Living the American Dream in Finland: The Social Mobility of Inventors*

Philippe Aghion Ufuk Akcigit Ari Hyytinen Otto Toivanen

July 18, 2016

STILL PRELIMINARY. COMMENTS WELCOME..

Abstract

In this paper we merge individual census data, individual patenting data, and individual IQ data from Finnish Defence Force to look at both, the selection into becoming an innovator and the returns to invention. On the former, we find that: (i) the probability of becoming an inventor is strongly correlated with parental income; (ii) this correlation is mostly driven by the fact that rich parents have more educated children: children's education explains 81.7% of the explained variation in the probability of becoming an inventor, followed by children's IQ (16.8%). In the second part of the paper we look at the returns to invention. Here, we find that: (i) inventing increases the annual wage rate of the inventor by a significant amounts over a prolonged period after the invention; (ii) returns are tied to the quality of innovation; (iii) coworkers in the same firm also benefit from an innovation, the highest returns being earned by entrepreneurs in the firm, especially in the long term. Finally, we find that becoming an inventor enhances both, intragenerational and intergenerational income mobility.

Keywords: Inventors, innovation, social mobility, IQ, education, parental background.

*Addresses- Aghion: College de France and London School of Economics. Akcigit: University of Chicago. Hyytinen: Jyvaskyla University. Toivanen: KU Leuven. This project owes a lot to early discussions with Raj Chetty, Xavier Jaravel and John Van Reenen when we were embarking on two parallel projets, them on US inventors and us on Finnish inventors. We are very grateful to our discussants Pierre Azoulay and Chad Jones for very helpful comments and suggestions. Finally, we also thank Harun Alp, Torsten Persson, Stefanie Stantcheva, and the seminar participants at the Kaufman Foundation, the London School of Economics, the IOG group at the Canadian Institute for Advanced Research, and College de France for very helpful discussions.

1 Introduction

New growth theories (e.g. see Romer (1990), Aghion and Howitt (1992), and Aghion, Akcigit and Howitt (2014)) typically assume an economy with ex ante identical individuals who freely decide whether or not to become innovators, and are indifferent in equilibrium between innovating or working in manufacturing. In practice, however, not everybody can become an innovator: whether one becomes an innovator or not, is likely to depend upon the social environment (parental resources and education, the individual's own education,..) and upon innate ability, both of which are unevenly distributed across individuals.

In this paper we look at what determines an individual's probability to become an inventor, and how inventing in turn affects the income of the inventor and the income of other employees in the same firm.

The following striking fact motivated our analysis. Figure 1 depicts the relationship between an individual's probability of becoming an inventor and his father's income: we see that the individual's probability of becoming an inventor increases with father's income, and that the effect is highly non-linear, being particularly steep at the highest levels of father's income. We also see that the probability of innovating for an individual whose father is at the very top of the income distribution is about ten times larger than the corresponding probability for an individual with a father at the bottom end of the income distribution. In fact this curve is remarkably similar to the findings in Bell et al (2015) and Akcigit et al (2016). And this is all the more remarkable that, unlike the US, Finland offers free education up to and including tertiary education. Moreover, Finland has among the lowest income inequality and highest social mobility among OECD countries (e.g. see Figure 2), whereas the opposite is true for the US. What lies behind this relationship in Figure 1 between father income and the probability of becoming an inventor?

In this paper, we merge individual census data, individual patenting data, and individual IQ data to look at both, the selection into becoming an innovator and the returns to invention in Finland.[1] More specifically, we merge three Finnish data sets: (i) individual data on income, education and other characteristics from Statistics Finland (SF) over the period between 1988 and 2012; (ii) individual patenting data from the European Patent Office (EPO); (iii) IQ data from the Finnish Defence Force. Our base data (i) consists of the whole Finnish work force. Given that conscription only affects males in Finland, we concentrate on the male work force in this paper.

In the first part of the paper we look at the selection into becoming an inventor. Here, we find that: (i) the probability of becoming an inventor is strongly correlated with parental income; (ii) this correlation is mostly driven by the fact that rich parents have more educated children:

[1] A parallel attempt at looking at the selection of inventors and the returns to invention, has been made by Bell et al (2015) using US data, see our discussion below.

children's education accounts for 81.7% of the explained variation in the probability of becoming an inventor, followed by children's IQ (16.8%); (iii) if we try to decompose the explained variation in children's education, we get that IQ accounts for 52.2% of that variation followed by parental education (24.6%) followed by parental wealth (15.7%) and parental income (7.4%).

In the second part of the paper we look at the returns to invention. Here, we find that: (i) making an innovation increases the annual wage rate of inventors by a significant amounts over a prolonged period; (ii) returns are tied the quality of innovation; (iii) coworkers in the same firm also benefit from an innovation, the highest returns being earned by entrepreneurs in the firm, especially in the longer term. Finally, we find that becoming an inventor enhances both, intragenerational and intergenerational income mobility, and that being an inventor drastically reduces the father-son income relation.

The paper relates to several strands of literature. There is first a theoretical literature on innovation incentives.[2] Then there is a recent literature on growth and reallocation (see Hsieh and Klenow, 2009; Acemoglu et al , 2013; Hsieh et al, 2013). We contribute to this literature by focusing on the selection of inventors and its relationship to parental wealth, education and IQ.

Aghion et al (2015) look at the relationship between innovation, inequality and social mobility using aggregate cross-state and cross-commuting-zone data. They show that innovation measured by the flow or quality of patents is positively correlated with the top 1% income share of income, is uncorrelated with broader measures of income inequality, and is positively correlated with social mobility (measured as in Chetty et al, 2014). In this paper we look at the relationship between innovation, income, and social mobility using individual data on income, patenting, education and IQ.

Closer to our analysis in this paper is a recent literature merging individual income data with individual patenting data. First, Toivanen and Vaananen (2012) use Finnish patent and income data to study the return to inventors of US patents. They find strong and long-lasting impacts, especially for the inventors of highly cited patents. Toivanen and Vaananen (2015) look at the effect of education on the probability of becoming an inventor and they find a positive and significant treatment effect, suggesting the one may increase innovation through education policy. Second, Celik (2015) matches inventors' surnames with socioeconomic background information inferred from those surnames by looking at the US census data back in 1930. His main finding is that individuals from richer backgrounds are far more likely to become inventors. Akcigit et al (2016) merge historical patent and individual census records and show that probability of becoming an inventor around 1940s was very highly correlated with father's income but this strong relationship

[2] In particular, see Holmstrom (1989), Lerner (2006), Manso (2011), and Aghion and Tirole (1994). However none of these papers looks at the effects of social background on the probability of inventing, nor do they analyze the social mobility of inventors and co-workers.

disappears once child's education is controlled for. Finally, Jaravel et al (2015) merge US individual tax data and individual patenting data to quantify the impact of coauthors in the career of inventors, finding evidence of large spillover effects.[3]

Most closely related to the present paper, is Bell et al (2015) who merge US individual fiscal data, test score information, and US individual patenting data over the recent period to look at the lifecycle of inventors and the returns to invention. These authors find that parental income, race and gender are important determinants of the probability of becoming an inventor. In particular, as we already stressed above, they obtain pretty much the same pattern as ours when depicting the probability of becoming an inventor on parental income. And they also find that when controlling for school performance at a later age, parental income has a more limited impact on the probability of becoming an inventor.

We contribute to this literature: (i) by adding family background variables, in particular parental education; (ii) by adding IQ variables; (iii) by adding information on invention returns for coworkers and entrepreneurs in the same firm as the inventor. This in turn allows us to: (a) better understand what lies behind the observed correlation between parental income and becoming an inventor; (b) analyze the returns to invention through the various layers of the firm.

The remaining part of the paper is organized as follows. Section 2 presents the data and shows some descriptive statistics. Section 3 analyzes the determinants of becoming an inventor. Section 4 analyzes the returns to invention and the effects of invention on income and social mobility. And Section 5 concludes.

2 Data and descriptive statistics

2.1 The data

Our data come from the following sources. First, Statistics Finland (SF). This dataset comprises: (i) the Finnish longitudinal employer-employee data (FLEED) which we exploit for the period 1988-2012; this annual panel is constructed from administrative registers of individuals, firms and establishments, maintained by SF. It includes information on individuals' labor market status, salaries and other sources of income extracted from tax and other administrative registers, it also includes information on other individual characteristics, and employer and plant characteristics. The FLEED contains the entire Finnish working age population; (ii) the population census 1975 and 1985. This informs us about parental education, and the location and income of social and biological parents. Only biological parents are considered in the present draft.

Second, the European Patent Office data which provides information on characteristics such as

[3]Jaravel's work builds on prior seminal work by Azoulay (2010) which examines the effect of the premature death of 112 eminent scientists on their co-authors.

the inventor names and applicant names.[4] We have collected patent information on all patents with at least one inventor who registers Finland as his or her place or residence. Data on all patents with a Finnish inventor up to and including 2012.

Third, the Finnish Defence Force, which provided us with information on IQ test results for conscripts that did their military service in 1982 or later; all conscripts take the IQ test in the early stages of the service. These data contains the raw test scores of spatial, verbal and quantitative IQ tests. The IQ test are a 2-hour multiple choice test containing sections for verbal, arithmetic and visiospatial reasoning. The latter is similar to the widely used Raven's Progressive Matrices – test. Overall, the Finnish Defense Force IQ test is similar to the commonly used IQ tests; moreover, a large majority of each male cohort performs the military service and therefore takes the test: most conscripts take their military service around the age of 20. We consider two IQ measures. The first measure uses the deciles in visiospatial IQ scores. The second measure uses the deciles of the sum of the three IQ test scores.

The linking of all other data but the patent data was done using individual and firm identifiers. The linking of patent data to individuals was done using the information on individual name (first and surname), employer name, individual address and/or employer's address (postcode, street name street number) and year of application. These were used in different combinations, also varying the year of the match to be before or after the year of application (e.g., matching a patent applied for in 1999 with the street address of the firm from the registry taken in 1998 or 2000).

For the "who becomes an innovator" regressions in Section 3, we used data on individuals for whom we have IQ data. This excludes women (since women do not go through military service) and it also excludes men born before 1961. The remaining sample comprises around 700,000 individuals. For the "return to invention" regressions in Section 4, we used data on individuals employed in the private sector (due to missing information on the number of employees in public sector organizations). This excludes roughly half of the working age male population, thus leading to a sample of about 900,000 individuals, and to more than 7 million observations. For the "intergenerational mobility" (or "social mobility") regressions, we used data on men for whom we observe IQ, father's wage and own wage at age 35. The corresponding sample comprises around 360,000 individuals. Finally for the "intragenerational mobility" (or "income mobility") regressions, we used data on men for whom we observe IQ, and the own wage at ages 35 and 45, which corresponds to about 120,000 individuals.

[4]Here we want to thank the research project "Radical and Incremental Innovation in Industrial Renewal" by the VTT Research Centre (Hannes Toivanen, Olof Ejermo and Olavi Lehtoranta) for granting us access to the patent-inventor data they compiled.

2.2 Some basic numbers and patterns

Our initial sample of consists of 12,575 inventors (6,799 in the IQ sample). 11% of them are females. The distribution of the number of patents per inventor is illustrated in Figure 3. Half of the inventors have one patent; another 19% two and 9% three patents. A total of 23 inventors have more than 50 patents. Inventors in our sample have more education compared to the whole population. Figure 4 plots the distribution of inventors versus the whole Finnish population with five education categories: base, secondary, college, master and PhD degrees. 56% of inventors have at least a master's degree, compared to 10% of non-inventors. Over 90% of inventors with at least a college degrees have a science education. Finnish inventors have also higher IQ than the population at large: as can be seen from Figure 5, more than 20% of inventors are in the 9th IQ decile and more than 35% in the highest IQ decile.

Finnish inventors have highly educated parents (Figures 6 and 7): about 5% of non-inventors, but 16% of inventors have a father with at least a master's degree. The comparable figures for mother's education are less than 3 and more than 7%. Finnish inventors have also richer parents on average. Figure 8 plots the income distribution of fathers of inventors and fathers of non-inventors. Around 40% of non-inventors have a father who is in the top two income quintiles. The same figure for inventors is 58%. Figure 9 does the same for mothers - 41% of non-inventors mothers and 48% of inventors' mothers are in the top two income quintiles. 70% of non-inventors and 85% of inventors come from urban areas. 94.4% of non-inventors and 94.5% of inventors speak Finnish, 5.5% and 5.3% speak Swedish and the rest some other language as their mother tongue.

2.3 Some illustrative graphs

Figure 10 shows the wage distribution conditional upon the inventor status: we see that the distribution moves to the right for inventors compared to non-inventors.

Figure 11 shows the correlation between an individual's probability of becoming an inventor and his own education at different levels, respectively for non-science and science education. We see that a high degree in science education greatly increases the probability of inventing.

Figure 12 shows the correlation between an individual's probability of becoming an inventor and his IQ: we see an increasing and convex effect of IQ on the probability of inventing.

Figure 13 shows the correlation between an individual's probability of becoming an inventor and both his IQ and educational level. We see that IQ matters at all levels of education, but proportionally more at lower levels of education.

Figure 14 shows the correlation between an individual's probability of becoming an inventor and his father's education, at different levels of the individual's own education. We see that father's education matters more, the more educated the individual, and the more so if he has pursued science studies.

Figure 15 shows similar findings when looking at the correlation between an individual's probability of becoming an inventor and his mother's education.

3 Becoming an inventor

In this section we estimate a linear probability model where we regress the probability of becoming an inventor on parental income, parental education, IQ, and own education. We first show some motivating evidence and then turn to the regressions.

3.1 Motivating evidence

Figure 16 reproduces Figure 1, but separately for individuals having a father with an Master in Science (MSc) and for individuals whose fathers do not have an MSc. We see that the non-linear effect of father income on the probability of becoming an inventor, is driven by individuals whose fathers have an MSc. In other words, having a wealthy father really helps only if the father has an MSc, not so much otherwise.

Figure 17 reproduces Figure 1, by breaking down individual according to both, whether they have a father with or without an MSc and whether the individual belongs or not to the top IQ decile. We see that the non-linear effect of father's income on the probability of becoming an inventor, is most strongly non-linear for individuals whose fathers have an MSc and who themselves belong to the highest IQ decile; next on the non-linearity scale come the individuals whose fathers have an MSc but who do not belong to the highest IQ decile; and then comes the individuals whose fathers do not have an MSc. In short, the nonlinearity in the relationship between father's income and the probability of becoming an inventor, seems to be mainly driven by father's education and to a lesser extent by the individual's IQ.

