Does the Cream Always Rise to the Top? The Misallocation of Talent in Innovation*

Murat Alp Celik†

November 1, 2021

Abstract

The misallocation of talent between routine production versus innovation activities is shown to have a first-order impact on the welfare and growth prospects of an economy. Surname level empirical analysis combining patent and inventor micro-data with census data reveals new stylized facts: (1) People from richer backgrounds are more likely to become inventors; but those from more educated backgrounds are not. (2) People from more educated backgrounds become more prolific inventors; but those from richer backgrounds exhibit no such aptitude. Motivated by this discrepancy, a heterogeneous agents model with financial frictions on the household side, and endogenous productivity growth on the firm side is developed. Individuals compete against each other for scarce inventor training in a tournament setting. Those from richer families can become inventors even if they are of mediocre talent by excessive spending on credentialing. This is individually rational but socially inefficient, and creates a negative link between wealth inequality and economic growth. The model is calibrated to match the new stylized facts via indirect inference. A thought experiment in which the credentialing spending channel is shut down (i.e., meritocratic allocation of talent) reveals that the rate of innovation can be increased by 10% of its value. Optimal progressive bequest taxes serve to increase social welfare by 6.20% in consumption-equivalent terms, increasing economic growth and decreasing consumption inequality simultaneously.

Keywords: Economic Growth, Inequality, Innovation, Misallocation, Patents, Optimal Taxation

JEL Classification: O15, O31, O41

*I would like to thank Ufuk Akcigit and Jeremy Greenwood; as well as Daron Acemoglu, Philippe Aghion, Harun Alp, Sina Ates, Salome Baslandze, Paul Beaudry, Gregory Clark, Harold Cole, Selman Erol, Nir Jaimovich, Charles Jones, Dirk Krueger, David Lagakos, Iourii Manovskii, Daniel Neuhann, Claudia Olivetti, Guillermo Ordonez, Serdar Ozkan, Jesse Perla, Michael Peters, Diego Restuccia, Jose-Victor Rios-Rull, Nicolas Roys, Peter Rupert, Felipe Saffie, Ananth Seshadri, Gustavo Ventura, Fabrizio Zilibotti, and the seminar and conference participants at Bilkent, FRB Philadelphia, Nottingham, Penn, FRB New York, UConn, Ohio State, CEMFI, Toulouse, IIES, Toronto, Rotman, McMaster, RIDGE, Queens, and Econometric Society Winter Meeting for helpful comments and discussions. I gratefully acknowledge financial support from SSHRC Insight Development Grant (503178)

†University of Toronto, Department of Economics, 150 St. George St., Toronto, ON, Canada, M5S 3G7. e-mail: murat.celik@utoronto.ca

1 Introduction

Albert Einstein was born in Ulm on March 14, 1879. His father was Hermann Einstein, a rich salesman and engineer, and owned a company called Elektrotechnische Fabrik J. Einstein & Cie that manufactured electrical equipment based on direct current. Albert received his education in various high quality schools in Germany, Italy, and Switzerland, and his alma mater was ETH Zurich. As a scientist and inventor, he produced over 300 scientific papers and 50 patented inventions. His groundbreaking contributions in the field of physics changed the technological landscape. What would happen, though, if his parents were poor and he could not receive the education he had? How would a world look like with Einstein as a factory worker instead of a scientist? Better yet, how do we know if we are not missing out on potential Einsteins right now?

Allocation of talent – assigning the right people to the right jobs – can have a first-order effect on the productivity of a society. The susceptibility of the allocation mechanism to be distorted away from the socially optimal outcome by private expenditures might create significant welfare losses in the presence of high levels of inequality in private resources. The losses are especially magnified if the best and the brightest of a society are not allocated to the professions where their social contributions would be the greatest. This paper aims to quantify the misallocation of talent in the United States due to economic inequality, with particular emphasis on its effects on innovation, and hence the long-run prospects of the country.

Parents spend considerable time and resources in order to improve the likelihood that their children end up with a desired job. The education system serves two main purposes in this regard: improving human capital, and credentialing people's talents.[1] The credentialing part can be seen as a tournament in which individuals seek to improve their overall ranking compared to others in order to improve their job market prospects. In 2010, the United States spent 7.3% of its gross domestic product on education, and the share of private spending was 7.7% for pre-college and 63.7% for college education (OECD, 2013). Annual expenditure per student in college education was $25,575, with total yearly cost going up to $60,000 for elite universities. At the same time, the net wage of the median worker was $26,364, whereas the median inventor earned above $100,000 per year. In such a high stakes environment where both the rewards and means to achieve them are unequal, financial frictions can easily prevent the talented children of poor families from being assigned to jobs suitable to their abilities since they are crowded out by the less talented children from richer families. Is this actually what happens in reality, or can we conclude that "the cream always rises to the top" regardless of inefficiencies of the system?

The first contribution of this paper is to provide empirical evidence on the misallocation of talent

[1] Work on education's role in improving human capital is discussed in detail in the related literature subsection. For the use of education in credentialing people's talents (as a "signaling" device), see Spence (1973), Stiglitz (1975) and Fernandez and Gali (1999).

in innovation. Information on innovation in the United States is obtained from inventor and patent level data from United States Patent and Trademark Office (USPTO). This data includes all patents granted in the United States between 1976 and 2006, as well as all registered inventors of these innovations. Inventors are identified uniquely throughout their careers, but direct information on their parental backgrounds is unavailable. In order to overcome this issue, surnames of the inventors are used as a proxy, and the inventor data is linked to socioeconomic background information at the surname level from U.S. census data (1930). The stylized facts obtained can be summarized as follows:[2]

Fact 1: Individuals from richer backgrounds are much more likely to become inventors (23.9%); whereas those from more educated backgrounds experience no similar advantage (0.1%).

Fact 2: Conditional on becoming an inventor, individuals from more educated backgrounds turn out to be much more prolific inventors (17.5%); whereas those from richer backgrounds exhibit no such aptitude (0.1%).

When the two facts are considered together, it appears that the misallocation of talent is an issue for inventors. Fact 2 shows that it is the education associated with the surname and not income that predicts higher inventor quality today. This is intuitive, since education and (unobserved) innate ability are likely to be complementary (or at least highly correlated), and in the presence of persistence of innate ability across generations, one would expect the descendants of the more educated to be better inventors today conditional on becoming one. However, Fact 1 shows that it is income and not education that predicts higher chances of becoming an inventor today. This can be interpreted as the allocation system choosing the wrong people as inventors. Those who come from families that were wealthier but had average education in the past have a higher chance of becoming inventors, but perform poorly conditional on becoming one. This observed discrepancy provides the motivation to investigate the issue of misallocation of talent in innovation quantitatively, so that its impact on the society can be assessed.[3]

In order to quantify the effects of the misallocation of talent in innovation and to analyze potential policy changes that might alleviate the inefficiency, a new model which can accommodate

[2]The numbers in parentheses correspond to how much one standard deviation increase in the independent variable causes the dependent variable to increase compared to its own standard deviation. The details of the empirical analysis can be found in Section 3.

[3]It is also noteworthy that the family background measures have such a high explanatory power. For instance, it is found that one standard deviation increase in the income associated with the surname in 1930 increases the relative probability of becoming an inventor by 23.9%. Given that the measures are constructed at the surname-level, and across two to three generations, just knowing the surname of an individual makes it possible to predict his or her chances of becoming an inventor to a high degree. This means the intergenerational mobility in socioeconomic status as captured by the relative probability of becoming an inventor is quite low, which is consistent with other studies that exploit the informational content of names and surnames [Clark (2014), Guell, Rodriguez-Mora, and Telmer (2015), Olivetti and Paserman (2015)].

the observed correlation patterns is developed. The firm side exhibits features found in the models from the endogenous growth literature: Firms undertake routine production using unskilled labor, and generate productivity-improving innovations (featuring positive intertemporal spillovers between firms) via research and development conducted by hired inventors. The household side is modeled in a detailed fashion, borrowing from heterogeneous agents models in order to make the model capable of replicating the patterns observed in the data.[4] The households are heterogeneous in wealth, education, and unobserved innate ability that is persistent across generations. Parents invest in the education of their offspring and leave bequests.[5] The training necessary to become inventors is scarce; hence individuals compete against each other in a tournament setting to receive it. Factors that improve inventor productivity such as innate ability and education increase the probability of receiving this training; but so does private credentialing spending which is unproductive by itself. Thus, individuals who inherit generous bequests can become inventors even if they are of mediocre talent through excessive spending on credentialing, preventing more talented individuals from poorer backgrounds from becoming one. This is individually rational but socially inefficient; reducing the quality of the inventor pool used in generating productivity-improving innovations that drive economic growth.

The tournament mechanism is the key ingredient that enables the model to replicate the stylized facts. In an ideal world, a social planner would prefer to allocate the best and the brightest of the society to the innovation sector, leading to a positive assortative matching between the talents of individuals and the (social as well as private) productivity of the jobs. However, if this were the case, the discrepancy between the parental backgrounds of those who become inventors and those who succeed as inventors would not be empirically observed. In order to allow the model to generate different correlation patterns at the two margins, individuals receive inventor training based on a score that depends differentially on innate ability, early childhood education and credentialing spending. The strength of each component in improving inventor probability as opposed to inventor productivity has different implications for the correlations of ancestor education and income with the two outcome variables, and this provides the main identification in the calibration of the model.

The model is calibrated to match the new stylized facts and data moments from the U.S. economy where an exercise in indirect inference pins down the influence of the new credentialing spending channel by replicating the two regressions from the empirical analysis using model-generated data. The calibrated model is then used to measure the economic importance of the misallocation of talent in innovation. A thought experiment in which the credentialing spending channel is shut down reveals that the aggregate growth rate of the economy can be increased by 10% of its value by

[4] This is in the spirit of Aiyagari (1994) since the heterogeneity of households is considered in a general equilibrium setting, where the time-invariant distributions of household characteristics affect the prices and the growth rate in the economy.

[5] Throughout the paper, the term bequest is used to refer to any transfer of resources to the descendants, including inter-vivos transfers.

assigning more talented and better educated individuals as inventors. As a result, the consumption inequality in the economy increases, which is detrimental to overall welfare; however the gain in output growth rate more than compensates for this loss, resulting in a welfare gain of 5.96% in consumption-equivalent terms.[6]

Seeking to alleviate the effects of misallocation in a decentralized economy, optimal progressive bequest taxes are calculated, which are found to increase output growth rate by 2.5% of its value. This increase is again through the allocation of higher innate ability individuals as inventors, who are also more educated on average. The progressive nature of the taxes causes the overall consumption inequality to remain the same. The increase in the output growth rate and the relatively unchanged consumption inequality lead to a social welfare gain of 6.20% in consumption-equivalent terms. This is higher compared to the credentialing spending shut-down experiment. The optimal bequest tax policy that achieves these results is quite progressive: The average bequest tax rate faced by the top 1% is 12.1%, whereas this number falls to 4.2% for the top 10%. The bottom 95% of the households are net recipients, whereas only the top 5% pay into the system.

Related literature: The paper relates to the growing literature on misallocation (some examples are Acemoglu, Akcigit, Alp, Bloom, and Kerr (2018), Akcigit, Celik, and Greenwood (2016), Guner, Ventura, and Xu (2008), Hsieh and Klenow (2009), Hsieh, Hurst, Jones, and Klenow (2019), Jones (2013), Jovanovic (2014), and Restuccia and Rogerson (2008)). One of the closest papers in this literature is Hsieh, Hurst, Jones, and Klenow (2019) where the misallocation of talent results from barriers to entry faced by distinct demographic groups based on gender and race for certain occupations. Guner, Parkhomenko, and Ventura (2018) show that selection and skill investments of managers in the presence of distortions can explain a large fraction of cross-country differences in output per worker. Another close paper is Jovanovic (2014) where workers and jobs are heterogeneous in quality, and are matched with each other under search frictions which affects the amount of on-the-job training, and the transition to the balanced growth path. This paper differs from these works by its emphasis on the financial frictions channel, and the focus on how innovation activities are influenced as a result. Empirically, there are three concurrent papers which reveal facts regarding inventors consistent with this study: Bell, Chetty, Jaravel, Petkova, and Van Reenen (2019) find out that individuals from higher income families are more likely to become inventors using U.S. social security data. Aghion, Akcigit, Hyytinen, and Toivanen (2017) use Finnish data to document the same, and show that controlling for the education of the inventor makes this correlation economically insignificant. Akcigit, Grigsby, and Nicholas (2017) show the positive correlation with parental income holds for historical inventor data from 1940 in the U.S.

Another closely related field is the modern literature on inequality and economic growth and development (see among others: Galor and Zeira (1988, 1994), Banerjee and Newman (1993), Maoz

[6]Welfare is defined as expected utility at the steady state.

and Moav (1999), and Galor (2009) for a literature review). This paper proposes a new mechanism through which wealth inequality can negatively influence long-term economic growth. It differs from the literature in that it acknowledges the scarce nature of training necessary to become an inventor, and focuses on how competition for this might create a misallocation of talent between routine production and innovation. Another difference is the source of economic growth. Unlike the previous literature which focuses on the accumulation of human capital, economic growth in the proposed model is driven by technological change as a result of firms investing in innovative activities, similar to the literature on endogenous growth with quality improvements pioneered by Aghion and Howitt (1992), and in the spirit of the broader endogenous growth literature [Lucas (1988), Romer (1990), Lucas (2009), Alvarez, Buera, and Lucas (2013), Lucas and Moll (2014). See Aghion and Howitt (2009), Acemoglu (2009) and Aghion, Akcigit, and Howitt (2014) for literature surveys]. The firm side of the model builds upon Akcigit, Celik, and Greenwood (2016). To my knowledge, this is the first paper to combine an innovation-based endogenous growth model with a general equilibrium heterogeneous agents model as in Aiyagari (1994). Recent work by Aghion, Akcigit, Bergeaud, Blundell, and Hemous (2019) investigates the relationship between innovation and top income inequality. The quantitative results of the current paper are in line with their empirical finding of a positive correlation between the two.

The focus on who become inventors versus who make prolific inventors conditional on becoming one links this work to the extensive literature on nature versus nurture, human capital and skill formation [Becker (1964), Ben-Porath (1967), Behrman, Taubman, and Wales (1977), Becker and Tomes (1979), Becker and Tomes (1986), Behrman, Rosenzweig, and Taubman (1994), Aiyagari, Greenwood, and Seshadri (2002), Heckman, Stixrud, and Urzua (2006), Cunha and Heckman (2007), Dahl and Lochner (2012), Lee and Seshadri (2019). See Cunha, Heckman, Lochner, and Masterov (2006) for a survey]. This literature is quite diverse, ranging from theoretical work such as the classic Becker and Tomes (1979) model, to empirical estimates exploiting rare datasets such as that on twins (Behrman, Rosenzweig, and Taubman, 1994) to separate the effects of nature and nurture. This paper investigates a related question, but focuses on inventors and their productivities in coming up with disruptive inventions as captured by patents.[7] This enables the use of the two new stylized facts obtained in the empirical analysis to tease out the persistence of innate ability versus the socioeconomic status persistence due to intergenerational wealth transmission. The model is close in spirit to Becker and Tomes (1979) type models, where parents cannot borrow against the future income of their dynasties, or insure themselves against idiosyncratic risks.

Finally, the policy experiment on optimal taxation of bequests links the paper to the literature on optimal taxation [Anderberg (2009), Bohacek and Kapicka (2008), Findeisen and Sachs (2016),

[7] While the focus of this paper is on inventors and patents, a misallocation of talent can occur in any high-paying profession. Applying the same empirical methodology to economists listed on IDEAS/RePEc, and top managers of publicly-listed US companies yields similar correlation patterns.

Grochulski and Piskorski (2010), Kapicka (2015), Kapicka and Neira (2019), Krueger and Ludwig (2013), Stantcheva (2017)]. Two close papers in this field are Krueger and Ludwig (2013) and Stantcheva (2017), where optimal progressive taxation and education subsidies are calculated in a model with heterogeneous households where human capital formation is also endogenous. The model in this paper also includes the endogenous human capital aspect, but enhances the problem by adding in the misallocation of talent dimension and its effects on innovation. This naturally affects the effectiveness of different policies in alleviating the inefficiencies that result from financial frictions.

Outline: The rest of the paper is organized as follows: Section 2 presents the theoretical model. Section 3 describes the datasets employed and variables constructed in the empirical analysis, and the resulting stylized facts. Section 4 describes the calibration of the model and the indirect inference. Section 5 presents and discusses the results of the quantitative results. Section 6 concludes.

2 Model

2.1 Environment and preferences

Time is discrete, and denoted by $t = 0, 1, 2, ...$ There is a continuum of households indexed by $m \in [0, 1]$. The households are modeled in an overlapping generations framework, where each generation lives for three periods: child, young adult and old adult. The children are born when their parents are young adults. The parents interact with their children in three ways: Parents (i) choose their consumption before they become adults, (ii) invest in their education,[8] and (iii) leave non-negative bequests to them upon death. The parents care about their children, and the relative weight of the utility of their offspring is denoted by the altruism parameter $\alpha > 0$. Preferences over consumption are time-separable with time discount factor β and exhibit constant relative risk aversion with parameter ω. Thus, lifetime utility of generation born at time t of household m can be expressed as

$$U_{m,t}(\vec{c}_{m,t}) = E_t\left[\frac{c_{c,m,t}^{1-\omega}}{1-\omega} + \beta\frac{c_{y,m,t}^{1-\omega}}{1-\omega} + \beta^2\frac{c_{o,m,t}^{1-\omega}}{1-\omega} + \alpha\beta U_{m,t+1}(\vec{c}_{m,t+1})\right] \tag{1}$$

where $c_{c,m,t}, c_{y,m,t}$ and $c_{o,m,t}$ denote the consumption of generation t of household m at child, young, and old periods respectively, and $\vec{c}_{m,t} = \{c_{c,m,T}, c_{y,m,T}, c_{o,m,T}\}_{T=t}^{\infty}$.

[8] The education investment of the parents is thought of as pre-college education or early childhood investment as discussed in Cunha and Heckman (2007). Individuals invest in their own college education when they become young adults, which is discussed later on.

2.2 Technology

2.2.1 Production and innovation

The final good is competitively produced by a continuum of firms indexed by $i \in [0, 1]$ which combine capital k and unskilled labor l_u according to the formula

$$o(z, k, l_u) = z^{\zeta} k^{\kappa} l_u^{\lambda} \tag{2}$$

where z stands for the firm-specific productivity, o denotes final good output, and $\zeta + \kappa + \lambda = 1$. Firms pay $r + \delta$ and w_u for capital and unskilled labor services respectively.

Firms can engage in risky innovation activities in order to increase their productivity if successful. Conditional on successful innovation, the productivity of the firm in the next period evolves according to

$$z' = z + \gamma \bar{z} \tag{3}$$

where z' and z are the new and old productivities, $\bar{z}$ is the average productivity in the economy, and $\gamma > 0$ is a scale parameter.[9] The firms which fail to innovate retain their old productivity, $z' = z$. In order to increase the probability of successful innovation, firms must hire skilled labor. For a firm which hires l_s amount of skilled labor, the probability of a successful innovation is given by

$$i(l_s) = \chi l_s^{\xi} \tag{4}$$

where $\chi > 0$ is a scale parameter and $\xi \in (0, 1)$ introduces diminishing returns.

2.2.2 Individual productivity, innate ability and early childhood education

Each generation t of each household m is heterogeneous in innate ability a, and early childhood education h. The individual productivity of generation t of household m is a constant elasticity of substitution (CES) aggregate of a and h given by

$$l_{m,t}(h_{m,t}, a_{m,t}) = \left(\psi h_{m,t}^{\frac{\epsilon-1}{\epsilon}} + (1 - \psi) a_{m,t}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{5}$$

where $0 < \psi < 1$ is the share of early childhood education, and ϵ is the elasticity of substitution. Innate ability a and early childhood education h remain constant as an individual gets older. Indi-

[9] Note that the $\bar{z}$ term in (3) introduces intertemporal spillover effects between the firms in the economy which is a salient feature of modern endogenous growth models. The additive structure is chosen over multiplicative because (i) it allows for solving the firm value functions in closed form and (ii) it ensures the existence of an invariant firm size distribution in a stationary equilibrium.

vidual productivity determines the effective labor supply of the individual. This labor contributes to the aggregate skilled or unskilled labor supply in the economy depending on the individual's job allocation.

The cost of endowing one's offspring with education level h in terms of the final good is given by the cost function

$$c_h(h, \Theta) = \kappa_h h^{\xi_h} \bar{z}^{\zeta/(\zeta+\lambda)} \tag{6}$$

where $\kappa_h > 0$ is a scale parameter, $\xi_h > 1$ introduces convexity, and $\bar{z}^{\zeta/(\zeta+\lambda)}$ ensures the cost scales up with aggregate output as the economy grows.

The innate ability of an individual is determined at the transition from childhood to young adult status, and depends on the innate ability of the parent. It is governed by a stochastic AR(1) process given by

$$\log a' = (1-\rho)\mu_a + \rho \log a + \epsilon_a, \;\; \epsilon_a \sim N(0, \sigma_a^2) \tag{7}$$

which has a mean of one. The variables a and a' denote the innate ability of the parent and the child respectively. The persistence parameter ρ determines how much of the parental ability the child inherits. The stochastic innate ability shock ϵ_a is normally distributed with a mean of zero and variance of σ_a^2.

2.2.3 Inventor training and job allocation

There are two types of jobs j in the economy: skilled/innovation jobs ($j = s$) and unskilled/production jobs ($j = u$). The job of an individual determines which pool his or her labor supply will contribute to, and hence the wage rate to be received per effective labor unit supplied (w_s if skilled and w_u otherwise). Any worker in the economy can get a production job. However, in order to get an innovation job, the individual needs to receive college education at a high-quality institution. This education provides the individual with (i) the training necessary to create innovations ("inventor training") and (ii) an increase in the individual productivity, given by $l' = \Lambda l$.[10]

[10] A good real world example of the described college education would be an MSc or PhD degree in a STEM field at a high quality institution, which itself usually requires having a prestigious BSc degree. NSF National Survey of College Graduates (2003) reveals that two thirds of inventors in the U.S. have a graduate degree, whereas one third are PhD holders. In addition, Aghion, Akcigit, Hyytinen, and Toivanen (2017) provide direct micro-evidence consistent with the described mechanism: Using data on inventors and their parents from Finland, the authors show that the probability of becoming an inventor is positively correlated with parental income and education when not controlling for the inventor's own education. However, once the latter is included, parental income and education become insignificant, and inventor's education captures virtually all of their predictive power. This is exactly in line with the predictions of the current model: Running the same regressions using model-simulated data produces the same pattern.

The ratio of inventor training available in the economy over total population is denoted by $\eta \in (0,1)$ and assumed to be fixed.[11] Since the innovation jobs pay better than production jobs in equilibrium, individuals would like to get innovation jobs.[12] Because of this, inventor training is sought after; and since the supply is fixed, there is competition among individuals to receive it, which is cleared by a score mechanism described below.

At the beginning of the young adult period and after observing the innate ability a, each individual receives a score given by

$$\tilde{s}(l(h,a),n) = (1-\nu)l(h,a) + \nu n + \epsilon_j,\ \epsilon_j \sim N(0,\sigma_j^2) \tag{8}$$

where $l(h,a)$ is individual productivity, n is credentialing spending (a choice variable), $\nu \in [0,1]$ is a parameter that governs the relative power of l versus n in determining the score, and ϵ_j is a normally distributed shock. After the scores for each individual are realized, the fraction η of the individuals with the highest scores receive inventor training, and are able to work in the innovation sector. The remaining $(1-\eta)$ fraction of the individuals do not receive inventor training and cannot create innovations, and thus have to work in the production sector.

In order to increase score upwards by the amount νn, the individual has to spend resources given by

$$c_n(n) = \kappa_n n^{\xi_n} \bar{z}^{\zeta/(\zeta+\lambda)} \tag{9}$$

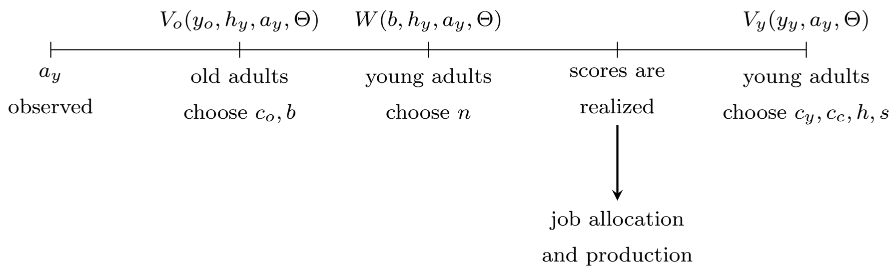
in terms of the final good, where $\kappa_n > 0$ is a scale parameter, $\xi_n > 1$ introduces convexity, and $\bar{z}^{\zeta/(\zeta+\lambda)}$ ensures the costs scale up with aggregate output as the economy grows. This choice variable n captures any real world spending that increases the chances of getting inventor training, such as hiring private tutors and private college counseling, re-taking standardized tests, covering tuition and living expenses out-of-pocket instead of relying on a scholarship, the opportunity cost of studying as opposed to joining the workforce, as well as high-cost extracurricular activities such as founding or leading one's own non-governmental organization.[13]

[11] An alternative assumption would be having no restrictions on η, but fixing the score threshold $\bar{s}$ instead, so that any individual who has a sufficiently high score would get the inventor training. The quantitative experiments replicated with this alternative model deliver higher growth rate and welfare responses to parameter and policy changes; so fixing η is the conservative assumption. See Appendix C.1 for details.

[12] This is not restriction of the model, but a result of the calibration exercise. See Section 4 for the details.

[13] Parents spend considerable time and resources to improve the likelihood that their children obtain the best credentials possible. The most direct way they can help is to pay the tuition and living expenses for prestigious colleges, which can go up to 3-4 times the net wage of the median worker. But much more extravagant options exist as well. In a 2018 lawsuit, New York-based private college counseling firm Ivy Coach was revealed to charge a client \$1.5 million to help their child with college and boarding school applications. Other families are known to found charitable non-governmental organizations with their children installed as the manager in an effort to give a boost to the extracurricular activities part of their application.

FIGURE 1: TIMING OF EVENTS WITHIN A PERIOD.



Since the upper η fraction of the score distribution receives inventor training, there exists a score threshold $\bar{s}$ such that individuals with $\tilde{s} \geq \bar{s}$ receive the education, and the rest do not. In equilibrium, individuals with the necessary education always choose the innovation sector over the production sector, so the probability of getting inventor training and that of being a skilled worker are the same. The implied probability distribution of having job j for an individual is denoted by $F(j; l(h, a), n, \Theta)$. The aggregate state matters, since the score of a worker is only meaningful compared to the score threshold $\bar{s}$, as relative rank determines job allocation. The probability of having a skilled job is increasing in innate ability a, early childhood education h and credentialing spending n, whereas it is decreasing in the score threshold $\bar{s}$.

2.3 Maximization problems

2.3.1 Timing of events

Before moving on to the decision problems of the firms and the households, the timing of events within a period are listed below, which are also summarized in Figure 1:

1. The innate ability of young adults a_y is observed.

2. Old adults choose their bequests b and consumption c_o.

3. Young adults decide on credentialing spending n to receive a better score $\tilde{s}$.

4. Scores $\tilde{s}$ are observed, inventor training is provided, and young adults are assigned to their jobs j.

5. Firms hire capital k and labor l_u and l_s for production and innovation. Production takes place and successful innovations are realized. Wages are paid.

6. Young adults choose how much to consume c_y, consumption of their children c_c, pre-college education investment in their children h, and savings s.

2.3.2 Firm decision problems

The static profit maximization problem of the firm is given by

$$\Pi(z,\Theta) = \max_{k,l_u \geq 0} \{z^{\zeta} k^{\kappa} l_u^{\lambda} - (r+\delta)k - w_u l_u\} \tag{10}$$

where the firm pays interest rate plus depreciation $(r+\delta)$ and unskilled wage w_u for capital and unskilled labor services respectively. The associated capital and labor demand policy functions are denoted by $\hat{k}(z,\Theta)$ and $\hat{l_u}(z,\Theta)$.

Given the period profits $\Pi(z,\Theta)$ from the static maximization problem and the innovation technology described in (4), the intertemporal maximization problem of a firm can be written in recursive form as follows:

$$V(z,\Theta) = \max_{l_s \geq 0} \left\{ \Pi(z,\Theta) + \frac{\chi l_s^{\xi}}{1+r} V(z+\gamma\bar{z},\Theta') + \frac{(1-\chi l_s^{\xi})}{1+r} V(z,\Theta') - w_s l_s \right\} \tag{11}$$

The firm chooses how much skilled labor l_s to hire, which increases the likelihood of successful innovation χl_s^{ξ}. If successful, the firm's productivity next period is increased by $\gamma\bar{z}$. The prospect of earning higher profits in the future due to higher productivity provide incentives for the firm to engage in costly innovation. The skilled labor demand that solves this problem is denoted by $\hat{l}_s(z,\Theta)$.

2.3.3 Household decision problems

Given the ingredients of the model, there are three relevant decision problems for each household in any given period: (i) the bequest decision of old adults, (ii) the credentialing spending decision of young adults before job allocation, (iii) the consumption, pre-college education investment and saving decisions of young adults after job allocation.[14] The associated value functions of the problems will be denoted by $V_o(.)$, $W(.)$ and $V_y(.)$ respectively.

2.3.4 Decision problem of the old

Let subscripts c, y and o stand for child, young and old respectively. Time subscripts will be suppressed for clarity. Let y denote wealth. Given the wealth of the old y_o, the early childhood education h_y and innate ability a_y of the young, and the aggregate state of the economy Θ, the

[14] Children in a household have no decision problems to solve. They receive pre-college education chosen by their parents and consume.

bequest decision problem of the old can be stated as

$$\begin{aligned} V_o(y_o, h_y, a_y, \Theta) &= \max_{c_o, b \geq 0} \{u(c_0) + \alpha W(b, h_y, a_y, \Theta)\} \text{ s.t.} \\ c_o + b &\leq y_0 \end{aligned} \tag{12}$$

where c_o is the consumption of the old, b is the bequest left to the descendants and $\alpha > 0$ is the altruism parameter. Old agents choose how much bequests b to leave to their children who are now young adults, at the cost of reducing the their own consumption c_0. The problem is solved by the choice of a single variable b since preferences ensure the budget constraint holds with equality. Note the financial restriction that the bequests must be positive. This disallows agents from borrowing against the future income of their dynasty to consume today, a seemingly mild, but important, financial friction. The associated policy function is denoted by $\hat{b}(y_o, h_y, a_y, \Theta)$.

2.3.5 Decision problem of the young before job allocation

Given the bequest amount b, the early childhood education h_y and innate ability a_y of the young, and the aggregate state of the economy Θ, the credentialing spending decision problem of the young before job allocation can be stated as follows:

$$\begin{aligned} W(b, h_y, a_y, \Theta) &= \max_{n \geq 0} \{E[V_y(y_y, a_y, \Theta)|\cdot]\} \text{ s.t.} \\ y_y &= \left(w_{j_y} + \frac{w'_{j_y}}{1+r'}\right) l_y(h_y, a_y) + b - c_n(n) \\ j_y &\sim F(j; l_y(h_y, a_y), n, \Theta) \end{aligned} \tag{13}$$

where j_y is a random variable that denotes job allocation and y_y stands for wealth as a young adult after job allocation. The wealth of the young y_y consists of the lifetime labor income and the bequests b received from parents, minus the cost of improving the score $c_n(n)$. The only choice variable is the resources spent on improving the score, denoted by n. Spending more resources increases the likelihood of getting a better job draw j_y distributed according to $F(j; l, n, \Theta)$ discussed earlier. The optimal n that solves this optimization problem is referred to as the credentialing spending policy function, $\hat{n}(b, h_y, a_y, \Theta)$.

Note that the young adults can borrow against their future lifetime labor income, so the model allows agents to borrow resources at the risk free interest rate r' to spend on credentialing which improves their chances of getting inventor training. On the other hand, they cannot insure themselves against the idiosyncratic risk of not getting inventor training, which is always positive due to the shock term ϵ_j in (8). This forces them to be more prudent in increasing credentialing spending n by borrowing due to risk aversion.

2.3.6 Decision problem of the young after job allocation

Given the wealth y_y and the innate ability a_y of the young, and the aggregate state of the economy Θ, the consumption, early childhood education investment and saving decision problem of the young after job allocation can be stated as follows:

$$\begin{aligned} V_y(y_y, a_y, \Theta) &= \max_{c_y, c_c, h'_y, s \geq 0} \{u(c_y) + \alpha u(c_c) + \\ &\quad \beta E[V_o(y'_o, h'_y, a'_y, \Theta')|\cdot]\} \text{ s.t.} \qquad (14) \\ y_y &\geq c_y + c_c + c_h(h'_y) + s \\ y'_o &= (1 + r')s \\ a'_y &\sim g(a_y) \\ \Theta' &= T(\Theta) \end{aligned}$$

Variables with primes indicate next period's values. The choice variables are the consumption of the young and their children, c_y and c_c, the early childhood education investment in the children h'_y which costs $c_h(h'_y)$ in terms of the final good, and the savings s. The sum of these expenditures must be below the wealth y_y. The expectation is over the innate ability a'_y of the child tomorrow, which depends on the innate ability of the parent a_y. The aggregate state of the economy evolves according to the transition function $T(.)$. The policy functions that solve this problem are given by $\hat{c}_y(y_y, a_y, \Theta), \hat{c}_c(y_y, a_y, \Theta), \hat{h}'_y(y_y, a_y, \Theta)$ and $\hat{s}(y_y, a_y, \Theta)$.

2.4 Balanced growth path equilibrium

Let $Z(z)$ denote the distribution of firm productivities in the economy. Labor market clearing implies

$$L_{u,t} \equiv \int_0^1 \hat{l}_{u,t}(z, \Theta) dZ(z) = 2(1-\eta) \int l(h, a) d\Phi_{u,t}(h, a), \text{ and} \qquad (15)$$

$$L_{s,t} \equiv \int_0^1 \hat{l}_{s,t}(z, \Theta) dZ(z) = 2\eta \int l(h, a) d\Phi_{s,t}(h, a) \qquad (16)$$

where $\Phi_{u,t}(h, a)$ and $\Phi_{s,t}(h, a)$ denote the joint distribution of early childhood education and innate ability at time t for unskilled and skilled workers respectively. The $(1-\eta)$ and η terms in the labor supply expressions are multiplied by average individual productivity because they designate the fraction of the population working in production and innovation sectors respectively. The terms are also multiplied by two since in any period both the young and old adults work. Aggregate savings in the economy is given by

$$A_{t+1} \equiv \int \tilde{a}_{m,t-1} d\tilde{A}(\tilde{a}) \qquad (17)$$

where $\tilde{a}_{m,t} \equiv s_{m,t} - l(h_{m,t}, a_{m,t}) w_{j_{m,t},t+2}/(1+r_{t+2})$ denotes the net savings of the young adults of household m born at time t.[15] There are two kinds of assets in the economy: physical capital and shares in the bundle of firms $i \in [0,1]$. Both assets pay the risk-free interest rate r_t.[16] The capital market clearing requires the physical capital supply in the economy to equal the aggregate capital demand of the firms given by

$$K_t \equiv \int_0^1 \hat{k}_{u,t}(z,\Theta) dZ(z). \tag{18}$$

Final good market clearing requires

$$O_t = C_t + K_{t+1} - (1-\delta)K_t + N_t + H_t \tag{19}$$

where O_t denotes aggregate output and C_t, N_t and H_t are aggregate spending on consumption, credentialing, and early childhood education investment at time t respectively. Finally, the number of people who receive inventor training must equal the exogenous restriction on their measure η. This imposes the condition

$$\eta = \int_{\bar{s}_t}^{\infty} \tilde{s} d\tilde{S}_t(\tilde{s}) \tag{20}$$

where $\tilde{S}_t(\tilde{s})$ is the score distribution at time t and $\bar{s}_t$ is the score cut-off above which agents get inventor training.

Given these ingredients, an equilibrium of this economy is defined as follows:

Definition 1 *An equilibrium is described by allocations* $[\{\vec{c}_{m,t}, b_{m,t}, n_{m,t}, h_{y,m,t}, s_{m,t}\}_{t=0}^{\infty}]_{m=0}^{1}$ *for households, allocations* $[\{z_{i,t}, k_{i,t}, l_{u,i,t}, l_{s,i,t}\}_{t=0}^{\infty}]_{i=0}^{1}$ *for firms, prices* $\{r_t, w_{u,t}, w_{s,t}\}_{t=0}^{\infty}$*, score cut-off* $\{\bar{s}_t\}_{t=0}^{\infty}$*, firm productivity distribution* $\{Z_t(z)\}_{t=0}^{\infty}$*, and joint distribution of jobs, early childhood education, and innate ability* $\{\Phi_t(j,h,a)\}_{t=0}^{\infty}$ *such that:*[17]

1. *Given prices and score cut-off, household allocations maximize* $V_o(b, h_y, a_y, \Theta)$, $V_y(y_y, a_y, \Theta)$*, and* $W(b, h_y, a_y, \Theta)$.

2. *Given prices and the productivity distribution, firm allocations maximize* $\Pi(z,\Theta)$ *and* $V(z,\Theta)$.

3. *All markets clear.*

[15] In order to calculate $\tilde{a}_{m,t}$, the labor income to be earned in the old adult stage is subtracted from $s_{m,t}$ because it was included in the expression y_y in the young agent's recursive decision problem. This was done to reduce the number of state variables to keep track of in the associated value function $V_y(.)$.

[16] Although each firm $i \in [0,1]$ faces idiosyncratic risk, aggregating over i makes profits received from the whole bundle a deterministic quantity due to the lack of aggregate fluctuations.

[17] Arguments of the allocations are suppressed for clarity.

Output growth in this economy is driven by improvements in the productivities of the firms given by the distribution $Z_t(z)$. This paper focuses on the balanced growth path equilibrium where aggregate variables O_t, K_t, N_t, H_t, and C_t grow at the constant rate g. Along the balanced growth path, it turns out that the mean of the firm productivity distribution, $\bar{z} \equiv \int z dZ(z)$, is a sufficient statistic to determine the growth rate of the economy. Let the growth rate of the mean productivity $\bar{z}$ be denoted by g_z. Define transformed variables $\hat{z} \equiv z/\bar{z}^{\lambda/(\lambda+\zeta)}$, $\tilde{z} \equiv \bar{z}^{\zeta/(\lambda+\zeta)}$ and $\tilde{w}_s \equiv w_s/\tilde{z}$. The balanced growth path equilibrium of this economy is described below.

Theorem 1 *The balanced growth path equilibrium of the economy has the following form:*

1. *Aggregate allocations O_t, K_t, N_t, H_t, and C_t, and wages $w_{u,t}$ and $w_{s,t}$ grow at constant rate g.*
2. *Aggregate labor allocations L_u and L_s, interest rate r, score cut-off $\bar{s}$, and joint distribution of jobs, early childhood education, and innate ability $\Phi(j, h, a)$ are time-invariant.*
3. *Mean of the firm productivity distribution $\bar{z}$ grows at constant rate g_z, where $1 + g = (1 + g_z)^{\zeta/(\lambda+\zeta)}$.*
4. *Period profits of the firm are linear in $\hat{z}$, given by $\Pi(z, \Theta) = \pi\hat{z}$.*
5. *The value function of the firm is linear in $\hat{z}$ and $\tilde{z}$, given by $V(z, \Theta) = v_1\hat{z} + v_2\tilde{z}$.*
6. *The constants v_1, v_2, π, prices $r, w_{u,t}, w_{s,t}$, growth rate g_z, and aggregate production factors K_t, L_u and L_s are jointly determined by a system of nonlinear equations given by (25), (26), (27), (29), (30), (31), and (32), and the market clearing conditions.*

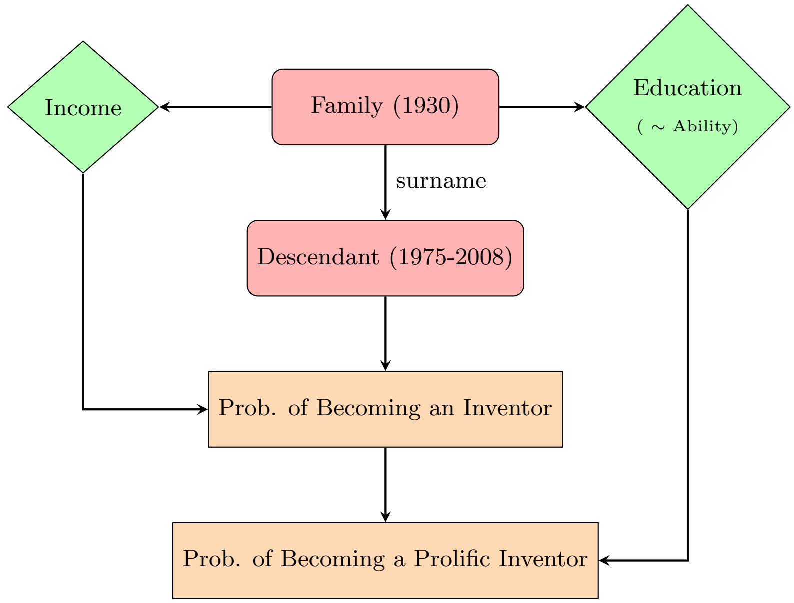
Proof. See Appendix A ■

3 Empirical Analysis

3.1 Overview

In order to assess whether there is any indication of a misallocation of talent in innovation, several different data sources are combined. Figure 2 presents a simple schema of the baseline empirical analysis. The information on the probability of becoming an inventor, and how well one performs conditional on becoming one is obtained from various datasets that cover the years 1976-2008. The information on the family backgrounds comes from the IPUMS-USA 5% sample of the U.S. census conducted in 1930. In order to link the recent patent and inventor micro-data to the older census data, surname information is used. Once the links between the families and the descendants are established at the surname level, the probability of becoming an inventor and the productivity as an inventor conditional on becoming one are regressed on family income and education. It is revealed that it is income and not education that predicts a positive probability of becoming an

Figure 2: Overview of the empirical analysis



inventor, whereas it is education and not income that predicts the probability of becoming a prolific inventor. This inconsistency between the extensive and the intensive margins is the main focus of the empirical analysis. Following sections discuss the data sources in detail, describe the variables created, present and discuss the baseline empirical results, and conclude with some robustness checks.

3.2 Data Sources

3.2.1 NBER USPTO Utility Patents Grant Data

Patents are exclusionary rights, granted by national patent offices, to protect a patent holder for a certain amount of time, conditional on sharing the details of the invention. United States Patent and Trademarks Office (USPTO) is the agency in the U.S. Department of Commerce that issues patents to inventors and businesses for their inventions. From the great amount of information available in the files of USPTO, a substantial subsample has been compiled in an easy-to-use format by a group of researchers from the National Bureau of Economic Research (NBER) under the name NBER Patent Data Project (PDP).[18]

This dataset contains detailed information on 3,210,361 utility patents granted by USPTO between the years 1976 and 2006. Each patent granted in the U.S. is assigned a unique patent number that makes it possible to link this dataset to many other datasets that contain information on patents some of which will be described further along. An important feature of this dataset is

[18] For more information, please visit https://sites.google.com/site/patentdataproject/

to provide citation links between individual patents. Similar to an academic paper, a new patent has to cite previous patents on which it builds, or other patents concerned with a similar but different invention, so that proper boundaries between the new and old patents can be established. The number of citations a patent receives from other patents is found in the literature to be a good proxy for its social and private value.[19] Since the citations a patent will receive throughout its lifetime cannot be known at a fixed point in time, and due to systematic citation differences between patents that belong to different technology classes, corrections need to be made to the citation numbers before using them as a proxy for patent quality. Hall, Jaffe, and Trajtenberg (2001) devise some correction weights to account for these biases, and their correction is used throughout this paper unless mentioned otherwise.

3.2.2 The Careers and Co-Authorship Networks of U.S. Patent-Holders

Filing a patent application in the U.S. requires providing the names of three types of individuals in the application form: The assignees who own the patent once granted; the applicants who are responsible for legal correspondence with USPTO; and the inventors who actually came up with the innovation.[20] Extensive information on the inventors of patents granted in the U.S. between years 1975-2008 is obtained from a dataset produced by Li, Lai, D'Amour, Doolin, Sun, Torvik, Amy, and Fleming (2014).[21] Unlike the PDP data, this dataset contains the names of every inventor who has worked on a patent granted in the U.S. between years 1975-2008. This is crucial, since 55.3% of the patents in the sample were created by a group of inventors. Another novel feature of this data is the provision of a unique inventor identifier which makes it possible to track the patent portfolio of individual inventors throughout their careers.

The dataset contains 8,031,908 observations at the patent $\times$ inventor level, and 2,229,219 unique inventors. Among other variables, the dataset contains address information of the inventors as well as their names and surnames. The address information is used to determine the country the inventor lives in at the registration date of the patent so that the analysis can be restricted to U.S. inventors only. The surname information is used to create a relative representation (among inventors) measure at the surname level and link the socioeconomic background data from 1930 to inventors today. Both of these will be discussed in detail.

[19] For instance, Hall, Jaffe, and Trajtenberg (2005) argue that the citation-weighted patent portfolio of a firm is a plausible indicator of the intangible knowledge stock of a private firm, and that this measure has additional explanatory power for the market value of the firm beyond the conventional discounted sum of R&D spending.

[20] Hence, the owner of a patent or the manager in an innovating firm are not listed as inventors unless they took part in the innovation process. USPTO explicitly states the following: "All inventor(s) named in the provisional application must have made a contribution, either jointly or individually, to the invention disclosed in the application."

[21] Please visit `http://hdl.handle.net/1902.1/12367` to access the data.

3.2.3 IPUMS-USA 1930 5% Sample

Integrated Public Use Microdata Series (IPUMS-USA) is a project dedicated to collecting and distributing United States census data, and it consists of more than fifty high-precision samples of the American population drawn from federal censuses. The particular sample used in this project is the 1930 sample which contains information on 5% of all Americans who were counted in the 1930 census. The 1930 sample is preferred over other samples since it is the most recent publicly available sample that contains name and surname information at the observation level.[22]

Since the dataset contains census information, the wealth of information at the individual level is immense. The main information derived from this dataset is on socioeconomic status of people with a particular surname, such as income and education collapsed at the surname level. Similar to other studies that use the IPUMS samples prior to 1940 (Olivetti and Paserman, 2015), income associated with a surname is measured using the OCCSCORE variable measured in hundreds of 1950 U.S. dollars. This variable includes income from non-wage activities such as interest income and dividends in addition to earnings. Finally, EDSCOR50 variable is used as the education variable which measures college attendance. Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek (2010) and the project website contain further details on the dataset and variables.[23]

3.2.4 Demographic Aspects of Surnames from Census 2000

This dataset released by the U.S. Census Bureau in 2007 contains information on the overall frequency of surnames in the U.S. constructed using the 2000 decennial census of population, based on approximately 270 million individuals with valid surnames.[24] It contains 151,671 unique surnames. Combined with the U.S. inventor data previously discussed, it is possible to create measures of probability of becoming an inventor at the surname level. This dataset further includes information on the ethnicity distribution for each surname broken down into six categories (White, Black, Hispanic, Asian or Pacific Islander, American Indian or Alaskan Native, or mixed). These variables are used to create dominant race fixed effects for race associated with a surname. One caveat of the data is that it only includes surnames that have a frequency above hundred, which makes it unsuitable to use in questions regarding extremely rare surnames. Such rare surnames are therefore excluded from the following analysis.

[22] Individual questionnaires of any specific census are not released by the National Archives until 72 years after that specific census has been taken due to confidentiality requirements. Name and surname information is also available for other samples spanning the years 1850-1920 in the IPUMS database; however they are less recent, and most of these samples are at 1% level instead of 5%.

[23] IPUMS-USA project website can be accessed at `https://usa.ipums.org/usa/index.shtml`.

[24] Refer to Word, Coleman, Nunziata, and Kominski (2008) for a detailed description, and `http://www.census.gov/genealogy/www/data/2000surnames/index.html` for the data.

3.3 Data construction and variables

The summary statistics for the variables used in the empirical analysis are presented on Table 1. The following subsections describe what they stand for and how they are generated.

