which today's topic share distribution is different to the past.

- Big shifts are 'exploration' while steady behavior is 'exploitation'. Implicit, stylised model is one of firms innovating aggressively and then hitting on 'cash-cows'.

We're adapting a literature outside econ that's emerged recently ...

ELSEVIER

Cognition
Volume 159, February 2017, Pages 117-126

Original Articles

Exploration and exploitation of Victorian science in Darwin's reading notebooks

Jaimie Murdock [a, b], Colin Allen [a, c, d], Simon DeDeo [a, b, e, f]

Show more

Share Cite

https://doi.org/10.1016/j.cognition.2016.11.012

Get rights and content

Highlights

- Cognitive study of Darwin's reading behavior via information foraging framework.
- Shows the existence of multiple timescales for exploration/exploitation behavior.
- Proposes framework for contrasting individual and collective behavior.
- Identifies shifts from exploitation to exploration in Darwin's reading.
- Darwin's reading order is more exploratory than the culture's publication order.

ELSEVIER

Vision Research
Volume 49, Issue 10, 2 June 2009, Pages 1295-1306

Bayesian surprise attracts human attention

Laurent Itti [a], Pierre Baldi [b, 1]

Show more

Share Cite

https://doi.org/10.1016/j.visres.2008.09.007

Get rights and content

Abstract

We propose a formal Bayesian definition of surprise to capture subjective aspects of sensory information. Surprise measures how data affects an observer, in terms of differences between posterior and prior beliefs about the world. Only data observations which substantially affect the observer's beliefs yield surprise, irrespectively of how rare or informative in Shannon's sense these observations are. We test the framework by quantifying the extent to which humans may orient attention and gaze towards surprising events or items while watching television. To this end, we implement a simple computational model where a low-level, sensory form of surprise is computed by simple simulated early visual neurons. Bayesian surprise is a strong attractor of human attention, with 72% of all gaze shifts directed towards locations more surprising than the average, a figure rising to 84% when focusing the analysis onto regions simultaneously selected by all observers. The proposed theory of surprise is applicable across different spatio-temporal scales, modalities, and levels of abstraction.

Growth through Heterogeneous Innovations

Ufuk Akcigit

University of Chicago, National Bureau of Economic Research, and Centre for Economic Policy Research

William R. Kerr

Harvard University and National Bureau of Economic Research

We build a tractable growth model in which multiproduct incumbents invest in internal innovations to improve their existing products, while new entrants and incumbents invest in external innovations to acquire new product lines. External and internal innovations generate heterogeneous innovation qualities, and firm size affects innovation incentives. We analyze how different types of innovation contribute to economic growth and the role of the firm size distribution. Our model aligns with many observed empirical regularities, and we quantify our framework with Census Bureau and patent data for US firms. Internal innovation scales moderately faster with firm size than external innovation.

ORGANIZATION SCIENCE
Vol. 2, No. 1, February 1991
Printed in U.S.A.

EXPLORATION AND EXPLOITATION IN ORGANIZATIONAL LEARNING*

JAMES G. MARCH

Graduate School of Business, Stanford University, Stanford, California 94305

This paper considers the relation between the exploration of new possibilities and the exploitation of old certainties in organizational learning. It examines some complications in allocating resources between the two, particularly those introduced by the distribution of costs and benefits across time and space, and the effects of ecological interaction. Two general situations involving the development and use of knowledge in organizations are modeled. The first is the case of mutual learning between members of an organization and an organizational code. The second is the case of learning and competitive advantage in competition for primacy. The paper develops an argument that adaptive processes, by refining exploitation more rapidly than exploration, are likely to become effective in the short run but self-destructive in the long run. The possibility that certain common organizational practices ameliorate that tendency is assessed.
(ORGANIZATIONAL LEARNING: RISK TAKING; KNOWLEDGE AND COMPETITIVE ADVANTAGE)

What We Find

- Method picks up sensible qualitative regime shifts. Key case study will be IBM and analogue-to-digital shift.

- Developmental S-shape for large, long-lived firms. Exploration tapers with age but size still grows. Also: exploration correlates with size measures over and above the effects of patents and R&D.

- Challenges: (1) relating this to a model of firm growth (2) developing aggregate measures, including by technology class.