All these figures are descriptive and do not control for any individual characteristic. In the next subsection we shall derive similar figures from linear regressions of the probability of becoming an inventor on parental income, parental education, IQ, and own education, where we control for a whole set of observables (see below).

3.2 Regression equation

The regression equation that will serve as the basis for the estimations in this section, can be written as:

$$D_i = \alpha + \sum_n \beta_{fn} fcontrols_{in} + \sum_n \beta_{mn} mcontrols_{in} + \sum_k \theta_{fk} IQ_{ik} + \sum_j \theta_{fj} educ_{ij} + \epsilon_i$$

where: (i) D_i is a dummy variable for i being inventor, MD, or lawyer; (ii) $fcontrols$ are the observable variables measuring father characteristics; (iii) $mcontrols$ are the variables measuring mother characteristics; (iv) IQ_{ik} is a dummy variable that stands for individual i belonging to IQ scale k, (v) $educ_{ij}$ is a dummy variable that stands for individual i belonging to the education category j.

3.3 Regression analysis for "who becomes an innovator"

Here we regress the probability of becoming an inventor on parental income, parental education, the individual's IQ and finally the individual's own education. The dependent variable D_i is equal to 1 if the individual ever invents during the observation period, and to zero otherwise. Parental income is calculated in 1975 and 1985 for those parents for whom wages are observed at least one of these dates. For fathers that are too young to have income in 1985 we use the first year we observe in the FLEED, i.e., starting in 1988. Parental income is taken as the residual of a log (wage) regression on years of birth and years of wage measurement dummies.

We first regress D_i on parental income. The excluded income group for both parents is the lowest quintile; we include but do not report dummies for the 2nd - 4th quintile. For education (both parents and own education), the excluded group is base education. For IQ, the excluded group is the 5th IQ decile; we also include dummies for 1st - 4th and 6th - 8th IQ deciles but for space reasons we do not report the coefficients.

In all specifications below we include: a 4th order polynomial in (log) age, r21 region dummies; dummies for suburban and urban areas; dummies for Swedish and other than Finnish language as mother tongue; and parental decade of birth dummies (separately for both parents).

The results from that regression are shown in Figure 18 (which focuses on father's income but at all levels) and in column 1 in Table 1 (which focuses on effect of parental income at the highest income percentile but for both, father and mother).[5] We see from column 1 in Table 1 that having either the father or the mother belong to the highest income percentile has a positive and significant effect on the probability of becoming an inventor. However, having a father who is in the highest income percentile has 3 times larger effect than having a mother in the highest income percentile (1.84% versus 1.37%). Now turning to Figure 18, we see that it matches quite well our motivating non-parametric Figure 1, with this highly non-linear profile which is particularly steep at the highest levels of father's income. But the difference is that Figure 18 and the subsequent figures in this section, are derived from regressions where we control for parental date of birth, regional and urban dummies, language dummies, age dummies, etc.

[5] Showing comprehensive tables for our regressions, would take too much space, thus we chose to show shorter tables focusing only on the most interesting variables (for instance on the top income or IQ deciles or quintiles, on the top educational levels, etc.).

The positive impact of parental income can emerge through a number of reasons. A first reason might be that high-income parents can afford better schooling possibilities for their kids, hence high-income parents have better educated kids. A second possibility is that high-income parents increase the returns of their children's becoming inventors. A third possibility is that high-income parents are more educated and more educated parents in turn train (homeschool) their kids better. Fourth, it could be that high-income parents have higher ability types and through the persistence of types, high-ability parents might have high-ability kids. To further explore what underlies the observed correlation between parental income and the probability of becoming an inventor, we introduce additional controls into the above regression.

We first control for parental education, which yields a number of new interesting results, shown in Figure 19 and column 2 in Table 1 (Figure 19 again focuses on father's income but at all levels whereas column 2 in Table 1 focuses on the effect of parental income and education at the highest levels but for both, father and mother). First, we see that Figure 19 mirrors Figure 16, with still a non-linear curve which becomes steeper at the highest levels of father's income, but less so than in Figure 18. Next, from column 2 in Table 1 we see that having a father with a PhD has a direct and important impact on the probability of making an invention. Second controlling for parental education reduces the effect of the father belonging to the highest income quintile by half, and it reduces the effect of the mother belonging to the highest income quintile by almost two thirds.

Next, we control for the individual's visiospatial IQ, and the results are shown in Figure 20 and column 3 in Table 1. From Figure 20, which mirrors the high IQ curve in Figure 17, we see that the father income curve moves further down at the higher income percentiles when controlling for the individual's IQ belonging to the highest percentiles. Next, looking at column 3 of Table 1, we first that visiospatial IQ has a direct effect on the probability of becoming an inventor. Second, controlling for visiospatial IQ further reduces the effect of parental income on the probability of becoming an inventor.

Finally, we control for the individual's own education in Figure 21 and column 4 in Table 1. From Figure 21 we see that the father income curve shifts further down and becomes essentially flat except for the highest percentiles of father income when we control for the individual's own education. And from column 4 in Table 1 we see first a large direct effect of the Science PhD dummy on the probability of making an invention. Second, even after controlling for the individual's own education, the effect of IQ on the probability of becoming an inventor remains positive and significant. Third, once we control for the individual's own education, the effect of parental income is reduced dramatically and becomes insignificant, and in particular well below the effect of belonging to the highest IQ percentile. Having a PhD in Science has the largest effect: it increases the likelihood of becoming an inventor by 20.9%.

These latter findings suggest a prominent role for own education and for IQ when explaining an individual's probability of becoming an inventor. To further test this conjecture, we now compute partial R^2's in order to assess the relative explanatory powers of the observable background variables in our data sample. The findings, summarized in the table below, indicate that out of the variation in the probability of becoming an inventor which we can explain using all our observed variables: (i) the individual's own education comes first, explaining 81.7% of that variation; (ii) second comes the individual's IQ (16.8%); (iii) each of the remaining variables accounts for less than 1% of the total explained variation in the probability of becoming an inventor.

Decomposing the Explained Variation (Table 1 Col 5)

Parental Income	Parental Education	Parental Wealth	Own Education	Own IQ
0.1%	0.4%	0.9%	81.7%	16.8%

These findings raise an interesting puzzle: why do children with rich parents end up being more/better educated?

As already hinted at above, a first candidate explanation is that education is costly and individuals face credit constraints which prevent them from financing their studies. But education is totally free in Finland from kindergarten up to PhD. Alternatively, it may be the case that returns to education are higher for children born from richer parents, as richer parents may help their educated children overcome credit constraints to start a business. It may also be the case that children born from richer parents face lower opportunity costs of education, as richer parents tend to also be more educated parents who can provide complementary input in their children's studies. Or else it may be the case that richer parents have children with higher IQ which also impacts of the child's education level.

To help us understand the relationship between parental income and the child's education level, in Table 2A and 2B we regress the individual's schooling level on all our background variables. More specifically, we estimate the regression equation:

$$ownedu_i = \alpha + \sum_n \beta_{fn} fcontrols_{in} + \sum_n \beta_{mn} mcontrols_{in} + \sum_k \theta_{fk} IQ_{ik} + \epsilon_i$$

where: (i) $ownedu_i$ is a variable that takes values between 1 and 5 (1= base; 2 = 2ndary; 3 = college; 4 = MSc; 5 = PhD) in 2A, it is a dummy for having at least MSc in 2B; (ii) $fcontrols$, $mcontrols$, and IQ_{ik} are as before.

In a first attempt to assess the relative explanatory power of these variables, we look at partial

R^2's for the regression in Table 2. The results are summarized in the table below.

Decomposing the Explained Variation of Education Regressions

Parental Income	Parental Education	Parental Wealth	Own IQ
	Table 2A Column 4		
7.4%	24.6%	15.7%	52.2%
	Table 2B Column 4		
7.4%	43.4%	7.6%	41.5%

In particular we see that the individual's IQ and parental education have by far the highest explanatory power in determining the individual's level of education: namely, 52.2% of the explained variation in the individual's level of education is explained by the individual's IQ, followed by parental education (24.6%) and then by parental wealth (15.7%) and by parental income (7.4%) in Table 2A. Corresponding fractions are 41.5%, 43%, 7.6%, and 7.4%, respectively, in Table 2B.

Overall, our analysis in this section leads to the following conclusions:

1. There is a strong link between parental income and the probability of becoming an inventor;

2. This correlation is mostly due to the fact that richer parents end up having more educated children;

3. That children born from richer parents achieve a higher of education, appears to be primarily associated with the fact that: (i) richer parents are also more educated parents which lowers their children's opportunity cost of acquiring education; richer parents tend to have higher IQ children on average (see Figure 22).

4. Parental wealth and parental income matters on top of the other background variables: this suggests that credit constraints may also be at work, with a resulting effect of parental wealth/income on the return to education; however this matters less than the individual's IQ and parental education.

3.4 Becoming an inventor versus becoming a lawyer or a medical doctor

To which extent what we said above regarding the determinants of becoming an inventor, should not equally apply to other high-earning professions such as lawyer or medical doctor? In this subsection we perform the same regression exercises as in the previous subsection, but replacing the probability of becoming an inventor on the left-hand side of the regression equation by the probability of becoming a medical doctor or a lawyer.

A first remark: in our cross-section data sample, 0.92% of individuals are inventors, whereas 0.38% are medical doctors and 0.39% are lawyers. This will help us compare the magnitudes of the effects of parental income, parental education, IQ,...on the probability of becoming a lawyer or a medical doctor with the magnitudes of the effects of the same variables on the probability of becoming an inventor. For example, if we find the same coefficient for parental education in the regression tables for becoming an inventor as in the regression tables for becoming a lawyer, that will mean that the actual effect of parental income is $.92/.38 \approx 2.4$ higher on the probability of becoming an inventor than on the probability of becoming a medical doctor.

Figure 23 shows the three curves depicting respectively the probability of becoming an inventor, the probability of becoming a lawyer and the probability of becoming a medical doctor, as a function of father income, not controlling for any individual characteristic. We see that all three curves have similar shapes, with the same non-linear effect which becomes steeper at the highest levels of father's income. However the probability of becoming an inventor starts increasing already at the lowest levels of father income and lies significantly above the probabilities of becoming a lawyer or a medical doctor until we reach the highest father income percentiles. In other words, becoming an inventor is easier than becoming a lawyer or a medical doctor at all except the highest father income percentiles.

Table 3 shows results from the linear probability regressions for becoming an inventor, a medical doctor and a lawyer respectively, on parental income, parental education, the individual's IQ, and the individual's own education. When comparing the coefficients across columns one should bear in mind that 0.92% of individuals in our estimation sample are inventors, whereas 0.38% are MDs and 0.39% are lawyers.

Rows 1-6 and first three columns of Table 3 compare the effects of father and mother income belonging to the higher wage percentiles on the three probabilities, controlling for parental education and the child's IQ. There we see that the coefficients for father or mother income belonging to the higher income percentiles, are larger in the regression for the probability of becoming a medical doctor or a lawyer than in the regression for the probability of becoming an inventor, although we should keep in mind that there more than twice as many inventors than lawyers or medical doctors in the population.

Rows 7-12 and first three columns of Table 3 compare the effects of father and mother education on the three probabilities, controlling for the child's IQ. We find larger coefficients for parental education in the lawyer and medical doctor regressions, but once again this is counteracted by the fact that substantially more inventors than lawyers or medical doctors in our sample.

Perhaps the most striking conclusion comes from looking at rows 13-15 in Table 3, which compare the effects of the individual's IQ belonging to the higher IQ percentiles on the probabilities to become an inventor, a medical doctor, and a lawyer respectively. We see that having a high IQ matters much more for the probability of becoming an inventor than for the probability of becoming a lawyer or a medical doctor.

Overall, the main takeaways from Table 3 are:

1. Parental, especially father's, income is more important for becoming an MD or a lawyer, than for becoming an inventor. This speaks against the interpretation that the father's income percentile coefficients reflect credit constraints. The reason for this is that both MDs and lawyers are well-paid professions. As for other university degrees, there are essentially no tuition fees, but students get government grants and can take government-backed (low-interest; the system has evolved somewhat across the cohorts we observe) loans. It is thus unlikely that the father income percentile coefficients reflect credit constraints for these two professions, yet father income seems to matter more for them than for becoming an inventor.

2. Parental education has a larger impact on the probability of becoming an MD or a lawyer, than on the probability of becoming an inventor. This is true even for mother's education once one scales the coefficients with the probabilities of becoming an inventor, an MD, or a lawyer.

3. Visiospatial IQ is a (much) more important determinant of becoming an inventor, than for becoming an MD or a lawyer.

4 The returns to invention and social and income mobility of inventors

What are the returns to invention? In this section we analyze this question from three different angles. First, we look at the effect of innovation on the log wage income of individual inventors and on the log of wage income of other employees and entrepreneurs in the same firm. Second, we look at the effect of innovation on the probability for the inventor (relative to non-inventors) to make it to top income brackets when starting from outside these brackets (income mobility). Third, we look at the effect of innovation on the correlation between the individual's income and his father's income (social mobility).

4.1 Returns to innovation

In this subsection we regress the log of wage income[6] in subsequent periods on making an invention in the current period. We consider two main groups of treated individuals, namely: (i) the inventors; (ii) other individuals in the same firm. For each group, we consider the impact of patent application this year on returns over the next ten years. Inventors in our sample earn 70,000 euro per annum on average, whereas non-inventors earn 25,000 euro per annum on average.

[6]We get very similar results by regressing the log of wage plus capital income on the same set of explanatory variables. Results are available upon request to the authors.

The basic regression equation to capture the dynamic returns from innovation, can be written as:

$$\ln inc_{it} = \alpha_i + \sum_{\tau=0,..,10} \beta_\tau pcount_{it,-\tau} + \sum_{\tau=0,..,10} \theta_\tau cowork_{it,-\tau} + X'_{it}\omega + \varepsilon_{it},$$

where α_i is the individual's fixed effect, inc_{it} denotes individual i's wage income at time t, $pcount_{it,-\tau}$ denotes the patent count of individual i at time $t-\tau$, $cowork_{it,-\tau}$ is a dummy equal to one if individual i was a coworker at time $t-\tau$, and X'_{it} is a vector of controls which includes log age (4th order polynomial), region dummies, urban dummies, year dummies, and individual fixed effects.