Table 1: Summary Statistics

	Observation	*Mean*	*St. Dev*
Panel A. Extensive Margin Analysis			
relative representation (1975-2008)	110,290	83.1	68.6
income (1930)	110,290	21.7	3.70
education (1930)	110,290	15.8	17.5
is black	110,290	2.28	14.9
is asian	110,290	1.52	12.2
is native	110,290	.048	2.20
is hispanic	110,290	11.4	31.8
is mixed	110,290	.002	.438
Panel B. Intensive Margin Analysis			
log quality weighted total patents (1975-2006)	81,348	3.65	.664
log average patent quality (1975-2006)	81,348	2.53	.387
log maximum patent quality (1975-2006)	81,348	2.95	.455
log total patents (renewed thrice) (1975-2006)	78,438	.695	.287
log total patents (top 10% only) (1975-2006)	81,348	.321	.200
log income (1930)	81,348	3.11	.158
log education (1930)	81,348	2.69	.494
is black	81,348	2.05	14.2
is asian	81,348	1.55	12.3
is native	81,348	.035	1.87
is hispanic	81,348	11.7	32.1
is mixed	81,348	.001	.38

Notes: Relative representation and dominant race indicator variables are multiplied by 100 for clarity. The means and standard deviations reported on the table are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population of 2000. Patent quality is measured by the number of patent citations corrected for truncation using the correction terms from Hall, Jaffe, and Trajtenberg (2001).

3.3.1 Surname level socioeconomic status variables (1930)

Socioeconomic status variables such as income, earnings, and education are constructed at the surname level by taking the averages of observations in the IPUMS-USA 1930 5% sample. In this process, observations without a valid occupation are not included.[25]

[25] These observations correspond to those with OCC1950 values between 980 and 999. Visit `https://usa.ipums.org/usa-action/variables/OCC1950#codes_section` for a complete list of OCC1950 values.

3.3.2 Relative representation of a surname among inventors (1975-2008)

The extensive margin analysis focuses on the question of how the socioeconomic background of an individual affects the probability of becoming an inventor – or using the model's terminology, the probability of being assigned to a job in the innovation sector. The Careers and Co-Authorship Networks of U.S. Patent-Holders data contains the names of all inventors who worked on patents granted in the U.S. between the years 1975 and 2008, from which it is possible to obtain the number of inventors with a particular surname. However, the fact that there are many inventors with the surname Smith does not mean that Smiths are more likely to become inventors by itself. In order to create a measure of the probability, the number of inventors with a particular surname is divided by the number of all people in the U.S. with the same surname obtained from Demographic Aspects of Surnames from Census 2000; i.e.

$$\text{inventor probability (surname)} = \frac{\text{number of inventors (surname)}}{\text{number of individuals (surname)}}$$

Relative representation of a surname among the inventor sample is then built simply by dividing the inventor probability associated with the surname with the unconditional probability of becoming an inventor in the U.S. given by

$$\text{relative representation (surname)} = \frac{\text{inventor probability (surname)}}{\text{unconditional inventor probability}}$$

Thus a relative representation score above unity means that individuals with that surname are more likely to become inventors than the average person, and vice versa.

3.3.3 Patent and inventor quality metrics (1975-2006)

The intensive margin analysis considers the question of how the socioeconomic background of an individual affects the productivity as an inventor conditional on becoming one. In order to conduct this analysis, it is necessary to come up with metrics that measure inventor productivity. The unique inventor variable allows tracking the patent portfolio of each inventor between the years 1975 and 2008. The productivity of an inventor can be calculated as a function of the information on all the patents he or she has worked on. This naturally leads to the question of how to assess the value of a patent. In line with the literature, the quality of a patent is proxied by the citations received by the patent, corrected for truncation bias and other concerns using the weights devised by Hall, Jaffe, and Trajtenberg (2001). The patent quality information from the PDP data is linked to the inventor data using the unique patent numbers granted by USPTO. The inventor quality metric that is used in the baseline analysis is the total quality weighted patents of an inventor throughout his or her career. Several additional alternative metrics are considered in the robustness analysis in Section 3.4.3.

Since the data contains all inventors who worked on patents registered in the U.S., it is necessary to separate the foreign inventors from the sample used to create surname level variables. The address information of an inventor is available for every patent, and there is considerable variation between the countries. For this study, only inventors who have stayed in the U.S. throughout their whole career are kept. Average inventor quality metrics at the surname level are constructed by taking the unweighted average of individual inventor qualities.

3.4 Empirical results

3.4.1 Extensive margin analysis

Table 2: Extensive Margin Baseline – Probability of Becoming an Inventor

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.239***		.239***
	(.010)		(.010)
education (1930)		.029***	.001
		(.006)	(.005)
Obs.	110,290	110,290	110,290
R^2	0.27	0.23	0.27

Notes: Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

In order to understand whether there is a misallocation of talent in the innovation sector or not, it is necessary to empirically demonstrate what is correlated with the probability of having a job in this sector. The surname level probability of being an inventor is used as a proxy to gauge this, although inventors are not the only individuals who work in the innovation sector. Socioeconomic background information at the surname level obtained from IPUMS-USA 1930 dataset is connected to these probabilities using surnames. Population-weighted ordinary least squares estimation is used where the relative representation rate is regressed on the socioeconomic variables: income and education.[26] The three columns of Table 2 correspond to regressions on income, education, and both variables at the same time respectively.

[26] Since all variables are averages calculated at the surname level, using a weighted ordinary least squares estimator is essentially identical to a two-sample two-stage least squares estimator. Alternatively, if an observation was created for each individual in the Census 2000 sample with an indicator variable for being an inventor or not, the unweighted OLS regression of this indicator variable on background information linked by surnames would yield the exact same coefficients as Table 2.

Looking at the first two columns, it is observed that both income and education associated with a surname in 1930 are positively correlated with the relative representation among inventors today (1975-2008) and statistically significant. A standard deviation increase in income increases the relative representation rate by 23.9% compared to its standard deviation, while a standard deviation increase in education increases it by 2.90%. Given that there are roughly three generations between 1930 and today, these numbers are quite substantial ($\sqrt[3]{23.9\%} = 62.1\%$), and hint towards low intergenerational mobility in social status, similar to the results in other studies that use surnames (Clark (2014), Olivetti and Paserman (2015)).

Looking at the third column tells another striking story: It is income and not education that is strongly correlated with the over-representation among inventors. In other words, people with surnames that were richer in the past are more likely to become inventors today; but controlling for income, education has no further prediction power.[27] This finding is the motivation behind the inclusion of the credentialing spending in the model, which enables agents to increase the probability of getting the inventor training necessary for innovation sector jobs by spending private resources.

3.4.2 Intensive margin analysis

Table 3: Intensive Margin Baseline – Inventor Productivity Conditional on Becoming One

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	.066***		.001
	(.009)		(.009)
log education (1930)		.176***	.175***
		(.008)	(.009)
Obs.	81,348	81,348	81,348
R^2	0.03	0.05	0.05

Notes: Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

[27] The insignificance of education is not driven by multicollinearity due to high correlation between the variables. In order to address such concerns, a variance inflation factor test is conducted after each regression in the paper in which both income and education are included as regressors. None of the tests result in a VIF large enough to be concerned about (< 3). The results are available upon request.

Having discovered that income associated with the surname is significantly positively correlated with the probability of being an inventor, the natural next step is to ask whether these individuals are the individuals who would make the best inventors. In order to investigate this question, the inventor quality metric described earlier is regressed on income and education. Table 3 displays the results of three OLS regressions: log inventor quality on log income, on log education, and on both variables at the same time.

Due to the log-log specification, the coefficients can be interpreted as elasticities. By themselves, both income and education turn out to be positively correlated with inventor quality, and the associated coefficients are statistically significant. Once again, given that there are three generations between the samples, the elasticity estimates are considerably high. However, this time the standard error for education is much smaller than that for income, the opposite of what was observed in the extensive margin analysis.

The last column regresses inventor quality on both income and education, and the results are striking. The elasticity of inventor quality with respect to education is very close to that on column 2, but the elasticity with respect to income vanishes, and is statistically insignificant. Conditional on becoming an inventor, it is the inventors with "more educated" surnames who are the most successful in creating new path-breaking innovations. This is in direct contrast to the extensive margin results, and suggests that the individuals who would make the best inventors might not be the same as those the society allocates as inventors. This fact is captured in the model by three ingredients: (i) education increases individual productivity in the innovation sector, (ii) education and innate ability are complementary in determining individual productivity, (iii) credentialing spending increases the probability of getting in an innovation sector job, but it does not increase individual productivity compared to other inventors (as opposed to education, which does both).

3.4.3 Alternative inventor quality measures

In the baseline intensive margin analysis, quality weighted total patents of an inventor was used as the inventor quality metric, where patent quality was measured by the citations a patent receives. This section establishes that the results are robust to using different measures of inventor quality. Results pertaining to additional alternative measures can be found on Table 12 in the empirical appendix.

Table 4 replicates the regression in column 3 of Table 3 using different inventor quality metrics.[28] The first two columns preserve the same patent quality metric (citations), but consider the average and maximum patent quality for inventors respectively. Compared to the baseline measure, the

[28] Results for replicating columns 1 and 2 are also very similar, but excluded to conserve space. These are available upon request from the author.

Table 4: Intensive Margin Robustness - Alternative Measures

	log avg. patent quality (1975-2006)	log max. patent quality (1975-2006)	log total patents (renewed thrice) (1975-2006)	log total patents (top 10% only) (1975-2006)
log income (1930)	.000	.013*	.031***	.033***
	(.008)	(.008)	(.009)	(.008)
log education (1930)	.130***	.142***	.098***	.085***
	(.008)	(.008)	(.008)	(.008)
Obs.	81,348	81,348	78,438	81,348
R^2	0.02	0.04	0.05	0.03

Notes: See notes for Table 3.

average patent quality measure puts less weight on inventors who come up with a high number of innovations which are of mediocre quality. Similarly, the maximum patent quality measure only considers the best invention of a given inventor, comparing inventors according to the best ideas they came up with and ignoring everything else. The results are very similar to the baseline analysis: log education dominates in both regressions, and log income is either statistically insignificant, or significant at the 10% level and economically insignificant.

In column 3, a new patent quality metric is introduced: patent renewal status. USPTO requires the patent holders to renew their patents 4, 8, and 12 years after the patent grant date by paying a small fee. If the patent holders do not renew their patents on these dates, they lose the monopoly rights on their invention. There is significant variation in how many times patents are renewed. The patent quality metric used on column 3 assigns a quality of 1 if the patent was renewed three times throughout its duration, and 0 otherwise. Hence only patents which were seen sufficiently valuable by their holders to renew three times are counted.[29] The results with this metric are similar in that education dominates income, but this time the effect of income is not statistically insignificant.

Last column does the opposite, and focuses on a patent quality measure that only puts weight on the best inventions produced in a year. For each year, the patents are ranked according to the citations they receive. Only the top 10% of the inventions in a given year are assigned a quality of 1, whereas the remaining 90% are assigned a quality of 0. Using the inventor quality measure derived from this new measure of patent quality, the results are similar to column 3: education is found to dominate income once again.

[29] Note that although this quality metric is very reliable in weeding out patents that turn out to be worthless over time, it provides no quality variation between patents which are sufficiently valuable to be renewed every single time. Hence it should be thought of as a quality measure that is more informative in the lower tail of the patent quality distribution as opposed to the upper tail.

3.4.4 Controlling for demographic changes and immigration between samples

The Unites States is a country of immigrants, and it has received significant immigration during the time period from 1930 to 2008. Many surnames that were very rare in the 1930s are now quite common. In contrast, some surnames are now less frequent, either due to being crowded out by the new or existing surnames, or due to low number of offspring or higher mortality rates. Could any of these demographic changes bias the obtained estimates in a particular direction, potentially causing wrong conclusions to be drawn? Recognizing this possible problem, this section is dedicated to investigating whether this is true.

In order to tackle this issue, a simple variable called population share ratio is constructed. The share of a surname in the population in 2000 is divided by that in 1930. This ratio is larger than unity if the surname has increased in frequency, which is the case for many immigrant surnames. Conversely it is smaller than unity for surnames which actually lost their prominence over time. Using this ratio as an additional explanatory variable, Table 5 repeats the extensive margin regression on column 3 of Table 2. Columns 1 and 2 repeat the regression after dropping the top and bottom 25% of the sample according to population share ratio. Hence they drop the extremely over- and under-achieving surnames from the sample respectively. Columns 3 and 4 repeat the same exercise keeping only the top and bottom halves of the sample respectively, i.e. looking at over- and under-achievers within their own groups. The last column retains the whole sample, but includes the population share ratio as a linear regressor. Although the magnitudes change, income is found to be dominant in all cases, whereas education is found to be either insignificant, or significant but negatively correlated. In addition, when included as a linear regressor, the population share ratio turns out to be insignificant. Consequently, the findings of the extensive margin analysis are found to be robust.

Table 6 repeats the same analysis done in Table 5 for column 3 of Table 3. The results are quite similar: Although the exact quantitative magnitudes may vary, the effect of education is always quite large and positive, dominating that of income. The effect of income is found to be statistically or economically insignificant in all cases. The only difference is observed when population share ratio is added as a linear regressor: Its coefficient turns out to be significant at the 5% level and positive. However, estimated at 1.6%, its coefficient is much smaller compared to the coefficient of education (17.7%). Hence, it is once again concluded that the findings in the intensive margin analysis are robust.

One could also be worried about another issue: It is possible that a surname the frequency of which is stable over the 1930-2008 time period actually belonged to people who were recent immigrants in 1930. Systematic differences between such surnames and those who were already largely stable in frequency prior to 1930 could lead to potential biases similar to those discussed

Table 5: Immigration Robustness (1930-2000) - Extensive Margin

	relative representation (1975-2008)				
	(1)	(2)	(3)	(4)	(5)
income (1930)	.341***	.249***	.232***	.158***	.239***
	(.009)	(.011)	(.013)	(.006)	(.010)
education (1930)	−0.021***	.003	.005	−.002	.001
	(.004)	(.005)	(.006)	(.004)	(.005)
pop. share(2000)/pop. share(1930)					−.025
					(.032)
Obs.	82,718	82,735	55,148	55,210	110,290
R^2	0.13	0.30	0.34	0.04	0.27

Notes: Columns 1 and 2 repeat the regression in the last column of Table 2 after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 3 and 4 repeat the same exercise for the top and bottom halves of the sample respectively. Column 5 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. All notes for Table 2 apply.

Table 6: Immigration Robustness (1930-2000) - Intensive Margin

	log quality wtd. total patents (1975-2006)				
	(1)	(2)	(3)	(4)	(5)
log income (1930)	−.015*	.011	.011	.037***	.001
	(.009)	(.010)	(.012)	(.008)	(.009)
log education (1930)	.162***	.173***	.177***	.145***	.177***
	(.006)	(.010)	(.012)	(.008)	(.009)
pop. share(2000)/pop. share(1930)					.016**
					(.032)
Obs.	61,013	61,011	40,684	40,676	81,348
R^2	0.03	0.06	0.07	0.03	0.05

Notes: Columns 1 and 2 repeat the regression in the last column of Table 3 after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 3 and 4 repeat the same exercise for the top and bottom halves of the sample respectively. Column 5 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. All notes for Table 3 apply.

earlier. Luckily, it is possible to construct a similar population share ratio using surname frequencies in 1930 and 1880, using an earlier IPUMS-USA sample. The cost of doing so is losing observations that belong to surnames which do not exist in the 1880 census sample. The results of this robustness analysis are qualitatively very similar, and can be found on Tables 13 and 14 in the empirical appendix.

3.4.5 Summary of empirical results

The two stylized facts obtained in the empirical analysis can be summarized as follows:

Fact 1: Individuals from richer backgrounds are much more likely to become inventors (23.9%); whereas those from more educated backgrounds experience no similar advantage (0.1%).

Fact 2: Conditional on becoming an inventor, individuals from more educated backgrounds turn out to be much more prolific inventors (17.5%); whereas those from richer backgrounds exhibit no such aptitude (0.1%).

However, these results by themselves would be insufficient to establish whether there is an economically significant misallocation of talent or not, given that innate ability is unobserved in the data. This is important, since (i) innate ability is likely to play a large role in determining the probability of becoming an inventor as well as success conditional on becoming one, (ii) innate ability is found to be very persistent across generations by other studies (Clark (2014), Olivetti and Paserman (2015)), and this might be causing the observed strong positive correlations. In order to measure the extent of the misallocation of talent in innovation, the model developed in Section 2 is employed, where the regressions run here are replicated within the model, targeting the empirical coefficient estimates. The next section describes this calibration exercise.

4 Calibration

4.1 Solution method

Computation of the solution requires value function iteration to solve for $V_o(y_o, h, a; \Theta), W(b, h, a; \Theta)$ and $V_y(y_y, a; \Theta)$ and the associated policy functions $\hat{b}(y_o, h, a; \Theta), \hat{n}(b, h, a; \Theta), \hat{h}(y_y, a; \Theta)$ and $\hat{s}(y_y, a; \Theta)$. Simulation of the joint stationary distribution of jobs, innate ability, and early childhood education as well as the stationary distribution of normalized savings are necessary to calculate the aggregate supplies as well as the cut-off score threshold $\bar{s}$. The results of the firm's maximization problem and the market clearing conditions boil down to analytical non-linear equations in K, L_u and L_s as discussed in Section 2. Then these are solved to obtain the balanced growth path equilibrium. The pseudo-code for the algorithm used to solve for the BGP equilibrium can be found in Appendix A.

4.2 Identification

The simulation of the model requires the assignment of values to several parameters. There are nineteen parameters to pick: $\beta, \omega, \alpha, \kappa, \lambda, \delta, \Gamma, \xi, \psi, \epsilon, \rho, \nu, \sigma_a, \eta, \kappa_h, \xi_h, \kappa_n, \xi_n, \sigma_j$. In order to select values for the parameters, a set of empirical targets are specified for the model to match. Some common parameters are chosen from existing studies, and the rest are internally calibrated by employing a minimization routine that seeks to match the data targets with the associated model-generated counterparts. In particular, some of the regressions found on Section 3 are replicated in the model, and the minimization algorithm attempts to achieve the same coefficients ("betas") with regressions run on model-simulated data, where the variables are normalized in the same manner. The summary of the calibration exercise is presented on Table 7. The details are as follows:

Table 7: Parameter Values

Parameter	*Description*	*Identification*
	External Calibration	
$\omega = 2.00$	CRRA parameter	Kaplow (2005)
$\alpha = 0.50$	Parental altruism	Aiyagari, Greenwood, and Seshadri (2002)
$\kappa = 0.25$	Capital's share in production	Corrado, Hulten, and Sichel (2009)
$\lambda = 0.60$	Labor's share in production	Corrado, Hulten, and Sichel (2009)
$\delta = 0.82$	Depreciation rate	U.S. NIPA
$\xi = 0.50$	Concavity of innovation production	Hall and Ziedonis (2001)
$\sigma_a = 0.70$	St. dev. of innate ability shock	Knowles (1999)
$\eta = 11.6\%$	Fraction of skilled jobs	U.S. Census (2013)
	Internal Calibration	
$\beta = 0.28$	Discount factor	Real interest rate
$\Gamma = 0.92$	Innovation productivity increase	GDP growth rate
$\rho = 0.70$	Persistence of innate ability	IG corr. of earnings
$\kappa_h = 0.04$	Cost of pre-college education investment	Education spending/GDP
$\kappa_n = 0.05$	Cost of credentialing investment	Inequality targets
$\xi_h = 1.30$	Convexity of pre-college education inv.	Inequality targets
$\xi_n = 2.50$	Convexity of credentialing inv.	Inequality targets
$\psi = 0.40$	Education share of ind. productivity	Regression targets
$\epsilon = 1.90$	Ind. productivity elasticity	Regression targets
$\nu = 0.89$	Influence of credentialing spending	Regression targets
$\sigma_j = 0.80$	St. dev. of job shock	Regression targets

Notes: All internally calibrated parameters are identified jointly; the moments in the internal calibration panel are provided for intuition.

1. *CRRA parameter:* This parameter is taken to be $\omega = 2.00$, consistent with the estimates listed in Kaplow (2005).

2. *Parental altruism parameter:* This variable is chosen to be $\alpha = 0.50$, following Aiyagari, Greenwood, and Seshadri (2002).

3. *Capital's and labor's share of income:* Corrado, Hulten, and Sichel (2009) calculate the shares of tangible capital, labor, and intangible capital to be $\kappa = 0.25$, $\lambda = 0.60$ and $\zeta = 0.15$ respectively. The share of intangible capital they calculate is mapped to the share of productivity of a firm in generating output in the model.

Table 8: Calibration Targets

Target	*U.S. Data*	*Model*
Aggregate targets		
Yearly real interest rate	4.00%	4.00%
Yearly GDP growth rate	2.00%	2.00%
Education spending/GDP	7.30%	8.55%
Intergenerational correlation targets		
IG corr. of earnings	0.70	0.70
IG corr. of wealth	0.37	0.33
Inequality targets		
Wage income Gini index	0.48	0.52
Log 90/10 ratio	1.08	1.17
Log 90/50 ratio	0.46	0.52
Log 50/10 ratio	0.62	0.65
Regression targets		
Extensive margin, income effect	0.24	0.19
Extensive margin, education effect	0.00	0.07
Intensive margin, income effect	0.00	0.08
Intensive margin, education effect	0.18	0.22

4. *Depreciation rate for capital:* The annual depreciation rate of physical capital is chosen as 6.9% which is consistent with the U.S. National Income and Product Accounts. Since each period lasts 25 years, $\delta = 0.82$.

5. *Concavity of innovation production:* Following Hall and Ziedonis (2001), the concavity parameter of the innovation production function is chosen as $\xi = 0.50$. This is the most widely used value in the literature.

6. *Standard deviation of innate ability shock:* This parameter is chosen to be $\sigma_a = 0.70$, in line with findings on empirical income distributions reported in Knowles (1999).

7. *Fraction of skilled jobs:* This parameter is chosen such that it equals the percentage of individuals in the U.S. with graduate degrees, which is 11.6% (U.S. Census, 2013).

8. *Long-run interest rate:* The long-run interest rate of 4.0% is targeted, which determines the discount factor β.

9. *Long-run output growth:* Since 1945, the the aggregate output in the U.S. grew at circa 2% per year. The parameter Γ determines the increase in productivity innovation generates, and hence it plays the foremost role in determining the output growth rate in the model.

10. *The ratio of education spending to GDP:* The ratio of the aggregate spending on education to GDP in the U.S. is around 7.30% (OECD, 2013).The model counterpart of this ratio is

the aggregate resources spent on education over total output.

11. *Intergenerational correlation of earnings:* The persistence of earnings across generations is an important statistic for the model to replicate, since it puts discipline on the persistence of innate ability which is unobserved. The value of 70% is targeted in the baseline analysis (Knowles (1999)).[30]

12. *Intergenerational correlation of wealth:* The persistence of wealth across generations is also an important statistic to replicate, since the mechanism that generates the misallocation of talent in the model works through the wealth inequality between households. This value is estimated to be 37% in Charles and Hurst (2003).

13. *Inequality targets:* The calibration procedure aims to generate a realistic income distribution. To this end, various inequality metrics are calculated using the model-generated distribution, and matched with their empirical counterparts. These are the Gini index, and log 90/10, 90/50 and 50/10 ratios.[31]

14. *Indirect inference:* The baseline extensive and intensive margin regressions in Section 3 are replicated in the model. Income is proxied by the income of the agents in the model, and education is proxied by pre-college education. Relative representation among inventors in the data is mapped to relative representation in the innovation sector jobs. Inventor quality in the data is mapped to individual productivity conditional on having an innovation job. As in the empirical analysis, all variables are normalized by subtracting the mean and dividing by the standard deviation. The coefficients of income end education in both margins are then targeted.

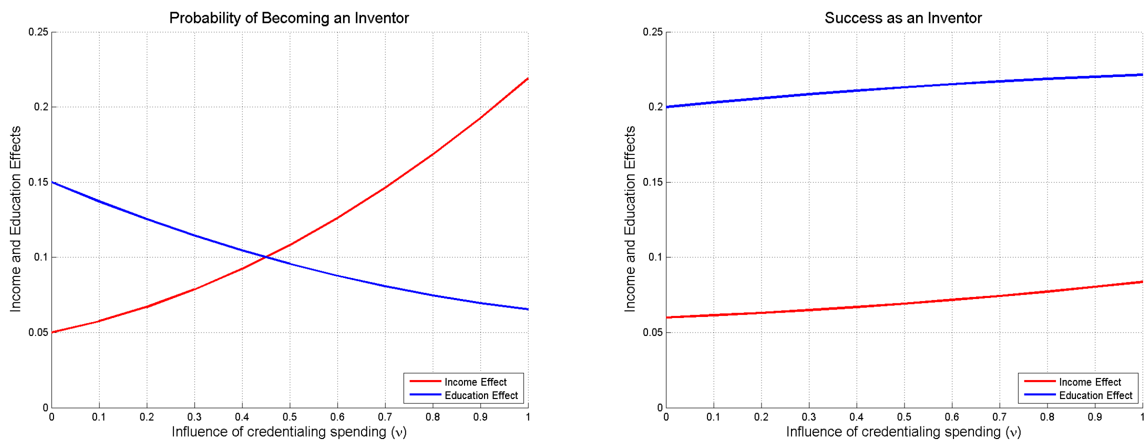
The success of the calibration exercise in matching the data targets is presented on Table 8. The interest rate and the yearly GDP growth rate are hit very precisely, and they determine the values of β and Γ respectively. The model generates an education spending to GDP ratio somewhat higher than what is observed in the U.S. data. Given that the number taken from the data does not include the opportunity cost of time spent by parents in order to nurture their children, overshooting might not be a significant problem.

The intergenerational correlation of earnings is hit precisely, which disciplines the persistence of (unobserved) innate ability ρ (positively related), but is also influenced by the standard deviation of the idiosyncratic job shock σ_j (negatively related). The intergenerational correlation of wealth the model produces is 0.33, which is somewhat lower than the value of 0.37 observed in the data, but still within a reasonable range.

[30]Since there are also estimates of intergenerational correlation of earnings as low as 40% in the literature, the model is re-estimated with a lower target as a robustness check in Section 5.4.

[31]Note that one of these three ratios is a deterministic function of the other two; so it provides no additional information.

FIGURE 3: CHANGES IN INCOME AND EDUCATION EFFECTS WITH VARYING VALUES OF ν



The model generates a wage income distribution slightly more unequal compared to the U.S. economy. For instance, the Gini index is calculated to be 0.52 as opposed to 0.48 observed in the data. However, the remaining inequality targets that measure the inequality in different sections of the distribution show that the model is successful in matching the shape. Log 90/10, log 90/50 and log 50/10 ratios are all slightly higher than their data counterparts by similar percentages.

The model is able to replicate the dominance of income on the extensive margin (the probability of getting an innovation sector job) and the dominance of education on the intensive margin (the success conditional on getting an innovation job). The coefficients of the dominated effects (education on the extensive margin, and income on the intensive margin) are not precisely zero, so the starkness of the differences are more similar to those observed in the regressions on Columns 3 and 4 of Table 4, as opposed to that on Column 3 of Table 3 on the intensive margin.

Generating the discrepancy between the effects of income and education on the two margins is made possible by the credentialing spending channel. Figure 3 plots the effects of income (red) and education (blue) while varying the influence of credentialing spending ν in the range of values it can take ($\nu \in [0, 1]$). The left panel plots the effects on the extensive margin, and the right panel plots the same on the intensive margin. As ν increases from 0 to 1, the predictive power of ancestor income on the probability of becoming an inventor increases, whereas that of education decreases. On the other hand, increasing ν from 0 to 1 does not change the predictive power of ancestor income and education in opposite directions, slightly increasing both at the same time.[32] This makes it possible to change the value of ν such that the dominance pattern observed in the data can be hit in the model generated regressions. This differential effect of ν on the two margins

[32]Whether income or education dominates on the intensive margin (i.e. inventor productivity) is determined by other parameters of the model. The model is able to generate any correlation pattern, including the exact opposite of the empirically observed pattern of dominance, by changing the parameter values.

provides the intuition on how targeting the dominance pattern helps pin down its value.[33]

5 Quantitative Results

In this section, using the parameter values estimated in Section 4, several quantitative experiments are conducted to better understand the mechanism of the model, to assess the welfare costs associated with the misallocation of talent due to the credentialing spending channel, and to determine socially optimal progressive bequest tax schedules.

The first subsection describes the social welfare function used in the study, and how two different steady-state economies are compared with each other. The following subsection conducts a hypothetical thought experiment where the credentialing spending channel is completely shut down, which results in an increase in the aggregate output growth rate as well as social welfare through a reduction in the misallocation of talent.

The third subsection focuses on how a benevolent government can increase social welfare and economic growth in a decentralized market economy through the policy tool of progressive bequest taxation. Although the growth effect is found to be around 25% of what can be achieved by shutting down the credentialing spending channel, the welfare increase is found to be larger and quite significant: 6.20% in consumption-equivalent terms

In Section 5.4, the model is recalibrated with a lower intergenerational correlation of earnings target of 0.45 in order to check whether the model generates similar quantitative implications. Repeating the credentialing spending shut-down experiment with the new calibration leads to higher growth and welfare effects, however the increase in magnitudes are not too large.

5.1 Welfare comparisons

In order to measure welfare, a utilitarian social welfare function is employed where each household is weighed equally. The social planner is assumed to assign equal value to the utility from consumption of all members of a household at a given time. The utility in the future is discounted by the discount factor β of the household. Hence, the social welfare in a balanced growth path equilibrium with

[33] One could be worried about whether other parameters that play a part in the determination of the individual score could generate a similar differential effect on the two margins. The prime candidates are the elasticity of substitution between innate ability and early childhood education ϵ, and the share of education in individual productivity ψ. It is found out that this is not the case. In particular, if the credentialing spending channel is shut down ($\nu = 0$), both ϵ and ψ change the effects of income and education in the same direction on both margins at the same time. Therefore if the credentialing spending channel is removed from the model without introducing any other mechanisms, the model is unable to mimic the domination patterns observed in the data.

output growth rate g is given by

$$\begin{aligned} W &= \sum_{t=0}^{\infty} \beta^t \int_{m=0}^{1} \left(\frac{c_{c,m,t}^{1-\omega}}{1-\omega} + \frac{c_{y,m,t-1}^{1-\omega}}{1-\omega} + \frac{c_{o,m,t-2}^{1-\omega}}{1-\omega} \right) dm \\ &= \frac{\int_{m=0}^{1} \left(c_{c,m,0}^{1-\omega} + c_{y,m,-1}^{1-\omega} + c_{o,m,-2}^{1-\omega} \right) dm}{(1-\omega)(1-\beta(1+g)^{1-\omega})} \end{aligned} \tag{21}$$

The welfare comparisons between different economies will be conducted by comparing the balanced growth path equilibria.[34] In order to make two different economies A and B comparable, both economies will be started at the same aggregate productivity level $\bar{z}_0^A = \bar{z}_0^B = 1$. Let $\nu > 0$ be the scalar such that multiplying every agent's consumption in economy A with ν results in a welfare number equivalent to the one in economy B. Simple algebra reveals that ν is given by

$$\nu = (W^B/W^A)^{1/(1-\omega)} \tag{22}$$

where W^A and W^B denote the welfare in economies A and B respectively. The welfare gain or loss a move from economy A to economy B provides in consumption-equivalent terms is given by $\nu - 1$. This welfare measure is used in all quantitative exercises.

5.2 Shutting down the credentialing spending channel

How does the misallocation of talent affect economic growth and social welfare? In order to address this question, a simple hypothetical thought experiment will be conducted. Recall that individuals heterogeneous in innate ability, early childhood education and wealth are able to receive inventor training if they can achieve a high enough score given by

$$\tilde{s}(l(h,a),n) = (1-\nu)l(h,a) + \nu n + \epsilon_j.$$

The score of an individual is partially influenced by the actual individual productivity $l(h,a)$, partially by the credentialing spending n, and partially by the random shock ϵ_j. Given the scarcity of inventor training, increasing the growth rate of the economy is only possible through improving the composition of the individuals who get inventor training in terms of individual productivity. If the influence of credentialing spending could be diminished such that $\nu = 0$, the scores of the individuals would be perfectly correlated with their actual individual productivity except for the random shock. This would result in highly talented individuals ending up in the innovation sector, where they can contribute to the aggregate productivity growth. Following this line of thought,

[34] Hence this analysis ignores the welfare effects of the transition to the new steady state. However, including the effects of the transition would amplify, rather than diminish, the calculated welfare numbers, since the steady-state capital stock is lower in the counterfactual economies considered, which would allow the consumption of the extra capital stock in the transition to the stationary equilibrium.

the economy calibrated in Section 4 is taken, and the parameter ν is set to 0. This hypothetical economy is then compared to the baseline economy.

Table 9: Shutting down the credentialing spending channel

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.19	0.05	-73.7%
Extensive margin, education effect	0.07	0.15	114%
Intensive margin, income effect	0.08	0.06	-25.0%
Intensive margin, education effect	0.22	0.20	-9.09%
Yearly GDP growth rate	2.00%	2.21%	10.4%
Education spending/GDP	8.55%	10.2%	19.1%
Aggregate skilled labor, L_s	0.48	0.62	28.4%
Aggregate unskilled labor, L_u	1.91	2.00	4.69%
Mean innate ability of skilled workers, a	2.08	2.57	23.4%
Mean early childhood education of skilled workers, h	2.27	2.96	30.1%
Mean parental wealth of skilled workers, y_o	0.87	0.84	-4.32%
Mean bequests received of skilled workers, b	0.49	0.25	-49.5%
Wage income Gini index	0.52	0.56	6.61%
Log 90/10 ratio	1.17	1.20	3.10%
Log 90/50 ratio	0.52	0.57	9.30%
Log 50/10 ratio	0.65	0.64	-1.88%

Table 9 displays the values of several statistics of interest in the baseline and hypothetical economies and how much they change in percentage terms. The first four rows display how the effects of income and education in the extensive and intensive margins change. In the baseline economy, income effect dominated in the determination of the chances of getting an innovation job, whereas education effect dominated in the prediction of productivity conditional on becoming an inventor. Now that the credentialing channel has been shut down, education effect dominates in both the extensive and intensive margins. Thus the people who would perform better as inventors and those who actually become inventors largely coincide.

The annual GDP growth rate changes from 2.00% to 2.21%, a large increase. This is caused by a 28.4% increase in the aggregate skilled labor supply L_s. Investigating the changes in the characteristics of the people who become inventors reveals that this is driven by higher quality individuals in terms of both innate ability and early childhood education. The mean innate ability a of inventors increases by 23.4%, indicating a better allocation of naturally talented individuals to where their contribution would be the greatest. Furthermore, these individuals also receive more early childhood education investment when they are children, further increasing the average individual productivity of inventors.

Looking at the parental backgrounds of the inventors, it is observed that the mean parental wealth is slightly lower by -4.32%. However the mean bequests received fall tremendously by

49.5%. This is driven by two effects working in the same direction: (1) since $\nu = 0$, it is no longer possible for less talented children with wealthier parents to outperform the more talented but less wealthy competitors in score by outspending them in credentialing, (2) given that their children do not need to spend any money on credentialing, parents do not deem it necessary to leave large bequests, spending some of the extra windfall for their own consumption, and the rest on the productive early childhood education investment which improves individual productivity l and score $\tilde{s}$ simultaneously.

The inequality measures tell a different story: The decrease in the misallocation of talent is beneficial for economic growth, but it also leads to a more unequal society in terms of income. The Gini index increases from 0.52 to 0.57. Examining the income ratios is more revealing: Log 90/10 ratio increases by 3.10%, exhibiting an increase in the gap between the rich and the poor. However log 90/50 ratio increases at a much higher rate of 9.30%, whereas log 50/10 ratio decreases by 1.88%. These results indicate that the increase in inequality is largely driven by the upper tail of the income distribution. As more naturally talented individuals have better chances at becoming inventors, they are also able to earn higher incomes, drifting away from the rest of the workers.

As a combined result of all of these changes, the welfare in the hypothetical economy is 5.93% higher than the baseline economy in consumption-equivalent terms. However, it is important to keep in mind that the hypothetical economy is still far away from the first best. Although the misallocation of talent in the college education stage is reduced to the effect of the randomness inherent in the allocation process only, the early childhood education investment in children is still a function of parental wealth. Thus there is still room for improvement. In addition, the egalitarian social welfare function assigns importance to equalizing outcomes between households in terms of consumption, so holding everything constant, there are also potential gains from redistribution of resources. The following subsection discusses a potential government policy which can address a combination of the listed concerns simultaneously.

5.3 Progressive bequest taxation

The previous thought experiment shows that reducing the misallocation of talent in the economy by shutting down the credentialing spending channel can lead to significant gains in growth and welfare. Can a benevolent government achieve similar gains by utilizing available policy options in a decentralized economy? To this end, socially optimal progressive bequest taxes will now considered. In order to reduce the cost of computation, a particular functional form is assumed with the scale parameter τ_s and the progressivity parameter τ_p such that the budget constraint of the old adults in the decision problem given in (12) becomes

$$c_o + \left(\frac{b}{1-\tau_s}\right)^{\frac{1}{1-\tau_p}} \leq y_o$$

which is equivalent to the old budget constraint if $\tau_s = \tau_p = 0$. All the collected taxes are then transferred to the young adults as a type-independent lump-sum transfer Tr, changing the equation that determines y_y in (13) to

$$y_y = \left(w_{j_y} + \frac{w'_{j_y}}{1+r'} \right) l_y(h_y, a_y) + b - c_n(n) + Tr.$$

In order to prevent lump sum taxes, $Tr \geq 0$ is imposed, and the government must balance its budget every period.

Table 10: Optimal progressive bequest taxation results

Variable	*Baseline*	*Optimal b tax*	Change
Extensive margin, income effect	0.19	0.17	-10.5%
Extensive margin, education effect	0.07	0.08	14.3%
Intensive margin, income effect	0.08	0.02	-75.0%
Intensive margin, education effect	0.22	0.27	22.7%
Yearly GDP growth rate	2.00%	2.05%	2.50%
Education spending/GDP	8.55%	9.13%	6.85%
Aggregate skilled labor, L_s	0.48	0.51	6.29%
Aggregate unskilled labor, L_u	1.91	1.93	0.94%
Mean innate ability of skilled workers, a	2.08	2.15	3.33%
Mean early childhood education of skilled workers, h	2.27	2.47	8.90%
Mean parental wealth of skilled workers, y_o	0.87	0.85	-3.05%
Mean bequests received of skilled workers, b	0.49	0.43	-10.6%
Wage income Gini index	0.52	0.53	1.92%
Log 90/10 ratio	1.17	1.17	0.54%
Log 90/50 ratio	0.52	0.52	0.00%
Log 50/10 ratio	0.65	0.66	0.01%

The welfare maximizing values of τ_s and τ_p are found to be 0.125 and 0.171 respectively. The bequest tax schedule implied by these two values is quite progressive: The average bequest tax rate faced by the top 1% is 12.1%, whereas this number falls to 9.70% for the top 5%, and 4.18% for the top 10%. In fact, when the transfers are also taken into account, the bottom 95% of the households are net recipients, whereas only the top 5% pay into the system. Furthermore, as it will be demonstrated later on, this progressive taxation scheme does not result in a less productive society: the aggregate productivity of the inventors and the growth rate of output are higher in this alternative economy. Hence the increased equity does not come at the cost of reducing efficiency.

Table 10 shows how the statistics of interest change compared to the baseline under the optimal progressive bequest taxation policy. Looking at the regression targets, and the extensive margin in particular, income loses its explanatory power by 10.5% of its value, whereas that of education increases by 14.3%. The effects on the intensive margin are much more pronounced, where income loses 75% of its explanatory power, and education completely dominates. All of these targets point towards a decrease in the misallocation of talent.

The growth rate of the economy increases to 2.05% from its baseline value of 2.00%, which corresponds to one quarter of the effect observed in the case of $\nu = 0$. This is caused by the increase in the aggregate skilled labor supply L_s by 6.29%. Examining the mean innate ability a and early childhood education h of inventors, the increase of quality in the composition is driven more by early childhood education (8.90%) rather than innate ability (3.33%). So it can be argued that the optimal bequest taxes contribute to the growth rate of the economy more through reducing the suboptimal investment in early childhood education rather than allocating higher innate ability people to the innovation sector. However, both channels have a positive contribution regardless of their relative power.

In contrast to the thought experiment where credentialing spending is shut down, the increase in the growth rate of the economy is not accompanied by a significant increase in income inequality. The inequality metrics under the optimal taxation policy have very similar values to their baseline values. This is caused by the redistributive nature of the optimal tax policy. As a result of this, even though the growth gain is one quarter of the $\nu = 0$ case, the welfare gain is calculated to be slightly higher: 6.20% in consumption-equivalent terms.

5.4 Recalibration with lower intergenerational earnings persistence

The intergenerational persistence of innate ability ρ is an important parameter of the model, the value of which has an important bearing on quantitative counterfactuals. Since innate ability is not directly observable, the value of ρ is indirectly inferred by trying to match the intergenerational correlation of earnings (IGE) generated in the model with that found in the data. However, the exact value of IGE in the U.S. over the time period is not a settled topic in the literature.[35] Although consistent with the highly persistent effects of income and education discovered in Section 3, the value of 0.70 targeted in the baseline analysis is on the higher end of the estimates found in the literature. This section repeats the calibration exercise in Section 4 with a lower IGE target of 0.45, and assesses its effects.

The calibrated values of most parameters remain very similar to the results in Table 7, with the exception of intergenerational persistence of innate ability, ρ. This falls from 0.70 to 0.40, a very significant decrease. As a result, the effects of income and education on both margins fall, as well as the differences between the effects for a given regression. The earnings inequality in the steady state is also lower.

[35] See the seminal work of Solon (1999) on the issue, and Black and Devereux (2010), Chetty, Hendren, Kline, and Saez (2014) and the references therein for a recent survey of the literature.

Table 11: Shutting down the credentialing spending channel - Low earnings persistence

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.05	0.03	-40.0%
Extensive margin, education effect	0.03	0.05	66.7%
Intensive margin, income effect	0.05	0.07	40.0%
Intensive margin, education effect	0.06	0.09	50.0%
Yearly GDP growth rate	2.00%	2.29%	14.5%
Education spending/GDP	7.09%	10.7%	51.5%
Aggregate skilled labor, L_s	0.37	0.52	42.5%
Aggregate unskilled labor, L_u	1.72	1.92	11.9%
Mean innate ability of skilled workers, a	1.85	2.45	32.5%
Mean early childhood education of skilled workers, h	1.43	2.18	52.7%
Mean parental wealth of skilled workers, y_o	0.70	0.74	5.34%
Mean bequests received of skilled workers, b	0.45	0.28	-38.3%
Wage income Gini index	0.49	0.52	6.37%
Log 90/10 ratio	1.03	1.00	-3.11%
Log 90/50 ratio	0.48	0.46	-3.71%
Log 50/10 ratio	0.56	0.54	-2.59%

How does the lower value of ρ effect the counterfactual experiments? In order to answer this question, the hypothetical thought experiment in Section 5.2 is repeated under the new calibration. Table 11 summarizes the results of shutting down credentialing spending by setting $\nu = 0$. The output growth rate of the economy increases from 2.00% to 2.29%, driven by a huge 42.5% increase in aggregate skilled labor supply. Compared to the baseline economy, the welfare gain is found to be 6.63% in consumption-equivalent terms.

These values are slightly higher compared to those found in Section 5.2. Why is this the case? Inspection reveals that this is caused by a higher degree of initial misallocation of talent in the low IGE economy. Under the baseline calibration, due to the higher persistence of innate ability ρ at 0.70, the rich and the talented largely coincide in the stationary equilibrium. When this persistence is lower at 0.40, the chances of a genius being born to a comparatively poor household are higher. As a result of this, the mean innate ability a of inventors is lower before the shutdown of the credentialing channel. Hence, the growth and welfare implications are amplified when ρ is lower.

6 Conclusions

This paper develops a model of misallocation of talent in the innovation sector. Workers in the economy are finitely-lived, and heterogeneous in terms of wealth, early childhood education, and innate ability. The sectors in the economy are separated into production and innovation, where the latter serves to improve the productivity of the prior. The training necessary to become a worker in the innovation sector is scarce. Agents compete against each other in order to acquire this scarce training so that they can get innovation sector jobs that pay more. They use productive early childhood education investment as well as (socially) unproductive credentialing spending in order to increase their chances. Financial frictions in the form of a non-negative bequest constraint and the inability to insure against idiosyncratic risk, coupled with the misalignment of private and social incentives result in a misallocation of talent across the two sectors. The nature and magnitude of this misallocation of talent are examined.

Empirical analysis makes use of three sets of micro-data—NBER USPTO Utility Patents Grant Database, The Careers and Co-Authorship Networks of U.S. Patent-Holders, and IPUMS-USA 1930 5% Sample—that were previously unlinked in order to establish two new stylized facts: (1) People from richer backgrounds are more likely to become inventors; but those from more educated backgrounds are not. (2) People from more educated backgrounds become more prolific inventors; but those from richer backgrounds exhibit no such aptitude. This discrepancy suggests a misallocation of talent in the innovation sector, which motivates the development of a model that can generate the correlation patterns observed in the data. The results are robust to the use of alternative patent and inventor quality measures, as well as potential biases that might be caused by immigration and similar demographic changes.

The developed model is calibrated to match data targets including aggregate moments of the U.S. economy such as the yearly long-run output growth and real interest rates and the ratio of education spending to GDP; moments obtained using micro data such as intergenerational correlation of earnings and wealth and various inequality measures regarding the earnings distribution; as well as data targets taken from the original empirical analysis such as the effect of income and education on the probability of getting an innovation sector job, and the productivity conditional on having one. The calibrated model is then used to explore how the misallocation of talent between the production and innovation sectors is generated, and the findings suggest that the welfare effects of this misallocation might be substantial.

The quantitative analysis reveals that if the credentialing spending channel could be shut down, the aggregate output growth rate would increase from 2.00% to 2.21%, leading to a welfare gain of 5.93% in consumption-equivalent terms. Another quantitative experiment that seeks to calculate the socially optimal bequest taxation policy reveals that the growth rate could be increased to 2.05% even in a decentralized market economy by leveling the playing field and reducing the effect of suboptimal early childhood education spending due to financial frictions. The resulting welfare gain is quite significant at 6.20%. A robustness analysis is conducted to show how the model performs when different calibration targets are chosen, and the quantitative results remain largely similar.

The stylized facts established in the empirical analysis are quite provoking, and the model suggests that reducing the existing misallocation of talent in the economy might yield significant welfare gains through an increase in the long-run output growth rate. Given how important the upper tail of the talent distribution is in generating the ideas that drive economic progress, it is likely that policies that alleviate the misallocation through reducing wealth inequality or financial frictions might be desirable. Future research is needed to establish more detailed policy responses that take additional life-cycle elements into account, as well as the welfare implications of the transition to the new steady-state. The empirical methodology used in the paper can also be applied in any other sector where surname-level information is available, which would considerably expand our understanding of the allocation of talent in other sectors, as well as the intergenerational dynamics of socioeconomic status.

References

Acemoglu, D. (2009): *Introduction to Modern Economic Growth.* Princeton University Press.

Acemoglu, D., U. Akcigit, H. Alp, N. Bloom, and W. Kerr (2018): "Innovation, Reallocation and Growth," *American Economic Review*, 108(11), 3450–91.

Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, and D. Hemous (2019): "Innovation and Top Income Inequality," *The Review of Economic Studies*, 86(1), 1–45.

Aghion, P., U. Akcigit, and P. Howitt (2014): "What Do We Learn From Schumpeterian Growth Theory?," in *Handbook of economic growth*, vol. 2, pp. 515–563. Elsevier.

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2017): "Living the American Dream in Finland: The Social Mobility of Inventors," Working Paper.

Aghion, P., and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60(2), 323–351.

——— (2009): *The Economics of Growth.* MIT Press.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Aiyagari, S. R., J. Greenwood, and A. Seshadri (2002): "Efficient Investment in Children," *Journal of Economic Theory*, 102(2), 290–321.

Akcigit, U., M. A. Celik, and J. Greenwood (2016): "Buy, Keep, or Sell: Economic Growth and the Market for Ideas," *Econometrica*, 84(3), 943–984.

Akcigit, U., J. Grigsby, and T. Nicholas (2017): "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age," Working Paper.

Alvarez, F. E., F. J. Buera, and R. E. Lucas (2013): "Idea Flows, Economic Growth, and Trade," National Bureau of Economic Research WP 19667.

Anderberg, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93(9-10), 1017–1026.

Banerjee, A. V., and A. F. Newman (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101(2), 274–298.

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* University of Chicago Press.

Becker, G. S., and N. Tomes (1979): "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87(6), 1153–1189.

——— (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), 1–39.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994): "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," *Journal of Political Economy*, 102(6), 1131–1174.

Behrman, J. R., P. Taubman, and T. J. Wales (1977): "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success," in *Kinometrics: Determinants of Socioeconomic Success within and between Families*, ed. by P. Taubman. North-Holland.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019): "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 134(2), 647–713.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), 352–365.

Black, S. E., and P. J. Devereux (2010): "Recent Developments in Intergenerational Mobility," National Bureau of Economic Research WP 15889.

Bohacek, R., and M. Kapicka (2008): "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55(1), 1–16.

Charles, K. K., and E. Hurst (2003): "The Correlation of Wealth Across Generations," *Journal of Political Economy*, 111(6), 1155–1182.

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," *The Quarterly Journal of Economics*, 129(4), 1553–1623.

Clark, G. (2014): *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press.

Corrado, C., C. Hulten, and D. Sichel (2009): "Intangible Capital and US Economic Growth," *Review of Income and Wealth*, 55(3), 661–685.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006): "Interpreting the Evidence on Life Cycle Skill Formation," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch. Elsevier.

Dahl, G. B., and L. Lochner (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *American Economic Review*, 102(5), 1927–1956.

Fernandez, R., and J. Gali (1999): "To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints," *Review of Economic Studies*, 66(4), 799–824.

Findeisen, S., and D. Sachs (2016): "Education and Optimal Dynamic Taxation: The Role of Income-contingent Student Loans," *Journal of Public Economics*, 138, 1–21.

Galor, O. (2009): *Inequality and Economic Development: The Modern Perspective*. Edward Elgar Publishing.

Galor, O., and J. Zeira (1988): *Income Distribution and Investment in Human Capital: Macroeconomics Implications*. Hebrew University of Jerusalem, Department of Economics.

——— (1994): "Income Distribution and Macroeconomics," *Review of Economic Studies*, 60(1), 35–52.

Grochulski, B., and T. Piskorski (2010): "Risky Human Capital and Deferred Capital Income Taxation," *Journal of Economic Theory*, 145(3), 908–943.

Guell, M., J. Rodriguez-Mora, and C. Telmer (2015): "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating," *The Review of Economic Studies*, 82(2), 693–735.

Guner, N., A. Parkhomenko, and G. Ventura (2018): "Managers and Productivity Differences," *Review of Economic Dynamics*, 29, 256–282.

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic Implications of Size-dependent Policies," *Review of Economic Dynamics*, 11(4), 721–744.

Hall, B. H., A. Jaffe, and M. Trajtenberg (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," National Bureau of Economic Research WP 8498.

——— (2005): "Market Value and Patent Citations," *RAND Journal of Economics*, 36(1), 16–38.

Hall, B. H., and R. H. Ziedonis (2001): "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995," *RAND Journal of Economics*, 32(1), 101–128.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019): "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 87(5), 1439–1474.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

Jones, C. I. (2013): "Misallocation, Economic Growth, and Input-Output Economics," in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. II, pp. 419–456. Cambridge University Press.

Jovanovic, B. (2014): "Misallocation and Growth," *American Economic Review*, 104(4), 1149–1171.

Kapicka, M. (2015): "Optimal Mirrleesean Taxation in a Ben-Porath Economy," *American Economic Journal: Macroeconomics*, 7(2), 219–48.

Kapicka, M., and J. Neira (2019): "Optimal Taxation With Risky Human Capital," *American Economic Journal: Macroeconomics*, 11(4), 271–309.

Kaplow, L. (2005): "The Value of a Statistical Life and the Coefficient of Relative Risk Aversion," *Journal of Risk and Uncertainty*, 31(1), 23–34.

Knowles, J. (1999): "Parental Decisions and Equilibrium Inequality," Ph.D. thesis, University of Rochester.

Krueger, D., and A. Ludwig (2013): "Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous," *American Economic Review*, 103(3), 496–501.

Lee, S. Y., and A. Seshadri (2019): "On the Intergenerational Transmission of Economic Status," *Journal of Political Economy*, 127(2), 855–921.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, Z. Y. Amy, and L. Fleming (2014): "Disambiguation and Co-authorship Networks of the US Patent Inventor Database (1975–2010)," *Research Policy*, 43(6), 941–955.

Lucas, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

——— (2009): "Ideas and Growth," *Economica*, 76(301), 1–19.

Lucas, R. E., and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1), 1–51.

Maoz, Y. D., and O. Moav (1999): "Intergenerational Mobility and the Process of Development," *Economic Journal*, 109(458), 677–697.

OECD (2013): *Education at a Glance 2013: OECD Indicators*. OECD Publishing.

Olivetti, C., and M. D. Paserman (2015): "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930," *American Economic Review*, 105(8), 2695–2724.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11(4), 707–720.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), 71–102.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Harvard Business School WP.

Solon, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, pp. 1761–1800. Elsevier.

Spence, A. M. (1973): "Job Market Signalling," *Quarterly Journal of Economics*, 87(3), 355–374.

Stantcheva, S. (2017): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 125(6), 1931–1990.

Stiglitz, J. E. (1975): "The Theory of "Screening," Education, and the Distribution of Income," *American Economic Review*, 65(3), 283–300.

Word, D. L., C. D. Coleman, R. Nunziata, and R. Kominski (2008): "Demographic Aspects of Surnames from Census 2000," Discussion paper, U.S. Census Bureau.

Appendices

A Theory Appendix

A.1 Proof of Theorem 1:

The static profit maximization of a firm is stated as follows:

$$\Pi(z,\Theta) = \max_{k,l_u \geq 0} \{ z^{\zeta} k^{\kappa} l_u^{\lambda} - (r+\delta)k - w_u l_u \} \tag{23}$$

First order conditions imply $l_u^* = \frac{\lambda o^*}{w_u}$ and $k^* = \frac{\kappa o^*}{r+\delta}$, hence we have

$$\begin{aligned} o^* &= z^{\zeta} \left(\frac{\kappa o^*}{r+\delta} \right)^{\kappa} \left(\frac{\lambda o^*}{w_u} \right)^{\lambda} \\ o^* &= \left[\left(\frac{\kappa}{r+\delta} \right)^{\kappa} \left(\frac{\lambda}{w_u} \right)^{\lambda} \right]^{1/\zeta} z \end{aligned} \tag{24}$$

and the profits are simply equal to $\Pi(z,\Theta) = \zeta o^*$. From the unskilled labor market clearing condition, we get

$$\begin{aligned} L_u &= \int l_u^*(z) dZ(z) \\ L_u &= \frac{\lambda}{w_u} \left[\left(\frac{\kappa}{r+\delta} \right)^{\kappa} \left(\frac{\lambda}{w_u} \right)^{\lambda} \right]^{1/\zeta} \int z dZ(z) \\ \left(\frac{w_u}{\lambda} \right)^{\frac{\lambda+\zeta}{\zeta}} &= \left(\frac{\kappa}{r+\delta} \right)^{\kappa/\zeta} \frac{\bar{z}}{L_u} \\ w_u &= \lambda \left(\frac{\kappa}{r+\delta} \right)^{\frac{\kappa}{\lambda+\zeta}} L_u^{-\frac{\zeta}{\lambda+\zeta}} \bar{z}^{\frac{\zeta}{\lambda+\zeta}} \end{aligned} \tag{25}$$

where $\bar{z} \equiv \int z dZ(z)$. Note that since L_u is constant along the balanced growth path, the unskilled wage rate grows with gross rate $(1+g_z)^{\zeta/(\lambda+\zeta)}$. Similarly, the capital market clearing condition yields

$$\begin{aligned}
K &= \int k^*(z)dZ(z) \\
K &= \frac{\kappa}{r+\delta}\left[\left(\frac{\kappa}{r+\delta}\right)^{\kappa}\left(\frac{\lambda}{w_u}\right)^{\lambda}\right]^{1/\zeta}\int zdZ(z) \\
\left(\frac{r+\delta}{\kappa}\right)^{\frac{\kappa+\zeta}{\zeta}} &= \left(\frac{\lambda}{w_u}\right)^{\lambda/\zeta}\frac{\bar{z}}{K} \\
r &= \kappa\left(\frac{\lambda}{w_u}\right)^{\frac{\lambda}{\kappa+\zeta}} K^{-\frac{\zeta}{\kappa+\zeta}}\bar{z}^{\frac{\zeta}{\kappa+\zeta}} - \delta \qquad (26)
\end{aligned}$$

This time since w_u, K and $\bar{z}$ grow over time at gross rates $(1+g_z)^{\zeta/(\lambda+\zeta)}$, $(1+g_z)^{\zeta/(\lambda+\zeta)}$, and $(1+g_z)$ respectively, the interest rate will be constant along the balanced growth path. Define $\tilde{K} = K/\bar{z}^{\frac{\zeta}{\lambda+\zeta}}$ as relative aggregate capital stock. Combining equations (25) and (26) yields the simplified expressions:

$$\begin{aligned}
w_u &= \lambda\tilde{K}^{\kappa}L_u^{\lambda-1}\bar{z}^{\frac{\zeta}{\lambda+\zeta}} \\
r+\delta &= \kappa\tilde{K}^{\kappa-1}L_u^{\lambda}
\end{aligned}$$

Plugging the expressions for the unskilled wage rate and the interest rate into the profits yields

$$\begin{aligned}
\Pi(z,\Theta) &= \zeta\tilde{K}^{\kappa}L_u^{\lambda}\frac{z}{\bar{z}^{\lambda/(\lambda+\zeta)}} \\
&\equiv \pi\frac{z}{\bar{z}^{\lambda/(\lambda+\zeta)}} \qquad (27)
\end{aligned}$$

where π is a time invariant constant.

Define $\hat{z} \equiv z/\bar{z}^{\lambda/(\lambda+\zeta)}$, $\tilde{z} \equiv \bar{z}^{\zeta/(\lambda+\zeta)}$ and $\tilde{w}_s \equiv w_s/\tilde{z}$. The guess and verify method will be used to solve the value function of the firm in the innovation decision problem. Assume the value function of the firm has the form $V(z,\Theta) = v_1\hat{z} + v_2\tilde{z}$ where v_1 and v_2 are scalars. Plugging the solution into the problem, we get:

$$\begin{aligned}
V(z,\Theta) &= \max_{l_s\geq 0}\left\{\Pi(z,\Theta) + \frac{\chi l_s^{\xi}}{1+r}V(z+\gamma\bar{z},\Theta') + \frac{(1-\chi l_s^{\xi})}{1+r}V(z,\Theta') - w_s l_s\right\} \\
&= \pi\hat{z} + \frac{v_1\hat{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}} + \frac{v_2(1+g_z)^{\zeta/(\lambda+\zeta)}\tilde{z}}{(1+r)} \\
&\quad + \max_{l_s\geq 0}\left\{\frac{\chi l_s^{\xi}}{1+r}\frac{v_1\gamma}{(1+g_z)^{\lambda/(\zeta+\lambda)}} - \tilde{w}_s l_s\right\}\tilde{z}
\end{aligned} \qquad (28)$$

$$
\begin{aligned}
&= \pi\hat{z} + \frac{v_1\hat{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}} + \frac{v_2(1+g_z)^{\zeta/(\lambda+\zeta)}\tilde{z}}{(1+r)} \\
&\quad + \left(\frac{\xi}{\tilde{w}_s}\right)^{\frac{\xi}{1-\xi}} \left[\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} (1-\xi)\tilde{z} \\
&= v_1\hat{z} + v_2\tilde{z}
\end{aligned}
$$

where

$$
v_1 = \frac{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}-1}\pi \tag{29}
$$

$$
v_2 = \frac{(1+r)}{(1+r)-(1+g_z)^{\zeta/(\lambda+\zeta)}}\left[\left(\frac{\xi}{\tilde{w}_s}\right)^{\frac{\xi}{1-\xi}} \left[\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} (1-\xi)\right] \tag{30}
$$

$$
l_s^* = \left[\frac{\xi}{\tilde{w}_s}\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}}
$$

It is required to verify that w_s grows at gross rate $(1+g_z)^{\zeta/(\lambda+\zeta)}$. Without loss of generality, assume $\Lambda = 1$. Market clearing for skilled labor requires

$$
\begin{aligned}
L_s &= \int l_s^* dZ(z) \\
L_s &= \left[\frac{\xi}{w_s}\frac{\chi\gamma v_1\tilde{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} \\
w_s &= \frac{\xi\chi\gamma v_1\tilde{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}L_s^{1-\xi}}
\end{aligned}
$$

$$
w_s = \frac{\xi\chi\gamma\pi\tilde{z}}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}-1)L_s^{1-\xi}} \tag{31}
$$

proving the statement. Also notice that $L_s = l_s^*$. The aggregate productivity evolves according to

$$
\bar{z}' = \bar{z} + \gamma\chi L_s^{\xi}\bar{z}
$$

$$
\Rightarrow g_z = \Gamma L_s^{\xi} \tag{32}
$$

where $\Gamma \equiv \gamma\chi$. This concludes the proof.

A.2 Computational algorithm

Given closed-form solutions for the firm's maximization problem and the resulting system of non-linear equations in Theorem 1, the following computational algorithm is used to solve for the BGP equilibrium of the model:

1. Create grids for y_o, h, b, y_y, a.

2. Guess initial $V_o(y_o, h, a), W(b, h, a), V_y(y_y, a)$.
3. Guess initial $w_u, w_s, r, g, \bar{s}$.
4. Until convergence in value functions according to the sup-norm is achieved, do:
   (a) Solve:

$$\begin{aligned} V_o(y_o, h_y, a_y, \Theta) &= \max_{c_o, b \geq 0} \{u(c_0) + \alpha W(b, h_y, a_y, \Theta)\} \text{ s.t.} \\ c_o + b &\leq y_0 \end{aligned}$$

   Details: Single variable maximization where $b \in [0, y_o]$. One dimensional interpolation is required for evaluation.

   (b) Solve:

$$\begin{aligned} W(b, h_y, a_y, \Theta) &= \max_{n \geq 0} \{E\left[V_y(y_y, a_y, \Theta)|\cdot\right]\} \text{ s.t.} \\ y_y &= \left(w_{j_y} + \frac{w'_{j_y}}{1 + r'}\right) l_y(h_y, a_y) + b - c_n(n) \\ j_y &\sim F(j; l_y(h_y, a_y), n, \Theta) \end{aligned}$$

   Details: Single variable maximization where $n \in [o, \bar{n}]$, where $\bar{n}$ assures positive y_y in the worst case scenario. One dimensional interpolation is required for evaluation. Normal cumulative distribution function is required for calculations. Expectation is calculated over j realization.

   (c) Solve:

$$\begin{aligned} V_y(y_y, a_y, \Theta) &= \max_{c_y, c_c, h'_y, s \geq 0} \{u(c_y) + \alpha u(c_c) + \\ &\quad \beta E[V_o(y'_o, h'_y, a'_y, \Theta')|\cdot]\} \text{ s.t.} \\ y_y &\geq c_y + c_c + c_h(h'_y) + s \\ y'_o &= (1 + r')s \\ a'_y &\sim g(a_y) \\ \Theta' &= T(\Theta) \end{aligned}$$

   Two variable maximization where $s \in [0, y_y]$, $h \in [0, (y_y/\kappa_h)^{1/xi_h}]$, and resulting c_y, c_c must be positive. Two dimensional interpolation is required for evaluation. Expectation is calculated over a' realization.
5. Simulate to calculate capital, skilled and unskilled labor, and fraction of population in each job. One uniform and one normal draw are required for each household and period.

6. Update $w_u, w_s, r, g, \bar{s}$ using simulation results, and go back to (4) up until they are consistent with the market clearing equations and η.

A.3 Aggregate factor demand equations

For computational purposes, it is useful to characterize aggregate factor demands in terms of only factor prices, and factor prices only in terms of aggregate factor demands. This section derives these algebraically using the equations from Appendix A.

In a stationary equilibrium, the aggregate demand for skilled and unskilled labor, L_s and L_u, and the capital rental rate r are constants. The aggregate demand for capital, K, and the wage rates for skilled and unskilled labor, w_s and w_u, grow at the same rate as aggregate output, in proportion to $\tilde{z} = \bar{z}^{\zeta/(\lambda+\zeta)}$. Define normalized aggregate capital stock, skilled and unskilled wage rates as $\tilde{K} = K/\tilde{z}, \tilde{w}_s = w_s/\tilde{z}$ and $\tilde{w}_u = w_u/\tilde{z}$ respectively. First, notice that by only using the definition for $\tilde{w}_u$ and Equation 24, the following identity for π is obtained:

$$\pi = \zeta \left[\left(\frac{\kappa}{r+\delta} \right)^{\kappa} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\lambda} \right]^{1/\zeta} \tag{33}$$

Then we have:

$$\begin{aligned} L_u &= \left(\frac{\kappa}{r+\delta} \right)^{\kappa/\zeta} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\frac{\lambda+\zeta}{\zeta}} && (34) \\ \tilde{K} &= \left(\frac{\kappa}{r+\delta} \right)^{\frac{\kappa+\zeta}{\zeta}} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\frac{\lambda}{\zeta}} && (35) \\ L_s &= \left[\frac{\xi}{\tilde{w}_s} \frac{\chi\gamma\pi}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)} - 1)} \right]^{\frac{1}{1-\xi}} && (36) \end{aligned}$$

Given these equations, it can be verified that:

$$\pi = \zeta \tilde{K}^{\kappa} L_u^{\lambda} \tag{37}$$

Then we have:

$$\begin{aligned} \tilde{w}_u &= \lambda \tilde{K}^{\kappa} L_u^{\lambda-1} && (38) \\ r+\delta &= \kappa \tilde{K}^{\kappa-1} L_u^{\lambda} && (39) \\ \tilde{w}_s &= \frac{\xi\chi\gamma\pi}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)} - 1) L_s^{1-\xi}} && (40) \end{aligned}$$

B Empirical Appendix

TABLE 12: INTENSIVE MARGIN ROBUSTNESS - ALTERNATIVE MEASURES II

	log total patents (renewed once) (1975-2006)	log total patents (renewed twice) (1975-2006)	log total patents (top 5% only) (1975-2006)	log total patents (top 20% only) (1975-2006)
log income (1930)	.037***	.033***	.029***	.033***
	(.009)	(.009)	(.008)	(.009)
log education (1930)	.099***	.099***	.075***	.096***
	(.008)	(.008)	(.008)	(.008)
Obs.	78,438	78,438	81,348	81,348
R^2	0.05	0.05	0.03	0.04

NOTES: ROBUST STANDARD ERRORS IN PARENTHESES. DOMINANT RACE FIXED EFFECTS ARE INCLUDED THE COEFFICIENTS OF WHICH ARE SUPPRESSED FOR BREVITY. ALL VARIABLES ARE NORMALIZED BY SUBTRACTING THE MEAN AND DIVIDING BY THE STANDARD DEVIATION. OBSERVATIONS ARE WEIGHTED BY THE SHARE OF THE SURNAME IN THE GENERAL POPULATION OBTAINED FROM THE U.S. DECENNIAL CENSUS OF POPULATION (2000). *, ** AND *** DENOTE SIGNIFICANCE AT 10, 5 AND 1% LEVELS RESPECTIVELY.

TABLE 13: IMMIGRATION ROBUSTNESS (1880-1930) - EXTENSIVE MARGIN

	relative representation (1975-2008)					
	(1)	(2)	(3)	(4)	(5)	(6)
income (1930)	.296***	.304***	.323***	.223***	.307***	.293***
	(.014)	(.014)	(.015)	(.017)	(.017)	(.013)
education (1930)	.004	.008	−.003	.003	.021*	.005
	(.007)	(.009)	(.010)	(.005)	(.012)	(.007)
pop. share(1930)/pop. share(1880)						.053***
						(.032)
Obs.	64,308	48,282	48,289	32,168	32,159	64,308
R^2	0.35	0.33	0.38	0.46	0.30	0.35

NOTES: COLUMN 1 REPEATS THE LAST COLUMN OF TABLE 2 FOR SURNAMES WHICH POPULATION SHARE RATIO IS NOT MISSING. COLUMNS 2 AND 3 REPEAT THE SAME REGRESSION AFTER DROPPING THE TOP AND BOTTOM 25% OF THE SAMPLE ACCORDING TO POPULATION SHARE RATIO RESPECTIVELY. COLUMNS 4 AND 5 REPEAT THE SAME EXERCISE FOR THE TOP AND BOTTOM HALVES OF THE SAMPLE RESPECTIVELY. COLUMN 6 REPEATS THE SAME REGRESSION WITH THE WHOLE SAMPLE WHILE INTRODUCING THE POPULATION SHARE RATIO LINEARLY AS A REGRESSOR IN ADDITION TO INCOME AND EDUCATION. ROBUST STANDARD ERRORS IN PARENTHESES. DOMINANT RACE FIXED EFFECTS ARE INCLUDED THE COEFFICIENTS OF WHICH ARE SUPPRESSED FOR BREVITY. ALL VARIABLES ARE NORMALIZED BY SUBTRACTING THE MEAN AND DIVIDING BY THE STANDARD DEVIATION. OBSERVATIONS ARE WEIGHTED BY THE SHARE OF THE SURNAME IN THE GENERAL POPULATION OBTAINED FROM THE U.S. DECENNIAL CENSUS OF POPULATION (2000). *, ** AND *** DENOTE SIGNIFICANCE AT 10, 5 AND 1% LEVELS RESPECTIVELY.

Table 14: Immigration Robustness (1880-1930) - Intensive Margin

	log quality wtd. total patents (1975-2006)					
	(1)	(2)	(3)	(4)	(5)	(6)
log income (1930)	.022*	.009	.045***	.106***	−.005	.022*
	(.012)	(.012)	(.014)	(.019)	(.017)	(.012)
log education (1930)	.142***	.138***	.141***	.089***	.140***	.141***
	(.011)	(.011)	(.013)	(.018)	(.011)	(.011)
pop. share(1930)/pop. share(1880)						−.003
						(.005)
Obs.	50,529	37,921	37,897	25,269	25,265	64,308
R^2	0.05	0.05	0.05	0.04	0.04	0.35

Notes: Column 1 repeats the last column of Table 3 for surnames which population share ratio is not missing. Columns 2 and 3 repeat the same regression after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 4 and 5 repeat the same exercise for the top and bottom halves of the sample respectively. Column 6 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 15: Extensive Margin - Male Only

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.259***		.259***
	(.008)		(.008)
education (1930)		.024***	.000
		(.004)	(.003)
Obs.	107,613	107,613	107,613
R^2	0.24	0.18	0.24

Notes: Data is obtained exclusively from the males in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 16: Intensive Margin - Male Only

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	.072***		.006
	(.008)		(.009)
log education (1930)		.176***	.173***
		(.009)	(.010)
Obs.	76,265	76,265	76,265
R^2	0.02	0.05	0.05

Notes: Data is obtained exclusively from the males in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 17: Extensive Margin - Female Only

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.075***		.082***
	(.013)		(.015)
education (1930)		.024*	−.015
		(.013)	(.014)
Obs.	67,240	67,240	67,240
R^2	0.16	0.15	0.16

Notes: Data is obtained exclusively from the females in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 18: Intensive Margin - Female Only

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	$-.025$		$-.072^{***}$
	(.016)		(.019)
log education (1930)		$.061^{***}$	$.103^{***}$
		(.016)	(.019)
Obs.	16,117	16,117	16,117
R^2	0.02	0.02	0.02

Notes: Data is obtained exclusively from the females in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 19: Additional Robustness Checks - Extensive Margin

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	$.239^{***}$	$.230^{***}$	$.202^{***}$	$.179^{***}$
	(.010)	(.011)	(.010)	(.011)
education (1930)	.001	$-.001$	$.012^{**}$	.008
	(.005)	(.005)	(.005)	(.005)
household size (1930)	$.012^{**}$			.006
	(.006)			(.006)
literacy rate (1930)		$.049^{***}$		$.108^{***}$
		(.010)		(.012)
non-native origin (1930)			$.246^{***}$	$.267^{***}$
			(.013)	(.014)
Obs.	110,290	110,289	110,290	110,289
R^2	0.27	0.27	0.31	0.32

Notes: Household size and literacy rates are averages at the surname level. Non-native origin is the fraction of people with the surname whose parents (at least one) were of foreign origin. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 20: Additional Robustness Checks - Intensive Margin

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	.001	−.001	.013	.013
	(.009)	(.009)	(.010)	(.010)
log education (1930)	.175***	.174***	.162***	.162***
	(.009)	(.009)	(.009)	(.009)
household size (1930)	−.008			−.007
	(.006)			(.006)
literacy rate (1930)		.009		−.001
		(.011)		(.011)
non-native origin (1930)			−.052***	−.052***
			(.010)	(.010)
Obs.	81,348	81,347	81,348	81,347
R^2	0.05	0.05	0.05	0.05

Notes: Household size and literacy rates are averages at the surname level. Non-native origin is the fraction of people with the surname whose parents (at least one) were of foreign origin. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 21: Extensive Margin Regressions with Intensive Margin Subsample

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.269***		.269***
	(.012)		(.012)
education (1930)		.030***	−.003
		(.007)	(.006)
Obs.	81,348	81,348	81,348
R^2	0.30	0.24	0.30

Notes: The regressions on Table 2 are replicated while restricting the sample of surnames to those on Table 3. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 22: Rare Surnames - Intensive and Extensive Margins

	(unweighted)		(without frequent surnames)		(inverse standard dev. wtd.)	
	relative represent.	log quality wtd. total patents	relative represent.	log quality wtd. log patents	relative represent.	log quality wtd. total patents
income (1930)	.072***		.181***		.080***	
	(.004)		(.006)		(.006)	
education (1930)	.005		.000		.004	
	(.003)		(.003)		(.003)	
log income (1930)		.012***		.049***		.004
		(.004)		(.007)		(.008)
log education (1930)		.061***		.118***		.068***
		(.004)		(.018)		(.008)
Obs.	110,290	81,348	109,052	80,243	89,099	69,160
R^2	0.03	0.01	0.16	0.04	0.03	0.01

Notes: Columns 1 and 2 repeat the regressions on the last columns of Tables 2 and 3 without using any weights. Columns 3 and 4 repeat the same regressions in a reduced sample where surnames that are more frequent than the median surname are dropped. Columns 5 and 6 repeat the same regressions where the inverse of the standard deviation in income and education (geometric average) is used as weight. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 23: Extensive Margin - 1975-1995 Only

	relative representation (1975-1995)	relative representation (1975-1995)	relative representation (1975-1995)
income (1930)	.245***		.245***
	(.008)		(.008)
education (1930)		.028***	−.000
		(.005)	(.003)
Obs.	110,290	110,290	110,290
R^2	0.21	0.16	0.21

Notes: Inventor data is obtained exclusively from the 1975-1995 period. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 24: Intensive Margin - 1975-1995 Only

	log quality wtd. total patents (1975-1995)	log quality wtd. total patents (1975-1995)	log quality wtd. total patents (1975-1995)
log income (1930)	.060***		.006
	(.009)		(.010)
log education (1930)		.151***	.149***
		(.009)	(.009)
Obs.	70,032	70,032	70,032
R^2	0.03	0.04	0.04

Notes: Inventor data is obtained exclusively from the 1975-1995 period. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 25: Becoming an Inventor - Ethnicity 1

	relative representation	relative representation	relative representation
income (1930)	.234***		.233***
	(.010)		(.010)
education (1930)		.034***	.007
		(.008)	(.006)
Obs.	94,241	94,241	94,241
R^2	0.78	0.80	0.78

Notes: Robust standard errors in parentheses. Dominant race and ethnicity fixed effects are included the coefficients of which are suppressed for brevity. Surnames not matched to an ethnicity are not included. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 26: Inventor Productivity - Ethnicity 1

	log quality wtd. total patents	log quality wtd. total patents	log quality wtd. total patents
log income (1930)	.062***		.013
	(.009)		(.010)
log education (1930)		.134***	.129***
		(.009)	(.010)
Obs.	72,018	72,018	72,018
R^2	0.07	0.08	0.08

Notes: Robust standard errors in parentheses. Dominant race and ethnicity fixed effects are included the coefficients of which are suppressed for brevity. Surnames not matched to an ethnicity are not included. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 27: Becoming an Inventor - Ethnicity 2

	relative representation	relative representation	relative representation
income (1930)	.208***		.207***
	(.009)		(.009)
education (1930)		.030***	.007
		(.006)	(.005)
Obs.	110,290	110,290	110,290
R^2	0.81	0.83	0.81

Notes: Robust standard errors in parentheses. Dominant race and ethnicity fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 28: Inventor Productivity - Ethnicity 2

	log quality wtd. total patents	log quality wtd. total patents	log quality wtd. total patents
log income (1930)	.054***		.009
	(.008)		(.009)
log education (1930)		.123***	.119***
		(.008)	(.009)
Obs.	81,348	81,348	81,348
R^2	0.08	0.09	0.09

Notes: Robust standard errors in parentheses. Dominant race and ethnicity fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard deviation. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

C Quantitative Appendix

C.1 Relaxing the scarce inventor training assumption

The fraction of inventor training available in the society η is assumed to be exogenously fixed in the model. This means that only a fraction η of the population can receive the education necessary to produce ideas and become inventors. As a result, this assumption implies that the output growth rate of the economy can only be increased by allocating more productive individuals as inventors rather than increasing the share of inventors in the population. How would the counterfactual exercises look like if this assumption was relaxed?

In order to answer this question, the opposite extreme will be considered. Recall that the score threshold $\bar{s}_t$ was chosen such that

$$\eta = \int_{\bar{s}_t}^{\infty} \tilde{s} d\tilde{S}_t(\tilde{s})$$

held. Consider setting the fraction η free and fixing $\bar{s}_t$ instead. In this alternative specification $\bar{s}_t$ denotes a fixed achievement rating in score. Individuals who go have scores greater than this threshold get inventor training, and the rest do not. The calibration of this alternative model is trivial: The parameter η which was externally calibrated becomes an additional targeted moment, and $\bar{s}$ becomes a parameter.

Table 29 presents the results of repeating the credentialing spending shut down experiment executed in Section 5.2 under this alternative model specification. The changes are quite significant: Now that η is freely chosen, its value increases from 11.6% to 44.9%. This means nearly half of the population is now allocated to the innovation sector. As a result, the skilled labor supply is quadrupled, and the output growth rate is nearly doubled, increasing from 2.00% to 3.45%. As one would expect, the welfare gain from this increase is also calculated to be huge at 107%.

Naturally, these numbers are not to be taken seriously, since the specification does not impose any additional cost on the society for quadrupling the amount of inventor training provided. Rather, these should be viewed as the extreme upper bound on the growth and welfare numbers that could be achieved by relaxing the fixed η assumption. This example also serves to illustrate the fact that exogenously fixing η is a conservative assumption in terms of putting a discipline on the growth and welfare numbers produced by the model in the counterfactual experiments.

TABLE 29: SHUTTING DOWN THE CREDENTIALING SPENDING CHANNEL - FREE η

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.19	0.11	-42.1%
Extensive margin, education effect	0.07	0.09	28.6%
Intensive margin, income effect	0.08	0.16	100%
Intensive margin, education effect	0.22	0.13	-40.9%
Yearly GDP growth rate	2.00%	3.45%	72.3%
Education spending/GDP	8.55%	11.8%	37.9%
Aggregate skilled labor, L_s	0.48	2.09	332%
Aggregate unskilled labor, L_u	1.91	1.39	-27.3%
Mean innate ability of skilled workers, a	2.08	1.78	-14.5%
Mean early childhood education of skilled workers, h	2.27	3.45	51.9%
Mean parental wealth of skilled workers, y_o	0.87	0.96	9.68%
Mean bequests received of skilled workers, b	0.49	0.46	-5.37%
Wage income Gini index	0.52	0.50	-4.91%
Log 90/10 ratio	1.17	1.82	55.5%
Log 90/50 ratio	0.52	1.17	125%
Log 50/10 ratio	0.65	0.65	-0.06%