Research Experience as Human Capital in New Business Outcomes[1]

Nathan Goldschlag[2]
Ron Jarmin[2]
Julia Lane [3]
Nikolas Zolas[2]

Abstract

Human capital is typically cited as an important contributor to the survival, growth and innovative activity of new businesses. This paper contributes to the literature by both developing novel measures of human capital and examining the link between those measures and the outcomes of young firms. It builds on several strands of the literature which emphasize the importance of employee workplace experience as a dimension of human capital. It shows that the effects of work experience differ substantially by where an employee worked and is valued differently by firms in different sectors. This is particularly true for research experience, which is consistent with the notion that on the job training in complex tasks should be valuable to firms with complex production technologies.

---

[1] Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. This research was supported by the National Center for Science and Engineering Statistics. NSF SciSIP Awards 1064220 and 1262447; NSF Education and Human Resources DGE Awards 1348691, 1547507, 1348701, 1535399, 1535370; NSF NCSES award 1423706; NIHP01AG039347; and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. Lane was supported through an Intergovernment Personnel Act assignment to the US Census Bureau. The research agenda draws on work with many coauthors, but particularly Bruce Weinberg and Jason Owen Smith.
[2] U.S. Census Bureau
[3] New York University and the US Census Bureau

1. Introduction

This paper contributes to the literature on the link between human capital and the survival, growth and innovative activity of new businesses. We develop new measures of workplace experience, particularly within R&D intensive and high tech firms. We also make use of an entirely new data source that directly measures research experience. We examine the relationship between those measures and startup survival, growth and innovative activities such as patenting and trademarks.

The paper describes the construction of four new human capital measures derived from two sources. The first is a direct measure of research experience derived from a new dataset drawn from the human resource files of a set of research-intensive universities. The data capture all payroll transactions for all individuals – including undergraduate students, graduate students, and postdoctoral fellows - employed on scientific projects at 22 major universities(*1*) These data are the first to directly measure the human capital developed through project level investments in university science. The second, third and fourth measures are indirect in nature. They are drawn from LEHD and W2 data, and create new worker-level measures of human capital based on whether each worker has worked in R&D labs, High Tech businesses and universities.

It also describes the construction of two new datasets on startups. The first of these is a Startup Firm History File drawn from the Longitudinal Business Database (LBD), supplemented with additional information from the Census Bureau's Business Register. In addition, we create a Startup Worker History File derived from worker level data on jobs and earnings. These new files provide a national frame of startups, their survival and their growth between the years 2005 and 2015, as well as a national frame of all workers affiliated with these startups.

Our results suggest that a one-worker increase in the number of high human capital employees in a startup firm's workforce is associated with a lower probability of survival to the next period by 1.7 to 4.1 percentage points, depending on the experience type. However, for startups that do survive to the first period, the hiring of one of these workers in the founding year is associated with a 1.5 to 4.7 percentage point increase in employment and a 2.8 to 6 percentage point increase in revenue in the following year. This is suggestive evidence that high human capital employees elect to go to more high-risk startups that exhibit "up or out" dynamics—either exiting or growing quickly. On the innovation side, the addition of one high human capital individual is positively related to patent and trademark outcomes in the next period, with patent filings increasing by 0.5 to 10.5 percentage points and trademark filings increasing by 3 to 8.5 points in the following year. Our measures of human capital also explains a significant amount of the variation in innovation outcomes, where the inclusion of our basic measures of human capital help explain an additional 40% of variation in patenting outcomes and 11% of variation in trademarking outcomes. These results are consistent with the view that there is a positive and significant relationship between workforce experience and business startup outcomes.

2. Background

Our focus on startups is informed by the literature which suggests that young entrepreneurial businesses are important for introducing and diffusing innovations in the economy. Several authors have shown indirect linkages between formal investments in research and innovation and entrepreneurship and economic growth (*2–4*). In particular, the work of Akcigit and Kerr(*5*) shows that the relative rate of major inventions is higher in small firms and new entrant firms. Scott Stern and coauthors note that the early stage choices of startups – their "digital signature" - is particularly important in predicting their future success.(*6*)

There has been a considerable literature linking human capital to the survival and growth of such new businesses (*6*, *7*). In particular, the decision to start a business, and its subsequent productivity and success is associated with having an entrepreneurial workforce (*8*, *9*). Related work also suggests that highly innovative individuals make "exceptional" contributions to economic growth (*10*). Indeed, the personnel economics and management literatures draw on extensive studies of businesses and human resource practices and suggests that many productive businesses either invest in job-based training or seek to hire well trained individuals(*11–13*) A related literature links external R&D investment and the success of the R&D efforts of individual firms (*14*). In depth studies of the components of intangible assets in contributing to firm productivity and success invariably mention the importance of training(*15*). There is a long literature on the effects of on-the-job training on firm productivity (*16*, *17*).
We draw on two sets of literature – one set that has studied human capital acquisition through learning by doing and experience, and another that has studied the transmission of new knowledge through the flows of individuals from one business to another.

The role of experience in terms of learning how to do complex new tasks through trial and error has been extensively discussed in the endogenous technical change literature (*18*). There is also a great deal of evidence to support the notion that past experience imparts valuable business skills(*19*), and that both firm and economic growth can be significantly affected by workers with experience in R&D activities (*20*, *21*).

The role of university research training on innovative activity and business startups is supported by the anecdotal evidence linking, for example, the growth of Silicon Valley to the presence of Stanford, Boston to the excellent set of universities in the area, and the Research Triangle to the research activity of Duke University, the University of North Carolina and North Carolina State. An extensive literature ties regional economic development clusters with the presence of active research universities, suggesting that research trained individuals flow into innovative new businesses (*8*, *22*, *23*, *4*) .

These various literatures are consistent with the notion that hiring workers with experience is a way firms gain tacit knowledge, particularly when ideas are complex (*24*, *25*). The work of Lee Fleming, for example, suggests that if there are impediments to research-experienced workers moving from one firm to another, less innovation occurs. (*26*, *27*). Our own work suggests that that research trained workers are more likely to work at firms with characteristics closely linked to productivity (*28*).

However, there has been little work done in terms of measuring the experience of workers at different types of firms. The Annual Survey of Manufactures provides counts of production and non-production workers; most other business data sources simply provide counts of employees. However, in principle, a particularly useful source of evidence in this context is economy-wide linked employer-employee data, such as the LEHD data (*29*). Abowd et al. have used linked data to compute person specific measures of human capital (*30*), but do not directly compute measures of research experience. While some work has shown that there are returns to experience at R&D performing firms (*31*), there has been no study to our knowledge that directly measures experience in High Tech firms, R&D labs or in scientific projects.

