The Lifecycle of Inventors*

Alex Bell, Harvard University
Raj Chetty, Harvard University and NBER
Xavier Jaravel, Harvard University
Neviana Petkova, Office of Tax Analysis, US Treasury
John Van Reenen, Centre for Economic Performance, LSE and NBER

March 19th, 2015; Preliminary: Not for circulation

Abstract

We use administrative records on the population of individuals granted patents between 1996 and 2012 to characterize the lives of more than 780,000 inventors in the United States. We show that children of low-income parents are much less likely to become inventors than their higher-income counterparts. Decompositions indicate that income-innovation gaps can largely be accounted for by differences in human capital acquisition while children are growing up rather than innate ability or differences in preferences between children from different backgrounds. Consistent with the importance of environment, we find that growing up in a high-innovation area (of a particular technology class) is associated with an increased probability of becoming an inventor (in that technology class). Using an event study methodology, we show that the returns to innovation are extremely skewed, with a small number of highly cited patents yielding very large returns for the inventors. While our analysis does not directly identify the causal mechanisms that drive innovation, our descriptive findings shed light on which types of policy tools are likely to be most effective in sparking innovation. In particular, our findings suggest that "supply side" policies targeted at drawing more talented individuals into innovation have greater potential than "demand side" policies such as reductions in tax rates on high-income individuals.

*The opinions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Internal Revenue Service or the U.S. Department of the Treasury. We thank Philippe Aghion, Ufuk Akcigit, Olivier Blanchard, Danny Kahnemann, Josh Lerner, Otto Toivanen and participants in seminars in Brussels, Harvard, LSE and the OTA for helpful comments and discussions. Augustin Bergeron, Jamie Fogel, Nikolaus Hidenbrand, and Benjamin Scuderi provided outstanding research assistance. This research was funded by the National Science Foundation, the Lab for Economic Applications and Policy at Harvard, and the Economic and Social Research Council.

I Introduction

Innovation is at the center of many modern theories of economic growth and advanced economies use a broad range of policies intended to spur innovation, ranging from subsidies for research to investments in technical education. Yet relatively little is known about the characteristics and life trajectories of inventors. Indeed, information on even the most basic demographic characteristics - such as the age distribution and gender composition of patent holders - is scarce because existing databases do not record such information.

In this paper, we present the first comprehensive portrait of inventors in the United States by linking data on all patents granted from 1996-2012 to federal income tax returns. Our linked dataset contains information on over 786,000 patent holders. We use this linked data to document a set of stylized facts about the lives of inventors that inform current theoretical debates and identify new patterns to be explained by the next generation of models of innovation and growth. We structure our analysis around the chronology of an inventor's life, starting from her family background and neighborhood at birth, on to her education and then labor market career, including the characteristics of the organization that employs her.

While our focus is primarily on descriptive facts rather than identification of causal mechanisms, the facts we document help discriminate between alternative theories in the literature and shed light on the types of policies that are likely to be most effective in sparking innovation. Most economic research on innovation policies focuses on what could be termed the "demand side" of innovation, such as changes in the tax system to increase innovation incentives. For example, the US Research and Experimentation (R&E) tax credit puts R&D expenditures in a tax privileged status compared to other forms of investment (see Bloom et al., 2013 for a recent assessment). Similarly, a leading argument for reducing top income tax rates on income is that doing so will increase incentives for innovation and therefore boost growth. One potential problem with such demand side policies is that they rely on those with the current potential to innovate to do more of it.[1] In light of these limitations, Romer (2000) recommends studying policies that focus on increasing the supply side of innovation. Our analysis of the lives of inventors yields a better understanding of what makes and constrains an innovator, which is an essential ingredient of formulating such supply side innovation policies.

We begin by studying the birth and origin of inventors. The first stylized fact we uncover is that

[1] In fact, there is a risk that with an inelastic supply of innovators, subsidies to research will simply increase the equilibrium wage of R&D scientists, rather than stimulate a greater volume of innovation (e.g. Goolsbee 1998).

the children of high-income parents are much more likely to be inventors: children born to parents in the top 1% of the income distribution are more than 10 times as likely to have a future patent as children born to families with below-median income. This relationship between parent income and children's patent rates could be driven by at least three different mechanisms. First, the children of the rich may be born with higher ability than those of the poor and so are naturally more likely to make technological breakthroughs. Second, low-income children may have different preferences than rich ones. For example, they may be more risk averse and so less likely to enter the innovation sector, which as we document below, yields highly variable payoffs. Thirdly, even when children begin with identical cognitive and non-cognitive traits, having low-income parents may hold children back from becoming innovators because of a lower quality education, neighborhoods, mentors, or jobs opportunities. While the first two channels do not call for obvious policy interventions, the third channel – a lack of opportunity for those who could have had a comparative advantage in innovation but for their financial position – is of particular policy interest because it would mean a loss of innovation and output due to a misallocation of talent. How many "lost Einsteins" could there be due to inequality of opportunity (e.g. Melik, 2015)?

We distinguish these three different mechanisms by using data from all those who went through the New York public school system between 1989 and 2009, from which we have standardized data on test score results in math and English in grades 3 through 8. We show that a substantial difference in educational outcomes between rich and poor families opens up as children progress through school, and it is this educational difference between the children of the rich and poor that accounts for most of the subsequent differences in the probability of becoming an innovator. Hence, it seems less likely that initial intrinsic differences in ability or preferences lie behind the innovation gap of the poor; rather, the differences are more closely linked to their worse educational outcomes.

We conduct a similar analysis to study the gap in innovation between men and women. We find that the vast majority of patents go to men; the share going to women is rising over time, but even in recent cohorts, less than 15% of patents go to women and at the current rate of convergence, it will take another 140 years for women to obtain 50% of granted patents. Unlike the gap in innovation by parents' socio-economic status, we find little evidence that the gender gap in innovation can be explained by differences in childhood environment, suggesting that a different set of policies is needed to draw more women into science and innovation.

In the next section of the paper, we characterize how the environment in which a child grows up affects his probability of becoming an inventor. We find that there is substantial variation across

states in the likelihood that children become innovators. Children who grow up in areas with a large number of universities and higher mean incomes tend to be much more likely to go into innovation. Perhaps more interestingly, we find that the propensity to become an inventor is strongly related to the density of inventors in the state where one grows up. This relationship holds even within patent technology classes, controlling for current state of residence. For instance, if we compare two individuals living in Massachusetts as adults, one of whom grew up in California (where there are many computer technology inventors) and another who grew up in Minnesota (where there are relatively more medical instrument inventors), the individual from Minnesota is more likely to have a patent in medical instruments than computer technology. This evidence suggests that mentoring effects or exposure to careers in science and innovation at young ages may play a key role in children's later outcomes. Insofar as children from low-income backgrounds are less likely to benefit from such exposure, this evidence reinforces the view that the innovation gap between the rich and the poor is driven by differences in environment and human capital accumulation. We go on to show that successful inventors were educated in a small number of elite colleges. Graduates from the top 10 US schools account for 23% of all citations and the top 20 schools account for a third. Controlling for college attendance, the gap in innovation rates between rich and poor children becomes much smaller, suggesting that the key source of the gap is before rather than after children enter the labor market.

In the third section of the paper, we turn to inventors' careers. Using event studies, we show that there is a significant increase in income after the patenting for the most highly cited patents, although this is very variable: a small fraction of highly-cited patents generate very large returns, but most patents have little or no return. The income gains are driven by a combination of wage earnings and non-wage income (e.g., royalties). Most of the earnings gains do not appear to be coming from starting up one's own firm, from switching employers, or from moving geographically.

The substantial variability in the returns to individual innovation has the implication that policies such as changes in top marginal tax rates are unlikely to have large effects on innovation, as the sensitivity of individuals' utility payoffs to marginal tax rates is small when returns are so skewed. We calibrate a simple model to our data, which shows that for plausible parameters, even large cuts in top income tax rates will only induce a small amount of innovation. Together with our preceding results, this finding suggests that focusing on the "supply side" by drawing more high ability children from low income families has greater potential to raise innovation than changes in ex-post incentives.

Our findings build and and contribute to several vast literatures. The theoretical literature on the individual incentives to innovate are summarized by Scotchmer (2004). For example, there is extensive work on how different types of employment contracts will alter innovation incentives (e.g. Pakes and Nitzan (1983); Franco and Mitchell (2008)). Second, there is a growing literature on how misallocation can be a first order constraint on economic performance. Hsieh and Klenow (2009) argue that reducing misallocation of resources to the level of the US would raise India's TFP by about 50%. More specifically, Hsieh et al (2013) argue that 15-20% of US GDP per worker growth 1960-2008 can be explained by the improved allocation of talent by race and gender. We find evidence that innovation (and growth) are held back because highly talented children from low income families are not becoming innovators as quickly as their richer, but less talented, peers. Although the link between inter-generational inequality and misallocation has been frequently discussed, it has not to our knowledge ever been examined in a systematic statistical manner. Third, there is a literature on academic scientists (e.g. Azoulay et al. (2010b); Azoulay et al. (2010a)) where biographies can be more easily built up. There is related work looking at sub-sets of patentors, especially star scientists in bio- and nano-technology (e.g. Zucker et al. (1998); Zucker et al. (1998)).

Finally, there is a literature on looking at the characteristics of patentors (see Jung and Ejermo, 2014, for a survey). The classic study is Schmookler (1957) who examined 87 US patenters and the most comprehensive recent work is the PatVal-EU data (e.g. Giuri et al, 2007) which covers 9,107 inventors filing at the European Patent Office (EPO). An issue with these studies is that sample responses are low and possibly non-random. In response to this issue researchers have recently started matching patent data to near population employer-employee administrative datasets.Toivanen and Vaananen (2012) match administrative wage data to 1,800 Finnish inventors at the USPTO (the US Patents and Trademark Office) 1988-1996 and find significant wage returns to patenting (as do Depalo and Addario, 2014, on matched Italian data 1987-2006). Using this Finnish data combined with a distance to college instrument, Toivanen and Vaananen (2015) argue that access to schools offering post-grad engineering training have a causal impact on becoming an inventor. Jung and Ejermo (2014) use Swedish patents to examine the issue of gender and age differences for just under 20,000 inventors between 1985 and 2007. Dorner et al. (2014) match the IEB employer-employee data (see Card et al., 2013) to a cross section of German patents in 2002 covering about 32,000 inventors. They find that inter-firm mobility is lower for the individuals who patent more.

The advantage of our data over these earlier efforts is that (i) it is far more comprehensive than these other studies being at least one order of magnitude larger; (ii) we focus on the US as the

country that is at the technology frontier in most industries and (iii) the US also has a relatively competitive labor market and so is less likely to depend on institutional idiosyncrasies.[2]

The structure of the paper is as follows. The next section describes the data construction and provides several new summary statistics that are of interest to the literature on innovation. Section III presents our main results on the lifecycle of innovation. Section IV discuss implications for innovation policy and Section V concludes.

II Data

II.A Patent Data

The raw source of the data is the several thousand weekly text and XML files of patent grant records hosted by Google. The files on this page contain the full text of about 5 million patents granted from 1976 to today, extracted from the USPTO's internal databases in weekly increments. The raw data is organized hierarchically by patent, and was then translated into relations such as inventors, citations, etc. for economic analysis.

II.B IRS Data

From the IRS Compliance Data Warehouse (CDW), we collect information on inventors' city/state, wages, employer ID, adjusted gross income, and royalties, as well as their current citizenship status and gender sourced from Social Security records. Most data are available starting in 1996, but wages and employer ID are available only starting in 1999. A great deal of information exists at the IRS on the employers of inventors in each year as well, including number of employees based on W2 counts, profits/losses, and R&E and other tax credits.

II.C Matching

We match inventors to taxpayers using inventors' name, city, and state. Any inventor whose given address is outside of the United States is excluded from the matching process and dropped. We find equivalent information for taxpayers on W2's, 1040's, and other information return forms. The iterative stages of the match algorithm are described in more detail in Appendix A. We match approximately 88% of inventors of patents applied for in the last decade (the period in which information returns are most available) and slightly above 80% in the late 1990s. We conduct various exercises to assess the quality of the match using additional data sources (e.g. data on

[2]For example, there are explicit laws in Finland that force firms to share the benefits of patenting with inventors making it harder to generalize the wage-innovation result.

inventor age from Jones (2010)). We also explore selection on observables and find no strong selection effects, although the match rate is slightly lower among self-assigned patents.

Because the taxpayer data is a source of linked observations of variants of a person's name and cities of residence lived in over time, the matching process provides a simple way to link different patents filed by the same inventor even if the inventor's name differs across patents or he has moved cities. Preliminary comparisons show that our matching technique yields similar results to the disambiguation routines of Lai et al. (2014). While the panel nature of the linked dataset allows us to see full income and patenting profiles for our sample, a drawback is that we cannot classify individuals as inventors who were active only prior to 1996.

II.D Inter-Generational Analysis

For the analysis in which we study parental income, we can only look at a sub-sample of the IRS data for adults who are born after 1980, the earliest cohort for which we have sufficient records to match them to parents through 1040 forms that record dependents. For years when parents file tax returns, we calculate parents' income as the pre-tax household income; we gather parent income from W2 and other information returns in years when a parent does not file. Further details on the process of matching kids to parents are outlined in Chetty et al. (2014b). In this sample of individuals, which are aged 32 or younger in 2012, we have only about 10,000 young inventors.

II.E Test Score Analysis

When we look at the test scores of elementary school children, we further condition on a sub-sample where we know whether children attended a New York City public elementary school between 1989 and 2009. We observe standardized test scores in grades 3 through 8 for these children. Approximately 350 of these children grew up to become inventors, or roughly 0.04 - 0.06% of children born in the early 1980s and attending these schools. Further details of the process by which these students were matched to taxpayers can be found in Chetty et al. (2014a).

II.F Initial Data Description

Table 1 contains descriptive statistics from the whole sample. There are a total of 786,656 inventors and 12.5m observations over the 1996-2012 period in our matched data. The innovation data is highly skewed as is typical in this kind of data. The average inventor has just under 4 patents and garnered 36 citations over the period. But the standard deviation is enormous: 9.21 for patents

and 169 for citations. With innovation there are many duds and a few bonanzas (e.g. Pakes, 1986) and we will return to the implications of this high skewness later in the paper.

The annual wage of an inventor is just under $140,000 at the mean and $100,000 at the median, with incomes of over $214,000 (mean) and $122,000 (median). Since inventors are higher human capital people, it is unsurprising that they are richer than the population as whole. Interestingly, royalties and self-employment income are relatively low (only positive in the top decile), suggesting that returns to innovation are mainly in the form of wages (at least for majority of patentors). 14% of the sample had changed jobs in the last 3 years, 5% had changed commuting zone and 19.1% are foreigners. These are higher than for US workers as a whole. Consistent with Hunt (2009) only 11.6% are women.

Figure 1 shows the distribution of inventors by age (blue line) compared to the age distribution of US workers in the whole IRS population. It is clear that most inventions tend to come from inventors who are older than the working population as a whole - e.g. there are a lot of workers in their 20s, but very few of them invent. By contrast innovation peaks in the early 40s and then trails off gradually. In Figure 2 we show the number of patents per wage earner (i.e. the blue line in Figure 1 divided by the red line). The upward sloping age profile can be seen again which declines a bit from the early 40s. In other words, although there are more patents from workers aged 35-45 than aged 45-55 this is mainly because there are fewer workers. Age-related declining research productivity is not very strong as the falling slope after the early 40s is not very steep. For example the ratio of patents per worker is 18.6 at age 45 and 16.3 at age 65, a fall of 12%. Does this over-estimate the abilities of the middle-aged? For example, could the patents of older workers be of lower quality and the most important work is still done when researchers are young? Figure 3 uses inventors with patents that are in the top 20% of the citations distribution as a measure of quality (i.e. a patent had to receive at least 20 citations to be counted as an "important" innovation on this definition). The slope does become more steep than in Figure 2, with a 33% fall in the innovation per worker ratio between age 45 and 65. This is broadly consistent with the "young and restless" hypothesis of Acemoglu et al. (2014) and the evidence in academia of "great" discoveries (e.g. Jones et al. (2014)).

III Results

III.A Organizing Framework

Our organizing framework to look at the lifecycle of an innovator is in Figure 4. We first look at characteristics around the time of birth and early childhood; then at neighborhoods and the end of formal education in college and finally at earnings and employment.

