

NBER WORKING PAPER SERIES

DANCING WITH THE STARS:
INNOVATION THROUGH INTERACTIONS

Ufuk Akcigit
Santiago Caicedo
Ernest Miguelez
Stefanie Stantcheva
Valerio Sterzi

Working Paper 24466
<http://www.nber.org/papers/w24466>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018

Ernest Miguelez gratefully acknowledges the French National Research Agency (ANR) for financial support (reference: ANR-17-CE26-0016). Marta Prato, Raphael Raux, Yulia Zhestkova, and Piotr Zoch provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Ufuk Akcigit, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Dancing with the Stars: Innovation Through Interactions

Ufuk Akcigit, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi

NBER Working Paper No. 24466

March 2018

JEL No. H25,L16,O31,O33,O41

ABSTRACT

An inventor's own knowledge is a key input in the innovation process. This knowledge can be built by interacting with and learning from others. This paper uses a new large-scale panel dataset on European inventors matched to their employers and patents. We document key empirical facts on inventors' productivity over the life cycle, inventors' research teams, and interactions with other inventors. Among others, most patents are the result of collaborative work. Interactions with better inventors are very strongly correlated with higher subsequent productivity. These facts motivate the main ingredients of our new innovation-led endogenous growth model, in which innovations are produced by heterogeneous research teams of inventors using inventor knowledge. The evolution of an inventor's knowledge is explained through the lens of a diffusion model in which inventors can learn in two ways: By interacting with others at an endogenously chosen rate; and from an external, age-dependent source that captures alternative learning channels, such as learning-by-doing. Thus, our knowledge diffusion model nests inside the innovation-based endogenous growth model. We estimate the model, which fits the data very closely, and use it to perform several policy exercises, such as quantifying the large importance of interactions for growth, studying the effects of reducing interaction costs (e.g., through IT or infrastructure), and comparing the learning and innovation processes of different countries.

Ufuk Akcigit
Department of Economics
University of Chicago
1126 East 59th Street
Saieh Hall, Office 403
Chicago, IL 60637
and NBER
uakcigit@uchicago.edu

Santiago Caicedo
Department of Economics
University of Chicago
5757 South University Avenue
Chicago, IL 60637
scaicedosoler@uchicago.edu
and Universidad de los Andes

Ernest Miguelez
GREThA UMR CNRS 5113 -
Université de Bordeaux
Avenue Leon Duguit
F-33608 Pessac Cedex
France
and AQR-IREA, University of Barcelona
ernest.miguelez@u-bordeaux.fr

Stefanie Stantcheva
Department of Economics
Littauer Center 232
Harvard University
Cambridge, MA 02138
and NBER
sstantcheva@fas.harvard.edu

Valerio Sterzi
GREThA UMR CNRS 5113 -
Université de Bordeaux
Avenue Leon Duguit
F-33608 Pessac Cedex
France
valerio.sterzi@u-bordeaux.fr

1 Introduction

Only a handful of ideas turn out to be transformative. Thomas Edison, Alexander Bell, or Nikola Tesla – all are familiar names thanks to the transformative inventions they introduced into our lives. When it comes to how these ideas are actually generated, the R&D investments behind them are only the tip of the iceberg. The inventor himself – with his own knowledge and human capital – is a key input. An entire process of knowledge accumulation by an inventor underlies the production of these new ideas. Inventors build their knowledge over time by interacting with others and learning from them. Universities, companies, and other research institutions try to create as many opportunities as possible for such interactions, from the informal common coffee room chat to formal international conferences. When it then comes to the production stage of ideas, inventors form research teams of varying sizes and use their knowledge to create ideas of heterogeneous qualities. These ideas lead to improved technologies and products, and, ultimately, translate into economic growth.

These observations lead to the following questions which we attempt to answer in this paper: What is the right theoretical framework that embeds interactions and knowledge diffusion at the inventor level into an innovation-based growth model? How can we micro-found the production of innovations at the level of inventors and research teams in a way that is disciplined by the data? How can we capture the processes through which inventors further develop their knowledge and productivity, which then translate into new ideas or innovations? Empirically, based on data on inventors, their interactions, and their innovations, what is the quantitative importance of interactions with others relative to other learning channels for individual productivity and growth?

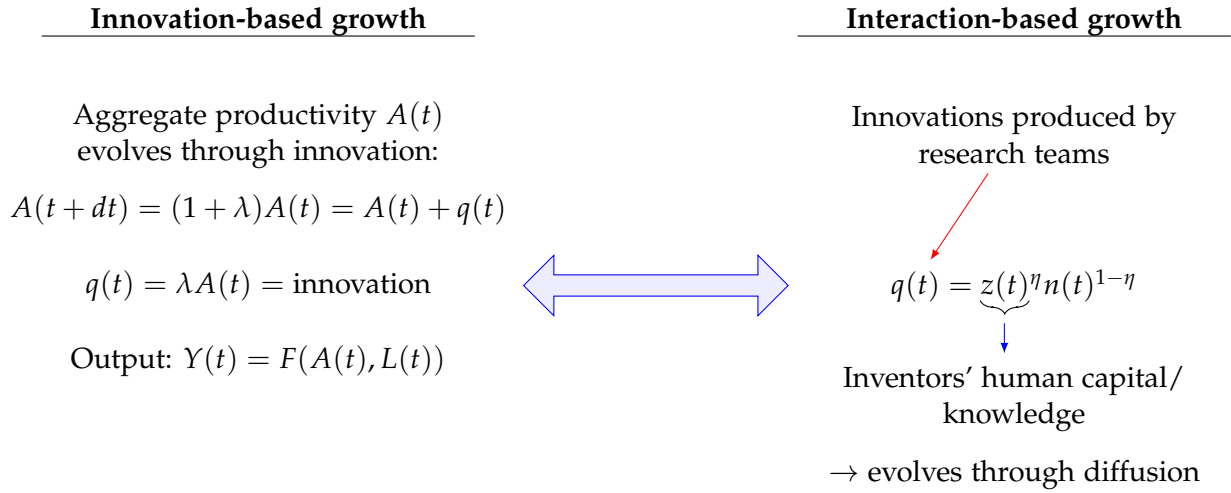
To answer these questions we bring together the recent knowledge diffusion models and innovation-based growth models. Empirically, we need detailed data on who interacts with whom, on their preceding and resulting productivity, their innovation and research teams. While such data is very challenging to find, we make use of brand-new data that allows us to bring empirics to what is thus far a mostly theoretical literature on diffusion, namely panel data of the universe of European inventors who patent and all European patents. Therefore, our contributions are both theoretical and empirical.

In our theoretical contribution, presented in Section 2, we build a new model with learning and human capital formation through interactions (i.e., knowledge diffusion) nested inside an innovation-based growth model.¹ The existing literature has focused mostly on the role of either innovation or knowledge and human capital accumulation as explanations for economic growth. Our analysis highlights the importance of individual inventors' human capital for innovation, therefore combining these two important forces for economic growth into a single framework.

Figure 1 provides a schematic illustration of our theory. Viewed through the lens of the innovation-based growth literature to the left, our model opens up the black box of ideas, which is represented by the “step size” of innovation λ and micro-founds this innovation production process at the research team levels, where individual researchers' heterogeneous productivities evolve endogenously

¹The knowledge diffusion model is inspired by Kortum (1997), Lucas (2009), Luttmer (2007), Lucas and Moll (2014), Perla and Tonetti (2014), and König, Lorenz, and Zilibotti (2016) among many others, while the innovation models are (Aghion and Howitt, 1992; Jones, 1995; Grossman and Helpman, 1991; Klette and Kortum, 2004; Aghion, Akcigit, and Howitt, 2014). We defer the detailed discussion of the related literature to Section 1.1.

FIGURE 1: COMBINING INTERACTION-BASED AND INNOVATION-BASED GROWTH MODELS



through interactions and knowledge diffusion. Viewed through the lens of the diffusion literature to the right, the endogenous interactions of researchers and knowledge diffusion contributes to their accumulated human capital (knowledge stock) or individual productivity z . Knowledge itself is not directly used to produce the output, but instead is first used inside research teams to create an idea or innovation. These in turn improve aggregate productivity through the step size. Knowledge is diffused among others through interactions; and more knowledge leads to the production of higher-quality innovations.

Our empirical contribution is to bring new data to a mostly theoretical literature. We leverage new large-scale micro data that allows us to give empirical content to our theory. It is difficult and rare to find a dataset that tracks many individuals, their productivity, and their interactions over time. We manage to do so thanks to the new panel data of European inventors with all European patents. Despite the fact that not all innovations are patented, our data carries information on millions of inventors who are responsible for a significant share of innovation around the world, which makes it very valuable for our purpose. A very recent special feature of our data is that inventors' names are uniquely disambiguated in order to turn the patent data into an inventor-level panel data. This major step allows us to observe and study the innovation and interaction histories of 826,878 inventors. In addition, inventors are carefully matched to their firms. There is also a compelling representation of numerous countries. In particular, we are able to quantify the productivity of individuals using the quality of their innovations over their life cycle. We can also measure interactions in several meaningful ways, such as past co-inventors, inventors in the same firm, and inventors in the same geographical region.

The data serves two major purposes in dialogue with our theory: First, it guides the key components of our theory and; second, it allows us to estimate the model and quantify the effects at play, and the impacts of changes. The model is disciplined by our main findings in the data – discussed in Section 3– which are used as building blocks:

1. Most patents are the result of collaborative work and produced by teams of heterogeneous sizes. The team size distribution is right-skewed.

2. Interactions among inventors are strongly correlated with higher subsequent inventor productivity (i.e., inventor innovation quality), even after including several detailed controls, such as inventor fixed effects. Inventors learn from each other to produce better innovations.
3. Interactions with inventors better than oneself are even more strongly correlated with higher subsequent productivity.
4. We observe a concave life-cycle profile of inventor productivity. This resembles the general earnings or wage life cycle profile documented by labor economists for other types of agents: First sharply increasing, then flattening. Through the lens of the model, inventors accumulate knowledge and become more productive over time, exactly according to this concave path.
5. Research productivity is positively associated with age, even after controlling for interactions with others and numerous controls. Thus, in our model, interactions should not be the only source of productivity improvements. We need to introduce an “external learning” channel, which serves as a catch-all for learning-by-doing, experience, or individual discovery.

Grounded in these empirical facts in our new data, our model can be viewed in more detail, but still at a single glance in Figure 2.

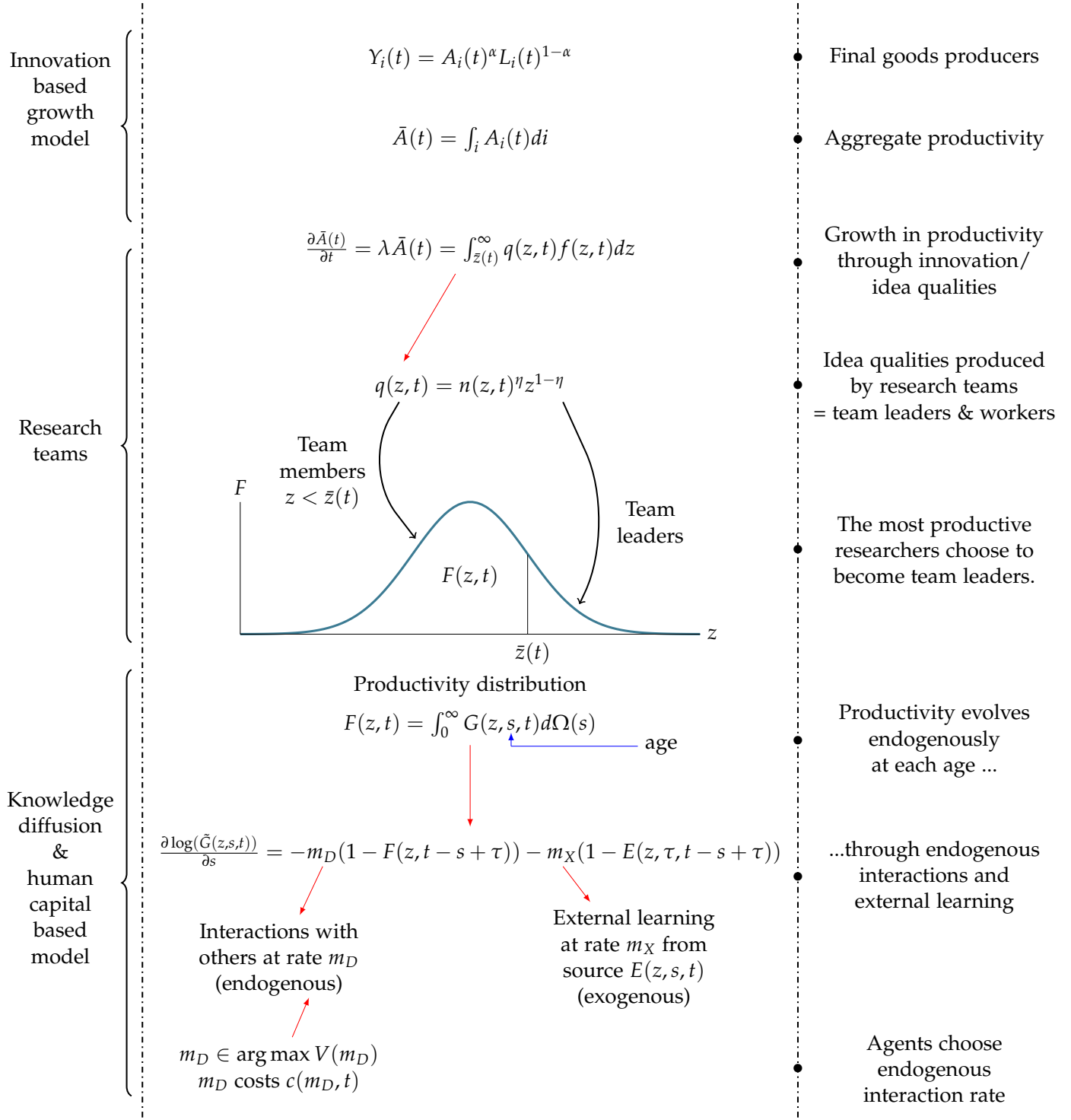
Building up from the individual inventor level (the bottom of the figure), inventors can learn, i.e., improve their productivity in two ways: They endogenously choose a meeting rate with others (interactions) at a search or effort cost, and they have access to an external, exogenous channel (which captures learning-by-doing, experience, individual discovery, etc.). In every period, given their realized productivity after learning, inventors form teams. To do so, they endogenously decide whether to become team leaders, who hire other inventors to their teams and reap the profits from innovation, or team members – earning the skilled researchers’ wage. This endogenous occupational choice induces a realistic team structure that mirrors the data and which will match the empirical team size distribution. Research teams produce innovations of heterogeneous qualities, increasing in the team leader’s productivity and in the team size. These innovations are sold to final good producers, who use them to increase their TFP and produce output. The growth in aggregate productivity, or TFP, every period will be the total innovation quality produced by research teams.

