# More than an Ivory Tower:
# The Impact of Research Institutions on the Quantity and Quality of Entrepreneurship[1]

*Version 1.0*

Valentina Tartari, CBS

Scott Stern, MIT and NBER

**Abstract**

While universities are often associated with strong entrepreneurial ecosystems, the underlying drivers of this relationship have proved more elusive. Universities are often located in economic and innovation environments conducive to growth-oriented entrepreneurial activity, are themselves a source of (potentially skill-biased) local demand, and produce knowledge, which might serve as the foundation for new ventures. The principal contribution of this paper is to provide systematic empirical evidence for the distinctive role universities play in shaping local entrepreneurial ecosystems. Our analysis includes three key steps. First, we combine comprehensive business registration records with a predictive analytics approach to measure both the quantity and (growth-oriented) quality of entrepreneurship at the level of U.S. zip codes on an annual basis. Second, we link each location to the presence or absence of research-oriented universities or national laboratories. Finally, we exploit significant changes over time in Federal commitments to both universities and national laboratories, as well as differences in whether Federal support is directed towards research versus other activities. These building blocks allow us to estimate the impact of changes in Federal support (for universities or national labs, for research or more general expenses) on the subsequent quantity and quality-adjusted quantity of entrepreneurship in the local environment. We find, first, that the presence of a university is associated not only with a higher level of entrepreneurship but also higher level of quality-adjusted entrepreneurship. Second, changes over time in Federal research commitments to universities (but not national laboratories) are uniquely linked to subsequent changes in the quality-adjusted quantity of entrepreneurship. These findings suggest both that universities as large economic institutions play a critical (and often underappreciated) role in local economic development, but that the norms and governance of universities play a unique role in promoting growth entrepreneurship conducive to long-term economic growth.

[1] Scott Stern, MIT, Sloan School of Management and NBER, sstern@mit.edu; Valentina Tartari, Copenhagen Business School, vt.si@cbs.dk. We would like thank Jorge Guzman, Pierre Azoualy, Fiona Murray, Jeff Furman, and seminar participants at Copenhagen Business School, DRUID, the AIEA-NBER Conferencs, and particpants in the MIT Regional Ecosystem Acceleration Program for useful comments and feedback, and Luca Gius for excellent research assistance. This work was supported by the Jean Hammond (1986) and Michael Krasner (1974) Entrepreneurship Fund and the Edward B. Roberts (1957) Entrepreneurship Fund at MIT, and the Kauffman Foundation. Valentina Tartari gratefully acknowledges the support from the Independent Research Fund Denmark through a Sapere Aude Research Talent Grant (DFF –4003-00137B), and Stern gratefully acknowledges the hospitality and hosting of Copenhagen Business School. All errors and omissions are of course our own.

## I. Introduction

A distinguishing feature of publicly funded research organizations such as universities or national laboratories is that they are situated in a single (or small number) of locations. As geographically bounded institutions, universities and national laboratories are often invoked, by both economists and policymakers, as drivers of local knowledge spillovers (Acs et al., 1992; Jaffe, 1989; Jaffe et al., 1993; Moretti, 2012) and local economic development (Adams et al., 2003; Hausman, 2012; Kantor and Whalley, 2014).

A particularly important route through which research institutions might influence local outcomes is through entrepreneurship. On the one hand, both universities and national laboratories serve as a source of local demand and produce frontier knowledge that might form the basis of new companies. But, while the impact of research institutions as a source of local demand is likely to enhance the quantity of entrepreneurs, the ability to leverage frontier research can play a critical role in the formation of companies with a high potential for significant growth at founding. In other words, public research institutions may not simply increase the quantity of entrepreneurs but also their average quality. As well, though both universities and national laboratories conduct research at significant scale, universities are distinctive in also producing students (who might launch careers as entrepreneurs in the local area) and being governed by policies and rules that encourage openness and diffusion to the local environment (Valero and Van Reenen, 2019).

To a significant extent, the reputation of universities and related institutions as drivers of local entrepreneurship and economic development is grounded in both academic and policy discussions of leading entrepreneurial ecosystems such as Silicon Valley (anchored by institutions such as Stanford, Berkeley, and UC-SF), Boston (with at least 10 research institutions), and the Research Triangle Park in North Carolina. For example, in the late 1980s, a group of five North Carolina State engineers conducting pioneering research on the advantages silicon carbide for light-emitting diode (LED) founded a start-up company, Cree. Given their ties to the local community, Cree was founded on the outskirts of the North Carolina State campus, and, over the next 25 years, established itself as a leader in the emerging LED industry, employing more than 6000 workers (most in the local region). Inspired by such examples, a significant body of work emphasizes the importance of geographically mediated academic entrepreneurship as an outcome of university and national laboratory research, with particular emphasis on the role of academic entrepreneurship in the clustering of innovation-oriented entrepreneurial activity (Audretsch and Stephan, 1996; Feldman, 1994; Kenney, 1986; Zucker et al., 1998).

Despite this significant interest, establishing a causal linkage between the activities of public research institutions (universities and national laboratories) and subsequent patterns of entrepreneurship has been elusive. Largely, prior research has focused on the linkage between a specific type of research conducted at a given university or national laboratory (e.g., life sciences) and subsequent entrepreneurship and economic activity (such as productivity growth) in industries linked to that research (e.g., biotechnology firms or agriculture) (Audretsch et al., 2005; Baptista and Mendonça, 2010; Kantor and Whalley, 2019). Though these sectoral studies provide evidence for the linkage between particular research streams and related start-up activity, they do not provide evidence about the overall impact of public research institutions on the general level and nature of entrepreneurship in a given location. This is particularly important as most research institutions conduct a wide range of types of research, and the average impact of research conducted at public research institutions may be quite different than the impact of research in a particularly fertile area (such as biotechnology). At the same time, while several studies examine the cross-sectional correlation between public research institutions (usually focusing on universities) and entrepreneurship (often focusing on entrepreneurship in R&D-intensive industries) (Audretsch et al., 2005; Audretsch and Lehmann, 2005; Bonaccorsi et al., 2014), existing research has not been able to disentangle the systematic impact of public research institutions on entrepreneurship from unobserved factors contributing both to the strength of the university system and to the rate of local entrepreneurship.[2] Alongside this work, another stream of research investigates more broadly the impact of different universities activities on a variety of economic outcomes. Adams (1990) investigates the impact of stocks of knowledge (embodied in academic publications) on productivity growth nationwide. Andersson and colleagues (2009, 2004) analyze the impact of a university decentralization policy in Sweden on the average

[2] An important partial exception is Baptista et al. (2011) that consider the impact of the entry of a small number of Portugese universities between 1992 and 2002 on the growth rate of young firms as a share of the overall business population, finding a positive impact on knowledge-intensive firms and a negative (and noisy) effect on the overall entry of firms in other sectors, such as low-tech manufacturing.

productivity of workers, finding important and significant positive effects. Moretti (2004) links the presence of land grant universities to economic growth through increased human capital in the area. Kantor and Whalley (2014) use changes in endowment spending caused by fluctuations in the stock market to identify the effects of university activities on local wages. They find that a 10% increase in higher education spending in an urban county increases the average wage in the non-education sector by 0.8%. Effects are persistent and stronger in industries that rely more on university outputs, such as graduates.

While existing literature has made important progress in linking universities and economic activity, the purpose of this paper is to look more specifically at the creation of new businesses by developing an integrated estimate of the impact of universities and national laboratories on both the quantity and quality-adjusted quantity of entrepreneurship. To do so, we undertake three key steps.

First, leveraging data from the Startup Cartography Project (Andrews, et al, 2020), we develop a consistent set of zip-code-level measures of both the quantity and quality-adjusted quantity of entrepreneurship across 34 U.S. states from 1988-2012. Second, for each of these zip codes, we identify both the presence and absence of universities and national laboratories co-located nearby to that zip code, as well as a rich set of covariates that allow us to observe zip codes that are otherwise "similar" to those that are located next to a research institution. Finally, and most critically, we take advantage of the fact that, for a given research institutions (university or national lab), there are significant changes over time in the allocation of Federal funds (both research-oriented and more general) to that institution, and the allocation of the bulk of these funds is through competitive grant processes that is independent of future changes to the local entrepreneurial ecosystem (Federal research funding is driven by grants based on scientific merit).

We combine these data to undertake two broad types of analyses. First, we undertake a cross-sectional evaluation of the link between research institutions and both the quantity and quality-adjusted quantity of entrepreneurship across locations. The cross-sectional analysis points to the broad colocation of university and a high quality-adjusted quantity of entrepreneurship, particularly in areas that are urban, in historically high-income zip codes, and for zip codes collocated with a top-tier university. Moreover, there has been a sizeable shift in the link between research institutions and entrepreneurship over time. First, regardless of whether there is a university or national lab in a location, the types of locations that host research institutions seem to have become more associated with the quantity of entrepreneurship over time. However, even after accounting for these underlying place-based characteristics, there is a sharp (and cyclical) uptick in the association between universities and the quality-adjusted quantity of entrepreneurship.

Building on these descriptive findings, we then turn to the heart of our analysis, where we focus on the causal impact of changes over time in the allocation of Federal resources (in terms of research and non-research funding) on the subsequent level of entrepreneurship. To address the main potential sources of bias, we implement both a fixed effects panel data estimator, as well as first-differences instrumental variables estimators that account for the potential for dynamics within a given entrepreneurial ecosystem. The results are striking. On the one hand, we find that a positive shock in the allocation of Federal resources (both research and non-research, to universities or national laboratories) is associated with a meaningful increase in the quantity of entrepreneurship. This is in line with the idea that, as large local economic institutions, universities induce local demand for entrepreneurship of all types. However, when we turn to an analysis of the impact of Federal resources on the quality-adjusted quantity of entrepreneurship (i.e., accounting for the potential for growth and spillovers from these start-ups), we find a very sizeable positive impact of Federal research expenditures, but a negative impact on non-research resources and a small (and negative) impact on research resources to national laboratories. From a quantitative perspective, an increase in $12.3 million for research to a university is linked to a sufficient level of growth-oriented entrepreneurial activity to seed the founding of one firm that will ultimately achieve a significant level of growth and scale (i.e., an IPO or successful acquisition within six years of founding).

Overall, these findings document the distinct impact of universities and national laboratories on local entrepreneurial ecosystem. There are two key distinctions that are worthwhile to note. First, there seems to be a meaningful distinction between research activity and non-research activity at universities; whereas non-research activity raises the quantity of entrepreneurship (but lowers its average quality), research activity increases both the quantity and quality-adjusted quantity of entrepreneurship. Second, there is an important distinction between universities and national laboratories. While these are the two primary institutions conductions publicly funded research in the United States, university research activities seem to be more closely linked to spillovers to local entrepreneurial ecosystems, whereas the more cloistered nature of national laboratories is associated with a much more muted (or even perhaps negative) impact.

These results highlight the important and unique role universities play in building and sustaining entrepreneurial ecosystems. Because of their norms and missions, universities seem to be able to spur economic development more widely than other research institutions, and also beyond the simple fact of being located in high-growth locations. We also highlight how universities have become more relevant over time for local entrepreneurial ecosystems, providing support for the increasing policy focus on the role of universities in shaping local economic development.

## II. The Impact of Public Research Institutions on Entrepreneurship

The central objective of this paper is to consider the empirical impact of research universities and national laboratories on entrepreneurship. We begin our analysis by briefly synthesizing g the distinctive economic properties of these public research institutions, and then draw out the implications of these properties for entrepreneurship in the context of a simple model.

It is useful to start by highlighting a central feature of the US national innovation system: the highly decentralized performance of publicly funded scientific and engineering research across a wide variety of institutions and domains (Mowery and Rosenberg, 1993; Rosenberg and Nelson, 1994). In particular, though the classical rationale for public investment in research arising from lack of appropriability is consistent with simply funding the research activities of private firms (Arrow, 1972; Nelson, 1959), the majority of public funding of research (as opposed to development) is concentrated in public and non-profit institutions, most notably universities and national laboratories. There are more than 260 research universities scattered across the United States receiving more than 30 billion USD annually from public research funds, and 46 national laboratories engaging in research with more than 19 billion USD research funds per year. When research is performed in an institution such as a national laboratory or university that only conducts research but face significant restrictions in direct commercialization, the social return from that public research investment depends crucially on whether research is transferred to the private sector. Though direct transfer to established firms is an important channel for such knowledge transfer (Furman and MacGarvie, 2007), the establishment of new firms leveraging new knowledge may be a particularly important channel for realizing the benefit of publicly funded research (Baptista et al., 2011; Kim et al., 2012).

As research-enhancing economic institutions (Furman and Stern, 2011; Valero and Van Reenen, 2019), universities and national labs share a number of distinctive properties and also display some key differences that shape their potential impact on entrepreneurship. On the one hand, both universities and national laboratories engage in a wide variety of different types of research, including both basic and applied research, and there are examples of important fundamental breakthroughs (as well as more modest discoveries) from both types of institutions. As well, both employ a workforce highly skewed towards those with a high level of formal training in scientific and engineering fields, the bulk of public research funding is assigned to individualized researchers pursuing well-defined research projects at a moment in time, and, to first approximation, researchers are allowed to disclose their research findings to the broader scientific and engineering community through publication and other types of disclosure (Dasgupta and David, 1994; Stephan, 2012).

Importantly, both universities and national laboratories are geographically bounded, with physical campuses in specific locations (and usually only a small number of satellite locations). Interestingly, the place-based nature of these research-enhancing institutions suggests that they will serve potentially as a source of knowledge for potential local spillovers. As well, universities and national laboratories are important economic agents within a given geography (e.g., as a source of local demand). As a source of economic density, universities concentrate demand in a particular place, thus encouraging the formation of new businesses near that place. This channel is of course particularly salient at the earliest stages of a university (Andrews, 2020) or in response to a change in the scale of university activities (i.e., in equilibrium, there would be a sustained flow of new businesses, but not necessarily different from other places).