III.B Birth and early experience: Inventors and parental income

III.B.1 Inventors, parental income and ability

Figure 5 shows that number of inventors per 10,000 individuals in percentiles of parental income distribution. A sharp, quite convex upward slope is apparent. For the lower three-quarters of the income distribution there are almost no inventors - a patent rate of under 5%. There is an average patent rate of 2.2. for those born to parents of below median income compared to a rate of 22.5 for those born into the top 1% of incomes, an order of magnitude difference.

There could be many reasons for the relationship in Figure 5. Since income is related to human capital and ability is partly genetic this relationship could be because of inherited innovative ability. To investigate this issue we turn to the New York City (NYC) schools data where we know test scores of children in grades 3 to 8. In Figure 6 we begin with looking at future innovation rates for children at different points of the ability distribution as measured by vingtile of math test scores in third grade. The pattern is striking. Kids in the top 5% of the ability distribution have a future innovation rate of over 20 (i.e. 20 patents per 10,000 people); those in the next 10% have innovation rates of just over 12 whereas those in the bottom 85% have innovation rates of 5 or less. Basically, unless kids are in the top 15% of the ability distribution there is little chance of them obtaining a patent as adults. Similar patterns are visible if we use other test scores (like English).[3]

Figure 7 then looks at the distribution of test scores by income, splitting parental income above and below the median and then plotting out the kernel density. Children from poorer backgrounds have lower test scores than those from richer backgrounds as the distribution is clearly shifted to the right for the wealthier kids. Hence there does seem to be some link between parental income and innovation operating through human capital. Another way of visualizing this is in Figure 8 where we plot the invention rate as a function of deciles of test scores split by whether a child is born to a poor or rich family. As before the relationship between future invention and early ability

[3]We can also instrument one test score with another to address the issue of noise in the text scores. See Kahneman (1962) for a discussion of the issues

is flat for those below average ability, but is clearly positive for those with test scores indicating above average ability. What is even more interesting is how this relationship differs by parental income. There is almost no innovation penalty for kids of equal ability except for children in the top 10% of the ability distribution. For the highest ability children in wealthier families the innovation rate is over 20 whereas for poorer families it is only 7.

How much of the innovation gap between rich and poor can be accounted for by differential ability in third grade? Table 2 provides the calculations based on the full distribution decomposition methodology of DiNardo et al. (1996). We give give the children below median income the same test scores as those with above median income across the whole distribution (we divide the distributions into vingtiles 5% bins - to do this). It is important to do this across the whole distribution rather than just at the mean as in a conventional Oaxaca-Blinder decomposition because most of the increase in innovation probability is in the right tail of the ability distribution. We calculate that the test score difference in third grade account for 16.7% of the difference in innovation propensities in later life. So this is a sizable proportion, but obviously much less than 100% (as shown in Figure 8). We can repeat these calculations, however, for every grade. Since we have currently only have data through grade 8 we show what fraction of the income-innovation gap we can explain by test scores at each age as shown in Figure 9. The fraction explained increases more or less monotonically as children go through school, by about 4.7 percentage points every year. By eighth grade 43% of the innovation-income difference can be accounted for by test scores. Extrapolating this forward would imply that by the time individuals would leave college the vast majority of the innovation-income relationship would be accounted for. Extrapolating this backwards to kindergarten implies that there is basically zero innovation-income gap when children start school. This suggests that the initial ability gap is not the reason for the innovation-income deficit.

What this decomposition suggests is that the key ingredient to understanding why kids from low income backgrounds do not become innovators is to do with the differences in attainment that open up when they are going through school from a very early age. At the end of formal education there appears to be no substantial income-related innovation deficit.

III.B.2 Inventors and gender

As we noted in Table 1, only 11% of inventors are female. Figure 10 slices this data in another way examining the proportion of inventors who are women by birth cohort. Clearly more women are becoming inventors over time: only 3% of inventors born in 1940 were female, whereas this

had risen to 15% for the 1980s birth cohort. This is a very slow rate of convergence, however. Extrapolating this line forward suggests it will take 140 years before women reach parity with their male counterparts.

We conduct an analogous decomposition exercise for women as we did for income in the previous sub-section (see Table 3). Figure 11 shows the distribution of math test scores for boys and girls in eighth grade. Although boys are slightly ahead of girls (this is less true in third grade and is the opposite way round for English with girls ahead of boys), what is striking about this distribution is how similar it is. Figure 12 shows that we can account for very little of the gender-innovation gap using math test scores.We account for nothing at all using third grade test scores and only about 3% using eight grade scores. So, in stark contrast to income, the gender gap in innovation does not appear to be ability related.

III.C Inventors as they grow up: Neighborhoods and College

III.C.1 Geography

There is a large literature[4] that examines the effects of place on economic outcomes in general and the effects on innovation in particular.Marshall's theory of "industrial districts" suggested that an advantage of geographical clusters was that there were "ideas in the air" and accounts of the success of Silicon Valley have also stressed the benefits of geographically local innovation spillovers. From our data we know where all individuals were born and in Figure 13 present the innovation rates by state using the future inventors born in the state. Note that the Figure does not present the patenting rates based on where inventors are *currently* residing (which is what is generally used because it is available in the USPTO inventor location), but rather the patenting in the states where inventors were born. Figure 13 shows an interesting spatial pattern. Innovation hotspots are in the North-East and Great Lakes which is expected as there are strong universities located there, but also in the North-West, Utah and Colorado which is not so obvious. Very low innovation states are those in the Deep South.

Figure 14 shows that children who grow up in innovation-intensive states are more likely to become innovators themselves. The correlation is 0.739 with a standard error of 0.149. Of course, one cannot take this relationship as causal as there are many other unobservable factors associated with place of birth and future outcomes. We probe this relationship a bit more in Figure 15 where

[4]There is a large literature on inventor mobility and agglomeration effects from localized knowledge spillovers. For example see Fernandez-Zubieta et al (2013), Hoisl (2007), Kaiser et al (2011), Kim and Marshke (2005), Marx et al (2009), Moene (2005), Saxenian (1994), Shalem and Trajtenberg (2009) and Stoyanov and Zubanov (2012).

we split innovations up into the 6 main technological categories in the USPTO and look at state-specific innovation within these classes.[5] Interestingly we see that the correlation survives even within technological class (slope 0.358 with standard error of 0.102). In other words, growing up in Minnesota (which is intensive in medical innovation) is associated with more individuals who become medical inventors (relative to computer innovations) than growing up in California which is relatively intensive in computer related innovation relative to medical innovation.

III.C.2 College

For the individuals who attended a US college between 1999 and 2012 (200,000 inventors) we can determine which college they attended. For inventors who were college graduates (the vast majority) we show that a few colleges produce graduates who make a remarkable number of successful innovations. Figure 16 shows, for example, that 23% of all citations come from just ten US schools (which account for only 2.7% of all college enrollees). A third of all citations come from the top twenty schools. This reinforces the earlier finding that educational attainment is likely to be the key factor linking parental background to whether or not a person grows up to be an inventor.

III.D Inventors and labor markets

III.D.1 Wages and Income

Figures A2 and A3 show age-income and age-wage profiles for inventors with different levels of patent citations. The usual concave pattern is evident over the live cycle for income and wages. Wage and income growth are higher for more successful inventors. This could be because of their innovations causally increasing wages, or it could be just a reflection of the fact that high citations reflect underlying higher human capital is observed by the market.

To investigate this more carefully, we focus on event studies before and after a (subsequently granted) patent application. Figure 17 gives an example of the analysis where we look at mean wages for patents broken down by their future citations. The event window is up to three years before the patent application and three years after the application. The bottom (blue) line are for low value patents with zero citations. Inventor income rises steadily before and after the patent event with no sign of a break in the trend. For higher value patents the picture is quite different, however. For patents which receive over 30 citations (top 1% of citation distribution, middle red line) or over over 50 citations (top 0.1% of citation distribution, top green line), there is a clear

[5]The classes are computers and communications; pharmaceuticals, non-pharmaceutical chemicals, engineering; electrical machinery and others.

break in the wage trend, with a big uptick to wages after the inventor patents. Note that the more successful the inventor the higher was his pre-patent wage as we might expect, but the thing we want to emphasis here is the change in the trend rather than the level.

The higher wages at the mean seem to be strongest at the top of the distribution. Figure 18 presents event studies for high value patents (95th percentile of citation distribution) for different quantiles of wages. The increase in the mean shown in Table 17 is clearly driven by the upper part of the wage distributions, with the main effects at the 75th percentile of wages and above (Panel B compared to Panel A). We obtain broadly the same patterns if we focus on gross income which sums over wages and other sources of income from capital, self-employment and royalties.

There are a variety of ways that income could increase after patenting. The worker could earn a bonus or promotion in the firm where s/he did the invention. But alternatively the worker could switch firms and obtain a higher wage in the new firm - a reward on the external labor market. Thirdly, the inventor could form his/her own start-up and exploit the technology. Fourthly, the inventor could license/sell the technology (if s/he has IP rights) and earn a return in this way. Empirically, all seem to matter to some degree. For example, royalties increase for inventors, but only if the patent rights are assigned to the individual (e.g. if the rights are assigned to the firm we do not see an increase in inventor royalties). But our preliminary investigation suggests the main returns come from on the same job.[6]

Looking at heterogeneity of the wage effects we currently find that things look broadly similar when cutting the data by gender, age, foreign or US citizen.

III.D.2 Inventor Mobility

Examining the probability of moving employers after patenting we find the interesting result that following a patent event inventors are *less* likely to switch employers. The more valuable the patent as indicated by future citations, the stronger is this effect. This might be surprising as the anecdotal view is that inventors can capitalize their knowledge by forming a start-up or moving up to a higher paying job. In fact it appears that part of the reward to invention is a greater chance of staying at the same job (and obtaining higher pay on the job). Similarly we do not find increases in geographic

[6] A concern with Figures 17 and 18 is that there appears to be a dip in wages in the pre-patent period for high value and higher wage quantiles (see also the patterns in Figures A3 and A4). We are investigating whether this is (i) a genuine economic phenomenon such as an inventor diverting work in research activity and away from "production" ; (ii) an adverse negative shock such as losing a lay-off which stimulates innovation (akin to the "virtues of bad times" models of e.g. Bloom, Romer, Terry and Van Reenen, 2014) or (iii) some other statistical phenomenon which is causing us to over-estimate the income returns.

mobility (between commuting zones) after patenting.

III.D.3 The firms for which inventors work

An advantage of the data we have is that we know what firms inventors work in from the EIN number attached to their W2s. Corporate structures are complex, however - for example, a US parent company may have multiple affiliates, the ultimate headquarters may be in another country and there are partnership structures as well as firms. For now, we simply work with the EIN numbers and discuss some of the data issues in Appendix B. There is a large literature on the distribution of innovation by firm size, so we begin the analysis here using employment as a measure of size as is conventional and intuitive. Figure 19 compares the distribution of employment by firm size class in the IRS CDW dataset, where each firm is assumed to be represented by a unique EIN, with the size distribution across the US economy from the Economic Census. Despite many reasons why the Census Bureau concept of an "enterprise" could be quite different from the tax based EIN, the match with the firm size distribution is very good. There appears to be a slight under-representation of firms with under 100 employees in the IRS data compared to the Census, and a slight over-representation in the 100-1000 employee range, but almost identical proportions in other size bins compared to the Census. And even the discrepancies are small. Narrowing the IRS sample to be closer to the Census in terms of industry composition does not fundamentally change this picture.

Figure A5 shows innovation by size class, but it is easier to see the patterns by dividing into deciles by employment, so that between each interval there are 10% of all employees (as in Figure 20). The exact size thresholds for the bins are given in Table A1 (for example the lowest bin has between 1 and 17 workers and the highest decile is of firms over 100,000 employees). Figure 20 shows innovation by firm size. There is the usual positive relationship between size and innovation, as firms become larger, the ratio of patents per employee becomes bigger. However, the relationship is relatively flat: up to the sixth decile there is only a slightly lower fraction of patents and citations in firms relatively to the number of employees. Firms in deciles 7, 8 and 9 have more patents and citations than those in lower deciles. This increase in patents per employee is not monotonic, however. It starts to fall off in the 9th decile and indeed, firms in the top decile produce less patents (and even less citations) than those in all other deciles.

The distribution of invention by firm size in our data is similar to what we observe in other near population datsasets where patents are matched to their applicants rather than by inventor

as we do (e.g.Akcigit and Kerr (2013)). We are able to disaggregate within the larger firm class, however, and this is where we observe a falling off of patent rates for the very largest firms. There does, however, appear to be a larger fraction of self-assigned patents in small firms (see Table A6) in our data which would lead to an underestimate in the amount of innovation in small firms in conventional datasets. However, the magnitude of this is not large as the fraction of self-assigned is small everywhere. There are also more single inventor patent applications in smaller firms than larger firms (again except for the largest firms) - see Table A7.

From the IRS data we also know which firms receive the R&E tax credit and how much they received. Figure 21 compares the proportion of our firms receiving any R&E tax credit compared with those with inventors. Only about 10% of firms receive the tax credit whereas about 27% of firms have at least one patent. Furthermore, not all the firms who receive the tax credit produce any patents. Figure 22 suggests only about 10% of firms who patented received any tax credit. This may be an underestimate as an EIN may be part of a larger consolidated group filing that did claim some tax credit. Nevertheless, it is clear that there are a larger number of seemingly innovative firms not receiving support (and vice versa).

III.E Some Further Analysis

A disadvantage of the data is that the first patents we observe are in 1996. Linking to an individual's entire history from 1964 (when the USPTO is in usable electronic form) requires name disambiguation. Fortunately, we can draw on Lee Flemming's work here. A concern is that we might be linking the wrong individuals over time (see Trajtenberg et al. (2006), for a discussion) solely using the patent data. The advantage of the IRS data is that once we have a match we know it is the same individual over time from their Social Security number. We are therefore checking the Flemming method within our sample to verify match quality and then using the individual's entire innovation history. We have generated many other results that are of interest in describing the inventor lifecycle. We have relegated these to Appendix C.

IV Discussion and implications for Innovation Policy

Our picture of inventor life-cycles has revealed many stylized facts of interest to researchers and policy makers. We believe that we are the first to document a strong relationship between parental income and future innovative success. Decomposition analysis suggests that this is working through education. Human capital as indicated by test scores diverges sharply between those from poor

and rich backgrounds as they progress through school and the importance of measured test scores for later outcomes also increases. The achievement gap in human capital is the main reason why so many more innovators come from wealthy backgrounds than from poorer backgrounds. If interventions could improve the position of potentially smart kids from poor backgrounds this would bring a whole new margin of individuals into the innovation pool who are being potentially lost. Card et al. (2013) report evidence from "gifted and talented" randomized control trials that suggest that although these programs do not work well for the typical student, those from poorer backgrounds do appear to particularly benefit. Changing to such policies has almost a zero financial cost in Card et al. (2013). This suggests that such interventions could have very large benefits in terms of growth as well as equity.

By contrast to these supply side policies, policies on the demand side may be much less effective. Although existing estimates of R&E tax credits do suggest increases in innovative activity (e.g. Bloom,. Schankerman and Van Reenen, 2013), they reach only a small fraction of the necessary population. We can examine this more carefully by calibrating a simple life-cycle model of innovation and tax rates. Appendix B sketches such a model, whose key feature is an examination of the marginal effects of top taxes on the incentive of individuals to become inventors. Although increasing top tax rates have a negative effect on the incentives to become an inventor, the magnitude of such effects is likely to be tiny. Calibrating the earnings return to innovation as we do in the event study confirms this intuition.

V Conclusions

In this paper we describe the lifecycle of inventors using a match between 780,000 US patentors during 1996-2012 with administrative tax data from the IRS. We are able to follow inventors from birth to their labor market career: from their conditions at birth (e.g. parental income, gender); while growing up (e.g. neighborhood, college attended) and finally in the labor market (e.g. their income profiles before and after inventing; the firms they work in). We uncovered some fascinating patterns. In particular, we find that children from high income families are much more likely to grow up to be innovators than those from less fortunate backgrounds. This appears to be closely linked to how their human capital accumulates from the time they start attending school to the time they enter the labor market. Using test score data matched to inventor and income records we show decompositions suggesting that the opening up of the test score gap between low and high socio-economic status groups accounts for about 5% of the innovation gap in each grade.

Policies to reduce the income-related test score divergence (particularly for kids in the top decile of achievement) are not only beneficial from an equity stance, they are also beneficial from an efficiency perspective as they have the potential to increase the degree of innovation and growth in the US economy. We argue that such "supply side" innovation policies may be much more effective in stimulating innovation than more traditional demand side policies through changing the tax code.

We also document a wide number of other new, stylized facts about the lifecycle of inventors that shed light on long-starting questions from a theoretical, empirical and policy perspective in the economics of innovation.