In our quantitative contribution, we map the theoretical model back to the data and estimate it via the simulated method of moments and indirect inference based on 86 distinct moments from the micro inventor data. As described in Section 4, we obtain an extremely close fit to the data – despite being highly over-identified – and are able to match very well some non-targeted moments.

We then put our estimated model to use and perform four quantitative policy exercises in Section 5: Shutting down interactions vs. external learning, reducing the search costs of meeting others, and reducing access to outside information (excessive agglomeration). We find that interactions with others are quantitatively very important for improving inventors’ productivity and, ultimately, for economic growth. In addition, there is a strong complementarity between access to external knowledge and learning from others: If others around oneself learn more from outside sources and then interact more, one will end up interacting with more knowledgeable people and learning more as well. We also estimate the model separately for the two major patenting countries: The U.S. and

Germany. We show that these two countries have very different learning environments, which map into different lifecycle evolutions of research productivity, thus also confirming findings on the wage lifecycle profiles in these two countries (Lagakos, Moll, Porzio, Qian, and Schoellman, 2017, 2018).

FIGURE 2: SUMMARY OF THE MODEL



1.1 Related Literature

On the theoretical side, our paper combines two strands of the economic growth literature: Innovation-based models, and technology or knowledge diffusion models.

The innovation-based endogenous growth models of [Aghion and Howitt \(1992\)](#), [Romer \(1990\)](#), [Jones \(1995\)](#), [Grossman and Helpman \(1991\)](#), [Klette and Kortum \(2004\)](#) view innovation as the result of costly investments. Innovation improves aggregate TFP.²

In the technology or knowledge diffusion literature, the focus is mostly on imitation and technology adoption, rather than on the production of innovation per se. “Innovation” in these models is most often modeled as a draw from an exogenous distribution (or a stochastic shock to productivity) in contrast to “imitation” or “technology adoption,” which are a draw from the endogenous technology distribution of firms (incumbents) or agents in the economy. [Kortum \(1997\)](#) builds a model of technological change where researchers sample ideas from a common distribution and pioneered the use of sources of ideas with fat tails that generate unbounded growth. [Lucas \(2009\)](#) introduces interactions between people who share ideas and a cohort structure. The starting distribution of productivities is unbounded and fat-tailed, so that even if agents only learn from each other, there can be sustained long-run growth.

More closely related to our work are the recent papers by [Lucas and Moll \(2014\)](#) and [Perla and Tonetti \(2014\)](#). [Lucas and Moll \(2014\)](#) allow for an endogenous resource allocation to production vs. interactions (i.e., imitation). They characterize the extent to which search and imitation externalities reduce welfare.³ [Staley \(2011\)](#) extends the Lucas model by allowing each agent’s productivity to experience stochastic shocks (termed “individual discovery”). [Perla and Tonetti \(2014\)](#) also feature a trade-off for firms between producing and paying a search cost to upgrade one’s technology through imitation, with the least productive firms choosing to imitate. Again, a thicker tailed productivity distribution leads to more growth (see also [Benhabib, Perla, and Tonetti \(2014\)](#)).

An interplay between innovation and technology diffusion is also explored also in [Luttmer \(2007, 2012, 2015\)](#). The second of these papers presents a growth model of imitation and selection where entrant and incumbent firms can imitate incumbents and there are Brownian shocks to productivity, which lead to firms remaining heterogeneous despite imitation.⁴ In the third paper, agents can make stochastic individual discoveries, yet also learn from others. Agents with more knowledge are teachers, while the others are students; higher ability students are assigned to more productive teachers. The resulting assortative matching will lead to significant differences in earnings.

Most tightly related are the very recent papers by [König, Lorenz, and Zilibotti \(2016\)](#) and [Benhabib, Perla, and Tonetti \(2017\)](#). [Benhabib, Perla, and Tonetti \(2017\)](#) propose a model that does not use an infinite support assumption, but which features an endogenously expanding finite technological frontier through innovation. Firms choose between innovating, adopting technology, or pro-

²See [Aghion, Akcigit, and Howitt \(2014\)](#) for a survey of the creative destruction literature and [Akcigit and Kerr \(2017\)](#) for a recent example.

³One of the focal points in the technology diffusion literature has been to what extent the models can endogenously generate unbounded growth. In these imitation models, since agents can only improve by imitating others around them, the productivity distribution would end up converging to a degenerate distribution with full mass at the most productive type. To avoid this, it is typically assumed that the initial productivity distributions have fat tails so that imitation (or interaction) opportunities never get exhausted.

⁴This is an alternative to the [Lucas and Moll \(2014\)](#) assumption that the stock of ideas is never exhausted.

ducing. In König, Lorenz, and Zilibotti (2016), firms can improve their heterogeneous productivities through two channels: Expanding resources on producing innovations – which generate stochastic productivity improvements, or imitating other firms. Imitating firms are randomly matched to others and may succeed in adopting (i.e., imitating) their technology. Firms choose whether to innovate or imitate depending on their position in the productivity distribution, with firms further away from the frontier choosing to imitate. They obtain convergence to long-run growth even from limited initial heterogeneity: Firms who are close to the frontier choose to innovate, which allows them to draw a new productivity that does not come from the existing pool of ideas.

Diffusion and innovation across countries, as a function of their distance to the frontier, is also explored in Acemoglu, Aghion, and Zilibotti (2006) and Benhabib, Perla, and Tonetti (2014).

In the trade literature, Alvarez, Buera, and Lucas (2017) and Buera and Oberfield (2016) develop semi-endogenous growth models, in which trade favors the international diffusion of technologies (see also Alvarez, Buera, and Lucas (2008)). Perla, Tonetti, and Waugh (2015) study the effects of opening up to trade on growth in a model where technology diffuses and heterogeneous firms can choose to adopt already existing technologies.

Our first main contribution is to bring empirical content to this class of models and is made possible by a new panel data on European inventors. The data both guides the main features of our model and enables us to quantify it. Regarding the theory contribution, note that in these papers, it is directly the technology used in production that is diffused. Technology is improved either through innovation (modeled as external draws or productivity shocks) and/or imitation (modeled as draws from the endogenous distribution). Instead, in our model the diffusion is of *knowledge* itself. Individual learning and discovery (external draws), and interactions (endogenous distribution draws) are two ways of diffusing knowledge. This is then used by research teams to produce ideas (“innovations”), which improve the technology of production and expand the finite technology frontier. Thus, the innovation-based growth models and diffusion models are truly linked and connected in our framework. We also model the micro-structure of the innovation production by introducing endogenously formed research teams as in the data, and a lifecycle of researchers.

On the empirical and quantitative side, we relate to several strands of the literature. The large and growing importance of teams for research has been documented in many empirical papers, recent examples of which are Wuchty, Jones, and Uzzi (2007), Jones, Wuchty, and Uzzi (2008), Azoulay, Fons-Rosen, and Zivin (2015), and Jaravel, Petkova, and Bell (2017). Jones (2009) in particular shows that “the burden of knowledge” over the past 50 years has increased and that, as a result specialization and teamwork have gained in importance. The relation between the life cycle stage (age) of inventors and research outcomes is also very important in the data, as documented by Jones (2010), who estimates shifts in life-cycle productivity of inventors and shows that they have become less productive at younger ages (see also Jones and Weinberg (2011)). We also are aligned with a copious empirical literature documenting technology spillovers (a prominent example of which is Bloom, Schankerman, and Van Reenen (2013)) by offering one possible explanation for why such spillovers may occur, namely, interactions between researchers. Our model can be used to study the effects of IT on the spread of knowledge and productivity as documented by Bloom, Sadun, and Reenen (2012) by modeling the latter as a reduction in the costs of interactions (see our analysis in

Section 5). Finally, our setting can also provide a microfoundation (see Section 5) for the heterogeneous productivity and life cycle profiles observed among skilled workers across different countries as documented in Lagakos, Moll, Porzio, Qian, and Schoellman (2017) and Lagakos, Moll, Porzio, Qian, and Schoellman (2018).

2 Model

In this section, we develop a model of innovation and growth through learning and interactions. We first lay out the model and then solve for the equilibrium on a balanced growth path. Because it facilitates comprehension of the main forces and patterns, we at times solve for intermediate steps and optimizations as we go.

2.1 Brief Model Overview

Figure 2 gives a brief schematic overview of the model’s structure. There are two sectors in the economy: A research sector that produces ideas (i.e., innovations) and a goods’ production sector whose output is the final consumption good. There are two types of workers: skilled research workers who produce innovations and unskilled labor used in production, with no endogenous transition between these two types.

We start (at the top of the figure) from a standard innovation-based growth model in which aggregate quality $\bar{A}(t)$ evolves according to

$$\frac{\partial \bar{A}(t)}{\partial t} = \lambda \bar{A}(t)$$

where λ is the so-called “step size.” This standard part is captured here by the goods production sector, in which final goods producers purchase the innovations (or ideas) from research teams and combine them with unskilled labor to produce the final good. Our contribution is to explicitly open up the black box of productivity growth and to model how the aggregate productivity improvement $\lambda \bar{A}(t)$ is produced through research teams in the research sector.

In the research sector, skilled workers of heterogeneous productivities decide whether to become team leaders or team workers; the most productive researchers become team leaders. Team leaders form research teams and hire team workers to produce an innovation or idea, the quality of which is a function of the team leader’s productivity and the team size.

Finally, the heterogeneous productivities of research workers evolve endogenously over time as a result of learning, through two channels: Endogenous interactions, i.e., by meeting others and acquiring their knowledge, and an exogenous external learning source.

In this way, we combine the knowledge-diffusion (or human capital-based models) with the innovation-based growth models. Researchers acquire knowledge and diffuse it by meeting others, and that knowledge is used to produce innovations which push the technological frontier.⁵

⁵Thus, researchers’ productivity and knowledge is different from the technology used in production of Lucas and Moll (2014) and the other papers described in Section 1.1. Here, productivity of researchers is research-related knowledge, which needs to go through the process of innovation by teams in order to be converted into technology, i.e., into an

2.2 Final Good Producers

There is a continuum of infinitely lived final good producers indexed by $i \in [0, 1]$. Their production function is:

$$y_i(t) = A_i(t)^\alpha L_i(t)^{1-\alpha}$$

where L_i is production (unskilled) labor and A_i is their total factor productivity (TFP), A_i . The total final good produced for consumption in the economy is:

$$Y(t) = \int_0^1 y_i(t) di.$$

A final good producer can improve his factor productivity $A_i(t)$ by purchasing innovations of varying qualities from the research teams. The process through which research teams produce ideas is described in detail below. A final good producer who purchases an innovation of quality q from research teams can improve his current factor productivity $A_i(t)$ by the increment q . This “market for ideas” can be modeled in many different ways without affecting our results (see below). Without loss of generality, the price of the final good is normalized to 1. The wage of unskilled production workers is $w_u(t)$, and the total supply of unskilled workers is 1.

At any instant t , a final good producer with existing factor productivity $A_i(t)$ chooses a level of unskilled labor $L_i(t)$ by solving the (static) maximization problem $\max_{L_i} \{A_i(t)^\alpha L_i^{1-\alpha} - w_u(t)L_i\}$. This yields the optimal unskilled labor input choice:

$$L_i(t) = \left[\frac{(1-\alpha)}{w_u(t)} \right]^{\frac{1}{\alpha}} A_i(t).$$

The flow profits of the final good producer are thus:

$$\Pi_i(t) = \left[\frac{(1-\alpha)}{w_u(t)} \right]^{\frac{1-\alpha}{\alpha}} \alpha A_i(t).$$

From the first order conditions and using the labor market clearing for unskilled labor for all t , $\int_0^1 L_i(t) di = 1$, we also obtain that:

$$w_u(t) = \bar{A}(t)^\alpha (1-\alpha); \quad L_i(t) = \frac{A_i(t)}{\bar{A}(t)} \quad y_i(t) = \frac{A_i(t)}{\bar{A}(t)^{1-\alpha}}$$

where

$$\bar{A}(t) = \int_0^1 A_i(t) di$$

is total aggregate factor productivity (or, TFP). Aggregating across all producers, we obtain the total final good produced as a function of aggregate quality.⁶

$$Y(t) = \bar{A}(t)^\alpha.$$

increase in aggregate productivity or TFP A .

⁶Note that if the level of unskilled labor were instead L , we would have $Y(t) = A(t)^\alpha L$.

The market for ideas on a balanced growth path

It will help to explain here how the market for ideas works, even though it requires anticipating some results presented later and assuming the economy is on a balanced growth path (which will be our focus throughout). The goal here is to show that it is irrelevant for aggregate productivity and growth in equilibrium, which final good producer purchases which team's innovation, and that the change in aggregate productivity will – under all surplus sharing rules – be equal to the sum of the qualities of the innovations produced by research teams.