3. Framework, Data and Measurement

We follow much of the literature (*1–3*) in adopting a simple reduced form framework examine outcomes for startups in terms of their survival, employment and revenue growth, and innovative activities such as producing patents and trademarks. Outcomes (Y) for startup firm *f* at time *t* are driven by the quantity and quality of human capital (HK) it employs as we as standard controls such as capital (K), technology (A), and external factors (X) such as macroeconomic conditions and industry factors.

$$\boldsymbol{Y_{ft}} = \boldsymbol{F}\left(\boldsymbol{A_{ft}}, \boldsymbol{K_{ft}}, \boldsymbol{HK}_{\mathbf{ft}}, \boldsymbol{X_{ft}}\right) \qquad \textbf{(1)}$$

There is some evidence that the effect of human capital will be important for businesses whose production processes involve performing complex tasks (*32*). As a result, the analysis that follows provides separate analyses for High Tech businesses– the scale of the data permit such a level of detail. The rest of this section describes how such businesses are identified, how the human capital measures are constructed, and how startup outcomes are measured.

3.1 Identifying and classifying startups

The Startup History file is constructed as a panel dataset. The primary frame for the data is the Longitudinal Business Database (LBD), supplemented with additional information from the Census Bureau's Business Register, upon which the LBD is based. We utilize this file to identify startups as age zero firms. Once the startups have been identified, we supplement the data with geocodes (state and county-level FIPS, along with Census Tract information if available) and EINs taken from the Business Register. These variables are used to subsequently characterize the workforce associated with each startup gathered from LEHD (Longitudinal Employee-Household Dynamics) and W2 records. The full file contains data on employment, payroll, industry, geography, firm-type and birth/death of the firm.

For the purpose of characterizing worker experience, firms are classified as R&D firms and labs, High Tech or universities. The R&D measure is created by creating identifiers based on the Business Innovation and Research and Development Survey (BRDIS) and Survey of Industrial Research and Development (SIRD)[4]. A firm is classified as an R&D firm if it has positive R&D

expenditures during the year the employee was affiliated with the firm. The R&D laboratory is then based on the industry code for the linked establishment of the employee, specifically NAICS 5417, which is defined as “Scientific research and development services”. The as High Tech based on the relative concentration of STEM employment by industry as in Hecker(*33, 34*). We use the High Tech classification to both characterize worker experience, identifying individuals with prior experience in High Tech industries, and to subset the universe of startups within a year.

The university measure is derived from data from IPEDS and the Carnegie Institute which provide a frame of universities in the United States. We use the national university research outlays collected by National Center for Science and Engineering Statistics at the National Science Foundation to subset our sample of universities to the top 130 research universities, which comprise of 90% of total federally funded R&D research.

While capital, financing, management and macroeconomic conditions are not directly measured in the data, because the data are longitudinal, we can include firm and time/industry/zip fixed effects.

3.2 Human Capital Measures

The first three human capital measures are derived from a new dataset called the Startup Worker History File, which characterizes the workforce associated with each startup. It is created from the universe worker level data on jobs derived from administrative records in both the LEHD and W2 records, and covers the period 2005-2015.

The frame is each paid job for each worker from 2005-2015 as reported at both the Employer Identification Number (EIN) level via IRS form W2 and state-level Unemployment Insurance wage records. The latter underlie the core LEHD infrastructure (*13*) and are necessary to identify the establishment for the bulk of multi-unit firms (*35*). The combined data includes more than 3 billion person-EIN-year observations (approximately 70% match across the W2 and LEHD/UI universes, 20% are found only in the W2 records and 10% are only found in LEHD). These data are enhanced with the LEHD Individual Characteristics File (ICF), which includes demographic data on persons including sex, age, race and place of birth.[5] We are able to link 43 million of the 3 billion person-EIN-year observations to startups in their birth year, giving us an average of nearly 4.5 million person-startup observations each year.[6]

The first three measures of human capital are indirect in nature, since they do not directly measure research experience. They are derived from an individual’s work history in the years prior to being employed at the startup in year 0 and reflect the degree of employment experience in R&D labs, high tech businesses and universities. In the case of R&D labs, we include all workers employed in an R&D performing firm in an R&D lab (NAICS code “5417”). For

[5] A detailed discussion on the matching process and match rates is provided in the appendix.

[6] This figure differs from the reported Business Dynamics Statistics (BDS), which calculate employment at startups at a specific point in time (March 12). Our figures are higher, reflecting employee-employer transitions (i.e. workers who work briefly for a startup and then move to a different job). The 48 million observations represent 37.8 million unique individuals.

workers in a High Tech industry, we include the workers that are in the top-half of the earnings distribution for that year, to minimize the likelihood of including support or administrative personnel. We construct the same measure for employs at national research universities.

The fourth measure is derived from UMETRICS data(*4*), which includes 22 universities accounting for about 26% of all federally funded research. The data are derived from universe personnel and financial records of participating universities. Although four files are provided by each university, the key file of interest in this project is the employee file. These workers are a subset of the previous university experience measure. Briefly, for each funded research project, both federal and nonfederal, the file contains all payroll charges for all pay periods (identified by period start date and period end date) with links to both the federal award id (unique award number) and the internal university identification number (recipient account number). In addition to first name and last name, and date of birth, the data include the employee's internal de-identified employee number, and the job title (which we mapped into broad occupational categories). Both federal and nonfederal funding is covered in the data. The Catalog of Federal Domestic Assistance (CFDA), which is included in each award identifier, provides a full listing of all Federal programs available to universities (and other types of organizations) and is captured in the UMETRICS data to be able to filter federal award expenditures by federal funding agency. Each university provided data as far back as their record keeping allowed.

3.3 The Startup Worker History File

The worker history file is constructed in three steps. The first steps involves identifying person and firm characteristics in the years prior to startup. The LEHD and W2 data provide worker histories for 260 million individuals for each employer (at the EIN level) for each year in the period 2005-2015. Their individual characteristics are captured by matching to the Individual Characteristics File (the ICF) – this file provide information on date of birth, foreign born status and sex.

The EIN of their employers is then matched to the BRDIS/SIRD data to determine whether the employer is an R&D performing firm. There are 74,000 of those EINS, and 420,000 resulting EIN-Year observations. The EIN is also matched to firms in 61 High Tech industries (6-digit NAICS). Actual employment on a grant is determined by a match to UMETRICS data; there are 340,000 research experienced individuals between 2005 and 2015.

Startups are identified as firms of age zero. The total worker history file thus has 530.3 million PIK-EIN-Year startup observations. Of those, 43.2 million observations are associated with startups in year 0.

Figure 1 provides a graphical illustration of the process.

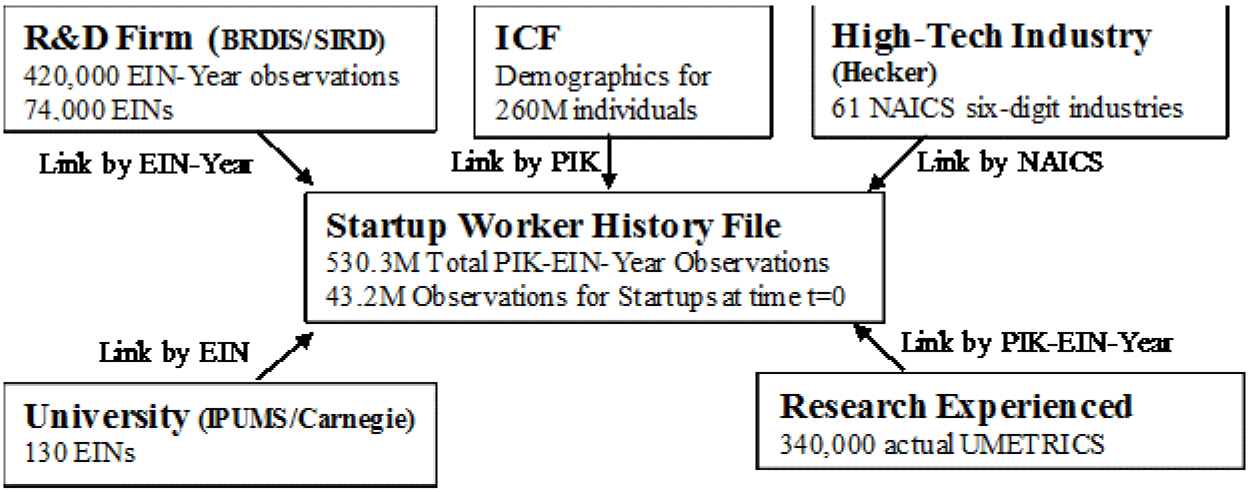
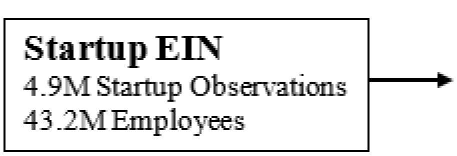


Figure 1: The startup worker history file

The second step involves collapsing and tabulating human capital totals for each startup EIN to create a startup file. There are a total of 4.9 million EINs of age zero in the data, of which about 35,000 have hired individuals with work experience in R&D labs – the number of such employees totals 67,000. 371,000 EINs have hired at least one individual with high tech experience – the number of these employees total 806,000. About 442,000 EINS have hired at least one university experienced employee; the number of these totals 882,000 There are about 11,000 startups that have hired individuals with research experience.at the UMETRICS universities; there are 13,000 such individuals . The process is described graphically in Figure 2.



	R&D Lab	High-Tech	University	Research Experience
Startup Count	35,000	371,000	442,000	11,000
Employee Count	67,000	806,000	882,000	13,000

Figure 2: Creating the Startup file

The third and final step involves merging startup EIN file with the Startup Firm History File and classify startup types and outcomes at time t=0 and calculating how many survive to the year subsequent to their birth.   That information is graphically presented in Figure 3.

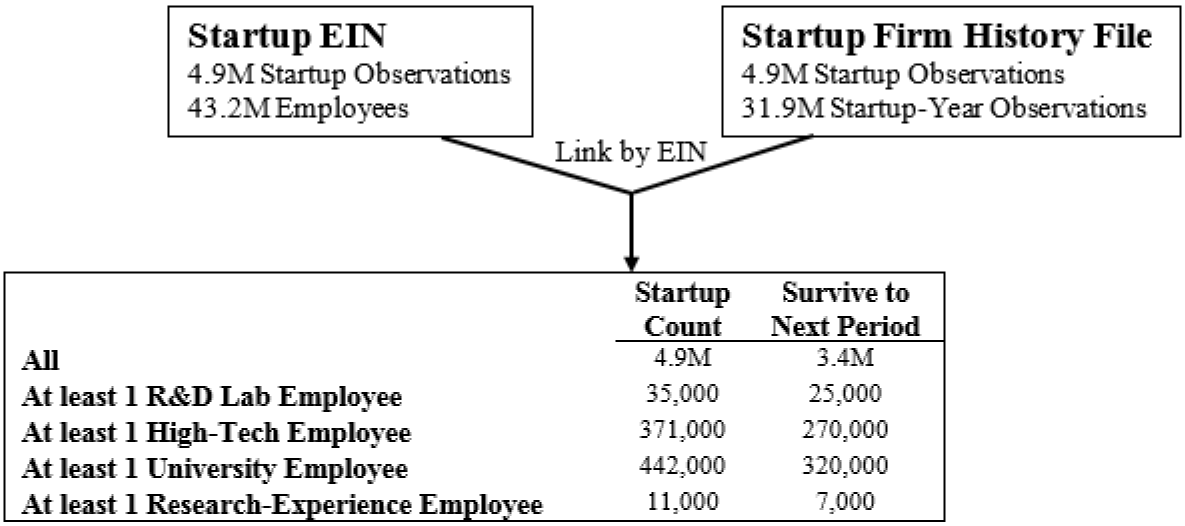


Figure 3: The Startup History file

3.3 Startup Outcomes

While a wide variety of outcome measures can be generated, we limit the measures to five: Survival to period *t+1*, Employment Growth between t and *t+1*, Revenue Growth between t and *t+1*, applying for a patent in *t+1* that is eventually granted, and filing for a trademark in *t+1* that is eventually registered.

Startups are linked to patent grants and trademark filings through existing crosswalks at the Census. Patent linkages are based on a triangulation methodology first described in Graham et al.(*35*) Their linkage methodology simultaneously leverages information on both patent inventors and assignees in combination with job-level information from the LEHD to distinguish between true and false matches. By using more information than traditional patent linkage efforts (e.g. fuzzy business name and geography), the triangulation match produces more and higher quality linkages. Trademarks are matched to startups using the match described in Dinlersoz et al. (*36*). The business name and address information found in the USPTO's Trademark Case File Database are used to create firm-trademark linkages. To measure innovative outcomes of startups we identify whether a startup applied for a patent in the year after its birth (t+1) that was eventually granted. Similarly, we identify whether each startup filed for a trademark in t+1 that was eventually registered.

4. Basic Facts

This section establishes some basic facts on the human capital composition of the startups and their outcomes.

4.1 Startup Facts

We begin by highlighting some facts regarding startups and their outcomes. Between 2005 and 2015, one-year survival rates typically hover around 68%, but are higher for high tech startups in every year. As is well known, the number of startups dropped in 2007 by 25% (relative to 2005) and by 33% the following year – and by 2013 was still at the same level. The same was also true for high tech startups, although the order of magnitude is not quite as great – the number of High Tech declined by around 25%. Startup employment follows a similar pattern: the total number of employees at t=0 declined by more than 30% between 2005 and 2014.

It is rare for startups to have high-human capital workers as employees in their first year[7]. Approximately 0.25% of employees at startups have experience working in an R&D laboratory, around 2.5% have experience working at a High Tech firm and 2% have been linked (through their earnings) with a research university. The proportion of startups that have individuals formerly paid on research grants is even smaller with fewer than 0.05% of employees being linked to a research grant from one of the 22 UMETRICS universities.

Table 1 provides some information about the characteristics of startups in their initial year of existence. The vast majority of startups, across all startup types, start off very small in their first year: 75% of all startups have fewer than 5 employees at time *t=0*; more than 50% of startups have 2 or fewer employees. Fewer than 5% of startups have more than 20 employees in the initial period. While the average revenue for startups exceeds half a million dollars per year, this measure is somewhat skewed as the median startup generates less then a quarter million dollars in their first year, with the revenue being even smaller in High Tech firms. While these size characteristics are mostly consistent across firm types, the payroll per employee and innovation measures are quite different. High Tech firms offer the highest mean payroll per employee, paying nearly twice as much as a typical startup and have innovation rates (as measured by patents and trademarks) that are 3-5x higher than the typical startup.

[7] It is important to keep in mind that the results are left-censored as the LEHD has somewhat limited coverage prior to 2002

Table 1: Startup Statistics at Year 0

All Startups	Mean	Fuzzy Median	Standard Deviation
Employment	5.6	2.0	16.5
Payroll per Employee (000s)	29.6	17.7	84.0
Revenue (000s)	540.2	232.5	958.7
Patents	0.02	-	3.1
TMs	0.06	-	0.7
High Tech Startups	**Mean**	**Fuzzy Median**	**Standard Deviation**
Employment	4.0	1.5	14.4
Payroll per Employee (000s)	54.4	39.8	64.8
Revenue (000s)	428.9	181.2	824.4
Patents	0.11	-	10.2
Trademarks	0.20	-	1.2

Source: LBD and author's calculations.
Note: Statistics calculated pooling 2005-2015 startups in the LBD and tabulating the first year statistics. Because employment figures are captured at a stationary point in time (March 12), if a firm is shown to have zero employment in their birth year, then the following year's employment is taken as the employment at *t=0*. Fuzzy medians are calculated by taking the mean of the 45th and 55th percentile levels.

The dataset also enables us to describe the human capital composition of the startup workforce. Table 2 documents the employment composition of all startups in the left hand panel and High Tech startups in the right hand panel. Individuals in startups that have at least one High Tech experienced employee are younger, less likely to be female or Black, more likely to be foreign and more likely to be Asian than other startups. Individuals in startups that have at least one university or research experienced employee are even younger, but are more likely to be female; research experienced startups are more likely to be Asian and less likely to be Black.

The demographic differences are even starker for High Tech startups. Overall employees in these startups are less likely to be female, more likely to be foreign, much less likely to be black and much more likely to be Asian. These patterns are even stronger for those with university experience and research experience.

Table 2: Startup Employee Mean Demographic Characteristics at time 0

	All Startups					High Tech				
		Startups with at least one worker with experience in:					Startups with at least one worker with experience in:			
	Total	R&D-	High Tech	University	Research	**Total**	R&D-	High Tech	University	Research
Count	**43.2M**	67,000	806,000	882,000	13,000	**1M**	21,000	416,000	48,000	1,000
Birth Year	**1973.8**	1968.7	1969.8	1980.4	1981.6	**1970.9**	1964.8	1968.6	1980.0	1979.2
Female	**45%**	44%	32%	54%	54%	**30%**	36%	27%	31%	26%
Foreign	**21%**	24%	24%	14%	18%	**25%**	24%	28%	25%	29%
White	**73%**	75%	75%	75%	70%	**74%**	80%	74%	72%	69%
Black	**12%**	7%	7%	12%	8%	**6%**	3%	5%	5%	2%
Hispanic	**16%**	10%	9%	8%	6%	**9%**	13%	8%	7%	4%
Asian	**6%**	13%	13%	8%	13%	**12%**	13%	15%	17%	19%
Other	**7%**	4%	4%	4%	8%	**7%**	2%	5%	5%	8%
Duration		4.73	5.29	2.46	1.85		5.93	6.05	2.42	2.20

Note that counts in this and subsequent tables are rounded for disclosure limitation reasons

Source: LBD combined with Individual Characteristics File (ICF)
Note: Statistics calculated pooling 2005-2015 startups in the LBD and tabulating the first year demographic statistics. Figures have been rounded for disclosure purposes. (D) indicates that the number has been suppressed for disclosure.

The literature suggests that high levels of human capital should be disproportionately valued by firms with complex production processes (32). That is borne out by our data. Even though High Tech startups account for only 4.4% of all startups in the US, they account for 17% of startups hiring at least one R&D experienced worker, 36% of startups hiring high tech workers, 6% of startups hiring university experienced workers and 8% of startups hiring research experienced workers.

Of course, the first three human capital measures, while extremely valuable in measuring potential research experience (in the same spirit, but in more detail, than older measures such as employment tenure and labor market experience), include a variety of workers. As such, a startup that hired a secretary who had been at an R&D lab would be classified as having hired an R&D experienced worker.

The direct measures offered by UMETRICS enable us to tease out the relationships in more detail. Table 3 looks at the subset of startups who hired workers who had been employed on research grants in the 22 UMETRICS universities and by funding source. In all cases, startups that hired funded researchers were more likely to be High Tech – the ratio is particularly high for those hiring individuals who worked on grants funded by the National Science Foundation, the Department of Defense and the Department of Energy.

Table 3: Distribution of Startups hiring research experienced workers by funding source

	NIH	NSF	DOD	DOE	Other Federal	Non-Federal
Number of startups hiring UMETRICS workers	3,500	1,900	700	400	5,400	3,000
Proportion of startups in High Tech	7.20%	16.80%	21.00%	17.40%	6.40%	9.40%
Ratio relative to proportion of all startups in high tech (4.4%)	1.64	3.82	4.77	3.95	1.45	2.14

Source: LBD combined with UMETRICS worker file

Note: Statistics calculated pooling 2005-2015 startups in the LBD and tabulating the funding sources for each of the UMETRICS experienced workers. UMETRICS workers can be funded through multiple agencies and startups can hire multiple UMETRICS experienced workers, so that the counts are not mutually exclusive. Figures have been rounded for disclosure purposes. (D) indicates that the number has been suppressed for disclosure.

A similar analysis can be done for the startups by the skill level of the individuals on the research grants, and is reported in Table 4. Startups hiring graduate students and faculty are much more likely to be High Tech than other startups; the pattern for undergraduate hiring is much more similar to the startup distribution as a whole.

Table 4: Distribution of Startups hiring research experienced workers by Occupation

	Faculty	Graduate Student	Post Graduate	Undergraduate	Other
Number of startups	3,500	1,900	700	400	5,400
Proportion of startups in High Tech	12.00%	15.20%	9.80%	6.00%	8.30%
Ratio relative to proportion of all startups in high tech (4.4%)	2.73	3.45	2.23	1.36	1.89

Source: LBD combined with UMETRICS worker file
Note: Statistics calculated pooling 2005-2015 startups in the LBD and tabulating the funding sources for each of the UMETRICS experienced workers. Startups can hire multiple UMETRICS experienced workers, so that the counts are not mutually exclusive. Figures have been rounded for disclosure purposes. (D) indicates that the number has been suppressed for disclosure.

Finally, the data enable us to drill down into the sectoral distribution of startups. An analysis of Table 5 shows vast compositional differences in the worker types of High Tech startups. More than 85% of all High Tech startups are in the fields of Computer Design (NAICS "5415"), Engineering (NAICS "5413") or R&D laboratories (NAICS "5417"). More than half of High Tech startups were in computer design. While there is some variation in the shares of each worker types across these industries, more than 80% of each of the worker types is affiliated with a startup in one of those 3 industries. Although only 5% of High Tech startups were R&D labs, almost two thirds of startups who hired workers with R&D experience and over one third of startups hiring workers with research experience were R&D labs.

Table 5: Industry sector of High Tech startups at Year 0

	All Startups		Startups hiring workers with			
Startup Sector	Counts	Distribution	R&D Experience	High Tech Experience	University Experience	Research Experience
AERO MANU	700	0.30%	0.18%	0.36%	0.34%	(D)
COMM MANU	700	0.30%	0.27%	0.36%	0.34%	(D)
COMP DESIGN	128,100	54.28%	14.64%	53.80%	46.21%	40.83%
COMP MANU	800	0.34%	0.27%	0.29%	0.34%	(D)
DATA PROCESS	6,700	2.84%	1.00%	2.99%	4.14%	4.17%
ENGINEER	61,500	26.06%	6.36%	28.47%	20.69%	14.17%
INFO SERVICE	8,800	3.73%	0.91%	1.82%	5.86%	5.00%
INSTRUM MANU	1,800	0.76%	0.91%	1.02%	1.03%	1.67%
INTERNET	1,300	0.55%	0.18%	0.58%	0.69%	(D)
ISP	2,600	1.10%	0.18%	1.09%	0.69%	(D)
OIL GAS	4,500	1.91%	0.18%	2.04%	1.03%	(D)
PHARMA	1,100	0.47%	1.64%	0.58%	1.03%	1.67%
RD LAB	12,900	5.47%	67.82%	3.80%	14.14%	28.33%
SEMI MANU	1,600	0.68%	0.91%	0.88%	1.03%	1.67%
SOFTWARE	3,500	1.48%	0.82%	1.75%	2.76%	4.17%
Total	**236,000**	**236,000**	**11,000**	**137,000**	**29,000**	**1,200**

Source: LBD combined with UMETRICS worker file
Note: Statistics calculated pooling 2005-2015 startups in the LBD. Figures have been rounded for disclosure purposes. (D) indicates that the number has been suppressed for disclosure.

4.3 Startup Outcomes and Human Capital Composition

This section provides some initial descriptive results about the link between workforce experience and startup outcomes (Survival to period *t+1*, Employment Growth to *t+1*,Revenue Growth to *t+1*, Patent in *t+1*, and Trademark in *t+1*). We start by first plotting the proportion of the all startups that survive to period *t+1*, have positive employment growth in *t+1*, positive revenue growth in *t+1*, file at least 1 trademark in *t+1* and file at least 1 patent in *t+1*.

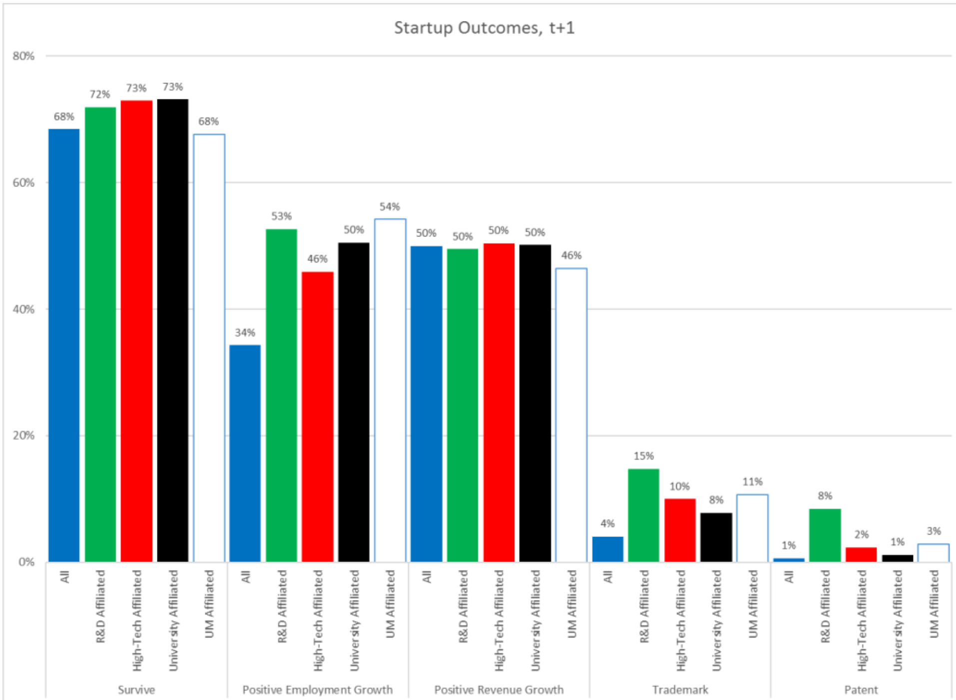


Figure 4: Outcomes of All Startups

Figure 4 provides some useful initial information about startup outcomes.   Although by and large, startups that hire workers with research experience are more likely to survive than those that do not, startups that hire UMETRICS experienced individuals are less likely to survive. This finding is consistent with those individuals being trained in basic research, which is riskier and more prone to failure than more applied work done in high tech industries and R&D Labs. Moreover, in the analyses that follow we find that higher surival rates for firms that hire high human capital workers is primarily a compositional effect. Controling for other characteristics of the startup, such as industry and size, these firms are generally less likely to survive. Consistent with an “up or out” trajectory, startups hiring high human capital individuals are more likely to see employment growth than those in the economy at large, and this is particularly true for UMETRICS startups.   The picture is a little different for revenue growth –UMETRICS startups have lower revenue growth. Patent and trademark activity are consistently substantially higher for all startups hiring experienced workers – and UMETRICS startups are second only to startups that hire R&D experienced workers in both of these dimensions of innovation. As Figure 5 shows, an almost identical pattern holds true, albeit at different levels, for high tech startups.

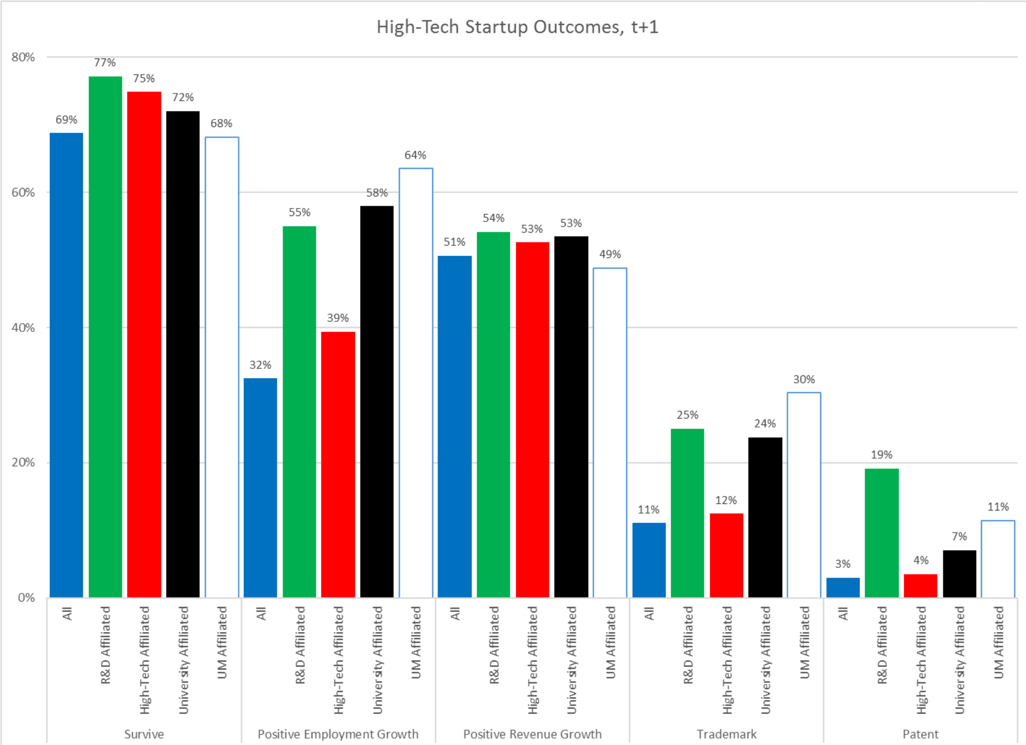


Figure 5Outcomes of High tech startups

For High Tech startups, we see a greater proportion of firms patenting and trademarking, with only modest differences across the other outcome measures.

5. Analysis

The basic framework was provided in Equation (1) and in this section, we formalize our model and control for a number of non-human capital characteristics. We assume that the functional form of Equation (1) is a linear combination of exponential functions, allowing us to use a log-linear estimation and calculate multiple outcome measures for each startup (survival, employment growth, revenue growth, patenting and trademarking) both one and five years after the birth of the firm. We regress these outcomes against the startup's workforce and other characteristics in the year of firm birth (t=0).

Our main empirical specification is as follows

$$\begin{aligned} Y_f = \alpha + \beta_1 \ln EARN_{f0} + \sum_{k=1}^{9} \delta_k SIZE_{kf0} + \beta_2 \ln \overline{AGE_{f0}} + \beta_3 \ln FEMALE_{f0} \\ + \beta_4 \ln FOREIGN_{f0} + \beta_5 \ln RD_{f0} + \beta_6 \ln HT_{f0} + \beta_7 \ln UNI_{f0} \\ + \beta_8 \ln Research\ Experience_{f0} + \varepsilon \end{aligned} \quad (2)$$

The key measures of interest are the workforce human capital measures – the number of workers who have worked in R&D performing firms, High Tech firms, universities – as well as the number who have direct research experience.

The richness of the data permit the introduction of many controls. In particular, we can include mean earnings of the firm workforce as well as firm employment size categories.
We interact demographics with each of the R&D worker types to identify potential non-linearities of being a certain type of worker (e.g. female University worker).[8]

Since the Census Bureau data does not have direct measures of technology, we control for industry, detailed geography and year using fixed effects. External macroeconomic conditions are proxied by zip code-year fixed effects and industry fixed effects.

5.1 Baseline Results

We begin by simply describing the contribution of each factor to startup outcomes. Table 6 describes the explanatory power of a group of covariates to the startup outcomes of survival, employment growth, revenue growth, patenting and trademarking in the next period
Table 6 shows that just controlling for location and industry fixed effects can explain a small share of the variance in outcomes. Including initial firm characteristics, such as employment size and mean earnings at *t=0*, contributes a significant share to all of the outcomes. Including demographic controls, such as the mean age of the employees, number of female employees, foreign-born status and race, has large explanatory power in future employment growth, but relatively little explanatory power on revenue, survival and innovation. Including our basic human capital measures leads to an insignificant increase in the explanatory power of the model in survival and employment growth across all firms, but has does have significant power in our model for revenue growth, patenting and trademarking. In particular, the human capital elements contribute an additional 40% in explanatory power for patenting outcomes in the following period and an additional 10% in explanatory power for trademarking. These patterns continue to hold for High Tech startups with human capital contributing an additional 25% in explanatory power for patents and an additional 4.5% in revenue and 4.7% in trademarking. This table highlights the explanatory power of human capital in relation to startup growth and innovative outcomes.

[8] Note that these interaction terms are the result of multiplying continuous counts of employees falling into each group and that any given employee may belong to any number of designated groups.

Table 6: Explanatory power (R^2) of Startup Covariates

All Startups	Survival, t+1	Employment Growth, t+1	Revenue Growth, t+1	Patent, t+1	TM, t+1
Geography-Year and Industry Dummies only	0.230	0.019	0.026	0.014	0.041
Geography-Year and Industry Dummies+ Initial Firm Characteristics	0.342	0.184	0.027	0.016	0.049
Geography-Year and Industry Dummies+ Initial Firm Characteristics + Demographics	0.344	0.303	0.031	0.017	0.050
Geography-Year and Industry Dummies+ Initial Firm Characteristics + Demographics + Human Capital	0.344	0.303	0.032	0.029	0.056
Share of Explained Variance Explained by Human Capital	**0.1%**	**0.3%**	**3.1%**	**41.4%**	**10.7%**
High Tech Startups	Survival, t+1	Employment Growth, t+1	Revenue Growth, t+1	Patent, t+1	TM, t+1
Geography-Year and Industry Dummies only	0.248	0.071	0.067	0.058	0.084
Geography-Year and Industry Dummies+ Initial Firm Characteristics	0.354	0.218	0.07	0.072	0.113
Geography-Year and Industry Dummies+ Initial Firm Characteristics + Demographics	0.355	0.371	0.085	0.078	0.123
Geography-Year and Industry Dummies+ Initial Firm Characteristics + Demographics + Human Capital	0.358	0.377	0.089	0.104	0.129
Share of Explained Variance Explained by Human Capital	**0.8%**	**1.6%**	**4.5%**	**25.0%**	**4.7%**

Table 7 provides the key results associated with the full regression. Briefly, the relationship between the different measures of human capital and startup survival and growth (both in terms of employment and revenue) is measurable and quite large. Startups that employ workers with experience working in R&D Labs, High Tech and universities are less likely to survive. Our human capital measures are clearly associated with positive employment and revenue growth. Using the fully controlled specification, our results suggest that employing 1 additional R&D worker is associated with a 1.6 percentage point increase in employment (conditional on survival). This figure increases to 4.7 percentage points for one additional High Tech worker, and 4.3 percentage point for a former university employee. We see similar patterns in revenue growth. For all startups, the hiring of one additional high human capital worker is associated with a 1.5 - 4.7 percentage point increase in employment and a 2.8 - 6 percentage point increase in revenue growth (conditional on survival). We see fairly large coefficients on the patenting and trademarking outcomes for R&D lab workers, with the addition of one R&D lab worker contributing a 10.5 percentage point increase in patent filing and 8.5 percentage point increase in trademark filing.

Table 7: OLS on All Startup Outcomes, 2005-2015

	Survival, t+1	Employment Growth, t+1	Revenue Growth, t+1	Patent, t+1	TM, t+1
$\ln RD_{f0}$	-0.0481***	0.0156*	0.0456***	0.105***	0.0849***
	(0.00407)	(0.00717)	(0.0127)	(0.0136)	(0.0134)
$\ln HT_{f0}$	-0.0268***	0.0474***	0.0596***	0.0121***	0.0488***
	(0.00333)	(0.00415)	(0.00384)	(0.000772)	(0.00311)
$\ln UNI_{f0}$	-0.0177***	0.0431***	0.0282***	0.00541***	0.0299***
	(0.00215)	(0.00416)	(0.00536)	(0.000915)	(0.00319)
Other Controls	Yes	Yes	Yes	Yes	Yes
Observations	4,930,000	3,370,000	1,910,000	4,930,000	4,930,000
R-squared	0.344	0.303	0.032	0.029	0.056
Startups that hired UMETRIC university employees: overall					
$\ln RESEARCH_{f0}$	-0.00902*	0.0204*	0.0272+	0.0139***	0.0180***
	(0.00357)	(0.00858)	(0.0161)	(0.00175)	(0.00396)
Firm controls	Yes	Yes	Yes	Yes	Yes
Observations	68,000	45,000	17,000	68,000	68,000
R-squared	.567	.397	.148	.109	.146
Startups that hire UMETRIC university employees: Decomposed by funding source					
NIH	-0.00662	0.0440**	-0.00850	0.0141***	0.0210**
	(0.00612)	(0.0144)	(0.0262)	(0.00299)	(0.00679)
NSF	-0.00852	0.0432*	0.0506	0.0259***	0.0313**
	(0.00864)	(0.0204)	(0.0381)	(0.00420)	(0.00954)
DOD	-0.00217	-0.0158	0.0615	0.0528***	0.0235
	(0.0134)	(0.0313)	(0.0551)	(0.00649)	(0.0147)
DOE	-0.0127	-0.0222	0.174*	0.0452***	-0.0432*
	(0.0177)	(0.0415)	(0.0787)	(0.00865)	(0.0196)
Other Federal Funding	-0.00594	0.0192+	-0.0109	-0.00605*	-0.00507
	(0.00486)	(0.0115)	(0.0212)	(0.00237)	(0.00538)
Non-Federal Funding	0.000349	0.0108	0.0558+	0.00217	0.0225**
	(0.00670)	(0.0161)	(0.0309)	(0.00326)	(0.00740)
ARRA	-0.00334	-0.0536	-0.0854	-0.0231	-0.0192
	(0.0296)	(0.0704)	(0.146)	(0.0144)	(0.0328)
Firm controls	Yes	Yes	Yes	Yes	Yes
R-squared	.567	.397	.148	.109	.146
Startups that hire UMETRIC university employees: Decomposed by Occupation					
Faculty	-0.0143	-0.0926**	-0.0151	0.0566***	0.00230
	(0.0146)	(0.0338)	(0.0586)	(0.00708)	(0.0161)
Graduate Student	-0.0204*	0.0225	0.0578	0.0416***	0.0289**
	(0.00921)	(0.0223)	(0.0429)	(0.00449)	(0.0102)
Post-Grads	-0.00804	-0.127***	-0.0297	0.0430***	-0.00418
	(0.0164)	(0.0383)	(0.0692)	(0.00800)	(0.0182)
Undergraduate	-0.00713	0.0784***	0.0461+	0.00192	0.00889
	(0.00525)	(0.0126)	(0.0241)	(0.00257)	(0.00583)
Other (Admin, Technician)	-0.00605	0.0251*	0.0237	0.00658**	0.0242***
	(0.00499)	(0.0118)	(0.0213)	(0.00244)	(0.00554)
Firm controls	Yes	Yes	Yes	Yes	Yes
R-squared	.567	.397	.148	.109	.146

Clustered Robust Standard Errors in Parentheses (by 4-digit Industry-Year). + p<0.10, *p<0.05, **p<0.01, ***p<0.001; controls included for size and average earnings, proportion of workforce that is female, foreign born, and interactions of female, foreign born with all of the different types of research experience (e.g. Foreign female R&D lab workers). Observations have been rounded for disclosure purposes.

The second panel of Table 7 reports the results for the subset of startups that hired employees from the 22 institutions that provided UMETRICS data. The interpretation of the coefficient is thus relative to the effects of hiring an individual trained on a research grant over and above those simply with experience of working in one of these 22 universities. The results are extremely consistent. Startups that hired research trained individuals were more likely to fail than those who simply hired only university experience individuals (which are in turn more likely to fail than other startups, as established in the first panel). However, those that survive are more likely to create jobs, have higher revenue, as well as patent and trademark. Again, these results are over and above the significantly higher relationship demonstrated by startups hiring university employees relative to all startups in the first panel.

The third and fourth panel of Table 7 delves more deeply into the types of projects and skill embodied within our human capital measure. Startups that hire workers funded DOD and DOE grants are much more likely to patent, again relative to startups that hire non-research trained workers at these universities. Startups that hire workers trained on NIH and NSF funded grants see greater employment growth. Interestingly, faculty, graduate students, and post-grads contribute more to patenting and trademark activity while undergraduates are associated with greater employment growth.

Table 8 reports estimates similar to the top panel of Table 7 (with the full set of controls) but for High Tech startups. The results are substantively unchanged. Our human capital measures have a negative impact on survival, but a significant and positive association with employment growth and revenue growth conditional on survival. The magnitude of the coefficients are also significantly larger than the coefficients in the previous table, which confirms our hypothesis that High Tech startups would be most sensitive to measures of human capital. In the case of employment growth, the hiring of a high human capital employee is associated with a 2.9 to 9.3 percentage point increase in employment growth and a 6.3 to 8.8 percentage point increase in revenue growth for High Tech firms. The addition of an R&D lab experienced worker is associated with an 18.2 percentage point increase in patenting and an 11.4 percentage point increase in trademarking.

Unfortunately, disclosure limitation protocols preclude us from doing a deeper dive using UMETRICS only data.

Table 8: OLS on High Tech Startup Outcomes, 2005-2015

	Survival, t+1	Employment Growth, t+1	Revenue Growth, t+1	Patent, t+1	TM, t+1
$\ln RD_{f0}$	-0.0515***	0.0287	0.0632*	0.182***	0.114***
	(0.00706)	(0.0146)	(0.0305)	(0.0211)	(0.0239)
$\ln HT_{f0}$	0.0423***	0.0823***	0.0865***	-0.00551*	0.00308
	(0.00549)	(0.00366)	(0.00638)	(0.00234)	(0.00417)
$\ln UNI_{f0}$	-0.00633	0.0933***	0.0879***	0.0142*	0.0711***
	(0.00429)	(0.00748)	(0.0127)	(0.00648)	(0.0137)
Other Controls	Yes	Yes	Yes	Yes	Yes
Observations	210,000	140,000	95,000	210,000	210,000
R-squared	0.358	0.377	0.089	0.104	0.129

Robust Standard Errors in Parentheses. *p<0.05, **p<0.01, ***p<0.001; controls included for size and average earnings, proportion of workforce that is female, foreign born, and interactions of female, foreign born with research experience.

In addition to these tables, we have estimated the same specification over different size groups of startups and find that the results are robust and do not differ greatly. To summarize our empirical findings, we find mostly positive and significant associations between R&D-experience, High Tech experience, university experience and research-trained experience with startup performance. These human capital measures are associated with much riskier outcomes: survival of such startups is significantly less likely. However, conditional on survival, these basic measures of human capital have positive and significant effects on employment growth and revenue growth for the following period. The explanatory power of these measures is surprisingly high, contributing more than 15% to the cumulative explanatory power of High Tech startup employment growth.

6. Conclusion

This paper leverages new data about workforce human capital that can be used to provide more insights into the survival, employment growth and innovative activity of new businesses. These results are consistent with the view that there is a relationship between workforce experience and business startup outcomes. While it is important to note that the cumulative magnitude of the effects of these human capital measures on startup outcomes is relatively small, it is important to consider that these are very basic measures of human capital (binary and extensive margin type measures),

As always, there is much more to be done. In future work we will expand the analysis of research experience to capture network effects as well as the effects of intensive exposure to research intensive environments. We will also examine a broader set of outcome measures, including for startups that went public or became exceptionally large. It is always difficult to identify causal relationships, but we have begun to investigate the effects of sharp changes in funding, such as the 2009 American Recovery and Reinvestment Act (ARRA), as well as changes in funding to different research areas.

References

1. J. I. Lane, J. Owen-Smith, R. F. Rosen, B. A. Weinberg, New linked data on research investments: Scientific workforce, productivity, and public value. *Res. Policy* (2015), doi:10.1016/j.respol.2014.12.013.
2. N. Bania, R. W. Eberts, M. S. Fogarty, Universities and the startup of new companies: can we generalize from route 128 and Silicon valley? *Rev. Econ. Stat.*, 761–766 (1993).
3. R. A. Lowe, C. Gonzalez-Brambila, Faculty entrepreneurs and research productivity. *J. Technol. Transf.* **32**, 173–194 (2007).
4. N. Hausman, “University Innovation, Local Economic Growth, and Entrepreneurship” (2012), (available at http://ideas.repec.org/p/cen/wpaper/12-10.html).
5. U. Akcigit, W. R. Kerr, “Growth through heterogeneous innovations” (National Bureau of Economic Research, 2010).
6. J. Guzman, S. Stern, “The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014” (National Bureau of Economic Research, 2016).
7. D. B. Audretsch, M. C. Keilbach, E. E. Lehmann, *Entrepreneurship and economic growth* (Oxford University Press, 2006).
8. H. McGuirk, H. Lenihan, M. Hart, Measuring the impact of innovative human capital on small firms’ propensity to innovate. *Res. Policy*. **44**, 965–976 (2015).
9. E. L. Glaeser, W. R. Kerr, G. A. M. Ponzetto, Clusters of entrepreneurship. *J. Urban Econ.* **67**, 150–168 (2010).
10. C. Syverson, “What determines productivity?” (National Bureau of Economic Research, 2010).
11. S. P. Kerr, W. Kerr, Ç. Özden, C. Parsons, Global talent flows. *J. Econ. Perspect.* **30**, 83–106 (2016).
12. E. P. Lazear, K. L. Shaw, Personnel Economics: The Economist’s View of Human Resources. *J. Econ. Perspect.* **21**, 91–114 (2007).
13. S. Bender, N. Bloom, D. Card, J. Van Reenen, S. Wolter, “Management practices, workforce selection and productivity” (National Bureau of Economic Research, 2016).
14. N. Bloom, R. Lemos, R. Sadun, D. Scur, J. Van Reenen, JEEA-FBBVA Lecture 2013: The new empirical economics of management. *J. Eur. Econ. Assoc.* **12**, 835–876 (2014).
15. P. Tambe, Big data investment, skills, and firm value. *Manage. Sci.* **60**, 1452–1469 (2014).
16. C. Corrado, C. Hulten, D. Sichel, in *Measuring capital in the new economy* (University of Chicago Press, 2005), pp. 11–46.
17. S. E. Black, L. M. Lynch, Human-capital investments and productivity. *Am. Econ. Rev.* **86**, 263–267 (1996).
18. A. P. Bartel, N. D. Beaulieu, C. S. Phibbs, P. W. Stone, Human capital and productivity in a team environment: evidence from the healthcare sector. *Am. Econ. J. Appl. Econ.*, 231–259 (2014).
19. P. M. Romer, Endogenous technological change. *J. Polit. Econ.* **98**, S71–S102 (1990).
20. F. Lafontaine, K. Shaw, Serial Entrepreneurship: Learning by Doing? *J. Labor Econ.* **34**, S217–S254 (2016).
21. C. I. Jones, Sources of US economic growth in a world of ideas. *Am. Econ. Rev.* **92**, 220–239 (2002).

22. D. Acemoglu, U. Akcigit, N. Bloom, W. R. Kerr, “Innovation, reallocation and growth” (National Bureau of Economic Research, 2013).
23. S. Kantor, A. Whalley, Research proximity and productivity: long-term evidence from agriculture. *Rev. Econ. Stat. Forthcom.* (2014).
24. S. Kantor, A. Whalley, Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks. *Rev. Econ. Stat.* **96**, 171–188 (2013).
25. G. Duranton, D. Puga, Micro-foundations of urban agglomeration economies. *Handb. Reg. urban Econ.* **4**, 2063–2117 (2004).
26. M. S. Gertler, Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *J. Econ. Geogr.* **3**, 75–99 (2003).
27. L. Fleming, I. I. I. Charles King, A. Juda, Small Worlds and Regional Innovation. *Organ. Sci.* **18**, 938–954 (2007).
28. M. Marx, J. Singh, L. Fleming, Regional disadvantage? Employee non-compete agreements and brain drain. *Res. Policy*. **44**, 394–404 (2015).
29. N. Zolas *et al.*, Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients. *Science (80-.).* **350**, 1367–1371 (2015).
30. J. M. Abowd, J. Haltiwanger, J. Lane, Integrated Longitudinal Employer-Employee Data for the United States. *Am. Econ. Rev.* **94**, 224–229 (2004).
31. J. M. Abowd *et al.*, in *Measuring capital in the new economy* (University of Chicago Press, 2005), pp. 153–204.
32. E. Barth, J. Davis, R. B. Freeman, Augmenting the Human Capital Earnings Equation with Measures of Where People Work. *Natl. Bur. Econ. Res. Work. Pap. Ser.* **No. 22512** (2016), doi:10.3386/w22512.
33. C. Ichniowski, K. Shaw, G. Prennushi, The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *Am. Econ. Rev.* **87**, 291–313 (1997).
34. D. E. Hecker, High-technology employment: a NAICS-based update. *Mon. Lab. Rev.* **128**, 57 (2005).
35. N. Goldschlag, J. Miranda, Business Dynamics Statistics of High Tech Industries (2016).
36. S. J. H. Graham, C. Grim, T. Islam, A. C. Marco, J. Miranda, Business dynamics of innovating firms: Linking US patents with administrative data on workers and firms (2015).
37. E. Dinlersoz, N. Goldschlag, A. Myers, N. Zolas, in *Measuring and Accounting for Innovation in the 21st Century*, C. Corrado, J. Miranda, J. Haskel, D. Sichel, Eds. (NBER, 2018).