Despite these significant similarities between universities and national laboratories, a number of key distinctions point towards the potential for differences in their likely impact on entrepreneurship in the local environment. Most notably, whereas universities combine the education and training of students alongside the pursuit of research, national labs by and large are focused more narrowly on research (Adams et al., 2003; Jaffe and Lerner, 2001). Not simply a difference in the scope of their activities, students may play an important (and often underappreciated) role in transferring knowledge produced by universities into private sector activity, including the founding of new firms by students directly as well as the hiring of students into start-ups for the purpose of leveraging their frontier knowledge.

Though the presence or absence of students and an educational function is likely the primary contrast between universities and national laboratories, universities also likely maintain practices and offer incentives that enhance the fluidity between research and industry, and may focus on a broader range of research problems that may be more applicable for local economic development. The most obvious channel through which knowledge flows from universities and other research organizations involves scientific research published in academic journals. This research is produced locally but it is distributed globally, and it is in principle available for anyone to use, independent of their geographical location. Yet, empirical evidence points to a different pattern. Jaffe and colleagues (1993) show that patents tend to cite other patents produced by organizations (both universities and corporations) that are located nearby. Even more strikingly, Adams (2002) finds that knowledge flows from universities tend to be much more local in nature than spillovers from firms, highlighting the apparent paradox that institutions whose mandate is to produce public knowledge, such as universities, tend to benefit disproportionately local businesses. Adams goes on arguing that it is precisely because of the open nature of the knowledge that is produced by universities, that we observe firms gravitating around academic institutions: as knowledge and information, especially if they are highly tacit in nature, do not transmit without costs, firms locate close to universities to absorb knowledge which is "reasonably current and not propriety" (Adams, 2002, p. 274). The importance of oral transmission of knowledge is indeed one of the main reasons why people tend to cluster in cities notwithstanding the increased costs (Lucas, 1988). This channel, which is present for all firms, may be even more relevant for start-ups (Audretsch and Thurik, 2001) as they rely more heavily on externally produced knowledge (Hall et al., 2003) due to the lower amount of resources they can devote to internal R&D. In other words, while both universities and national labs conduct research, the dual mission of universities and related practices may have the consequence that universities are more likely to serve as a source of knowledge for start-up activity in the local environment, at least in the short to medium term.

We draw out the implications of these distinctive properties of universities and national laboratories as local economic institution in the context of a simple model identifying the impact of each type of research-enhancing institution on entrepreneurship. Suppose that, in each period, there exists within a given location $j$, $d_j$, a potential set of entrepreneurs normalized to be equal to a unit mass. Each of these entrepreneurs draws an idea whose operating profits are equal to the quality of the idea, $\theta_i \sim f(\theta;\overline{\theta})$, scaled by the size of the market, $d_j$.[3] Entry into the market involves a sunk cost, $F$, so that the expected profits from entering are simply equal to $\theta_i d_j - F$. For entrepreneurs in a given period and location, then, the decision to enter will depend on whether the underlying quality of the idea is sufficiently high and the market is sufficiently large to pay for the sunk cost of entry (i.e., $\theta > \theta_j^* = \frac{F}{d_j}$), a cut-off determining the overall entry level $\lambda_j = \int_{\theta_j^*}^{\infty} f(\theta) d\theta$.

Publicly funded research institutions play two distinct roles – as a source of knowledge and demand -- in shaping entrepreneurship. Specifically, increased resources devoted to a research institution (e.g., through an external increase in funding or increased activity resulting from educating additional students) serves to increase the level of local demand, $d_j$. However, a simple increase in $d_j$ does not enhance the quality of the distribution from which ideas are drawn. Indeed, while an increase in $d_j$ increases $\lambda_j$, the marginal entrants induced by such an expansion will be of low quality (i.e., $\frac{\partial \theta_j^*}{\partial d_j} \leq 0$). In contrast, a shift in the level of *research-oriented* resources towards a public research institution also has the potential to shift the distribution from which ideas are drawn (by shifting $\overline{\theta}$). By shifting the distribution of the quality of ideas itself, research aimed at universities can actually shift not only the quantity but the quality-adjusted quantity of

[3] We further assume that the private value and social value of ideas are proportional to each other. In other words, though entry exerts additional surplus to consumers (and perhaps to follow-on entrepreneurs through imitation), the ratio of the private and social value to entrepreneurship are fixed.

entrepreneurship ( $\frac{\partial \theta_j^*}{\partial \bar{\theta}_j} \geq 0$ ). Of course, this comparative static only holds in the case when research conducted in the public institution is accessible to local entrepreneurs, either through embodied human capital in the form of students or through disembodied knowledge spillovers (which might themselves depend on the presence of open institutions). In either case, we hypothesize that the impact of a shift in research resources for a public institutions on local quality-adjusted entrepreneurship will be larger for universities than for national laboratories.

Putting these ideas together, our analysis so far highlights the idea that realizing the social return from public investments in universities and national laboratories requires the transfer of the research conducted in these institutions into private sector activity, including entrepreneurship. Specifically, our model suggests that an increase in the level of economic activity at a public research institution is predicted to have an increase on the quantity of entrepreneurship. However, only an increase in research-focused expenditures is predicted to impact the quality-adjusted quantity of entrepreneurship, and this linkage between research expenditures and subsequent quality-adjusted quantity is predicted to be higher for universities as the result of their degree of openness and their engagement in teaching.

## III. The Measurement of Entrepreneurship

The first task that we face is how to develop consistent measures of entrepreneurial quantity and quality on a consistent and granular basis so that we can examine how the presence or absence of a research institution, or changes in the economic and research activities of research institutions over time, influence subsequence entrepreneurship within a given local area. To do so, we leverage data from the Startup Cartography Project (Andrews et al., 2020). Building on Guzman and Stern (2017, 2015), the Startup Cartography Project data leverage a systematic approach for measuring the founding quality of startups using predictive analytics and business registration records. The approach combines three interrelated insights. First, as the challenges to reach a growth outcome as a sole proprietorship are formidable, a practical requirement for any entrepreneur to achieve growth is business registration (as a corporation, partnership, or limited liability company). This practical requirement allows to form a population sample of entrepreneurs "at risk" of growth at a similar (and foundational) stage of the entrepreneurial process. Second, it is possible to distinguish among business registrants through the measurement of founding choices observable at or close to the time of registration. For example, Guzman and Stern measure start-up characteristics such as whether the founders name the firm after themselves (eponymy), whether the firm is organized in order to facilitate equity financing (e.g., registering as a corporation or in Delaware), or whether the firm seeks intellectual property protection (e.g., a patent or trademark). Third, though rare, some firms experience observable meaningful growth outcomes (such as achieving an IPO or high-value acquisition). Combining these insights, it is possible to estimate entrepreneurial quality by estimating the relationship between observed growth outcomes and start-up characteristics.

That is, for a firm *i* born in region *r* at time *t*, with at-birth start-up characteristics $K_{i,r,t}$, we observe growth outcome $g_{i,r,t+l}$ *l* years after founding and estimate:

$$\theta_{i,r,t} = P\left(g_{i,r,t+l} \middle| K_{i,r,t}\right) = f\left(\alpha + \beta K_{i,r,t}\right) \qquad (1)$$

and use the *predicted* value of this regression the measure of entrepreneurial quality.  As long as the process by which start-up characteristics map to growth remain stable over time (an assumption which is itself testable), this mapping allows to form an estimate of founding characteristics to entrepreneurial quality for any business registrant within any given sample.

$$\hat{\theta}_{i,r,t} = P\left(g_{i,r,t+l} \middle| K_{i,r,t}\right) = f\left(\alpha + \beta K_{i,r,t}\right) \qquad (2)$$

These estimates can be used to generate three core entrepreneurship statistics capturing the level of entrepreneurial quantity, and the potential for growth entrepreneurship within a given geographical area and start-up cohort.

*Entrepreneurial Quantity and Quality Indices (N and EQI).*  With a set of business registrants over a given time period within a given geographic region, it is straightforward to measure entrepreneurial quantity which is simply equal to the count of business registrants:

$$N_{r,t} = \#i \in \{I_{r,t}\} \qquad ] \qquad (3)$$

where $\{I_{r,t}\}$ represents the set of all new business registrations in ZTCA *r* and year *t.*

The predictive analytics approach described above is then combined with this population of business registrants to create an index of entrepreneurial quality for any group of firms (e.g., all the firms within a particular cohort or a group of firms satisfying a particular condition). Specifically, the Entrepreneurial Quality Index (EQI) is defined as an aggregate of quality at the ZTCA-year level by simply estimating the average of $\hat{\theta}_{i,r,t}$ over that ZTCA:

$$EQI_{r,t} = \frac{1}{N_{r,t}} \sum_{i \in \{I_{r,t}\}} \hat{\theta}_{i,r,t} \qquad (4)$$

To ensure that the estimate of entrepreneurial quality for a region *r* reflects the quality of start-ups in that location rather than simply assuming that start-ups from a given location are associated with a given level of quality, location-specific measures are excluded from the vector of observable start-up characteristics.

*The Regional Entrepreneurship Cohort Potential Index (RECPI).* From the perspective of a given geographical area, the overall inherent potential for a cohort of start-ups combines both the quality of entrepreneurship in a region and the number of firms in such region (a measure of quantity). *RECPI* is therefore defined as simply $EQI$ multiplied by the number of firms in that region-year:

$$\mathrm{RECPI}_{\mathrm{r,t}} = \mathrm{EQI}_{\mathrm{r,t}} \times \mathrm{N}_{\mathrm{r,t}} \qquad (5)$$

Since this index multiplies the average probability of a firm in a region-year to achieve growth (quality) by the number of firms, it is, by definition, the expected number of growth events from a region-year given the start-up characteristics of a cohort at birth. This measure of course abstracts away from the ability of a region to realize the performance of start-ups founded within a given cohort (i.e., its ecosystem performance), and instead can be interpreted as a measure of the "potential" of a region given the "intrinsic" quality of firms at birth, which can then be affected by the impact of the entrepreneurial ecosystem, or shocks to the economy and the cohort between the time of founding and a growth outcome. Together, this use of business registration data allows the construction of measures of both entrepreneurial quality and quantity-adjusted quality, giving us the ability to undertake a time-sensitive and granular evaluation of the interplay between research institutions and entrepreneurial activity.

## IV. Empirical Approach

The primary objective of our analysis is to leverage the SCP measures to examine the impact of public research resources on the quantity and quality-adjusted quantity of entrepreneurship. Our data are composed of zip-code level measures of the quantity and quality-adjusted entrepreneurship over time at a granular geographic level, these geographic units are co-located (or not) with a public research institution (university or national laboratory or both), and measures of the resources utilized by those public research institutions (including their overall budget, but also specific time-varying measures of resources such as the annual level of Federal research funding allocated to a given university). In an ideal case, we would be simply able to allocate public research resources randomly (e.g., to some locations but not others) and then measure the quantity and quality-adjusted quantity of entrepreneurship over time for those locations that had received public research resources versus those that had not. Indeed, the first stage of our empirical analysis is purely descriptive, and simply documents the differences in the quantity and quality of entrepreneurship for locations that are proximate or not to universities, with a focus on how this cross-sectional empirical relationship has changed over time.

However, to assess the causal impact of universities and national laboratories on their local entrepreneurial ecosystem, we must overcome three distinct but related empirical challenges. First, the co-location of public research institutions and vibrant entrepreneurial ecosystems is not random. While it is instructive in a descriptive sense to simply examine the cross-sectional correlation between the presence of public research institutions and entrepreneurship, it is possible that the presence of a positive correlation may be driven by potentially unobserved (to the econometrician) location-specific infrastructure or amenities. For example, the area surrounding Boulder, Colorado contains both the University of Colorado (which might influence that ecosystem) but also has specific private sector initiatives (such as the entrepreneurial accelerator TechStars) as well as other nearby institutions (e.g., the National Renewable Energy Laboratory is only 25 miles away). Both public research institutions and strong entrepreneurial ecosystems are likely elements of the "Creative Economy" (Florida, 2002), and the correlation between public research activity and entrepreneurship may be driven by selection rather than a causal link between the two.

Second, even if one controls for selection of public research institutions into favorable locations (e.g., through location-specific fixed effects), the activities of these public research institutions themselves may be endogenous to the activities of the local entrepreneurial ecosystem. Knowledge produced by firms within a local entrepreneurial ecosystem might simultaneously promote entrepreneurial activity within that ecosystem, and also exert spillovers onto university or national laboratory research itself. For example, the strategic shift by Monsanto (headquartered in St. Louis) into agricultural biotechnology research not only impacted that firm but also resulted in a shift towards agricultural biotechnology research at co-located research institutions such as Washington University of St. Louis (Sohn, 2020). More generally, a discovery or invention made by the entrepreneurs in a given location might facilitate both a higher level of research output at the university and at the same time serve as a spur for subsequent local entrepreneurship. The potentially bidirectional nature of impact between public research institutions and local entrepreneurial activity means that a simple examination of positive correlation over time of public research institutions research and local entrepreneurial activity cannot identify the causal linkage of public research institutions on entrepreneurship.

And, finally, entrepreneurial ecosystems are themselves dynamic. A positive shock to the quantity and quality of entrepreneurship in a given location (e.g., as the result of favorable economic conditions) may not only influence the current rate of entrepreneurial activity but may also exert a positive influence on the follow-on period. This would occur, for instance, if there was a potential for a two-period lag between the an initial unobserved shock and its influence on the rate of entrepreneurship (e.g., increasing the rate most saliently in the period immediately following the period but also having a smaller influence on entrepreneurship in the second period after the shock). Accounting for this dynamic would result in a lagged dependent variable structure. Specifically, if the lagged dependent variable is itself correlated with the changing level of local university research, abstracting away from the dynamic impact of entrepreneurial ecosystems themselves might result in overestimating the impact of university research itself.

To address these interrelated empirical challenges, we leverage the fact that the level of Federal research resources allocated to any given public research institution varies considerably over time. For example, between 1988 and 2012, the average coefficient of variation of Federal real research expenditures across 80 research universities is 1.86 (i.e., the standard deviation in research expenditures within a given university (measured in 2012 dollars) is nearly double its mean over the period). While the receipt of Federal research funding is of course not randomly allocated (indeed, it is dependent on the choices and grantsmanship of public researchers), our key assumption is that the *changes* over time in the level of overall Federal funding (research and otherwise) to a given university or national laboratory is exogenous to the existing quantity and quality of the local entrepreneurship. More precisely, the ability to attract a Federal grant to a particular institution indicates an increase in both the resources and the knowledge produced by that institution. As such, a consistent estimate of the impact of those additional resources on subsequent local entrepreneurship can be interpreted as the impact of the time-varying strength of that institution on the local environment.

To take advantage of this variation over time in the level of Federal resources allocated to universities and national laboratories, we first link our zip-code-level measures of the quantity and quality-adjusted quantity of entrepreneurship to the presence or absence of universities themselves. To do so, we define a zip-code $r$ in state $m$ to be proximate to a university or national laboratory if the center of that zip code is within five miles of the university or national laboratory (the next section details this procedure in greater detail). For each university or national laboratory, we observe a vector of inflation-adjusted annual Federal resource commitments, $X_{rt}$, which includes the level of research funding to a university, the level of non-research funding to the university, and the overall Federal budget allocation to a given national laboratory. In order to allow a meaningful gap of time to elapse between the allocation of Federal funding and its impact on the local entrepreneurial ecosystem, we group the annual zip-code-level data from the SCP, as well as measures of Federal resource allocation, into a smaller number of multi-year periods (either 2 years or 3 years). Denoted each of these multi-year periods by $s$, we can thus observe a multi-year measure of local entrepreneurial activity, $Y_{rs}$ (measured as either $N_{rs}$ or $RECPI_{rs}$), for each zip code, as well as measures of the level of resources provided by the Federal government to universities or national laboratories in that zip code, $X_{rs}$, and other time-varying location-specific characteristics, $Z_{rs}$.

We implement two empirical approaches to estimate the impact of changes over time in the Federal resources allocated to research institutions, $X_{rs-1}$, on the quantity and quality-adjusted rate of entrepreneurship in $s$. We first examine a simple fixed effects panel estimator:

$$Y_{rs} = \alpha_r + \beta_t + \delta X_{rs-1} + \gamma Z_{rs-1} + \varepsilon_{rs} \quad (6)$$

This simple specification allows us to estimate the relationship between entrepreneurship and the lagged level of resources, while controlling for fixed differences between locations and year in terms of the rate of entrepreneurship, and controlling for other observable location-specific measures. This specification goes beyond a simple cross-sectional correlation between the colocation of research institutions and entrepreneurship by focusing exclusively on variation in resources that are structurally independent from local entrepreneurship: changes in the level of Federal resources are the result of grant writing rather than the result of a flow from entrepreneurship towards the university. However, this specification does not address the potential for dynamics within an entrepreneurial ecosystem insofar as it is possible that the rate of entrepreneurship is not only impacted by the level of resources from the university but also the existing level of entrepreneurship itself. In principle, we can account for this possibility by including the lagged level of dependent variable:

$$Y_{rs} = \alpha_r + \beta_t + \lambda Y_{rs-1} + \delta X_{rs-1} + \gamma Z_{rs-1} + \varepsilon_{rs} \quad (7)$$

Though an appealing method for accounting for dynamics, the assumptions required for consistent estimation of (7) using a simple fixed effects panel estimator are demanding (Angrist and Pischke, 2008, p. 243): (a) the elements of *X* and *Z* must themselves by serially independent (i.e., conditional on $\alpha_r$, there can be no correlation over time in the *X* and *Z* themselves), and (b) the elements of $\varepsilon_r$ must be independent of each other (i.e., the shocks are not only independent of *X* and Z, but there are also no dynamics to the shocks over time), and. Specifically, in the case where there is positive correlation across sequential time periods (e.g., an AR(1) process)), then (7) is likely to result in an overestimate of (the absolute value of) δ (our parameter of interest).

We can significantly relax the assumptions required for identification of δ by transforming (7) through first-differences:

$$\Delta Y_{rs} = \Delta\beta_t + \lambda \Delta Y_{rs-1} + \delta \Delta X_{rs-1} + \gamma \Delta Z_{rs-1} + \eta_{rs} \quad (8)$$

Specifically, this specification allows us to assess how *changes over time* in the resources allocated to a research institution influences subsequent *changes* over time in the quantity or quality-adjusted quantity of entrepreneurship. And, since we are focusing only on changes in the level of $X$ and $Z$, this specification allows for correlation over time in the level of $X$ and $Z$. However, both $\Delta Y_{rs-1}$ and $\eta_{rs}$ contain $\varepsilon_{rs-1}$ and so there is a mechanical (negative) correlation between $\Delta Y_{rs-1}$ and $\eta_{rs}$; as such, the estimates from an OLS implementation are biased (Arellano and Bond, 1991; Nickell, 1981). To address this, we can use the level of $Y_{rs-2}$ as an instrument for $\Delta Y_{rs-1}$ to implement a consistent estimator for (3) (Anderson and Hsiao, 1981).[4] As highlighted by Angrist and Pischke (2008), this approach may lead to an underestimate of the (absolute value of) δ due to attenuation bias.

Putting these ideas together, the key empirical challenges of estimating the impact of universities on local entrepreneurial ecosystems arises from the fact that the colocation of universities and start-ups is neither random nor unidirectional, and local entrepreneurial ecosystems are themselves dynamic. To address these concerns, we take advantage of the fact that Federal resources towards local research institutions are essentially independent of start-ups in that region (conditional on the pre-existing level of entrepreneurship), and we will take advantage of this insight by considering the impact of changes over time in Federal resources to local research institutions. To account for different potential biases, our analysis compares both fixed effects and panel data estimators to assess the robustness of our core results to different sets of assumptions required for consistent estimation of this fixed effects dynamic panel data model.

## V. Data

Implementing our approach requires the combination of systematic measures of quantity and quality-adjusted quantity of entrepreneurship at a granular geographic level with measures of universities and national laboratories activities over time, as well as locational characteristics (e.g., urban versus rural, income, etc). We therefore combine data at the zip code level from the Startup Cartography Project, the National Center for Education Statistics, the NSF Survey of

[4] Accounting for the interplay between dynamics over time within a panel and the presence of fixed differences is complex and involves tradeoffs (Angrist and Pischke, 2009, Section 5.3). To simplify exposition and focus on estimates that most directly connect to our research question, we focus on the fixed effects panel estimator and OLS and instrumental variables first-differences specifications. We have also explored using an Arellano-Bond GMM estimator (which allow for more efficiency by using all the lags of the levels of dependent and independent variables as instruments for the changes in those variables over time; however, we found that the estimates varied considerably depending on the precise specifications (and in ways that were not easily interpretable except perhaps reflecting the imprecision arising from a "weak instruments" problem).

R&D Expenditures, Census socioeconomic data, as well as auxiliary sources concerning research university quality and type.

*A. The Startup Cartography Project*[5]

The first building block of our dataset consists of systematic and granular measures of the quantity and quality-adjusted quantity of entrepreneurship drawn from the Startup Cartography Project (Andrews et al., 2020, SCP hereafter). Building on Guzman and Stern (2015, 2017, 2020), these measures are calculated by aggregating, to the zip code level, state-level business registration data with the predictive analytics approach discussed in Section 2 allowing for an estimate of entrepreneurial quality for all new registrants at founding. The data are drawn from 32 states (corresponding to more than 81% of the national GDP) from 1988 to 2012, and include, for any year, all new business registrants that are either a for-profit firm in the local jurisdiction, or a for-profit firm registered in Delaware but whose principal office address is in the focal state.[6]

For each business registrant, the SCP constructs variables related to: (i) growth outcomes (IPO or significant acquisition); (ii) firm characteristics based on business registration observables; (iii) firm characteristics based on secondary data that can be directly linked to the firms itself, such as patents and trademarks.

*i.* *Growth Outcomes*. The growth outcome, *Growth*, is a dummy variable equal to 1 if the firm has an initial public offering (IPO) or is acquired at a meaningful positive valuation within 6 years of registration as reported in the Thomson Reuters SDC database.[7]

*ii.* *Firm Characteristics Based on Business Registration Data.* Two binary measures relate to how the firm is registered: *Corporation*, which captures whether the firm is a corporation rather than an LLC or partnership, and *Delaware*, equal to one if the firm is registered in Delaware. Five additional measures are furthermore based directly on

[5] This section builds on Andrews, et al (2020), which itself builds on Guzman and Stern (2015, 2017, 2020), which introduces the details regarding state-level business registration records and the predictive analytics approach used to measure the quantity and quality of entrepreneurship.

[6] While the updated version of SCP (Andrews, et al, 2020) contains data for 48 states, we focus on the 32 states for which data were available in 2019 (in line with Guzman and Stern (2020)).

[7] Although the coverage of IPOs is essentially comprehensive, the SDC data set excludes some acquisitions. Guzman, et al (2020) and Andrews, et al (2020) discuss robustness to alternative datasets and measures of growth outcomes.

the name of the firm. *Eponymous* is equal to 1 if the first, middle, or last name of the top managers is part of the name of the firm itself, which may be associated with lifestyle businesses rather than growth-oriented entrepreneurs (Belenzon et al., 2020, 2017), and *Short Name* is equal to one if the entire firm name has three or less words, and zero otherwise (most growth-oriented ventures have names of three words or fewer). Seven measures are based on how the firm name reflects the industry or sector within which the firm is operating, taking advantage of the industry categorization of the US Cluster Mapping Project ("US CMP") (Delgado et al., 2016) and a text analysis approach. The first three are associated with broad industry sectors and include whether a firm can be identified as local (*Local*), traded (*Traded*) or resource intensive (*Resource Intensive*). The other five industry groups are narrowly defined high technology sectors that are typically associated with high growth firms, including whether the firm is within the biotech (*Biotech Sector*), e-commerce (*E-Commerce*), other information technology (*IT*), medical devices (*Medical Devices*) or semiconductors (*Semiconductor*) space.

iii. *Firm Characteristic Based on External Observables.* Two measures are related to intellectual property measures based on data from the U.S. Patent and Trademark Office. *Patent* is equal to 1 if a firm is associated with a patent application within the first year and 0 otherwise, including patents filed by the firm within the first year of registration and patents that are assigned to the firm within the first year from another entity (e.g., an inventor or another firm). The second measure, *Trademark*, is equal to 1 if a firm applies for trademark protection within a year from registration.

As reported in Appendix A, these data are used to estimate a logit regression model that allows one to examine how the presence or absence of a startup characteristic correlates with the probability of growth (conditioning on the presence or absence of other startup characteristics). There is an extremely strong (and robust) correlation between startup characteristics and the probability of growth. Substantial changes in the predicted likelihood of a growth outcome are associated with characteristics observable in real time from business registration records as well as characteristics observable with a lag (e.g., patent and trademark applications). On the one hand, startups founded as corporations are almost 220% more likely to grow. Similarly, firms with short names are 79% *more* likely to grow, while eponymous firms are 69% *less* likely to grow. Finally, the interplay between corporate form and formal intellectual property protection is particularly predictive: a startup that registers in Delaware and applies for a patent in its first year is 83 times more likely to grow than a firm that only registers in its home state and does not apply for intellectual property protection.[8] For each firm, these estimates can be used to calculate the probability of growth at founding. While this measure is low on average (on the order of one in 1400), firms with particular combinations of startup characteristics are far more likely to grow (e.g., firms in thet top 1 percent of the out-of-sample estimated quality distribution have a better than 1 in 100 chance of achieving a growth outcome).

These firm-level estimates can then be aggregated at an arbitrary level of geographic and temporal granularity to form annual measures of entrepreneurial quantity and quality-adjusted quality. To capture the local relationship of university and national laboratory activity on local entrepreneurship, we focus our geographic analysis at the ZCTA level.[9] Specifically, for each year and ZCTA, we construct two core measures (see Table 1 for means and standard deviations). # of Ventures is simply equal to the number of newly formed ventures in that ZCTA (mean = 41.8), while RECPI is # of Ventures multiplied by the average estimated probability of growth (EQI) within that cohort multiplied by 10,000 (for interpretability). The average probability of a growth event of the startups within a ZTCA-year cohort is 0.029.[10]

----------------------------------------

*Insert Table 1 about here*

----------------------------------------

[8] It is important to emphasize that these startup characteristics are not the *causal* drivers of growth, but instead are "digital signatures" that allow to distinguish firms in terms of their entrepreneurial quality. Registering in Delaware or filing for a patent will not guarantee a growth outcome for a new business, but the firms that historically have engaged in those activities have been associated with skewed growth outcomes

[9] Zip Code Tabulation Areas (ZCTA) are generalized area representations of USPS Zip Code service areas. While standard zip codes represent geographic areas with well-defined geographical boundaries, zip code boundaries are potentially changed over time, and also there are zip codes for PO Boxes, military, and large customers. To fix zip code boundaries over time, we rely on the 2015 Zip Code to ZCTA crosswalk by the HRSA UDS mapper (https://www.udsmapper.org), which contains 41251 unique ZIP Codes, the correspondent ZCTAs, the type of ZIP Code, city and State. All but two ZIP codes in the Startup Carrtography Project dataset were matched using the cross walk (these two were then assigned to a ZCTA manually).

[10] We then group these annual measures into two-year or three-year aggregates when we conduct fixed effects panel effects and first-differences estimation of the causal impact of changes in Federal resources towards research institutions on the subsequent changes in N and RECPI.

*B. University and National Laboratory Measures*

The second component of the data include measures related to universities and national laboratories. We first develop a list and location for all higher education institutions from the National Center for Education Statistics. From this list, we define a Research University those universities rated as either "high research activity" or "very high research activity" (the two top categories of research orientation) according to the standard Carnegie Classification system (Shulman, 2001). For each research university *j* and year *t*, we measure Federal R&D $funds_{jt}$ and Federal Non-R&D $Funds_{jt}$ from the NSF Survey of Federal Funds for Research and Development (in current dollars). Moreover, for each research university *j* and year *t* we measure funds received from other source than federal: Institutional $funds_{jt}$, State $funds_{jt}$, Business $funds_{jt}$ and Other $funds_{jt}$. As well, for each research university, we include several dummies for university characteristics. We define a research university as being in the Top 10 or Top 50 based on the 2017 Times Higher Education rankings (while there is year-to-year variation in the ranking the groupings of universities into the Top 10 or Top 50 are quite stable over time). As well, we collected information from Wikipedia about each university, denoting whether the university has specialized professional schools, including a Business School, Medical School, and/or Law School (each coded as a separate dummy variable).

We then use the NSF NCSES Survey of R&D Expenditures at Federally Funded R&D Centers to construct a list and location data for all national laboratories that receive Federal funding in any year. For each national laboratory *k* in year *t*, we use the NSF NCSES Survey of R&D Expenditures at Federally Funded R&D Centers to measure the Federal R&D $Funds_{kt}$.

With this set of geocoded universities and national laboratories, we then create the core measures used in our analysis by aggregating these institution-level measures to the ZCTA level using these annual data. Specifically, for each ZCTA, we create a circle with a 5 mile radius centered at the ZCTA centroid; we then aggregate the number, nature and funding of research universities that fall within the circle for each ZCTA. There is a research university proximate to 6.4% of ZTCA, but only 0.6% of ZTCA have a proximate national laboratory. The incidence rate of professional schools is between 0.05 and 0.076, in part because there exist many professional schools that are independent of research universities (and are not themselves research-oriented) and also because professional schools often have separate campuses, often in urban locations. The average level of research university funding proximate to a ZCTA is $10.5 million (with not surprisingly a very wide variance), while the average level of Non-R&D Funds is only about 11% of this level, while R&D Funds to National Laboratories registers an average of $3.3 million.

### *C. ZCTA Characteristics*

While the centerpiece of our analysis focuses on changes over time within a ZCTA, we also report cross-sectional estimates of the university and national laboratory "premium" in terms of the level of entrepreneurial quantity and RECPI (and examine how this premium has changed over time). We include a set of ZCTA characteristics, drawn from the 1990 Census, in that analysis to control for the presence of locational characteristics that might also explain the premium. These include population, density, male percentage, urban, fraction aged between 18 and 65, white, born outside the U.S., percentage of population holding a college degree, private versus public sector employment, income per sector, and housing value (See Appendix Table B.1 for definitions, means and standard deviations). When undertaking the matching between ZCTA and the 1990 Census, we fail to match 14% of ZCTAs, and so exclude these zip codes throughout the analysis (all of the panel and long-differences results are robust to their inclusion or exclusion). Overall, the final sample consist of 15950 ZCTAs in the time period 1988-2012.

## VI. Results

We now proceed to assess the relationship between research institutions and regional entrepreneurship. Our analysis proceeds in three steps. We first consider how the quantity and quality-adjusted level of entrepreneurship varies in the cross-section with the presence of research institution, and how this relationship varies by the type of university, geographic characteristics, and over time. A key insight from this analysis is that while the positive correlation between the presence of a university and entrepreneurial quality was negligible as of the late 1980s, there was a striking increase over the course of our sample in the importance of colocation with research universities for high-quality ventures. These cross-sectional observations motivate our core empirical analysis, where we consider the impact of changes over time in research and non-research funding to universities and national labs on both the quantity and quality-adjusted quantity of entrepreneurship. Across a broad range of approaches, we document a sizeable positive relationship between shifts in research funding to universities and subsequent entrepreneurial quality; however, we find no impact of shifts in non-research funding to universities or overall funding changes to national laboratories. We conclude the analysis by illustrating the concrete consequences of the estimates for policy; for example, taken at face value, our estimates suggest that a doubling of expenditures on university research would enhance the overall quality-adjusted quantity of entrepreneurship by 23% on an ongoing basis.

*A. The Cross-Sectional Relationship Between Research Institutions and Entrepreneurship*

We begin in Table 2 where we decompose Number of Ventures and RECPI by the presence or absence of at least one research university or national laboratory. Overall, there is a striking correlation between the presence of a research institution and entrepreneurial activity. The number of ventures more than doubles in the presence of a national laboratory and more than triples for those zip codes within a five-mile radius of a research university. Even more strikingly, RECPI shows a 350% increase in locations near national labs and more than a 700% boost in in locations near research universities. Each of these conditional means is statistically significantly different than the figures for locations that are not co-located with at least one research institution.

----------------------------------------
*Insert Table 2 about here*
----------------------------------------

Of course, the geography of universities or national laboratories is not random. Even when performing a cross-sectional analysis, it is important to account for locational differences that may also be correlated with the level and quality of entrepreneurship. More subtly, many national laboratories are themselves co-located with research universities, and so the presence of a national laboratory may in fact reflect the simultaneous presence of a research university rather than an independent correlation between national laboratories and entrepreneurship.

Table 3 explores this possibility by examining the relationship between entrepreneurship and the presence of research universities and national laboratories at the same time, both excluding and including a rich set of locational characteristics (population, density, income, etc), and also a set of year dummies. Three relationships stand out. First, regardless of the inclusion or exclusion of other measures, there is a quantitatively large and statistically significant relationship between both Number of Ventures and RECPI in a Zip code and the presence of a nearby research university. For example, in (3-4), where we control for the presence of national laboratories, as well as locational characteristics, the presence of a research university is associated with an increase of 351 in RECPI, corresponding to 40% of the standard deviation of that measure. Second, and in sharp contrast, the positive pairwise correlation between national laboratories and entrepreneurship is nullified out after controlling for the simultaneous presence of a research university, and there is actually a negative relationship between the presence of a national laboratory and Number of Ventures after the inclusion of locational characteristics. Finally, it is useful to note that the inclusion of locational controls have a significant impact on the relationship between research institutions and entrepreneurship; for example, the relationship between the existence of a research university and RECPI is nearly halved after the inclusion of these controls. This is consistent with the idea that the areas where universities are located are indeed different from the areas without a research institution nearby, and they present features that are correlated with increased entrepreneurial activity (Florida, 2005).

---------------------------------------

*Insert Table 3 about here*

----------------------------------------

We emphasize that these relationships are not causal, but are simply intended to highlight the cross-sectional correlation (or lack thereof) between research institutions and both the quantity and quality-adjusted quantity of entrepreneurship. Figure 1 extends this logic by examining how the correlation between research institutions and RECPI have changed over time. To do so, we plot the regression coefficients of the interaction effects between research university and national laboratory, respectively, with a full set of year dummies (controlling for locational characteristics). There are two sharp insights from this figure. First, while the relationship between the presence of a research university and RECPI in the late 1980s was limited, there was a steady increase in the premium associated with university colocation through the 1990s, which has persisted through the end of our observation period. In other words, the now-common linkage between research universities and growth-oriented entrepreneurship is not an historical constant but a phenomenon that emerged during the last thirty years. Put another way, the incidence for the colocation of universities and locational clusters of innovation and entrepreneurship, highlighted first in areas such as biotechnology (Feldman, 2000; Zucker et al., 1998), seems to have increased over time. Second, and in sharp contrast, the overall relationship between RECPI and national laboratories is very noisy; there is no discernible robust relationship between these dedicated research facilities and entrepreneurial quality.

---------------------------------------

*Insert Figure 1 about here*

---------------------------------------

We complete this descriptive analysis in Table 4 by further exploring the interplay between university, locational characteristics and entrepreneurship. Each of these specifications includes a set of year fixed effects, a full set of locational controls, and are clustered at the zip code level. We first examine, in (4-1) and (4-4) whether "elite" universities (those in the top 10 or top 50) are linked to Number of Ventures or RECPI. Interestingly, there is no significant relationship between university quality and the quantity of entrepreneurs but there is a significant impact in terms of RECPI. Top 10 universities are associated with more than a five-fold increase in RECPI, reflecting the importance of entrepreneurial clusters surrounding Stanford, MIT, and others. When we examine the role of different professional schools ((4-2) and ((4-5)), the only significant relationship is between the presence of a law school and RECPI (perhaps reflecting a certain degree of professionalization of entrepreneurship in those local areas, highlighting the need for the changes over time analysis we conduct below). Finally, in (4-3) and (4-7), we see that even after controlling for locational characteristics, high-income zip codes near universities are associated with an increase in the Number of Ventures, and both high-income and urban zip codes near universities are associated with a higher level of RECPI. Taken together, these descriptive results highlight the positive correlation that has emerged over the past thirty years between the presence of a research university and growth-oriented entrepreneurship, the importance of other locational attributes attenuating (but in no way eliminating) this correlation, and, finally, the lack of a robust relationship between the presence of a national laboratory and either the quantity or quality-adjusted quantity of entrepreneurship.

---------------------------------------

*Insert Table 4 about here*

---------------------------------------

*B. The Impact of Changes in Federal Resources Towards Research Institutions on Local Entrepreneurial Ecosystems*

We now turn to the centerpiece of our analysis where we move beyond cross-sectional correlations and turn to how changes in public funding to research institutions impact the subsequent quantity and quality of entrepreneurship. As highlighted in our descriptive findings, the key challenge of identifying the impact of research institutions on local entrepreneurship is that the location of universities and national laboratories is not random; research universities and national laboratories are associated with other locational characteristics (including the presence of vibrant entrepreneurial clusters) that themselves encourage entrepreneurship. Accordingly, we focus on leveraging variation in the level and nature of activities in research institutions over time that is independent of those locational characteristics in order to identify the causal impact of research institutions on entrepreneurship.

The specifications we implement follow the logic laid out in Section IV. We first group each of the annual measures into two-year or three-year aggregates (depending on the specification)/ We then focus on the impact of changes in Federal resources to research institutions on first the quantity and then quality-adjusted quantity of entrepreneurship.

---------------------------------------
*Insert Table 5 about here*
----------------------------------------

Our analysis begins in Table 5 with a focus on the overall quantity of entrepreneurship. We begin in (5-1) with a panel fixed effects specification, where we examine the impact of funding on the quantity of entrepreneurship (Number of Ventures), controlling for year and ZTCA fixed effects (with standard errors clustered at the zip code level). There is a large and statistically significant impact of both R&D and non-R&D funding to universities, with a much lower impact for funding for national laboratories. Indeed, the quantitative estimate in (5-1) seems implausibly large: the estimates suggests that a $1,000 increase in Federal funds to a university (either research-oriented or not) is associated with ~0.3 new ventures in the subsequent time period in zip codes co-located with that university. Of course this specification abstracts away from the potential dynamics within locations, where there might be contemporaneous increases in the level of the entrepreneurial ecosystem and the ability of universities within that ecosystem to attract Federal resources. As such, we then include the lagged level of entrepreneurship (Number of $\text{Ventures}_{s-1}$) in (5-2) (once again with ZTCA and year fixed effects). The coefficients on resources towards universities remain positive and statistically significant but are much smaller (though still sizeable) in magnitude. For R&D funds to universities, an increase of $70,000 is associated with a subsequent increase of one business registrant in zip codes co-located to that university, and the estimated impact for non-R&D funds to universities is estimated to be nearly three times as large. This evidence is consistent with the idea that expenditures by universities serve as a source of local demand (Andrews, 2018), and so an increase in university expenditures might encourage a higher quantity of entrepreneurship (without having an impact on quality).

Given the inclusion of ZTCA fixed effects and the inclusion of lagged values of Number of Ventures, a causal interpretation of the coefficients on Federal resources in (5-2) depends on assuming that changes in the levels of Federal resources over time within a location are independent of each other. If there are trends over time in the level of Federal resources (and the impact of those resources extend over time, then the (absolute value) of the coefficient on Federal resources in (5-2) will be positively biased, leading to an overestimate of the impact of Federal resources on the quantity of entrepreneurship.

Following the logic of Section IV, we therefore turn in the remaining columns of Table 5 to a first-differences specification, including both an OLS and instrumental variables implementation (Anderson and Hsiao, 1981). The use of first-differences allows us to exclude the ZTCA fixed effects. However, to account for differences in the trend in entrepreneurship across regions, we include county-level fixed effects, and to account for changing regional economic conditions, we include state-year fixed effects (all of our broad findings are robust to the inclusion or exclusion of these fixed effects). Before turning to the results of interest, it is useful to note that the coefficient on the change in Number of $\text{Ventures}_{s-1}$ (i.e., the lagged first-difference) is positive, and is of similar magnitude in the case of both the OLS and Anderson-Hsiao IV estimator. Moreover, the results provide support for the hypothesis that resources directed at local research institutions encourage a higher level of entrepreneurship. While the coefficient on non-R&D Federal research funding is somewhat reduced (but remain quantitatively and statistically significant for two-year first-differences), the coefficients on R&D funding to universities and overall Federal funding of national laboratories actually increase relative to (5-2). According to these estimates, a $100,000 "shock" to university resources in a two-year period is associated with between 4 and 8 additional business registrants in adjacent zip codes in the subsequent two-year period (accounting for the overall trend in entrepreneurship in that county and the overall shift in entrepreneurship in that state-year). It is also useful to note that the impact of resources seems to be relatively rapid: the magnitudes of the effect are somewhat smaller for the case where we group the data into three-year time periods (and the coefficient on non-R&D resources to universities becomes insignificant).

---------------------------------------

*Insert Table 6 about here*

---------------------------------------

While the analysis in Table 5 focuses on the impact of resources on the quantity of entrepreneurship, Table 6 focuses attention on potentially a more critical measure, RECPI, which directly accounts for the growth potential of start-ups at the time of their founding. Table 6 follows the same empirical logic as Table 5. The first two specifications (6-1 and 6-2) present the fixed effects panel estimator (excluding or including a lagged dependent variable), the second two specifications (6-3 and 6-4) examine a first-differences specification (implemented with OLS and an Anderson-Hsiao instrumental variables estimator), and (6-5) and (6-6) present the first-differences specification with the data grouped into three-year rather than two-year intervals.

The results are striking and robust. First, across both the fixed effects and first-differences specifications (both OLS and IV), there is a positive and quantitatively significant impact of Federal research funds for universities on the subsequent change in *RECPI*, the quality-adjusted quality of entrepreneurship. After controlling for $RECPI_{s-1}$, the coefficient on *Federal Research Funds* varies from 0.81 to 1.26. From a quantitative perspective (and taking 0.81 as the baseline estimate), this implies that an increase of $1 million of Federal R&D resources (approximately the size of an extra 2-3 additional average-sized NIH grants) is associated with a shift of 810 in *RECPI*: this can be interpreted as increasing the potential of the birth of a single growth firm by 0.08 within that zip code over that two-year time period. Put another way, at face value, the *Federal Research Funds* coefficient implies that a positive shock of $12.3 million is linked to an increase of 10,000 in *RECPI,* or the equivalent of one successful growth outcome for that entrepreneurial ecosystem.[11] This estimate is particularly striking given that the median returns to a single successful growth outcome (either an IPO or successful acquisition > $10M within six years of founding) are approximately $80 million of private value to the entrepreneur and investors (and there is likely at least an equally likely positive impact in terms of employment growth and follow-on regional knowledge and supply chain spillovers).[12]

In contrast to the positive impact of Federal funds for research, the estimated impact of *Federal Non-Research Funds for Universities* is negative across all specifications. Moreover, the absolute value of these coefficients are from 2 – 6 times as large (ranging from -2.19 in (6-2) to -7.34 in (6-3)). Though a large estimated impact, it is useful to note that both the mean and standard deviation of this measure is about 20% as large as *Federal Research Funds for Universities*. While the results from Table 5 identify the impact of non-R&D funding on the quantity of entrepreneurship, these results suggest that non-R&D funding actually has a negative net impact on the average quality of entrepreneurship. Finally, there is a negative and much smaller in magnitude estimated coefficient across all specifications for *R&D Funds to National Laboratories.* Despite being focused on research expenditures, this small and negative finding suggests that the cloistered nature of national laboratories (less open to local knowledge spillovers, less involved in teaching) results in a much more limited (or even negative) impact of these institutions on their local entrepreneurial ecosystems.

---------------------------------------
*Insert Table 7 about here*
---------------------------------------

In Table 7 we examine the interplay between federal research funding to universities, locational and university characteristics and entrepreneurship. The interactions terms are added to the baseline models corresponding to (5-4) and (6-4), and include year and county fixed effects. When we interact federal research funding to a variable indicating if there is a high-

[11] Recall that *RECPI* is the probability of growth at the time of founding as measured using the predictive analytic model from the SCP (Andrews, et al, 2020), multiplied by 10,000. Given that most start-ups fail, a shift in *RECPI* of 10,000 implies the founding of multiple ventures, with the sum of *RECPI* across those ventures equal to 10,000.

[12] It is useful to recall that medium-sized universities often experience significant variation in their receipt of Federal research funds. To give but one concrete example, consider Weber State University in Ogden, Utah, where a 249% increase in research funding between 1997 and 1998 was accompanied by a 40% increase in RECPI (including a 17% increase in # of Ventures) in zip codes less than five miles from campus. Interestingly, this burst in entrepreneurial activity was not experienced in more distant zip codes in the state (e.g., zip codes between 5 to 50 miles (e.g., Salt Lake City). See Figure D for calculations.

ranking university near the ZIP code we find a negative effect both on the # of ventures and RECPI. We find a similar negative effect on both dependent variables interacting funding with the number of research universities. It seems that the increases in federal research funding to universities have a greater effect in zip codes where lower-tier universities are located, and that funding is more beneficial in increasing entrepreneurial quality if it is concentrated, rather than spread across a number of different universities. Finally, urban zip codes seem to benefit less from increases in federal research funding in term of # of ventures: these are likely areas where it is more costly to set up a business, and therefore even as demand increases, there may be fewer entrepreneurs able to start their business in the area.

---------------------------------------
*Insert Table 8 about here*
---------------------------------------

Finally, in Table 8 we control for the possible confounding effect of other funding sources to universities. (8-1) replicates the baseline model corresponding to (5-3), and includes year, county and state-year fixed effects. Only State funds to universities have a significant effect of # of ventures, while the effect of R&D federal funds remains the same. (8-2) replicates the baseline model corresponding to (6-3), and again includes year, county and state-year fixed effects. In this model, Business funds to universities play a very significant and positive role in increasing RECPI, but again the effect of R&D federal funds is very similar to the one in the main specification, confirming the robustness of our results.

Overall, these results provide evidence about the relative salience of alternative channels of impact of research institutions on entrepreneurship. Specifically, the framework in Section II highlights that one channel by which research institutions might impact entrepreneurship is through their role as a source of local demand (Andrews, 2020). And, that would have the consequences that increased economic activity by research institutions might encourage an increase in the quantity of entrepreneurship; however, an increase in non-R&D funds to universities nor research funds to national laboratories is associated with an increase in the quantity of entrepreneurship. However, a second channel by which research institutions might impact entrepreneurship is through the generation of knowledge, and that channel is likely to be salient for research institutions where it is possible for potential entrepreneurs to carry that knowledge outside the boundary of the institution (or, similarly, if the boundaries of the institution are porous and open enough to support local knowledge spillovers). The broad pattern of findings in Tables 5-8 are consistent with the idea that research funding to universities plays a special role in shaping local entrepreneurship by inducing the creation of ventures with high potential for growth (and local economic impact).

## VII. Conclusions

In this paper we provide systematic empirical evidence for the unique role that universities play in shaping local entrepreneurial ecosystems, and in so doing allow for a reconceptualization of the underlying theoretical and policy drivers of the impact of universities on entrepreneurship. We present three main sets of results. First, universities are associated with not only a higher level of entrepreneurship but in particular a higher level of quality-adjusted entrepreneurship, and this relationship has strengthened over time. Second, relative to the direct impact of universities, demographic and economic factors associated with the presence of a university (income levels, diversity) are even more strongly associated with the presence of a strong entrepreneurial ecosystem. Finally, even after controlling for such factors, changes over time in resources enhance entrepreneurship but only increases in research-oriented funding enhance entrepreneurial quality. Together, these findings suggest not only that universities as large economic institutions have a critical (and often underappreciated) role in local economic development, but also that the specific norms and governance of universities play a unique role in promoting growth entrepreneurship conducive to long-term economic growth.

While we believe our works shows a clear link between university researcher activities and local high-growth entrepreneurship, future work would benefit from a careful micro-level analysis of exactly how these knowledge spillovers influence local ventures. As universities are increasingly asked to contribute to actively economic growth through commercialization activities, it would be interesting to understand the extent that faculty-generated academic spin-offs contribute to the high-growth entrepreneurial activities in the vicinity of universities. Moreover, future work could focus on potential industry-related contingencies of this link between universities and entrepreneurial activities. In particular, understanding how firms in certain industries may benefit from scientific specialization of universities activities would help shed further light on the role of universities in stimulating economic activity.

## References


Acs, Z.J., Audretsch, D.B., Feldman, M.P., 1992. Real Effects of Academic Research: Comment. Am. Econ. Rev. 82, 363–367.

Adams, J.D., 2002. Comparative localization of academic and industrial spillovers. J. Econ. Geogr. 2, 253–278. https://doi.org/10.1093/jeg/2.3.253

Adams, J.D., 1990. Fundamental stocks of knowledge and productivity growth. J. Polit. Econ. 98, 673–702.

Adams, J.D., Chiang, E.P., Jensen, J.L., 2003. The influence of federal laboratory R&D on industrial research. Rev. Econ. Stat. 85, 1003–1020.

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. J. Am. Stat. Assoc. 76, 598–606.

Andersson, R., Quigley, J.M., Wilhelmson, M., 2004. University decentralization as regional policy: The Swedish experiment. J. Econ. Geogr. 4, 371–388. https://doi.org/10.1093/jnlecg/lbh031

Andersson, R., Quigley, J.M., Wilhelmsson, M., 2009. Urbanization, productivity, and innovation: Evidence from investment in higher education. J. Urban Econ. 66, 2–15.

Andrews, M., 2020. How do institutions of higher education affect local invention? Evidence from the establishment of US colleges.

Andrews, R.J., Fazio, C., Guzman, J., Liu, Y., Stern, S., 2020. The Startup Cartography Project. Working Paper.

Angrist, J.D., Pischke, J.-S., 2008. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58, 277–297.

Arrow, K.J., 1972. Economic welfare and the allocation of resources for invention, in: Readings in Industrial Economics. Springer, pp. 219–236.

Audretsch, D.B., Lehmann, E.E., 2005. Do University policies make a difference? Res. Policy 34, 343–347. https://doi.org/10.1016/j.respol.2005.01.006

Audretsch, D.B., Lehmann, E.E., Warning, S., 2005. University spillovers and new firm location. Res. Policy 34, 1113–1122. https://doi.org/10.1016/j.respol.2005.05.009

Audretsch, D.B., Stephan, P.E., 1996. Company-scientist locational links: The case of biotechnology. Am. Econ. Rev. 86, 641–652.

Audretsch, D.B., Thurik, A.R., 2001. What's New about the New Economy? Sources of Growth in the Managed and Entrepreneurial Economies. Ind. Corp. Chang. 10, 267–315. https://doi.org/10.1093/icc/10.1.267

Baptista, R., Lima, F., Mendonça, J., 2011. Establishment of higher education institutions and new firm entry. Res. Policy 40, 751–760. https://doi.org/10.1016/j.respol.2011.02.006

Baptista, R., Mendonça, J., 2010. Proximity to knowledge sources and the location of knowledge-based start-ups. Ann. Reg. Sci. 45, 5–29. https://doi.org/10.1007/s00168-009-0289-4

Belenzon, S., Chatterji, A.K., Daley, B., 2020. Choosing between growth and glory. Manage. Sci. 66, 2050–2074.

Belenzon, S., Chatterji, A.K., Daley, B., 2017. Eponymous entrepreneurs. Am. Econ. Rev. 107, 1638–1655.

Bonaccorsi, A., Colombo, M.G., Guerini, M., Rossi-Lamastra, C., 2014. The impact of local and external university knowledge on the creation of knowledge-intensive firms: Evidence from the Italian case. Small Bus. Econ. 43, 261–287. https://doi.org/10.1007/s11187-013-9536-2

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Res. Policy 23, 487–521.

Delgado, M., Porter, M.E., Stern, S., 2016. Defining clusters of related industries. J. Econ.

Geogr. 16, 1–38.

Feldman, M.P., 2000. Location and innovation: the new economic geography of innovation, spillovers, and agglomeration, in: The Oxford Handbook of Economic Geography.

Feldman, M.P., 1994. The geography of innovation. Springer Science & Business Media.

Florida, R., 2002. Bohemia and economic geography. J. Econ. Geogr. 2, 55–71.

Florida, R.L., 2005. Cities and the creative class. Psychology Press.

Furman, J.L., MacGarvie, M.J., 2007. Academic science and the birth of industrial research laboratories in the US pharmaceutical industry. J. Econ. Behav. Organ. 63, 756–776.

Furman, J.L., Stern, S., 2011. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. Am. Econ. Rev. 101, 1933–1963.

Guzman, J., Stern, S., 2017. Nowcasting and Placecasting Entrepreneurial Quality and Performance, in: Haltiwanger, J., Hurst, E., Miranda, J., Schoar, A. (Eds.), Measuring Entrepreneurial Businesses: Current Knowledge and Challenges.

Guzman, J., Stern, S., 2015. Innovation economics. Where is Silicon Valley? Science (80-. ). 347, 606–9. https://doi.org/10.1126/science.aaa0201

Hall, B.H., Link, A.N., Scott, J.T., 2003. Universities as Research Partners. Rev. Econ. Stat. 85, 485–491. https://doi.org/10.1162/rest.2003.85.2.485

Hausman, N., 2012. University Innovation, Local Economic Growth, and Entrepreneurship. SSRN Electron. J. 48. https://doi.org/10.2139/ssrn.2097842

Jaffe, A.B., 1989. Real effects of academic research. Am. Econ. Rev. 79, 957–970. https://doi.org/10.2307/1831431

Jaffe, A.B., Lerner, J., 2001. Reinventing public R&D: Patent policy and the commercialization of national laboratory technologies. RAND J. Econ. 167–198.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge

Spillovers as Evidenced by Patent Citations. Q. J. Econ. 108, 577–598. https://doi.org/10.2307/2118401

Kantor, S., Whalley, A., 2019. Research proximity and productivity: long-term evidence from agriculture. J. Polit. Econ. 127, 819–854.

Kantor, S., Whalley, A., 2014. Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks. Rev. Econ. Stat. 96, 171–188. https://doi.org/10.1162/REST_a_00357

Kenney, M., 1986. Biotechnology: The university-industrial complex. Yale University Press.

Kim, Y., Kim, W., Yang, T., 2012. The effect of the triple helix system and habitat on regional entrepreneurship: Empirical evidence from the U.S. Res. Policy 41, 154–166. https://doi.org/10.1016/j.respol.2011.08.003

Lucas, R.E., 1988. On the mechanics of economic development. J. Monet. Econ. 22, 3–42. https://doi.org/10.1016/0304-3932(88)90168-7

Moretti, E., 2012. The new geography of jobs. Houghton Mifflin Harcourt.

Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. J. Econom. 121, 175–212. https://doi.org/10.1016/J.JECONOM.2003.10.015

Mowery, D.C., Rosenberg, N., 1993. The US national innovation system. Natl. Innov. Syst. A Comp. Anal. 29–75.

Nelson, R.R., 1959. The simple economics of basic scientific research. J. Polit. Econ. 67, 297–306.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econom. J. Econom. Soc. 1417–1426.

Rosenberg, N., Nelson, R.R., 1994. American universities and technical advance in industry. Res. Policy 23, 323–348. https://doi.org/10.1016/0048-7333(94)90042-6

Shulman, L.S., 2001. The Carnegie classification of institutions of higher education. Menlo Park Carnegie Publ.

Sohn, E., 2020. How Local Industry R&D Shapes Academic Research: Evidence from the Agricultural Biotechnology Revolution. Organ. Sci. https://doi.org/10.1287/orsc.2020.1407

Stephan, P.E., 2012. How economics shapes science. Harvard University Press Cambridge, MA.

Valero, A., Van Reenen, J., 2019. The economic impact of universities: Evidence from across the globe. Econ. Educ. Rev. 68, 53–67.

Zucker, L.G., Darby, M.R., Armstrong, J., 1998. Geographically localized knowledge: spillovers or markets? Econ. Inq. 36, 65–86.

## TABLE 1

### Variable Definition and Summary Statistics

| | Definition | Source | Mean | Std. Dev. |
|---|---|---|---|---|
| ***Outcome Variables*** | | | | |
| # of Ventures | Number of newly-formed ventures in each ZCTA, by year (1989-2012) | Guzman and Stern (2016) | 41.78 | 101.20 |
| RECPI | [10,000 * average quality * # of Ventures], within a ZCTA and by year (1989-2012) | Guzman and Stern (2016) | 262.50 | 1,074.00 |
| ***Public Research Measures*** | | | | |
| National Lab | 1 if at least one National Research Laboratory is in a 5-miles radius of the ZCTA | National Science Foundation | 0.006 | 0.079 |
| Research University | 1 if at least one university classified as “Doctoral/Research Universities” or “Research Universities – (very) high research activity" is in a 5-miles radius of the ZCTA | National Center for Education Statistics, Carnegie Classification | 0.062 | 0.241 |
| # of Research Universities | Number of Research Universities in a 5-miles radius of the ZCTA | Own elaboration | 0.098 | 0.471 |
| Top50 | 1 if at least one of the universities in a 5 miles radius of the ZCTA is in the top 50 of the *Times Higher Education 2017 ranking* of US universities | Times Higher Education | 0.018 | 0.133 |
| Top10 | 1 if at least one of the universities in a 5 miles radius of the ZCTA is in the top 10 of the Times Higher Education 2017 ranking of US universities | Times Higher Education | 0.006 | 0.080 |
| Business School | 1 if there is at least one business school in a research university in a 5 miles radius of the ZCTA | Wikipedia | 0.048 | 0.214 |
| Medical School | 1 if there is at least one medical school in a research university in a 5 miles radius of the ZCTA | Wikipedia | 0.036 | 0.187 |
| Law school | 1 if there is at least one law school in a research university in a 5 miles radius of the ZCTA | Wikipedia | 0.040 | 0.197 |
| R&D funds to universities | Total amount of federal funds ($k) earmarked for R&D activities received by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 10,490 | 68,942 |
| Non-R&D funds to universities | Total amount of federal funds ($k) earmarked for non R&D activities received by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 1,247 | 7,912 |
| R&D funds to national laboratories | Total amount of federal funds ($k) received by the national laboratories in a 5 miles radius of the ZCTA, per year | National Science Foundation | 3,319 | 68,956 |
| | Total number of ZCTA | | 16,320 | |
| | Total number of observations (panel) | | 408,000 | |

Note: All dollar amounts are inflation-adjusted and standardized to 2012 dollars.

## TABLE 2

## Quantity and Quality-Adjusted Quantity of Entrepreneurship: With and Without a Research Institution

| **Research University = 0 (15,307 ZCTAs)** | | | **Research University = 1 (1,013 ZCTAs)** | | | |
|---|---|---|---|---|---|---|
| *Variable* | *Mean* | *Std. Dev.* | *Variable* | *Mean* | *Std. Dev.* | *T-test of means* |
| # of Ventures | 35.6 | 92.0 | # of Ventures | 134.6 | 167.5 | *** |
| RECPI | 200.0 | 751.7 | RECPI | 1208 | 3015.1 | *** |

| **National Laboratory = 0 (16,220 ZCTAs)** | | | **National Laboratory = 1 (100 ZCTAs)** | | | |
|---|---|---|---|---|---|---|
| *Variable* | *Mean* | *Std. Dev.* | *Variable* | *Mean* | *Std. Dev.* | *T-test of means* |
| # of Ventures | 41.5 | 101.2 | # of Ventures | 89.6 | 89.7 | *** |
| RECPI | 258.4 | 1069.9 | RECPI | 929.6 | 1452.1 | *** |

* p<0.05 ** p<0.01 *** p<0.001

## TABLE 3

## Number of Ventures and RECPI By Presence of Research Institutions: Year and Zipcode Controls

| | *# of Ventures* | | *RECPI* | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Research University | 99.67*** | 22.20*** | 1,003.23*** | 399.98*** |
| | (4.74) | (4.84) | (81.14) | (74.57) |
| National Laboratory | -10.87 | -39.05*** | 77.70 | -235.72* |
| | (9.15) | (6.77) | (123.38) | (112.88) |
| Constant | 12.54*** | -87.88*** | 79.28*** | -585.07*** |
| | (0.38) | (10.47) | (4.41) | (142.31) |
| Observations | 408,000 | 408,000 | 408,000 | 408,000 |
| Number of ZCTAs | 16,320 | 16,320 | 16,320 | 16,320 |
| Census 1990 Controls | | Yes | | Yes |
| R-squared | 0.0896 | 0.362 | 0.0577 | 0.236 |
| Years FE | Yes | Yes | Yes | Yes |
| | | | w | |

All specifications include year fixed effects. Columns 2 and 4 also include controls for the 1990 census variables presented in Appendix B, Table B.1. Robust standard errors in parenthesis, clustered at the ZCTA level. * p<0.05 ** p<0.01 *** p<0.001

# TABLE 4

## Number of Ventures and RECPI By Institutional and Geographic Characteristics

Columns 1 and 5 examine the effect of the quality of the universities in a 5 miles radius of the focal ZCTA. Columns 2 and 6 examine the effect of the presence of specialized schools in a 5 miles radius of the focal ZCTA. Columns 3 and 7 examine the interaction between the presence of at least one research university and the urbanization/income of the focal ZCTA. Columns 4 and 8 present the full specification.

| | **# of Ventures** | | | | **RECPI** | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Research university | 24.40*** | 11.25 | 2.04 | -7.32 | 260.19*** | 84.52 | -167.90* | -166.48 |
| | (5.14) | (7.47) | (7.01) | (8.88) | (71.13) | (160.68) | (67.68) | (145.72) |
| National Laboratory | -35.92*** | -42.52*** | -40.92*** | -38.14*** | -765.70*** | -275.49* | -258.86* | -760.65*** |
| | (7.38) | (7.02) | (7.37) | (7.82) | (184.71) | (118.28) | (118.11) | (184.13) |
| ***Quality of the university*** | | | | | | | | |
| Top 50 | -6.26 | | | -10.01 | -45.18 | | | -116.26 |
| | (11.82) | | | (12.14) | (152.09) | | | (162.51) |
| Top 10 | -9.16 | | | -22.16 | 2,229.9*** | | | 2,094.4*** |
| | (17.56) | | | (17.73) | (445.11) | | | (438.82) |
| ***Specialized School*** | | | | | | | | |
| Business School | | -9.85 | | -7.18 | | 74.20 | | -84.19 |
| | | (8.66) | | (8.84) | | (138.70) | | (139.62) |
| Medical School | | 3.31 | | 6.06 | | 64.68 | | 197.95 |
| | | (9.57) | | (9.86) | | (142.49) | | (144.38) |
| Law School | | 27.19** | | 30.48** | | 366.37** | | 88.91 |
| | | (10.43) | | (10.56) | | (137.23) | | (130.07) |
| ***Characteristics of the ZCTA*** | | | | | | | | |
| Research University*Urban ZCTA | | | 10.12 | 4.34 | | | 398.48*** | 177.84 |
| | | | (8.57) | (9.01) | | | (106.04) | (120.50) |
| Research University*High Income ZCTA | | | 40.53*** | 45.19*** | | | 948.25*** | 742.77*** |
| | | | (11.78) | (12.13) | | | (204.00) | (192.95) |
| Constant | -87.49*** | -89.43*** | -92.36*** | -92.99*** | -645.46*** | -612.27*** | -615.81*** | -664.86*** |
| | (10.47) | (10.43) | (10.57) | (10.48) | (141.48) | (142.25) | (145.89) | (146.18) |
| Observations | 408,000 | 408,000 | 408,000 | 408,000 | 408,000 | 408,000 | 408,000 | 408,000 |
| R-squared | 0.362 | 0.363 | 0.367 | 0.368 | 0.257 | 0.239 | 0.247 | 0.265 |
| Number of ZCTAs | 16,320 | 16,320 | 16,320 | 16,320 | 16,320 | 16,320 | 16,320 | 16,320 |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Census 1990 Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Robust standard errors in parenthesis, clustered at the ZCTA level. * p<0.05 ** p<0.01 *** p<0.001

# TABLE 5

## The Impact of Funding on the Quantity of Entrepreneurship: Fixed Effects, Long First Differences and Anderson-Hsiao Instrumental Variable Specification

| | *Long Panel with FE [2 years]* | | *Long First Differences [2 years]* | | *Long First Differences [3 years]* | |
|---|---|---|---|---|---|---|
| | | | | With Instrument | | With Instrument |
| | $(\text{\# of Ventures})_t$ | $(\text{\# of Ventures})_t$ | $\Delta(\text{\# of Ventures})_t$ | $\Delta(\text{\# of Ventures})_t$ | $\Delta(\text{\# of Ventures})_t$ | $\Delta(\text{\# of Ventures})_t$ |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ***Covariates in levels*** | | | | | | |
| $(\text{\# of Ventures})_{t-1}$ | | 0.981*** | | | | |
| | | (0.003) | | | | |
| *Federal funds* | | | | | | |
| $(\text{R\&D funds to universities})_{t-1}$ | 0.348*** | 0.018*** | | | | |
| | (0.030) | (0.002) | | | | |
| $(\text{Non R\&D funds to universities})_{t-1}$ | 0.362*** | 0.056*** | | | | |
| | (0.090) | (0.017) | | | | |
| $(\text{R\&D funds to national laboratories})_t$ | 0.014* | 0.001 | | | | |
| | (0.006) | (0.001) | | | | |
| ***Covariates in first differences*** | | | | | | |
| $\Delta(\text{\# of Ventures})_{t-1}$ | | | 0.646*** | 0.537*** | 0.714*** | 0.574*** |
| ` | | | (0.009) | (0.021) | (0.012) | (0.020) |
| *Federal funds* | | | | | | |
| $\Delta(\text{R\&D funds to universities})_{t-1}$ | | | 0.102*** | 0.120*** | 0.049*** | 0.081*** |
| | | | (0.011) | (0.013) | (0.008) | (0.010) |
| $\Delta(\text{Non R\&D funds to universities})_{t-1}$ | | | 0.079* | 0.087* | 0.027 | 0.043 |
| | | | (0.040) | (0.041) | (0.043) | (0.046) |
| $\Delta$(R&D funds to national laboratories | | | 0.016*** | 0.017*** | 0.011*** | 0.013*** |
| | | | (0.002) | (0.003) | (0.002) | (0.002) |
| Constant | 15,100*** | -253.88** | 9,173*** | 9,520*** | 130.71 | 19,722*** |
| | (521.51) | (97.83) | (1,049.34) | (1,122.13) | (2,136.67) | (2,008.57) |
| Observations | 375,360 | 359,040 | 103,230 | 103,230 | 57,350 | 57,350 |
| R-squared | 0.169 | 0.935 | 0.594 | 0.589 | 0.636 | 0.629 |
| Number of ZCTAs | 16,320 | 16,320 | 11,470 | 11,470 | 11,470 | 11,470 |
| Year FE | YES | YES | | | | |
| ZCTA FE | YES | YES | | | | |
| Counties FE | | | YES | YES | YES | YES |
| State*Year FE | | | YES | YES | YES | YES |

Note: In column (1), both endogenous variable and covariates are expressed as 2-year averages of the original variables: $\overline{X}_t = \frac{x_t + x_{t-1}}{2}$. In columns (2), (3), we reduce the 25-year panel into 12 periods and difference out the resulting panel: $\Delta\overline{X}_t = \overline{X}_t - \overline{X}_{t-4}$. In column (4), we re-transform all our original variables as 3-year averages ($\overline{\overline{X}}_t = \frac{x_t + x_{t-1} + x_{t-2}}{3}$), reduce the 25-year panel into 8 periods and difference out: $\Delta\overline{\overline{X}}_t = \overline{\overline{X}}_t - \overline{\overline{X}}_{t-6}$. Notation simplified and Number of Ventures scaled by 1000 for exposition. Robust standard errors in parenthesis. * p<0.05 ** p<0.01 *** p<0.001

.

## TABLE 6

### The Impact of Funding on RECPI: Fixed Effects, Long First Differences and Anderson-Hsiao Instrumental Variable Specification

| | *Long Panel with FE [2 years]* | | *Long First Differences [2 years]* | | *Long First Differences [3 years]* | |
|---|---|---|---|---|---|---|
| | | | | With Instrument | | With Instrument |
| | $(\text{RECPI})_t$ | $(\text{RECPI})_t$ | $\Delta(\text{RECPI})_t$ | $\Delta(\text{RECPI})_t$ | $\Delta(\text{RECPI})_t$ | $\Delta(\text{RECPI})_t$ |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ***Covariates in levels*** | | | | | | |
| $(\text{RECPI})_{t-1}$ | | 0.862*** | | | | |
| | | (0.017) | | | | |
| *Federal funds* | | | | | | |
| $(R\&D\ funds\ to\ universities)_{t-1}$ | 6.134*** | 1.132*** | | | | |
| | (0.734) | (0.137) | | | | |
| $(\text{Non R\&D funds to universities})_{t-1}$ | -5.470** | -2.185** | | | | |
| | (1.917) | (0.775) | | | | |
| $(\text{R\&D funds to national laboratories})_{t-1}$ | -0.199** | -0.039 | | | | |
| | (0.074) | (0.021) | | | | |
| ***Covariates in first differences*** | | | | | | |
| $\Delta(\text{RECPI})_{t-1}$ | | | 0.346*** | -0.044 | 0.436*** | 0.230* |
| | | | (0.024) | (0.118) | (0.038) | (0.09) |
| *Federal funds* | | | | | | |
| $\Delta(\text{R\&D funds to universities})_{t-1}$ | | | 0.806* | 1.257*** | 2.034*** | 2.509*** |
| | | | (0.336) | (0.363) | (0.461) | (0.47) |
| $\Delta(\text{Non R\&D funds to universities})_{t-1}$ | | | -7.335*** | -7.270*** | -13.652*** | -13.060*** |
| | | | (1.921) | (1.806) | (3.199) | (3.11) |
| $\Delta(\text{R\&D funds to national laboratories})_{t-1}$ | | | -0.233* | -0.312* | -0.297** | -0.322** |
| | | | (0.098) | (0.125) | (0.111) | (0.117) |
| Constant | 96,977*** | 31,599*** | 28,433** | | 21,010*** | 95,049*** |
| | (8,208) | (5,802) | (10,458) | | (19,717) | (20,057) |
| Observations | 375,360 | 359,040 | 103,230 | 103,230 | 57,350 | 57,350 |
| R-squared | 0.057 | 0.719 | 0.165 | 0.041 | 0.239 | 0.216 |
| Number of ZCTAs | 16,320 | 16,320 | 11,470 | 11,470 | 11,470 | 11,470 |
| Year FE | YES | YES | | | | |
| ZCTA FE | YES | YES | | | | |
| Counties FE | | | YES | YES | YES | YES |
| State*Year FE | | | YES | YES | YES | YES |

Note: In column (1), both endogenous variable and covariates are expressed as 2-year averages of the original variables: $\overline{\text{X}}_t = \frac{x_t + x_{t-1}}{2}$. In columns (2), (3), we reduce the 25-year panel into 12 periods and difference out the resulting panel: $\Delta\overline{\text{X}}_t = \overline{\text{X}}_t - \overline{\text{X}}_{t-4}$. In column (4), we re-transform all our original variables as 3-year averages ($\overline{\overline{\text{X}}}_t = \frac{x_t + x_{t-1} + x_{t-2}}{3}$), reduce the 25-year panel into 8 periods and difference out: $\Delta\overline{\overline{\text{X}}}_t = \overline{\overline{\text{X}}}_t - \overline{\overline{\text{X}}}_{t-6}$. Notation simplified and RECPI scaled by 1000 for exposition. Robust standard errors in parenthesis.

* p<0.05 ** p<0.01 *** p<0.001.

## TABLE 7

## The Impact of Funding on RECPI and Number of Ventures: Interactions with Institutional and Geographic Characteristics

| | $\Delta(\text{\# of Ventures})_t$ | | | $\Delta(\text{RECPI})_t$ | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\Delta(\text{\# of Ventures})_{t-1}$ | 0.535*** | 0.512*** | 0.530*** | | | |
| | (0.022) | (0.023) | (0.022) | | | |
| $\Delta(\text{RECPI})_{t-1}$ | | | | -0.082 | -0.086 | -0.075 |
| | | | | (0.122) | (0.124) | (0.121) |
| **Federal Funds** | | | | | | |
| $\Delta(\text{R\&D funds to universities})_{t-1}$ | 0.169*** | 0.075 | 0.180*** | 1.987** | 1.882** | 1.551* |
| | (0.027) | (0.041) | (0.024) | (0.708) | (0.686) | (0.767) |
| $\Delta(\text{Non R\&D funds to universities})_{t-1}$ | 0.071 | 0.091* | 0.079 | -7.821*** | -7.222*** | -7.368*** |
| | (0.042) | (0.041) | (0.040) | (1.851) | (1.802) | (1.813) |
| $\Delta(\text{R\&D funds to national laboratories})_{t-1}$ | 0.016*** | 0.016*** | 0.016*** | -0.350* | -0.362** | -0.382** |
| | (0.003) | (0.003) | (0.003) | (0.138) | (0.130) | (0.129) |
| ***Interactions*** | | | | | | |
| $\Delta(\text{R\&D funds univ})_{t-1} * \text{top50 Univ}$ | -0.098** | | | -4.728*** | | |
| | (0.030) | | | (1.123) | | |
| $\Delta(\text{R\&D funds univ})_{t-1} * \text{Urban ZCTA}$ | | 0.028 | | | -1.235 | |
| | | (0.044) | | | (0.788) | |
| $\Delta(\text{R\&D funds univ})_{t-1} * \text{\# of Research Uni}$ | | | -0.030*** | | | -0.879*** |
| | | | (0.005) | | | (0.177) |
| Observations | 103,230 | 103,230 | 103,230 | 103,230 | 103,230 | 103,230 |
| R-squared | 0.589 | 0.587 | 0.588 | 0.030 | 0.025 | 0.031 |
| Number of ZCTAs | 11,470 | 11,470 | 11,470 | 11,470 | 11,470 | 11,470 |
| State*Year FE | YES | YES | YES | YES | YES | YES |
| Counties FE | YES | YES | YES | YES | YES | YES |

Note: The baseline specification for columns (1)-(3) is in Table 5, col. (4). The baseline specification for columns (4) - (7) is in Table 6, col. (4). Main effects on Top 50 Univ, Urban ZCTA, # of Research Universities, and the Constant are omitted for exposition.
All endogenous and federal funds variables are in a long difference format. Notation simplified and endogenous variables scaled by 1000 for exposition. Robust standard errors in parenthesis.
* p<0.05 ** p<0.01 *** p<0.001
.

## TABLE 8
## The Impact of Funding on RECPI and Number of Ventures: All Sources of Funding

| | $\Delta(\text{\# of Ventures})_t$ | $\Delta(\text{RECPI})_t$ |
|---|---|---|
| $\Delta(\text{\# of Ventures})_{t-1}$ | 0.536***<br>(0.021) | |
| $\Delta(\text{RECPI})_{t-1}$ | | -0.090<br>(0.121) |
| **Federal Funds** | | |
| $\Delta(\text{R\&D funds to universities})_{t-1}$ | 0.097***<br>(0.013) | -0.074<br>(0.572) |
| $\Delta(\text{Non R\&D funds to universities})_{t-1}$ | 0.104*<br>(0.041) | -8.067***<br>(1.912) |
| $\Delta(\text{R\&D funds to national laboratories})_{t-1}$ | 0.018***<br>(0.003) | -0.322*<br>(0.134) |
| **Other Funding Sources** | | |
| $\Delta(\text{State funds to universities})_{t-1}$ | 0.164***<br>(0.050) | -0.283<br>(2.043) |
| $\Delta(\text{Institutional funds to universities})_{t-1}$ | -0.011<br>(0.033) | 2.102<br>(1.573) |
| $\Delta(\text{Business funds to universities})_{t-1}$ | -0.081<br>(0.051) | 23.913***<br>(6.149) |
| $\Delta(\text{Other funds to universities})_{t-1}$ | 0.116**<br>(0.037) | 4.653<br>(3.230) |
| Observations | 103,230 | 103,230 |
| R-squared | 0.589 | 0.028 |
| Number of ZCTAs | 11,470 | 11,470 |
| Year*State FE | YES | YES |
| Counties FE | YES | YES |

Note: The baseline specification is in Table 5 and 6, col. (4). Constant omitted for exposition.
All variables are in a long difference format. Notation simplified and endogenous variables scaled by 1000 for exposition.
Robust standard errors in parenthesis.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

## TABLE 9

## The Impact of Awarded Degrees on RECPI and Number of Ventures: Natural Sciences, Social Sciences and Humanities

| | $\Delta$(# of Ventures)$_t$ (1) | $\Delta$(# of Ventures)$_t$ (2) | $\Delta$(RECPI)$_t$ (3) | $\Delta$(RECPI)$_t$ (4) |
|---|---|---|---|---|
| $\Delta$(# of Ventures)$_{t-1}$ | 0.537*** | 0.538*** | | |
| | (0.021) | (0.021) | | |
| $\Delta$(RECPI)$_{t-1}$ | | | -0.051 | -0.061 |
| | | | (0.119) | (0.120) |
| *Federal funds* | | | | |
| $\Delta$(R&D funds to universities)$_{t-1}$ | 0.120*** | 0.133*** | 1.072** | 1.307** |
| | (0.013) | (0.015) | (0.394) | (0.445) |
| $\Delta$(Non R&D funds to universities)$_{t-1}$ | 0.097* | 0.046 | -6.582*** | -6.268*** |
| | (0.041) | (0.041) | (1.754) | (1.739) |
| $\Delta$(R&D funds to national laboratories)$_{t-1}$ | 0.017*** | 0.015*** | -0.328* | -0.276* |
| | (0.003) | (0.003) | (0.133) | (0.117) |
| *Degrees Awarded* | | | | |
| $\Delta$(Bachelor students natural science)$_{t-1}$ | | 6.262** | | 318.719*** |
| | | (1.970) | | (72.103) |
| $\Delta$(Master students natural science)$_{t-1}$ | | 5.452 | | -83.530 |
| | | (3.067) | | (154.357) |
| $\Delta$(PhD students natural science)$_{t-1}$ | | -13.911** | | -45.943 |
| | | (4.973) | | (228.721) |
| $\Delta$(Total students natural science)$_{t-1}$ | 3.528** | | 231.991*** | |
| | (1.257) | | (57.938) | |
| $\Delta$(Bachelor students social sciences and humanities)$_{t-1}$ | | -0.653 | | -73.245* |
| | | (0.981) | | (31.414) |
| $\Delta$(Master students social sciences and humanities)$_{t-1}$ | | -1.105 | | 4.776 |
| | | (0.788) | | (20.504) |
| $\Delta$(PhD students social sciences and humanities)$_{t-1}$ | | -4.949 | | 1,059.452** |
| | | (7.211) | | (393.664) |
| $\Delta$(Total students social sciences and humanities)$_{t-1}$ | -0.804* | | -34.959** | |
| | (0.396) | | (13.259) | |
| Observations | 103,230 | 103,230 | 103,230 | 103,230 |
| R-squared | 0.589 | 0.589 | 0.038 | 0.035 |
| Number of ZCTAs | 11,470 | 11,470 | 11,470 | 11,470 |
| Counties FE | YES | YES | YES | YES |
| State*Year FE | YES | YES | YES | YES |

Note: The baseline specification is in Table 5 and 6, col. (3). Constant omitted for exposition.
All variables are in a long difference format. Notation simplified and endogenous variables scaled by 1000 for exposition. Robust standard errors in parenthesis.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

## TABLE 10

## Case Study: Weber State University (UTAH)

In 1998, WSU's R&D federal funding more than tripled. The Table highlights the substantial increase in RECPI and in the Number of Ventures in the immediate vicinity of the university.

| Distance | Year | # of Ventures | RECPI | Lagged R&D fed funding | Growth # of Ventures | Growth RECPI | Lagged Growth R&D fed funding |
|---|---|---|---|---|---|---|---|
| 5-50 miles | 1998 | 9,479 | 14,093 | 2,148,481 | | | |
| 5-50 miles | 1999 | 10,528 | 14,464 | 2,242,103 | 11.1% | 2.6% | 4.4% |
| <5 miles | 1998 | 686 | 564 | 1,502 | | | |
| <5 miles | 1999 | 803 | 798 | 5,247 | 17.1% | **41.7%** | **249%** |

Note: R&D funding to WSU is lagged one year to mitigate reverse causality concerns: the 249% increase in funding refers to the 1997-1998 period. Again, all funding amounts are in thousands (2012 baseline year).

**FIGURE 1a**

***Impact of the Presence of a RESEARCH INSTITUTION on Number of Ventures over time***

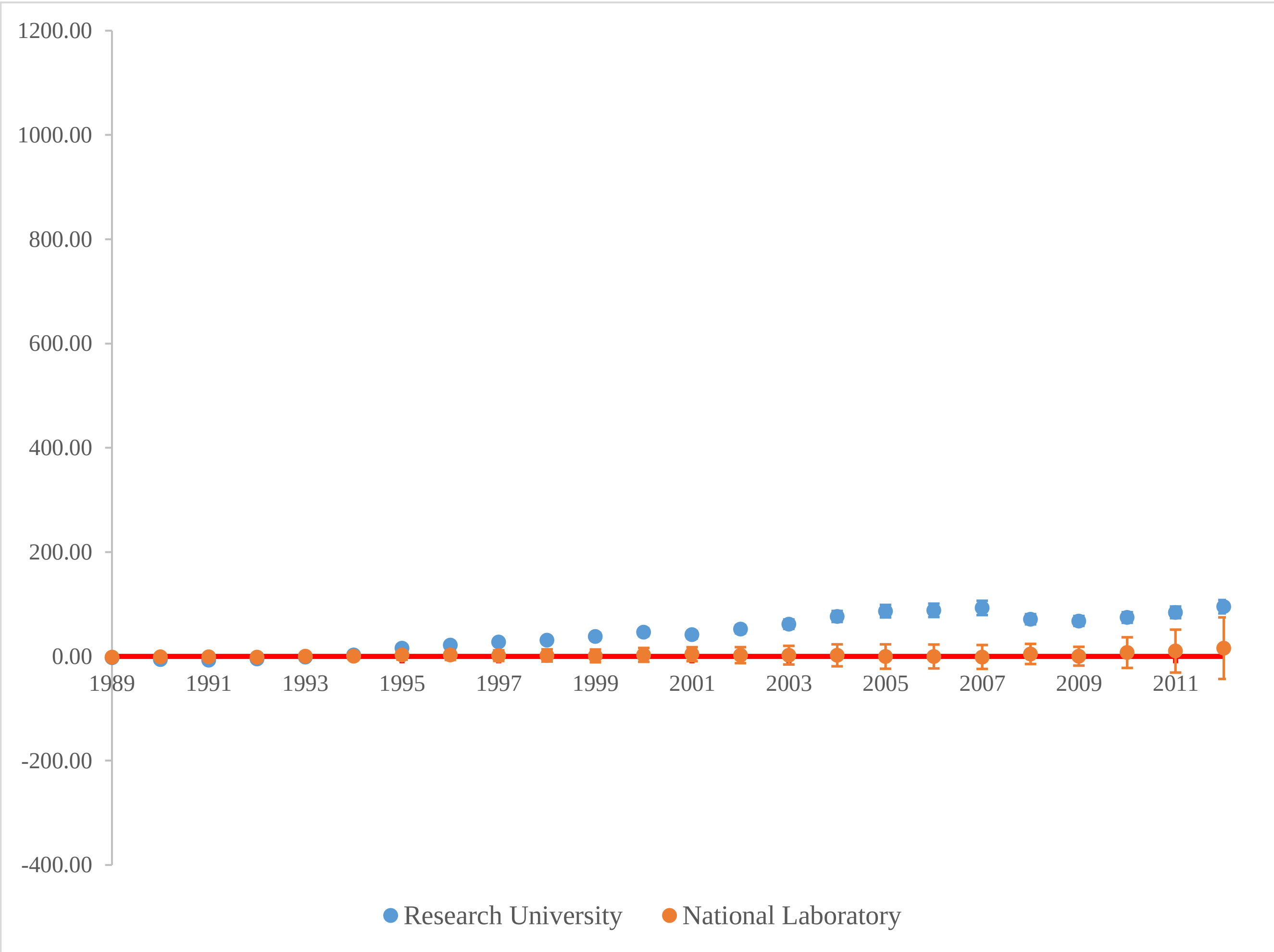
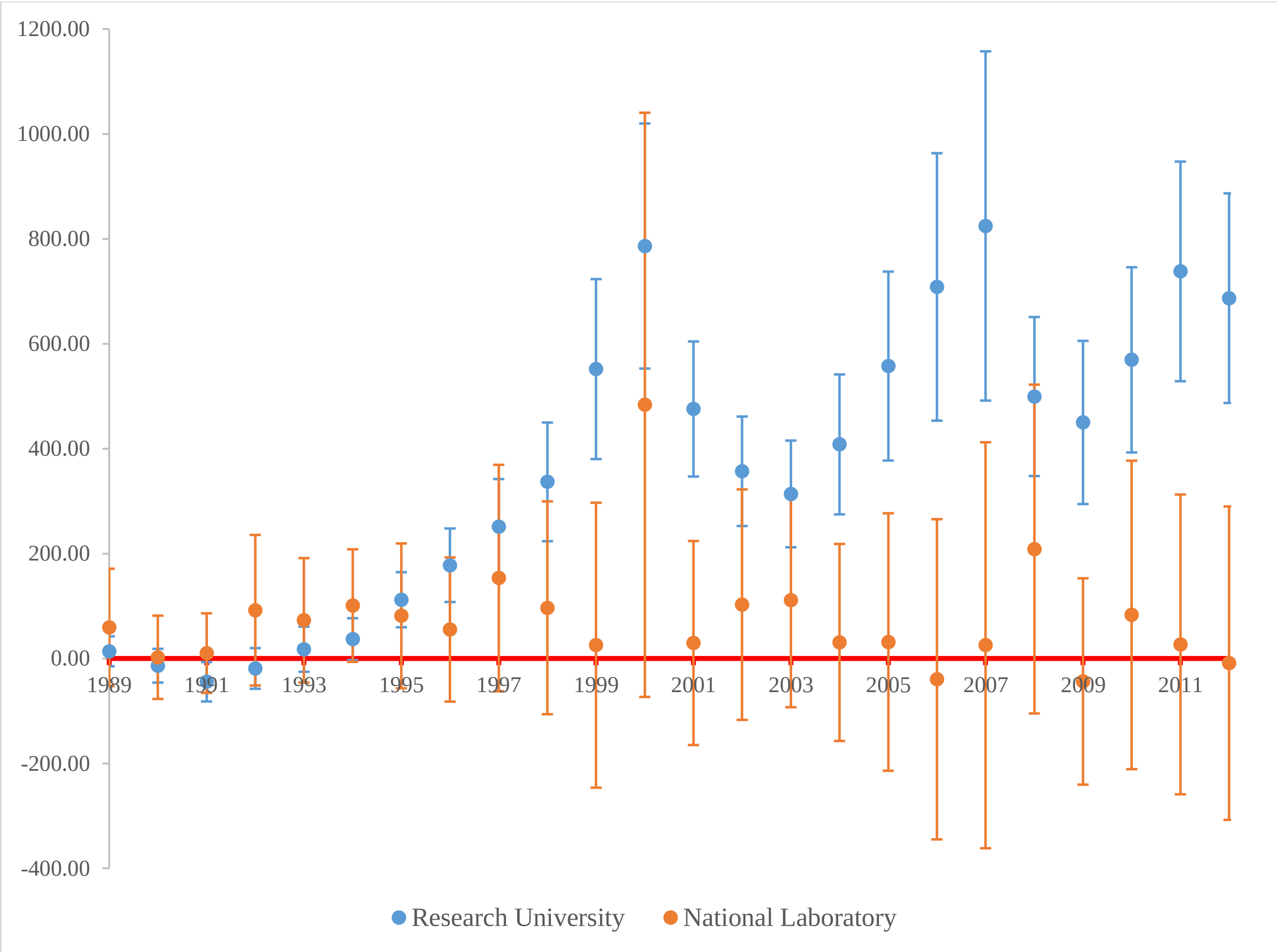


**FIGURE 1b**

***Impact of the Presence of a RESEARCH INSTITUTION on RECPI over time***



# APPENDIX A

## TABLE A.1

### Entrepreneurial Quality Logit Model

Predictive Analytics Model of Equity Growth
Dependent Variable: Equity Growth
Logit model. Incidence Rate Ratios Reported

| | Full model<br>(1) |
|---|---|
| *Corporate Governance Measures* | |
| Corporation | 3.202*** |
| | (0.0650) |
| | |
| *Name-Based Measures* | |
| Short Name | 1.786*** |
| | (0.0208) |
| | |
| Eponymous | 0.312*** |
| | (0.0150) |
| *Intellectual Property Measures* | |
| | |
| Trademark | 4.288*** |
| | (0.200) |
| *Patent - Delaware Interaction* | |
| Patent Only | 20.24*** |
| | (0.847) |
| | |
| Delaware Only | 15.26*** |
| | (0.294) |
| | |
| Patent and Delaware | 84.08*** |
| | (2.720) |
| | |
| US CMP Clusters | Yes |
| US CMP High-Tech Clusters | Yes |
| N | 26,051,461 |
| R-squared | 0.194 |

We estimate a logit model with Growth as the dependent variable. Growth is a binary indicator equal to 1 if a firm achieves IPO or acquisition within 6 years and 0 otherwise. Growth is only defined for firms born in the cohorts of 1988 to 2010. This model forms the basis of our entrepreneurial quality estimates, which are the predicted values of the model. Incidence ratios reported; Robust standard errors in parenthesis. * $p<0.05$ ** $p<0.01$ *** $p<0.001$

## TABLE A.2

## Funding and Degrees Controls: Definition and Summary Statistics

| | Definition | Source | Mean | Std. Dev. |
|---|---|---|---|---|
| **Funds to universities** | | | | |
| State funds to universities | Total amount of state funds ($k) received by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 981 | 7,389 |
| Institutional funds to universities | Total amount of funds ($k) received from institutional donors by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 3,136 | 18,785 |
| Business funds to universities | Total amount of funds ($k) received from business donors by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 1,148 | 8,175 |
| Other funds to universities | Total amount of funds ($k) received from other sources by the research universities in a 5 miles radius of the ZCTA, per year | National Science Foundation | 1,719 | 11,998 |
| **Degrees Awarded** | | | | |
| Bachelor students natural science | Total number of natural sciences bachelor degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 108.1 | 446.4 |
| Master students natural science | Total number of natural sciences Master degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 44.34 | 241.7 |
| PhD students natural science | Total number of natural sciences PhD degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 13.6 | 97.1 |
| Total students natural science | Total number of bachelor, Master and PhD degrees in natural sciences awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 166.0 | 753.7 |
| Bachelor students social sciences and humanities | Total number of humanities and social sciences bachelor degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 302.4 | 1,219 |
| Master students social sciences and humanities | Total number of humanities and social sciences Master degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 149.1 | 838.8 |
| PhD students social sciences and humanities | Total number of humanities and social sciences PhD degrees awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 14.0 | 90.9 |
| Total students social sciences and humanities | Total number of bachelor, Master and PhD degrees in humanities and social sciences awarded in a 5 miles radius of the ZCTA, per year | National Science Foundation | 465.4 | 2,067.0 |
| | Total number of ZCTA | | 16,320 | |
| | Total number of observations (panel) | | 408,000 | |

Note: All dollar amounts are inflation-adjusted and standardized to 2012 dollars.

# APPENDIX B

## TABLE B.1

## Census Controls: Summary Statistics

| | Definition | Source | Mean | Std Dev |
|---|---|---|---|---|
| ***Census Variables by ZCTA (1990)*** | | | | |
| Total population | Total population | Decennial Census | 9,367 | 12,893 |
| Density | Total population divided by the geographical area | Decennial Census | 0.0004 | 0.001 |
| Male | % of men in the total population | Decennial Census | 0.496 | 0.038 |
| Urban | % of the population residing in urban areas | Decennial Census | 0.349 | 0.431 |
| Age between 18 and 65 | % of the population between 18 and 65 years old | Decennial Census | 0.594 | 0.068 |
| White | % of the population of white ethnicity | Decennial Census | 0.884 | 0.182 |
| Born outside the US | % of the population born outside of the United States | Decennial Census | 0.042 | 0.077 |
| College | % of the population with a college degree or higher | Decennial Census | 0.210 | 0.131 |
| Private sector | % of the population employed in the private sector | Decennial Census | 0.720 | 0.118 |
| Public sector | % of the population employed in the public sector | Decennial Census | 0.153 | 0.083 |
| Income per capita | Median per capita income | Decennial Census | 12,506 | 5,780 |
| Housing values | Median house value | Decennial Census | 72,684 | 67,492 |
| | Total number of ZCTA | | 16,320 | |
| | Total number of observations (panel) | | 408,000 | |

## TABLE B.2

## # of Ventures and RECPI By Presence of Research Institutions: Census Controls

| | *# of Ventures* | *RECPI* |
|---|---|---|
| | (1) | (2) |
| Research University | 22.20*** | 399.98*** |
| | (4.84) | (74.57) |
| National Laboratory | -39.05*** | -235.72* |
| | (6.77) | (112.88) |
| Total population | 0.00*** | 0.01*** |
| | (0.00) | (0.00) |
| Density | -2,728.18* | -25,777.99+ |
| | (1,086.45) | (13,878.21) |
| Male | -61.10** | 351.21 |
| | (21.98) | (250.15) |
| Urban | 20.52*** | 42.62+ |
| | (2.30) | (24.01) |
| Age 18 - 65 | 29.98* | 810.16*** |
| | (13.10) | (168.38) |
| White | 16.24*** | -148.40*** |
| | (4.83) | (42.27) |
| Born outside the US | 269.78*** | 2,033.19*** |
| | (33.93) | (244.01) |
| College | 84.88*** | 146.21 |
| | (11.01) | (177.85) |
| Private sector | 18.61*** | -461.43*** |
| | (4.88) | (57.39) |
| Public sector | -20.01** | -820.14*** |
| | (7.13) | (84.63) |
| Income per capita | 0.00*** | 0.02* |
| | (0.00) | (0.01) |
| Housing values | -0.00*** | 0.00*** |
| | (0.00) | (0.00) |
| Constant | -87.88*** | -585.07*** |
| | (10.47) | (142.31) |
| Observations | 408,000 | 408,000 |
| Number of ZCTAs | 16,320 | 16,320 |

All specifications include year fixed effects. Robust standard errors in parenthesis, clustered at the ZCTA level. * p<0.05 ** p<0.01 *** p<0.001