Data

Patent Data

- Constructed from Iaria, Schwarz, and Waldinger (2018), Bergeaud, Potiron, and Raimbault (2017) and Google webscraping.
- Use abstracts or 'pseudo-abstracts' (first 300 words).
- Standard text pre-processing.

Firm Data

- CRSP and Compustat data. Use match from Kogan et al (2016) QJE.

Final Data Set

- Contains 1,830 unique firms and 27,760 firm years from 1920 to 2004. Key point is the size of the patent portfolio. We require 11 years of pateting to estimate one of our measures (successful exploration).

Basic Ingredients

- We put the text into firm-year cells and run topic models. We run a topic model for each firm across the firm's lifetime where the number of topics k is set to 50, 100 and 150 topics for firm corpora consisting of more than 100, 1000 and 10000 patents, respectively, and 10 topics for smaller corpora.

- This gives us a distribution of topic shares across topics (denoted θ_k) per firm year. Intuitively, this will pick up how firms transition across different technological topics (eg: analogue to digital) over their lifetime.

- We then plug these topic shares θ_k into a 'Bayesian Surprise' measure. Idea is to develop a measure of how the probability distribution wrt topic structure is changing. This will be measured in 'bits'.

Bayesian Surprise

- In a Bayesian framework, uncertainty is expressed in terms of prior and posterior beliefs. Thus, surprise is related to the expectations of the observer and thus relative.

- An observer may experience varying amounts of surprise at different points in time. Naturally, they update their information.

- 'Bayesian Surprise' (Itti and Baldi, 2009) is hence defined as the difference between an observer's prior and posterior beliefs. We'll express this in terms of two different probability distributions.

Kullback-Leibler Divergence

- Formally, this is measured as the KL divergence from a prior distribution q to posterior distribution p:

$$D_{\mathrm{KL}}(p||q) = \sum_{i=1}^{N} p(x_i)\log_2 \frac{p(x_i)}{q(x_i)}.$$

- Rewriting the above equation yields

$$D_{\mathrm{KL}}(p||q) = \sum_{i=1}^{N} p(x_i)\left[\log_2 p(x_i) - \log_2 q(x_i)\right].$$

  ie: Expectation of the log difference between the prior and the posterior.

- Practically, we define our different $p(\cdot)$ and $q(\cdot)$ according to firm topic share distributions ...

Basic Exploration Measure

- We define our exploration measure as

$$\eta_t := D_{\mathrm{KL}}\left(\theta_t \middle\| \bar{\theta}_{-t}\right),$$

 where

$$\bar{\theta}_{-t} = \frac{1}{t-1}\sum_{j=1}^{t-1}\theta_j$$

 denotes the average topic distribution up until year t. The topic distribution θ_t for each year t is based on the collection of all documents filed by the firm in a given year.

- This goes into the previous KL formula and we get a measure of change in units of 'bits'.

Cumulative Exploration

- In addition to the above flow measure of exploration, we also compute cumulative exploration or the 'exploration stock' in year t defined as

$$H_t := \sum_{t=1}^{t} \eta_j.$$

- Allows us to track different phases of exploration over a firm's lifetime. Specifically, how exploration varies with firm age.
- But exploration isn't necessarily always 'good'. Experiments can fail. We try to capture this via 'surprise asymmetry'.

Successful Exploration

- Following the Barron, Huang, Spang, and DeDeo (2018) ‘resonance’ measure:

$$\rho_w(t) := \frac{1}{w} \sum_{d=1}^{w} \left[\mathrm{KL}\left(\theta_t \| \theta_{t-d}\right) - \mathrm{KL}\left(\theta_t \| \theta_{t+d}\right) \right],$$

  where w is the window size. We set this at $w = 10$ such that there is five years either side.

- First term in brackets is difference between today and yesterday. Second term is the difference today and tomorrow. That is we compare today to the past and the future.

- Consider if there’s a big difference between now and the past, but minimal difference with respect to the future. This means there’s been an episode of exploration that has ‘stuck around’.

Ok, let's implement.

- We'll begin with a case study of IBM.
- Big firm with a long history that was at the centre of ICT's evolution over the 20th century.
- General theme: Since we need decent-sized patent portfolios our approach is best-suited to studying long-lived firms.

Start with basic word frequencies by decade...

