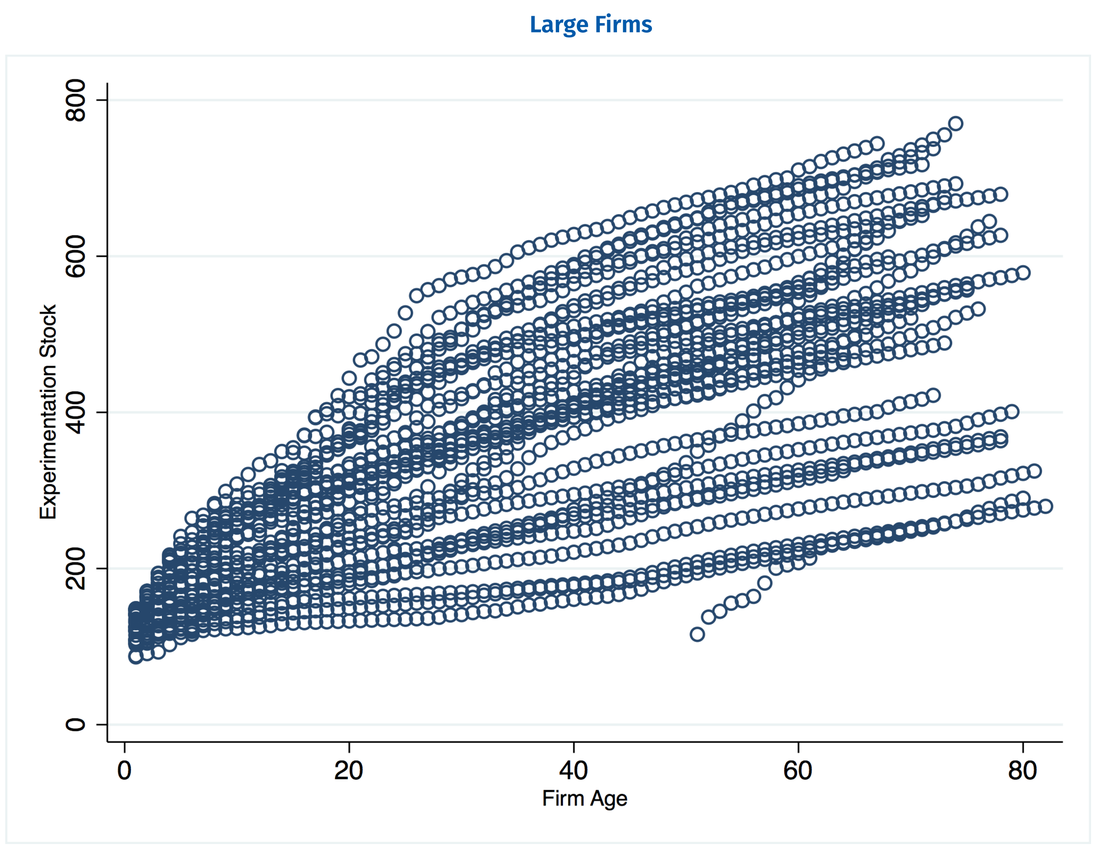
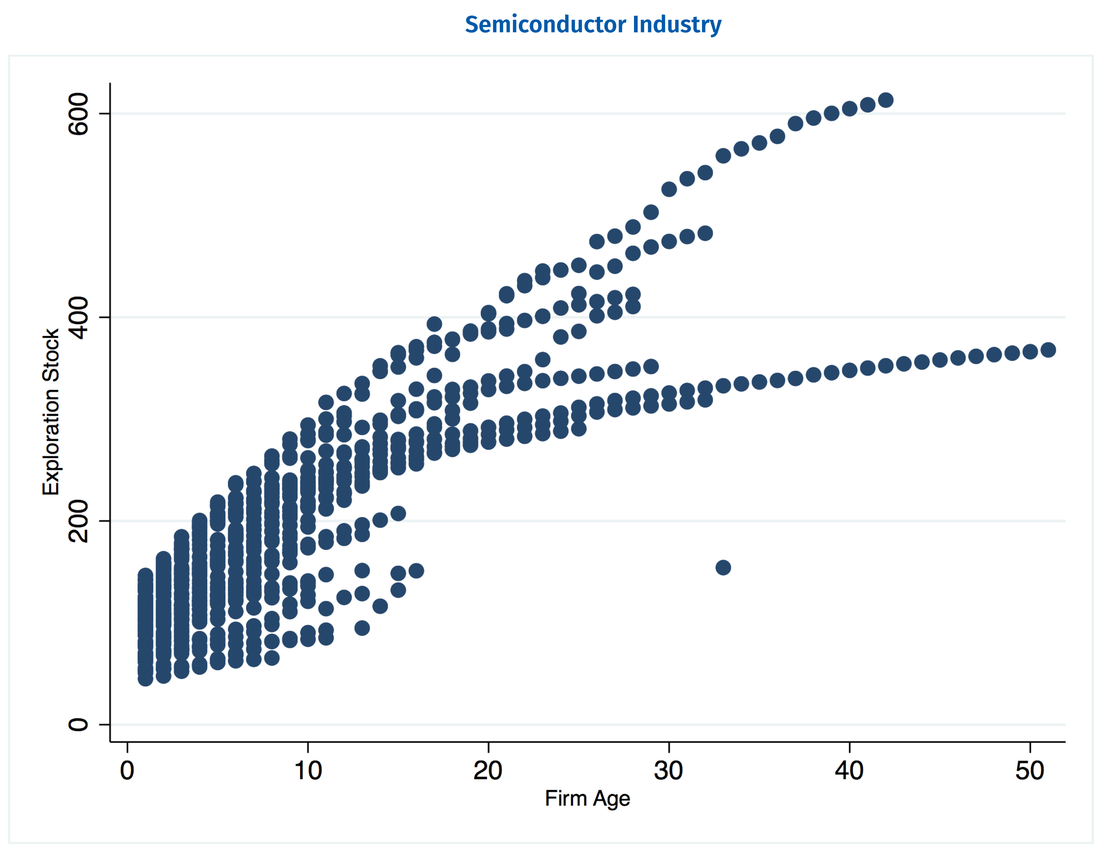
Exploration and Exploitation in US Corporate Research

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

[a] The Alan Turing Institute (nkuhlen@turing.ac.uk); [b] University of Cambridge; [c] University of Warwick

1 **Exploration Stock and Firm Age Over Time.** Left graph shows the evolution of our 'experimentation stock' measure for firms that are aged 60 or more as of 2007 and are above the 95th percentile in the firm-level distribution of total cumulated patents. Right graph shows the evolution for firms in the semiconductor industry.



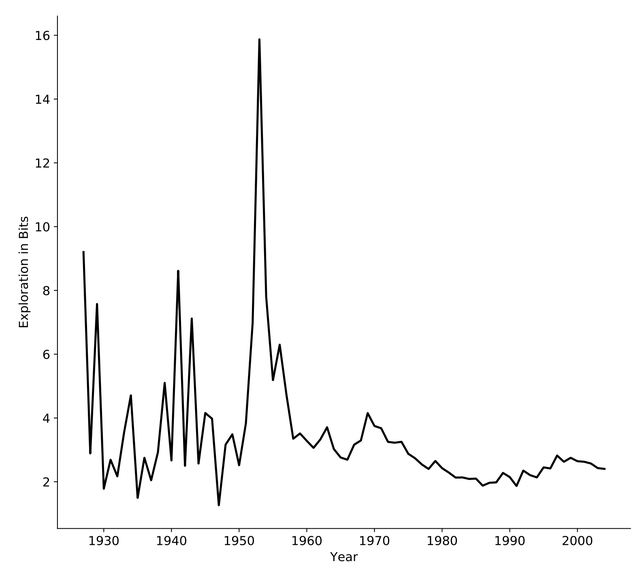


2 **Relationship between Cumulative Exploration and Firm Age.** The table shows the results of regressions of the cumulative exploration measure on linear and quadratic firm age.

	(1) Baseline	(2) +SIC4	(3) +Mktcap	(4) +PatStock	(5) +Sales
age	13.28***	13.06***	13.06***	13.59***	13.04***
	(0.411)	(0.366)	(0.389)	(0.444)	(0.398)
age2	-0.0879***	-0.0818***	-0.0819***	-0.0863***	-0.0812***
	(0.0101)	(0.00813)	(0.00823)	(0.00858)	(0.00833)
log marketcap			0.722		
			(1.543)		
log patstock				-5.105*	
				(2.279)	
log sales					0.144
					(1.732)
R-sq	0.727	0.787	0.788	0.788	0.787
N	27760	27760	27514	27760	23804

Notes: Standard errors clustered by firm in parentheses. log marketcap is the logarithm of market capitalization, and log patstock is the logarithm of the patent stock. Year effects in all regressions, SIC4 fixed effects from col(2) onwards.

3 **IBM's Exploration.** Shows the exploration for IBM from 1927 to 2004.



Introduction

- We propose a method for tracking patterns of 'exploration and exploitation' in firm patenting behaviour in the US for the period since 1920.
- Our first objective is to construct an exploration measure from the text of patents and involves the use of Bayesian surprise (Itti and Baldi, 2009) to measure how different current patent-based innovations are from the firm's existing portfolio.
- Our second objective is to characterise, empirically, the implications of exploration versus exploitation choices for a firm.
- Our empirical implementation of this approach uses a database built up from a match of US Patents and Trademark Office (USPTO) records on the abstracts of patents from 1920 onwards to listed company information.

Methodology

- We use LDA (Blei et al., 2003) to describe patent texts in terms of their latent topic structure. We then use Bayesian surprise to quantify the extent to which a firm's patents contain a new mixture of topics.

1 For each firm, estimate topics for the patent subcorpus using Latent Dirichlet Allocation (LDA).

2 Measure the KL distance between the topic distributions of a given year to the previous average.

- More specifically, we define exploration as

$$\eta_{i,t} := D_{\mathrm{KL}}\left(\theta_{i,t}\middle\|\bar{\theta}_{i,-t}\right) = \sum_{k=1}^{K} \theta_{i,t}(k) \log_2\left(\frac{\theta_{i,t}(k)}{\bar{\theta}_{i,-t}(k)}\right),$$

where $\bar{\theta}_{i,-t}$ denotes the average topic distribution up until year t for firm i in year t and K is the number of topics. This definition builds on the study by Murdock et al. (2017).

Results

- Paths of the exploration measure for a sample of large, long-lived firms – aged 60 or older by the end of the sample and included in the top 5% of firms in terms of total patents – show evidences of clearly defined trends at the firm-level, including indications of classic 'S-shaped' developmental behaviour.
- Firms in the semi-conductor industry show a pattern of dispersion whereby firms with higher exploration trajectories appearing to 'breakaway' after surviving their first 10 years.
- Regressions: exploration is indeed parabolic in age and, interestingly, age explains exploration over and above any correlation with firm size. Columns (3)-(5) control for market capitalisation, the firm patent stock and firm sales in succession with minimal effects on the coefficients of the two age variables.

Conclusion

- The initial work we present here uncovers credible evidence of exploration patterns in firm behaviour that are distinct from other potentially correlated aspects of firm performance.
- Using a measure of 'cumulative exploration' we are able to trace out development patterns in a firm's innovation behaviour. That is, there are clear phases of faster and slower exploration.
- Future directions: First, deepen the present analysis and further characterising the prevalence of exploration versus exploitation across the size and age distribution of firms.
- Second,we plan to aggregate our firm-level measures at the industry and economy level.
- Third, we consider integrating our measure into quantitative endogenous growth environments.

References

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022. arXiv: 1111.6189v1.

Itti, L., and P. Baldi. 2009. "Bayesian surprise attracts human attention." *Vision Research* 49 (10): 1295–306.

Murdock, J., C. Allen, and S. Dedeo. 2017. "Exploration and exploitation of Victorian science in Darwin's reading notebooks." *Cognition* 159: 117–26.