Technology and Real Options: Evidence from Patent Text

Davidson Heath, USC Marshall

Motivation

- Technology is the central fact of long-run growth at the aggregate and the firm level
- Hard to measure, often as a residual (TFP)

This Paper

- Measure firms' technology using patent text
  - Open up the black box, exploit vast textual data

Research Question:

- How do firms' technology positions affect risk and returns?

Patent Text

- Body text of every patent granted by USPTO from 1926-2010
  - online at the Google Patents Project
- 6.2 million patents, 240GB text (~134,000,000 pages)
- Select one, two and three word phrases (“ngrams”)
  - English nouns
  - Standard text screens

Ngrams

Year first appearing:		
1950	1970	1990
control circuitry	acid sequence	email address
clock cycle	interferon	notebook computer
substrate material	nucleotide	remote memory storage
epoxy resins	bus interface	jpeg image
computer networks	programming language	terabyte
command signal	interface card	picture experts
remote computer	plasma display	picture experts group
enantiomers	cholesterol	email addresses
breast cancer	mosfet	sound card
polymer matrix	cpu controls	multiple servers

Patent Data

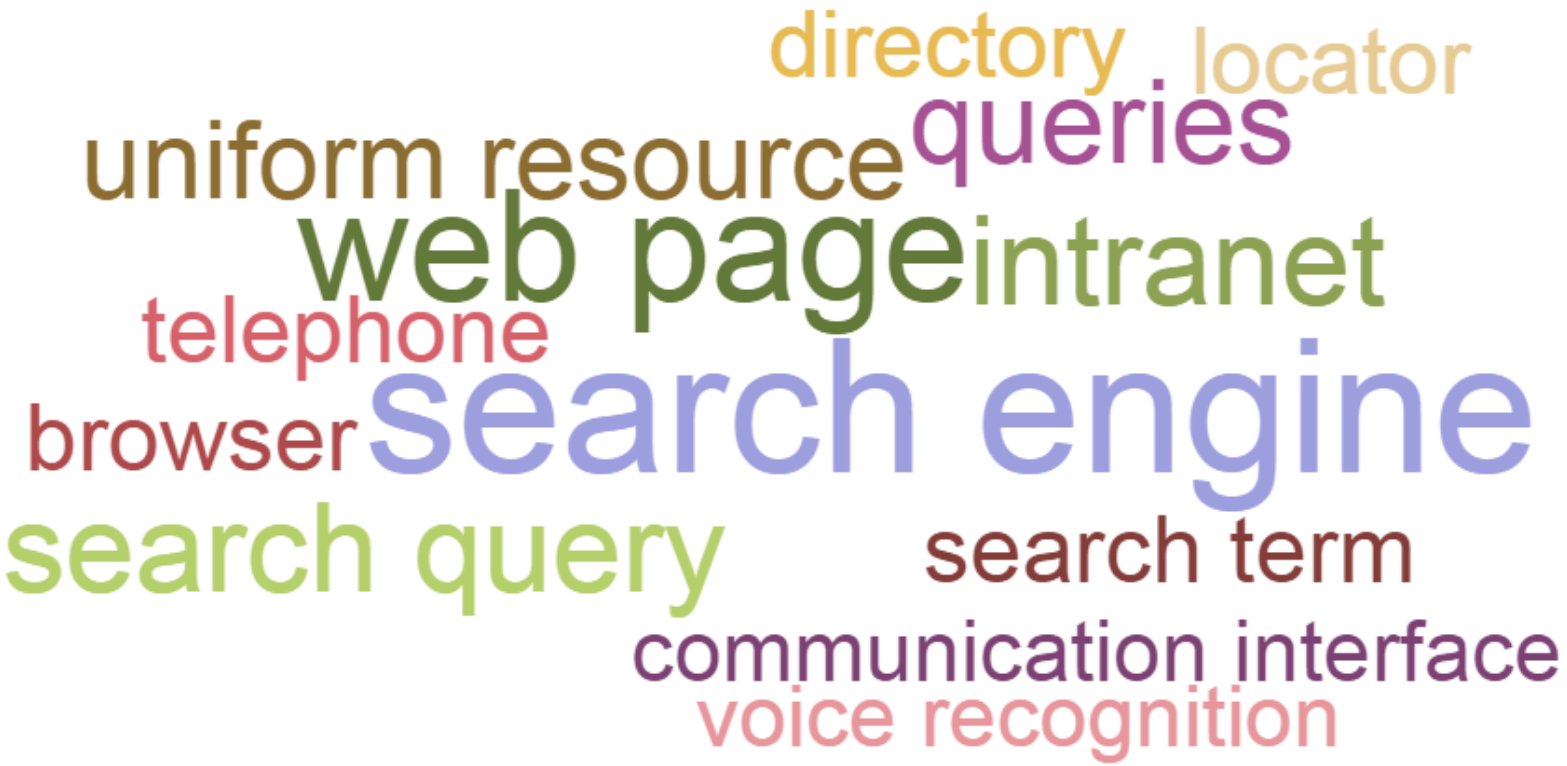
- Each patent:
  - Evaluate presence (=1) or absence (=0) of each ngram
  - Boolean vector with 64,000 entries
  - Normalize to unit length
- Maps each patent to a point in 'word space'
- Text-based patent measures predict patent renewals