We extend this basic regression equation: (i) by adding citation variables to capture the quality of the invention; (ii) by distinguishing between different types of agents in the same firm as the inventor, in particular: senior managers, junior managers, base blue collar workers, senior and junior white collar workers, and entrepreneurs in the same firm;[7] (iii) the control variables in X'_{it} include: log age (4th order polynomial), regional dummies, urban dummies, year dummies.

In Table 4, each row represents a different lag of the "treatment" variable in question. Thus, in the first column of panel A of Table 4, the coefficient in the first row (0.0461) represents the OLS estimate of the wage increase per patent application to an individual in the year his patent application has been submitted. The coefficient in the second row of the same column (0.0275) is the estimated wage increase one year after the patent application. Finally, the coefficient in the last row of the first column (0.0060) is our OLS estimate of the wage increase of an inventor 10 years after the patent application.

Each column in Table 4 represents the coefficients for a different (vector of) treatment variables. The first one gives the coefficients on the patent count of the individual himself (contemporaneous plus 1st to 10th period lag); the second the coefficients for the dummy for an individual being the coworker of an inventor - our base coworker category is a blue-collar worker; the third the extra return on top of the blue-collar coworker's return for a senior manager; the fourth similarly the extra return for a senior white-collar worker on top of the blue-collar coworker's return; and finally, the fifth column the extra returns that an entrepreneur who is a coworker of the inventor gets on top of the return to the blue-collar worker. In addition to the reported coworker types, we include separate (vectors of) dummies for junior managers, junior white-collar workers, "other" (= the residual category in Statistics Finland socioeconomic grouping) coworkers, and (though there are extremely few), agricultural coworkers.

Table 4 has three panels: panel A reports the OLS results; panel B the results from a specification that includes individual fixed effects; and panel C the results from a specification that includes

[7] The entrepreneur category in our database, comprises the self-employed plus all the individuals who (alone or with family) own at least one half of a company subject to limited liability, and who work for that company.

not only the patent count, but also the citation count to measure the returns to the quality of invention.

Tables 4A, 4B and 4C summarize the most interesting regression results. First, we see that innovating and thereby increasing the patent count by one unit at any year u induces a significant wage increase over the ten year period starting in year u. When controlling for individual fixed effects (Table 4B), we see that the per annum wage increases for the inventor lie between 0.5% and 1.87%.

A quick comparison between panels A and B of Table 4 reveals that individual fixed effects are needed to control for the fact that individuals who at some point invent, or work with an inventor, would be earning higher wages even without inventing - the OLS coefficients are throughout larger than the fixed effects coefficients.

Column 2 of Table 4B shows the returns from innovation for base blue-collar coworkers of the inventor in the same firm. We see that their wage is enhanced during the first three years after the innovation year, but is reduced thereafter. This in turn may reflect the fact that innovation leads firms to eventually replace existing blue-collar workers by competing workers from outside. The most interesting column is column 5 which shows the returns to being an entrepreneur in the same firm (the actual return to entrepreneurs each year, is obtained by adding the coefficients columns 2 and column 5 in the row corresponding to that year). In particular we see that after six years, entrepreneurs earn far more than the inventor (compare between the coefficients in columns 1 and the sum of the coefficients from columns 2 and 5 for each row). Namely, entrepreneurs always gain between 3.9% (=0.0395-0.0004) and 6.25% (=0.0639-0.0014) over years 6 to 10 after the invention year.

Column 1 of Table 4C shows the results from regressing the log of wage income on all previous variables plus the citation counts to capture the intensive margin of innovation

We see that each additional citation to a patent applied for in year u, increases the per annum wage by between 0.45 and 0.67% over the four year period starting in year u.

To summarize our findings in this subsection: (i) an increase in patent count has significant and sizeable effects on the wage of the inventor through the ten year period starting in the invention year; (ii) the invention benefits coworkers during the first years after the invention, but in the longer term it has contrasting effects on blue collar workers versus more upstream agents in the firm: it decreases the wage of blue collar workers whereas it considerably enhances the wage of entrepreneurs in the firm.

Final remark; since we controlled for individual fixed effects, we capture the return from invention beyond any potential selection effect.

4.2 Innovation and income mobility

Here we look at the extent to which innovation helps an individual's wage move upward between ages 35 and 45 compared the dynamic wage profile of a non-inventor. More formally we estimate the following equation

$$owninc45_i = \alpha + \beta owninc35_i + \theta owninc35_i \times inventor_i + \gamma inventor_i + X_i'\omega + \epsilon_i$$

where: (i) $owninc45$ and $owninc35$ are the individual's own income percentile at age 45 and 35, respectively; (ii) *inventor* is a dummy that takes value 1 if an individual invents before age 33; (iii) X is the same vector of controls as in the "who becomes an inventor" regression.

The base sample for intragenerational (income) mobility includes all individuals for whom we have IQ data, irrespective of their employer. We initially include in our estimation sample all individuals from the base sample from whom we have their income at age 35 and at age 45. Our control group consists of individuals who never invent. Our "treatment" group consists of individuals who: (i) had not invented by age 33 (= 35 – 2); (ii) invented by age 43 (= 45 – 2). We exclude from the estimation sample those individuals who invent before age 33, and those that invent for the first time after age 43 to make the overall sample comparable across individuals. We thus regress the wage percentile of a 45 years old individual on his wage percentile at age 35, an inventor dummy, and interactions between the inventor dummy and initial income characteristics. Table 5 shows the most interesting results from this regression.

Column 1 of Table 5 takes the whole population of non-inventors as the control group. From there we see that for non inventors the wage at age 35 is a main determinant of the wage at age 45. But by far the dominant coefficient is on the inventor dummy: in other words, inventing at age 33 has a large effect on the wage at age 45, and conditional upon inventing, the initial wage matters very little for the wage at age 45.

Column 2 of Table 5 narrows down the control group of non-inventors using the Coarsened Exact Matching (CEM) methodology. The idea is that the control group of non-inventors should share the maximum possible number of observable characteristics with the inventor, thereby helping us argue that in the regression we are capturing the effects of innovation on income mobility beyond selection. We did not need to resort to any such methodology in the previous subsection as the wage regression in that subsection was performed using panel data. This in turn allowed us to control for individual fixed effects, and thereby to deal with the selection issue. But in the mobility regressions we perform in this and the next subsection, we cannot use panel data and therefore we must find another way to address the selection issue. Therefore what we do is to construct coarsened exact matching cells using the discrete variables corresponding to father income quintiles, mother income quintiles, father education levels, mother education levels, IQ levels, father date of birth, mother date of birth, etc. Then we throw away all those cells for which we do not have at least one inventor and one non-inventor. And then we run the wage regression described above taking as control group for each inventor the non-inventor(s) in the same cell, and we weight the various cells by the number of individuals in that cell.

Comparing between column 1 and column 2 of Table 5, we see that all the effects remain almost identical when moving from a control group comprising all non-inventors in the sample to a more restricted control group constructed through the CEM method. This in turn allows us to argue that the above effects of innovation on income go beyond selection.

4.3 Innovation and social mobility

In this subsection we look at the extent to which innovation increases cross-generational mobility, measured as in Chetty et al (2014). Here we look at the extent to which innovation increases cross-generational mobility, measured as in Chetty et al (2014).

More specifically, we estimate the regression equation:

$$owninc35_i = \alpha + \beta fatherinc_i + \theta fatherinc_i \times inventor_i + \gamma inventor_i + X_i'\omega + \epsilon_i$$

where: (i) $owninc35$ is the individual's own income percentile at age 35; (ii) $fatherinc$ is the father's income percentile; (iii) $inventor$ is a dummy that takes value 1 if an individual invents before age 33; (iv) X is the same vector of controls as in the "who becomes an inventor" regression.

The base sample for intergenerational (social) mobility is the same sample as for intragenerational mobility. We then include all individuals for whom we observe: (i) the father's income; (ii) the individual's own income at age 35.

The individual's own income is measured at ages 34, 35, and 36 and we take the mean over the 3 years if all these are observed. If income at age 36 is not observed, we take the average over wages at ages 34 and 35. And if wage at age 35 is not observed, the individual is not in the sample.

Next, we compute the father's percentile rank based on the residual from a regression of father income on father year of birth dummies and year of wage measurement dummies. We measure father income by wage in 1975 if father is no longer working in 1985, or by the average of wages in 1975 and 1985 if father is working in both periods, or by the wage in 1985 if the father is not working in 1975, or by the first observed wage in FLEED (almost always 1988) if father is not yet working in 1985.

Table 6 shows the most interesting results from regressing an individual's wage percentile at age 35 on his father's wage percentile, an inventor dummy, an interaction between the inventor dummy and the father's wage percentile, and an interaction between the individual's being in the top IQ percentile and his father's income percentile. Column 1 of Table 6 includes all non-inventors in the control group, whereas column 2 uses the CEM method to restrict the control group to non-inventors that essentially share the same observable characteristics with inventors

The results are very similar to those in Table 5, but here we consider intergenerational (social) mobility rather than intragenerational (income) mobility. First, for non-inventors, the father's income percentile has a determinant effect on the individual's wage percentile. Second, the correlation between father and son income is greatly reduced for inventors, as the coefficient on the inventor dummy is far greater than the coefficient on the father's income percentile. Finally, moving from a broad control group comprising all non-inventors to a more restricted control group using the CEM method, leads to almost identical regression coefficients, which in turn implies that the effects uncovered here go beyond selection.

5 Conclusion

In this paper we have exploited the merging between three data sets -namely individual income data, patenting data, and IQ data- to analyze the selection into becoming an inventor and the returns to invention in Finland over the period 1988-2012. First, looking at the effects of parental income, parental education and the child's IQ and education on the probability of becoming an inventor, we found that:

1. There is a strong link between parental income and the probability of becoming an inventor;
2. This correlation is mostly due to the fact that richer parents end up having more educated children;
3. That children born from richer parents achieve a higher of education, appears to be primarily associated with the fact that: (i) richer parents are also more educated parents which lowers their children's opportunity cost of acquiring education; richer parents tend to have higher IQ children on average (see Figure 22).
4. Parental wealth and parental income matters, which suggests that credit constraints are also at work (with a resulting effect of parental wealth/income on the return to education) although to a much lower extent than the individual's IQ and parental education.

In the second part of the paper we looked at how becoming an innovator affects an individual's income and on income and social mobility. First, in wage/returns regressions, we found that: (i) making an innovation increases the annual wage rate by significant amounts over a prolonged period after the invention; (ii) more highly cited innovations yield a higher revenue to the innovator; (iii) coworkers in the same firm also benefit from an innovation, the highest returns being earned by entrepreneurs in the firm, especially after a few years; moreover, the returns are higher for entrepreneurs than for the inventor herself.

Next, looking at income and social mobility, we showed that making an innovation enhances both, intragenerational and intergenerational social mobility, and that being an inventor very effectively reduces the father-son income relation. In that section we also showed that the effects of innovation on wage income and income and social mobility, remain strong when controlling for individual fixed effects (in the wage regressions) or when restricting the control group of non-inventors to non-inventors with identical observable background characteristics (in the income and social mobility regressions). In other words, these effects of innovation on returns and mobility go beyond selection, i.e. beyond the fact that inventors would share particular characteristics among themselves and/or with other high-return activities.

Overall, our analysis in this second part suggests that inventing boosts an individual's position in the income distribution and that the observed rise in top income inequality is not solely driven by forces related to misallocation. At the same time, inventing act as an equalizer that (almost completely) destroys the link between father income and son income, and much more so than education or IQ.

We plan to extend our current analysis in several directions. A first extension is to replicate our analysis for other countries: do we get a pattern always similar to that in Figure 1[8] for the relationship between parental income and the probability of becoming an inventor, and do we explain it primarily by education and IQ (as we did here for Finland) or more by credit constraints? A second extension would be to look at how income mobility of inventors depends upon characteristics of the firm or the sector, in particular firm size, firm age, the degree of competition in the firm's sector. These and other extensions of the analysis in this paper are left to future research.

[8] We already know that Bell et al (2015) and Akcigit et al (2016) obtain a similar pattern in the US, respectively using contemporaneous data and historical data.

References

[1] Acemoglu, D., Akcigit, U., Bloom, N., and Kerr, W (2013). "Innovation, Reallocation and Growth." NBER Working Paper 18993.

[2] Aghion, P., Akcigit, U., and Howitt, P (2014), "What Do We Learn from Schumpeterian Growth Theory?", in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Vol 2B: 515-563.

[3] Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., and Hemous, D (2015), "Innovation and Top Income Inequality", mimeo Harvard.

[4] Aghion, P., and Howitt, P (1992), "A Model of Growth Through Creative Destruction", *Econometrica*, 60, 323-351

[5] Aghion, P., and Tirole, J (1994), "The Management of Innovation", *Quarterly Journal of Economics,* 109, 1185-1209

[6] Akcigit, U., Grigsby, J., and Nicholas, T (2016), "The Birth of American Ingenuity: Innovation and Inventors of the Golden Age", University of Chicago Working Paper.

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

[8] Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J (2015), "The Lifecycle of Inventors", mimeo Harvard.

[9] Black, S. , and Devereux, P (2011), "Recent Developments in Intergenerational Mobility", In O. Ashenfelter and D. Card, eds., *Handbook of Labour Economics*, Amsterdam: Elsevier, Vol. 4B, 1487-1541.

[10] Celik, M (2015), "Does the Cream Always Rise to the Top? The Misallocation of Talent in Innovation", mimeo University of Pennsylvania

[11] Jaravel, X., Petkova, N., and Bell, A (2015), "Team-Specific Capital and Innovation", mimeo Harvard

[12] Holmstrom, B (1982) "Managerial Incentive Problems: A Dynamic Perspective," in *Essays in Economics and Management in Honor of Lars Wahlbeck,* Helsinki: Swedish School of Economics.

[13] Holmstrom, B (1989), "Agency Costs and Innovation", *Journal of Economic Behavior and Organization*,12, 305-327.

[14] Hsieh, C. and Klenow, P (2009). "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124, 1403-1448.

[15] Hsieh, C., Hurst, E., Jones, C. and Klenow, P (2013), "The Allocation of Talent and US Economic Growth", NBER Working Paper 18693.