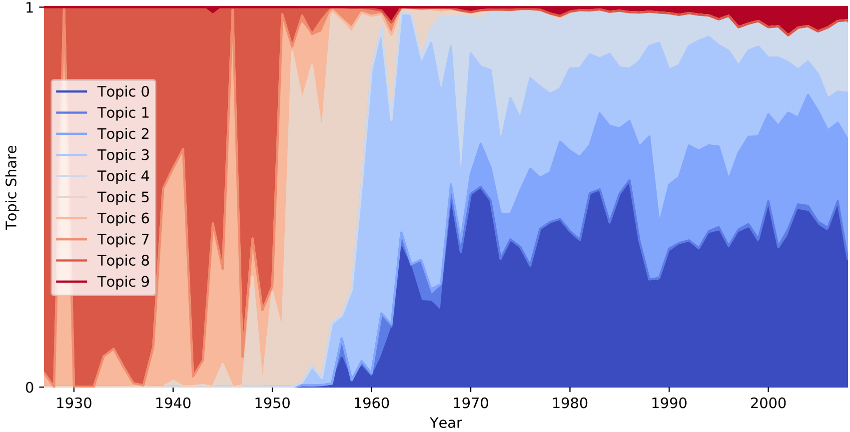
IBM Case Study 1: Fastest Growing Unigrams

Overall		**1930s**		**1940s**		**1950s**	
Word	Share	Word	Change	Word	Change	Word	Change
data	2.59	mean	1.64	card	2.81	circuit	2.52
system	1.45	feed	0.85	machin	1.68	magnet	1.63
layer	1.26	select	0.61	tape	1.10	memori	1.38
first	1.23	new	0.58	perfor	0.97	data	1.19
devic	1.13	gear	0.58	electron	0.69	signal	0.94
circuit	1.02	sheet	0.55	number	0.61	input	0.90
signal	0.94	time	0.55	sens	0.56	puls	0.87
second	0.92	applic	0.47	column	0.47	line	0.77
memori	0.84	charact	0.46	digit	0.47	devic	0.76
control	0.76	invent	0.43	valu	0.46	binari	0.63

IBM Case Study 2: Fastest Growing Unigrams

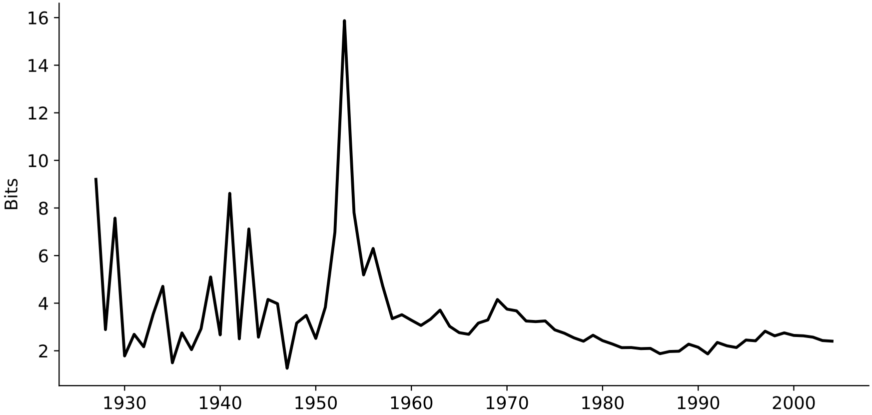
1960s		1970s		1980s		1990s	
Word	Change	Word	Change	Word	Change	Word	Change
surfac	0.73	silicon	0.85	data	1.18	user	0.73
cell	0.60	line	0.78	system	1.04	layer	0.59
metal	0.58	layer	0.72	imag	0.53	system	0.56
control	0.55	print	0.55	comput	0.52	first	0.40
substrat	0.54	address	0.52	first	0.49	one	0.37
code	0.50	data	0.52	document	0.44	content	0.36
error	0.46	chip	0.50	access	0.42	request	0.34
wave	0.35	region	0.50	user	0.38	method	0.32
member	0.34	generat	0.40	circuit	0.35	process	0.31
mean	0.34	ribbon	0.38	optic	0.34	inform	0.30

IBM Case Study 3: Topic Evolution



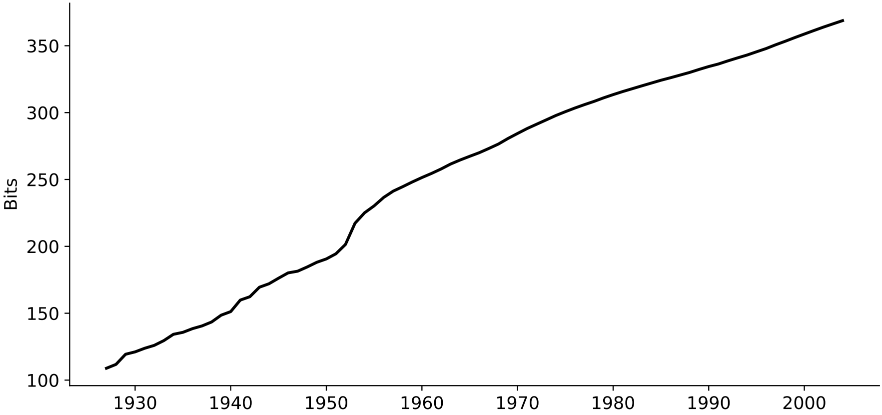
10-topic model to ‘eyeball’ underlying structure.

IBM Case Study 4: Exploration



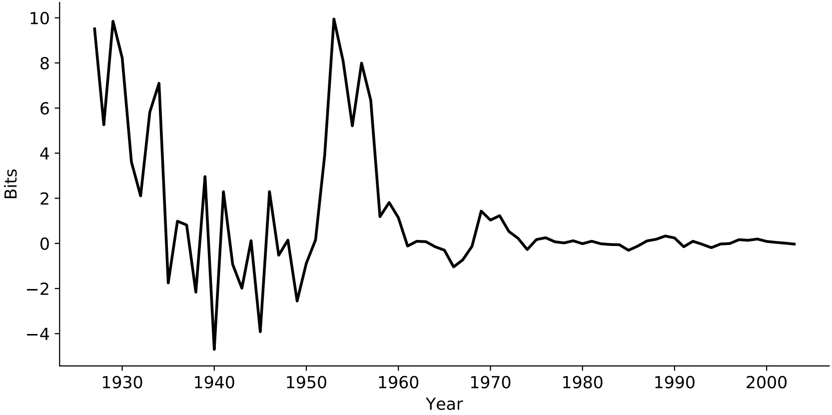
Exploration measure picks up a topic structure shift circa 1953.

IBM Case Study 5: Cumulative Exploration



Stock measure suggests a major 1950s shift with plausible ‘exploitation’ phase from early 1970s onwards.

IBM Case Study 6: Successful Exploration

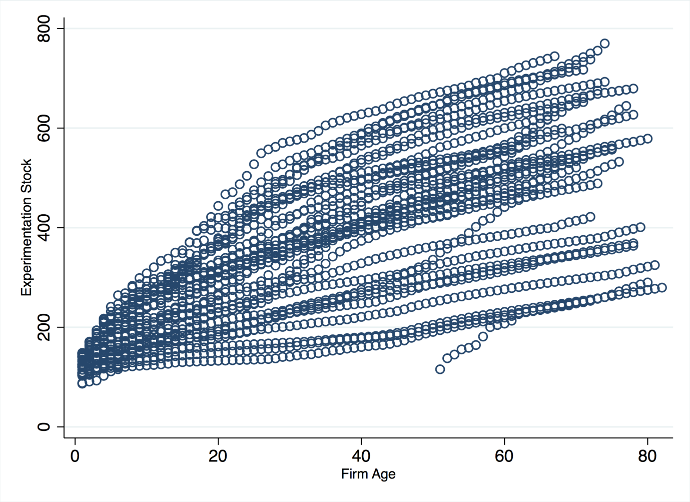


Resonance indicates the late 1950s exploration spike was also important.

Generalising

- How typical is the experience of IBM? We now want to look at the bigger sample of long lived firms.
- Specifically, does exploration taper with age faster than growth in firm size (sales, market cap)? That is, does exploration slow down as firms start to exploit past innovative efforts?
- And what happens to firm R&D intensity over the firm's life-cycle?