We can show that, on a balanced growth in which aggregate productivity and the unskilled wage grow at a constant rates, the value of any final good producer will be linear in his productivity $A_i(t)$. More precisely, in Appendix A-1, we demonstrate that the value of a final good producer with productivity $A_i(t)$, on a balanced growth path on which the wage $w_u(t)$ grows at constant rate g_w and given interest rate r is:

$$V_i(t) = v \frac{A_i(t)}{\bar{A}(t)^{1-\alpha}} + \tilde{v} \bar{A}(t)^\alpha \quad \text{where} \quad v \equiv \frac{\alpha}{\left[r + \frac{1-\alpha}{\alpha} g_w\right]}, \quad \text{and} \quad \tilde{v} \text{ is constant.}$$

The surplus or change in value from buying an idea quality q from research teams is thus:

$$\Delta V_i(t) = \frac{v}{\bar{A}(t)^{1-\alpha}} q$$

This surplus has to be shared between the final good producer and the research team. In fact, the exact market structure in the market for ideas will be irrelevant. This is because the value function is linear in $A_i(t)$ for all i , so that the return to an additional unit of productivity is the same across all final goods producers and independent of their current level of factor productivity $A_i(t)$. Thus, regardless of which final good producer purchases which team's innovation – and of how the surplus is shared – aggregate productivity growth will be the same.⁷ We can for instance assume that final goods producers are randomly matched to research teams and that the research team receives a fraction β of the value generated. The price per unit of innovation quality is then simply: $p = \beta \frac{v}{\bar{A}(t)^{1-\alpha}}$. For all surplus sharing rules, the change in aggregate productivity equals the sum of innovation qualities produced by research teams. If research team j produces innovation quality q_j , then:

$$\frac{\partial \bar{A}(t)}{\partial t} = \int_i \frac{\partial A_i(t)}{\partial t} = \int_j q_j dj. \tag{1}$$

This completes the link between the research sector – which develops, then sells innovations— and the production sector – where those innovations are bought, then used to produce with higher factor productivity.

⁷Since there is a fixed number of researchers and no extensive margin switches between skilled and unskilled labor, the wage of skilled workers will scale with the price of innovation and the latter will not affect the optimal innovation intensity of research teams.

2.3 Research Teams

In the research sector, researchers produce innovations of varying quality. At a given time t all researchers have heterogeneous innovation productivities, $z(t) \in [0, \infty)$, distributed according to a cumulative density function $F(z, t)$.⁸ Variable $z(t)$ can be called human capital or researcher knowledge. Researchers endogenously group into *research teams*, which consist of a *team leader* and n *team members* (or team workers). Each researcher inelastically provides one unit of labor. A team leader with productivity z who hires n members produces an idea of quality q , where:

$$q = z^{1-\eta} n^\eta. \quad (2)$$

The quality of ideas thus increases in the team leader's productivity and in the number of members. We use $\eta \in [0, 1]$ to denote the team leader's span of control (as in Lucas (1978)). Let the wage of research team members at time t be denoted by $w(t)$. The price at which each unit of idea or innovation quality can be sold is $p(t)$.⁹ A team leader with productivity z chooses the number of members n that maximize his profits,

$$\max_{n \geq 0} p(t) z^{1-\eta} n^\eta - w(t) n.$$

The optimal number of members that a team leader with productivity z hires at time t , $n(z, t)$, the idea quality produced $q(z, t)$, and profits $\pi(z, t)$ are given by:

$$n(z, t) = \left(\frac{p(t)\eta}{w(t)} \right)^{\frac{1}{1-\eta}} z; \quad q(z, t) = \left(\frac{p(t)\eta}{w(t)} \right)^{\frac{\eta}{1-\eta}} z; \quad \pi(z, t) = p(t) \left(\frac{p(t)\eta}{w(t)} \right)^{\frac{\eta}{1-\eta}} (1-\eta)z. \quad (3)$$

More productive team leaders have more researchers in their team, produce better quality ideas, and make larger profits.

Occupational choice

An individual with productivity z can decide whether to become a team member or a team leader. If he chooses to be a member, he receives wage $w(t)$ for skilled workers in exchange for his unit of labor. On the other hand, if he chooses to be a team leader, he receives profits $\pi(z, t)$ from building a research team and producing ideas.

An individual of productivity z chooses to be a team leader if the profits he would make are larger than the wage he would receive as a team member, i.e., if

$$\pi(z, t) = \max\{w(t), \pi(z, t)\}.$$

Since $\pi(z, t)$ is an increasing function of z , and $\pi(0, t) = 0 < w$, there exists a cutoff $\bar{z}(t)$ such that any individual $z(t) > \bar{z}(t)$ chooses to be a team leader, and any individual $z(t) \leq \bar{z}(t)$ chooses to be

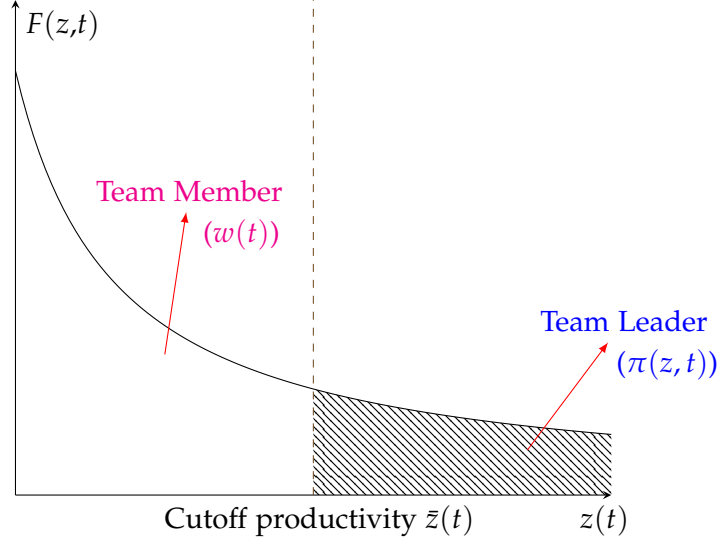
⁸The productivity distribution will be endogenized below.

⁹This price is determined in the final goods market according to some surplus sharing rule, i.e., some market structure, as described in Section 2.2.

a member, as illustrated in Figure 3. The relation between the wage and the cutoff is:

$$\frac{w(t)}{p(t)} = (1 - \eta)^{1-\eta} \eta^\eta \bar{z}(t)^{1-\eta}. \quad (4)$$

FIGURE 3: OCCUPATIONAL CHOICE: TEAM LEADERS AND TEAM MEMBERS



We can use the labor market clearing conditions to solve for the productivity cutoff $\bar{z}(t)$ by setting the labor supplied by team members to be equal to the labor demanded by team leaders:

$$F(\bar{z}(t), t) = \int_{\bar{z}(t)}^{\infty} n(z, t) f(z, t) dz,$$

or, substituting for the wage from equation (4):

$$(1 - \eta) \bar{z}(t) F(\bar{z}(t), t) = \eta \int_{\bar{z}(t)}^{\infty} z f(z, t) dz. \quad (5)$$

For any productivity distribution $F(\cdot, t)$ and span of control parameter $\eta \in (0, 1)$, there exists a cutoff $\bar{z}(t)$ that solves (5).¹⁰

There are two limit cases. When the team's product grows linearly in the number of members, i.e. $\eta \rightarrow 1$ so that there is perfect span of control for the team leader, the productivity cutoff $\bar{z} \rightarrow \infty$. This implies that only the most productive agent would be a team leader and hire everyone else as workers. Conversely, if $\eta \rightarrow 0$, the span of control becomes very restrictive and the productivity cutoff $\bar{z} \rightarrow 0$. Each team will then consist of only the team leader himself.

¹⁰This is because the left-hand side is equal to zero at $\bar{z}(t) = 0$ and is increasing in $\bar{z}(t)$, while the right-hand side is equal to the average productivity when $\bar{z}(t) = 0$, is decreasing in $\bar{z}(t)$, and converges to 0 as $\bar{z}(t) \rightarrow \infty$.

Team size distribution

The model produces a non-degenerate team size distribution, which will be one of the moments we will target in the data. Let $\mathcal{F}(n, t)$ denote the cumulative distribution function of team size. For all $n \geq \left(\frac{p(t)\eta}{w(t)}\right)^{\frac{1}{1-\eta}} \bar{z}(t)$, the team size distribution is related to the productivity distribution through:

$$\mathcal{F}(n, t) = Pr\{n(z, t) \leq n | z(t) \geq \bar{z}(t)\} = \frac{F\left(\left(\frac{w(t)}{p(t)\eta}\right)^{\frac{1}{1-\eta}} n, t\right) - F(\bar{z}(t), t)}{1 - F(\bar{z}(t), t)}.$$

2.4 Learning and the Evolution of the Productivity Distribution

We now describe how inventors “learn”, i.e., improve their productivity z over time. The cumulative distribution of productivity of an individual of age s born in period t_b is $\tilde{G}(z, s, t_b)$. When individuals are born, they draw from a non-degenerate initial distribution, $\tilde{G}(z, 0, t_b)$, $\forall z, t_b$. Individuals die at a rate δ , which implies that the age distribution is stationary, with cumulative probability distribution $\Omega(s) = 1 - e^{-\delta s}$ over $[0, \infty)$. Throughout their life, individuals can learn in two ways: through endogenous interactions with others and through external, individual sources.

External or individual learning: The external or individual learning channel captures the possibility of an individual improving his productivity even without meeting other individuals. At a general level, this is a catch-all term for all learning that occurs outside of interactions, such as learning-by-doing, learning through general work experience, reading, and individual discovery. This ingredient makes the model more realistic *per se*; in reality, individuals can learn on their own and in the data, productivity is strongly correlated with age, even after controlling for interactions (see Section 4). In addition, introducing a learning channel other than interactions with others is key for the quantitative analysis. It allows us to empirically quantify the importance of each of the two learning channels for growth, without mechanically loading all productivity improvements on the interaction channel. Put differently, we give the model the chance to show that interactions do not matter for productivity improvements.

Learning from the external source occurs through a Poisson process. With arrival rate m_X , an agent gets a productivity draw from an exogenous distribution of $E(z, s, t)$. For generality, the latter is allowed to depend on an agent’s age s , as well as on calendar time t ; how much an agent is able to learn from external sources can change over time and can be different for younger and older agents.

Endogenous learning through interactions: The second way in which individuals learn is by interacting with others. Meetings occur at a Poisson arrival rate (m_D), which is endogenous and will be chosen optimally by agents. When individual i meets individual j , he comes out of the meeting with the max productivity among them, i.e, after i meets j in the time interval Δt :¹¹

$$z_i(t + \Delta t) = \max\{z_i(t), z_j(t)\}.$$

¹¹As in Lucas and Moll (2014) and the other papers described in Section 1.1, these meetings are not assumed to be symmetric, although this is a feasible extension.

Evolution of Individual Productivity:

For a small time interval Δs , the probability that an individual born in period t_b and of age $s + \Delta s$ has productivity no greater than z , is equal to the probability that he starts from productivity no greater than z at time $t_b + s$ and, during the time interval Δs , either i) meets no other agent and gets no draw from the external learning source, or ii) meets someone, but with lower productivity than his, iii) or gets a draw from the external learning source that does not teach him anything (i.e., a draw lower than or equal to z). Formally:

$$\tilde{G}(z, s + \Delta s, t_b) = \tilde{G}(z, s, t_b) (1 - m_D \Delta s - m_X \Delta s + m_D \Delta s F(z, t_b + s) + m_X \Delta s E(z, s, t_b + s)) + o(\Delta s)$$

where $o(\Delta s)$ represents the second-order terms that vanish as $\Delta s \rightarrow 0$. Taking the limit as $\Delta s \rightarrow 0$ yields the law of motion for a cohort born at time t_b :

$$\frac{\partial \log(\tilde{G}(z, s, t_b))}{\partial s} = -m_D(1 - F(z, t_b + s)) - m_X(1 - E(z, s, t_b + s)). \quad (6)$$

We can use the change of variable $t = t_b + s$ to map age and birth data into calendar time. Let $G(z, s, t) := \tilde{G}(z, s, t - s)$ denote the cumulative probability distribution of an individual of age s at calendar time t . Taking the limit as $\Delta s \rightarrow 0$, integrating, and using the change of variable yields:

$$\log G(z, s, t) = -m_D \int_0^s (1 - F(z, t - s + \tau)) d\tau - m_X \int_0^s (1 - E(z, \tau, t - s + \tau)) d\tau + \log G(z, 0, t - s). \quad (7)$$

The cross-sectional age-conditional and unconditional productivity distributions are related through:

$$F(z, t) = \int_0^\infty G(z, s, t) d\Omega(s). \quad (8)$$

Choosing the endogenous meeting rate m_D :

Meeting others is costly and requires time and effort. At the beginning of their lives, individuals choose the rate at which they meet others, m_D optimally, weighting the cost and benefits of interactions. The cost of achieving a meeting rate m_D is given by $C(m_D, t)$ and can depend on calendar time. The exogenous interest rate is r . Hence, at time t , the expected payoff from choosing an interaction rate m_D of an individual of age s , across all possible realizations of productivity z , is:

$$W(m_D, s, t) = \int_0^{\bar{z}(t)} w(t) dG(z, s, t) + \int_{\bar{z}(t)}^\infty \pi(z, t) dG(z, s, t) - C(m_D, t), \quad (9)$$

and the individual born at time t_b chooses his meeting rate m_D so as to maximize his lifetime utility:

$$\begin{aligned} \max_{m_D} V(m_D, t_b) &= \int_0^\infty e^{-(r+\delta)s} W(m_D, s, t_b + s) ds, \\ &\text{subject to (7), (8), and (9).} \end{aligned} \quad (10)$$

Note that, in our model, interactions are distinct from working in teams. Put differently, the learning or knowledge diffusion stage is distinct from the idea production stage in teams; both occur each period. In the empirical analysis in Section 3, we thus provide several measures of

interactions, such as inventors in the same firm or region. This feature, which could be generalized in future work, is done to stick with the tradition of pairwise learning that has been prevalent in the literature, and which is much more tractable.

2.5 Balanced Growth Path

We now characterize the balanced growth path of the economy, on which the economy's growth rate is constant and all quantiles of the productivity distributions grow at the same rate g (following a "travelling wave" pattern).¹² Let de-trended productivity be denoted by $x = ze^{-gt}$. On the BGP, the age-conditional cumulative probability distribution is $\Gamma(x, s)$ and the cross-sectional distribution is $\Phi(x)$ (with density $\phi(x)$) such that, for all $z, s, t \geq 0$:

$$\Gamma(x, s) = G(z, s, t) \quad \text{and} \quad \Phi(x) = F(z, t), \quad \text{where} \quad x := ze^{-gt}.$$

Assumption 1. *On the BGP, the exogenous learning distribution $E(z, s, t)$ is invariant, i.e., there is a distribution $\Psi(x, s)$ such that, for all t, s , and z , $\Psi(x, s) = E(z, s, t)$ with $x = ze^{-gt}$.*

The de-trended cutoff $\bar{x} := \bar{z}e^{-gt}$ is constant along the BGP, since all quantiles of the productivity distributions grow at the same rate g . The labor market clearing real wage is given by:

$$\frac{w(t)}{p(t)} = \frac{w_0}{p_0} e^{g(1-\eta)t} \quad \text{with} \quad \frac{w_0}{p_0} = (1-\eta)^{1-\eta} \eta^\eta \bar{x}^{1-\eta}; \quad (11)$$

and the unique de-trended cutoff by:

$$(1-\eta)\bar{x}\Phi(\bar{x}) = \eta \int_{\bar{x}}^{\infty} x\phi(x)dx. \quad (12)$$

We see from equation (3) and the fact that individual productivity grows at rate g that idea quality grows at rate $(1-\eta)g$. It follows immediately that the growth rate of aggregate productivity is $\frac{\partial \bar{A}(t)}{\partial t} \frac{1}{\bar{A}(t)} = g(1-\eta)$ and the growth rate of aggregate output is thus $\frac{\partial Y(t)}{\partial t} \frac{1}{Y(t)} = \alpha g(1-\eta)$.