[16] Jäntti, M., Österbacka, E., Raaum, O., Eriksson, T. and Björklund, A (2002), "Brother correlations in earnings in Denmark, Finland, Norway and Sweden Compared to the United States", *Journal of Population Economics,* 15, 757-772.

[17] Lazear, E (1997), "Incentives in Basic Research," *Journal of Labor Economics* 15(1): S167-S197.

[18] Lerner, J., and Malmendier, U (2010), "Contractibility and the Design of Research Agreements," *American Economic Review,* 100, 214-246.

[19] Lerner, J., and Wulf, J (2007), "Innovation and Incentives: Evidence from Corporate R&D", *Review of Economics and Statistics,* 89, 634-644.

[20] Manso, G (2011), "Motivating Innovation", *Journal of Finance,* 66, 1823-1860.

[21] Novak, S., and Stern, S (2008), "How Does Outsourcing Affect Performance Dynamics? Evidence from the Automobile Industry", *Management Science,* 54, 1963-1979.

[22] Pekkarinen, T., Uusitalo, R., and Pekkala Kerr, S (2009), "School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform", *Journal of Public Economics,* 93, 965-973.

[23] Romer, P (1990). "Endogenous Technical Change." *Journal of Political Economy*, 98, 71-102.

[24] Scotchmer, S (2004), *Innovation and Incentives.* Cambridge, MA: The MIT Press.

[25] Solon, G (1999), "Intergenerational Mobility in the Labor Market", in O.C. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3A. Amsterdam: Elsevier, 1761-1800.

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

[27] Toivanen, O., and Vaananen, L (2015), " Education and Invention", *Review of Economics and Statistics,* forthcoming.

Tables

Table 1: WHO BECOMES REGRESSIONS

Variable	Group	Parental Income	+Parental Educ	+IQ	+Own Educ	+Parental Wealth
Father Wage	w(father)_90-94	0.0127***	0.00727***	0.00587***	0.00250***	0.00295***
		(0.000747)	(0.000743)	(0.000740)	(0.000717)	(0.000747)
	w(father)_95-99	0.0167***	0.00866***	0.00699***	0.00302***	0.00370***
		(0.000838)	(0.000830)	(0.000827)	(0.000805)	(0.000854)
	w(father)_100	0.0184***	0.00698***	0.00538***	-0.000527	-0.000262
		(0.00210)	(0.00209)	(0.00208)	(0.00203)	(0.00222)
Mother Wage	w(mother)_90-94	0.00418***	0.00140**	0.000582	-1.69e-05	-0.000166
		(0.000675)	(0.000684)	(0.000682)	(0.000662)	(0.000690)
	w(mother)_95-99	0.00782***	0.00206**	0.00111	-0.000447	-0.000642
		(0.000802)	(0.000844)	(0.000841)	(0.000818)	(0.000860)
	w(mother)_100	0.0137***	0.00436**	0.00309	-0.000584	-0.00204
		(0.00205)	(0.00208)	(0.00207)	(0.00201)	(0.00206)
Father Education	Secondary(father)		-0.000293	-0.00156***	-0.00177***	-0.00166***
			(0.000397)	(0.000396)	(0.000387)	(0.000394)
	College(father)		0.00598***	0.00330***	-0.000491	-0.000493
			(0.000715)	(0.000713)	(0.000696)	(0.000713)
	Master(father)		0.0105***	0.00758***	0.000798	0.000768
			(0.000880)	(0.000877)	(0.000855)	(0.000869)
	PhD(father)		0.0291***	0.0256***	0.0104***	0.0113***
			(0.00287)	(0.00287)	(0.00277)	(0.00288)
Mother Education	Secondary(mother)		0.00390***	0.00294***	0.000806***	0.000658**
			(0.000277)	(0.000276)	(0.000268)	(0.000276)
	College(mother)		0.00688***	0.00493***	0.000393	0.000303
			(0.000665)	(0.000663)	(0.000648)	(0.000661)
	Master(mother)		0.0104***	0.00814***	0.00223*	0.00242**
			(0.00121)	(0.00120)	(0.00117)	(0.00118)
	PhD(mother)		0.0121*	0.00923	-0.00114	-0.00197
			(0.00652)	(0.00651)	(0.00631)	(0.00650)
IQ	IQ_90-94			0.0119***	0.00616***	0.00569***
				(0.000891)	(0.000856)	(0.000879)
	IQ_95-99			0.0200***	0.00904***	0.00896***
				(0.000977)	(0.000931)	(0.000961)
	IQ_100			0.0388***	0.0193***	0.0196***
				(0.00266)	(0.00255)	(0.00270)
Own Education	Secondary(own)				-7.94e-05	-0.000517
					(0.000298)	(0.000329)
	College(own)				-0.00264***	-0.00307***
					(0.000343)	(0.000380)
	Master(own)				-0.000331	-0.000986
					(0.000604)	(0.000647)
	PhD(own)				-0.000404	-0.00402*
					(0.00274)	(0.00236)
Own (Science) Education	Sci_Secondary(own)				-0.00295***	-0.00351***
					(0.000205)	(0.000231)
	Sci_College(own)				0.0135***	0.0121***
					(0.000575)	(0.000605)
	Sci_Master(own)				0.0918***	0.0911***
					(0.00175)	(0.00185)
	Sci_PhD(own)				0.209***	0.204***
					(0.00716)	(0.00755)
Father Wealth	wealth(father)_90-94					0.000110
						(0.000753)
	wealth(father)_95-99					-0.000103
						(0.000749)
	wealth(father)_100					0.00141
						(0.00161)
Mother Wealth	wealth(mother)_90-94					0.000616
						(0.000759)
	wealth(mother)_95-99					-0.000303
						(0.000806)
	wealth(mother)_100					0.000773
						(0.00177)
	Observations	696,348	696,348	696,348	696,348	625,609

Notes: Robust standard errors in parantheses. Additional controls include 4th order polynomial in log inventor age, 21 region dummies, urban/suburban/rural dummies, language dummies, and father and mother birth decades.

Table 2A: Determinants of Education Regressions:
$ownedu$=1(BASE); =2(2NDARY); =3(COLLEGE); =4(MSC); =5(PHD)

Variable	**Group**	**Parental Income**	**+Parental Educ**	**+IQ**	**+Parental Wealth**
Father Wage	w(father)_90-94	0.406***	0.240***	0.201***	0.179***
		(0.00509)	(0.00508)	(0.00497)	(0.00524)
	w(father)_95-99	0.522***	0.288***	0.244***	0.216***
		(0.00550)	(0.00560)	(0.00550)	(0.00589)
	w(father)_100	0.723***	0.402***	0.361***	0.341***
		(0.0133)	(0.0131)	(0.0129)	(0.0140)
Mother Wage	w(mother)_90-94	0.208***	0.0969***	0.0727***	0.0890***
		(0.00494)	(0.00495)	(0.00485)	(0.00510)
	w(mother)_95-99	0.334***	0.146***	0.120***	0.135***
		(0.00547)	(0.00569)	(0.00560)	(0.00596)
	w(mother)_100	0.413***	0.177***	0.148***	0.167***
		(0.0123)	(0.0125)	(0.0124)	(0.0131)
Father Education	Secondary(father)		0.159***	0.123***	0.120***
			(0.00305)	(0.00300)	(0.00306)
	College(father)		0.312***	0.245***	0.246***
			(0.00496)	(0.00490)	(0.00502)
	Master(father)		0.358***	0.288***	0.290***
			(0.00605)	(0.00599)	(0.00612)
	PhD(father)		0.550***	0.472***	0.476***
			(0.0149)	(0.0148)	(0.0153)
Mother Education	Secondary(mother)		0.155***	0.128***	0.110***
			(0.00217)	(0.00211)	(0.00218)
	College(mother)		0.267***	0.220***	0.196***
			(0.00460)	(0.00455)	(0.00465)
	Master(mother)		0.260***	0.209***	0.185***
			(0.00768)	(0.00763)	(0.00775)
	PhD(mother)		0.327***	0.270***	0.249***
			(0.0368)	(0.0366)	(0.0381)
IQ	IQ_90-94			0.159***	0.151***
				(0.00515)	(0.00531)
	IQ_95-99			0.302***	0.284***
				(0.00545)	(0.00564)
	IQ_100			0.448***	0.434***
				(0.0119)	(0.0126)
Father Wealth	wealth(father)_90-94				0.0596***
					(0.00496)
	wealth(father)_95-99				0.0641***
					(0.00502)
	wealth(father)_100				0.0442***
					(0.0108)
Mother Wealth	wealth(mother)_90-94				0.0400***
					(0.00509)
	wealth(mother)_95-99				0.0587***
					(0.00539)
	wealth(mother)_100				0.0880***
					(0.0117)
	Observations	696,348	696,348	696,348	625,609

Notes: Robust standard errors in parantheses. Additional controls include 4th order polynomial in log inventor age, 21 region dummies, urban/suburban/rural dummies, language dummies, and father and mother birth decades.

Table 2B: DETERMINANTS OF EDUCATION REGRESSIONS: *ownedu*=MASTER AND ABOVE DUMMY

Variable	Group	Parental Income	+Parental Educ	+IQ	+Parental Wealth
Father Wage	w(father)_90-94	0.122***	0.0601***	0.0510***	0.0440***
		(0.00202)	(0.00198)	(0.00196)	(0.00205)
	w(father)_95-99	0.177***	0.0849***	0.0743***	0.0646***
		(0.00227)	(0.00226)	(0.00223)	(0.00236)
	w(father)_100	0.271***	0.139***	0.129***	0.120***
		(0.00597)	(0.00581)	(0.00573)	(0.00624)
Mother Wage	w(mother)_90-94	0.0584***	0.0190***	0.0136***	0.0158***
		(0.00192)	(0.00192)	(0.00189)	(0.00199)
	w(mother)_95-99	0.119***	0.0445***	0.0385***	0.0393***
		(0.00231)	(0.00235)	(0.00233)	(0.00247)
	w(mother)_100	0.172***	0.0668***	0.0592***	0.0583***
		(0.00540)	(0.00545)	(0.00540)	(0.00570)
Father Education	Secondary(father)		0.0402***	0.0320***	0.0311***
			(0.00115)	(0.00114)	(0.00116)
	College(father)		0.107***	0.0902***	0.0900***
			(0.00205)	(0.00203)	(0.00209)
	Master(father)		0.150***	0.132***	0.131***
			(0.00255)	(0.00252)	(0.00258)
	PhD(father)		0.242***	0.221***	0.223***
			(0.00654)	(0.00647)	(0.00668)
Mother Education	Secondary(mother)		0.0391***	0.0329***	0.0290***
			(0.000753)	(0.000741)	(0.000764)
	College(mother)		0.0959***	0.0840***	0.0785***
			(0.00191)	(0.00189)	(0.00193)
	Master(mother)		0.106***	0.0924***	0.0854***
			(0.00326)	(0.00323)	(0.00329)
	PhD(mother)		0.147***	0.130***	0.124***
			(0.0159)	(0.0158)	(0.0164)
IQ	IQ_90-94			0.0576***	0.0556***
				(0.00213)	(0.00220)
	IQ_95-99			0.109***	0.105***
				(0.00230)	(0.00238)
	IQ_100			0.181***	0.181***
				(0.00540)	(0.00568)
Father Wealth	wealth(father)_90-94				0.0177***
					(0.00199)
	wealth(father)_95-99				0.0168***
					(0.00203)
	wealth(father)_100				0.0111**
					(0.00442)
Mother Wealth	wealth(mother)_90-94				0.0169***
					(0.00205)
	wealth(mother)_95-99				0.0274***
					(0.00222)
	wealth(mother)_100				0.0387***
					(0.00499)
	Observations	696,348	696,348	696,348	625,609

Notes: Robust standard errors in parantheses. Additional controls include inventor age, region, urban/rural, language, and father and mother birth decades.

Table 3: WHO BECOMES INVENTOR, DOCTOR, OR LAWYER?

Variable	**Group**	**Inventor**	**Medical Doctor**	**Lawyer**
Father Wage	w(father)_90-94	0.00580***	0.00218***	0.00273***
		(0.000772)	(0.000511)	(0.000552)
	w(father)_95-99	0.00714***	0.00441***	0.00626***
		(0.000878)	(0.000624)	(0.000705)
	w(father)_100	0.00558**	0.0281***	0.0144***
		(0.00228)	(0.00271)	(0.00231)
Mother Wage	w(mother)_90-94	0.000566	0.000938*	0.00192***
		(0.000708)	(0.000495)	(0.000527)
	w(mother)_95-99	0.000998	0.00207***	0.00349***
		(0.000884)	(0.000665)	(0.000707)
	w(mother)_100	0.00169	0.0121***	0.00503***
		(0.00213)	(0.00213)	(0.00191)
Father Education	Secondary(father)	-0.00149***	0.00229***	0.00398***
		(0.000404)	(0.000269)	(0.000304)
	College(father)	0.00333***	0.00419***	0.00517***
		(0.000731)	(0.000499)	(0.000523)
	Master(father)	0.00741***	0.00908***	0.0117***
		(0.000890)	(0.000700)	(0.000793)
	PhD(father)	0.0262***	0.0279***	0.0131***
		(0.00299)	(0.00269)	(0.00215)
Mother Education	Secondary(mother)	0.00247***	0.00141***	0.00130***
		(0.000284)	(0.000159)	(0.000171)
	College(mother)	0.00442***	0.00455***	0.00360***
		(0.000676)	(0.000505)	(0.000521)
	Master(mother)	0.00774***	0.00604***	0.00587***
		(0.00122)	(0.000945)	(0.00102)
	PhD(mother)	0.00807	0.0175**	0.00654
		(0.00673)	(0.00690)	(0.00551)
IQ	IQ_90-94	0.0113***	0.00163***	-0.000315
		(0.000914)	(0.000536)	(0.000492)
	IQ_95-99	0.0196***	0.00437***	-0.000157
		(0.00101)	(0.000617)	(0.000526)
	IQ_100	0.0392***	0.00416***	-0.00111
		(0.00282)	(0.00143)	(0.00114)
Father Wealth	wealth(father)_90-94	0.00143*	0.00126**	0.000766
		(0.000776)	(0.000512)	(0.000501)
	wealth(father)_95-99	0.00103	0.000385	0.000368
		(0.000774)	(0.000520)	(0.000522)
	wealth(father)_100	0.000768	-0.00247**	0.00177
		(0.00165)	(0.00123)	(0.00140)
Mother Wealth	wealth(mother)_90-94	0.00207***	0.000571	0.00110**
		(0.000782)	(0.000513)	(0.000550)
	wealth(mother)_95-99	0.00158*	0.00286***	0.00185***
		(0.000830)	(0.000627)	(0.000626)
	wealth(mother)_100	0.00150	0.000131	0.00294*
		(0.00184)	(0.00143)	(0.00158)
	Observations	625,609	625,609	625,609

Notes: Robust standard errors in parantheses. Additional controls include 4th order polynomial in log inventor age, 21 region dummies, urban/suburban/rural dummies, language dummies, and father and mother birth decades.