References

Acemoglu, D., U. Akcigit, and M. Celik. 2014. ""Young, restless and Creative: Openness to Disruption and Creative Innovations"." *NBER Working Paper No. 19894 [currently cite mimeo].*

Akcigit, U. and W. Kerr. 2013. ""Growth through Heterogeneous Innovations"." *Bank of Finland Discussion Paper No. 1328.*

Azoulay, P., Graff Zivin J.S., and Manso G. 2010. "Incentives and Creativity: Evidence from Academic Life." *RAND Journal of Economics.*

Azoulay, P., J. S. Graff Zivin, and J. Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics, 125 (2), 549-589.*

Bloom, N., M. Schankerman, and J. Van Reenen. 2013. "Technology Spillovers and Product Market rivalry." *Econometrica (2013) 81 (4) 1347–1393.*

Card, D., J. Heining, and P. Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics, 128(3): 967-1015.*

Chetty, R., J. N. Friedman, and J. E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review, American Economic Association*, *104* (9): 2593–2632.

Chetty, R., N. Hendren, P. Kline, and E. Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." (19843).

DiNardo, J., N. M. Fortin, and T. Lemieux. 1996. ""Labor Market Institutions and the Distribution of Wages, 1973-1992: A semiparametric Approach"." *Econometrica, Econometric Society, vol. 64(5), pages 1001-44, September.*

Dorner, M., D. Harhoff, and K. Hoisl. 2014. "Patterns and Determinants of Inventor Mobility – Evidence from the Employment Biographies of Inventors in Germany." *mimeo.*

Franco, A. M. and M. F. Mitchell. 2008. "Covenants not to compete, labor mobility, and industry dynamics." *Journal of Economics and Management Strategy 17, 581–606.*

Goolsbee, A. 1998. ""Does Government R&D Policy Mainly Benefit Scientists and Engineers?"." *American Economic Review, 88(2), 298-302.*

Hunt, J. 2009. "Why Don't Women Patent?" *NBER Working Paper No. 17888.*

Jones, Benjamin F. 2010. "Age and Great Invention." *Review of Economics and Statistics.*

Jones, B.F., E. J. Reedy, and B.A. Weinberg. 2014. "Age and Scientific Genius." *NBER Working Paper No. w19866.*

Lai, R., A. D'Amour, D. M. Doolin, G.-C. Li, Y. Sun, V. Torvik, and A. Yu. 2014. "Disambiguation and co-authorship networks of the U.S. patent inventor database (1975-2010)." *Research Policy, Elsevier*, *43* (6): 941–955.

Pakes, A. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica,54(4), 755-84.*

Pakes, A. and S. Nitzan. 1983. "Optimum Contracts for Research Personnel, Research Employment, and the Establishment of Rival Enterprises." *Journal of Labor Economics, 1(4), 345-65.*

Rao. 2013.

Romer, P. 2000. ""Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers ?"." *NBER Chapters, in : Innovation Policy and the Economy, Volume 1, pages 221-252.*

Scotchmer, S. 2004. "Innovation and incentives." *Cambridge: MA Free Press.*

Toivanen, O. and L. Vaananen. 2012. "Returns to Inventors." *Review of Economics and Statistics, 94(4), 1173-1190.*

Trajtenberg, M., G. Shiff, and R. Melamed. 2006. "The "names game" : harnessing inventors' patent data for economic research." *NBER Working Paper No. 12479, Cambridge, MA.*

Zucker, L. G., M. R. Darby, and M. B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review, 88(1), 290-306.*

APPENDICES

A. Data Construction

A.1 Data Preparation

- **Suffix Standardization**. Suffixes may appear at the end of taxpayers' first, middle, or last name fields. Any time any of these fields ends with a space followed by "JR", "SR", or a numeral I-IV, the suffix is stripped out and stored separately from the name[7].

- **First name to imputed first/middle name**. The USPTO separates inventor names into "first" and "last," but the IRS often separates names into first, middle, and last. In practice, many inventors do include a middle initial or name in the first name field. Whenever there is a single space in the inventor's first name field, for the purposes of matching, we allow the first string to be an imputed first name, and the second string to be an imputed middle name or initial. The use of these imputed names is outlined below.

A.2 Pseudocode for Match on Name and Location

The exact matching stages are as follows. We conduct seven progressive rounds of matching. Inventors enter a match round only if they have not already been matched to a taxpayer in an earlier round. Each round consists of a name criterion and a location criterion. The share of data matched in each round is noted, with an impressive 49% being exact matches on the first stage.

- The matching algorithm takes as input a relation of inventor data and five relations of IRS data:

  - Input relations:

    * Inventors(inv_id, first, last, imputed_first, imputed_middle, suffix) - directly from USPTO
    * NamesW2(irs_id, first, middle, last, suffix) - all names used by individual on W2 information returns; name field is recorded as first, middle, and last
    * Names1040(irs_id, first, middle, last) - all self-reported names from 1040 forms[8]
    * Nameln1W2(irs_id, fullname) - all names from W2, but a separate variable not recorded as first, middle, last that was more frequently present

[7] Numerals I and V are only permissive suffixes at the end of a last name field, as these may be middle initials in a middle name field.

[8] We only take names off of 1040s for those who file singly because it proved difficult to parse names of those list them jointly

    - CitiesW2(irs_id, city, state) - all cities reported on W2
    - Zips1040(irs_id, name) - all zip codes reported on 1040
  - Output relation:
    - Unique-Matches (inv_id, irs_id)

- **Stage 1**: Exact match on name and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. Either the inventor's first name field exactly matches the concatenation of the IRS first and middle name fields or the IRS middle name field is missing, but the first name fields match. If an imputed middle name is available for the inventor, candidate matches are removed if they have ever filed at the IRS with a middle name or initial that conflicts with the inventor's.
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 49% of patents are uniquely matched in this stage.

- **Stage 2**: Exact match on imputed name data and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name and the taxpayer's last name is the same as the inventor's imputed first name. Either the inventor's imputed middle name/initial matches one of the taxpayer's middle/initial name fields, or one of the two is missing. For inventors with non-missing imputed middle names, priority is given to matches to correct taxpayer middle names rather than to taxpayers with missing middle names. As above, candidate matches are removed if they have ever filed at the IRS with a conflicting middle name or initial.
  - Location match: As above, the inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 12% of patents are uniquely matched in this stage.

- **Stage 3**: Exact match on actual or imputed name data and 1040 zip crosswalked.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. The inventor's first name matches the taxpayer's first name in one of the following situations, in order of priority:

1. Inventor's firstname is the same as the taxpayer's combined first and middle name.
2. Inventor's imputed firstname matches taxpayer's and middle names match on initials.
3. The inventor has no middlename data, but inventor's firstname is the same as the taxpayer's middle name.

   - As always, taxpayers are removed if they are ever observed filing with middle names in conflict with the inventor's.Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
   - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
   - 3% of patents are uniquely matched in this stage.

- **Stage 4**: Same as previous stage, but using 1040 names instead of names from W2's.
  - Name match: The inventor's name matches the name of a 1040 (or matches without inventor's middle initial/name, and no taxpayer middle initials/names conflict with inventor's).
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 6% of patents are uniquely matched in this stage.

- **Stage 5**: Match using W2 full name field.
  - Name match: The inventor's FULL name exactly matches the FULL name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 8% of patents are uniquely matched in this stage.

- **Stage 6**: Relaxed match using W2 full name field.
  - Name match: The inventor's full name (minus the imputed middle name) exactly matches the full name of a taxpayer on a W2.

  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 1% of patents are uniquely matched in this stage.

- **Stage 7**: Match to all information returns.
  - Name match: The inventor's full name exactly matches the full name of a taxpayer on any type of information return form.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's information return forms.
  - 6% of patents are uniquely matched in this stage.

B. A simple model

We sketch a simple lifecycle model of invention with an emphasis on how tax incentives such as cuts in the top marginal tax rates, might affect these incentives. Consider a two period model. In period 1, individuals acquire human capital through a mixture of permanent genetic ability, environmental inputs (such as school and neighborhood) and a stochastic shock. For now, assume these are exogenous. They then enter the labor market in period 2. There are three sectors in the labor market: a low skilled sector, an high skilled "R&D" sector and a high "non-R&D" sector. Expected wages are higher in the high skilled sectors. If human capital is below an exogenous threshold, workers are unproductive in the high skill sectors, so will work in the low skilled sector. If a worker's human capital is above a threshold she has the choice of whether to work in the R&D or non-R&D sector.

How will different levels of tax (or subsidy) effect the decision of whether or not a high skilled worker will enter the R&D sector? In the R&D sector there is a probability p of innovating leading to a pay-off of π and a probability 1- p of not innovating leading to the pay-off w. So Expected utility E(U) is

$$E(U_i) = pu(\pi) + (1-p)u(w)$$

The utility that the individual i obtains in the non-R&D sector is heterogeneous according to individual taste which is idiosyncratic across individuals. Denote this utility as the "outside option", $\widetilde{U_{0,i}}$ with respect to the R&D sector. A worker will choose to work in the R&D sector if $E(U_i) > \widetilde{U_{0,i}}$, or equivalently $pu(\pi) > \widetilde{U_{0,i}} - (1-p)u(w)$. Denote $U_{0,i} = \widetilde{U_{0,i}} - (1-p)u(w)$ and assume $U_{0,i} \sim F(U_{0,i})$. Given this set-up, the probability of an individual joining the R&D sector is $\phi = F(pu(\pi))$, i.e. a function of the probability of innovating and the utility from the pay-off if the individual is successful.

We parametrize the pay-off to innovation as earnings of tax,$\pi = (1-\tau)\pi_0$ where τ is the tax rate and π_0 is gross earnings. Assume for simplicity that the tax rate is only relevant above a certain income threshold which is higher than the individual would earn in the non-R&D sector and if he doesn't innovate.

Consider the marginal impact of lowering the tax rate on probability of entering the R&D sector:

$$\frac{\partial \phi}{\partial(1-\tau)} = \frac{\partial[F(pu(\pi)]}{\partial(1-\tau)} = f(pu(\pi))pu'(\pi)\pi_0 = f_0 p\pi_0 u'(\pi)$$

If we denote expected revenue gross of taxes as $R = p\pi_0$, then the effect of taxes on the incentive to enter the R&D sector asymptotes to zero as the p goes to zero holding R fixed if u(.) is concave. The intuition for this result is simple. If the "prize" for innovating is very large, but the probability of successfully achieving it is very low, then in the presence of concave utility it makes little difference to innovation incentives if the individual has a slightly smaller amount of a very large prize in the (low probability) event that they win it.

To see this more clearly, consider the example of CRRA utility, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ so that marginal utility of consumption, c, is $u'(c) = c^{-\gamma}$. Then

$$\frac{\partial \phi}{\partial(1-\tau)} = \frac{f_0 p\pi_0}{\pi^\gamma(1-\tau)^\gamma}$$

If log utility then $\gamma = 1$ and

$$\frac{\partial \phi}{\partial(1-\tau)} = \frac{f_0 p}{1-\tau}$$

which converges to zero at rate p. As the probability of innovating becomes lower, the marginal effect of taxes on innovation becomes insignificant.

We can use some of our existing knowledge to bound the effects. In the conventional labor supply elasticity case we assume p=1. In this case we have estimates of the elasticity ε_c and a we have an upper bound for this (say 1). Now the elasticity we are interested in is:

$$\varepsilon = \frac{\partial \phi}{\partial(1-\tau)} = \frac{f_0 p}{1-\tau} \leq p\varepsilon_c$$

So if p = 1% an upper bound for ε is 0.01, very low.

It is easy to generalize this to the continuous case (instead of a binary model) to show that the effect of taxes on innovation are lower when the returns to innovation are more skewed (as they are in the data).

The key aspect of the result is that for very skewed pay-off function, taxes will matter less than for less skewed ones in the presence of concave utility. Given the data on earnings before and after the patent event, we can characterize the distribution of returns to an individual of innovating. Hence, we can calibrate the model by using the distribution of returns and show that there is quantitatively only a very small effect of top taxes on incentives to innovate. This is consistent with some of the evidence in Akcigit, Baslandze and Stantcheva (2015) who find that differences in international taxation have a statistically significant, but economically small effect on inventor mobility across OECD countries.

C. Further Results

Citations conditional on being an inventor

We have argued that some of the innovation-parental income relationship in Figure 5 could be because of financial constraints having a causal impact on children's probability of becoming inventors (rather than inherent ability or preferences at birth). An implication of the financial constraints story would be that some talented low income children are not living up to their potential to become inventors. This could be because they end up in worse schools that do not develop their potential to become inventors. Another way to investigate this is to examine citations per inventor split by parental income. In Figure A1 we examine the fraction of inventors who manage to produce high value innovations (as indicated by having a patent in the top quintile of the citation distribution). There is a clear downwards sloping line - inventors from poorer families appear to be far more successful than those from richer ones. If the picture in Figure 1 was all driven by inherited ability we would not expect to see such a downward slope. Indeed, we would be more likely to see an upwards sloping line as the most wealthy - and therefore smarter - parents produced the smartest

inventors.

Co-Invention Patterns

Most inventors are not "lone scholars" but in teams of two or more co-inventors. Of our 1,375,587 patents 55% are in teams and we have a total of 362,000 teams. The age, wage and income distributions of these teams are in appendix Table A2 to A5. One might imagine that there is assortative mating in the sense that inventors who are similar in characteristics like age and rank would tend to work together. This is true, but only up to a point: there is quite a wide dispersion in characteristics within teams. Consider two person teams, which account for a quarter of all teams. Table A2 shows that the most common combination of ages is actually those in the 9th and tenth deciles (average ages 55 and 64) and the next most common is the "young guns" in the first and third deciles (average ages 28 and 33). In fact, two person teams where there is at least a 10 year age gap (about 4 deciles) account for almost half of all these teams (Table A3). Tables A4 and A5 conduct the same analysis for wages. Again, the most frequent combinations are in the top and bottom of the wage hierarchy, but not in exactly the same deciles. Table 6 shows that only 17% of teams are in the same decile and over 42% of teams are in wage deciles four deciles apart (about $60,000), which is quite a large spread.

Repeating the analysis for larger teams (3,4,5, etc co-inventors). Shows a similar pattern that there is great diversity in the team characteristics.

Continuing to look at teams Table A7 shows that over a quarter of teams are spread across two firms. Indeed, when we get to teams of 5 almost a half of all patents have inventors in more than one firm and 28% have inventors spread across 2 employers (2% actually have 5 different firms!). Figure 8 shows exactly the same pattern when we condition on the 20% of most highly cited patents.

This is a remarkable degree of collaboration across inventors as usually it is assumed that since the property rights of a patent are usually assigned to a single firm, all the inventors work in the same firm. This does not seem to be the case. This also implies that the view that patenting is heavily concentrated in large firms may be misleading. If many of these are collaborations between smaller and larger firms where the larger partner receives the IP rights, but the smaller firm is fully involved in creating the innovation, this would generate quite a different picture to the conventional view.

With better firm data we can also match in a number of firm characteristics (such as profits and

productivity measures). We can then investigate what happens when a firm receives an exogenous increase in innovation (due, for example to changes in past rules for R&E tax credits as in Bloom et al. (2013); Rao (2013), etc.). We expect from previous work that firm profits and productivity rise, but to what extent are these shared with employees? The above spillover analysis can indicate how different types of worker benefit (including the inventor themselves), but with firm data we will also be able to see the extent to which innovation rents go to holders of capital.

FIGURE 1: Age Distribution of Patent Applicants W-2 Wage Earners, 2000

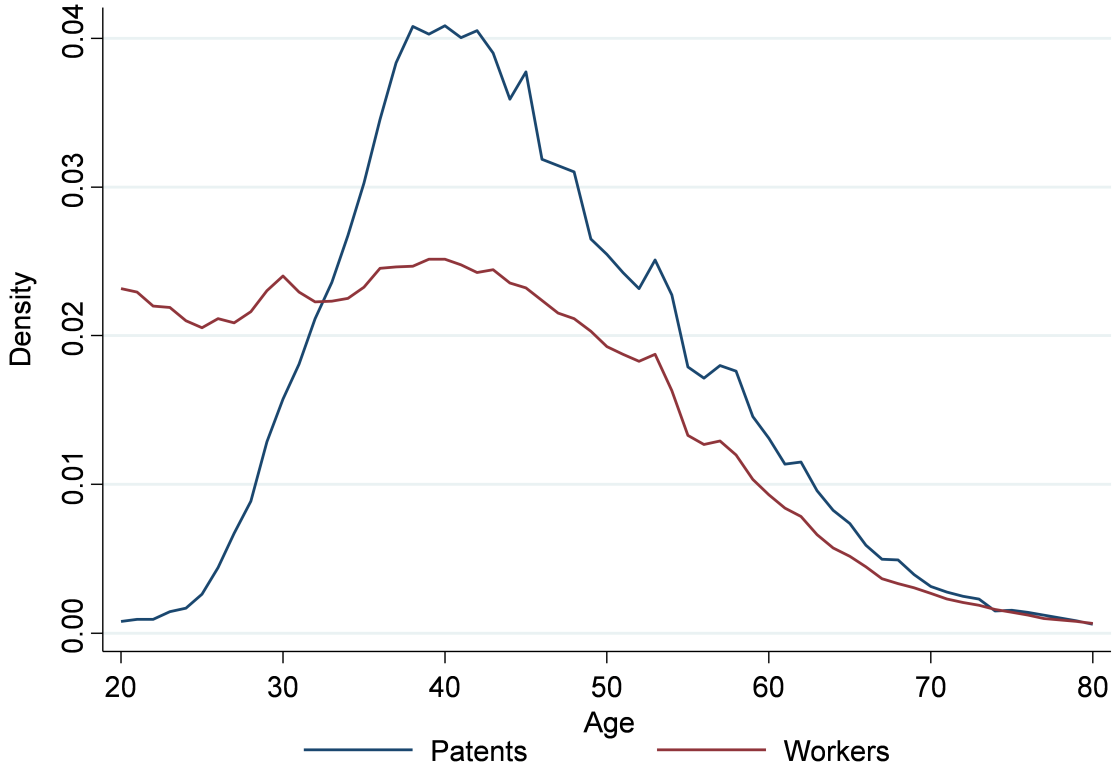


FIGURE 2: Patents Per Wage Earner by Age in 2000

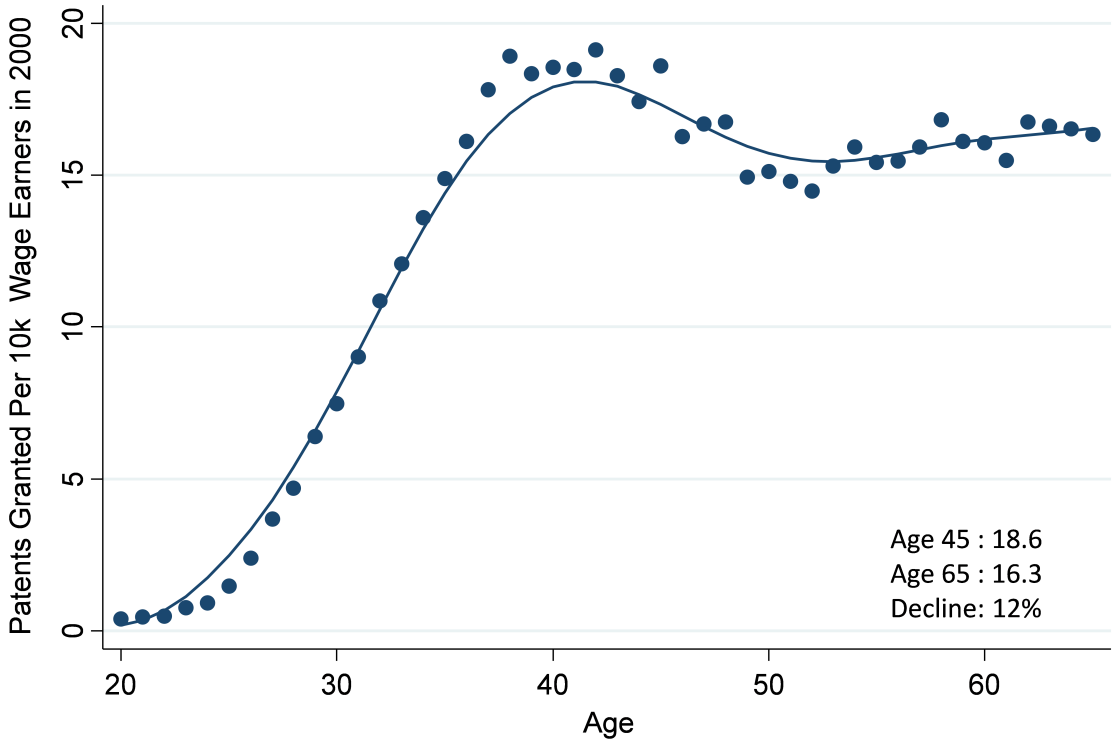


FIGURE 3: Highly Cited (Top 20%) Patents Per Wage Earner by Age in 2000

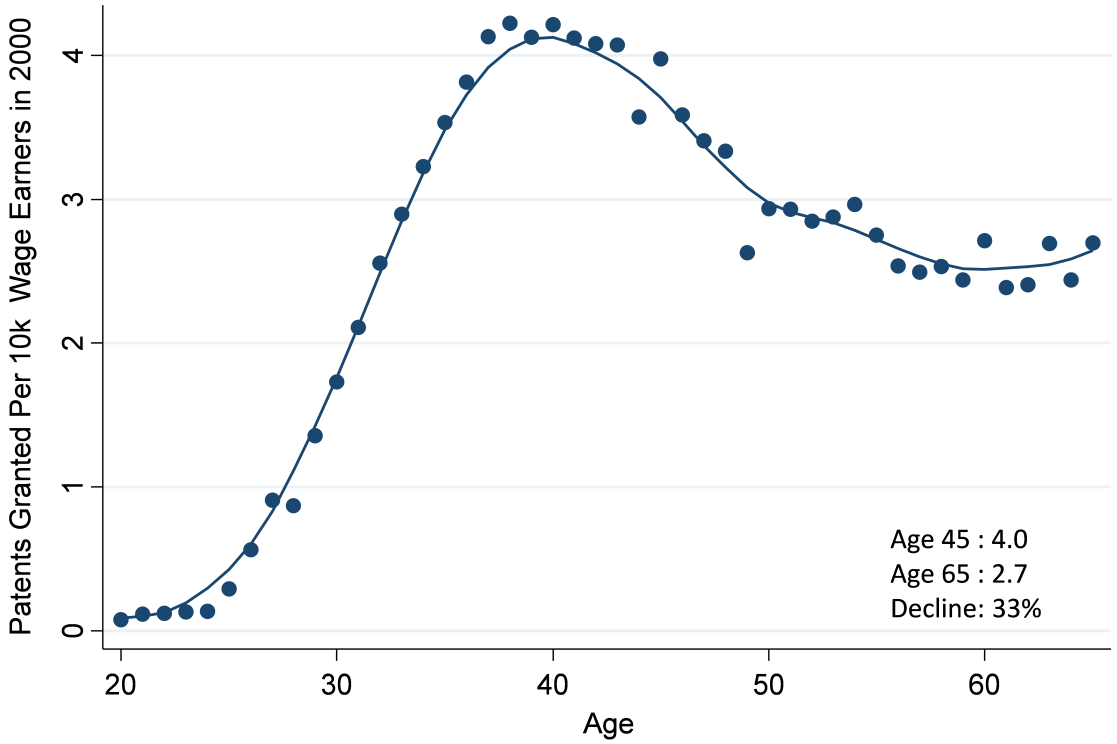


FIGURE 4: Lifecycle of an innovator



FIGURE 5: Patent Rates vs. Parent Income

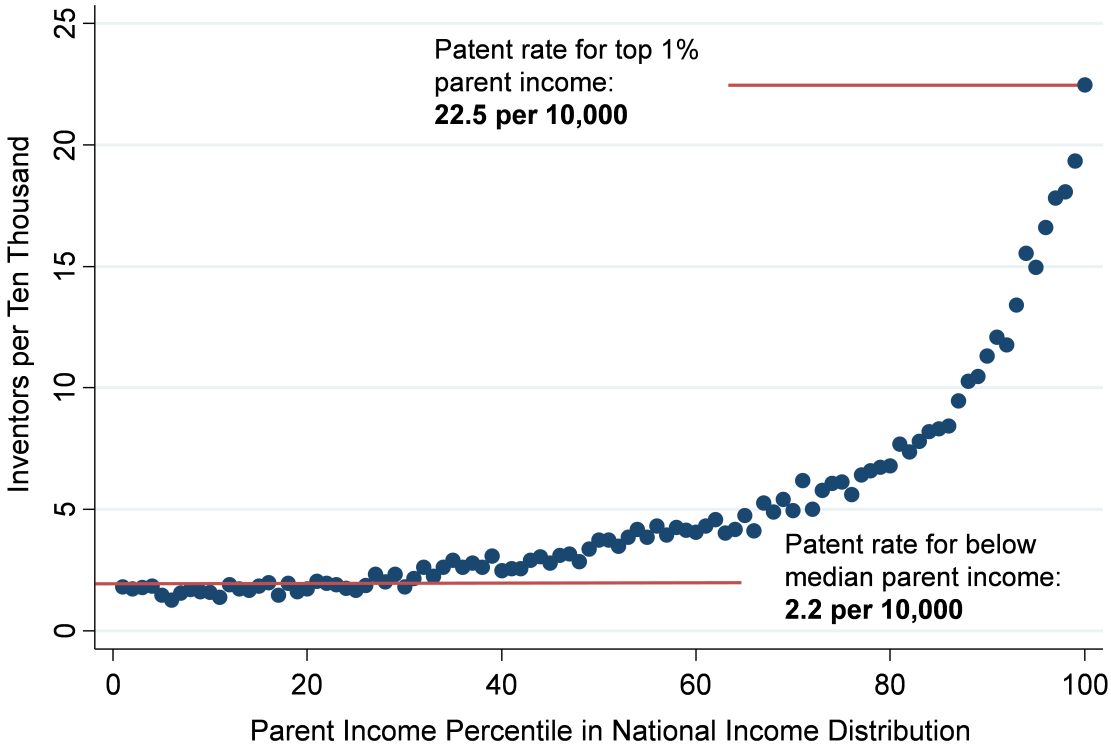


FIGURE 6: Patent Rates vs. 3rd Grade Test Scores

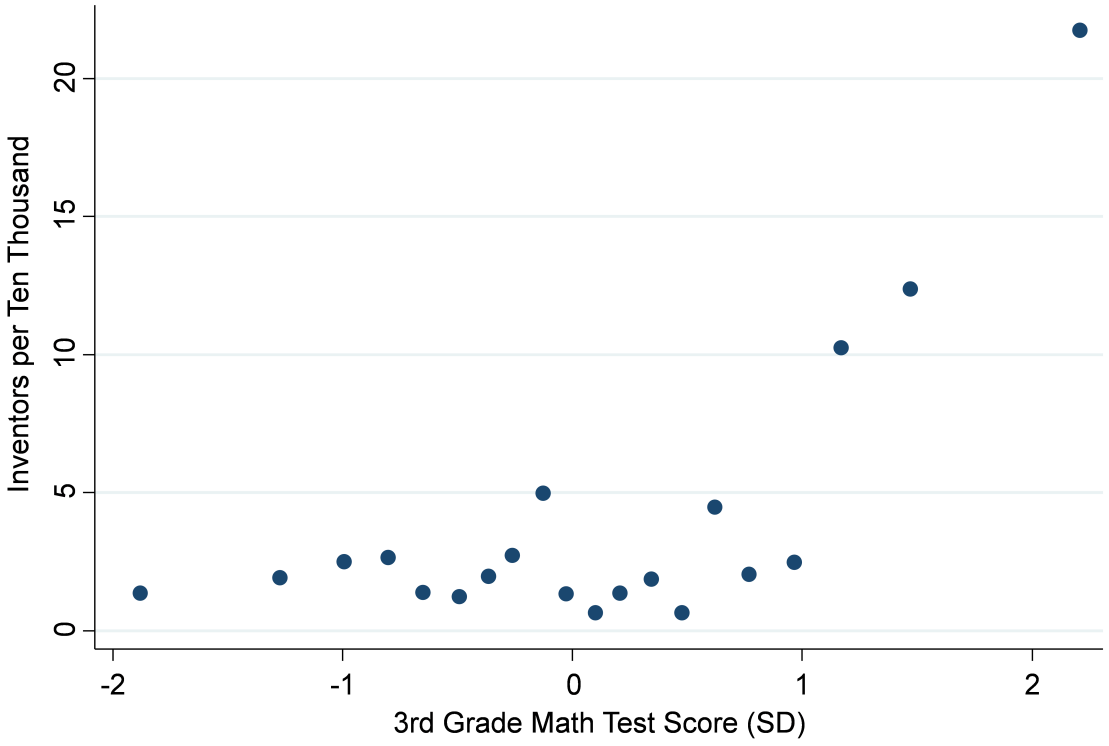


FIGURE 7: Distribution of Math Test Scores in 3rd Grade for Children with Low vs. High Income Parents

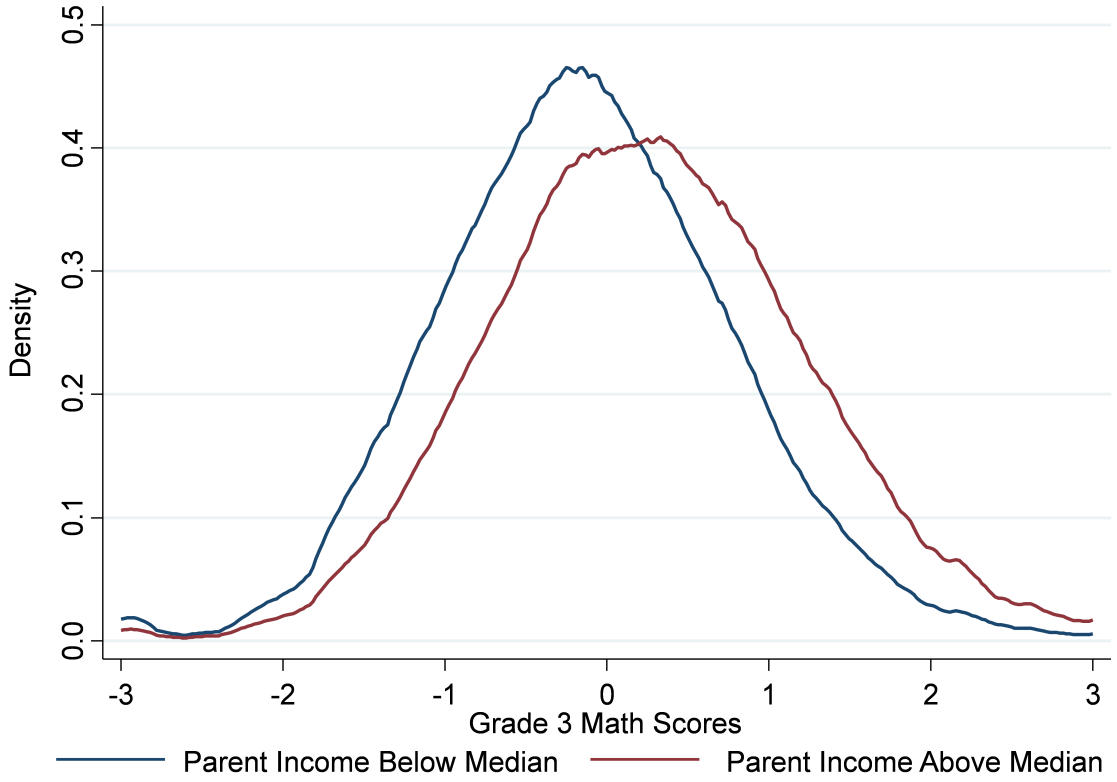


FIGURE 8: Patent Rates vs. Third Grade Test Scores for Children with Low vs. High Income Parents

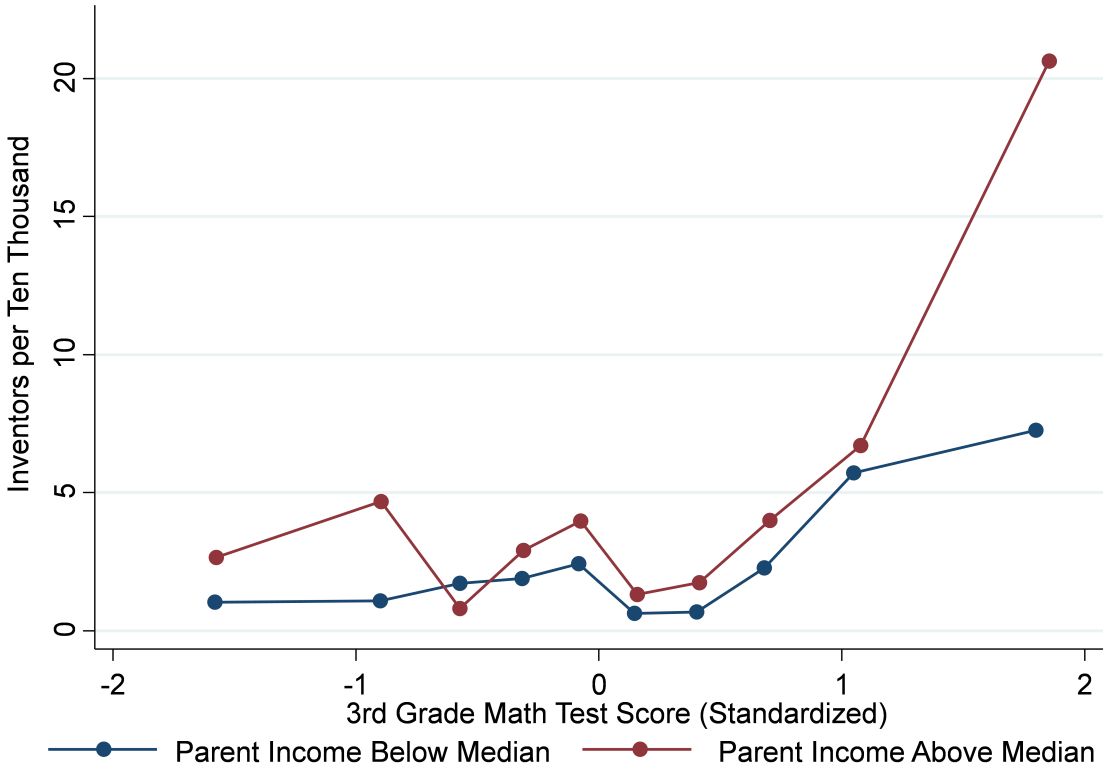


FIGURE 9: Percent of Gap in Patent Rates Between Low vs Hihg-Income Students Explained by Test Scores in Grades 3-8

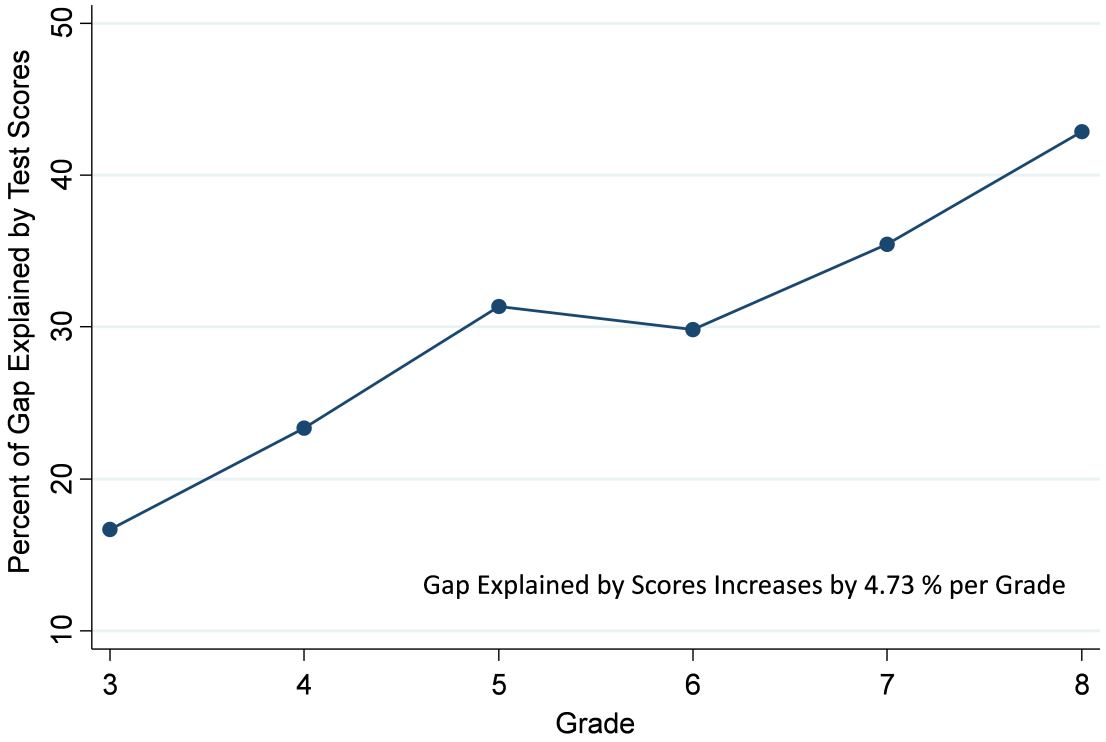


FIGURE 10: Percentage of Female Inventors by Birth Cohort

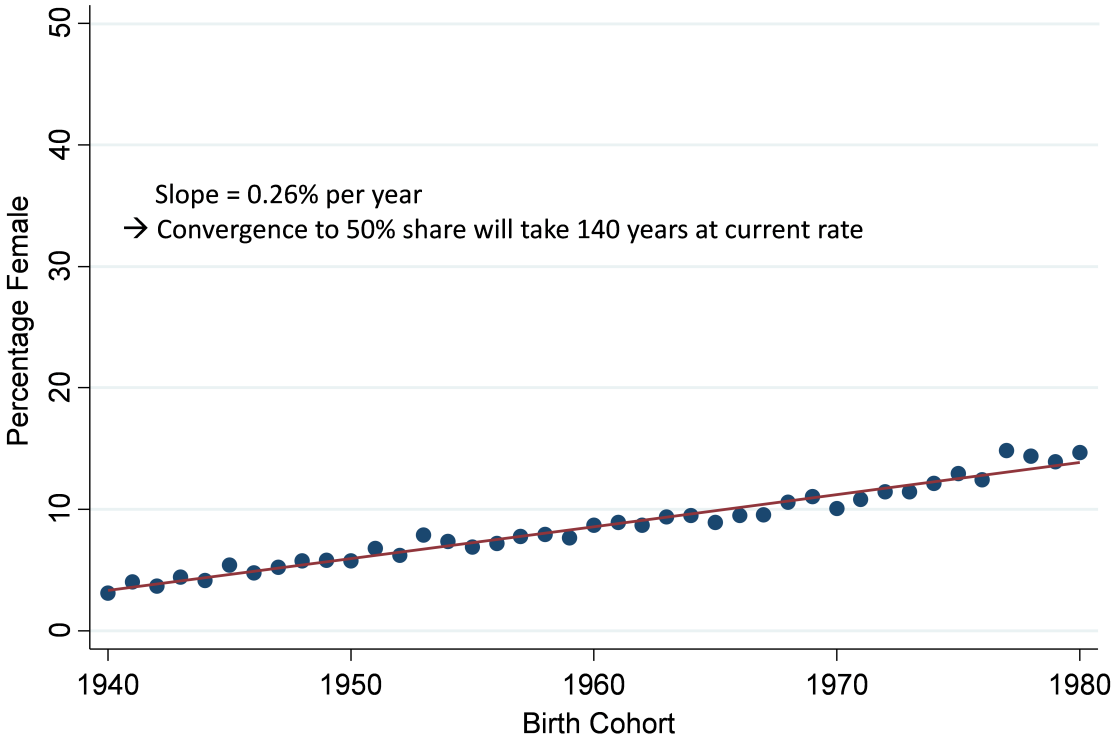


FIGURE 11: Distribution of Math Test Scores in 8th Grade for Males vs. Females

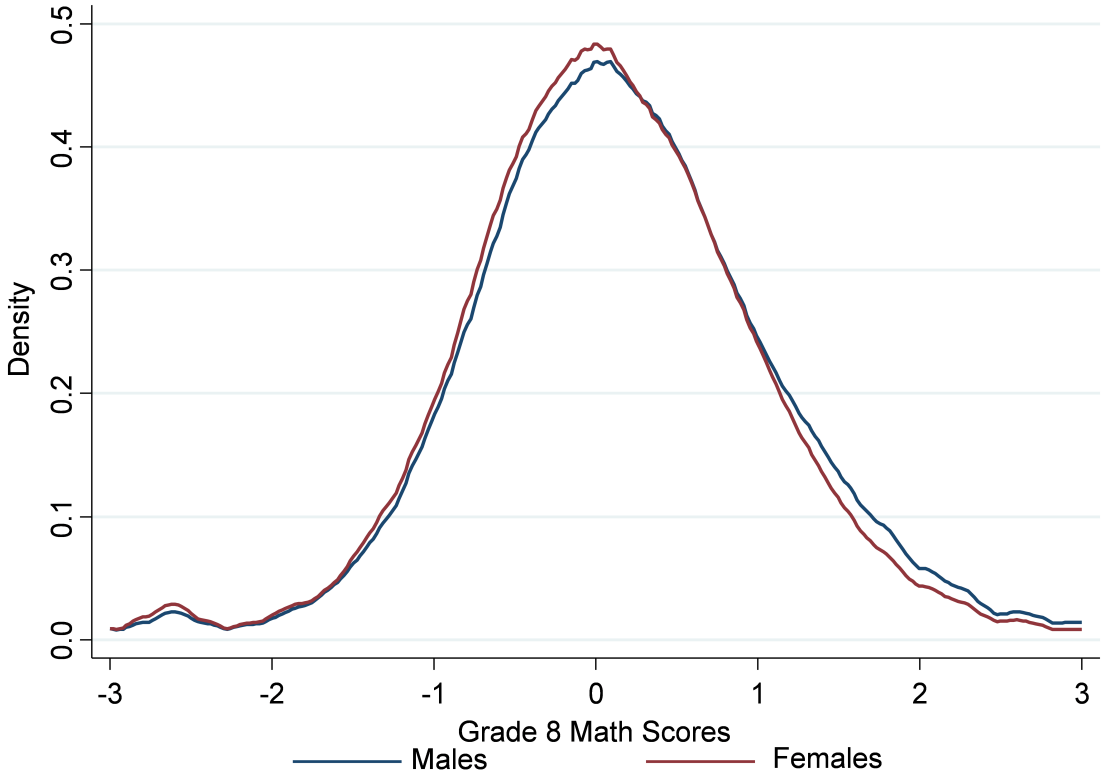


FIGURE 12: Percent of Gender Gap in Patent Rates Explained by Test Scores in Grades 3-8

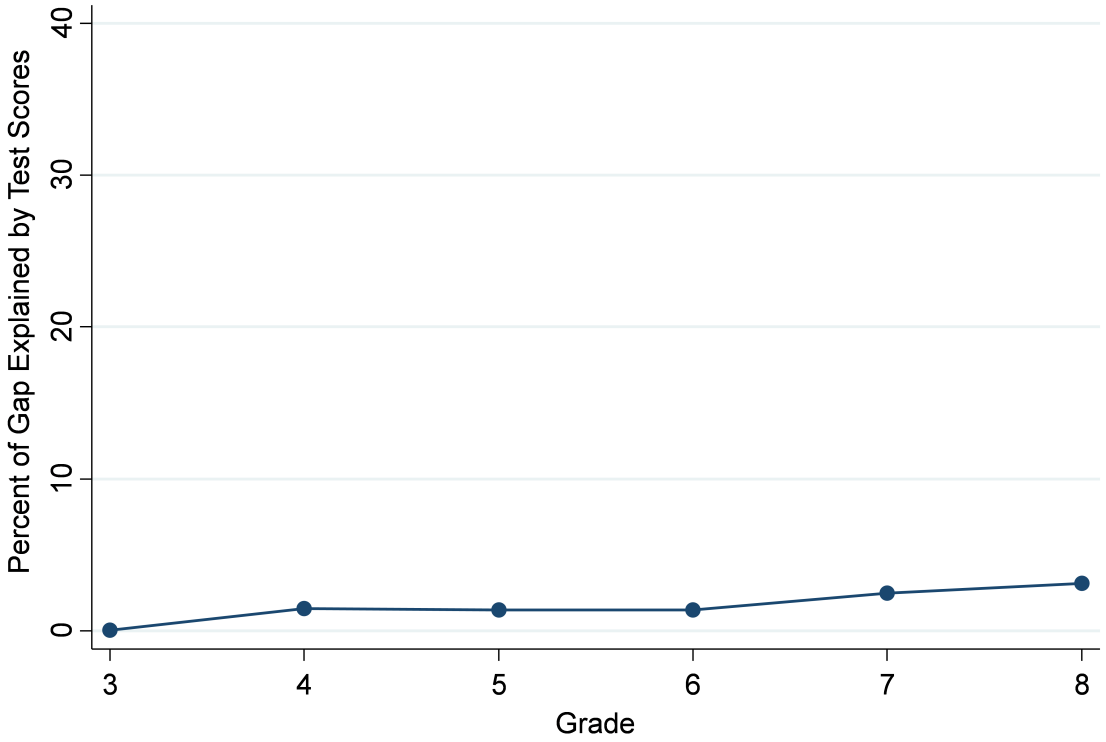
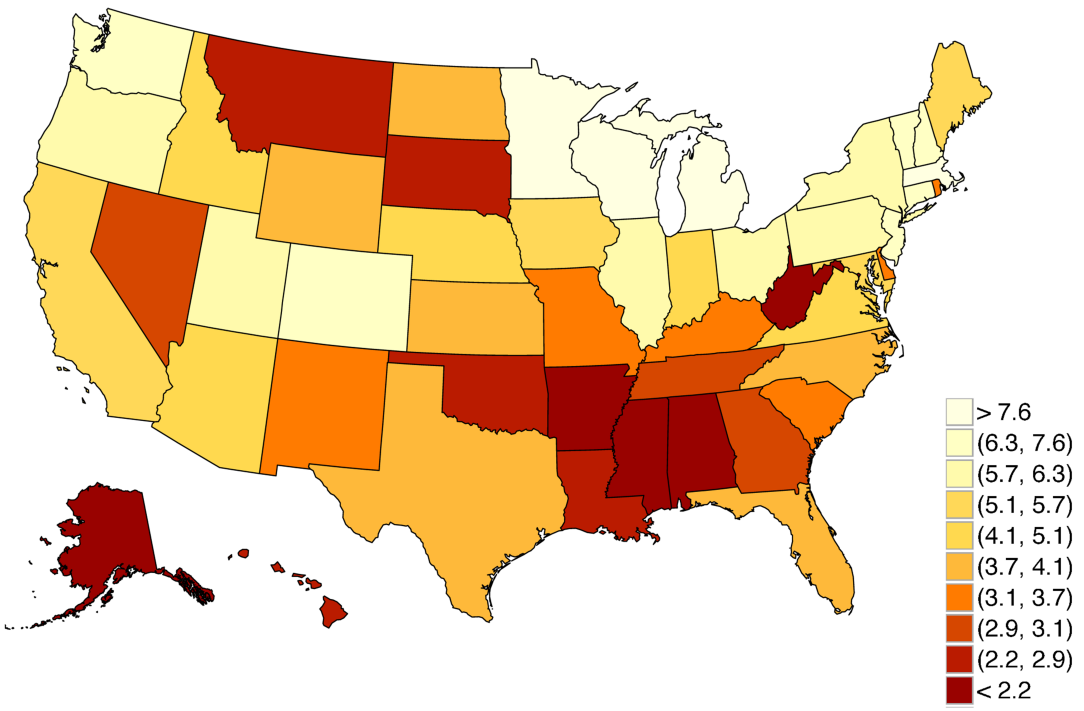


FIGURE 13: The Origins of Inventors: Patent Rates (per 10,000 Children) by State where Child Grew Up



Note: Ligher colors represent areas with more inventors

FIGURE 14: Children Who Grow Up Near Inventors are More Likely to be Inventors

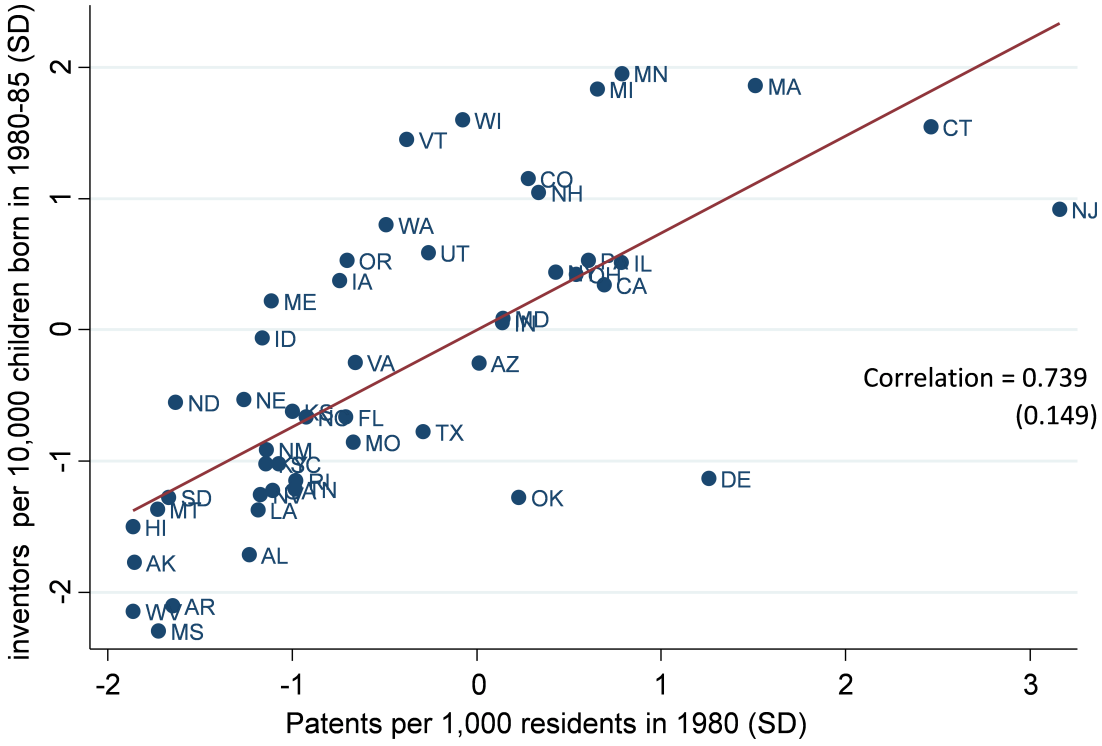


FIGURE 15: Correlation between Patent Rates of Children and Prior Residents - Residuals Controlling for State and Tech Category Fixed Effects

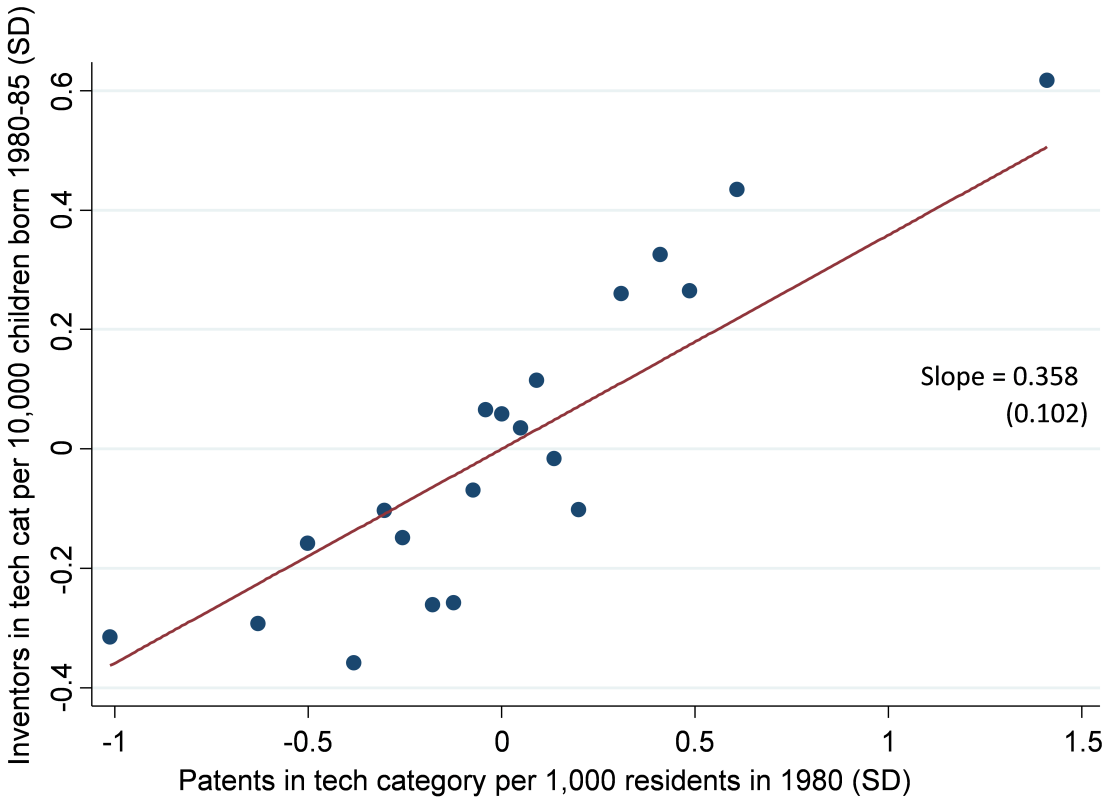
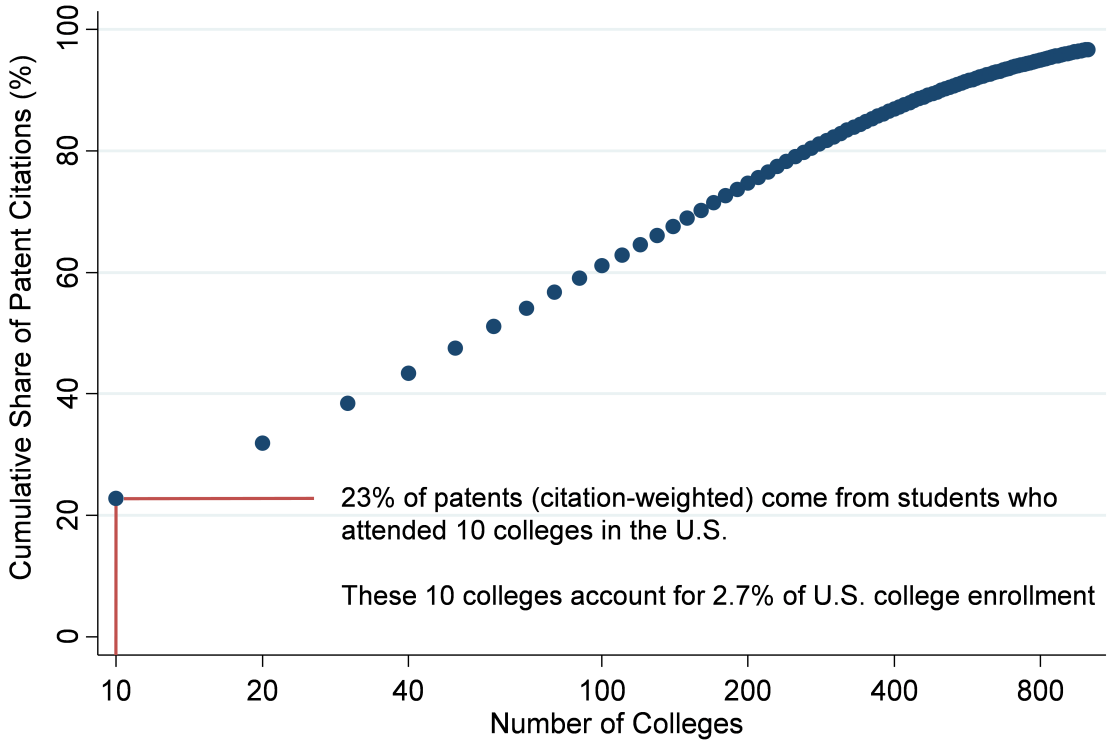


FIGURE 16: Concentration of Innovation Among Graduates of Selected Colleges



Sample: 200,000 inventors enrolled in college between 1999-2012

FIGURE 17: Returns to Innovation - Event Studies of Earnings Around Patent Application

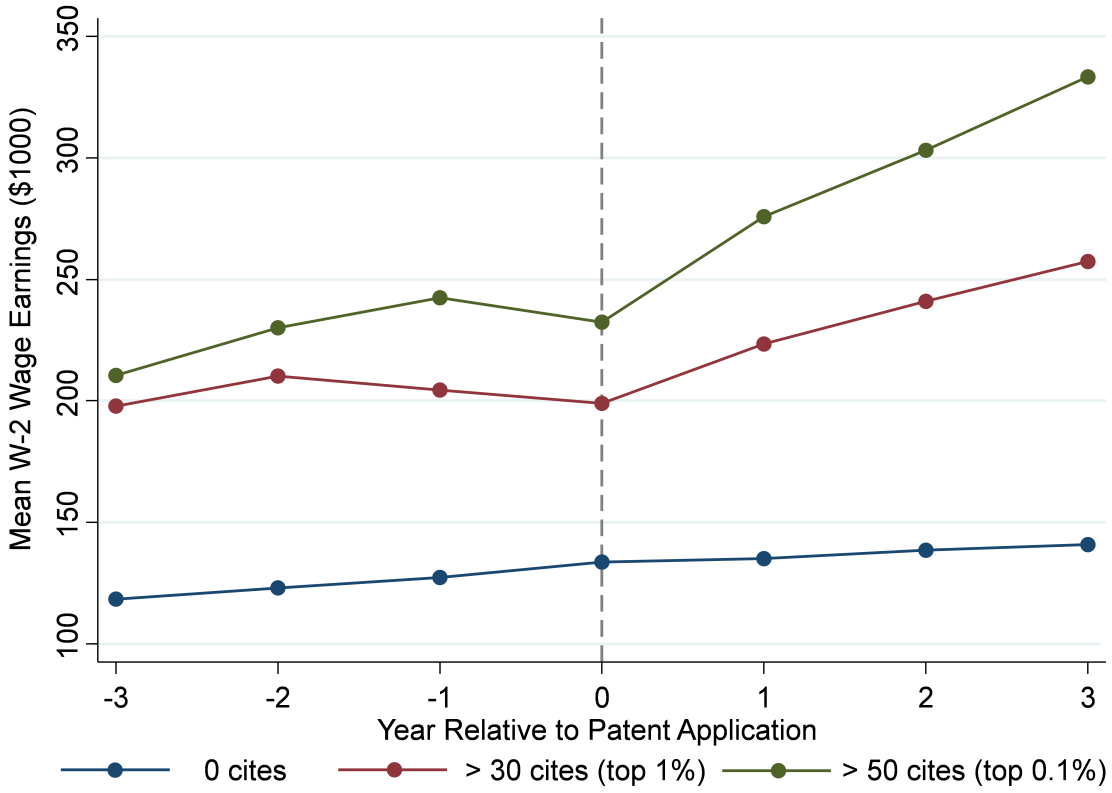
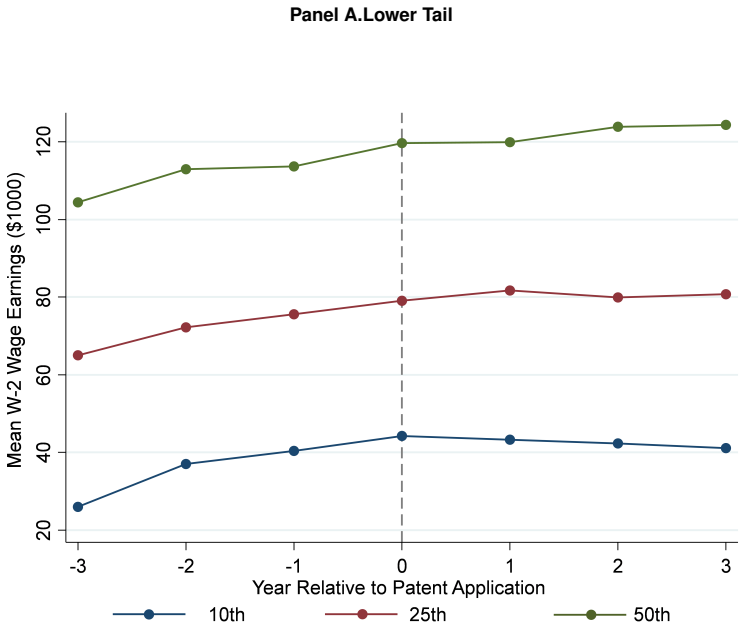
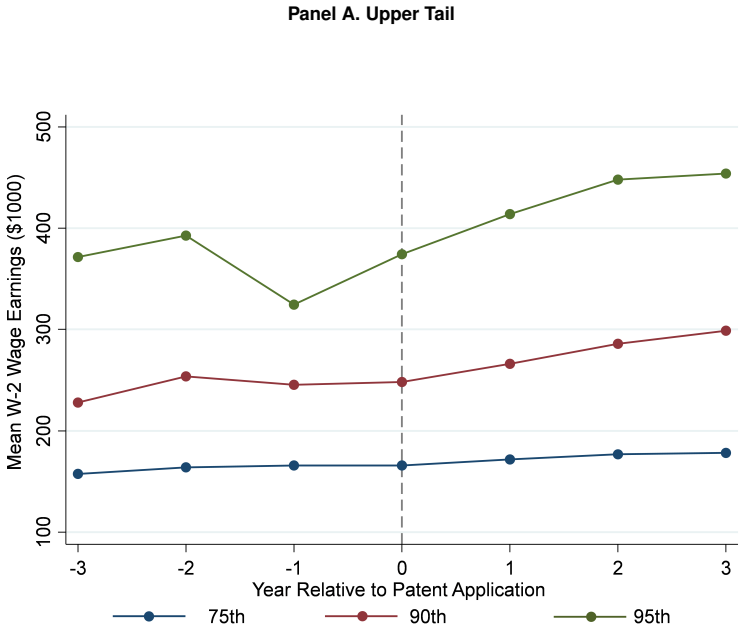


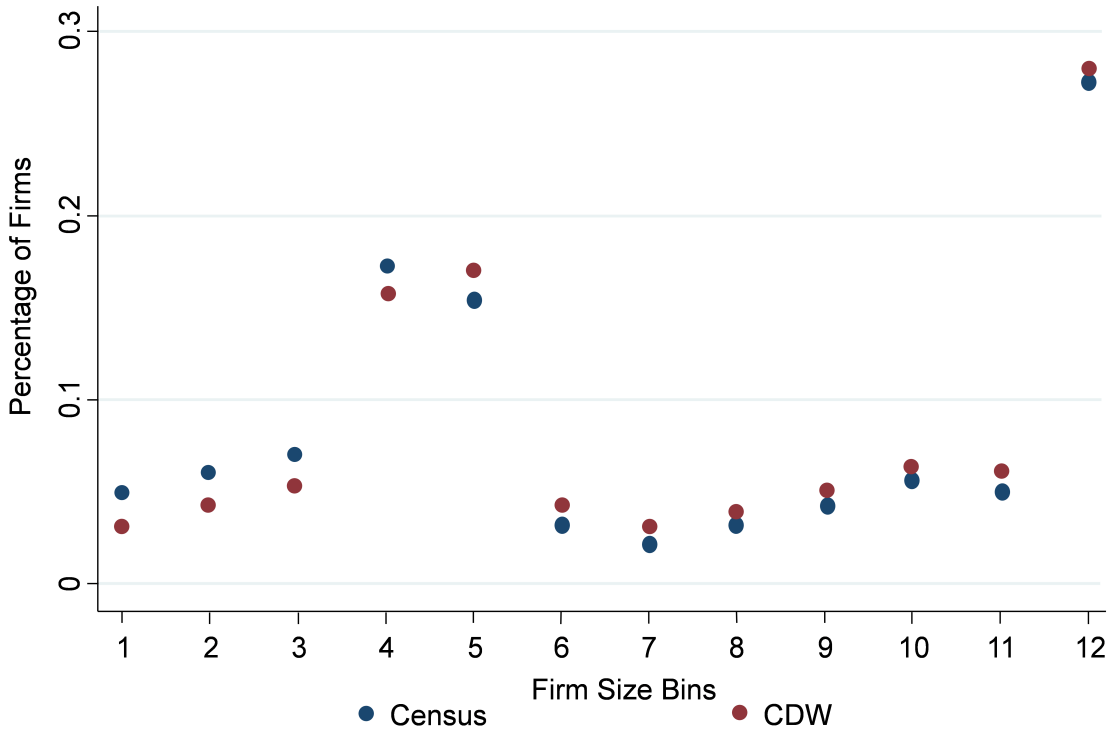
FIGURE 18: Returns to Innovation - Event Studies of Patents of Different Citations and Wages at the 95th Percentile of Citation Distribution





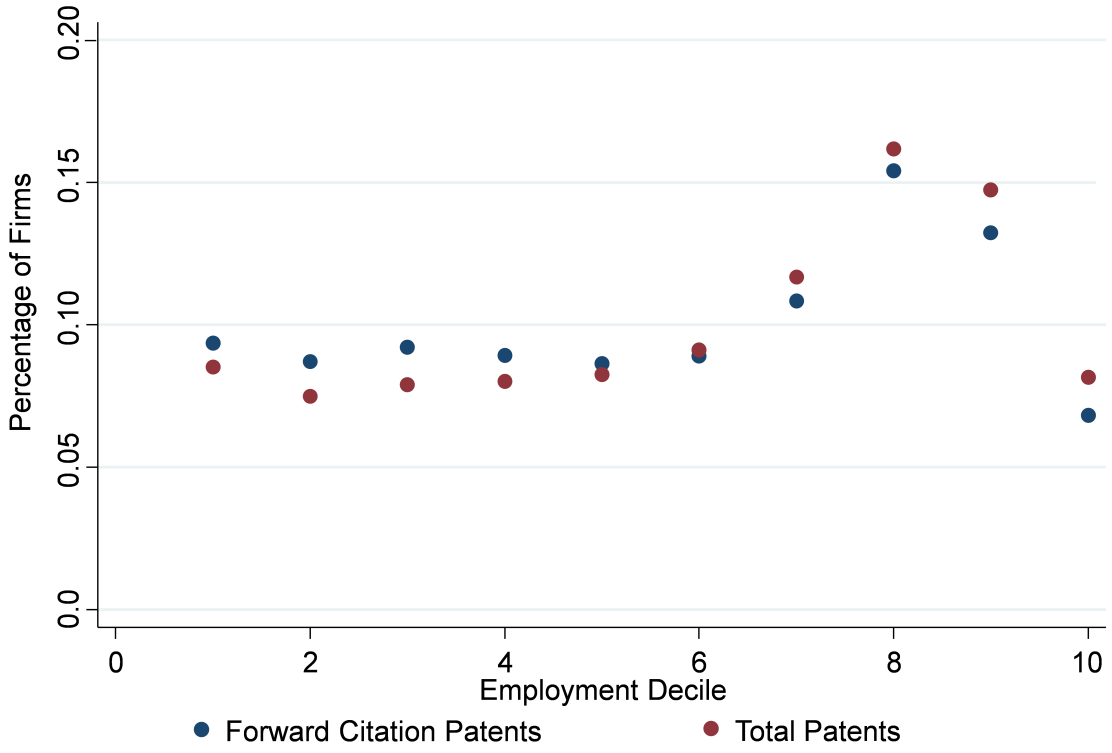
Notes: The y-axis is the wage at differential quantiles. Event window is three years before and three years after the patent filing (year of filing is the vertical red line). The biggest effects are for wages at the top of the distribution.

FIGURE 19: Comparing Firm Size Distribution in Census and CDW



Notes: Blue dots are the fraction of employees in different firm sizes according to Census (in 2003) and red dots are the same in CDW. Size bins are used by the Census. They match up very well despite the fact that EIN may not correspond to Census "enterprises" and includes some industries that Census does not. Firm size bins in number of employees from 1 to 10: 1-4; 4-9; 9-19; 19-99; 99-499; 499-749; 749-999; 999-1,499; 1,499-2,499; 2,499-4,999; 4,999-9,999; 9,999+.

FIGURE 20: Innovation Activity by Firm (EIN) Size - Firms in Deciles 7-9 are More Innovation Intensive than Smaller Firms



Notes: Employment deciles are based on the distribution of employment across all EINs in the population (with and without inventors).

FIGURE 21: Number of Firms with Patents vs. R&E Credits

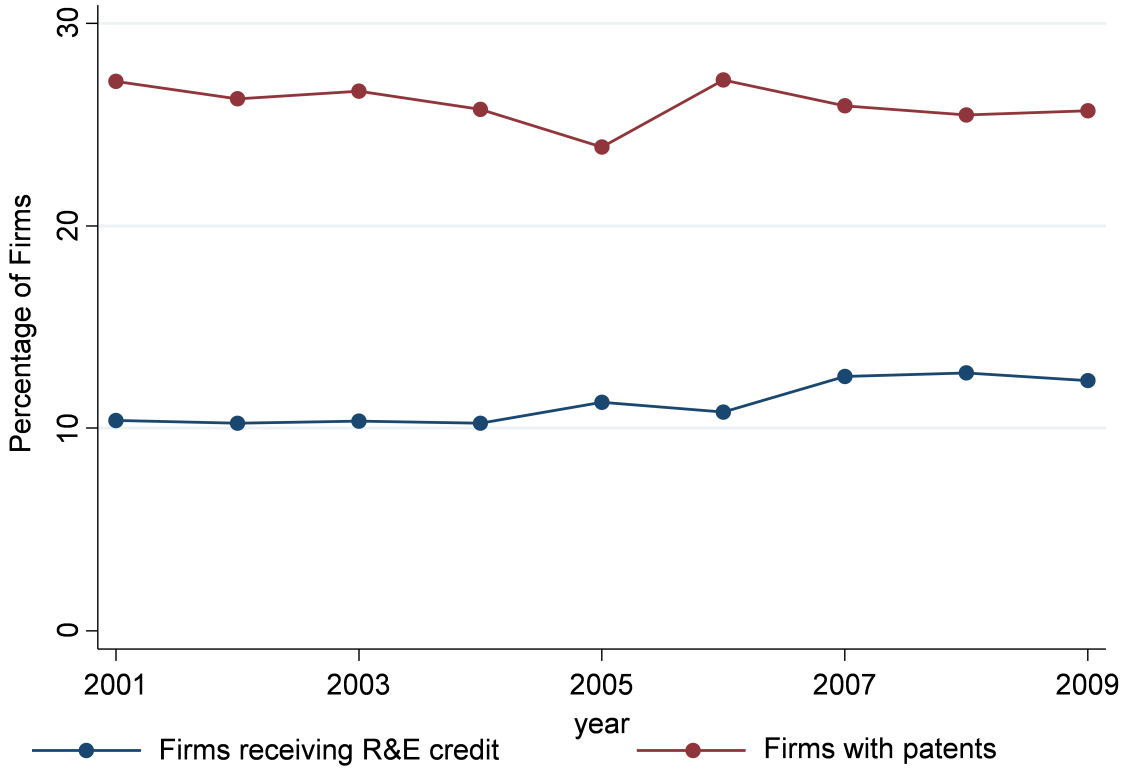


FIGURE 22: Percentage of Firms with Patents receiving R&E Credits

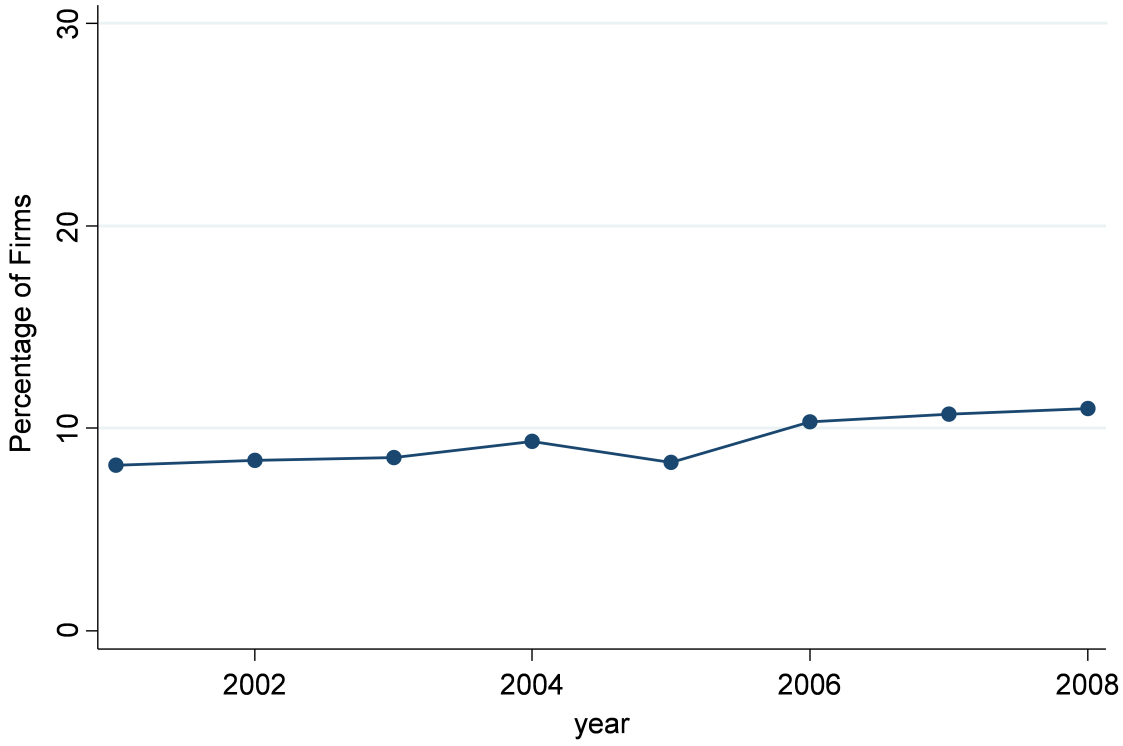


TABLE 1: Summary Statistics for Inventors

	Mean	Median	Standard Deviation
	(1)	(2)	(3)
Patents	3.93	2	9.21
Citations	36.19	6	168.96
Age at Patent	46.55	45	19.95
Wage Income ($)	139,427	100,082	558,855
Total Income ($)	214,424	122,346	1,564,315
Female (%)	11.60		

Number of Inventors: 786,656

TABLE 2: What Fraction of the Gap in Patenting by Parent Income is Explained by Differences in Test Scores ?

	Patent Rate (per 10,000 Individuals) (1)	Gap Relative to Above-Median Income (2)
Above Median Income	5.82	
Below Median Income	2.15	3.67
Below Median Income (Reweighting Scores)	2.76	3.06
Fraction of gap explained by 3rd grade scores (%)		16.70

TABLE 3: What Fraction of the Gender Gap in Patenting is Explained by Differences in Test Scores ?

	Patent Rate per 10,000 Individuals (1)	Gap Relative to Males (2)
Above Median Income	8.26	
Below Median Income	1.62	6.63
Below Median Income (Reweighting Scores)	1.83	6.42
Fraction of gap explained by 3rd grade scores (%)		3.10

FIGURE A1: Conditional on Becoming an Inventor, Probability of Being a Star Inventor if Born Into Low Income Family

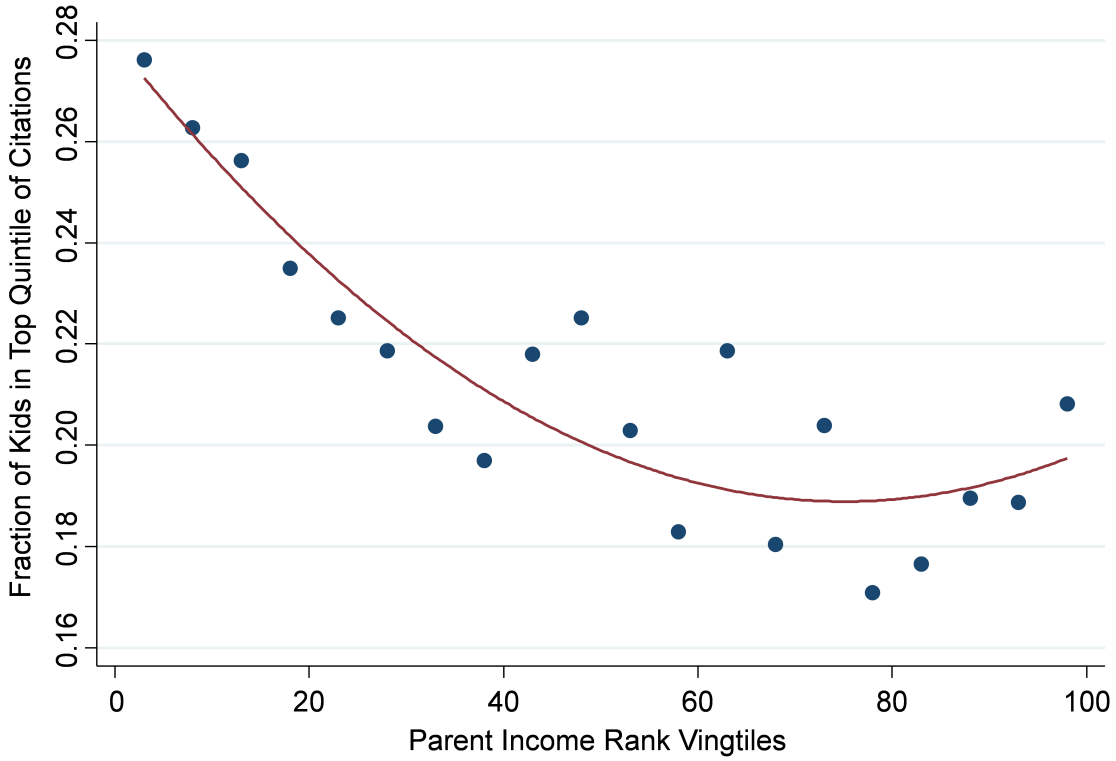
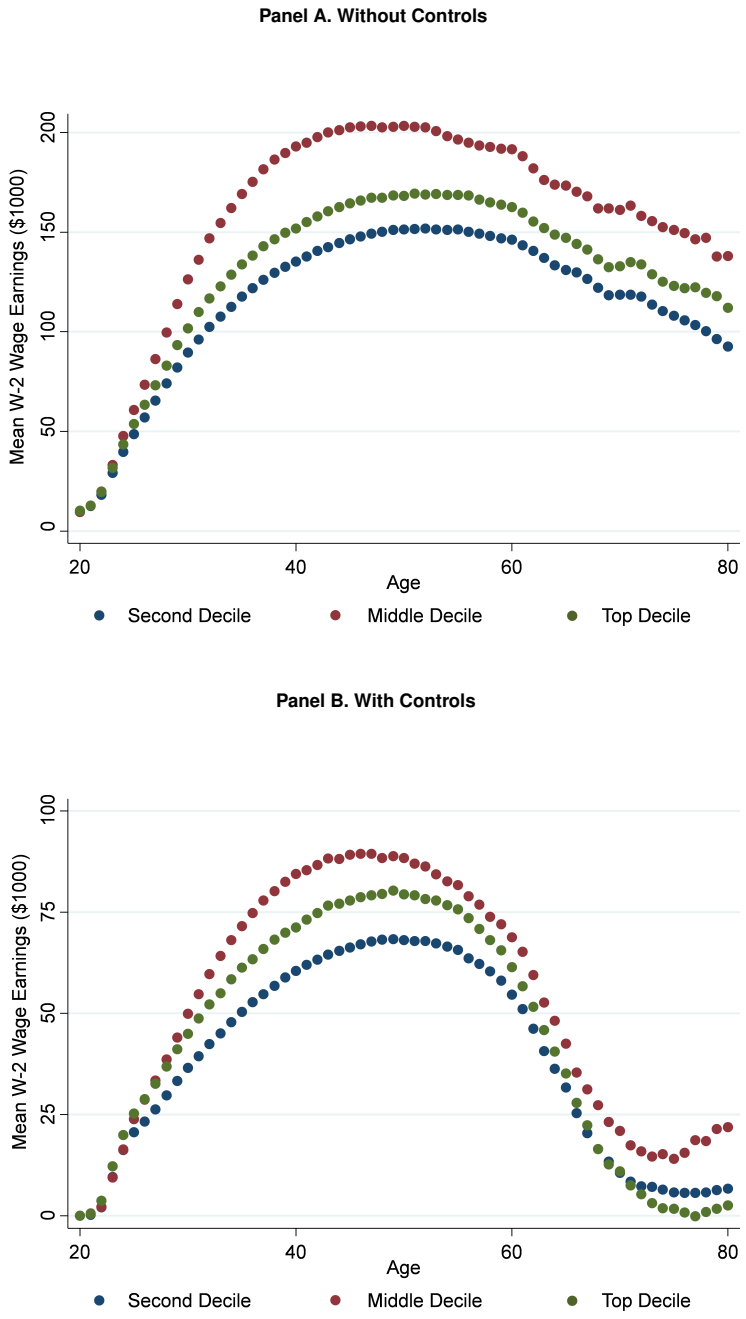


FIGURE A2: Income-Age Profiles for Inventors (Depending on Citations)



Notes: These are the average wages of inventors normalized to 0 at age 20 but cut by citation bin. Sample is all individuals who had a patent granted 1996-2012. Lines at top decile (10); middle decile and second decile. Panel A has no controls, panel B has controls for year, state, foreigner, tech category, self-assigned indicator, married and gender dummies.

FIGURE A3: Event Study of Non-Wage Income Around Time of First Application

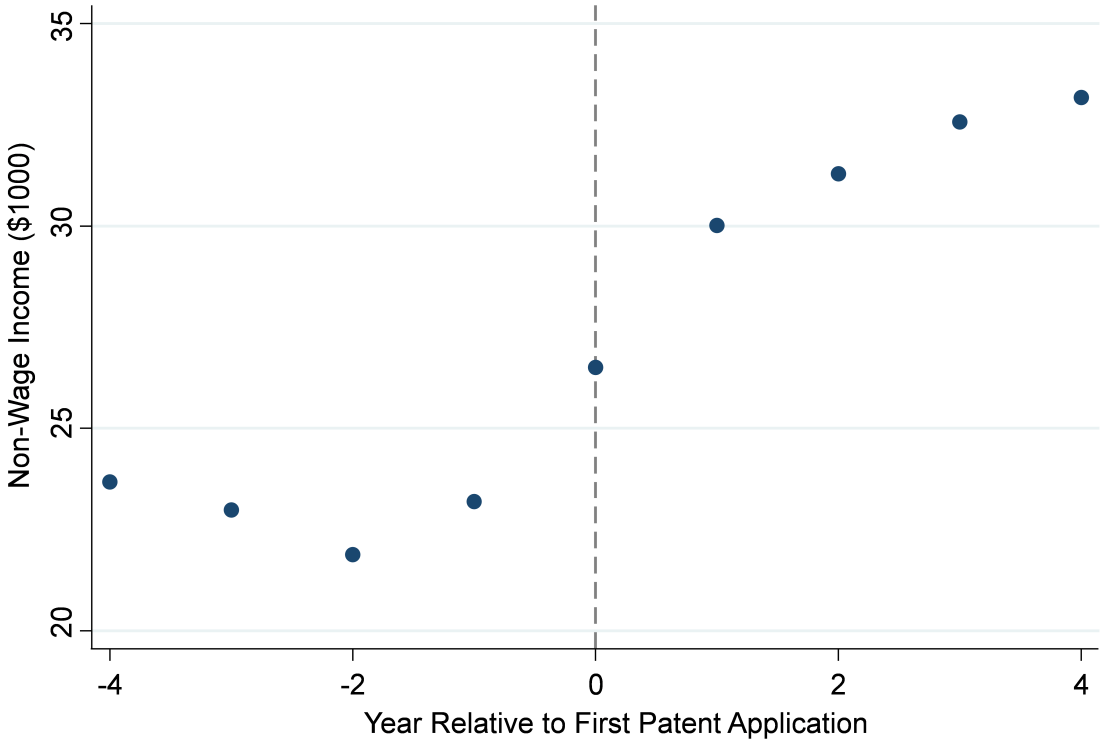


FIGURE A4: Event Study of Wage Earnings Around Time of First Application

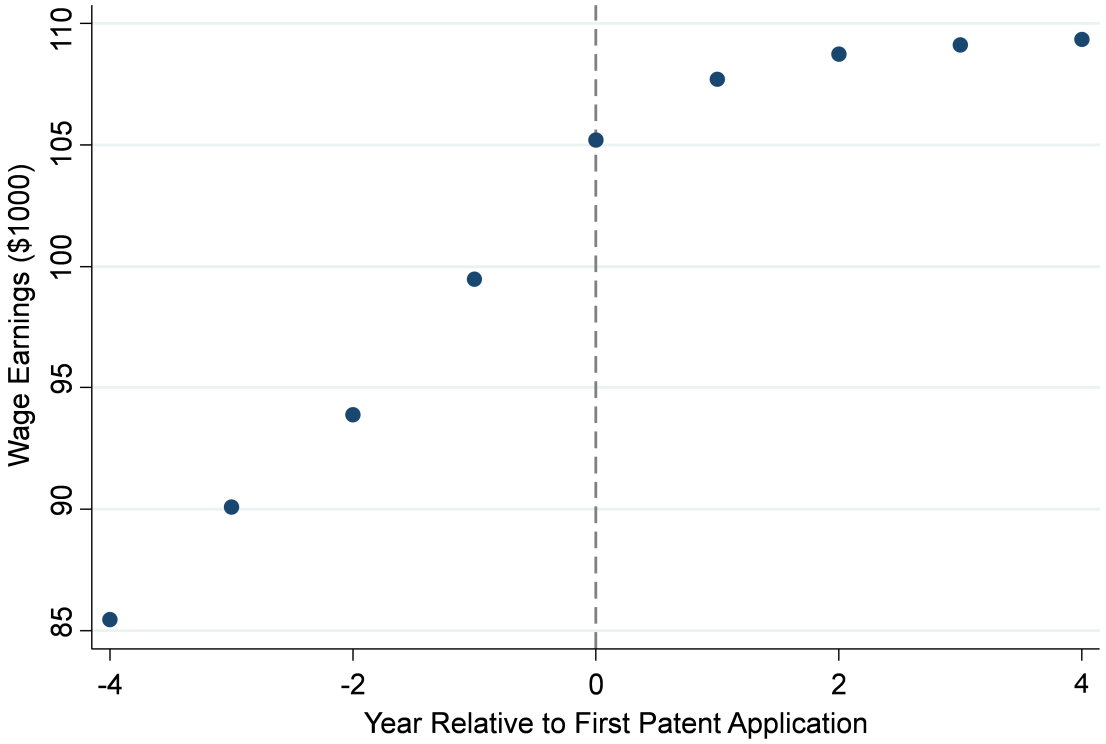
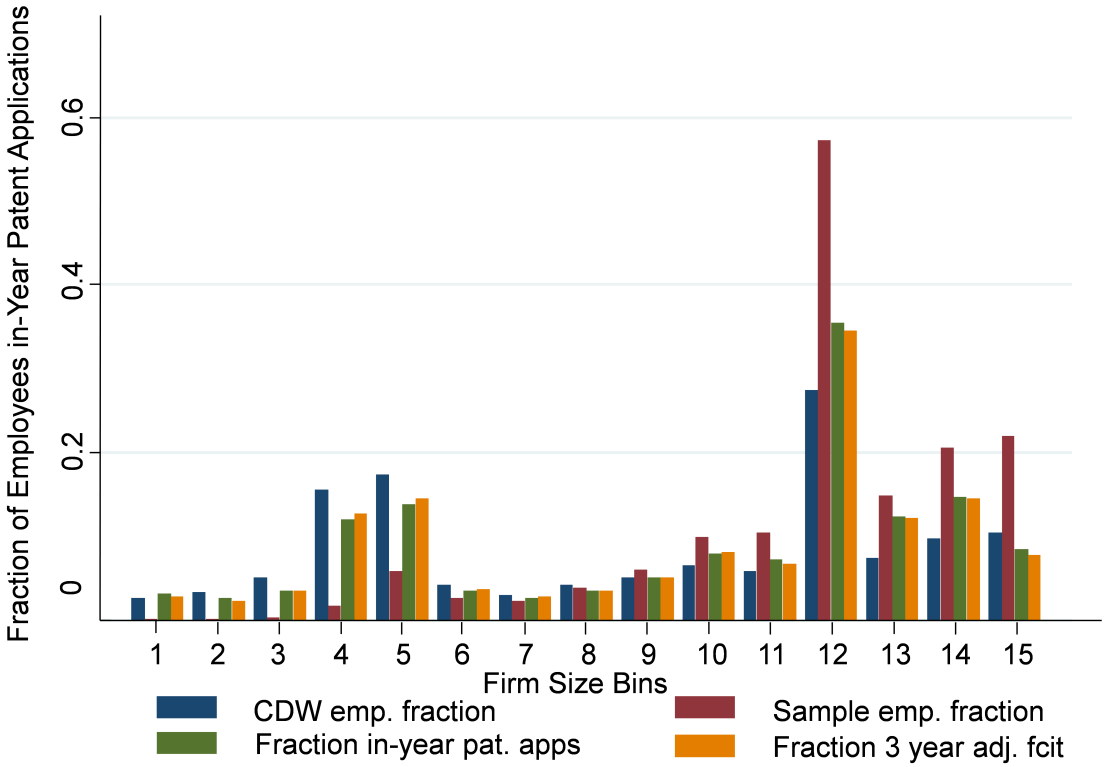


FIGURE A5: What Firms Employ Inventors ?



Notes: Firm size bins in number of employees from 1 to 15: 1-4; 4-9; 9-19; 19-99; 99-499; 499-749; 749-999; 999-1,499; 1,499-2,499; 2,499-4,999; 4,999-9,999; 9,999+; 9,999-25,000; 25,000-100,000; 100,000+

FIGURE A6: Small Firms Have Greater Fraction of Self-Assigned Patents

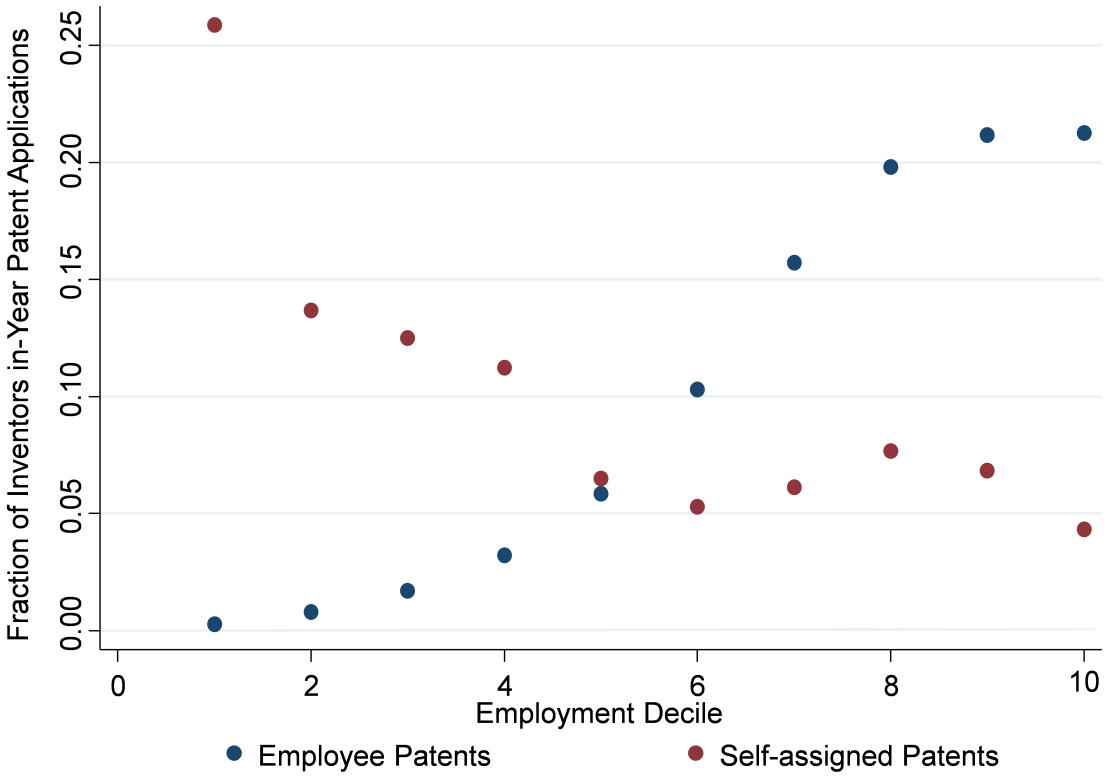


FIGURE A7: Single-Inventors vs Multiple-Inventor Patent Applications

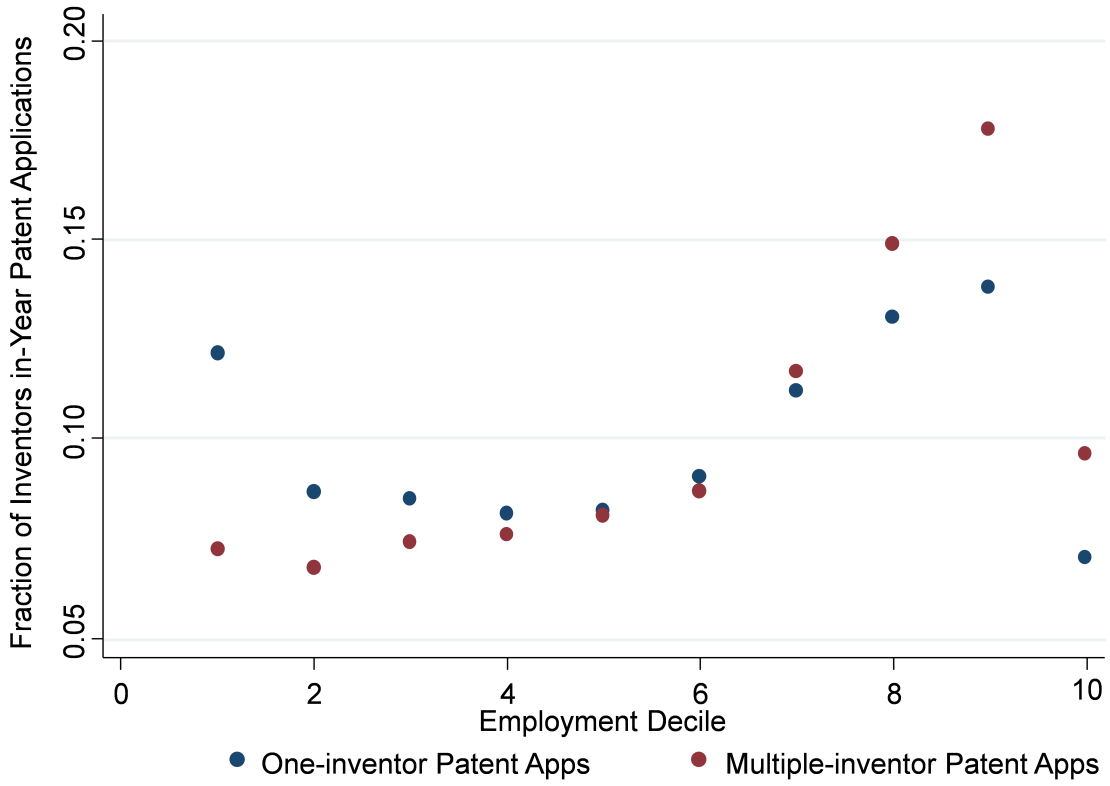


TABLE A1: Employment Deciles

Decile	Min. Employment (1)	Max. Employment (2)
1	1	16
2	17	52
3	53	135
4	136	339
5	340	899
6	900	2,499
7	2,500	7,299
8	7,300	24,999
9	25,000	99,999
10	100,000	

TABLE A2: Most Frequent Age Distribution, Teams of Two Inventors

1	2	3	4	5	6	7	8	9	10	Number	Frequency
0	0	0	0	0	0	0	0	1	1	10,364	11.55%
1	0	1	0	0	0	0	0	0	0	9,284	10.35%
0	1	1	0	0	0	0	0	0	0	9,012	10.05%
0	0	1	0	1	0	0	0	0	0	8,960	9.99%
0	0	0	0	0	0	0	1	1	0	8,788	9.80%
0	0	0	0	0	0	0	1	0	1	8,754	9.76%
0	0	0	0	0	0	0	0	0	2	8,728	9.73%
1	1	0	0	0	0	0	0	0	0	8,706	9.71%
0	0	0	0	0	0	1	1	0	0	8,604	9.59%
0	0	1	1	0	0	0	0	0	0	8,498	9.47%

TABLE A3: Age Spread of Two Person Teams

Age Spread is at least X deciles	Share
4	46.50%
5	33.10%
6	22.50%
7	14.50%
8	8.50%
9	4.00%
10	1.30%

TABLE A4: Most Frequent Wage Distribution, Teams of Two Inventors

1	2	3	4	5	6	7	8	9	10	Number	Frequency
0	0	0	0	0	0	0	0	1	1	8,129	11.83%
0	1	1	0	0	0	0	0	0	0	7,385	10.75%
0	0	0	0	0	0	0	0	0	2	7,365	10.72%
0	0	0	0	0	0	0	1	1	0	7,272	10.58%
0	0	0	0	0	0	1	1	0	0	6,739	9.81%
1	1	0	0	0	0	0	0	0	0	6,697	9.75%
0	0	1	1	0	0	0	0	0	2	6,614	9.63%
0	0	0	0	1	1	0	0	0	0	6,246	9.09%
0	0	0	0	0	0	1	0	1	0	6,163	8.97%
0	0	1	1	0	0	0	1	0	1	6,096	8.87%

TABLE A5: Wage Spread of Two Person Teams

Age Spread is at least X deciles	Share
2	83.00%
3	59.00%
4	42.00%
5	29.00%
6	19.00%
7	12.00%
8	7.00%
9	3.50%
10	1.20%