Firm-Years

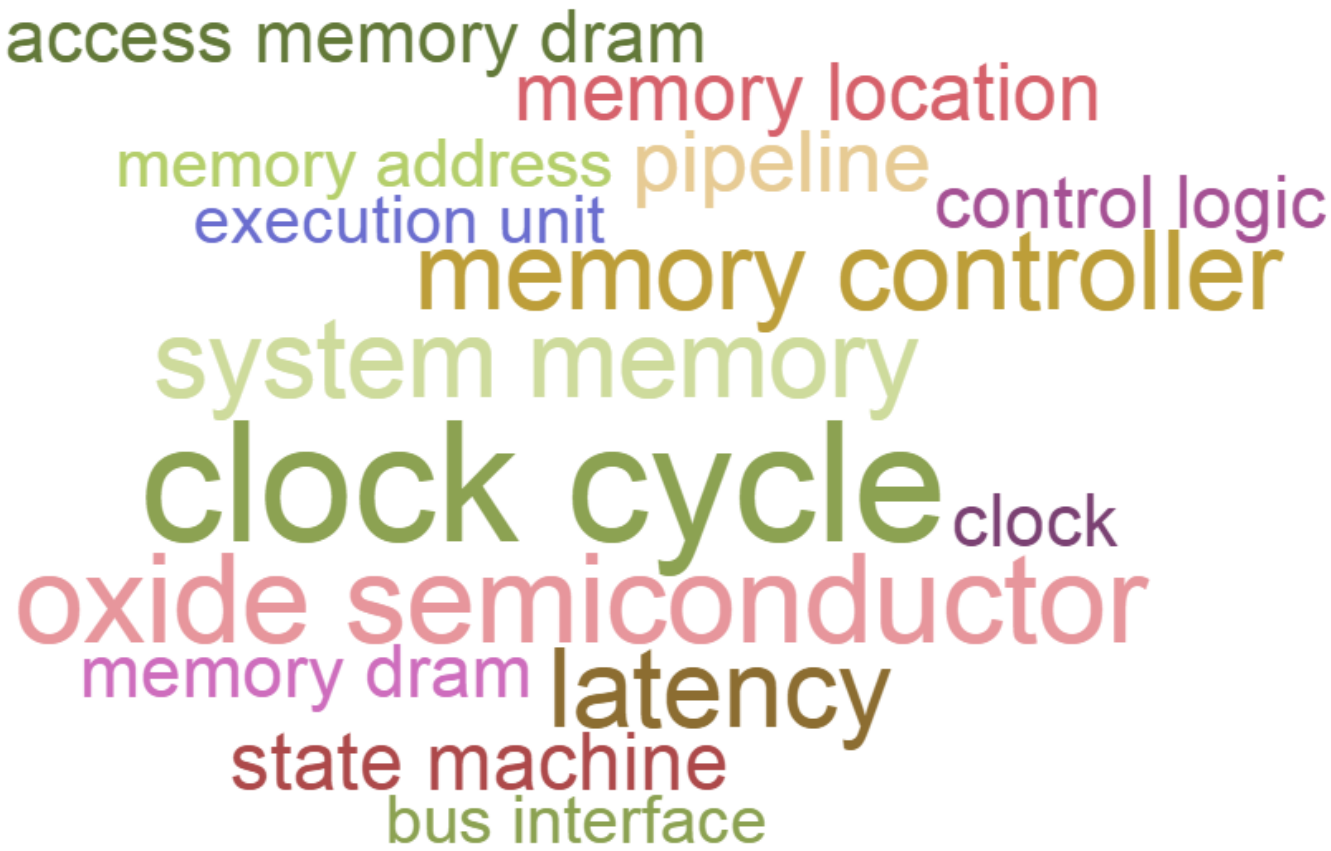
- 18-58% (mean=29%) of Compustat firms in each year 1961-2008 have at least one patent to their name

- Sum the vectors of all patents for each firm-year
  - Normalize to unit length
  - Maps firm-years in 'patent word space'

GOOG in 2000


directory
locator
queries
uniform resource
web page
intranet
telephone
search engine
browser
search query
search term
communication interface
voice recognition

INTL in 2000

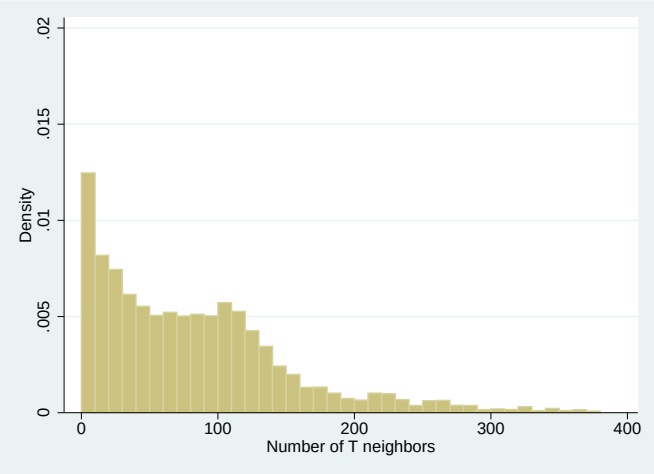

access memory dram
memory location
memory address
pipeline
control logic
execution unit
memory controller
system memory
clock cycle
clock
oxide semiconductor
memory dram
latency
state machine
bus interface

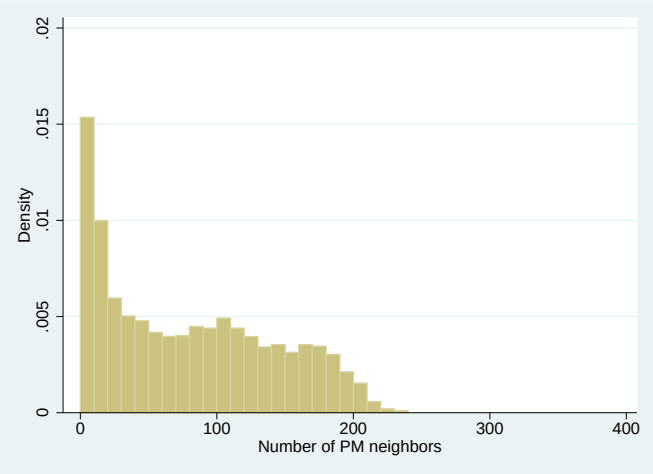
KO in 2000


departure
cabinet
drink
water line
bottle
consumer
dispenser
shelve
carbonation
beverage dispenser
refrigeration
beverage bottle
beverage
refrigeration system
cooler
flavor
beverage container
beverage industry

Complementarity with TNIC

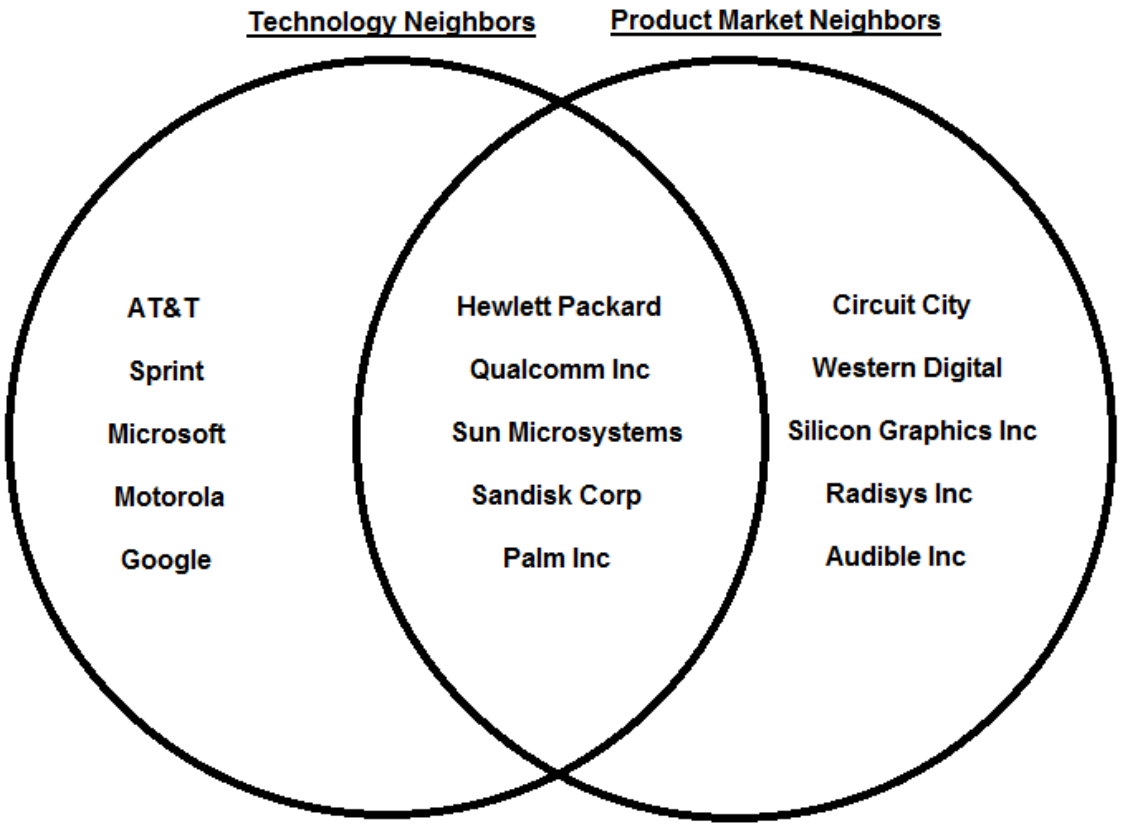
- Hoberg and Phillips TNIC uses 10-K product descriptions
  - "PROD" neighbors
- I assign firms to be neighbors if similarity > cutoff
  - "TECH" neighbors
- Both networks are ~3.1% saturated





Complementarity with TNIC

- $Pr(TECH_t|PROD_t) = 32.7\%$, $Pr(PROD_t|TECH_t) = 31.8\%$
- Apple in 2006: (pre-iPhone):



Complementarity with TNIC

- Dummy for whether firms are PROD and TECH neighbors:

	(1) $PRODNEIGHBOR_{t+1}$	(2) $TECHNEIGHBOR_{t+1}$
$TECHNEIGHBOR_t$	0.024***	0.66***
	(0.0075)	(0.012)
$PRODNEIGHBOR_t$	0.32***	-0.025***
	(0.021)	(0.0018)
Observations	3,461,540	3,461,540
Fixed Effects	Year	Year
Pseudo R-squared	0.14	0.38

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Main Measure: *techdiff*

- For each firm-year, compute its cosine similarity with all other firms in the same year
- Define

$$techdiff_{it} = 1 - quantile_j^{99}(cos_{ijt})$$

- Of 1000 firms in that year, how close is the 10th closest firm

Associated with Q

	(1) $log(Q_t)$	(2) $log(Q_t)$
$techdiff_t$	0.078***	0.17***
	(0.018)	(0.040)
$log(cites_t)$	0.014**	0.017
	(0.0059)	(0.011)
$HHI_{SIC,t}$	0.011	-0.0064
	(0.0093)	(0.020)
$tnic3tsimm_t$		0.090***
		(0.034)
$RDstock/AT_t$	-0.0072	-0.059**
	(0.0068)	(0.030)
$log(AT_t)$	-0.23***	-0.35***
	(0.023)	(0.049)
Observations	38,762	12,946
R-squared	0.041	0.078
No of firms	4,038	2,262
Fixed Effects	Firm + Year	Firm + Year

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Associated with *ROA*, *CAPEX*

	(1) ROA_t	(2) ROA_t	(3) $CAPX/AT_t$	(4) $CAPX/AT_t$
$techdiff_t$	0.030***	0.038***	0.0042***	0.0053***
	(0.0042)	(0.0079)	(0.00094)	(0.0015)
$log(cites_t)$	-0.00079	-0.0010	0.00027	-0.00044
	(0.0016)	(0.0031)	(0.00044)	(0.00067)
$HHI_{SIC,t}$	-0.0032	-0.011*	0.00094	0.000045
	(0.0021)	(0.0055)	(0.00089)	(0.0014)
$tnic3tsimm_t$		-0.0072		-0.0023*
		(0.0058)		(0.0014)
$RD/sales_t$	-0.061***	-0.073***	0.00035	-0.00034
	(0.0052)	(0.010)	(0.00053)	(0.00077)
$log(AT_t)$	0.095***	0.23***	0.00065	-0.0011
	(0.012)	(0.021)	(0.0018)	(0.0028)
$Mkt/Book_t$	0.016***	0.023***	0.0030***	0.0018***
	(0.0037)	(0.0062)	(0.00056)	(0.00042)
Observations	39,957	13,069	39,542	12,994
R-squared	0.134	0.177	0.006	0.005
No of firms	4,080	2,289	4,057	2,278
Fixed Effects	Firm + Year	Firm + Year	Firm + Year	Firm + Year

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Associated with TFP

	(1) TFP_t	(2) TFP_t
$techdiff_t$	0.098***	0.094***
	(0.011)	(0.025)
$log(cites_t)$	0.00094	0.0072
	(0.0032)	(0.0052)
$HHI_{SIC,t}$	-0.012*	-0.0085
	(0.0068)	(0.019)
$tnic3tsimm_t$		0.010
		(0.025)
$RDstock/AT_t$	-1.09***	-1.48***
	(0.12)	(0.24)
$log(AT_t)$	0.27***	0.44***
	(0.020)	(0.046)
$Mkt/Book_t$	0.17***	0.16***
	(0.012)	(0.020)
Observations	30,603	9,041
R-squared	0.182	0.180
No of firms	2,974	1,561
Fixed Effects	Firm + Year	Firm + Year

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Research Question

How do firms' technology positions affect their expected returns?

- High Q firms have low returns (value premium)
- Tuzel and Imhoroglu (2013): high TFP firms have low returns
  - Over and above the value premium

TFP, Q, and Expected Returns

Two big streams:

- Adjustment Costs (Novy-Marx, Tuzel and Imohoroglu)
  - Endogenous heterogeneity in position on production function
  - No optionality, NPV rule holds
- Growth Options (Berk Green Naik, Carlson Fisher Giammarino, Garleanu Panageas Yu)
  - Uncertainty + irreversibility
  - Firms have heterogeneous investment *options*

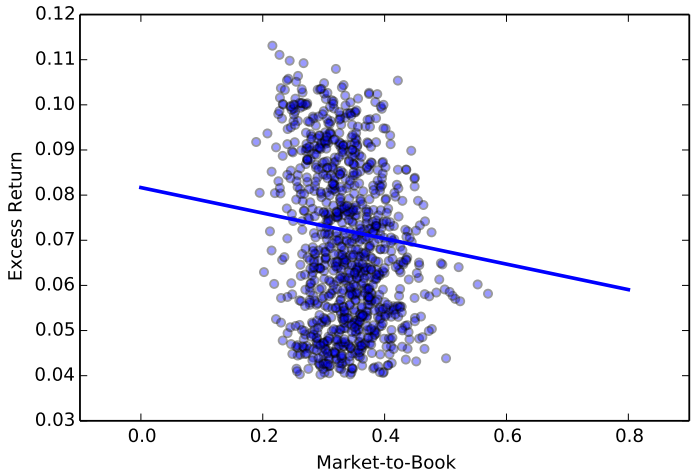
Model

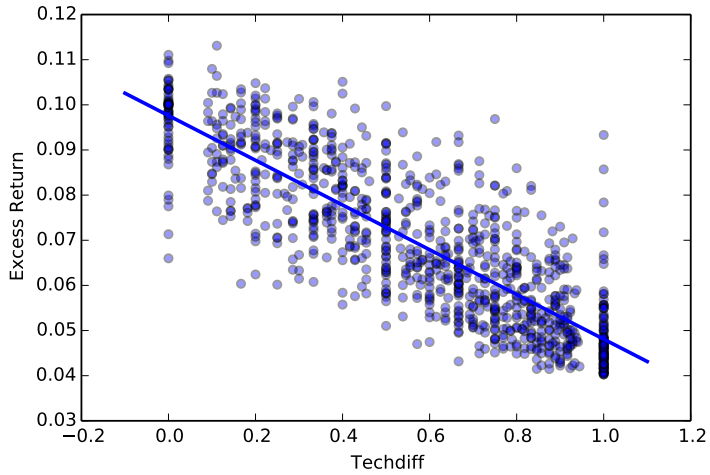
- Firms have growth options on differentiated products
- Invest I to get a product that yields stochastic cash flows
- More differentiated products have lower risk, lower returns (Hou and Robinson)

Model

- Firms wait until $NPV \geq D^* > I$, so NPV rule does not hold
- Once exercised the product is on the balance sheet

Predictions





- Market-to-book ratio reflects product market position
  - Mechanical value premium
- Technology position is correlated but separate
  - Tech differentiation also predicts lower returns

Predictions

1. Higher *techdiff* predicts lower returns, over and above the value premium
2. Product market position does not
3. Pattern stronger in growth firms

Returns

	(1) $rtn - rf$	(2) $rtn - rf$	(3) $rtn - rf$
techdiff	-0.032***	-0.034***	-0.061***
	(0.0068)	(0.0067)	(0.016)
$log(cites)$		0.020***	
		(0.0068)	
$RD/sales$		0.15	-0.0100
		(0.13)	(0.015)
HHI_{SIC}		-0.0033	-0.0054
		(0.0043)	(0.011)
TNIC3tsimm			0.017
			(0.023)
$log(Size)$	-0.063***	-0.064***	-0.087***
	(0.013)	(0.013)	(0.029)
$Book/Mkt$	0.018**	0.017**	0.000060
	(0.0081)	(0.0085)	(0.018)
rtn_{t-1}	-0.12***	-0.12***	-0.080***
	(0.0087)	(0.0094)	(0.016)
$rtn_{t-12,t-1}$	0.036***	0.040***	-0.018
	(0.011)	(0.011)	(0.022)
Observations	496,741	406,698	156,266
R-squared	0.060	0.085	0.075

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Returns, Interacted

	(1) $rtn - rf$	(2) $rtn - rf$
$techdiff$	-0.053***	-0.0035**
	(0.010)	(0.0015)
$techdiff \times Book/Mkt_{quintile}$	0.0070***	
	(0.0026)	
$Book/Mkt_{quintile}$	0.0057	
	(0.0056)	
$techdiff \times Size_{quintile}$		0.00044
		(0.00032)
$Size_{quintile}$		-0.0029***
		(0.00072)
$log(Size)$	-0.065***	
	(0.013)	
$Book/Mkt$		0.0019***
		(0.00065)
rtn_{t-1}	-0.12***	-0.0098***
	(0.0086)	(0.00073)
$rtn_{t-12,t-1}$	0.037***	0.0030***
	(0.011)	(0.00089)
Observations	497,311	496,741
R-squared	0.064	0.062

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Portfolio Sort

		techdiff Portfolio					
		1	2	3	4	5	5 minus 1
	1	0.0129	0.0079	0.0095	0.0071	0.0047	-0.0082
		(0.0070)	(0.0044)	(0.0041)	(0.0035)	(0.0033)	(0.0054)
	2	0.0090	0.0026	0.0003	-0.0011	-0.0026	-0.0116
Size		(0.0048)	(0.0036)	(0.0033)	(0.0032)	(0.0031)	(0.0034)
Portfolio	3	0.0038	0.0001	-0.0024	-0.0027	-0.0035	-0.0073
		(0.0036)	(0.0033)	(0.0028)	(0.0029)	(0.0029)	(0.0023)
	4	-0.0001	-0.0027	-0.0032	-0.0032	-0.0039	-0.0038
		(0.0028)	(0.0027)	(0.0027)	(0.0028)	(0.0026)	(0.0016)
	5	-0.0041	-0.0047	-0.0051	-0.0042	-0.0049	-0.0008
		(0.0022)	(0.0024)	(0.0022)	(0.0023)	(0.0023)	(0.0014)

Conclusion

- Use patent text to map firms' technological position
  - Dynamic (year by year)
  - Distinct from product market position
  - Predicts product market position

- Measure of tech differentiation, *techdiff*
  - Associated with higher Q, ROA, Capex, TFP
  - Predicts lower returns
    - Stronger for growth firms
    - Fits with model of growth options on heterogeneous products