Table 4: RETURNS TO INNOVATIONS

PANEL A: OLS

time	inventor	coworker	senior manager	senior w-c	entrepreneur
t=0	0.0461***	0.0282***	0.0292***	0.0002	0.1547***
	(0.0036)	(0.0006)	(0.0026)	(0.0013)	(0.0529)
t=1	0.0275***	0.0266***	0.0171***	-0.0007	0.1891***
	(0.0035)	(0.0005)	(0.0020)	(0.0010)	(0.0433)
t=2	0.0199***	0.0092***	0.0208***	0.0000	0.1018***
	(0.0037)	(0.0005)	(0.0019)	(0.0010)	(0.0330)
t=3	0.0204***	0.0120***	0.0025	-0.0094***	0.0001
	(0.0035)	(0.0005)	(0.0020)	(0.0010)	(0.0438)
t=4	0.0244***	0.0113***	0.0134***	-0.0081***	0.0295
	(0.0033)	(0.0005)	(0.0020)	(0.0010)	(0.0396)
t=5	0.0328***	0.0103***	0.0111***	-0.0058***	0.0544*
	(0.0030)	(0.0005)	(0.0021)	(0.0010)	(0.0288)
t=6	0.0317***	0.0074***	0.0203***	-0.0034***	0.0606***
	(0.0037)	(0.0005)	(0.0022)	(0.0010)	(0.0209)
t=7	0.0378***	0.0188***	0.0246***	-0.0088***	0.0319
	(0.0038)	(0.0005)	(0.0024)	(0.0011)	(0.0218)
t=8	0.0413***	0.0163***	0.0244***	-0.0041***	0.0766***
	(0.0040)	(0.0005)	(0.0025)	(0.0011)	(0.0164)
t=9	0.0400***	0.0290***	0.0133***	-0.0128***	0.0384***
	(0.0038)	(0.0005)	(0.0029)	(0.0012)	(0.0146)
t=10	0.0540***	0.0456***	0.0103***	-0.0095***	0.0511***
	(0.0034)	(0.0006)	(0.0036)	(0.0015)	(0.0165)
Observations			7,285,011		

PANEL B: FIXED EFFECT REGRESSION

time	inventor	coworker	senior manager	senior w-c	entrepreneur
t=0	0.0187***	0.0089***	-0.0037*	-0.0019*	0.0763
	(0.0020)	(0.0005)	(0.0019)	(0.0011)	(0.0516)
t=1	0.0116***	0.0080***	0.0077***	0.0030***	0.1695***
	(0.0020)	(0.0005)	(0.0017)	(0.0009)	(0.0336)
t=2	0.0071***	0.0027***	-0.0011	0.0015	0.0630**
	(0.0023)	(0.0005)	(0.0017)	(0.0009)	(0.0294)
t=3	0.0063***	0.0008*	0.0012	0.0020**	-0.0276
	(0.0022)	(0.0005)	(0.0018)	(0.0009)	(0.0311)
t=4	0.0059**	-0.0023***	0.0037**	0.0030***	0.0438
	(0.0027)	(0.0004)	(0.0018)	(0.0009)	(0.0317)
t=5	0.0099***	-0.0012***	0.0051***	0.0022**	0.0256
	(0.0022)	(0.0004)	(0.0019)	(0.0010)	(0.0249)
t=6	0.0072***	-0.0012***	0.0076***	0.0042***	0.0535***
	(0.0025)	(0.0004)	(0.0020)	(0.0010)	(0.0196)
t=7	0.0089***	-0.0004	0.0137***	0.0023**	0.0395**
	(0.0025)	(0.0004)	(0.0020)	(0.0010)	(0.0178)
t=8	0.0073***	-0.0014***	0.0093***	0.0053***	0.0639***
	(0.0026)	(0.0004)	(0.0023)	(0.0010)	(0.0130)
t=9	0.0049	0.0057***	0.0002	0.0007	0.0562***
	(0.0032)	(0.0004)	(0.0025)	(0.0010)	(0.0134)
t=10	0.0060**	0.0010**	-0.0056**	0.0019*	0.0404***
	(0.0025)	(0.0005)	(0.0026)	(0.0011)	(0.0129)
Observations			7,285,011		

- Table 4 continued on next page -

Table 4 (CONT'D): RETURNS TO INNOVATIONS

PANEL C: FIXED EFFECT WITH INTENSIVE MARGIN

time	inventor (Pat count)	inventor (cites)	coworker	senior manager	senior w-c	entrepreneur
t=0	0.0167***	0.0061**	0.0089***	-0.0037*	-0.0019*	0.0763
	(0.0020)	(0.0029)	(0.0005)	(0.0019)	(0.0011)	(0.0516)
t=1	0.0103***	0.0055**	0.0080***	0.0077***	0.0030***	0.1695***
	(0.0020)	(0.0025)	(0.0005)	(0.0017)	(0.0009)	(0.0336)
t=2	0.0055**	0.0067***	0.0027***	-0.0012	0.0015	0.0630**
	(0.0023)	(0.0024)	(0.0005)	(0.0017)	(0.0009)	(0.0294)
t=3	0.0050**	0.0045*	0.0008*	0.0012	0.0020**	-0.0276
	(0.0022)	(0.0024)	(0.0005)	(0.0018)	(0.0009)	(0.0311)
t=4	0.0062**	0.0004	-0.0023***	0.0037**	0.0030***	0.0438
	(0.0027)	(0.0023)	(0.0004)	(0.0018)	(0.0009)	(0.0317)
t=5	0.0095***	0.0026	-0.0012***	0.0051***	0.0022**	0.0256
	(0.0023)	(0.0025)	(0.0004)	(0.0019)	(0.0010)	(0.0249)
t=6	0.0071***	0.0020	-0.0012***	0.0076***	0.0042***	0.0535***
	(0.0024)	(0.0019)	(0.0004)	(0.0020)	(0.0010)	(0.0196)
t=7	0.0072***	0.0050**	-0.0004	0.0137***	0.0023**	0.0395**
	(0.0026)	(0.0020)	(0.0004)	(0.0020)	(0.0010)	(0.0178)
t=8	0.0102***	-0.0025	-0.0014***	0.0093***	0.0053***	0.0639***
	(0.0027)	(0.0020)	(0.0004)	(0.0023)	(0.0010)	(0.0130)
t=9	0.0076**	-0.0014	0.0057***	0.0002	0.0007	0.0562***
	(0.0034)	(0.0018)	(0.0004)	(0.0025)	(0.0010)	(0.0134)
t=10	0.0069***	0.0019	0.0010**	-0.0056**	0.0019*	0.0404***
	(0.0026)	(0.0021)	(0.0005)	(0.0026)	(0.0011)	(0.0129)
Observations	7,285,011					

Notes: Robust standard errors in parantheses. Additional controls include 4th order polynomial in log inventor age, 21 region dummies, urban/suburban/rural dummies, year dummies.

Table 5: SOCIAL MOBILITY

	All	CEM
father wage percentile	0.131***	0.0993***
	(0.00184)	(0.0101)
inventor × father percentile	-0.138***	-0.126***
	(0.0139)	(0.0158)
high IQ × father percentile	0.00520	0.0215*
	(0.00383)	(0.0122)
inventor	31.62***	30.96***
	(0.980)	(1.100)
observations	359,861	82,583

Notes: Robust standard errors in parantheses. Additional controls include language dummies, year dummies, father and mother birth decades, father and mother education, own IQ, and own education.

Table 6: INCOME MOBILITY

	All	CEM
wage percentile at 35	0.629***	0.633***
	(0.00270)	(0.0117)
inventor × p35	-0.159***	-0.150***
	(0.0497)	(0.0527)
high IQ × p35	-0.0260***	-0.0438**
	(0.00621)	(0.0200)
inventor	14.76***	14.07***
	(4.405)	(4.677)
observations	117,493	28,044

Notes: Robust standard errors in parantheses. Additional controls include language dummies, year dummies, father and mother birth decades, father and mother education, own IQ, and own education.

Figures

Figure 1:

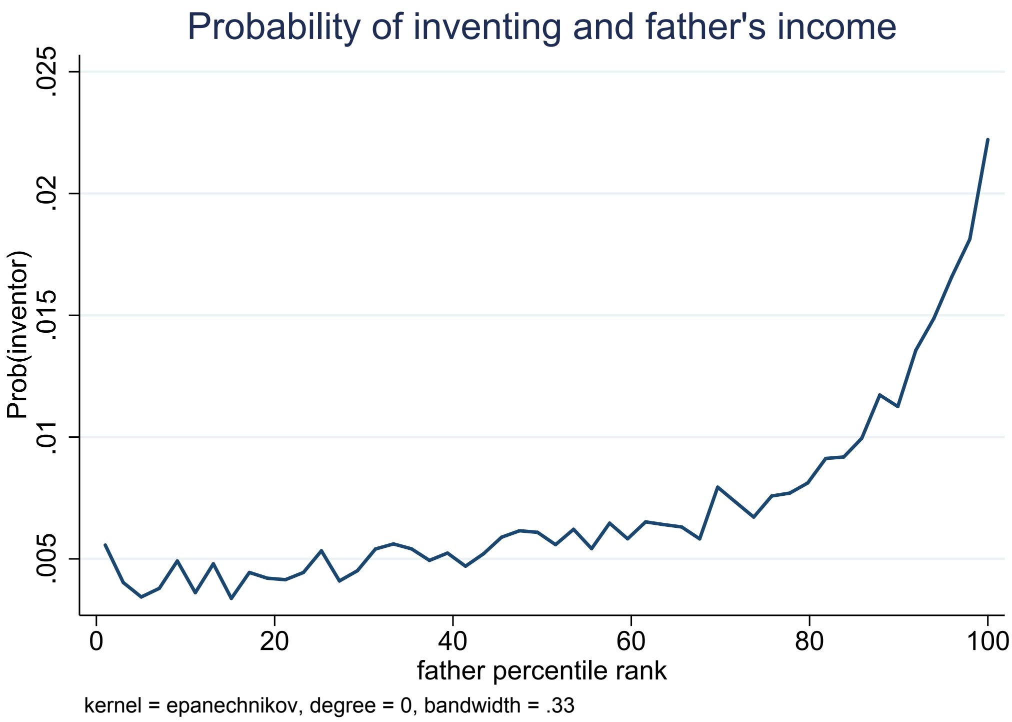
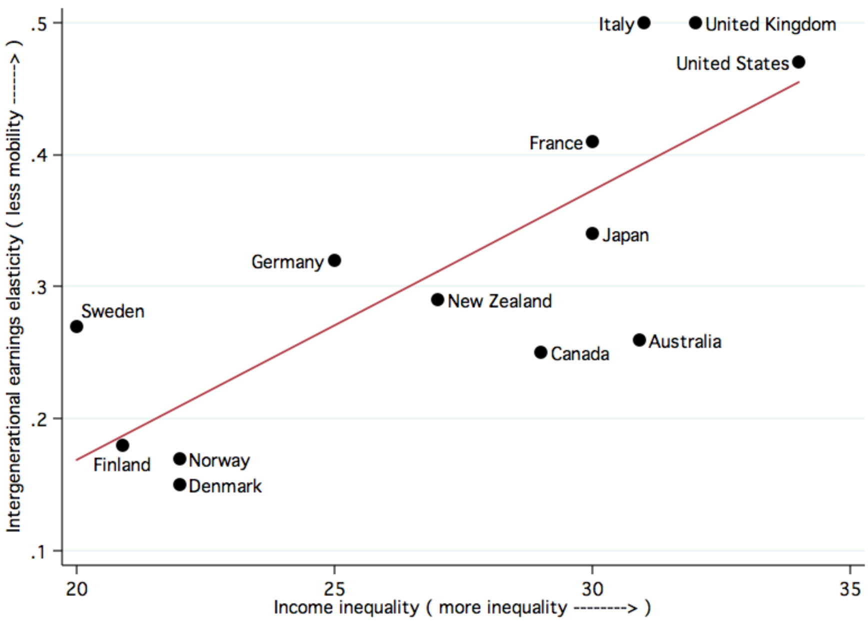


Figure 2: THE GREAT GATSBY CURVE



Source: Corak (2013)

Figure 3:

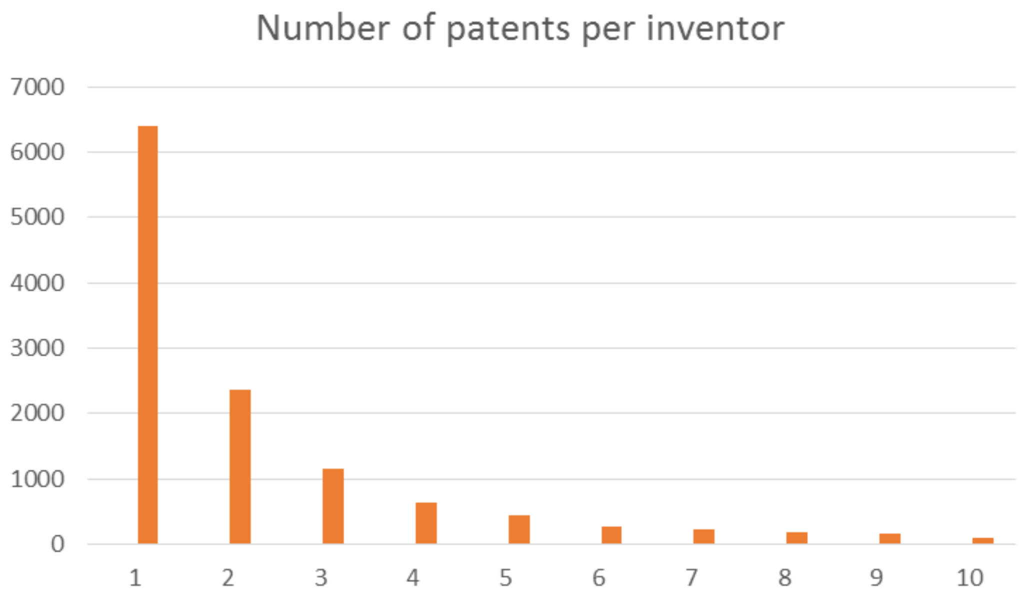


Figure 4:

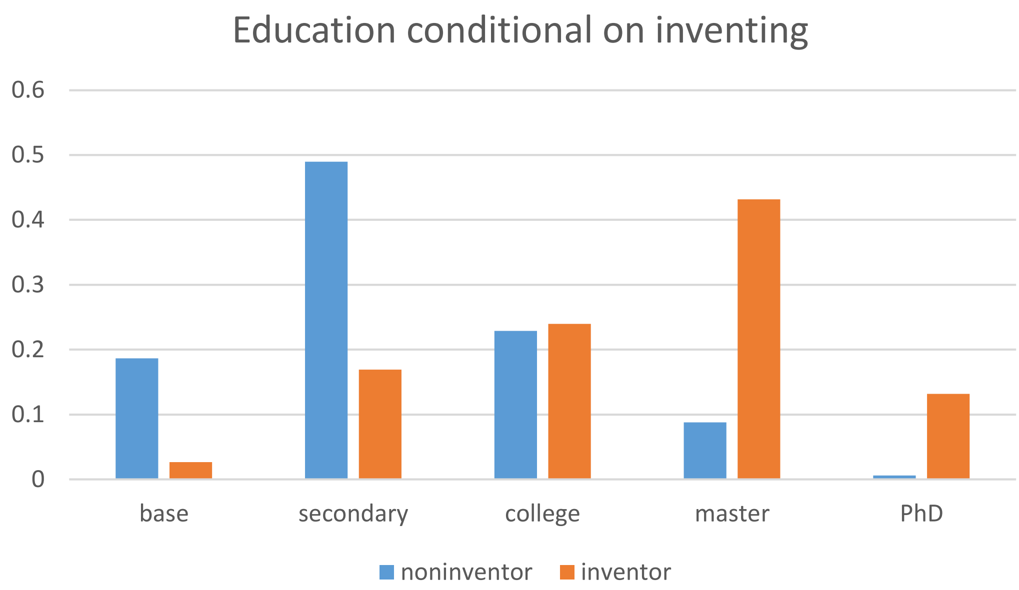


Figure 5:

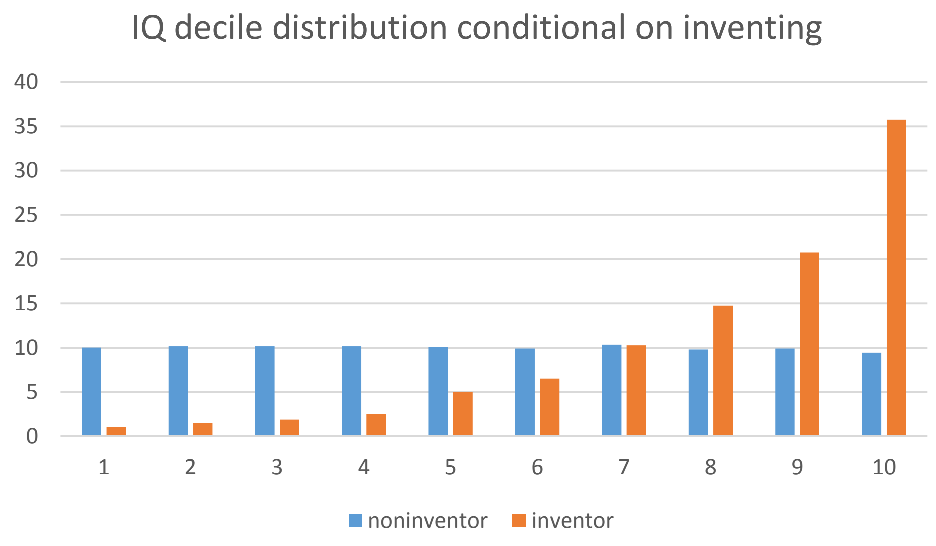


Figure 6:

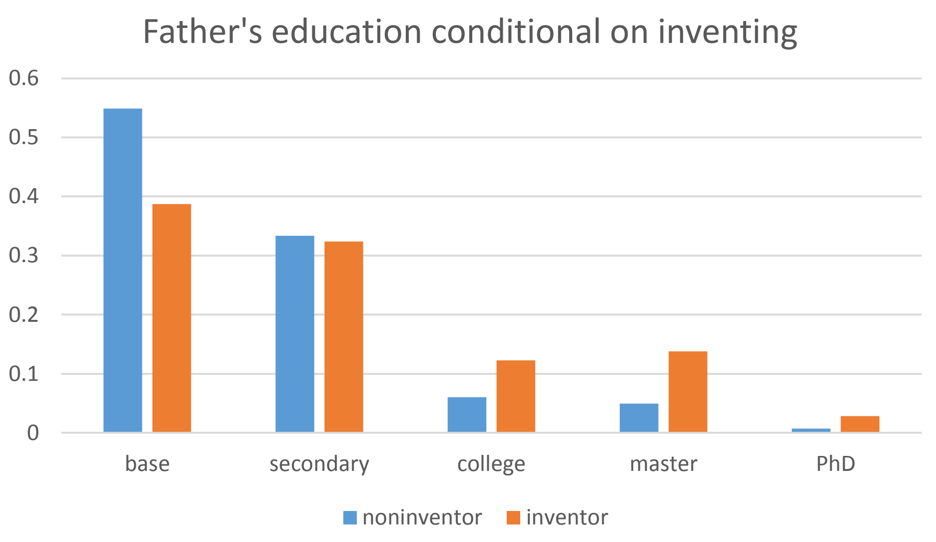


Figure 7:

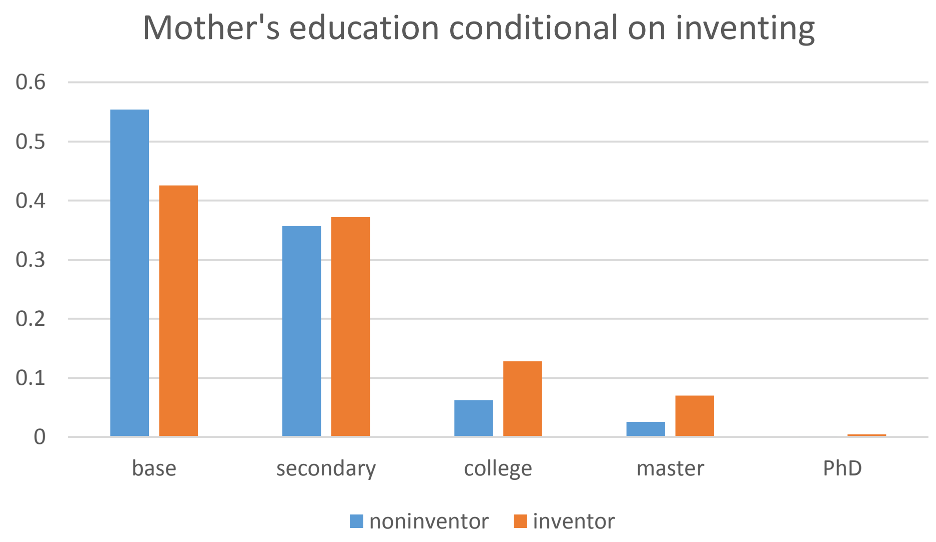


Figure 8:

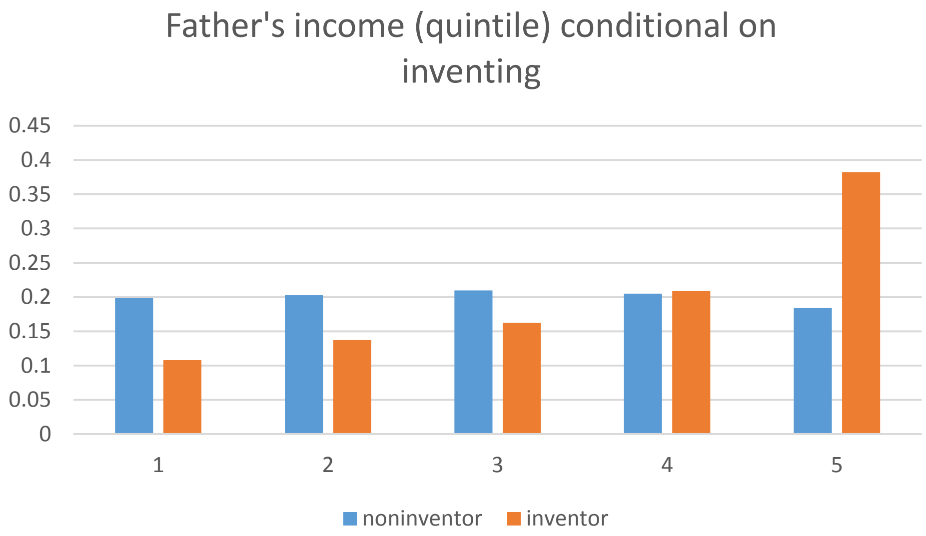


Figure 9:

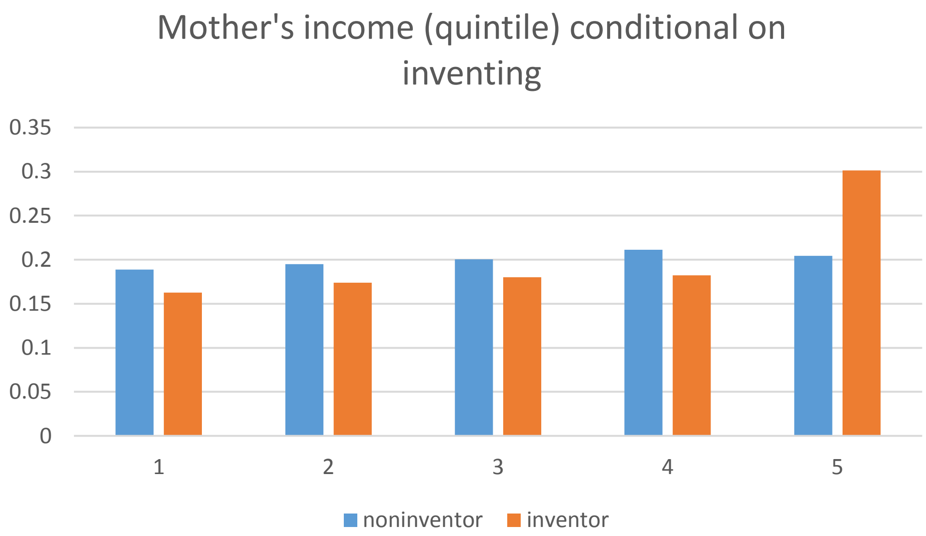


Figure 10:

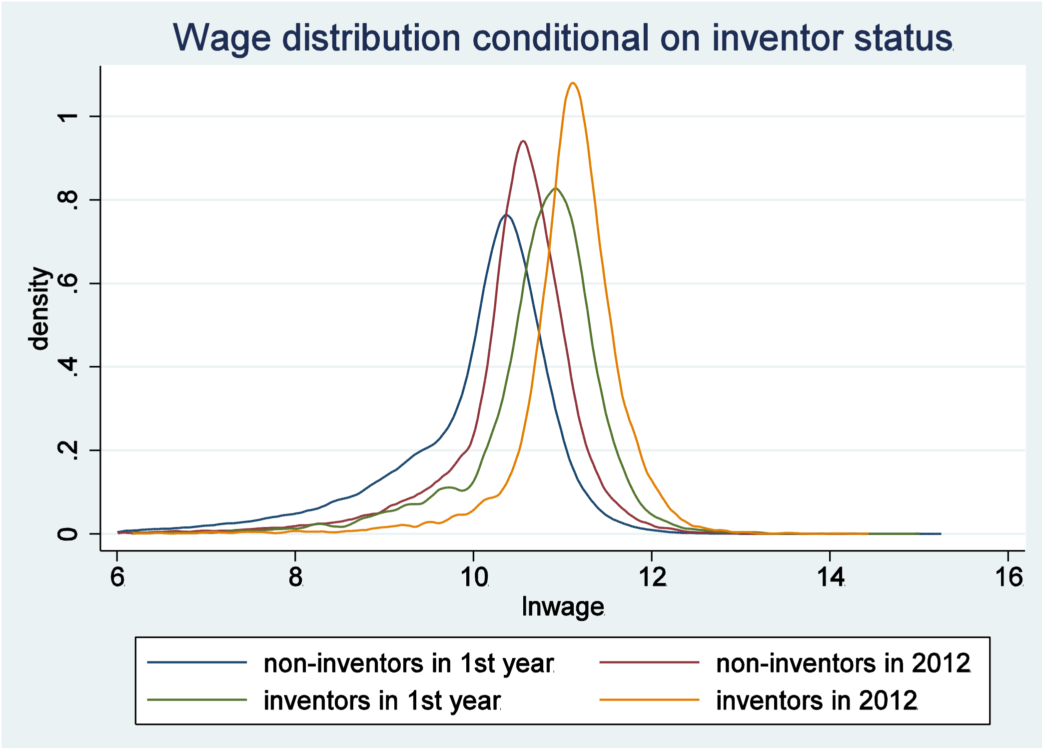


Figure 11:

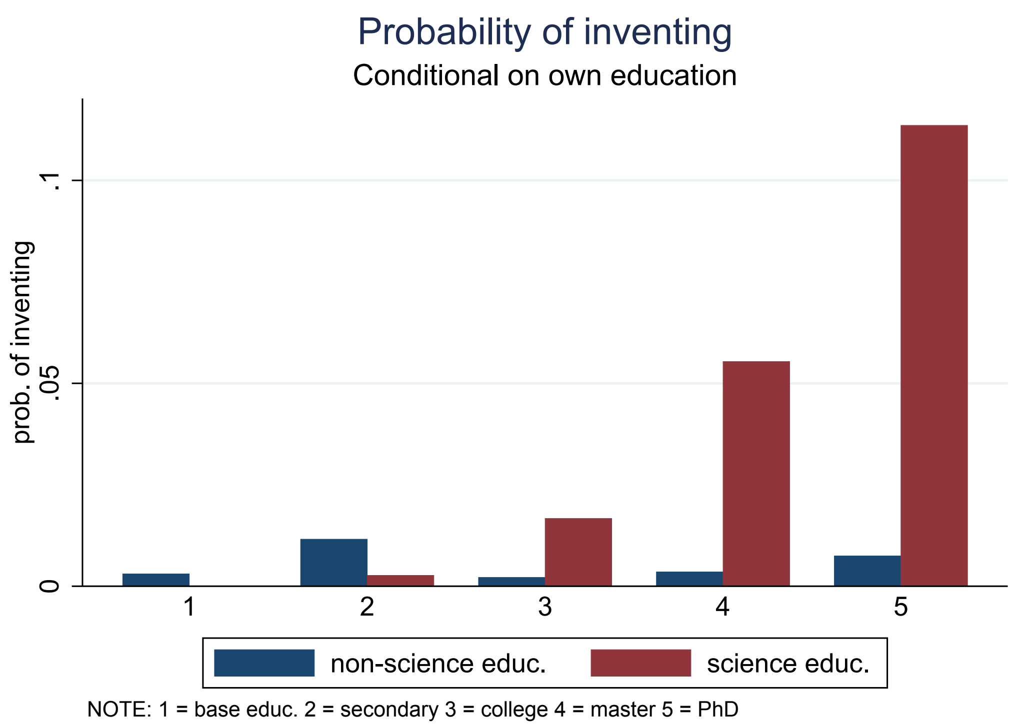


Figure 12:

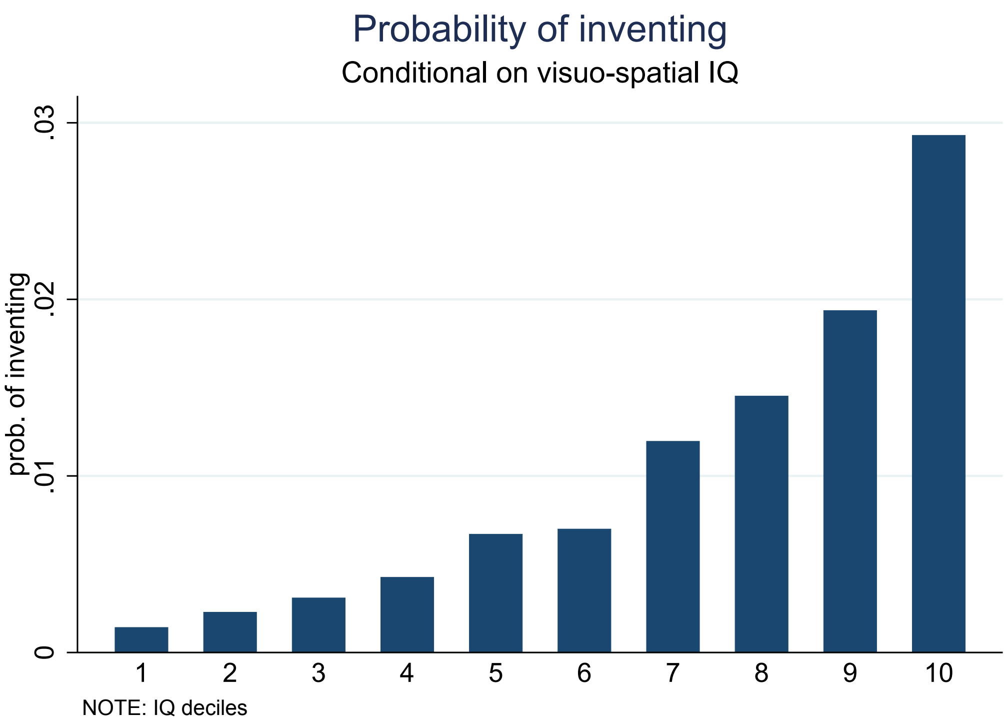


Figure 13:

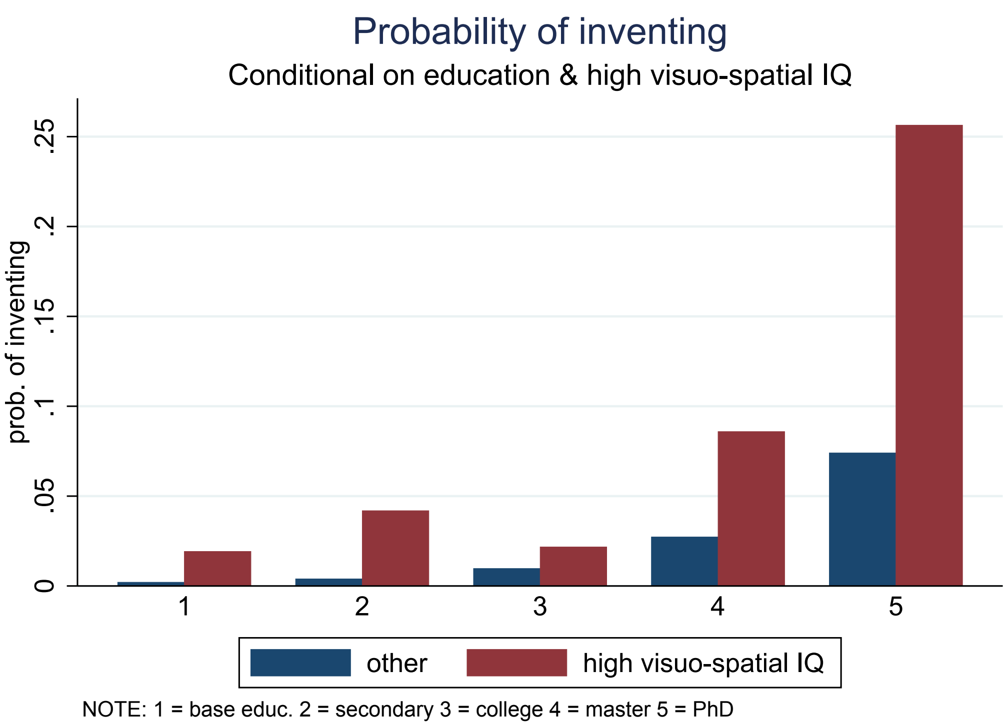


Figure 14:

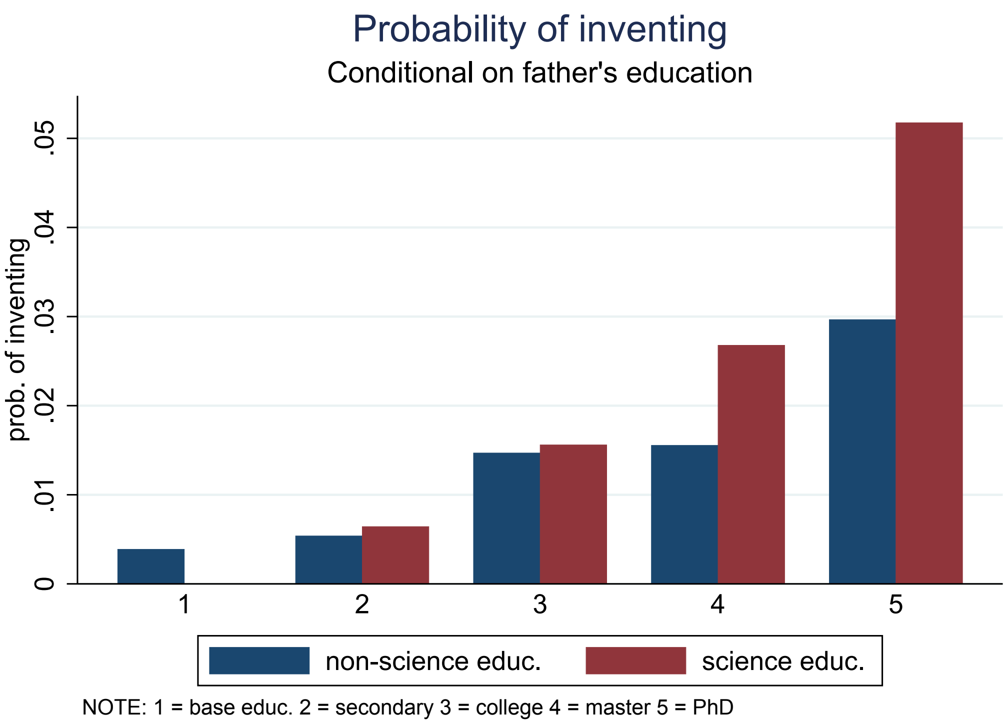


Figure 15:

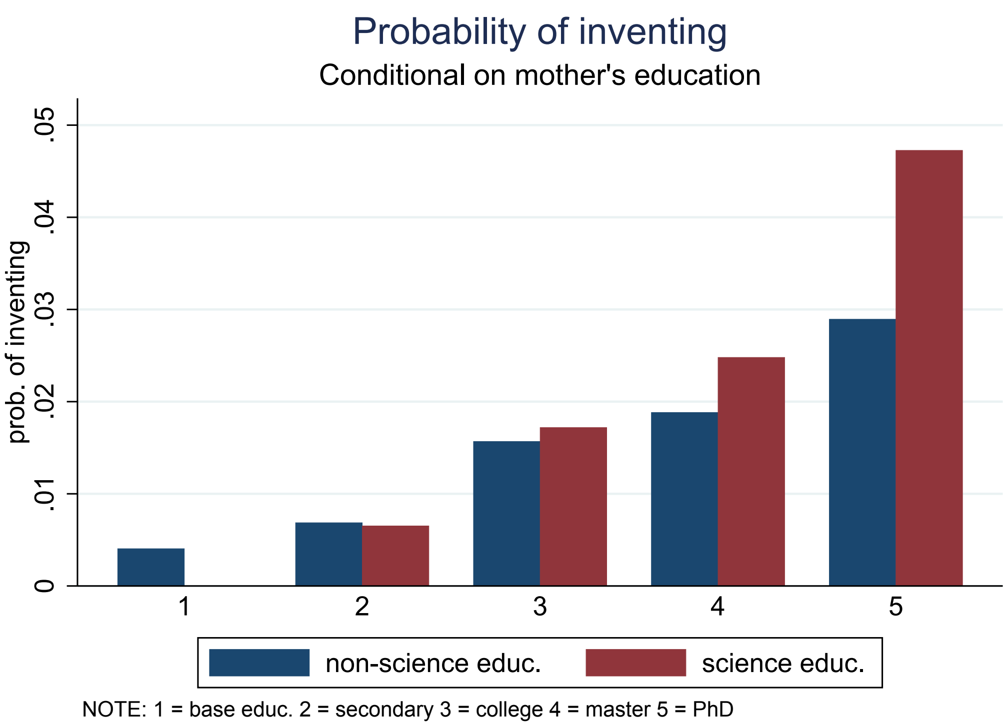


Figure 16:

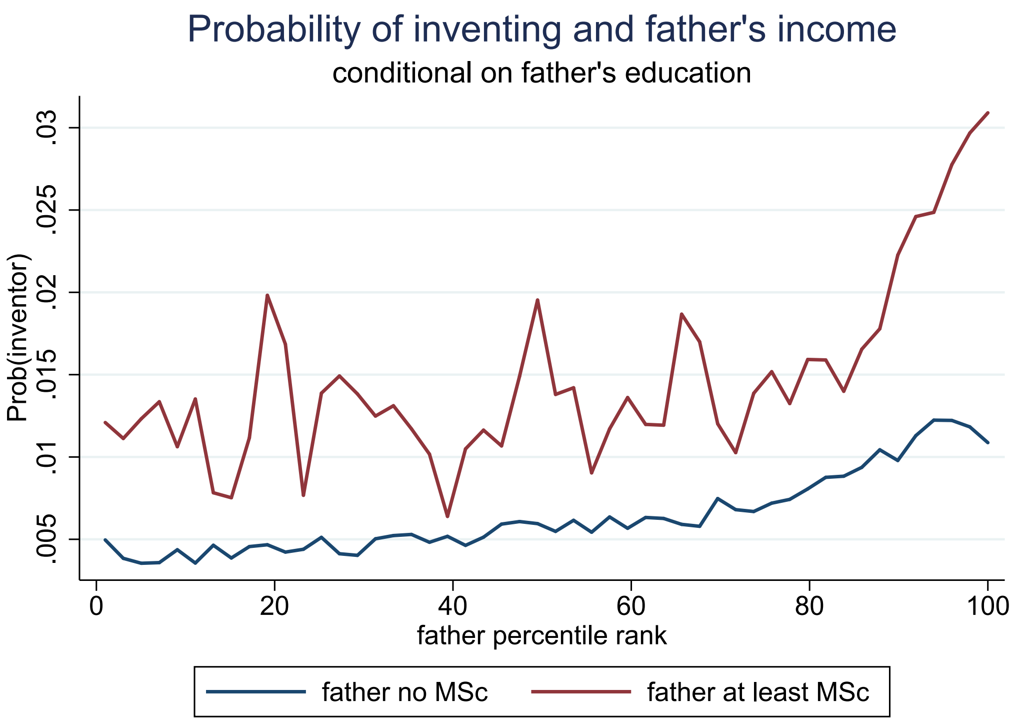


Figure 17:

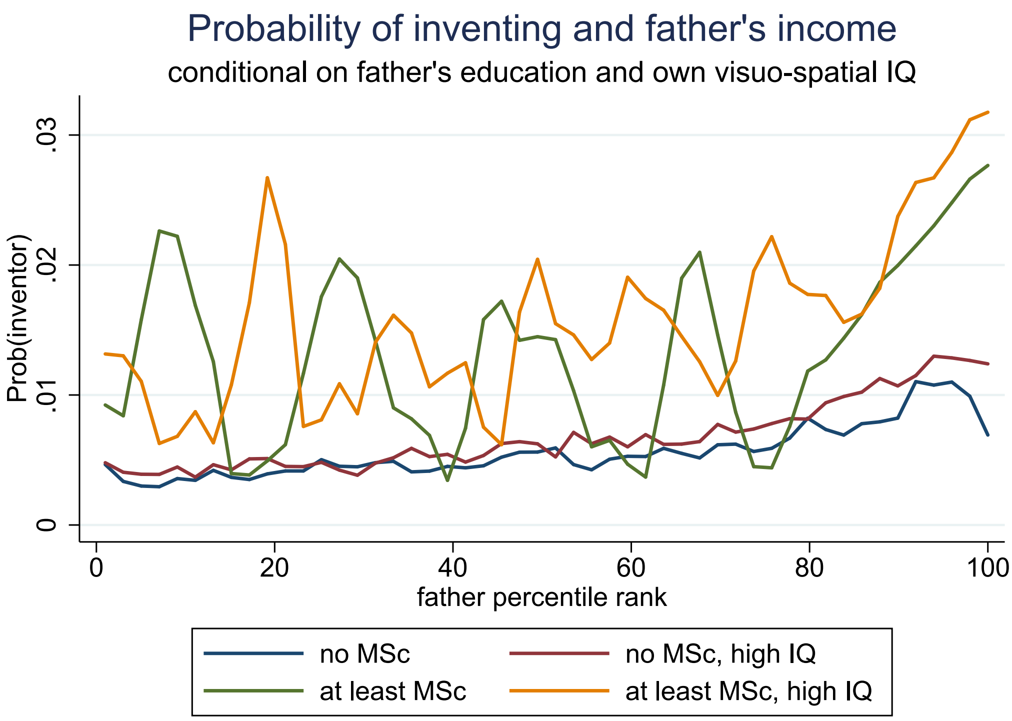


Figure 18:

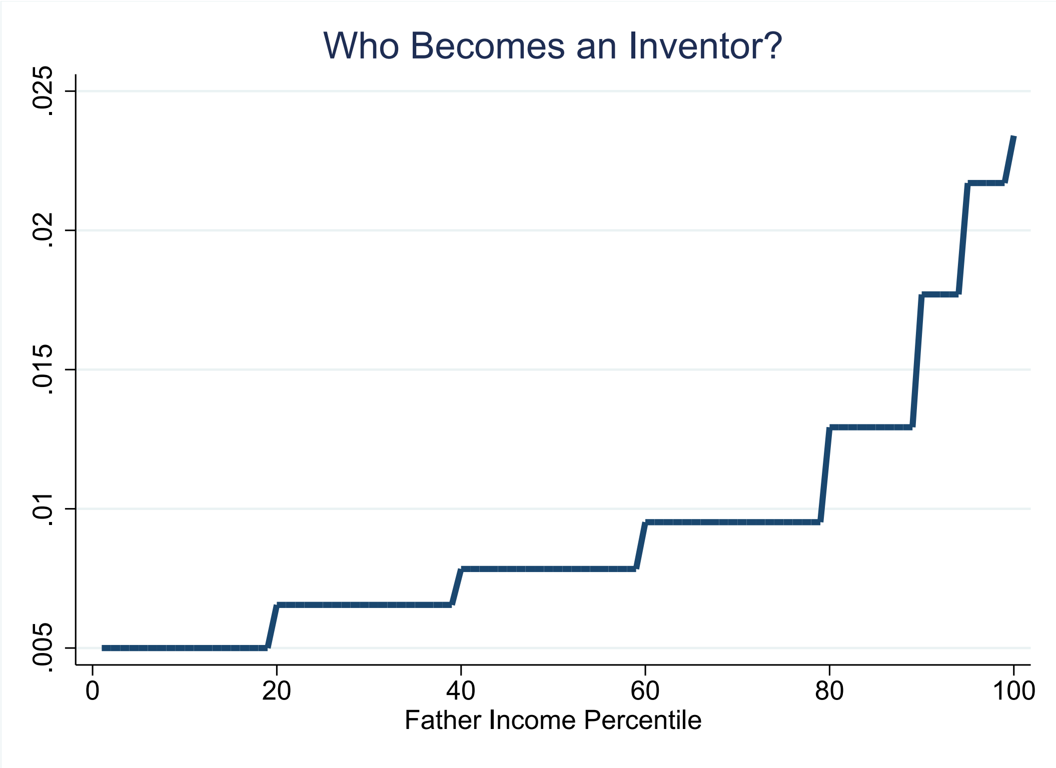


Figure 19:

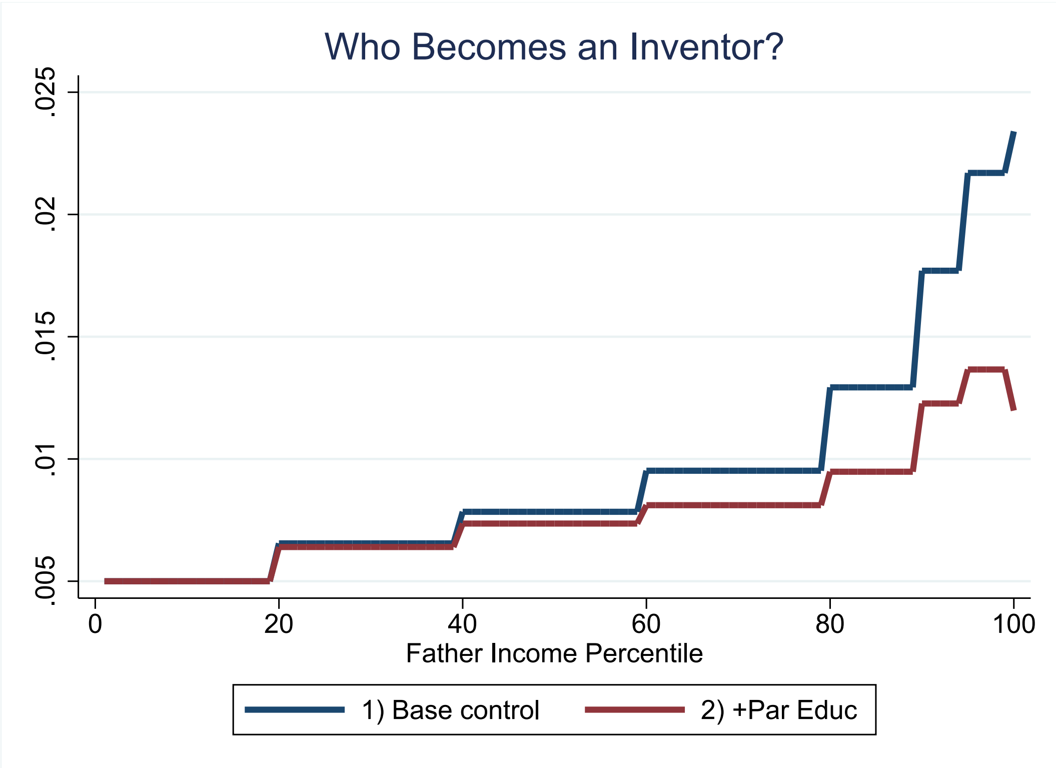


Figure 20:

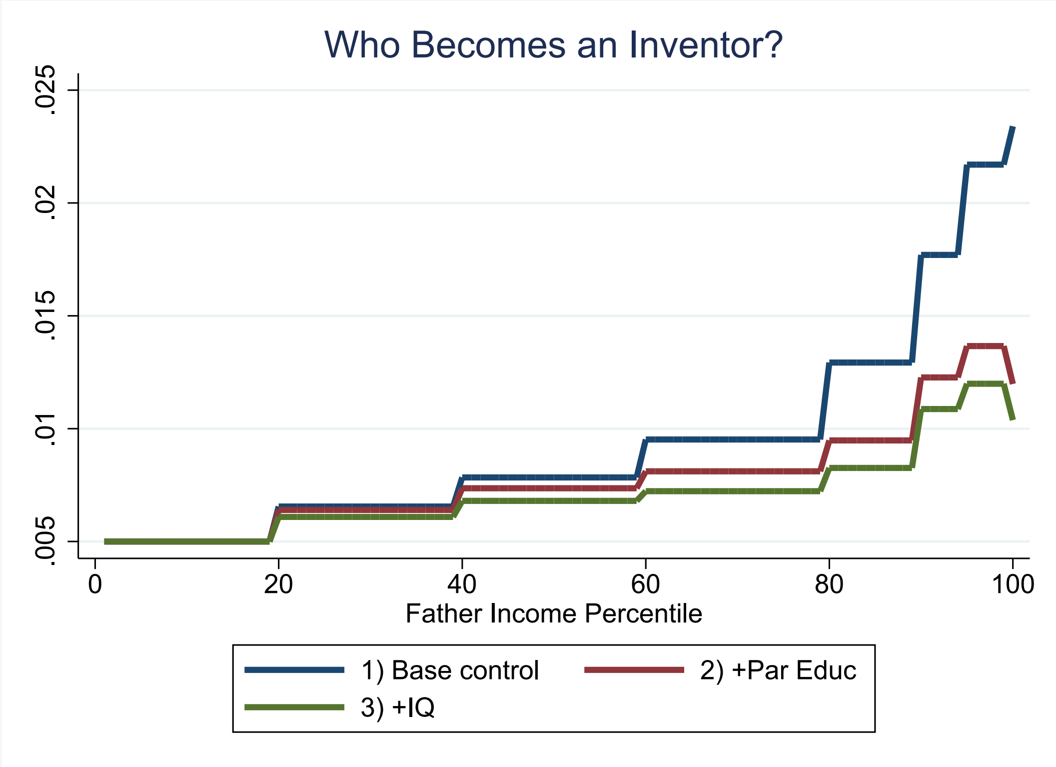


Figure 21:

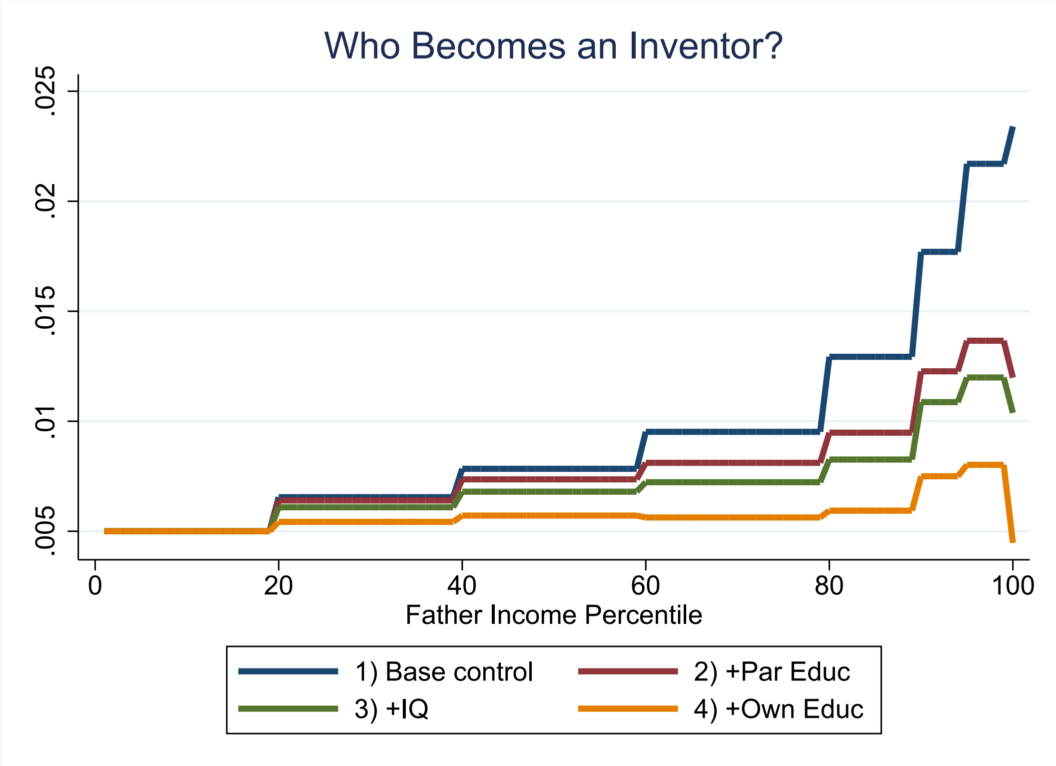


Figure 22:

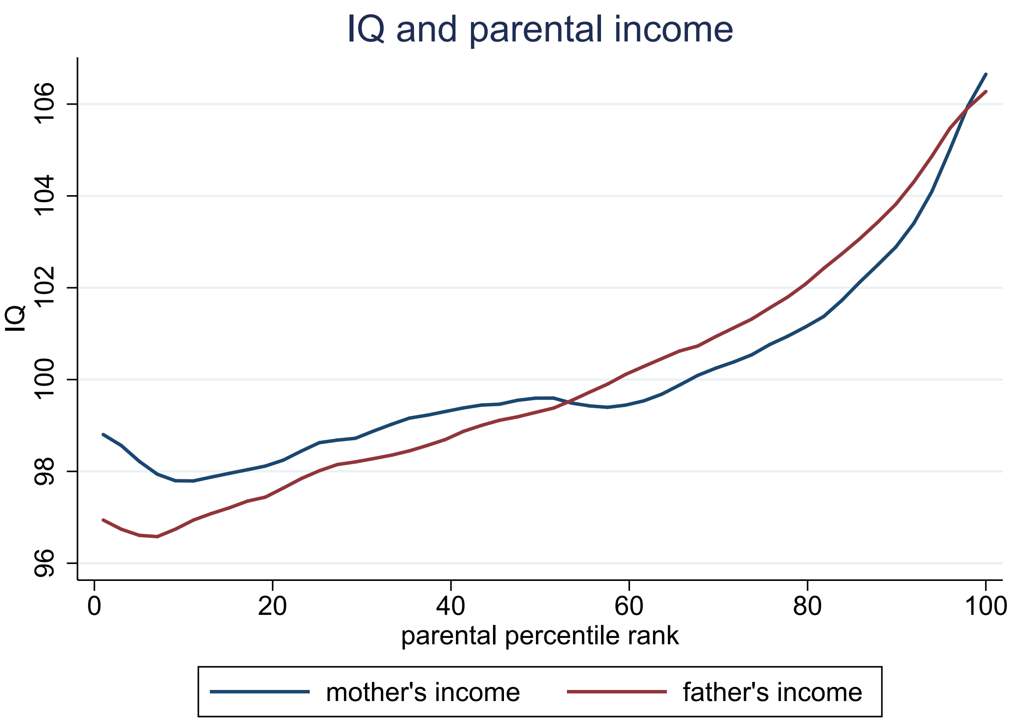


Figure 23: