

Entrepreneurial Spillovers from Corporate R&D

March, 2018

Tania Babina and Sabrina T. Howell*

Abstract

How does corporate innovation investment affect employee departures to entrepreneurship? Research and development (R&D) spending may decrease spawning if it increases firm's growth options or makes it a more interesting workplace. Conversely, R&D may increase spawning if it generates new ideas that employees can appropriate, or if it makes employees more entrepreneurial. Using U.S. employer-employee matched Census data, we show that R&D investment increases spawning. We identify the causal effect of R&D with changes in federal and state tax incentives. The effect is driven by high-tech parents and by departures to high-growth and VC-backed entrepreneurship. Intellectual rather than human capital seems to explain the spawning (i.e., new ideas rather than skills). The effect does not impose observable costs on the parent, leading us to conclude that entrepreneurial spawning is a source of knowledge spillovers from corporate R&D.

*Columbia GSB, NYU Stern. Emails: tania.babina@gsb.columbia.edu; sabrina.howell@nyu.edu. We are grateful to Manuel Adelino, Bronwyn Hall, Nick Bloom, Alon Brav, Wes Cohen, Song Ma, and Brian Lucking.

1 Introduction

Investing in innovation yields knowledge spillovers, which are difficult to observe but crucial to explaining economic growth (Marshall 1920, Krugman 1991). Knowledge spillovers are benefits from one firm's innovation efforts that accrue to other firms, and which are not embodied in products and services (Griliches 1992).¹ Despite their incorporeal nature, knowledge spillovers are known to be large in magnitude and seem to decline with geographic distance.² However, there is little evidence about their transmission channel. We also do not know much about the identity of spillover recipients; the literature has typically assumed that potential recipients are close in technological or geographic space.

This paper offers a new channel for knowledge to spill from one firm to another: We show that corporate investment in research and development (R&D) increases the rate at which employees depart to found their own firms. This effect is not obvious, as R&D might increase the firm's growth options or make it a more interesting place to work, leading to greater employee retention. Evidence that many successful entrepreneurs are former employees of high-tech, large firms motivates our research (Bhide 2000, Klepper 2001). Especially relevant is Gompers, Lerner & Scharfstein (2005), who observe that around 40 percent of venture capital-backed executive teams previously worked at a public company, often those with entrepreneurial, focused cultures. They note that one limitation of their analysis is that entrepreneurial individuals may select into working at companies with entrepreneurial learning opportunities. We depart from their approach, as well as other prior work, by demonstrating that greater

¹These pecuniary externalities could in theory ultimately can be measured through hedonic prices, as producer and consumer surplus. For example, after accounting for the improved quality of the R&D-intensive good, a downstream firm will have lower quality-adjusted input costs. Embodied spillovers may be difficult to measure - for example if price indices are not adjusted for quality changes - but they are nonetheless pecuniary externalities (Griliches 1992).

²On the first point, see Bernstein & Nadiri (1989), Jones & Williams (1998), Griffith, Harrison & Van Reenen (2006), and Bloom et al. (2013). On the second, see Jaffe et al. (1993) and Greenstone et al. (2010).

innovation inputs at public companies increases entrepreneurial spawning.³

We measure the propensity of employees to join startups' founding teams using U.S. Census employer-employee matched panel data.⁴ We first demonstrate a robust relationship between firm R&D investment and entrepreneurial spawning. We find that a one standard deviation increase in R&D is associated with a 18.7 percent increase in entrepreneurial spawning, relative to sample mean of 1.3 percent.

Our results are robust to including firm, state-year, and industry-year fixed effects, as well as time-varying firm and establishment characteristics such as age, diversification, Tobin's Q, total investment, size, and intangible assets. Further, the results are robust to including four-digit SIC code fixed effects, suggesting that narrow industries do not explain the result. Finally, they are robust to several measures of entrepreneurial spawning.

However, there is concern that a new technological opportunity not captured by these controls may lead to both the parent R&D and the spawn. To address this concern, we instrument for R&D using changes in state and federal R&D tax credits, following Bloom et al. (2013). Changes in tax credits affect the firm's tax price of R&D (or equivalently its user cost of R&D capital), and thus its incentives to invest in R&D. We provide exhaustive detail on the sources of within-firm variation for both instruments. The federal tax credit is firm-specific for five reasons, importantly because it depends on firm age, with annual changes for most firms. The state instrument is firm-specific because it is calculated using the time-varying share of the firm's patent inventors located in a given state. The instruments satisfy the relevance condition and are likely to satisfy the exclusion restriction.⁵

³Existing work on spawning includes Hellmann (2007), Sørensen (2007), Chatterji (2009), Nanda & Sørensen (2010), Campbell et al. (2012), Babina (2015), and Babina, Ouimet & Zarutskie (2015). Research on knowledge spillovers at the micro-level has focused on inventor networks, particularly in academia, and exploiting immigration waves or superstar scientist deaths for exogenous variation in (e.g. Waldinger (2012), Azoulay et al. (2010)). By focusing on private firms rather than academics, and a real effect of R&D rather than patent citations, we extend this individual-level literature.

⁴For each public firm establishment-year, we follow departing workers and examine whether they are on the founding team of a new firm (top five earners of a firm founded within three years of when R&D is measured).

⁵To satisfy the relevance condition, we present extensive evidence that the elasticity of R&D spending to tax credits is at least one. To satisfy the exclusion restriction, we show empirically that there is no relationship between the tax credit and spawning, and present evidence from the legal literature that R&D tax credits are not in general useful to startups.

The instrumental variables (IV) effect of R&D on spawning is larger than the main effect, and equally robust. This offers strong evidence that the relationship is causal: firm R&D leads to new firm creation. It is logical that the IV estimate is larger. While the OLS strategy measures the effect of an additional dollar of average R&D, the IV strategy approximates increased R&D spending on the margin, capturing the effect on spawning of the “last” R&D dollar. This marginal R&D is likely farther from the parent’s core focus than average R&D, explaining why the growth options it creates more often optimally reside outside the firm’s boundary.

This framing demands evidence of a specific mechanism. Corporate R&D might increase spawning through what we term an “intellectual capital” channel. In this channel, R&D leads serendipitously to new ideas, some of which are deployed in new firms by departing employees. The alternative channel is “human capital,” where learning-by-doing during the R&D process increases employees’ entrepreneurial skills. For example, a manager overseeing a new R&D project might learn to guide employees under high uncertainty, and become better suited to an innovative startup.

Overall, our cross-sectional evidence is more consistent with the intellectual capital channel. First, this channel should be more associated with the “research” part of R&D. We find that high-tech parents and more inventive R&D (parents with above-median forward patent citations) drive the effect. Second, it should be associated with new-to-the-world ideas, rather than “Main street” businesses. We find that within the population of spawns, more parent R&D is associated with venture capital-backed spawns. Third, we find that spawns tend to be in different industries from parents, suggesting that the spawn has a new idea and is not replicating the parent’s business. We find further cross-sectional evidence that is inconsistent with the human capital channel.

The intellectual capital channel is especially consistent with the spawning effect of R&D being a new avenue for knowledge spillovers. This requires that the cost to the parent of the R&D-induced spawn does not outweigh the spawns’s combined private and social value. In light of the large benefits of new firms (Kortum & Lerner 2000, Decker et al. 2014, Glaeser et al. 2015), it almost certainly does not. However, we consider two ways that control rights may matter: (a) If the spawn is very costly to

the parent; and (b) if the parent fully internalizes the spawn’s benefits.

If R&D-induced spawning is very costly, the effect should be attenuated in states that enforce non-compete covenants. Instead, the effect persists and is not much smaller in these states. Second, the effect should be weaker in sectors where intellectual property is easier to protect. We do not find that the effect varies with a measure of industry patentability. Third, costly spawns may compete in product markets with parent. Instead, R&D-induced spawns tend to be in different industries from their parents. We also make a revealed preference argument. By virtue of observing the persistent phenomenon of R&D-induced spawning, the parent either chose not to develop the idea in house, or chose not to take ex-ante steps to prevent the spawn.

The spawning effect of R&D is less likely to be a new source of R&D spillovers if the parent internalizes the spawn’s private benefits by investing in or acquiring it. We conduct an out-of-sample test based on the underlying spawn-parent pair data in Gompers et al. (2005). Of the 9,152 unique parents in their data, just 2.3 percent invest in or acquire their spawns. This small percentage is evidence that parents do not usually internalize spawns by investing in or acquiring them.

Our finding sheds light on how the optimal boundaries of the firm depend on the nature of the growth options it creates (Seru 2014, Bernstein 2015). We find that more inventive R&D drives our effect, and further that R&D-induced spawns are much more likely to be VC-backed than the average spawn. Venture capital backing implies both a high-growth business and high-powered owner incentives (Metrick & Yasuda 2011). Gromb & Scharfstein (2002) show how growth options with high risk and high potential payoffs benefit from higher-powered incentives.⁶ In our setting, high-risk, high-reward growth options are reallocated from large incumbents to startups. This relates to Robinson (2008), who shows that it can be optimal to locate risky projects outside the firm boundary in formal alliances.

Finally, the spawning effect of R&D offers a new source for where ideas for startups come from, a direct channel of absorptive capacity (Cohen & Levinthal 1989, Aghion & Jaravel 2015). R&D is the only corporate variable that predicts VC-backed employee spawns, a fact that contributes to efforts such as Guzman & Stern (2017) to

⁶See also Zingales 2000 and Sevilir (2010).

predict which new firms will be high-growth. We show how high-skill R&D labor can be reallocated from incumbents to new firms, which Acemoglu et al. (2013) argue is likely welfare-enhancing. More generally, our paper is related to work on knowledge diffusion through labor mobility, including Almeida & Kogut (1999), Kim & Marschke (2005), Matray (2015), and Herkenhoff et al. (2018).

2 Data

We use data from five sources: Compustat, Census LBD, Census LEHD, VentureXpert, and the NBER Patent Data Project. This section describes each source of data and explains the key variables we use in analysis. It also discusses concerns with the data and sample.

2.1 Compustat

Our measure of corporate innovation investment is R&D expenditure as reported in 10K filings and provided by Compustat. As R&D expenditure is only available for public firms, they form our universe of firms at hazard of being parents, or spawning entrepreneurs out of their labor force. We primarily use log R&D, but also test whether the results are robust to using R&D divided by total assets. We also obtain balance sheet and income statement data about the potential parents from Compustat.

We consider only firms with positive R&D, for three reasons. First, firms that report R&D are likely qualitatively different from firms that do not in ways that might affect spawning, despite rigorous controls and fixed effects (Lerner & Seru 2017). Second, our primary specification will be focused on the intensive margin; since we use firm fixed effects, firms with zero R&D provide no variation. Third, Bloom et al. (2013) restrict the sample to firms with positive R&D, and we try to be as consistent with their first stage approach as possible.

2.2 Census LBD

We merge Compustat to the restricted-access U.S. Census Bureau’s Longitudinal Business Database (LBD), and establishment-level panel, using the internal Census Compustat/LBD crosswalk. The LBD covers all U.S. business establishments with paid employees beginning in 1976, described in detail in Jarmin & Miranda (2002). We use data for all 50 states from 1990 to 2011.

Establishments and firms are tracked consistently over time. We use data on establishment age, industry, physical location, total employment, payroll, birth, and death. We can therefore identify new employer firms and their future employment growth, payroll, and exit. We define age as the oldest establishment that the firm owns in the first year the firm is observed in the LBD, as in Haltiwanger et al. (2013). A firm birth is thus defined when all of its establishments are new, preventing us from misclassifying an establishment that changes ownership as a startup.

2.3 Census LEHD

A challenge when studying how R&D affects employee departures to entrepreneurship is that we must observe employees and track them from firm to firm. We solve this with the Longitudinal Employer-Household Dynamics (LEHD) program at the U.S. Census Bureau provides firm-worker matched data, permitting us track salaried employees over time across firms on a quarterly basis using worker social security numbers. In addition to wages, the data contain employees’ gender, race, place and date of birth, and citizenship status. Coverage starts in 1990 for several states and increases over time, ending in 2008. We have access to 31 states, shown in Figure 6, in which we observe all spawns. The LEHD covers over 96 percent of all private-sector jobs and over 96 percent of total wage and salary civilian jobs (BLS 1997, Abowd et al. 2009). Thus we do not face employee selection concerns present in other data sources, such as LinkedIn.

The LEHD connects quarterly earnings from the state Unemployment Insurance programs to the Quarterly Census of Employment and Wages Program. Abowd et al. (2009) describe the construction of this data in detail. Workers’ employers are identified

with State Employer Identification Numbers (SEIN, the state equivalent to EIN), but the data also include the federal employer identification number (EIN), which we use to link the LEHD to the LBD and thus follow employees from firm to firm. The LBD infrastructure is based on physical establishments while the LEHD infrastructure uses reporting units. We match LEHD SEINs to LBD EINs in the first quarter of each year, using an internal Census bridge file. We drop establishments with less than ten employees. This yields an annual panel of SEINs of the LBD firms, in which employees are observed as of the first quarter of each year. For ease of exposition, we term SEINs “establishments.”

2.3.1 Identifying spawns

To identify spawns, we begin by observing worker identities at public firm establishments in the first quarter of year t , and the quantity of R&D investment in year $t - 1$. We denote an establishment e . Using longitudinally consistent individual identifiers available in the LEHD data, we follow the establishment e ’s employees one, two, and three years after year t .

The LEHD data do not designate the founder(s) of a new firm. We proxy for an individual being on the founding team using the highest earners at a young firm. While this is an imperfect measure of entrepreneurship, we believe that it is the best available in administrative data. Founders may not pay themselves the highest wage as they seek to attract high-skill employees. Therefore, using only the highest earner is unlikely to capture all founders. Our definition captures both founders and the early employees who are important to the startup’s initial success. It is also in line with prior research focusing on the executive team, including Gompers et al. (2005). Our primary definition of a spawn is a firm founded between t and $t + 3$ in which any of the e ’s employees at year t is among the top five earners as of $t + 3$. To arrive at our primary outcome variable – an establishment’s spawning rate – we divide the number of spawn founders by e ’s total number of employees in year t .

There are four other outcomes for e ’s year t employees. First, they may remain at the firm. Second, they may be employed at a different firm that existed before year

t (other incumbents). Third, they may be employed at a firm whose age is unknown.⁷ Finally, the employee may no longer be observed in the data. They may have left the work force, no longer earn a wage, or otherwise fail to be covered by the LEHD (see concerns below). We use these outcomes in robustness tests, for example to test whether R&D also leads to greater labor mobility to other incumbent firms.

The final sample consists of an annual panel of public firm establishments in 31 states between 1990–2003. We follow startup creation from 1990 to 2006.

The LEHD has been widely used in economic research, for example in Tate & Yang (2015) and Goldin et al. (2017). There is nonetheless some concern about its coverage. About 17 percent of workers in year t are not in the LEHD in year $t + 3$. This may initially seem high, but the U.S. Current Population Survey (CPS) has a similar attrition rate. The CPS tracks workers for a maximum of 16 months. In the CPS data, among private sector employees who are observed 16 months later, about 11 percent drop out from the employment sample.⁸

A final concern is that parent R&D is correlated with worker mobility to or from uncovered state. If this is the case, then R&D should correlate with the fraction of workers who drop out of sample. We find that this is not the case.

2.4 VC and patent data

We use a linking between ThomsonOne VentureXpert and the Census Business Register to identify VC-backed startups from Puri & Zarutskie (2012). We use patent data from the NBER Patent Data Project, which includes patent and citation variables through 2006. The NBER data includes Compustat identifiers, allowing us to link to the parent firms (publicly traded firms) in the LBD. We use patent data to construct the instrument for R&D.

In our OLS regressions, we modulate the effect of R&D with annual patent-based variables at both the firm and industry level. These are the number of patent classes a firm or industry patents in, the number of patents, the number of forward

⁷ This is because some LEHD employers are non-profit and government entities not covered by the LBD, which is used to determine employer age.

⁸ Authors calculations based on IPUMS-CPS data, available at <https://cps.ipums.org/cps/>.

and backward citations, and the average, maximum, and median patent generality and originality. Generality is higher (closer to one than zero) when forward citations are in many classes, and originality is higher when backward citations are in many classes. In analysis, we use indicators for having an above median value for each patent variable within a year.

2.5 Summary statistics

Table 1 panels 1-6 show summary statistics at the parent firm, parent establishment, and spawn levels. For all levels, we divide the sample into low and high R&D changes, since the main analysis is identified using changes in R&D, and present the p-value for a difference of means test across the two groups.⁹ We measure entrepreneurial spawning at the establishment level (panels 3 and 4). These are the set of establishments of public firms with positive R&D and at least 10 employees, between 1990 and 2003. On average, 1.3 percent of an establishment’s employees separate and are identified as entrepreneurs three years later. Similarly, using the LBD/LEHD matched data Kerr et al. (2015) find that 1.7 percent of workers transition to entrepreneurship over a four-year period.

Panels 5 and 6 Table present summary statistics of the 108,000 spawns identified in the LBD. When we observe a former public firm employee first working at a spawn (three years after R&D is measured), the spawn is on average 1.6 years old and have 14.6 employees. Just 2 percent of spawns received VC funding. Among all new firms founded in 2016 nationwide, about 0.9 percent of new businesses raised VC.¹⁰ Thus the rate of VC backing among firm founders who formerly worked at public firms is higher than that in the whole population of firm founders.

⁹Change in $R\&D_{t-1}$ is defined as: $\frac{R\&D_{t-1}-R\&D_{t-2}}{.5 \cdot (R\&D_{t-1}+R\&D_{t-2})}$.

¹⁰According to the NVCA 2017 Yearbook, 6,611 unique companies raised angel, seed, or early VC funding in 2016, albeit not all for the first time. According to the U.S. BLS, in 2015 there were about 675,000 new businesses (allowing new firms one year to raise VC financing). See <https://www.bls.gov/bdm/entrepreneurship/entrepreneurship.htm> and <https://nvca.org/blog/nvca-2017-yearbook-go-resource-venture-ecosystem/>

3 Empirical Approach

This section first describes our primary estimation strategy, a tightly controlled ordinary least squares (OLS) regression. In Section 3.2, we explain our instrumental variables strategy.

3.1 Reduced form relationship between R&D and spawning

We estimate variants on Equation 1, where e denotes an establishment, f a firm, i an industry, s state, and t the year. The dependent variable is, as described in Section 2.3.1, is the percent of e_t 's employees who are among the top five earners at startups as of $t + 3$.

$$\begin{aligned} \text{Pct Entrepreneurs}_{efist+3} = & \beta \ln(\text{R\&D}_{f,t-t}) + \text{Firm FE}_f + \text{Industry-year FE}_{it} \\ & + \text{State-year FE}_{st} + \text{Controls}_{ft} + \text{Controls}_{et} + \varepsilon_{efist}. \end{aligned} \quad (1)$$

We employ either firm or establishment fixed effects to control for time-invariant differences across firms and establishments. We expect omitted variables to be correlated within the firm, so we cluster standard errors by firm. Industry-year fixed effects control for changes in investment opportunities, and also subsume industry and year effects. We use SIC four-digit industry codes in our most stringent specifications, and SIC three-digit codes elsewhere. State-year fixed effects control for regional shocks, which may affect investment opportunities at incumbents as well as entrepreneurship.

Time-varying establishment and firm controls address other concerns. First, we control for establishment size, in case, for example, smaller establishments have more focused or autonomous cultures and thus lead to more spawning. Second, we control for the establishment's average wage, in case spawning is driven by higher skilled workers rather than R&D. We also include the following firm-level controls, which might correlate with R&D and spawning: return on assets, sales growth, Tobin's Q, asset tangibility (measures as PPE investment divided by total assets), size (log total assets), cash holdings, age, and diversification (indicator for firm having establishments in more than one SIC 3-digit industry).

3.2 Instrument for R&D

The central challenge to Equation 1 is that an unobserved demand shock or new technological opportunity, not captured by our granular industry-year fixed effects, may jointly engender parent R&D and spawning. This is a version of the Manski (1993) reflection problem. The ideal experiment would randomly allocate R&D to firms, and observe whether firms assigned to more R&D have more employees that leave to found their own firms. This is infeasible, so we use the best available instrument for R&D expenditure: changes in state and federal R&D tax credits, following Bloom et al. (2013).

One concern with the use of these instruments in Bloom et al. (2013) is that they are used to instrument for both R&D and constructed spillover variables. The instrument may be correlated across firms in the same industry (Acemoglu 2011). We need only instrument for R&D, and thus do not face this issue.

This section first describes the motivation for the instrument (Section 3.2.1), and then addresses our expected direction of endogeneity (Section 3.2.2). In Section 3.2.3, we briefly explain the two tax prices of R&D that we use. Appendix Section 6 contains exhaustive details about this process, with subsections describing the federal tax credit and its calculation, the state tax credits, and concerns with instrument validity. While imperfect, we show that the instrumental variables strategy is well-suited to our context and is likely to satisfy the exclusion restriction.

3.2.1 Instrument motivation

We use two instruments: federal tax credit changes, and state tax credit changes. These are plausible sources of variation in corporate R&D expenditure. The federal R&D tax credit has been robustly shown to have a strong effect on corporate R&D in the short and long term. A substantial literature finds an elasticity of at least one, such that an extra dollar of federal tax R&D credits stimulates roughly a dollar of additional R&D expenditure (or much more, in some studies). This evidence includes Hall (1993), McCutchen (1993), Mamuneas & Nadiri (1996), Hall & Van Reenen (2000), Billings et al. (2001), Bloom et al. (2002), Klassen et al. (2004), and Clausen (2009).

Buttressing this evidence is the fact that firms claim the tax credit, or expense

essentially all of their qualified research expenditures (Guenther 2015). The relative sensitivity to the R&D tax credit may reflect the fact that firms tend to finance R&D out of free cash flows (Brown & Petersen 2011). We are confident that the federal tax price of R&D, if it has adequate firm-level variation, should predict R&D in our sample.

Paff (2005), Wu (2008) and Wilson (2009), among others, find that state R&D tax credits increase R&D within the affected state. The most conservative finding is in Wilson (2009), where a one percentage point increase in the state tax credit rate increases R&D by 1.7 percent in the short term and 3-4 percent in the longer term. However, Wilson (2009) also finds that the tax credits cause a reallocation of R&D activity geographically. Since large, multi-state firms are responsible for most R&D expenditure, we expect the state instrument to be generally weaker than the federal one.

3.2.2 Expected direction of endogeneity

There are two major sources of endogeneity that may bias our ordinary least squares (OLS) estimates: 1) technology shocks to the firm’s industry would bias the estimates upwards, and 2) the firm’s inability to fully appropriate the benefits of R&D would bias the estimates downwards. An example of the first source is a scientific discovery at a university that creates new opportunities for the firm’s industry. This may increase both firm R&D and entrepreneurship rates. The second source of endogeneity stems from firms’ investment being correlated with their ability to appropriate the investment’s benefits. This point is widely used to justify government subsidy of corporate R&D (cite here). The presence of a spawning effect represents benefits that the parent firm is not appropriating. This second source implies that if firms were randomly assigned R&D expenditures, we would expect a larger fraction of that R&D output to ultimately be developed outside of the firm’s boundaries in spawned startups.

In our setting, do we expect that endogeneity biases the OLS result upwards or downwards? While it is possible to tell stories going both ways, we believe it is more likely that endogeneity biases the OLS result down. Two facts suggest that positive bias due to technology shocks is unlikely. First, when we add industry-year fixed effects

to specifications with firm fixed effects, our estimates do not attenuate. Second, an opportunity shock in a given sector should lead to both more R&D and more startup formation in that sector. We find that the R&D-induced spawn’s line of business tends to be unrelated to the parent’s. While it is more difficult to test for the negative bias due to appropriation concerns, the instruments proposed in the following sections will address this concern. The instrumented effect’s magnitude relative to the OLS effect will be a test for the direction of bias.

3.2.3 Summaries of the tax credits

Changes in tax credits affect firm incentives to invest in R&D, because they change the firm-specific tax price of R&D. The lower the user cost of R&D capital, the more likely firms to invest in R&D. The first instrument is the federal tax price of R&D, which we denote ρ_{ft}^F . The Appendix contains a detailed description of the calculation, which draws from Hall (1993). We explain in the Appendix that the federal tax credit value depends on the firm’s qualified research expenditures and, crucially, a fixed base R&D spending. The credit is firm-specific for five reasons, including because it depends on firm age, with annual changes for most firms. We find substantial within-industry variation in the tax price of R&D, as well as the necessary variation within firm over time. We ensure that relevant current year variables, including R&D, do not have strong explanatory power over the tax price of R&D.

The state instrument, also described in more detail in the Appendix, requires two objects: the state tax price component of the R&D user cost of capital, and a measure of the share of a firm’s R&D that occurs in a given state. First, we use the state tax price of R&D in Wilson (2009). He incorporated state level corporate income taxes, depreciation allowances, and R&D tax credits into this tax price component, which we call ρ_{st}^S . These credits vary across states and time. To build the second object, θ_{fst} , we follow Bloom et al. (2013). θ_{fst} is a proxy for a firm’s R&D share in a given state-year calculated using the share of the firm’s patent inventors located in state s . The firm’s state-level tax price is then $\rho_{ft}^S = \sum_s \theta_{fst} \rho_{st}^S$.

3.2.4 First stage estimation

Having constructed firm-level federal and state tax prices of R&D (ρ_{ft}^F and ρ_{ft}^S , respectively), we estimate the following first stage regression:

$$\ln(R\&D_{ft}) = \beta_1 \ln(\rho_{ft}^S) + \beta_2 \ln(\rho_{ft}^F) + \text{Firm FE}_f + \text{Industry-year FE}_{it}. \quad (2)$$

$$+ \text{State-year FE}_{st} + \text{Controls}_{ft} + \varepsilon_{eft} \quad (3)$$

We cluster standard errors by firm. The results are in Table 2. We show all of the specifications that we will show in our main instrumented results table. The R^2 values are around 0.16, indicating that we are capturing a reasonable amount of variation in R&D. The instruments are strong, yielding F-statistics of about 25, well above the rule-of-thumb cutoff of ten (these are shown in Table 4). The federal instrument is stronger than the state instrument, which in part reflects the fact that the state instrument is identified by firms with patents. As we show below, our main result is not driven by firms with patents, but rather by firms in high-tech sub-sectors.

Note that Bloom et al. (2013) use only firm and year fixed effects. This is equivalent to column 1, though we also use firm time-varying controls (the results are essentially the same without these controls). We show a variety of specifications; our preferred specification, with SIC 3-digit industry-year and state-year fixed effects, along with firm time-varying controls and firm fixed effects, is in column 5. The results are also robust to using SIC 4-digit industry fixed effects (column 6).

3.2.5 Concerns

There are four potential concerns, which we describe in detail in the appendix. Here, we summarize the two more important ones. First, the exclusion restriction is that tax credits cannot affect entrepreneurial spawning. We show empirically that there is no relation between the state tax credits and startup creation. More generally, the legal literature has argued that R&D tax credits are not useful to startups because they usually do not have taxable income. (Bankman & Gilson 1999).

The second concern is that changes in state-level R&D tax credits may lead firms to reallocate R&D (or misreport it such that it appears reallocated). Any such

reallocation should reduce the power of the instrument. This leads us to expect that the federal instrument will have more power than the state instrument, which is indeed what we find. In sum, R&D tax credits offer the best available source of variation driving corporate R&D, which is plausibly unrelated to technological or demand shocks that could jointly give rise to parent R&D and entrepreneurial spawning.

4 Results

This section first explains our main results. We present the instrument result in Section 4.2. In Section 4.4, we consider reverse causation, and in Section 4.3, we consider alternative outcome variables and a restructuring hypothesis.

4.1 Main results

We present the main results from estimating Equation 1 in Table 3. Our preferred specification in column 4 includes firm, industry-year, and state-year fixed effects. The coefficient of 0.109 implies that a one standard deviation increase in R&D is associated with a 18.9 percent increase in entrepreneurial spawning, relative to sample mean of 1.3 percent. Alternatively, the coefficient implies that a 50 percent increase in R&D is associated with a 3 percent increase in entrepreneurial spawning. The main result is remarkably robust to a wide array of alternative controls and fixed effects as shown across the nine models in panels 1 and 2. For example, the result is robust to using SIC 4-digit industry fixed effects (panel 1 column 3 and panel 2 column 3).

Our baseline set of firm-level controls are reported in Panel 1. We do not report them in further results because we are strictly limited by the Census Bureau in the number of coefficients we may disclose. The controls are at the firm level, except for employment and payroll which are at the establishment level. The only control with any predictive power is employment; spawning is negatively associated with the establishment's number of employees. This could in part reflect the spawning effect itself, if replacement employees are not immediately hired. Panel 2 shows a different set of establishment employee level controls in column 4; establishments with older and longer

tenure workers are associated with slightly less spawning, while establishments with a higher share of white workers or foreign-born workers are associated with dramatically more spawning. In panel 2 column 5, we show that patents and citations do not predict spawning. However, patenting in more classes is associated with spawning.

Our results are also robust to alternative measures of R&D, shown in Appendix Table 2. When the independent variable is an indicator for the firm having had an above median change in last year’s R&D, the effect is .089, significant at the .01 level (column 1). This implies that moving from the bottom to the top half of R&D changes increases the rate of entrepreneurial spawning by 6.9 percent. The effect is naturally stronger using the top 10 percentiles of R&D change (columns 3 and 4). We also find that the effect is robust to using R&D divided by total assets (column 7).

4.2 IV Result

The results from the instrumented second stage are in Table 4. (The first stage results are described in Section 3.2.4 and are in Table 2.) We repeat the specifications from Table 3. The coefficients in all models are statistically significant, and they are also uniformly much larger than the OLS results. Our preferred specification, in column 5, implies that a 50 percent increase in R&D is associated with a 18.3 percent increase in spawning, or a one standard deviation increase in R&D is associated with a 100 percent increase.

The larger instrumented effect indicates that the subset of R&D expenditure affected by the tax credits leads to greater spawning than the average increase in R&D. This could reflect endogeneity that biases the OLS result downward, as discussed in Section 3.2.2 above. However, the local average treatment effect for the complier subset may also be larger than the population average treatment effect. As Angrist & Imbens (1995) and Jiang (2015) explain, this can lead an IV strategy to produce larger effects than the true effect, even if the exclusion restriction is satisfied. That is, compliers with the instrument (in our case, the change in the tax price of R&D), may be those firms with a higher causal effect of R&D on spawning.

There are three possible explanations for such a phenomenon. First, there may be a correlation between propensity to spawn and adjustable R&D. That is, firms whose

R&D is more sensitive to its tax price may also be doing the sort of R&D that leads to more spawning. Adjustable R&D may tend to be more general or inventive, and thus more often yield new ideas best suited to development outside the firm. Indeed, we find in our patent-type interactions that firms whose R&D yields more forward citations have higher R&D-induced spawning. It is not obvious why adjustable R&D would be more inventive, but we cannot rule out this possibility.

The second, more plausible explanation is that adjustable R&D is less crucial to the firm. The loss of its idea output to spawns would then be less costly, and these firms would have lower ex-ante incentives to prevent spawning. If the managers making R&D investment decisions are rational and have some information about the expected treatment effect, then costly spawning should lead them to increase R&D less in response to the tax price shock than a firm for which spawning is less costly. If R&D-induced spawning is less costly to compliers with the treatment (the tax shock), we expect the IV estimate to exceed the OLS estimate.

The third possibility is that the IV estimate represents the marginal effect of R&D, which is higher than the average effect. Note that OLS estimates the effect of an additional dollar of average R&D. The IV strategy, which uses additional R&D tax subsidies to approximate increased R&D expenditure on the margin, better captures the effect on spawning of the “last” R&D dollar. This marginal R&D is likely farther from the parent’s core focus than average R&D, which may make it either less costly to lose or harder to protect.

Which of these explanations for the larger IV estimate is true is economically important. In cases where endogeneity biases the OLS result down, our IV better approximates the true effect. In the case where we capture the marginal effect of R&D better in the IV, the IV also better approximates the true effect. However, in the case where the IV isolates those firms whose cost of R&D-induced spawning is especially low, or for which adjustable R&D is otherwise correlated to spawning, then the LATE in the IV is biased upward, and we should assume that OLS yields a better approximation of the true effect. The true economic magnitudes likely lie between the OLS and IV estimates.

4.3 Alternative outcomes

We consider alternative measures of spawning in Table 5. First, columns 1 and 2 use only spawns founded by years $t + 1$ and $t + 2$, respectively (our main variable uses spawns founded in any year through $t + 3$). In both cases, we continue to find positive, significant coefficients. We then examine whether the results are driven by team exits, in which multiple employees leave to found a startup together. This is possible because our definition of “entrepreneur” includes the top five earners at a new firm. In this case, the number of spawns should be less than the number of spawning employees. The dependent variable in column 3 is the number of unique startups spawned from an establishment, based on primary spawning dependent variable. The coefficient is 0.067, indicating that team exits are not explaining the main results.

We turn to a different, “flow” measure of spawning in column 4. This is departed employees who are among the top five earners at a one-year-old spawn in year $t + 1$, at a two-year-old spawn in year $t + 2$, or at a three-year-old spawn in year $t + 3$. That is, we consider cumulative departures. (Our primary dependent variable took a snapshot of the workers in year $t + 3$.) We normalize by the movers, or by the number of workers who, as of the 1st quarter in year 3, have left the firm. The coefficient increases dramatically to 0.272.

An alternative hypothesis to the spawning effect of R&D is that R&D leads to restructuring, or greater turnover in general. Table 6 column 1 shows that R&D in year t has no effect on the percent of employees who remain with the parent by year $t + 3$ (the same time period in which we measure spawning). Similarly, columns 2, 3, and 4 show that R&D has no effect on the percent of employees who move to another incumbent firm, drop out of the labor force, or move to firms whose age is unknown.

4.4 Reverse causation

If our effect is causal – that is, if R&D leads to spawning – then it cannot be the case that spawning predicts R&D. To test this, we project current-year R&D (in year t) on past spawning in Table 7. In column 1, we include four years of spawning, from year $t - 4$ to $t - 1$. In column 2, we include three years ($t - 3$ to $t - 1$), and in column 3, two

years ($t - 2$ to $t - 1$). In all cases, the coefficient is insignificant. This provides strong evidence for causality of our main effect, beyond the instrumental variables approach. It especially obviates the primary endogeneity concern, which is that a technological opportunity jointly causes R&D and spawning. The very nature of a startup is to be adaptable and responsive to new opportunities. We would thus expect startup founding to respond to the new opportunity faster than corporate R&D. In contrast, we find that the entrepreneurial spawning occurs after the R&D.

5 Mechanism

This section considers two not mutually exclusive ways that corporate R&D might increase entrepreneurial spawning. One is intellectual capital, or new ideas generated by R&D, which an employee takes to his new firm. The other is human capital, or entrepreneurial skills that make employees more likely to launch their own ventures.

5.1 Intellectual capital

Our cross-sectional evidence supports the intellectual capital channel, as it satisfies three hypotheses. R&D induced spawning is associated with (a) the “research” part of R&D; (b) new-to-the-world ideas; and (c) R&D generating some ideas that are too far afield for the firm to benefit from.

5.1.1 “Research” part of R&D

First, we expect that the intellectual capital channel will be more associated with the “research” part of R&D, rather than the “development” part. Indeed, we find that high-tech parents, high earner workers, and more innovative firms are more responsible for the spawning effect of R&D. These are less well associated with the commercialization, or development, aspect of R&D.

In Table 8, we interact R&D with a parent firm-level cross-sectional variable. We identify an establishment as “high-tech” using four-digit SIC codes corresponding

to high-tech manufacturing and R&D.¹¹ The effect is 0.083 larger for high-tech parents than non-high-tech parents. The effect for non-high-tech parents (the coefficient on Log R&D) is small and insignificant, indicating that despite having positive R&D, non-high-tech firms do not generate a spawning effect of R&D.

In Table 9, we examine which types of employees depart to found new firms in response to greater parent R&D. Column 1 considers only employees in the top half of the establishment’s wage distribution, while column 2 considers employees in the bottom half. The effects lose significance with this sample split, but the coefficient among high wage employees is larger than for low wage employees. We split employees by tenure at the establishment in columns 3 and 4. The effect appears to be driven more by employees with longer tenure.

We find some evidence that the effect is driven by firms with patents that are more valuable because they are more general purpose, used by a wider array of fields (Hall & Trajtenberg 2004). We interact R&D with an indicator for the firm having above-median patent generality, which means that future cites of its patents are from a wider array of patent classes. The effect is significantly higher for these firms (Table 8 column 4).¹² Table 3 panel 2 column 5 shows that firms that patent in more classes tend to have higher spawning rates. Thus it seems that firms doing more basic and broad research have higher entrepreneurial spawning effects per dollar of R&D.

5.1.2 New-to-the-world ideas

The intellectual capital channel should yield spawns with new-to-the-world ideas, rather than “Main street” type businesses. If R&D stimulates restaurants or plumbing companies, it seems unlikely that ideas and inventions created by the R&D investment are the mechanism. We find that within the population of spawns, more parent R&D is associated with high-tech and VC-backed spawns. We examine in Table 10 whether parent R&D is associated with certain spawn characteristics.

The dependent variable in Table 10 panel 1 column 1 is one if the spawn receives venture capital backing, which is the case for only 2 percent of spawns. Strikingly in

¹¹These are 3200-3299, 3500-3599, 3700-3899, and 8732-8734.

¹²We do not find significant differences using the number of patents, number of citations, or patent originality.

light of this low incidence rate, the coefficient on R&D is 0.007, significant at the .01 level. Though only a tiny fraction of new businesses, VC-backed startups are widely thought to be strongly associated with new-to-the-world ideas. Gornall & Strebulaev (2015) show that among U.S. public companies, those with venture capital are responsible for 44 percent of research and development expenditure, and Kaplan & Lerner (2010) show that over 60 percent of IPO issuers have venture capital backing. While there are other firm variables that are weakly associated with spawning, such as investment, R&D is the only variable that predicts VC-backed spawns.

In panel 2 column 3, the dependent variable is one if the spawn is in a high-tech industry, and zero if it is not. Parent R&D is strongly associated with spawns being high-tech, consistent with the R&D-induced spawns being driven by new ideas. In column 4, we show that R&D induces spawns with higher wages than the average spawn. In unreported analysis, we do not find that R&D induces spawns with more initial employees. Thus R&D seems to induce spawns with high-skill labor, but that do not start at a larger than average size.

We finally turn to exit (IPO or acquisition) rates. Having shown that R&D induces spawns with higher wages, we are interested in exit because it serves as a proxy for high-impact, high-risk startups. In column 4, the dependent variable is one if the startup exits within five years (starting from year $t + 3$, where t is the year in which we measure R&D). We find a positive, significant effect of R&D. In sum, this offers evidence that relative to the average spawn, R&D-induced spawns tend to be in the far right-tail in terms of growth potential.

5.1.3 R&D generates some ideas that are too far afield for the firm to benefit from

We expect that the intellectual capital channel reflects R&D generating many new ideas, some of which are far from the parent’s ken, and are easily appropriated by an employee. If the spawns are simply replicating the parents’ business models, then they likely do not have a new idea. Instead, we find that spawns tend to be in different industries from parents, and more parent R&D makes it less likely that the spawn is in same industry as the parent. This suggests that the spawn has a new idea and is

not replicating the parent’s business.

In column 1 of Table 10, the dependent variable is one if the spawn is in the same 2-digit SIC classification as its parent (examples of 2-digit industries are “Business Services” and “Coal Mining”). The coefficient is negative and significant; more parent R&D reduces the chances that a spawn is in the same industry as its parent. Only 16.8 percent of spawns are in their parent’s 2-digit industry.

Thus, our effect is driven by high-tech parents and high-tech spawns, while the R&D-induced spawns tend to be in different industries from their parents. The intellectual capital channel can reconcile these facts. The parent firm R&D creates growth options far from its core focus, which the employee can deploy in a new firm. The intellectual capital mechanism also fits well with our interpretation of the IV results. The ideas generated by R&D that wind up in departing employees’ startups are much more likely to come from the last dollar of R&D than the first. In this light, the IV strategy yields an effect that isolates the driving mechanism: marginal R&D generates ideas, some of which spill over into startups founded by employees.

5.2 Human capital

In the human capital channel, R&D induces employee learning, which makes the employee more productive as an entrepreneur. In this channel, R&D leads to spawning not because of new ideas that it generates, but rather because of new skills that it generates. We cannot rule out that this channel plays a role, but three pieces of cross sectional evidence are inconsistent with it.

First, in a human capital channel, we would expect R&D-induced spawns to come from small parents. This is because small firm employees tend to have a broader scope of work (Stuart & Ding 2006, Sørensen 2007). Instead, we find that large firms drive the effect (Table 8 columns 2 and 3). These are defined as having above-median total assets within a given year. This finding contrasts with Elfenbein et al. (2010), who find using survey data on scientists that entrepreneurs are more likely to be spawned from small firms. Second, we might also expect that there is more opportunity for entrepreneurial learning at young firms. However, when we interact R&D with an indicator for being young (below median age), we find no effect.

Third, we would expect that capital expenditure would have a similar effect on spawning if the channel were skills, because new capital investment seems likely to create similar project management skills as R&D projects. Instead, while capital expenditure has a positive effect on spawning, it is far smaller on a per-dollar basis than the effect of R&D, and is not robust to many of our standard specifications. In sum, our evidence strongly supports the intellectual capital channel and is inconsistent with the human capital channel.

5.3 Control rights and spillovers

The intellectual capital channel implies that the spawning effect of R&D is a new avenue for knowledge spillovers. While knowledge spillovers have long been known to exist, it has been challenging for the literature to identify the channel of transmission (Jaffe et al. 1993, Greenstone et al. 2010). For the spawning effect of R&D to be a knowledge spillover, the cost to the parent of the R&D-induced spawn cannot outweigh the spawns's combined private and social value. Its private value is to the entrepreneur and other equity holders. Its social value comes from new jobs created or unpriced benefits from commercializing a new idea.

A strong argument for the effect being a spillover is that R&D induces high-growth new firms. Young firms are more productive and grow faster than incumbents, as shown by a voluminous literature including Acs & Audretsch (1990), Kortum & Lerner (2000), Akcigit & Kerr (2010) Decker et al. (2014), and Glaeser et al. (2015). These large social benefits imply that even in the presence of costs to the parent firm, the effect is likely still be a spillover.

The only ways that control rights may matter for the effect being a positive externality is if either: 1) The spawn is very costly to the parent; 2) The parent fully internalizes the spawn's benefits, implying no ex-ante underinvestment relative to the social optimum. In the subsections below, we consider each of these in turn. Note, however, that the ex-post split of the surplus from new idea split between the parent and spawn is not relevant from a social welfare perspective. Strictly speaking, to be a positive externality it is only necessary that the spawning effect of R&D implies greater ex-ante underinvestment in R&D relative to the social optimum, which it almost

certainly does.

5.3.1 Is spawning costly to the parent?

We find no evidence of large costs to parent. One test comes from non-compete covenants, which restrict employees from working for a competing firm within the state for a specified period of time. Literature has found that non-compete enforcement reduces local knowledge spillovers (Belenzon & Schankerman 2013, Matray 2015), and reduce within-state inventor mobility (Marx et al. 2015). In light of this literature, we expect that if the R&D effect on spawning is very costly to the parent, it should be attenuated in states that enforce non-competes. Instead, the main result persists in states that enforce non-competes, and there is no significant effect on an interaction between R&D and an indicator for being in a weak enforcement state.

Second, if R&D-induced spawning were costly to parent, it should be less feasible in sectors where intellectual property is easier to protect. We did not find that the effect varied with a measure of industry “patentability”. Third, costly spawns may compete in product markets with the parent. Instead, we found that spawns tend to be in different industries from parents. This helps explain why the parent does not value the idea enough to keep it in-house. Finally, spawning may be costly if the parent cannot find a replacement for the lost employee. While we cannot test this directly, in unreported tests we do not find that R&D predicts establishment employment growth or labor productivity, measured as payroll divided by employment.

Finally, we make a revealed preference argument. By virtue of observing the persistent phenomenon of R&D-induced spawning, the parent either chose not to develop the idea in house, or chose not to take ex-ante steps to prevent the spawn. These steps could include increasing the employee’s compensation to retain him, or even not conducting R&D at all.

It is possible that the parent does not possess the option to prevent the spawn. For example, the employee may fear expropriation and not disclose ideas, or he may be able to steal an idea that the firm deems valuable.¹³ In the case of such contract-

¹³The low costs of information and resource sharing (including teamwork) are reasons the firm exists in the first place. Giving employees the right incentives to innovate – that is, high-powered incentives – would make it impossible to manage the larger R&D process. For example, the firm will find it

ing frictions, the parent firm should predict the loss of some innovative employees to spawning. It might price this cost into their compensation ex-ante. Regardless of these considerations, by virtue of observing spawning, any costs of preventing it must exceed the benefits.

5.3.2 Do parents internalize the benefits of the spawn?

The spawning effect of R&D would be less obviously a new source of R&D spillovers if the parent internalizes, or appropriates, the spawn’s private benefits. This might occur if the parent invests in or acquires the spawn. Two pieces of cross-sectional evidence make this unlikely. First, firms with corporate venture capital programs are not more responsible for spawning. Second, we expect parent-supported spinoffs to start at a larger scale than a typical bootstrapped startup. Conversely, we find no relation between initial spawn size and parent R&D.

To provide more concrete evidence, we directly assess the possibility that parents internalize spawns’ benefits using an out-of-sample test based on the underlying data in Gompers et al. (2005). They connected all VC-backed startup executives in the VentureOne database between 1986 and 1999 to their prior employers.¹⁴ We hypothesize that this data should provide an upper bound on possible internalized spawning; since these spawns by definition received external investment, they are more likely than the average spawn to have received investment from their former employer.

We begin with 13,612 entrepreneur-parent pairs. The entrepreneurs are founders of 6,499 unique spawns. There are 9,152 unique parents. In most cases spawns have multiple parents (that is, there are multiple executives with prior jobs). We linked all of the spawn parents to VentureXpert acquisition and investment data. We successfully matched 4,786 unique spawns to at least one investor or acquirer, a match rate of 74 percent. There are 20,478 unique spawn-investor pairs.¹⁵

difficult to figure out ex-post exactly who is responsible for the innovation, and individuals will have incentives to hoard information. Note that the contracting challenges arise in large part from the inalienability of human labor (no slavery). This relates to the property rights literature associated with Grossman and Hart (1986).

¹⁴This time period overlaps with our primary Census data (1990 to 2006).

¹⁵Note that the underlying dataset, from Dow Jones Venture Source, is of VC-backed startups. In theory, if we used VentureSource, we should match 100 percent to initial investors. However, as Kaplan & Lerner (2016) and Maats et al. (2011) explain, VentureXpert’s coverage is much better than Venture

A merge of these investors and acquirers to the parents yields 266 unique spawns where the parent matches an investor or acquirer, out of 4,786 spawns that we matched to VentureXpert, or 5.6 percent.¹⁶ Of these, 192 are investment deals, and 74 are acquisitions. There are 208 unique parents that are matched to investors/acquirers. Note that some parents have multiple spawns, such as IBM and Highland Capital Partners, so the parent and spawn numbers do not match. Some parents that invested in or acquired their spawns are corporates, including Seagate, Xerox, Monsanto, Johnson & Johnson, and Microsoft. Others are asset managers, including Accel Partners, Softbank, and Equus Capital. Still others are non-corporates, including Boston University. We identified 41 spawning parents that are clearly venture funds or other asset managers. This leaves 167 parents that are plausibly corporates, though this is generous as we retained financial services companies such as Goldman Sachs.

To interpret this exercise, we return to the total parent population. Of the 9,152 unique parents in the original Gompers et al. (2005) data, just 2.3 percent (208) invest in or acquire their spawns. This small percentage is evidence that parents do not usually internalize spawns by investing in or acquiring them. One concern may be that perhaps many corporate parents are not covered as investors or acquirers in VentureXpert. We can match 2,617 of the parents to investors or acquirers in VentureXpert. The most conservative framing of our results, then, restricts the parent population to firms that ever invested in or acquired a startup in VentureXpert. In this case, 7.9 percent of parents (208 out of 2,617) invest in or acquire their spawns. This extreme upper bound is still small, and confirms that it is unlikely that parents generally internalize the benefits of their spawns.

The parent could also appropriate the spawn’s benefits through technology licensing deals. We cannot assess this possibility with our data, but we think it unlikely

Source (more than 40 percent more investments). VentureXpert also has superior acquisition data, and Venture Source’s data quality has declined over time. We are most interested in whether parents ultimately invested in (and especially acquired) spawns, so VentureXpert seems like the optimal data set to use. If there is any bias, it should be the case that the spawns that do not match have lower rates of subsequent investment and acquisition, since the commercial databases often backfill based on exit events.

¹⁶We matched on the company’s first word, which yielded 275 matches. This enables successful matches such as “Xerox Venture Capital” to “Xerox.” We then manually removed obviously wrong matches, erring on the side of leaving the match to be conservative in ambiguous cases.

that the parent can fully internalize the spawn’s social benefits through such arms-length contracts.

Thus parents tend not to invest in or acquire their spawns; consistent with this out-of-sample test, within our data we find no interaction effect on spawning between R&D and the parent having a corporate venture capital program. These results are consistent with Ma (2016), who finds that public firms launch corporate venture capital (CVC) programs when internal innovation is poor, invest in startups in their own industries, and invest in geographically distant startups. That is, CVCs are a way to outsource innovation. This is the opposite of the corporate environment that yields R&D-induced spawning. Instead, when corporate R&D increases at innovative firms, it seems to serendipitously produce “extra” growth options, and spawning is an unintended consequence.

In sum, it appears likely that R&D-induced spawning is a direct form of knowledge spillover. A remarkable 88 percent of spawns are located in the same state as the parent. The literature on knowledge spillovers has focused on their ability to explain industrial agglomeration, or the spatial concentration of firms. Knowledge spillovers have been found to be quite local and to decline with distance (e.g. Jaffe et al. 1993, Belenzon & Schankerman 2013, Kantor & Whalley 2014).

Our result offers another channel for the link between industrial agglomeration and knowledge spillovers, which is often attributed in part to the importance of tacit information (Audretsch & Feldman 1996, Glaeser 1999, Duranton & Puga 2001). Since spawns tend not to be in the same industry as their parents, our data suggest another reason for the connection between spillovers and agglomeration, more along the lines found in Ellison et al. (2010): moving may be privately costly to the spawned entrepreneur, or he may have relevant networks in the location of his former firm.

6 Conclusion

This paper shows that corporate R&D investment leads to entrepreneurial spawning, in which employees depart to found their own firms. Thus corporate R&D is both a

new channel for R&D spillovers, and is also a new source of high-tech startups. We find that there are no obvious contractual benefits or observable costs to the firm of this phenomenon. This implies that the R&D effect of spawning is likely a positive externality, in which case there is greater corporate underinvestment in R&D relative to the social optimum than previously thought.

Our finding also speaks to a core aspect of the theory of the firm. Zingales (2000) writes that

“Entrepreneurship is the process by which new firms are created. But new firms are created to exploit growth options existing firms cannot or do not want to exploit. Thus, a theory able to explain what growth options existing firms are willing and able to exploit will also identify the opportunities for entrepreneurial activity.”

While we do not offer such a theory, we offer a concrete mechanism tying entrepreneurship to the growth options that an incumbent does not to exploit.

References

- Abowd, J. M., Stephens, B. E., Villhuber, L., Andersson, F., McKinney, K. L., Roemer, M. & Woodcock, S. (2009), The lehd infrastructure files and the creation of the quarterly workforce indicators, *in* ‘Producer dynamics: New evidence from micro data’, University of Chicago Press, pp. 149–230.
- Acemoglu, D. (2011), ‘Technological change, lecture 3, knowledge spillovers and diffusion’. MIT Lecture Notes.
URL: <https://economics.mit.edu/files/6963>
- Acemoglu, D., Akcigit, U., Bloom, N. & Kerr, W. (2013), ‘Innovation, reallocation and growth’, *NBER Working Paper No. 18993*.
- Acs, Z. J. & Audretsch, D. B. (1990), *Innovation and Small Firms*, MIT Press, Cambridge, MA.
- Adelino, M., Ma, S. & Robinson, D. (2017), ‘Firm age, investment opportunities, and job creation’, *The Journal of Finance*.
- Aghion, P. & Jaravel, X. (2015), ‘Knowledge spillovers, innovation and growth’, *The Economic Journal* **125**(583), 533–573.
- Akcigit, U. & Kerr, W. (2010), ‘Growth through heterogeneous innovations’, *NBER Working Paper 16443*.
- Almeida, P. & Kogut, B. (1999), ‘Localization of knowledge and the mobility of engineers in regional networks’, *Management science* **45**(7), 905–917.
- Angrist, J. D. & Imbens, G. W. (1995), ‘Two-stage least squares estimation of average causal effects in models with variable treatment intensity’, *Journal of the American statistical Association* **90**(430), 431–442.
- Audretsch, D. B. & Feldman, M. P. (1996), ‘R&d spillovers and the geography of innovation and production’, *The American economic review* **86**(3), 630–640.
- Azoulay, P., Zivin, J. G. & Wang, J. (2010), ‘Superstar extinction’, *Quarterly Journal of Economics* **25**.
- Babina, T. (2015), ‘Destructive creation at work: How financial distress spurs entrepreneurship’. Working Paper.
- Babina, T., Ouimet, P. P. & Zarutskie, R. (2015), ‘Going entrepreneurial? ipos and new firm creation’. Working Paper.
- Bankman, J. & Gilson, R. J. (1999), ‘Why start-ups?’, *Stanford Law Review* **51**, 289, 293–95.

- Belenzon, S. & Schankerman, M. (2013), ‘Spreading the word: Geography, policy, and knowledge spillovers’, *Review of Economics and Statistics* **95**(3), 884–903.
- Bernstein, J. I. & Nadiri, M. I. (1989), ‘Research and development and intra-industry spillovers: An empirical application of dynamic duality’, *Review of Economic Studies* **56**(186), 249–268.
- Bernstein, S. (2015), ‘Does going public affect innovation?’, *The Journal of Finance* **70**(4), 1365–1403.
- Bhide, A. V. (2000), *The Origin and Evolution of New Businesses.*, Oxford University Press.
- Billings, A., Glazunov, S. & Houston, M. (2001), ‘The role of taxes in corporate research and development spending’, *R&D Management* **31**(4), 465–477.
- Bloom, N., Griffith, R. & Van Reenen, J. (2002), ‘Do r&d tax credits work? evidence from a panel of countries 1979–1997’, *Journal of Public Economics* **85**(1), 1–31.
- Bloom, N., Schankerman, M. & Van Reenen, J. (2013), ‘Identifying technology spillovers and product market rivalry’, *Econometrica* **81**(4), 1347–1393.
- BLS (1997), *BLS Handbook of Methods*, Bureau of Labor Statistics, Washington DC. U.S. Bureau of Labor Statistics, Division of Information Services, Washington D.C.
- Brown, J. R. & Petersen, B. C. (2011), ‘Cash holdings and r&d smoothing’, *Journal of Corporate Finance* **17**(3), 694–709.
- Campbell, B. A., Ganco, M., Franco, A. M. & Agarwal, R. (2012), ‘Who leaves, where to, and why worry? employee mobility, entrepreneurship and effects on source firm performance’, *Strategic Management Journal* **33**(1), 65–87.
- Chatterji, A. K. (2009), ‘Spawned with a silver spoon? entrepreneurial performance and innovation in the medical device industry’, *Strategic Management Journal* **30**(2), 185–206.
- Chirinko, R. S. & Wilson, D. J. (2008), ‘State investment tax incentives: A zero-sum game?’, *Journal of Public Economics* **92**(12), 2362–2384.
- Chirinko, R. S. & Wilson, D. J. (2011), ‘Tax competition among us states: racing to the bottom or riding on a seesaw?’, *Working Paper* .
- Clausen, T. H. (2009), ‘Do subsidies have positive impacts on r&d and innovation activities at the firm level?’, *Structural Change and Economic Dynamics* **20**(4), 239–253.
- Cohen, W. M. & Levinthal, D. A. (1989), ‘Innovation and learning: the two faces of r & d’, *The economic journal* **99**(397), 569–596.
- Decker, R., Haltiwanger, J., Jarmin, R. & Miranda, J. (2014), ‘The role of entrepreneurship in us job creation and economic dynamism’, *The Journal of Economic Perspectives* **28**(3), 3–24.

- Duranton, G. & Puga, D. (2001), ‘Nursery cities: Urban diversity, process innovation, and the life cycle of products’, *American Economic Review* pp. 1454–1477.
- Elfenbein, D. W., Hamilton, B. H. & Zenger, T. R. (2010), ‘The small firm effect and the entrepreneurial spawning of scientists and engineers’, *Management Science* .
- Ellison, G., Glaeser, E. L. & Kerr, W. R. (2010), ‘What causes industry agglomeration? evidence from coagglomeration patterns’, *The American Economic Review* **100**(3), 1195–1213.
- Glaeser, E. L. (1999), ‘Learning in cities’, *Journal of urban Economics* **46**(2), 254–277.
- Glaeser, E. L., Kerr, S. P. & Kerr, W. R. (2015), ‘Entrepreneurship and urban growth: An empirical assessment with historical mines’, *Review of Economics and Statistics* **97**(2), 498–520.
- Goldin, C., Kerr, S. P., Olivetti, C., Barth, E. et al. (2017), ‘The expanding gender earnings gap: evidence from the lehd-2000 census’, *American Economic Review* **107**(5), 110–14.
- Gompers, P., Lerner, J. & Scharfstein, D. (2005), ‘Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999’, *The Journal of Finance* **60**(2), 577–614.
- Gornall, W. & Strebulaev, I. A. (2015), ‘The economic impact of venture capital: Evidence from public companies’, *Working Paper* .
- Greenstone, M., Hornbeck, R. & Moretti, E. (2010), ‘Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings’, *Journal of Political Economy* **118**(3), 536–598.
- Griffith, R., Harrison, R. & Van Reenen, J. (2006), ‘How special is the special relationship? using the impact of us r&d spillovers on uk firms as a test of technology sourcing’, *American Economic Review* **96**(5), 1859–1875.
- Griliches, Z. (1992), ‘The search for r&d spillovers’, *The Scandinavian Journal of Economics* pp. S29–S47.
- Gromb, D. & Scharfstein, D. (2002), Entrepreneurship in equilibrium, Technical report, National bureau of economic research.
- Guenther, G. (2015), ‘Research tax credit: Current law and policy issues for the 114th congress’. Congressional Research Service Report 7-5700.
- Guzman, J. & Stern, S. (2017), *Nowcasting and Placecasting Entrepreneurial Quality and Performance*, Vol. 75, University of Chicago Press.
- Hall, B. H. (1993), ‘R&d tax policy during the 1980s: success or failure?’, *Tax policy and the economy* **7**, 1–35.
- Hall, B. H. & Trajtenberg, M. (2004), Uncovering gpts with patent data, Technical report, National Bureau of Economic Research.

- Hall, B. & Van Reenen, J. (2000), ‘How effective are fiscal incentives for r&d? a review of the evidence’, *Research Policy* **29**(4), 449–469.
- Haltiwanger, J., Jarmin, R. S. & Miranda, J. (2013), ‘Who creates jobs? small versus large versus young’, *Review of Economics and Statistics* **95**(2), 347–361.
- Hellmann, T. (2007), ‘When do employees become entrepreneurs?’, *Management science* **53**(6), 919–933.
- Herkenhoff, K., Lise, J., Menzio, G. & Phillips, G. (2018), ‘Worker mobility and the diffusion of knowledge’.
- Jaffe, A. B., Trajtenberg, M. & Henderson, R. (1993), ‘Geographic localization of knowledge spillovers as evidenced by patent citations’, *the Quarterly journal of Economics* pp. 577–598.
- Jarmin, R. S. & Miranda, J. (2002), ‘The longitudinal business database’.
- Jiang, W. (2015), ‘Have instrumental variables brought us closer to the truth’, *The Review of Corporate Finance Studies* .
- Jones, C. I. & Williams, J. C. (1998), ‘Measuring the social return to r&d’, *Quarterly Journal of Economics* pp. 1119–1135.
- Kantor, S. & Whalley, A. (2014), ‘Knowledge spillovers from research universities: evidence from endowment value shocks’, *Review of Economics and Statistics* **96**(1), 171–188.
- Kaplan, S. N. & Lerner, J. (2010), ‘It ain’t broke: The past, present, and future of venture capital’, *Journal of Applied Corporate Finance* **22**(2), 36–47.
- Kaplan, S. N. & Lerner, J. (2016), Venture capital data: Opportunities and challenges, in ‘Measuring Entrepreneurial Businesses: Current Knowledge and Challenges’, University of Chicago Press.
- Kerr, S., Kerr, W. R. & Nanda, R. (2015), House money and entrepreneurship, Technical report, National Bureau of Economic Research.
- Kim, J. & Marschke, G. (2005), ‘Labor mobility of scientists, technological diffusion, and the firm’s patenting decision’, *RAND Journal of Economics* pp. 298–317.
- Klassen, K. J., Pittman, J. A., Reed, M. P. & Fortin, S. (2004), ‘A cross-national comparison of r&d expenditure decisions: Tax incentives and financial constraints’, *Contemporary Accounting Research* **21**(3), 639–680.
- Klepper, S. (2001), ‘Employee startups in high-tech industries’, *Industrial and corporate change* **10**(3), 639–674.
- Kortum, S. & Lerner, J. (2000), ‘Assessing the contribution of venture capital to innovation’, *The Rand Journal of Economics* **31**(4), 674.
- Krugman, P. R. (1991), ‘Geography and trade’.

- Lerner, J. & Seru, A. (2017), The use and misuse of patent data: Issues for corporate finance and beyond, Technical report, National Bureau of Economic Research.
- Ma, S. (2016), ‘The life cycle of corporate venture capital’. Working Paper.
- Maats, F., Metrick, A., Yasuda, A., Hinkes, B. & Vershovski, S. (2011), ‘On the consistency and reliability of venture capital databases’, *Unpublished working paper*.
- Mamuneas, T. P. & Nadiri, M. I. (1996), ‘Public r&d policies and cost behavior of the us manufacturing industries’, *Journal of Public Economics* **63**(1), 57–81.
- Manski, C. F. (1993), ‘Identification of endogenous social effects: The reflection problem’, *The Review of Economic Studies* **60**(3), 531–542.
- Marshall, A. (1920), *Principles of economics*, MacMillan, London.
- Marx, M., Singh, J. & Fleming, L. (2015), ‘Regional disadvantage? employee non-compete agreements and brain drain’, *Research Policy* **44**(2), 394–404.
- Matray, A. (2015), ‘The local innovation spillovers of listed firms’, *Working Paper*.
- McCutchen, W. W. (1993), ‘Estimating the impact of the r&d tax credit on strategic groups in the pharmaceutical industry’, *Research Policy* **22**(4), 337–351.
- Metrick, A. & Yasuda, A. (2011), ‘Venture capital and other private equity: a survey’, *European Financial Management* **17**(4), 619–654.
- Miller, C. R. & Richard, B. (2010), ‘The policy diffusion of the state r&d investment tax credit’, *State and Local Government Review* **42**(1), 22–35.
- Nanda, R. & Sørensen, J. B. (2010), ‘Workplace peers and entrepreneurship’, *Management Science* **56**(7), 1116–1126.
- Paff, L. A. (2005), ‘State-level r&d tax credits: A firm-level analysis’, *The BE Journal of Economic Analysis & Policy* **5**(1).
- Puri, M. & Zarutskie, R. (2012), ‘On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms’, *The Journal of Finance* **67**(6), 2247–2293.
- Robinson, D. T. (2008), ‘Strategic alliances and the boundaries of the firm’, *Review of Financial Studies* **21**(2), 649–681.
- Seru, A. (2014), ‘Firm boundaries matter: Evidence from conglomerates and r&d activity’, *Journal of Financial Economics* **111**(2), 381–405.
- Sevilir, M. (2010), ‘Human capital investment, new firm creation and venture capital’, *Journal of Financial Intermediation* **19**(4), 483–508.
- Sørensen, J. B. (2007), ‘Bureaucracy and entrepreneurship: Workplace effects on entrepreneurial entry’, *Administrative Science Quarterly* **52**(3), 387–412.

- Stuart, T. E. & Ding, W. W. (2006), ‘When do scientists become entrepreneurs? the social structural antecedents of commercial activity in the academic life sciences’, *American Journal of Sociology* **112**(1), 97–144.
- Tate, G. & Yang, L. (2015), ‘The bright side of corporate diversification: Evidence from internal labor markets’, *The review of financial studies* **28**(8), 2203–2249.
- Waldinger, F. (2012), ‘Peer effects in science: Evidence from the dismissal of scientists in nazi germany’, *The Review of Economic Studies* **79**(2), 838–861.
- Wilson, D. J. (2009), ‘Beggar thy neighbor? the in-state, out-of-state, and aggregate effects of r&d tax credits’, *The Review of Economics and Statistics* .
- Wilson, D. J. et al. (2005), ‘The rise and spread of state r&d tax credits’, *FRBSF Economic Letter* .
- Wu, Y. (2008), ‘State r&d tax credits and high-technology establishments’, *Economic Development Quarterly* **22**(2), 136–148.
- Zingales, L. (2000), ‘In search of new foundations’, *The journal of Finance* **55**(4), 1623–1653.

Table 1: Summary Statistics

Panel 1: Firm-year level indicator variables

Sample:	All	Low Δ R&D $_{t-1}$	High Δ R&D $_{t-1}$	P-value for difference [(2)-(3)]
	Mean (St Dev)	Mean (St Dev)	Mean (St Dev)	Mean (St Dev)
	(1)	(2)	(3)	(4)

Δ R&D $_{t-1}$	0.492 (0.500)			
Made corporate VC investments $_t$	0.038 (0.192)	0.040 (0.195)	0.037 (0.189)	0.559
Had ≥ 1 patent $_{t-10,t}$	0.601 (0.490)	0.615 (0.487)	0.586 (0.493)	0.009
Diversified	0.789 (0.408)	0.822 (0.382)	0.755 (0.430)	0.000

Panel 2: Firm-year level continuous variables

	All	Low Δ R&D _{<i>t</i>-1}	High Δ R&D _{<i>t</i>-1}	
	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	P-value for difference [(2)-(3)]
	(1)	(2)	(3)	(4)
R&D/Total Assets _{<i>t</i>-1}	0.085 0.052 (0.102)	0.067 0.040 (0.084)	0.103 0.067 (0.115)	0.000
Log R&D _{<i>t</i>-1}	2.53 2.45 (2.25)	2.54 2.43 (2.30)	2.51 2.46 (2.20)	0.615
Tobin's Q _{<i>t</i>-1}	2.12 1.65 (1.59)	1.87 1.50 (1.33)	2.37 1.84 (1.78)	0.000
Age _{<i>t</i>}	20.03 21.03 (6.18)	20.64 21.55 (5.76)	19.40 20.49 (6.53)	0.000
Number of establishments	3.20 2.36 (4.05)	3.41 2.64 (4.24)	2.98 2.34 (3.84)	0.000
Number of industries (SIC3)	1.54 1.99 (1.01)	1.59 1.99 (1.04)	1.49 1.99 (0.97)	0.000
Number of states	2.74 2.35 (3.02)	2.91 2.36 (3.15)	2.57 2.34 (2.87)	0.000
Total Assets _{<i>t</i>-1}	3,483 529 (12,630)	4,291 670 (14,560)	2,650 412 (10,190)	0.000
Employment _{<i>t</i>-1}	835 326 (1,603)	889 354 (1,662)	780 299 (1,537)	0.015

<i>Panel 3: Establishment-year level indicator variables</i>			
Sample:	All	Low Δ R&D _{<i>t</i>-1}	High Δ R&D _{<i>t</i>-1}
	Mean	Mean	Mean
	(1)	(2)	(3)
Δ R&D _{<i>t</i>-1}	0.458		
State has weak non-compete enforcement	0.613	0.614	0.611
In high-tech industry	0.641	0.621	0.664
Made corporate VC investments _{<i>t</i>}	0.079	0.083	0.074
Had ≥ 1 patent _{<i>t</i>-10,<i>t</i>}	0.709	0.711	0.706

Panel 4: Establishment-year level continuous variables

	All	Low Δ R&D _{<i>t</i>-1}	High Δ R&D _{<i>t</i>-1}	
	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	P-value for difference [(2)-(3)]
	(1)	(2)	(3)	(4)
Entrepreneurial spawning rate _{<i>t</i>+3}	1.31 0.82 (2.43)	1.24 0.80 (2.29)	1.39 0.85 (2.59)	0.000
# spawned firms _{<i>t</i>+3}	1.15 0.78 (1.91)	1.10 0.76 (1.81)	1.21 0.80 (2.02)	0.000
Stayers _{<i>t</i>+1}	47.77 52.30 (25.98)	47.72 52.21 (26.17)	47.83 52.39 (25.76)	0.886
Movers to old firms _{<i>t</i>+3}	26.29 22.51 (18.10)	26.60 22.78 (18.35)	25.93 22.28 (17.80)	0.221
Unemployed _{<i>t</i>+3}	12.39 11.11 (7.78)	12.46 11.19 (7.79)	12.30 11.05 (7.78)	0.245
Movers to firms of unknown age _{<i>t</i>+3}	9.73 6.65 (12.28)	9.47 6.52 (12.03)	10.05 6.79 (12.56)	0.037
Log R&D _{<i>t</i>-1}	3.70 3.67 (2.30)	3.72 3.70 (2.33)	3.67 3.62 (2.27)	0.656
Average worker wage _{<i>t</i>}	17.53 15.50 (10.56)	17.01 15.21 (10.06)	18.13 15.86 (11.09)	0.002
Average worker age _{<i>t</i>}	40.08 40.27 (4.76)	40.25 40.52 (4.82)	39.89 39.98 (4.68)	0.032

...Continued

<i>...Continued</i>				
	All	Low Δ R&D _{<i>t</i>-1}	High Δ R&D _{<i>t</i>-1}	
	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	P-value for difference [(2)-(3)]
	(1)	(2)	(3)	(4)
Average employee tenure _{<i>t</i>}	2.69	2.73	2.66	0.092
	2.40	2.43	2.37	
	(1.88)	(1.91)	(1.85)	
Δ R&D _{<i>t</i>-1}	0.088	-0.098	0.307	0.000
	0.071	-0.034	0.213	
	(0.342)	(0.238)	(0.314)	
Federal R&D tax price	-0.047	-0.042	-0.053	0.000
	-0.034	-0.025	-0.044	
	(0.050)	(0.050)	(0.049)	
State R&D tax price	0.170	0.172	0.167	0.000
	0.170	0.172	0.168	
	(0.033)	(0.033)	(0.033)	
Share employees female _{<i>t</i>}	0.333	0.326	0.341	0.020
	0.313	0.304	0.323	
	(0.192)	(0.192)	(0.192)	
Share employees white _{<i>t</i>}	0.795	0.796	0.793	0.404
	0.835	0.837	0.832	
	(0.171)	(0.172)	(0.170)	
Share employees foreign _{<i>t</i>}	0.062	0.057	0.067	0.000
	0.031	0.028	0.034	
	(0.098)	(0.095)	(0.101)	
Number employees _{<i>t</i>}	329	344	311	0.268
	122	123	120	
	(1,698)	(1,872)	(1,465)	

Panel 5: Spawn level indicator variables

Sample:	All	Low Δ R&D _{<i>t</i>-1}	High Δ R&D _{<i>t</i>-1}	P-value for difference [(2)-(3)]
	Mean (St Dev)	Mean (St Dev)	Mean (St Dev)	Mean (St Dev)
	(1)	(2)	(3)	(4)
Above median Δ R&D _{<i>t</i>-1}	0.442 (0.497)			
Same industry (SIC2) as parent	0.168 (0.374)	0.158 (0.365)	0.182 (0.386)	0.039
Same state as parent	0.876 (0.330)	0.877 (0.328)	0.874 (0.332)	0.302
High-tech industry	0.494 (0.500)	0.477 (0.499)	0.514 (0.500)	0.042
Ever received VC	0.020 (0.139)	0.015 (0.122)	0.025 (0.157)	0.006
Exit _{<i>t</i>+5}	0.525 (0.499)	0.516 (0.500)	0.536 (0.499)	0.031
Employee female	0.331 (0.471)	0.339 (0.473)	0.322 (0.467)	0.268
Employee white	0.799 (0.401)	0.795 (0.404)	0.804 (0.397)	0.447
Employee foreign	0.077 (0.267)	0.072 (0.259)	0.084 (0.277)	0.071
Employee born in state	0.475 (0.499)	0.488 (0.500)	0.459 (0.498)	0.017

Panel 6: Spawn level continuous variables

	All	Low Δ R&D _{t-1}	High Δ R&D _{t-1}	
	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	Mean Quasi-med (St Dev)	P-value for difference [(2)-(3)]
	(1)	(2)	(3)	(4)
Spawn employment _{t+3}	14.59 5.92 (39.48)	14.15 5.67 (38.74)	15.15 5.94 (40.39)	0.162
Spawn age _{t+3}	1.59 1.99 (1.01)	1.59 1.99 (1.01)	1.59 1.99 (1.01)	0.919
Spawn payroll _{t+3}	511 126 (1,603)	475 118 (1,554)	556 137 (1,662)	0.027
Employee age _t	35.16 34.64 (10.94)	35.07 34.64 (11.11)	35.27 34.71 (10.71)	0.794
Employee education	13.66 14.36 (2.49)	13.60 14.34 (2.48)	13.74 14.39 (2.50)	0.236
Employee tenure _t	2.07 1.58 (2.25)	1.99 1.30 (2.22)	2.17 1.59 (2.28)	0.189
Employee's log wages (at parent firm) _t	57.80 39.12 (71.70)	53.02 35.07 (67.49)	63.85 44.31 (76.27)	0.072
Employee's log wages (at spawn) _{t+3}	51.84 33.60 (60.99)	48.56 31.53 (58.18)	55.99 36.55 (64.12)	0.022
Parent firm log R&D _{t-1}	377 70 (885)	376 55 (965)	379 90 (772)	0.975

Note: In panels 1-2 (firm-year), all variables have 10,500 observations. In panels 3-4 (establishment-year), all variables have 36,000 observations. In panels 5-6 (spawn), all variables have 108,000 observations. Change (Δ) in R&D_{t-1} is defined as: $\frac{R\&D_{t-1} - R\&D_{t-2}}{.5 \cdot (R\&D_{t-1} + R\&D_{t-2})}$. Low Δ R&D_{t-1} is below-median, and high Δ R&D_{t-1} is above median. In panel 2, number of industries and number of states are within the 31 states covered by our LEHD data. In panels 5 and 6, “employee” variables refer to the spawning employee who (who left for the new firm).

Table 2: First Stage Results

Dependent variable: Log R&D _{t-1}						
	(1)	(2)	(3)	(4)	(5)	(6)
Federal R&D tax price	-2.020*** (0.295)	-1.504*** (0.231)	-1.504*** (0.231)	-1.470*** (0.225)	-1.363*** (0.168)	-1.424*** (0.199)
State R&D tax price	-1.158* (0.691)	-0.950** (0.476)	-0.956** (0.476)	-0.978** (0.471)	-0.303 (0.375)	-0.947** (0.420)
Controls		Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes		Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE			Yes	Yes		Yes
Industry (SIC3) FE				Yes		
Industry-year FE					Yes	
State-year FE					Yes	
Industry (SIC4) FE						Yes
N	36000	36000	36000	36000	36000	36000
R ²	0.971	0.981	0.981	0.981	0.985	0.982

Note: This table shows the first stage of the instrumental variables analysis (Table 4), in which we predict parent firm R&D using federal and state tax prices of R&D, which are partially determined by tax credits that change across time, states, and depending on firm age. The federal R&D tax price is the log firm-level tax price of R&D, based on the federal tax credit, and following Hall (1993) and Bloom et al. (2013). The state R&D tax price is the log state-level tax price of R&D, following Bloom et al. (2013). See Section 3.2 and Appendix 6 for details. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 3: Effect of R&D on Entrepreneurial Spawning

<i>Panel 1</i>				
Dependent variable: Entrepreneurial spawning rate _{t+3}				
	(1)	(2)	(3)	(4)
Log R&D _{t-1}	0.108** (0.051)	0.106** (0.051)	0.099* (0.052)	0.109* (0.060)
Log Employment _t	-0.181*** (0.018)	-0.181*** (0.019)	-0.174*** (0.018)	-0.179*** (0.019)
Log payroll _t	-0.056 (0.053)	-0.057 (0.054)	-0.082 (0.056)	-0.033 (0.054)
Age _t	-0.032 (0.032)	-0.033 (0.033)	-0.021 (0.028)	-0.003 (0.030)
Diversified _t	-0.125 (0.096)	-0.130 (0.095)	-0.135 (0.095)	-0.141 (0.100)
Sales growth _{t-1}	0.127 (0.090)	0.130 (0.090)	0.124 (0.091)	0.129 (0.099)
EBITDA _{t-1}	0.133 (0.262)	0.127 (0.260)	0.155 (0.261)	-0.112 (0.294)
Investment/Total _{t-1}	0.819 (0.541)	0.811 (0.543)	0.731 (0.553)	0.508 (0.617)
Log Tobin's Q _{t-1}	0.031 (0.066)	0.032 (0.067)	0.027 (0.067)	0.044 (0.077)
Log Total Assets _{t-1}	-0.040 (0.069)	-0.033 (0.069)	-0.054 (0.070)	-0.001 (0.066)
PPE Investment/Total Assets _{t-1}	-0.068 (0.385)	-0.058 (0.385)	-0.050 (0.393)	-0.063 (0.424)
Cash _{t-1}	-0.498 (0.307)	-0.502 (0.307)	-0.506 (0.315)	-0.521 (0.320)
Debt _{t-1}	0.059 (0.220)	0.052 (0.220)	0.069 (0.225)	0.187 (0.203)
Year FE	Yes	Yes	Yes	
Firm FE	Yes	Yes	Yes	Yes
State FE		Yes	Yes	
Industry (SIC3) FE	Yes	Yes		
Industry (SIC4) FE			Yes	
Industry-year FE				Yes
State-year FE				Yes
N	36000	36000	36000	36000
R ²	0.172	0.176	0.184	0.180

Panel 2

Dependent variable: Entrepreneurial spawning rate_{t+3}

	(1)	(2)	(3)	(4)	(5)
Log R&D _{t-1}	0.096** (0.045)	0.105** (0.050)	0.102** (0.052)	0.104** (0.051)	0.101** (0.051)
Average employee age _t				-0.036*** (0.007)	
Share employees female _t				-0.084 (0.165)	
Share employees white _t				0.713*** (0.169)	
Share employees foreign _t				0.508** (0.251)	
Average employee education _t				-0.055 (0.043)	
Average employee tenure _t				-0.023* (0.013)	
Average employee experience _t				0.004 (0.017)	
Log patent classes					0.227* (0.120)
Log patents					-0.137 (0.091)
Log forward citations					-0.006 (0.022)
Log backward citations					-0.005 (0.038)
Controls		Yes	Yes		
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
State FE				Yes	Yes
Industry (SIC3) FE				Yes	Yes
Industry (SIC4) FE			Yes		
N	36000	36000	36000	36000	36000
R ²	0.156	0.167	0.181	0.179	0.176

Note: This table shows the effect of corporate R&D on entrepreneurial spawning. The dependent variable is the fraction of an establishment's workers as of first quarter of year zero who are entrepreneurs as of 1st quarter of year 3. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. In panel 2, we do not display controls for models 1-3 because we are limited by Census in the number of coefficients we may disclose. Controls are the same as in panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 4: Instrumental Variables Effect of R&D on Entrepreneurial Spawning

Dependent variable: Entrepreneurial spawning rate _{t+3}						
	(1)	(2)	(3)	(4)	(5)	(6)
Instrumented log R&D _{t-1}	0.577*** (0.207)	0.719*** (0.274)	0.659** (0.271)	0.648** (0.270)	0.587* (0.317)	0.598** (0.276)
Controls	None	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes		Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE				Yes		Yes
Industry (SIC3) FE			Yes	Yes		
Industry-year FE					Yes	
State-year FE					Yes	
Industry (SIC4) FE						Yes
N	36000	36000	36000	36000	36000	36000
R ² (partial from first stage)	0.032	0.027	0.026	0.026	0.022	0.025
F-test for weak instrument	24.70	22.23	22.25	22.37	34.11	27.64

Note: This table shows the effect of instrumented R&D on entrepreneurial spawning. The first stage predicting R&D is shown in Table 2. The dependent variable is the fraction of an establishment's workers as of first quarter of year zero who are entrepreneurs as of 1st quarter of year 3. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. In panel 2, we do not display controls for models 1-3 because we are limited by Census in the number of coefficients we may disclose. Controls are the same as in panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 5: Effect of R&D on Alternative Measures of Entrepreneurial Spawning

Dependent variable:	Entrepreneurial spawning rate _{t+1}	Entrepreneurial spawning rate _{t+2}	Number of spawned firms _{t+3}	Flow entrepreneurial spawning rate _{t+3}
	(1)	(2)	(3)	(4)
Log R&D _{t-1}	0.055** (0.025)	0.076* (0.042)	0.067* (0.037)	0.272* (0.144)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes	Yes
N	36000	36000	36000	36000
R ²	0.090	0.131	0.154	0.092

Note: This table shows the effect of R&D on alternative measures of entrepreneurial spawning. In column 1 (2), the dependent variable is the fraction of an establishment's workers as of first quarter of year zero who are entrepreneurs as of 1st quarter of year 1 (2). In column 3, the dependent variable is the number of distinct new firms associated with workers in our baseline "Entrepreneurial spawning rate_{t+3}" variable. In column 4, the dependent variable is the number of establishment workers as of first quarter of year zero who are entrepreneurs as of the 1st quarter in either year 1, 2, or 3, normalized by the number of workers who, as of the 1st quarter in year 3, have left the firm. In column 1 (2), an entrepreneur is defined as a person at a firm no more than 1 (2) years old who is among the top 5 earners at that new firm. In columns 3-4, an entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 3 Panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 6: Effect of R&D on Non-entrepreneurial Employee Outcomes

Dependent variable:	Stayers _{t+1}	Movers to old firms _{t+3}	Unemployed _{t+3}	Movers to firms of unknown age _{t+3}
	(1)	(2)	(3)	(4)
Log R&D _{t-1}	-1.133 (0.715)	0.485 (0.608)	-0.004 (0.133)	0.506 (0.452)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes	Yes
N	36000	36000	36000	36000
R ²	0.385	0.356	0.222	0.207

Note: This table shows the effect of R&D on alternative employee outcomes. In column 1, the dependent variable is the fraction of an establishment's workers in the 1st quarter of year zero who remain at the firm in the 1st quarter of year 3. In column 2, the dependent variable is the fraction of an establishment's workers in the 1st quarter of year zero who move to a firm that is more than 3 years old by the 1st quarter of year 3. In column 3, the dependent variable is the fraction of an establishment's workers in the 1st quarter of year zero who drop out of the employment sample by the 1st quarter of year 3 (note they may have moved to an uncovered state). In column 4, the dependent variable is the fraction of an establishment's workers in the 1st quarter of year zero who move to an organization whose age is unknown by the 1st quarter of year 3. Controls are the same as in Table 3 Panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 7: Reverse Causality Test (Effect of Entrepreneurial Spawning on R&D)

Dependent variable: Log R&D _t			
	(1)	(2)	(3)
One-year entrepreneurial spawning rate _{t-1}	0.008 (0.005)		
Two-year entrepreneurial spawning rate _{t-2}		0.001 (0.006)	
Three-year entrepreneurial spawning rate _{t-3}			-0.005 (0.003)
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes
N	36000	36000	36000
R ²	0.879	0.879	0.879

Note: This table shows that current entrepreneurial spawning does not predict corporate R&D. The independent variables are lagged variations on our main entrepreneurial spawning rate measures used as the dependent variable in Tables 3 and 4. The one-year entrepreneurial spawning rate_{t-1} is the fraction of an establishment's workers as of first quarter of year $t - 1$ who are entrepreneurs as of 1st quarter of year t , which is the year that R&D is measured (the dependent variable). The two-year entrepreneurial spawning rate_{t-2} is the fraction of an establishment's workers as of first quarter of year $t - 2$ who are entrepreneurs as of 1st quarter of year t . The three-year entrepreneurial spawning rate_{t-3} is the fraction of an establishment's workers as of first quarter of year $t - 3$ who are entrepreneurs as of 1st quarter of year t . An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 3 Panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 8: Parent Variation in Effect of R&D on Entrepreneurial Spawning

<i>Panel 2</i>				
Dependent variable: Entrepreneurial spawning rate _{t+3}				
	(1)	(2)	(3)	(4)
Log R&D _{t-1}	0.048 (0.057)	0.016 (0.062)	0.035 (0.066)	0.099* (0.052)
Log R&D _{t-1} ·High Tech	0.083*** (0.029)			
High Tech	1.378*** (0.351)			
Log R&D _{t-1} ·Large		0.130** (0.056)		
Large		-0.333** (0.149)		
Log R&D _{t-1} ·Old			0.098 (0.067)	
Old			-0.315 (0.491)	
Log R&D _{t-1} ·High patent generality				0.027* (0.016)
High patent generality				-0.093 (0.076)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes	Yes
N	36000	36000	36000	36000
R ²	0.176	0.176	0.176	0.176

Note: This table shows how the effect of corporate R&D on entrepreneurial spawning varies by parent firm characteristics. High Tech is 1 if the parent is in a high-tech industry, and 0 if not. Large is 1 if the parent has above-median total assets (calculated at the firm-year level), and 0 if below-median. Old is 1 if the parent is of above-median age (calculated at the firm-year level), and 0 if below-median. High patent generality is 1 if the parent has above-median patent generality (calculated at the firm-year level), and 0 if below-median. The dependent variable is the fraction of an establishment's workers as of first quarter of year zero who are entrepreneurs as of 1st quarter of year 3. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 3 panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 9: Employee Variation in Effect of R&D on Entrepreneurial Spawning

<i>Panel 1</i>				
Dependent variable: Entrepreneurial spawning rate _{t+3}				
Establishment employees restricted to:	Above median wage	Below median wage	Above median tenure	Below median tenure
	(1)	(2)	(3)	(4)
Log R&D _{t-1}	0.125 (0.081)	0.090 (0.062)	0.113* (0.061)	0.099 (0.061)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes	Yes
N	36000	36000	36000	36000
R ²	0.164	0.085	0.160	0.091

<i>Panel 2</i>				
Dependent variable: Entrepreneurial spawning rate _{<i>t</i>+3}				
Establishment employees restricted to:	Above median age	Below median age	Male	Female
	(1)	(2)	(3)	(4)
Log R&D _{<i>t</i>-1}	0.144** (0.062)	0.070 (0.066)	0.079 (0.070)	0.089 (0.063)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes	Yes	Yes	Yes
N	36000	36000	36000	36000
<i>R</i> ²	0.149	0.110	0.159	0.033

Note: This table shows the effect of R&D on entrepreneurial spawning within subsets of employees. Median is defined within the establishment-year. Thus in panel 1 column 1 the dependent variable is the fraction of employees with above median earnings (calculated within the establishment-year) in first quarter of year zero who are entrepreneurs as of 1st quarter of year 3. Tenure is measured as the duration of employment at the establishment. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 3 Panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 10: Effect of R&D on Entrepreneurial Spawning by Spawn Characteristics

Panel 1: What predicts venture capital-backed spawns?

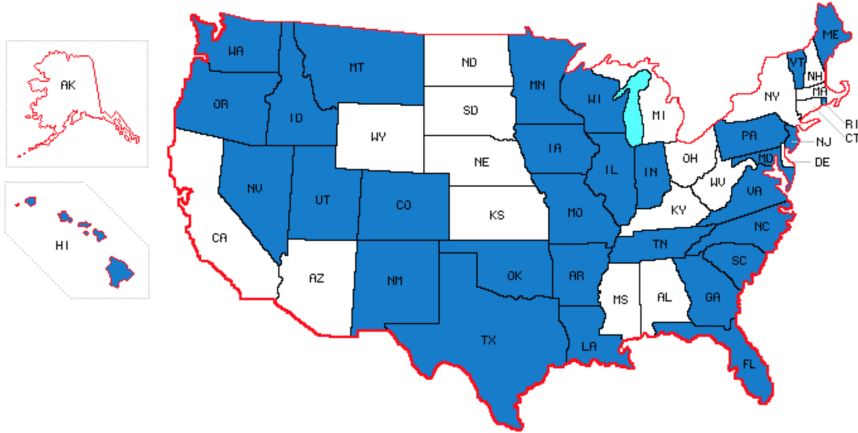
Dependent variable: Spawn ever received VC

	(1)	<i>...Continued</i>	
Log R&D _{t-1}	0.007*** (0.001)		
Employee age _t	0.001** (0.000)	Establishment Log Employment _t	0.001 (0.001)
Employee age ² _t	-0.000** (0.000)	Establishment average employee wage _t	0.012*** (0.003)
Employee female	-0.013*** (0.002)	Firm Age _t	-0.002*** (0.001)
Employee white	0.003** (0.001)	Firm Diversified	-0.003 (0.006)
Employee foreign	-0.002 (0.004)	Firm Sales growth _{t-1}	0.004 (0.006)
Employee born in state	-0.007*** (0.001)	Firm EBITDA _{t-1}	-0.008 (0.016)
Employee education	0.001*** (0.000)	Firm Investment/Total _{t-1}	-0.013 (0.041)
Employee experience _t	-0.000 (0.001)	Firm Log Tobin's Q _{t-1}	0.002 (0.004)
Employee tenure _t	-0.000 (0.000)	Firm Log Total Assets _{t-1}	-0.006*** (0.002)
Employee log earnings _t	0.008*** (0.002)	Firm PPE Investment/Total Assets _{t-1}	-0.004 (0.012)
Spawn age _{t+3}	0.007*** (0.001)	Firm Cash _{t-1}	0.076*** (0.015)
Spawn initial employment	0.008*** (0.002)	Firm Debt _{t-1}	0.009 (0.007)
<i>Continued...</i>		Year-state FE	Yes
		Year-Industry (SIC3) FE	Yes
		N	108000
		R ²	0.079

<i>Panel 2: Other spawn characteristics</i>					
Dependent variable:	Spawn in same industry (SIC2) as parent	Spawn in same state as parent	Spawn in a high-tech industry	Spawning employee's log wages _{t+3}	Spawn exit _{t+5}
	(1)	(2)	(3)	(4)	(5)
Log R&D _{t-1}	-0.007** (0.003)	0.002 (0.002)	0.009*** (0.004)	0.028*** (0.006)	0.007** (0.003)
Controls	Yes	Yes	Yes	Yes	Yes
Year-state FE	Yes	Yes	Yes	Yes	Yes
Year-industry (SIC3) FE	Yes	Yes	Yes	Yes	Yes
N	108000	108000	108000	108000	108000
R ²	0.206	0.053	0.102	0.318	0.083

Note: This table shows the effect of R&D on types of entrepreneurial spawning. Based on the main variable used in Table 3, “Entrepreneurial spawning rate_{t+3}”, we identify whether the new firm associated with the spawning employee has a given characteristic. The dependent variable in panel 1 column 1 is 1 if the spawn ever received VC backing (either before or after the spawn is identified in year $t + 3$), and 0 if not. The “Employee...” controls in panel 1 column 1 refer to the spawning employee who left the parent. The dependent variable in panel 2 column 1 (2) (3) is 1 if the spawn is in the same 2-digit SIC code as the parent (is in the same state as the parent) (is in a high-tech industry), and 0 if not. The dependent variable in panel 2 column 4 is the spawning employee’s log wages at the new firm in the 1st quarter of year 3. The dependent variable in panel 2 column 5 is 1 if the spawn experienced an acquisition or IPO by year 5, and 0 if not. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 10 Panel 1, except that we include the indicator for being VC-backed as an additional control in columns 1-4. Standard errors are clustered by parent firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Figure 1: Map of States with LEHD (spawn) Data



Note: This figure shows the LEHD states that we have access to. We observe all spawns located in these states, of public parent establishments in all 50 states.

Appendix

(for online publication)

A1 Instrumental variables calculation and discussion

A1.1. The Federal R&D tax credit

The first instrument is the federal tax price of R&D, which we denote ρ_{ft}^F . Implemented in 1981, the federal “Research and Experimentation” tax credit permits firms to reduce their corporate income tax liability by the value of the credit. The credit was extremely complex to calculate (leading to a substantial simplification in 2009), and has changed over time. To extract value from the tax credit, a firm must have substantial after tax profit.¹⁷ In the early 2000s, the total value of the federal credits was about \$5 billion per year (Wilson et al. 2005).

In this description, we focus on the calculation of the credit between 1990 and 2006, which is the sample period for which we need to predict public firm R&D.¹⁸ The general formula for the R&E tax credit is as follows, for tax year t and firm f :

$$R\&E\ Tax\ Credit\ Value_{tf} = 20\% \cdot [QRE_{tf} - Base_{tf}] + 20\% \cdot [Basic\ Research_{tf}] \quad (4)$$

The last element, basic research expenditures, must be paid to a qualified organization, which is either a research university or tax-exempt scientific organizations. The other, more complex type of research costs are qualified research expenditures (QRE). These must occur within the U.S., and have three categories: salaries and wages, supplies, and contract research. The law is quite specific about what counts and what does not count as QRE. For example, QRE must be technological in nature and relate to new or improved function, performance, reliability, or quality. Among other excluded

¹⁷The amounts have generally not been large enough for firms without sufficient profits to sell their credits in the tax equity market.

¹⁸The calculation was quite different before 1989. In practice, we draw heavily from code originally written for Hall (1993).

types, research after commercial production of a component, survey research, and social science research do not count.¹⁹

The “base” amount is by far the most complicated element. It is constructed using the following equation:

$$Base_{tf} = Fixed\ Base\ \%_{tf} \cdot Sales_t$$

The complexity lies in the fixed base percentage, which varies by a firm’s “startup” status (number of years since the firm’s first instance of QRE). It is calculated as follows (firm index omitted for simplicity):

$$Fixed\ Base\ \% = \begin{cases} \max \left[\frac{\sum_{t=1984}^{1988} \frac{QRE_t}{Sales_t}}{5}, 0.16 \right] & \text{if } QRE_{1983} > 0 \ \& \ Sales_{1983} > 0 \\ 0.03 & \text{if } QRE_{t-6} \in [0, \emptyset] \\ \frac{1}{6} \left[\frac{\sum_{t=-2}^{-1} \frac{QRE_t}{Sales_t}}{2} \right] & \text{if } QRE_{t-7} \in [0, \emptyset] \ \& \ QRE_{t-6} > 0 \\ \frac{1}{3} \left[\frac{\sum_{t=-2}^{-1} \frac{QRE_t}{Sales_t}}{2} \right] & \text{if } QRE_{t-8} \in [0, \emptyset] \ \& \ QRE_{t-7} > 0 \\ \frac{1}{2} \left[\frac{\sum_{t=-3}^{-1} \frac{QRE_t}{Sales_t}}{3} \right] & \text{if } QRE_{t-9} \in [0, \emptyset] \ \& \ QRE_{t-8} > 0 \\ \frac{2}{3} \left[\frac{\sum_{t=-4}^{-1} \frac{QRE_t}{Sales_t}}{4} \right] & \text{if } QRE_{t-10} \in [0, \emptyset] \ \& \ QRE_{t-9} > 0 \\ \frac{5}{6} \left[\frac{\sum_{t=-5}^{-1} \frac{QRE_t}{Sales_t}}{5} \right] & \text{if } QRE_{t-11} \in [0, \emptyset] \ \& \ QRE_{t-10} > 0 \\ \min \left[\frac{QRE_t}{Sales_t} \right]_{t-6}^{t-1} & \text{if } QRE_{t-x} \in [0, \emptyset] \ \& \ QRE_{t-x-1} > 0 \ \forall \ x \geq 12 \end{cases}$$

¹⁹The complete legal text is here: <https://www.law.cornell.edu/uscode/text/26/41>.

In words, the first row is interpreted in the following way. For firms that had positive QRE and sales in 1983, the fixed base percentage is the maximum of 16% and the average of R&D intensity over the five years between 1984 and 1988. All the subsequent rows in the above equation pertain to what the law terms “startups.” For example, for the first five taxable years after the first year in which a firm has positive QRE, the fixed base is 3%. In the 6th such year, it is one-sixth the average of the R&D intensity over the previous two years. The following rows are similarly calculated. Starting in the eleventh such year, firm may choose the percentage from any of the prior fifth through tenth years.

A few other details bear mention. The expense deduction for R&D is recaptured, reducing the effective credit rate from 20% to about 13.5%. Also, in the fiscal year 1995-6, the credit lapsed entirely. Additionally, when the credit value is larger than taxable profits, it can be carried forward for ten years. Finally, between 1990 and 1996, the only option was the R&E tax credit. Starting in 1996, firms could elect the alternative incremental credit (AIC), in lieu of the R&E tax credit. This has 3 tiers depending on R&D intensity (QRE relative to sales); if intensity is 1-1.5% (1.5-2%) (>2%), the AIC rate is 2.65% (3.2%) (3.75%), respectively. These rates have varied over time; they were lower in the late 1990s, and have increased in recent years.

The credit is firm-specific for a number of reasons. First, it depends on firm age, with annual changes for most firms. Second, the “base” amount of R&D is calculated using a firm’s past R&D and current-year sales. Third, the base amount of the tax credit is the difference between realized R&D and the base. Fourth, there is a lower implicit value of the credit among tax exhausted firms because the value of the carry forward must be discounted. Finally, the lapse in 1995-96 generates additional within-firm variation, only for firms with R&D expenditures that year.

The R&E tax credit (denoted ERC_t) is in practice considerably more complicated to calculate than Equation 4, and follows Equation 7 in Hall (1992) and underlying equations not shown in her paper; these are available in Stata code on request. Calculating ERC_t begins with the tax credit rate (constant across firms), and multiplies by a categorical variable derived from QRE. This is then deducted from corporate tax liability. Then, a 3-year carry-back and a 15-year carry-forward are added in cases

of no taxable income this year. Once this tax credit is arrived at, the tax price of R&D is calculated following Equation 6 in Hall (1992). This is:

$$\rho_{ft}^F = \rho_t^R \left[1 - T_t (1 + r)^{-J_t} \tau \right] - \eta ERC_t \quad (5)$$

Here, ρ_t^R is an R&D deflator divided by a GDP deflator, or the "price" of R&D investment in the absence of taxes, T_t is an indicator for whether the firm has taxable income in the current year, J_t is the number of years until loss carry-forwards will be exhausted, τ is the corporate tax rate, and η is QRE. If $\rho_{ft}^F = 1$, then the firm should not treat R&D differently than other expenditure. If $\rho_{ft}^F < 1$, R&D is less expensive than other expenditure because of the tax credit.

In practice, we find substantial within-industry variation in ρ_{ft}^F , especially in manufacturing and services. The median tax price is well below 1 on average, so that R&D is cheaper than other spending. Within industries, the distributions have negative skew (i.e., a longer right tail). We also ensure that relevant current year variables, including R&D, do not have strong explanatory power over the tax price of R&D. Within firms, we find small positive correlations (all less than 0.1) between ρ_{ft}^F and employment, assets, and R&D. In regressions, we verify substantial firm-level variation in the tax price of R&D. Firms in high tech areas such as pharmaceuticals and electronics, tend to have the most variation.

A1.2 State R&D tax credits

State R&D tax credits have been generally modeled on the federal one. The first state R&D tax credit was implemented in 1982 by Minnesota; by the end of our sample period, forty states had some sort of R&D tax credit. The calculation of the base amount, and the definition of qualified R&D, can vary across states (Wilson et al. 2005). According to Miller & Richard (2010), manufacturing-intensive states, and those with one-party political control, are more likely to pass R&D tax credits. They argue that the tax credits primarily support incumbent R&D-conducting firms.

The state instrument requires two objects: the state tax price component of the R&D user cost of capital, and a measure of the share of a firm's R&D that occurs in

a given state. For both, we follow Bloom et al. (2013). First, we use the state tax price of R&D in Wilson (2009). He incorporated state level corporate income taxes, depreciation allowances, and R&D tax credits into this tax price component, which we call ρ_{st}^S .²⁰ These credits vary across states and time. They allow a firm to offset its state-level corporate tax liabilities, and they are calculated by weighting total firm profits according to the location of the firm’s sales, employment, and property. Thus firms with R&D activities in the state will likely both have tax liability and R&D tax credit eligibility there.

The second object, θ_{fst} , is a proxy for a firm’s R&D share in a given state-year. It is the 10-year moving average of the share of the firm’s patent inventors located in state s .²¹ The firm’s state-level tax price is then $\rho_{ft}^S = \sum_s \theta_{fst} \rho_{st}^S$.

A1.3 Concerns

There are four potential concerns. Most importantly, the exclusion restriction is that tax credits cannot affect entrepreneurial spawning. We show empirically that there is no relation between the state tax credits and state-level startup creation, or the federal tax credit and national startup creation. We do this using two data sources, each of which have limitations. The first is the Business Dynamics Statistics (BDS), which contains firm entry by state for our entire sample period, but does not have state-industry data.²² The second is the Quarterly Workforce Indicators (QWI), a publicly available dataset derived from the LEHD. While the QWI has state-industry level data, its coverage is poor in the early years of our data, with counties being added over time.²³

At the state level, using the BDS sample, we regress either the log number of new firms or the change in firm entry rates year to year on the tax price of R&D, as well as state and year fixed effects. The results are in Table 1. We cluster errors by state. Regardless of the fixed effects or standard error assumptions, we find that the

²⁰Specifically, it is roughly: $\frac{1-(tax\ credits+depr.\ allowances)}{1-tax\ rate}$.

²¹The data is from NBER patent data, available at <https://sites.google.com/site/patentdatapoint/Home/downloads>.

²²This public version of the LBD is available at https://www.census.gov/ces/dataproducts/bds/data_firm.html.

²³We used a transformed version of the data used in Adelino et al. (2017), courtesy of Song Ma.

tax credits have no correlation with startup entry (panel 1). Using the QWI sample, our dependent variable is either the logged new jobs created in new firms in the past two years, or the change in the number of new jobs created in new firms in the past two years. We consider only R&D-intensive industries.²⁴ Again, regardless of whether we use year and/or state fixed effects, and regardless of the standard error assumptions, we find no effect of the tax price of R&D on these measures. This is in Table 1 Panel 2.

At the federal level, we regress either the log number of new firms or the change in firm entry rates on the statutory federal R&D tax credit. This is, of course, very different from the firm-specific tax price of R&D that is calculated per the description in Section A1.1. This reflects baseline changes in the rate, which is then applied to a firm’s specific situation. There are very few observations, and we do not use robust standard errors. The results, in Table 1 Panel 3, again show no correlation.

More generally, the legal literature has argued that R&D tax credits are not useful to startups, as they have no or little taxable income against which to offset losses from failed R&D efforts (Bankman & Gilson 1999).²⁵ Perhaps in response to this, a few states have recently made their R&D tax credits transferable, so that firms without revenue can potentially derive value from them. However, these policies occurred after the end of our sample period.

The second concern is that changes in state-level R&D tax credits may lead firms to reallocate R&D (or misreport it such that it appears reallocated). For studies evaluating how a state-level R&D tax credit affects national R&D, this is a central concern. In our case, however, such reallocation will simply reduce the power of the instrument. As long as the combined instruments have adequate power, some degree of reallocation should not bias our findings. It does lead us to expect that the federal instrument will have more power than the state instrument, which is indeed what we find. This is because it should have a larger effect on firms that only operate in the

²⁴NAICS codes 31-33, 51, and 54.

²⁵Bankman & Gilson (1999) note that “the U.S. tax code subsidizes R&D by existing successful companies by allowing losses from failed attempts at innovation to offset otherwise taxable income from other activities. Since startups have no other income against which their losses from a particular project may be set off, the government in effect gives established companies with a stable source of income an R&D tax subsidy that is not available to a startup entity.”

affected state, but most firms with positive R&D operate in multiple states.

The third concern is that the tax credits may not be large enough to affect R&D. The above sections pointed to substantial literature finding R&D responses to R&D tax credits that are large in economic magnitude and quite robust, especially for the federal instrument. The literature examining the state instrument finds large within-state elasticities, but also finds evidence of reallocation across states.

Finally, the fourth concern is that state decisions to adopt R&D tax credits could be endogenous, reflecting recent declines in R&D. Bloom et al. (2013) consider this possibility at length, and show that the results are robust to lagging the tax credit instruments one and two periods. They also point out that cross-sectional variation in the state R&D tax credit rates is very large relative to the average rate within states, and also large relative to the secular increase in the tax credit generosity that has occurred over time. Finally, Chirinko & Wilson (2008), Chirinko & Wilson (2011), and Bloom et al. (2013) show that the level and timing of R&D tax credit adoption is uncorrelated with local economic observables like state R&D expenditure or per capita GDP, once year and state fixed effects are included.

In sum, we believe that R&D tax credits offer the best available source of variation driving corporate R&D that is plausibly unrelated to technological opportunities that could jointly give rise to parent R&D and entrepreneurial spawning.

Table 1: Relationship between state tax price of R&D and state startup formation

<i>Panel 1: Quarterly Workforce Indicator (LEHD) data</i>						
Dependent variable	2-year employment growth		Log 2-year employment growth		Change in 2-year employment growth	
	(1)	(2)	(3)	(4)	(5)	(6)
State tax price of R&D	-20068 (21295)	4754 (9035)	-.74 (.59)	.33 (.36)	-117 (7912)	-6.5 (57677)
State f.e.	Y	Y	Y	Y	Y	Y
Year f.e.	Y	N	Y	N	Y	N
N	449	449	449	449	448	447
R^2	.21	.2	.44	.43	.11	.11

<i>Panel 2: Business Dynamics Statistics Data</i>						
Dependent variable	2-year employment growth		Log 2-year employment growth		Change in 2-year employment growth	
	(1)	(2)	(3)	(4)	(5)	(6)
State tax price of R&D	-1650 (3570)	-493 (756)	-.11 (.37)	.036 (.084)	188 (1619)	-583 (981)
State f.e.	Y	Y	Y	Y	Y	Y
Year f.e.	Y	N	Y	N	Y	N
N	1530	1530	1530	1530	1529	1529
R^2	0.1585	0.0016	.24	0.0012	0.0204	0.0005

Note: This table shows estimates of the relationship between last year's state tax price of R&D (from Wilson), and employment growth at new firms. Panel 1 uses data from the QWI, courtesy of Song Ma. Firms are limited to R&D-intensive (high tech) sectors. Panel 2 uses data from the BDS, where all firms are used as the data do not include industry information. Errors are clustered at the state *** indicates p-value<.01.

<i>Panel 3:</i>				
Data source:	Quarterly Workforce Indicator (LEHD) data		Business Dynamics Statistics Data	
Dependent variable	Log 2-year employment growth	Change in 2-year employment growth	Log 2-year employment growth	Change in 2-year employment growth
	(1)	(2)	(3)	(4)
Federal R&D credit	4.4	-39912	-.19	-377227
	(7.3)	(885697)	(.16)	(274243)
N	16	15	30	37
R^2	.026	.00016	.05	.051

Note: This panel shows estimates of the relationship between last year's federal tax price of R&D, and employment growth at new firms. *** indicates p-value<.01.

Table 2: Effect of Alternative Measures of R&D on Entrepreneurial Spawning

Dependent variable: Entrepreneurial spawning rate _{t+3}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Above median Δ R&D _{t-1}	0.089*** (0.033)	0.078** (0.032)					
Top 10 pct Δ R&D _{t-1}			0.132** (0.067)	0.157** (0.070)			
Bottom 10 pct Δ R&D _{t-1}					-0.105** (0.053)	-0.114* (0.060)	
R&D _{t-1} /Total Assets _{t-2}							1.020** (0.495)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes		Yes		Yes		Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry (SIC3) FE	Yes		Yes		Yes		Yes
State FE	Yes		Yes		Yes		Yes
Year-industry FE		Yes		Yes		Yes	
Year-state FE		Yes		Yes		Yes	
N	36000	36000	36000	36000	36000	36000	36000
R ²	0.176	0.180	0.176	0.180	0.176	0.180	0.175

Note: This table shows the effect of alternative measures of R&D on entrepreneurial spawning. Change (Δ) in R&D is defined as: $\frac{R\&D_{t-1} - R\&D_{t-2}}{.5 \cdot (R\&D_{t-1} + R\&D_{t-2})}$. Top 10 pct Δ R&D_{t-1} is 1 if the firm had a change in R&D that is in the top 10 percentiles (relative to all firm-years), and 0 if in the bottom 90 percentiles. Bottom 10 pct Δ R&D_{t-1} is defined analogously. The dependent variable is the fraction of an establishment's workers as of first quarter of year zero who are entrepreneurs as of 1st quarter of year 3. An entrepreneur is defined as a person at a firm no more than 3 years old who is among the top 5 earners at that new firm. Controls are the same as in Table 3 Panel 1. Standard errors are clustered by firm. *, **, and *** denote significance at the 10%, 5%, and 1% levels.