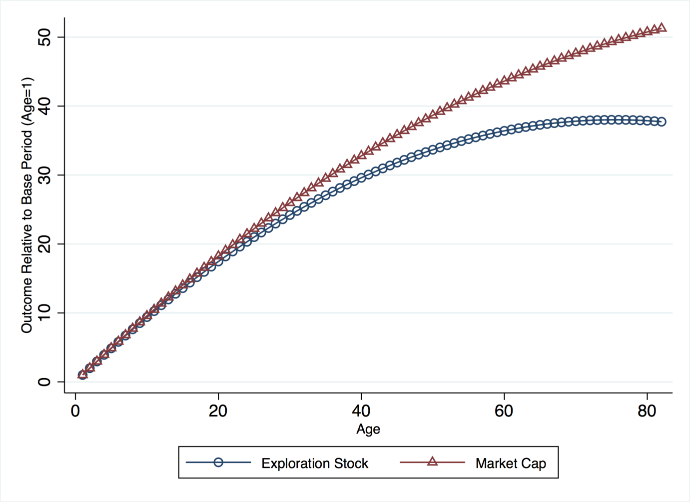
Empirical Results 1: Large Firms



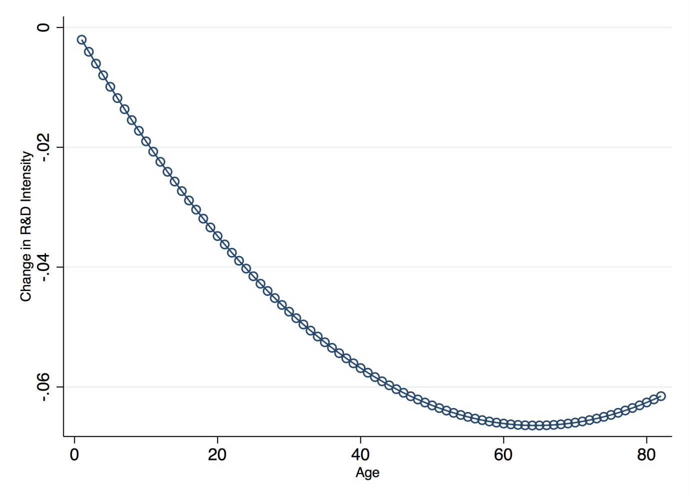
Empirical Results 2: Cumulative Exploration and Firm Age

	(1) Baseline	(2) +SIC4	(3) +Mktcap	(4) +PatStock	(5) +Sales
age	12.03***	12.31***	11.89***	10.94***	11.73***
	(0.533)	(0.505)	(0.511)	(0.574)	(0.576)
age2	-0.0645***	-0.0644***	-0.0633***	-0.0562***	-0.0607***
	(0.00781)	(0.00742)	(0.00735)	(0.00807)	(0.00803)
log marketcap			7.156***		
			(1.746)		
log patstock				14.09***	
				(2.238)	
log sales					7.890***
					(2.001)
R-sq	0.620	0.718	0.720	0.728	0.726
N	26,727	26,721	26,375	26,721	23,009

Empirical Results 3: Gradients of Exploration Stock and Firm Size



Empirical Results 4: Change in R&D Intensity



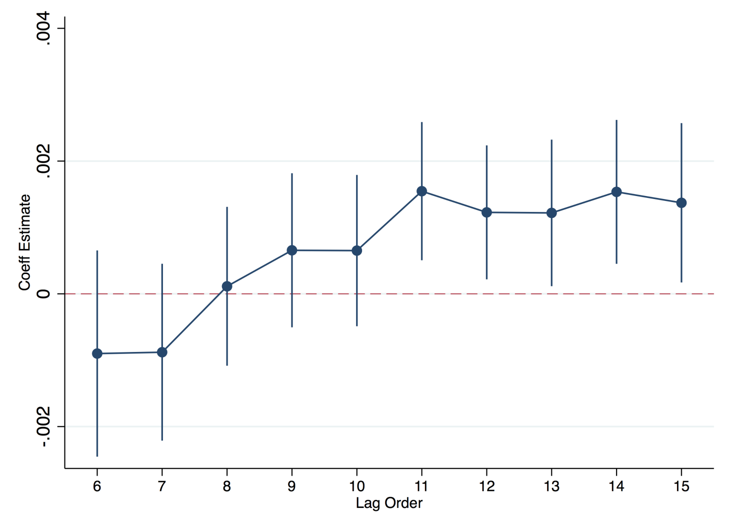
Empirical Results 5: Sales Regressions

We now connect our exploration measure to firm outcomes in a regression framework. The basic model that we adopt is as follows:

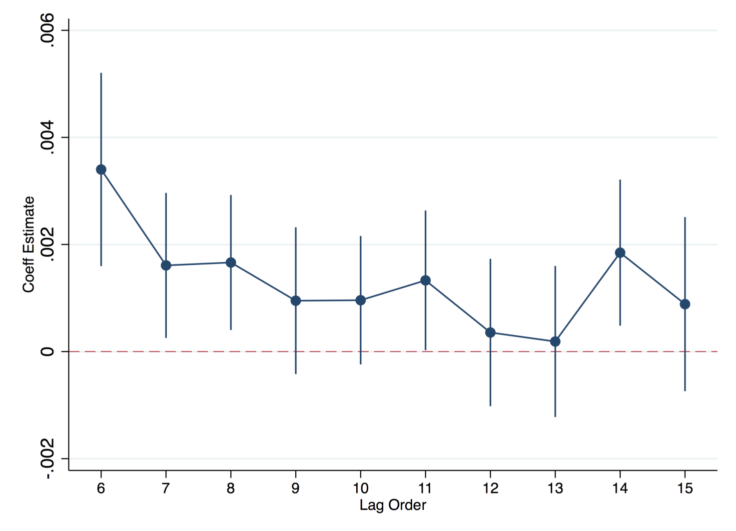
$$\Delta_k \ln(\text{Sales})_{ijt} = \alpha + \sum_{L} \beta_{k-1} \text{KL}_{t-l} + \tau_t + \mu_j + \tau_{jt} + \varepsilon_{ijt}$$

where $\Delta_k \ln(\text{Sales})_{ijt}$ is the k-year change in firm i log sales measure in period t, KL_{t-l} is an l-period lagged exploration measure, τ_t are time effects, μ_j are industry effects, τ_{jt} are industry trends, and ε_{ijt} is an error term. We use different lag orders L to understand the dynamic relationship across specifications.

Empirical Results 6: Lagged Exploration



Empirical Results 7: Lagged Successful Exploration



Exploration and Firm Dynamics

Nesting within a theoretical or empirical model faces challenges.

- Our exploration measure is effectively a time-varying firm fixed effects.
- Furthermore, it has a time-varying impact on firm performance. Specifically, it precedes revenue-rich exploitation phases.
- And methodology is best focused on the large firms due to data reasons. Hard to study the small firms (big focus of new heterogeneous growth models).

Aggregate Technology

- Any pooling across lots of patents is heavily influenced by ‘patent explosions’ like the recent ICT surge. This might be legitimate (ie: we are currently

- Looking at the class level faces the problem of identifying the ‘birthdates’ of classes.

- So it’s turned out that big companies like IBM are good summary vehicles - they nest a lot of relevant, closely-related technologies together.

Conclusion

- We build a new measure of innovative behavior from the patent text.

- It picks up an S-shape in behavior for large firms. The measure explains growth even after controlling for traditional innovation measures.

- Lot of stuff in this research area: The big challenge for the future is understanding evolution year-by-year

References

Barron, Alexander T.J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo. 2018. "Individuals, institutions, and innovation in the debates of the French Revolution." *Proceedings of the National Academy of Sciences of the United States of America* 115 (18): 4607–12. DOI: 10.1073/pnas.1717729115. arXiv: 1710.06867. [PDF p. 18]

Bergeaud, Antonin, Yoann Potiron, and Juste Raimbault. 2017. "Classifying patents based on their semantic content." *PLoS ONE* 12 (4): 1–22. DOI: 10.1371/journal.pone.0176310. arXiv: 1612.08504. [PDF p. 12]

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* 28 (4): 1661–707. [PDF p. 2]

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production." *Quarterly Journal of Economics,* (June): 927–91. DOI: 10.1093/qje/qjx046.Advance. [PDF p. 12]

Itti, Laurent, and Pierre Baldi. 2009. "Bayesian surprise attracts human attention." *Vision Research* 49 (10): 1295–306. DOI: 10.1016/j.visres.2008.09.007. [PDF p. 14]

Kelly, Bryan T., Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2020. "Measuring Technological Innovation over the Long Run." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3279254. [PDF p. 5]