Assumption 2. *On the BGP, the real cost of meetings is proportional to aggregate productivity $A(t)$ (which grows at rate $g(1-\eta)$). Thus, on the BGP, the nominal cost can be written as: $C(m_D, t) = c(m_D)e^{g(1-\eta)t}p(t)$. We further assume that $c(m_D) := \frac{\kappa}{2}m_D^2$, $\kappa > 0$.*

The cost of meetings is expressed in units of the final good (the numeraire) like all other prices. The analysis does not depend on the parametric specification, but we use this particular one to illustrate the results and to estimate the model later on.

Making use of the characteristics of the BGP, we can rewrite the system of functional equations

¹²As has been discussed before (Lucas and Moll, 2014), there are no general existence or uniqueness theorems for partial differential equations systems, and our analysis does not focus on that, nor on computing the solutions to all possible initial distributions. Instead, we focus on the analysis of a balanced growth path.

in terms of de-trended productivities:

$$\log \Gamma(x, s) = -m_D \int_0^s \left(1 - \Phi(xe^{g(s-\tau)})\right) d\tau - m_X \int_0^s \left(1 - \Psi(xe^{g(s-\tau)}, \tau)\right) d\tau + \log \Gamma(xe^{gs}, 0) \quad (13)$$

$$\Phi(x) = \int_0^\infty \Gamma(x, s) d\Omega(s). \quad (14)$$

On the BGP, the expected payoff at time t of an individual of age s can be rewritten as:

$$\begin{aligned} W(m_D, s, t) &= \int_0^{\bar{z}(t)} w(t) dG(z, s, t) + \int_{\bar{z}(t)}^\infty \pi(z, t) dG(z, s, t) - c(m_D) e^{g(1-\eta)t} p(t) \\ &= e^{g(1-\eta)t} p(t) \left[\int_0^{\bar{x}} w_0 d\Gamma(x, s) + \int_{\bar{x}}^\infty \pi_0(x) d\Gamma(x, s) - c(m_D) \right]. \end{aligned}$$

Since $p(t)$ also grows at a constant rate, this implies that an agent's optimal choice of m_D will not depend on calendar time t , as it should only maximize the term inside the square brackets. As a result, the choice $m_D(t_b)$ will be constant for all birth dates t_b . Without loss of generality, using $t_b = 0$, the optimal choice of m_D maximizes the individuals' expected discounted value of their life time earnings net of the searching cost:

$$\begin{aligned} \max_{m_D \in [0, \infty)} V(m_D) &\propto \int_0^\infty e^{-(r+\delta)s} \left(\int_0^{\bar{x}} \frac{w_0}{p_0} d\Gamma(x, s) + \int_{\bar{x}}^\infty \pi_0(x) d\Gamma(x, s) - c(m_D) \right) ds \\ &\text{subject to (13) and (14)} \end{aligned} \quad (15)$$

where $\pi_0(x) = \left(\frac{p_0 \eta}{w_0}\right)^{\frac{\eta}{1-\eta}} (1-\eta)x$.

Let $Z(t)$ denote average productivity at time t . Letting i index skilled researchers, we can write the proportional change in average productivity in a small time interval Δt , $\frac{Z(t+\Delta t) - Z(t)}{Z(t)\Delta t}$ and obtain the growth rate of productivity on the BGP by letting Δt go to zero, i.e.:

$$\begin{aligned} g &= \lim_{\Delta t \rightarrow 0} \frac{Z(t+\Delta t) - Z(t)}{Z(t)\Delta t} \\ &= \left[\begin{aligned} &m_d \times \left[\frac{1}{Z(t)} [z_i F(z_i, t) + [1 - F(z_i, t)] \mathbb{E}_F(z'_i(t) | z'_i(t) > z_i)] - 1 \right] \\ &+ m_x \times \left[\frac{1}{Z(t)} [z_i F(z_i, t) + [1 - F(z_i, t)] \mathbb{E}_E(z'_i(t) | z'_i(t) > z_i)] - 1 \right] \\ &+ \delta \times \left[\frac{1}{Z(t)} \mathbb{E}_0(z'_i(t)) - 1 \right] \end{aligned} \right] \quad (16) \end{aligned}$$

where \mathbb{E}_F , \mathbb{E}_E , and \mathbb{E}_0 denote the expectations relative to the, respectively, cross-sectional, external, and age zero distributions.

We now summarize these characteristics of the BGP in the following definition.

Definition 1. Balanced Growth Path. A balanced growth path (BGP) consists of a constant g , a set of invariant distributions $(\Phi(z), \Gamma(z, s))$, paths for the real wage $\frac{w(t)}{p(t)}$, team sizes $n(z, t)$, idea qualities $q(z, t)$, profits $\pi(z, t)$, a cutoff for productivity \bar{x} , and an interaction rate m_D , such that:

(i) All quantiles of the invariant distributions grow at the same rate g , i.e.:

$$\Gamma(x, s) = G(z, s, t), \quad \forall z, t, s \geq 0$$

$\Phi(x) = F(z, t)$, $\forall z, t, s \geq 0$, with density $\phi(x)$.

(ii) $\Gamma(x, s)$ and $\Phi(x)$ satisfy the system of functional equations composed of (13) and (14) for all (x, s) .

(iii) The growth rate satisfies (16).

(iv) The real wage $\frac{w(t)}{p(t)}$ that clears the labor market for researchers satisfies (11).

(v) Profits $\pi(z, t)$ (as well as team size $n(z, t)$ and innovation qualities $q(z, t)$) are given by (3).

(vi) The cutoff \bar{x} is solution to (12).

(vii) The meeting rate m_D solves (15).

(viii) Aggregate productivity $\bar{A}(t)$ grows at rate $(1 - \eta)g$.

Aggregate output is $Y(t) = \bar{A}(t)^\alpha$ and grows at rate $\alpha g(1 - \eta)$.

We can, in addition, derive a useful relation between the economy's growth rate, the rate of endogenous interactions m_D , and the external learning rate m_X . To do so, we assume that, on the BGP, the external and the age zero productivity distributions have a common tail parameter, θ :

Assumption 3. Assume that $\Psi(x, s)$ and $\Gamma(x, 0)$ have a Pareto tail, i.e.

$$\lim_{x \rightarrow \infty} \frac{1 - \Psi(x, s)}{x^{-1/\theta}} = \rho(s), \quad \lim_{x \rightarrow \infty} \frac{1 - \Gamma(x, 0)}{x^{-1/\theta}} = k_0$$

for some $\theta, k_0 > 0$, and $\rho(s) > 0 \forall s$.

Note that the external source distribution has a location parameter $\rho(s)$ that depends on age.

We can then show that the BGP cross-sectional productivity distribution $\Phi(x)$ also has a Pareto tail and derive an expression for the growth rate g .

Proposition 1. Along a BGP, the cross-sectional productivity distribution $\Phi(x)$ has a Pareto tail, i.e.:

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1/\theta}} = k > 0$$

and the following relation holds:

$$\frac{g}{\theta} = \frac{m_D \int_0^\infty \left(1 - e^{-\frac{g}{\theta}s}\right) d\Omega(s) + m_X \frac{g}{k\theta} \int_0^\infty e^{-\frac{g}{\theta}s} \int_0^s \rho(\tau) e^{\frac{g}{\theta}\tau} d\tau d\Omega(s)}{1 - \frac{k_0}{k} \int_0^\infty e^{-\frac{g}{\theta}s} d\Omega(s)}. \quad (17)$$

Proof. See Appendix A-1. □

All else equal, this relation shows that the growth rate is increasing in the rate of interactions m_D , the rate of external learning m_X , the scale of the productivity distribution at birth relative to the cross-sectional distribution, k_0/k , and the thickness of the productivity distributions' tails as captured by θ . It is decreasing in the death hazard rate δ .

Let us consider a few key special cases. If the location parameter of the external distribution does not depend on age, i.e., $\rho(s) = \rho$ for all s , the growth rate is:

$$g = \left[(m_D + \frac{\rho}{k} m_X) - \delta \left(1 - \frac{k_0}{k} \right) \right] \theta.$$

In addition, if the scale parameters of the external distribution and the age zero distributions are the same, $\rho(s) = k_0 = k$ for all s , the growth rate depends only on the sum of the rates of interaction and external learning, and on the tail parameter:

$$g = (m_X + m_D)\theta.$$

If the external distribution Ψ were not growing (or if it were growing at a rate lower than g), the growth rate would simply be determined by the tail parameter and the endogenous interaction rate:^{13,14}

$$g = m_D\theta.$$

When is there positive growth? A major focal point in the diffusion literature is whether sustained long-run growth can occur. It is important to note that our model is rich enough to allow for many different cases, with positive or zero growth. This is why we start from the most general case and cover several special cases above. Ultimately, the model will be disciplined empirically in Section 4. For the sake of completeness, we draw out all the possible cases that can arise with respect to the growth rate here:

(i) If either one of the initial productivity distribution $F(z, 0) = \Phi(z)$ or the external distribution $\Psi(z, s)$ has a fat tail, i.e., if $\theta > 0$, there is positive growth.¹⁵

(ii) If neither of these distributions has a fat tail ($\theta = 0$), and the external distribution is growing at a rate strictly below g or not growing at all, then there is no sustained growth in the long-run.

(iii) If the external distribution is also growing at the endogenous rate g , growth can be sustained regardless of the shape of the distributions Φ and Ψ .

3 Data, Measurement and Empirical Findings

In this section, we give empirical content to the theory presented in Section 2. One of our main contributions is to leverage a new large-scale micro dataset – the European Patent Office data – that allows us to give concrete empirical counterparts to the concepts of innovations, research teams, interactions, and productivity of the model.

We first describe the data and empirically document some key facts about researchers and their teams, which were already used to inform the key elements of the model in Section 2. These facts

¹³Although our model does not nest and is not nested in Lucas and Moll (2014), their growth rate most closely corresponds to our special case with $k_0 = k$ and no external ideas source or an external idea source that grows at rate $g_e < g$ strictly smaller than g . In this special case of our model, $g = m_D\theta$, as in Lucas and Moll (2014). This same growth rate also arises in an extension in Lucas and Moll (2014), in which there is a fixed (not growing) source of external learning. The rate of arrival of external ideas m_X is irrelevant, justifying the authors' claim that it is really interactions that drive growth. This is also true in our model as shown here, in the case where the external distribution is not growing or is growing at rate $g_e < g$. We allow for $g_e = g$ here, to ensure a fair quantitative comparison between the external learning source and the endogenous interactions. If we did not do this, we would always conclude mechanically that in the long-run, the external distribution loses its importance.

¹⁴Note also that, as in Lucas and Moll (2014), if the external distribution has a fat tail with tail parameter $\zeta > \theta$, then the growth rate would be $g = m_D\zeta$ (even if the external distribution were not growing at rate g).

¹⁵This is true even if the external distribution is not growing or growing at a rate smaller than g .

also will be used as the main data moments to estimate the model in Section 4. To not repeat the figures twice, we will point the reader directly to each relevant figure highlighting a significant data moment in the estimation Section 4. One of the core facts of our model which we study in this section is the link between interactions and productivity improvements.

3.1 The European Patent Office Data

There are many patent offices in the world. Typically, academic papers on productivity have been based on patents filed with the United States Patent and Trademark Office (USPTO) (made available first by Hall et al. (2001)). Instead, our data is derived from the European Patent Office (EPO) data – an under-explored and novel source.

Patent data is ideal for studying the effect of interactions on learning and innovation. First, patent documents contain a multitude of information on the patent assignee (the original owner of the patent, which can be a firm, an individual, or a university), all the inventors who contributed to the innovation, and the innovation itself. Second, it allows us to track individual inventors over time. To do so, the first step is to disambiguate inventor names, in order to turn the list of inventors into an inventor level panel data. Disambiguation consists of identifying two or more inventors listed on several patents as the same person. This is based on their homonymy or quasi-homonymy (identity or similarity of names and surnames), together with related information (e.g., residential address, co-inventors, citations, or technological classes). We describe the disambiguation algorithm in Appendix OA-1. Because the EPO data was only very recently disambiguated, we have improved upon the disambiguation over time.¹⁶

A fundamental reason to use the EPO data is that it provides better representation of different countries. This mitigates the “home bias” in patenting. Indeed, a patent only protects an innovation within the geographical jurisdiction of the patent office where it is filed. The USPTO would thus be preferred by U.S. firms (which can be seen in the large representation of U.S. firms’ patents in the USPTO data). A patent in the EPO can come from any country in Europe, as well as from other patenting countries interested in exporting to the European market. Figure A-1 panel (A) depicts the number of EPO applications per country for the major patenting countries. The U.S. is the largest contributor to innovations at the EPO, followed by Germany, Japan, other European countries (France, Great Britain, Italy, the Netherlands, and Switzerland), and South Korea.

Overall, our data contains 2,955,055 applications. There are fewer patent applications at the EPO than at the USPTO (see Table OA-1 for a summary of the number of worldwide applications, as well as the number of applications to the EPO and the USPTO per year). Panel (B) of Figure A-1 also shows that the number of EPO patents has steadily increased over the years.

The disambiguated EPO data contains 3,474,514 unique inventors, of which 1,145,185 (32.96%) are listed in two or more applications.¹⁷ The average number of patents per inventor is 2.2. However, patent production, like other scientific or academic endeavors, is a very skewed phenomenon.

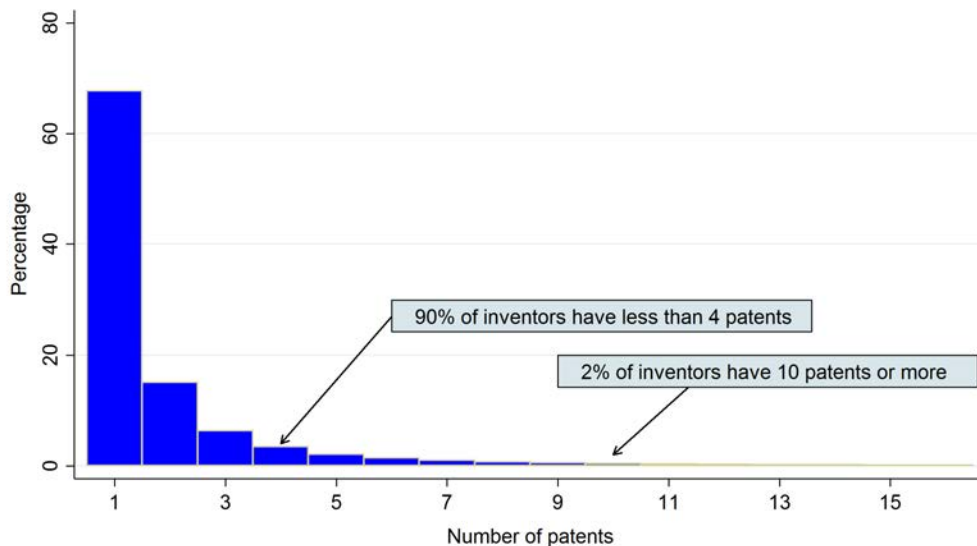
¹⁶The raw data used is the CRIOS-PatStat dataset described in Coffano and Tarasconi (2014). See also Breschi et al. (2016) and Akcigit et al. (2016).

¹⁷This is very comparable to the USPTO disambiguation by Li et al. (2014), which identifies 3,124,041 unique inventors, out of which 37.39% (1,168,208) have filed more than one patent.

A substantial proportion of inventors has only one patent and very few super-productive inventors have many patents. Figure 4 shows the distribution of inventors with no more than 20 patent applications (99.33% of the sample). The majority (around 91% of the total) produce 1 to 4 patents during their career. Only 80,000 (2% of the sample) of inventors have more than 10 patent applications.¹⁸

We assign patent applications to specific years according to their “priority year,” i.e., the first year of application to any patent office worldwide (thus, the priority date may differ from the application date at the EPO). This procedure ensures that the patent is assigned to the year closest to the actual innovation. We assign patents to origin countries using the inventors’ country of residence.¹⁹ Finally, each patent is also assigned a technology field. We adopt the OST (2008) classification of IPC codes into 30 fields. Online Appendix OA-1 provides more information on IPC codes and their technology classification, as well as summary statistics on the distribution of patents across fields in the EPO data.

FIGURE 4: NUMBER OF PATENTS PER INVENTOR AT THE EUROPEAN PATENT OFFICE



Notes: The figure shows the percent of inventors with a given number of patents in their lifetime. Sample contains 3,474,514 unique inventors from patents filed with the EPO between 1977 and 2010.

¹⁸Figure OA-1 shows that there is some variation in the number of patents per inventor across the top 10 patenting countries, ranging from 1.5 per inventor in South Korea to 3.1 in Germany. These differences reflect the differential propensities to work in teams, as well as the propensity to file with the EPO.

¹⁹There are different methods for assigning patents to countries. Traditionally, the applicant’s address reported in the patent document has been used (usually, the firm’s address). However, this approach can give a biased measure of where the actual innovation was developed, since many firms may assign their patents to their headquarters rather than to the branch where the innovation was produced. A more meaningful alternative approach – which we use – is to use the addresses of the inventors listed in the application. When inventors reside simultaneously in several countries – a relatively rare case – we can assign the patent probabilistically to each country.

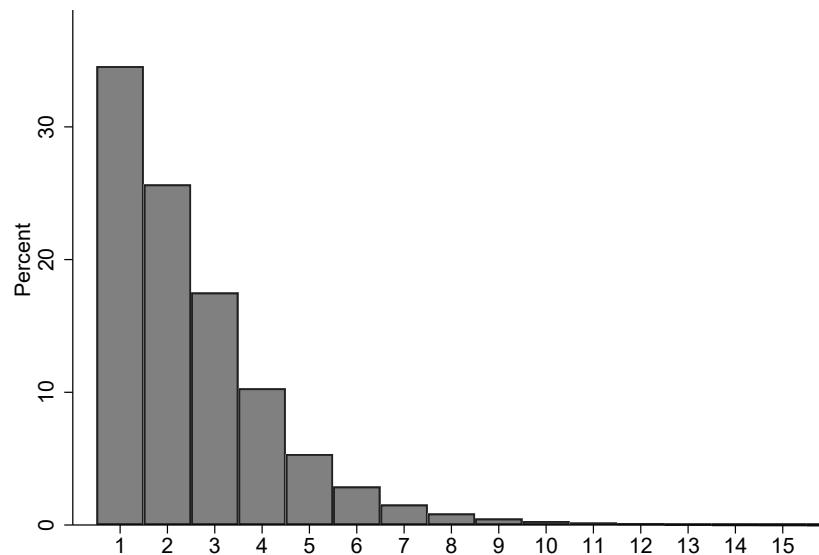
3.2 Measuring Productivity, Interactions, and Team Composition

Our empirical analysis sheds light on the key channel highlighted by the model, namely on how interactions with others contribute to human capital accumulation and enhanced productivity. We now map each of the model's concepts of teams, team leaders, individual productivity, idea quality and interactions to the data. In doing so, we remain as true as possible to the spirit of the model. It is worth noting outright that the data is very rich and that there are several meaningful ways of measuring these variables. In each case, we provide several possible robustness checks using alternative definitions.

Inventor teams

As in the model, research teams play a key role in the data. The patent applications at the EPO reflect the increasing importance of teams in patent production. In the time around 2010, more than 70% of patents were produced by teams with at least two inventors. Figure 5 depicts the distribution of team size and highlights that teams of one, two and three inventors account for the majority of patents. The average team size is 2.6 inventors; conditioning on multi-inventor patents, the average team size is 3.4 inventors.

FIGURE 5: TEAM SIZE DISTRIBUTION



Notes: The figure shows the team size distribution, i.e., the number of co-inventors per patent in the EPO data, based on 3,474,514 unique inventors from patents filed between 1977 and 2010.

Idea quality

To measure the quality of the idea produced by a team, we use the quality of patents produced by the team. As is standard in the literature (see [Hall et al. \(2001\)](#); [Pakes \(1986\)](#); [Schankerman and Pakes \(1986\)](#); [Trajtenberg \(1990\)](#)), patent quality is measured by the number of forward citations. We count the citations a given patent receives in a time window of three years after the priority date. We use

the three year window to account for truncation in the citations, i.e., the fact that more recent patents have had less time to accumulate citations, as described in Hall et al. (2001). Thus, our benchmark measure of the idea quality of team j at time t , $q_{j,t}$, is the citations of the patent(s) produced by team j at time t :²⁰

$$q_{j,t} = \sum_{\tau=t}^{t+2} \text{citations}_{\tau}.$$

An improvement we make over earlier work is that we consider forward citations coming from patents filed with the EPO, as well as all patents from a family in which one patent is filed with the EPO.²¹ As a robustness check, we also provide results using a window of five years, as well as removing self-citations (see Table 3).

Panel (B) of Figure 6 shows the idea quality q of teams as a function of the team leader’s age, i.e., the life cycle of idea qualities of team leaders. This life cycle profile will be one of the moments targeted in the estimation.

Individual productivity measures

To map the individual inventor’s productivity in the model to its counterpart in the data at each time, we compute a set of different measures. These capture slightly different notions of productivity and allow us to check that our results are robust to various inventor productivity definitions. One natural measure of current (or flow) productivity of inventor i in year t , denoted by $p_{i,t}$, is the number of citations-weighted patents produced by inventor i at time t :

$$p_{i,t} = \sum_j q_{j,t}.$$

This accounts for both innovation quantity and quality produced by an inventor. The cumulative (or stock) productivity of inventor i at time t is his citations-weighted patent stock to date, $P_{i,t}$:

$$P_{i,t} = \sum_{s=t_0}^t p_{i,s}.$$

This stock measure takes into account the quantity and quality of an inventor’s past innovations, in addition to current ones. This is our benchmark measure of an individual inventor’s productivity.

Additional quality measures used in the analysis – some reported as robustness checks, others available upon demand – are the unweighted number of patents produced at time t (which captures only the quantity of innovation, regardless of quality), as well as average citations per patent (received within a 3-year window) of the patents produced at time t (which captures purely the quality of innovation, regardless of quantity). We also experiment with different truncation windows (e.g. five years) and with counting or excluding self-citations.

Panel (C) of Figure 6 shows the life cycle profile of productivity for all inventors, according to our benchmark measure. It is in line with those found in the recent literature on life cycle wages

²⁰If the same team produces two patents in the same year, we sum the citations to both of these patents in a 3 year window.

²¹On patent families, see Martinez (2010).

and productivity (Lemieux, 2006; Lagakos et al., 2017): It is increasing and concave. We will use the productivities at different ages as moments to target in the estimation of the model in Section 4.

Assigning team leaders

Our theory suggests that the team leader is the most productive inventor. There are several ways of identifying in the data who is “the most productive” inventor of a team.

Our first and main method for picking the patent team leader is to simply compare productivities, as defined above, and to choose the team member among all inventors on a given patent who had the highest cumulative productivity in period $t - 1$. Thus, inventor i will be the team leader of the team that files patent j if

$$P_{it-1} = \max_n (P_{n,t-1}),$$

where n indexes all inventors in the team.²² The remaining inventors of the team, if there are any, are considered to be the “team members” equivalent of the model.

We also explored two alternative and complementary ways of identifying the team leader: As the most senior inventor on the team, or as the first inventor listed in the application. Our results (not shown) are robust to these alternative team leader assignments.

Measuring interactions

There are many possible ways to measure interactions – one of the key variables of the model. We show results for several different measures going from least to most broad, which are:

1. Past co-inventors.
2. Inventors in the same firm.
3. Inventors in the same geographical region.

The first measure only counts as interactions past co-inventors, i.e., inventors listed on the same patents at some point in the past. This is a strong definition of interactions since these are people one has worked with with certainty. Also, it likely underestimates total interactions: Inventors have probably interacted with more people than just those with whom they have filed patents. To be exactly in line with the model, we could also count only past co-inventors who were better than the focal inventor. This measure of interactions will be our benchmark which corresponds to the model’s idea that one learns only from better people, which we test and confirm below. The second and third definitions are broader. They count inventors who were better than oneself at the time of the encounter, and with whom one has been in the same firm (definition 2), or with whom one has lived in the same region (definition 3), at some point over one’s past career. These are broader measures of interaction in the sense that one cannot be sure of the intensity of direct contact that actually happened with someone working in the same firm or living in the same region. The results below show, as is intuitive, that the effects on productivity from more interactions increase

²²According to this measure, there could be more than one team leader for a given patent.

progressively as one moves from the weakest interaction measure (measure 3) to the strongest one (measure 1).

Regarding our benchmark measure, formally, we can count the interactions with co-inventors better than oneself, and worse than oneself, that one has had in the past. The high-quality interactions of inventor i at time t are defined as the number of unique co-inventors j of inventor i in all past years, who had a productivity $P_{j,s}$, larger than that of inventor i in the period right before the encounter or interaction took place:

$$\text{High Quality Interactions}_{i,t} = \sum_{s=t_0}^t \text{unique more productive co-inventors}_{i,s}.$$

Conversely, “Low Quality Interactions” are defined as the number of unique co-inventors of inventor i in all past years, who had a productivity $P_{j,s}$, lower than that of inventor i in the period right before their encounter. As a robustness check in Table 3, we also count repeat interactions, i.e., if inventor i interacts with inventor j in two different years, that interaction is counted twice.

The regressions also include a number of important additional controls, such as the team leader’s age, which is the number of years he has already spent in the sample, the square of age, and controls for team characteristics, such as the team size, i.e., the number of inventors listed in a given patent. We also control for firm size using the number of inventors active in the firm at each time. To assign active inventors to their firms, we reconstruct the employment histories of all inventors based on the assignees (firms) listed on their patents.²³ The strategy used to build inventors’ employment histories is described in Appendix OA-1.²⁴

The sample consists of all inventors and patents between 1977 and 2010.²⁵ The first year of each inventor is omitted from the regressions in order to be able to construct the necessary quality and interaction measures. Table 1 presents summary statistics of the variables used in the regression analysis.

The effect of age

Intuitively, the effect of the external learning channel on productivity will be disciplined in the data by the effect of age. Importantly this effect must be measured while controlling for interactions and conditional on many detailed fixed effects. The age effect then captures the learning that happens over time, not through interactions, but through learning-by-doing, individual discovery and learning, or exposure to outside knowledge sources. The effect of age is consistently positive and significant. The lifecycle of productivity and idea quality is concave as was described before, and as can be seen in panels (B) and (C) of Figure 6.

²³For similar approaches, see Hoisl (2007) and Nakajima et al. (2010).

²⁴The assignee, could also be a university, hospital, research centre, or even an individual inventor. We treat them all as the inventors’ de facto employers for our purposes, which makes most sense for the variables we need to construct.

²⁵We remove the patents after 2010 to minimize the truncation issue in citations.

TABLE 1: SUMMARY STATISTICS

	Mean	Standard Deviation	Min	Max
<i>Idea Quality</i>				
<i>Conditional on patenting:</i>				
3-year citations	1.4	3.09	0	401
5-year citations	2.2	4.53	0	421
3-year citations (excluding self citations)	1.2	2.76	0	401
<i>Unconditional on patenting (at individual level):</i>				
3-year citations	0.3	2.30	0	302
<i>Interactions</i>				
High Quality Interactions	1.7	3.25	0	68
Low Quality Interactions	8.9	18.0	0	578
Total interactions	10.6	19.9	0	605
High Quality Interactions in the firm	673	1848	0	45443
High Quality Interactions in the region	2876	7581	0	184077
<i>Team and firm characteristics</i>				
Team leader age	5.8	5.23	1	34
Team size	3.1	2.08	1	10
Firm Size	525	964	1	5843

Notes: Summary stats based on 1,891,694 inventor-year observations used in the regressions, corresponding exactly to the sample used in the regressions in Tables 2 and 3 and in the estimation in Section 4. “Conditional on patenting” refers to the quality of patents produced by inventors. “Unconditional on patenting” refers to the quality of patents produced by inventors, where years with zero patents are counted as “zero” quality (i.e., zero citations received). “Total interactions” is the sum of High and Low Quality interactions, as defined in the text. “Firm size” is the number of active inventors in a firm at a given time. See the text for a detailed description of these variables.

3.3 The Link Between Interactions and Productivity

We now present our key results on the link between past interactions and subsequent productivity, controlling for characteristics of the team leader, the team’s inventors, the firm, and the research environment.

Specification and estimation

Our baseline specification is as follows:

$$\begin{aligned}
 q_{j,t} = & \beta_1 \text{Interactions}_{i[j],t-1} + \beta_2 \text{Age}_{i[j],t} + \beta_3 \text{Age}^2_{i[j],t} + \beta_4 \text{Team Size}_{j,t} + \beta_6 \text{Firm Size}_{j,t} \\
 & + B_t + B_{s[j]} + B_{f[j]} + B_{r[j]} + B_{r[j]} \times B_t + B_{s[j]} \times B_t + B_{s[j]} \times B_{r[j]} + B_{i[j]} + \varepsilon_{j,t}, \quad (18)
 \end{aligned}$$

where $q_{j,t}$ is the standardized (z-score) idea quality of team j at time t . The team leader of team j is denoted by $i[j]$, employed by firm $f[j]$, in region $r[j]$ and technology sector $s[j]$. Idea or innovation quality is measured in turn by one of the several measures described above and transformed into a z-score by subtracting the mean and dividing by the sample standard deviation to yield $q_{j,t}$. “Interactions” is a measure of the past interactions of the team leader; recall that we explore several measures. *Age* is the age of the team leader. With some abuse of notation for simplicity, B_t denotes a set of year fixed effects, $B_{s[j]}$ are sector fixed effects, $B_{f[j]}$ are firm fixed effects, and $B_{r[j]}$ are region fixed effects.²⁶ In order to absorb more of the inventor and team heterogeneity, we add region-year, sector-year, sector-region, and, most importantly, individual inventor fixed effects.

Inventor fixed effects, in particular, go a substantial way towards alleviating a potential endogeneity problem, namely that more productive inventors may be able to both interact with more people and have more co-inventors. Conditional on inventor fixed effects, a positive coefficient on lagged interactions could still arise because of endogeneity, but in much less likely cases. It would have to be that an inventor who expects to be better than his own average productivity in year t intentionally starts interacting with more people at time $t - 1$. Alternatively, there would have to be inventor-specific shocks which simultaneously increase an inventor’s productivity in a given year year t and make him co-invent with more high quality inventors in the previous year $t - 1$.

Instrumental variable strategy

To further address the remaining potential heterogeneity, we adopted an instrumental variable (IV) strategy that uses ethnic networks as an instrument. More precisely, there is a strong first stage of the number of interactions an inventor has at a given time, and the share of inventors of his ethnicity in his firm or region at that time, conditional on all previously described fixed effects as well as region-ethnicity and firm-ethnicity fixed effects. Conditional on these latter fixed effects, the instrument is likely to satisfy the exclusion restriction.²⁷ Using this IV strategy, we find that the effect of interactions is strongly significant – even larger in magnitude than in the OLS specification. Ethnic networks are a relatively distinct topic from the one covered in this paper. In addition, this IV estimator measures a LATE effect on those who are induced to interact more from having more ethnically similar people around them. We thus prefer the strategy with very detailed fixed effects that we use here. It is, however, possible to include the IV results and describe their construction in more detail in a separate section in a revision of the paper.

Main results

Table 2 contains the results from the benchmark regressions. In this table, idea quality (the left-hand side variable) is always measured according to the benchmark measure, which is the forward

²⁶Regions r are defined at comparable administrative levels across countries. For instance, for the US, we use the BEA economic areas, which are equivalent to Metropolitan Statistical Areas (MSAs) in the case of metropolitan regions. For Europe, we use the Nomenclature of Territorial Units for Statistics (NUTS3) level. For all remaining countries, we use the smallest level used by the OECD classification of regions (Territorial Level 3) - which corresponds to, e.g., Statistical Divisions in Australia, Census Divisions in Canada, or Prefectures in Japan.

²⁷Some regions or firms may be better suited to some ethnicities, leading to higher productivity and more interactions for those ethnicities, but that is captured by these additional fixed effects.

TABLE 2: MAIN RESULTS

	Benchmark		High Tech Sector	Broader Interactions Measures	
				Firm	Region
High Quality Interactions (t-1)	0.021*** (0.001)	0.027*** (0.001)	0.035*** (0.004)	0.0000385*** (0.000001)	0.0000266*** (0.0000008)
Low Quality interactions (t-1)		-0.004*** (0.000)			
Team Size	0.028*** (0.001)	0.028*** (0.001)	0.034*** (0.004)	0.027*** (0.001)	0.034*** (0.001)
Log Firm Size	-0.022*** (0.002)	-0.021*** (0.002)	-0.045*** (0.010)		-0.028*** (0.003)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year \times Sector FE	Yes	Yes	Yes	Yes	Yes
Year \times Region FE	Yes	Yes	Yes	Yes	Yes
Sector \times Region FE	Yes	Yes	Yes	Yes	Yes
Team Leader FE	Yes	Yes	Yes	Yes	Yes
<i>N</i>	1574216	1574216	286231	1574216	1338075
adj. R^2	0.187	0.187	0.133	0.187	0.158
F	88.25	103.9	16.52	164.6	177.4

Notes: All models include in addition year, sector, and region fixed effects (not listed). Column (1) uses the benchmark measure of interactions (past co-inventors). Column (2) includes also lower quality past co-inventors. Column (3) limits the sample to the high-tech sector. Column (4) and (5) consider, respectively, interactions measured as inventors in the same firm or in the same region. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

citations of the patent in a 3-year window. In column (1), interactions are measured in the benchmark way, as the total number of past higher quality co-inventors.

The effect of high quality interactions on patent quality is consistently positive and significant: One additional high quality interaction increases subsequent patent quality by 0.02 standard deviations, which corresponds to a 0.06 level increase in citations per patent in a 3-year window. What does this mean economically? If we use the mean patent quality for existing patents in the sample (i.e., mean citations per inventor in a 3-year window conditional on patenting) from Table 1, this represents a 4% increase. However, since inventors do not patent every year, a more accurate comparison would use the mean citations per patent (in a 3 year window), per inventor, per year, in the sample (unconditional on patenting): One interaction represents a 21% increase in citations relative to that unconditional mean (see Table 1).

In column (2), we simultaneously include those past co-inventors who were more productive and those who were less productive than the inventor at the time of their encounter. If anything, the effect of high quality interactions increases to 30%. In the model, individuals benefit more from interacting with people better than they are, i.e., who have higher productivity.²⁸ In the data, low quality interactions, conditional on inventor fixed effects, have a very slight negative effect on

²⁸In fact, in the model, agents gain nothing by interacting with less productive people.

TABLE 3: ROBUSTNESS CHECKS

	Repeated	Productivity		Low
	Interactions	5-year cit.	Excl. Self-cit.	Tech
High Quality Interactions (t-1)	0.010*** (0.001)	0.019*** (0.001)	0.020*** (0.001)	0.019*** (0.001)
Team Size	0.028*** (0.001)	0.029*** (0.001)	0.018*** (0.001)	0.027*** (0.001)
Log Firm Size	-0.022*** (0.002)	-0.029*** (0.002)	-0.016*** (0.002)	-0.019*** (0.002)
Firm Fixed Effects	Yes	Yes	Yes	Yes
Year \times Sector FE	Yes	Yes	Yes	Yes
Year \times Region FE	Yes	Yes	Yes	Yes
Sector \times Region FE	Yes	Yes	Yes	Yes
Team Leader FE	Yes	Yes	Yes	Yes
<i>N</i>	1574216	1574216	1574216	1233323
adj. R^2	0.186	0.225	0.177	0.209
F	82.58	118.0	63.15	80.36

Notes: All models include in addition year, sector, and region fixed effects. Column (1) counts repeated interactions twice. Column (2) uses a 5-year window to measure citations (i.e., to measure idea quality); column (3) excludes self-citations. Column (4) limits the sample to low-tech sectors. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

productivity: They may be crowding out more productive interactions and wasting an inventor's time.

Column (3) restricts the sample to high-tech sectors according to the Eurostat classification.²⁹ In high tech sectors, the effect of one more high quality interaction is even stronger, equal to 8% of the sample mean conditional on patenting, and 39% unconditional on patenting.³⁰

Columns (4) and (5) use the broader measures of interactions. Interactions as measured by the number of inventors in the firm (column (4)) and in the region (column (5)) who are more productive than the focal inventor, also significantly affect productivity. Naturally, the effects of these interactions are smaller than the effect of the benchmark measure, since the interaction with someone in the same firm or region is more distant and thus weaker. An inventor who has 10 more higher quality inventors in his firm has a 0.4% higher productivity. Additionally, an inventor with 10 more higher quality inventors in his region has a 0.3% higher productivity (both effects measured relative to the unconditional mean). However, these interactions can still be significant because of the sheer number of inventors in a firm or region (see Table 1). For instance, in the median region by inventor density, there are 506 inventors. In the top 25% regions, there are 2100 inventors; in the top 5% there are 13,200 inventors. Our estimates indicate that – even conditional on individual

²⁹The Eurostat classification uses specific sub-classes of the International Patent Classification (IPC) as defined in the trilateral statistical report of the EPO, JPO and USPTO. Macro high-tech technical fields are: Computer and automated business equipment; Microorganism and genetic engineering; Aviation; Communications technology; Semiconductors; Lasers. The list of sub-classes and their definition is provided by Eurostat at: http://ec.europa.eu/eurostat/cache/metadata/Annexes/pat_esms_an2.pdf

³⁰See Table 3 for the effect in low-tech sectors, which is still strongly significant, but smaller in magnitude.

fixed effects – inventors who live in regions with many productive inventors, or work in firms who employ more productive inventors, are significantly more productive.

Robustness checks

To further test the robustness of the relation between interactions and productivity improvements, we consider additional measures of productivity and interactions in Table 3. In column (1), we count repeated interactions: If inventor i interacts with inventor j in two different years, that interaction is counted twice. In column (2), we use a five-year window to count forward citations. In column (3), we exclude self-citations. In column (4), we focus only on low tech sectors. In all cases, the effect of interactions is significant and the magnitudes are very robust.

4 Estimation

In this section, we bring together the model from Section 2 and the empirical results from Section 3 to estimate the model using the new data on productivity, innovation, and interactions. We explain our numerical algorithm, the parameters estimated, and the moments that we target. We then show that the model fits the data very closely, not only for targeted moments, but also for non-targeted ones.

4.1 Computational Algorithm, Functional Forms and Parameters

We start by describing our numerical algorithm. For any given set of parameters, we simulate the model for $n_{sim} = 1,000,000$ individuals for $T = 25$ years.³¹ We initialize the simulations with a given initial cross-sectional distribution $F(z, 0)$ (see Table 4).

TABLE 4: FUNCTIONAL FORMS

Function	Description
$q(z) = z^{1-\eta} n^\eta$	Idea production function.
$c(m_D) = \frac{\kappa}{2} m_D^2$	Interactions cost function.
$F(z, 0) = \frac{1}{1+\lambda z^{-1/\theta}}$	Initial cross-sectional productivity distribution.
$\Gamma(x, 0) = \frac{1}{1+k_0 x^{-1/\theta}}$	Age zero productivity distribution.
$\Psi(x, s) = \frac{1}{1+\rho s^v x^{-1/\theta}}$	External (age-dependent) learning distribution.

In each period, there is a constant probability δ that an individual survives to the next period. We can compute the number of successful interactions (i.e., the number of times an agent improves his productivity by learning from someone better than him) and the number of successful draws from the external distribution. We thus obtain the realized productivity distribution $F(z, t)$ at each

³¹We discretize each year with a size step of $\Delta t = 0.2$.

time t , as well as the age-conditional productivity distributions, $G(z, s, t)$. We then use the distribution $F(z, t)$ to compute the paths of researchers' wages $w(t)$ and productivity cutoffs $\bar{z}(t)$ in the occupational choice problem. In the BGP, the economy reaches the invariant distributions $\Phi(z)$ and $\Gamma(z, s)$. At age zero, new inventors draw their productivity from the distribution $\Gamma(z, 0)$.³² When an individual dies, he is replaced by another individual with a productivity draw from the same initial distribution $\Gamma(z, 0)$.

Using the simulated model, we compute the values of key data moments, described below, and iterate on the parameters until we find the ones for which the model-generated moments best match the data moments. We match the moments in the data when the economy is on a BGP. To parameterize the model we suppose that the initial cross-sectional distribution $F(z, 0)$, the age zero distribution $\Gamma(x, 0)$, and the external source distribution $\Psi(x, s)$ are log-logistic. In Table 4, we summarize all the functional forms relevant for the calibration of the model.

We estimate nine parameters of the model – listed in Table 5 – and captured by the vector:

$$\chi := (\kappa, m_X, \lambda, \theta, \nu, \eta, \delta, k_0, \alpha).$$

TABLE 5: ESTIMATED PARAMETERS

Parameter	Description	Value
κ	Cost of interactions parameter.	1.1906
m_X	Rate of draws from external learning source.	0.5281
λ	Location parameter of initial productivity distribution $F(z, 0)$.	0.5439
θ	Tail parameter of productivity distributions $F(z, 0)$, $\Gamma(x, 0)$ & $\Psi(x, s)$.	0.4503
ν	Exponent on location parameter $\rho(s) = \rho s^\nu$ of $\Psi(x, s)$.	0.5987
η	Team leader's span of control.	0.2698
δ	Death hazard rate.	0.2171
k_0	Location parameter of age zero distribution $\Gamma(x, 0)$.	0.0492
α	Final good elasticity with respect to aggregate productivity.	0.1220

4.2 Data Moments and Estimation

We estimate the nine-parameter vector χ by matching key data moments. We target the following nine groups of moments, for a total of 86 moments:

1. Average number of high-quality interactions of team leaders by age (estimated at 18 age values, each corresponding to one moment).
2. Average idea (patent) quality of team leaders by age (estimated at 18 age values).

³²In the data, age zero corresponds to the first year in the sample, i.e., the first year in which an inventor patents.

3. Average productivity of all inventors by age (estimated at 18 age values).
4. Team size distribution (discretized and estimated at 20 team size values).
5. The age distribution (discretized and estimated at 18 age values).
6. The share of inventors who are team leaders.
7. The marginal effect of high-quality interactions on idea quality.
8. The marginal effect of age on idea quality.
9. The economy's growth rate.

Some of these groups of moments are life cycle profiles (1-3) or entire distributions (4 and 5), which are matched at discretized values. Table 6 summarizes the moments we target.

Two of the moments are the marginal effects of, respectively, age and interactions on productivity (moments 7 and 8). Intuitively, conditional on the number of interactions and all our other fixed effects and controls (including also individual fixed effects), the effect of age on productivity helps calibrate the strength of the external learning channel since it captures improvements in productivity that arise over time, but do not directly come through interactions. The effects of age and interactions were estimated in the regressions in Section 3 with many different measures of productivity and interactions, as well as with several different specifications. We run an equivalent regression using data generated from the simulated model:

$$\text{Productivity}_{it} = \beta_1 + \beta_2 \text{High-quality interactions}_{it-1} + \beta_3 \text{Age}_{it} + \beta_4 \text{Age}_{it}^2 + \varepsilon_{it} \quad (19)$$

where $i \in \{1, \dots, n_{sim}\}$ indexes simulated inventors and $t = \{1, \dots, 25\}$ age. We then match the marginal effect of interactions β_2 and age β_3 from the actual, and moment simulated, data.

TABLE 6: TARGETED MOMENTS

Moment	Weight ω_i	Description	Model	Data
1-18	$\frac{1}{9} \frac{1}{18}$	Average high-quality interactions of team leaders by age	see Fig. (6a)	see Fig. (6a)
19-36	$\frac{1}{9} \frac{1}{18}$	Average Idea quality of team leaders by age	see Fig. (6b)	see Fig. (6b)
37-54	$\frac{1}{9} \frac{1}{18}$	Productivity of all inventors by age	see Fig. (6c)	see Fig. (6c)
55-74	$\frac{1}{9} \frac{1}{20}$	Team size distribution	see Fig. (6d)	see Fig. (6d)
75-82	$\frac{1}{9} \frac{1}{18}$	Age distribution of team leaders	see Fig. (6e)	see Fig. (6e)
83	$\frac{1}{9}$	Fraction of team leaders	0.5739	0.5793
84	$\frac{1}{9}$	Regression coefficient on high-quality interactions, β_2	0.0561	0.0561
85	$\frac{1}{9}$	Regression coefficient on age, β_3	0.0389	0.0378
86	$\frac{1}{9}$	Growth rate	0.025	0.025

We choose the values for all the estimated parameters in Table 5 that minimize the weighted sum

$\mathcal{L}(\chi)$ of the absolute difference between the model-implied moments and the data moments:

$$\mathcal{L}(\chi) := \min \sum_{i=1}^{86} \omega_i \frac{|\text{model}(i) - \text{data}(i)|}{\frac{1}{2}|\text{model}(i)| + \frac{1}{2}|\text{data}(i)|},$$

where ω_i is the weight assigned to moment i , reported in Table 6. We assign the weights ω_i so that all nine groups of moments (each represented by a row in Table 6) receive equal weight. Thus, moments within a group receive a weight of $1/9$ multiplied by $1/n_p$ where n_p is the number of discretized points of the distribution that we match.

4.3 Parameter Estimates

Table 5 reports the estimates we obtain from our calibration procedure. The meeting cost parameter is equal to $\kappa = 1.1906$. This leads to an endogenous interaction rate of $m_D = 0.5620$, slightly larger than the external source meeting rate $m_X = 0.5281$. The location parameter of the initial productivity distribution is equal to $\lambda = 0.5439$, and the tail parameter of the distributions is $\theta = 0.4503$. Parameter ν determines the concavity in age of the age-dependent location parameter of the external distribution $\rho(s) = \rho s^\nu$. A larger value of ν means that the average productivity of external ideas grows faster during the first years of experience. In our calibration $\nu = 0.5987$. The relative importance of team members in patent production is $\eta = 0.2698$, which implies that the team leader's span of control is equal to $(1 - \eta) = 0.7302$. The death hazard rate is equal to $\delta = 0.2171$. Finally, the elasticity of final good production to productivity is equal to $\alpha = 0.1220$.³³

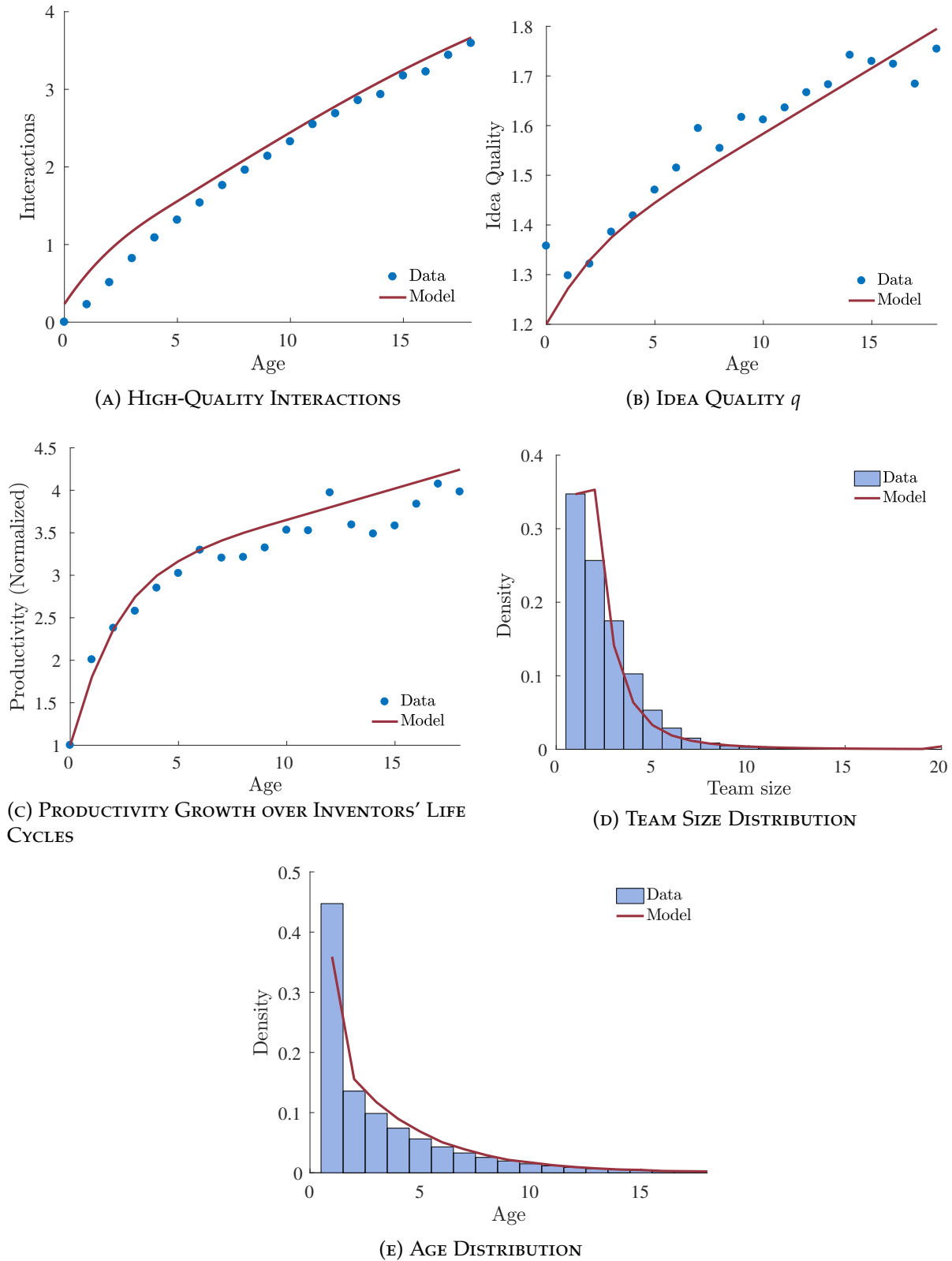
4.4 Goodness of Fit

Table 6 reports the target for the moments from the data and the values we obtain in the estimated model. In cases in which the moment is a distribution function, we do not report a value, but instead provide the reference to the relevant figure that shows how the distribution generated from the simulated model matches the one in the data.

Our calibration closely fits all the targeted moments; we are able to capture most of the salient and peculiar features of the data. This can be seen most vividly in Figure 6. Panel (A) shows that the model implies that the average number of high quality interactions for team leaders is increasing in age and that the relation is close to linear; this is also exactly what we see in the data. The model also predicts that the average idea quality produced by team leaders and their teams is increasing and concave in age, which is also true of the data (see panel (B)). The productivity of all inventors is increasing and concave in age in both the model and the data (see panel (C)). Team leaders account for 57% of individuals and the team size distribution is decreasing (see panel (D) and Table 6). The age distribution is also decreasing, with an initial mass of individuals of age 0 (see panel (E)).

³³Intuitively, this parameter is calibrated to match the empirical growth rate.

FIGURE 6: TARGETED MOMENTS – FIT BETWEEN MODEL AND DATA



Panel (A) represents the high-quality interactions of a team leader as a function of his age (age 0 is the first year in the patent data). Panel (B): represents the de-trended idea quality $q(z, t)e^{-g_q t} = q(z, t)e^{-g(1-\eta)t}$, where $g_q = g(1-\eta)$ is the growth rate of idea quality, as a function of team leader's age. Panel (C) depicts average inventor productivity z across all inventors at a given age. It is normalized by productivity at age 0. Panel (D) shows the distribution of the team size n and Panel (E) shows the distribution of age (time in the sample).

4.5 Fit for Non Targeted Moments

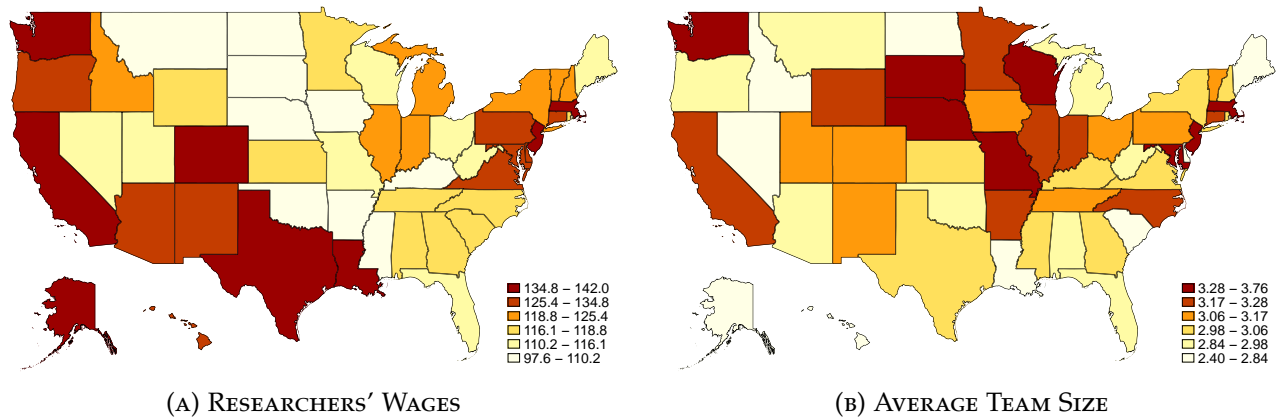
We now check the goodness-of-fit of the estimated model for some non-targeted moments.

R&D tax credits, wages and team size

We start with two moments which are interesting per se for policy reasons, and which are intentionally very distinct from the targeted moments (thus providing a particularly convincing non-targeted fit): The elasticity of wages to R&D tax credits, and the elasticity of team size with respect to R&D tax credits.

Goolsbee (1998, 2003) argued that government support for R&D would simply push up scientists' and researchers' wages. In our model, we can mimic an R&D tax credit through a subsidy on the costs of the teams' innovation (i.e., a subsidy on total wage payments to team researchers (see Appendix OA-2 for a formal treatment)). The model-generated elasticity of wages with respect to R&D tax credits is 0.026. We compare this to the elasticity in the data for the U.S. We compute average researchers' wages at the state level in the CPS March supplement (ASEC file). Panel (A) of Figure 7 illustrates the large variation in research wages across U.S. states. We then regress the wages on state-level R&D tax credits, which vary across states and years. We find that the overall elasticity of researcher's wages to the R&D tax credit is 0.023, extremely close to the one implied by our model. We also computed the elasticity in the data of researchers' wages to total government R&D funding (rather than just R&D tax credits) and found very similar results.

FIGURE 7: RESEARCHERS' WAGES AND RESEARCH TEAM SIZES ACROSS U.S. STATES



Notes: Panel (A) depicts the average wages of researchers at the state level (from the CPS March supplement) for the period 2003-2016. Panel (B) depicts average team size of teams listed on patent across U.S. states, based on the EPO data.

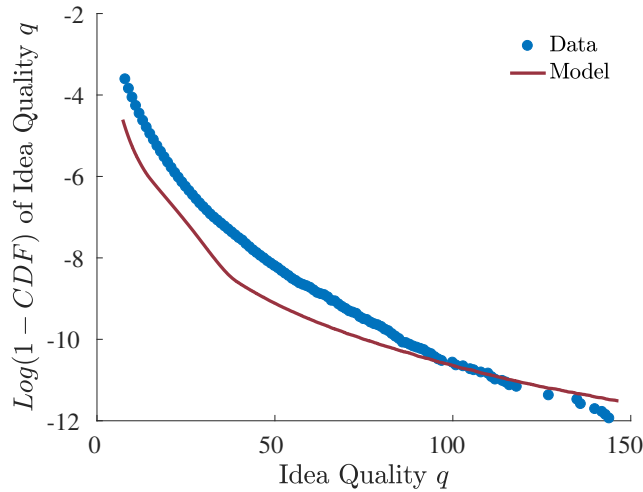
Teams are key in both our model and in the data. Their size is related to the allocation of inventors to team leaders or team members, as well as to the quality of innovation. Panel (B) of Figure 7 shows the extreme variation in team sizes across U.S. states. Thus, one may wonder about the effect of government support for research on team size. We find that the elasticity of team size with respect to R&D tax credits is 0.057 in the model and 0.044 in the data. Again, our model is able

to closely replicate a non-targeted, and policy-relevant, elasticity.

Idea quality Pareto tail

In Figure 8 we check another non-targeted moment, namely the tail of the idea quality distribution. The figure depicts the log counter cumulative distribution ($\log(1 - CDF)$) for idea qualities above the 99th percentile.³⁴ The tail of the distribution generated by our estimated model strongly resembles the one observed in the data.

FIGURE 8: TAIL OF IDEA QUALITY DISTRIBUTION



Note: The figure plots the log of the counter CDF ($\log(1 - CDF)$) of the idea quality q for idea quality above the 99th percentile.

5 Quantitative Analysis and Policy Experiments

As shown in Sections 4.4 and 4.5, the estimated model fits the data very well for both targeted and non-targeted moments. The model is complete enough to be used to understand past and contemporaneous changes in knowledge diffusion, productivity, and innovation. To illustrate this, we now put our estimated model to work and quantitatively explore four policy experiments:

1. **Importance of interactions:** We quantify the importance of interactions for productivity and growth, relative to other learning channels.
2. **Reducing interaction costs:** What happens when interaction costs are reduced? Concretely, this can be interpreted as companies or industries adopting a “Google model,” with open-floor offices and designated hang-out times that foster interactions. It also can be mapped to the rise of IT which facilitates interactions.

³⁴This corresponds to patents with more than 6.9 citations in the 3-year window.

3. **Reduced access to external ideas:** We study the effect of reduced access to external ideas. This highlights potential downsides of agglomeration and the “paradox of proximity.”
4. **U.S. vs. Germany:** We estimate the model separately for two large innovating countries, Germany and the U.S., to compare their innovation production functions and team dynamics. What would happen to U.S. growth were it to inherit the (lower) interaction costs of Germany?

There are many other possible counterfactual analyses that can fruitfully be done using our estimated model. Moreover, the data can be sliced in several other ways: e.g., the model could be separately estimated by technological field or time period to see how the innovation production and research teams dynamics differ.

5.1 Interactions versus External Learning

We start by computing the contributions of interactions to economic growth.³⁵ For this exercise, we shut off, in turn, the endogenous interaction channel (setting $m_D = 0$) and the exogenous learning channel (setting $m_X = 0$). Table 7 shows the resulting growth rates in these two cases.

TABLE 7: THE ROLE OF INTERACTIONS AND EXTERNAL LEARNING

	Interactions + External Learning	No interactions: $m_D = 0$	No external learning: $m_X = 0$
Growth	2.5%	0.1%	0.62%

There are three key findings.

1. Shutting down any of the two learning channels significantly reduces growth.
2. The endogenous interaction channel contributes more to growth than the exogenous learning channel.
3. There is a strong complementarity between the two learning channels.

Clearly, both channels significantly contribute to growth; when either one is shut off, growth is reduced by at least 75%. Nevertheless, when the endogenous interaction channel is shut down (case with $m_D = 0$), growth is reduced to a striking 0.1%. If the external learning channel is shut down (case with $m_X = 0$), the growth rate drops by less, to 0.62%. In fact, recall from Section 2 that in the special case in which the external productivity distribution $E(z, s, t)$ is not growing, the growth rate is $m_D\theta$. In this case, shutting down interactions reduces growth to 0 (but, conversely, shutting down external learning would still lead to positive growth). This case, akin to the one in Lucas and Moll (2014), emphasizes even more starkly our quantitative finding that interactions are vital for growth.

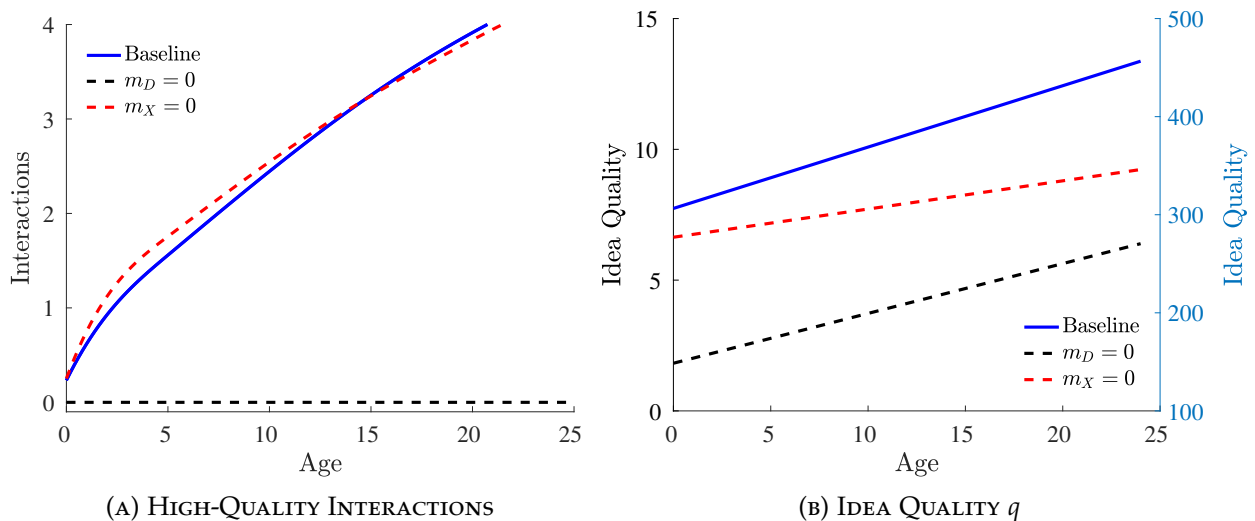
The third finding is that the contribution of each channel individually adds up to much less than the total growth rate when both channels are active. This complementarity between the two learning

³⁵For this, it is useful to recall the discussion in Section 2, which described in what ways the growth rate depends on the rate of external learning and endogenous interactions and some special cases.

channels is key: When agents learn more from the external source, each endogenous interaction becomes more valuable in expectation as one can hope to acquire that new external knowledge indirectly by meeting others. Individuals then optimally choose a higher endogenous meeting rate.

Figure 9 shows the effect of shutting down the endogenous and exogenous channels on the number of interactions and on idea quality. The solid blue curve represents the estimated benchmark case from Section 4 (i.e., the case in which the two channels are active). Panel (A) highlights that when the exogenous learning rate is zero ($m_X = 0$) team leaders have more interactions during their initial career years than in the benchmark case. However, interactions eventually become lower (here, after age 14). This highlights how younger individuals can have some successful interactions initially while their productivity is still low, but eventually the absence of an external source of new knowledge causes the pool of knowledge of people around them to become stagnant and less interesting. As a result, it is less worthwhile to incur the costs of interaction, as agents correctly anticipate that a given new interaction has little chance of teaching them something new. Panel (B) shows that the average idea quality is an order of magnitude lower when either the external or interaction learning channel is shut off, and that the absence of interactions ($m_D = 0$) leads to the lowest idea quality. This immediately translates into the lower growth rates illustrated in Table 7. Also note that, without an external source of learning, the average quality of idea over an inventor's life is flatter, which again illustrates the importance of an external pool of ideas to keep up learning opportunities. Put simply, with no external sources of ideas, a group of individuals will eventually learn most (albeit not all) there is to be learned from each other and innovation will slow down.

FIGURE 9: THE ROLE OF INTERACTIONS

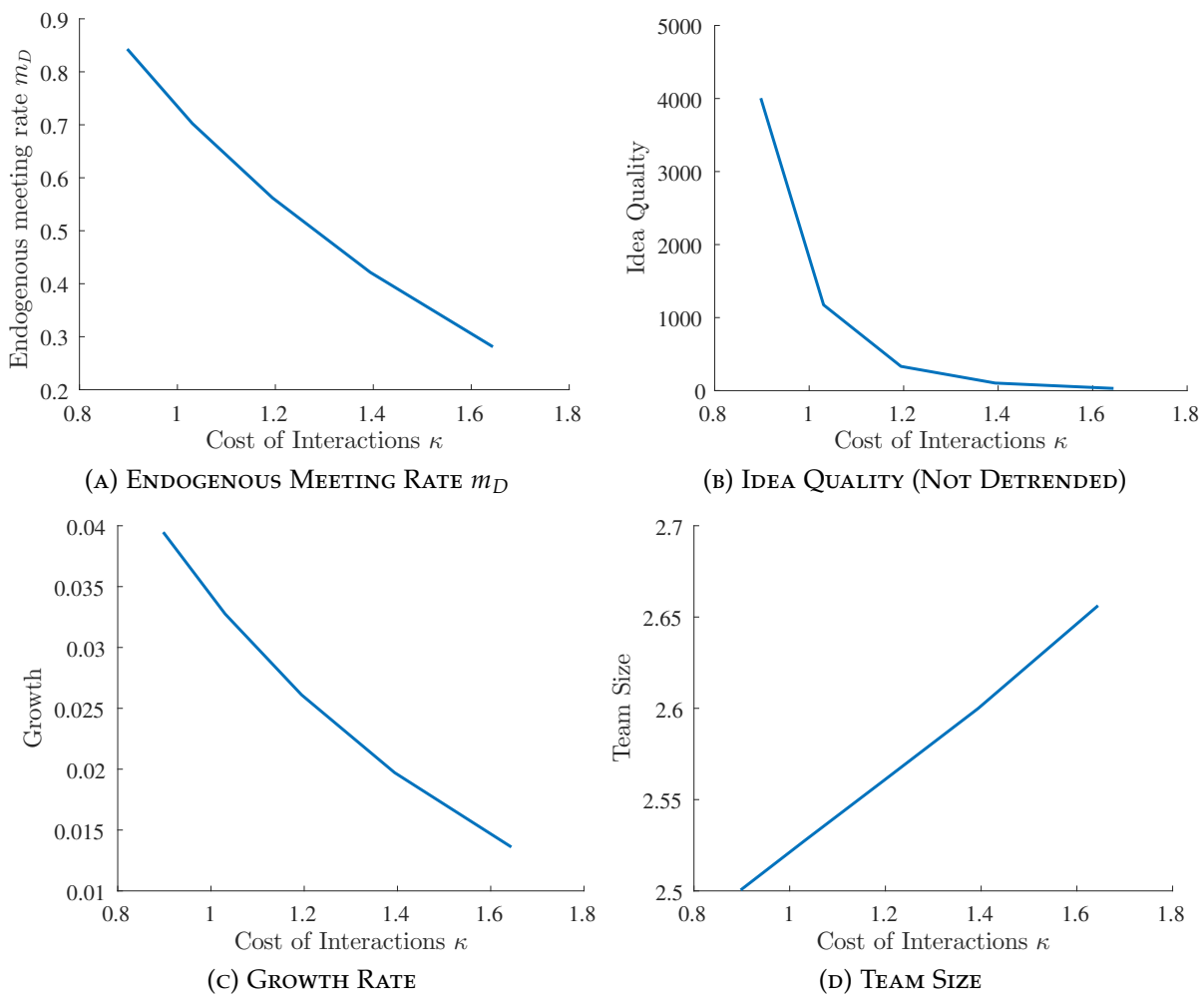


Panel (A) represents the high-quality interactions of a team leader as a function of his age (age 0 is the first year in the patent data). Panel (B): represents quality $q(z, t)$ as a function of team leader's age. In this panel, the right vertical axis is for the baseline scenario. In each panel, the solid blue line represents the baseline case, with both the endogenous interaction channel and the external learning channel active. The black dashed line corresponds to the case with no endogenous interactions ($m_D = 0$) and the red dashed line to the case with no external learning ($m_X = 0$).

5.2 Lowering Interaction Costs: Information Technology or the “Google Model”

What happens if the cost of interactions, κ , decreases? This scenario can mimic the spread of information technologies (IT) that make communication and, hence, interactions, easier, even across larger geographical distances. IT can also break down language barriers (e.g., through real-time translation or easy tools such as Google translate), or matching frictions (through online platforms and specialized forums). Many dedicated programs and apps have appeared to make interacting easier and cheaper, e.g., Slack, Skype, or FaceTime.³⁶ A decrease in interaction costs can also mimic the rise of new business models, such as the “Google model” that allows for designated spaces and times for interactions or, more generally, the open space office.

FIGURE 10: THE EFFECTS OF INTERACTION COSTS κ



Panel (A) represents the endogenous meeting rate m_D . Panel (B): represents the idea quality $q(z, t)$. Panel (C) depicts the yearly growth rate of product $g_y = \alpha g$ and Panel (D) the average team size. All plots as functions of interaction costs κ .

A lower interaction cost κ will, all else equal, increase the endogenous meeting rate m_D . As

³⁶On the differential impact of I.T. in Europe and in the U.S., see Bloom, Sadun, and Reenen (2012).

researchers have an easier time meeting others, they learn more and the quality of innovation and the growth rate are higher. These results lend support to the many attempts in the real world to make interactions easier through conferences, socializing events, and common spaces. A perhaps less immediate result is that lower interaction costs decrease the average size of teams (Figure (10)). The decrease in κ only very slightly increases the cutoff \bar{x} above which individuals become team leaders (see equation (5)). At the same time, the productivity distribution improves, creating more mass to the right of the (slightly increased) cutoff. On balance, there are more team leaders, and thus more competition for skilled researchers hired as team members. Therefore, teams become smaller on average.³⁷

5.3 Limiting Access to External Knowledge

What happens if access to external learning sources is reduced? There are many concrete situations in which individuals become less exposed to external sources of knowledge. First, there may be a strong agglomeration and geographical concentration of talent in some areas, which, paradoxically, may lead to a great deal of interactions with similarly-minded people, but little inflow of new knowledge. This is a version of the “proximity paradox,” whereby too much cognitive or geographical proximity with the same group of people – importantly, without additional external inflow of new knowledge – can hinder innovation. Second, there may be very intensive specialization that makes it hard to understand or adopt external knowledge (even if interactions with others from the same specialization are very productive, as everybody speaks the same “language.”) Third, protectionism, with an ensuing reduction in trade and foreign direct investment (FDI), can lead to lower inflow of new knowledge.³⁸ In all these settings, each specialized or concentrated group of researchers becomes “an island” that gets draws from the outside world’s knowledge only rarely; almost all of the learning happens from interacting with other people.

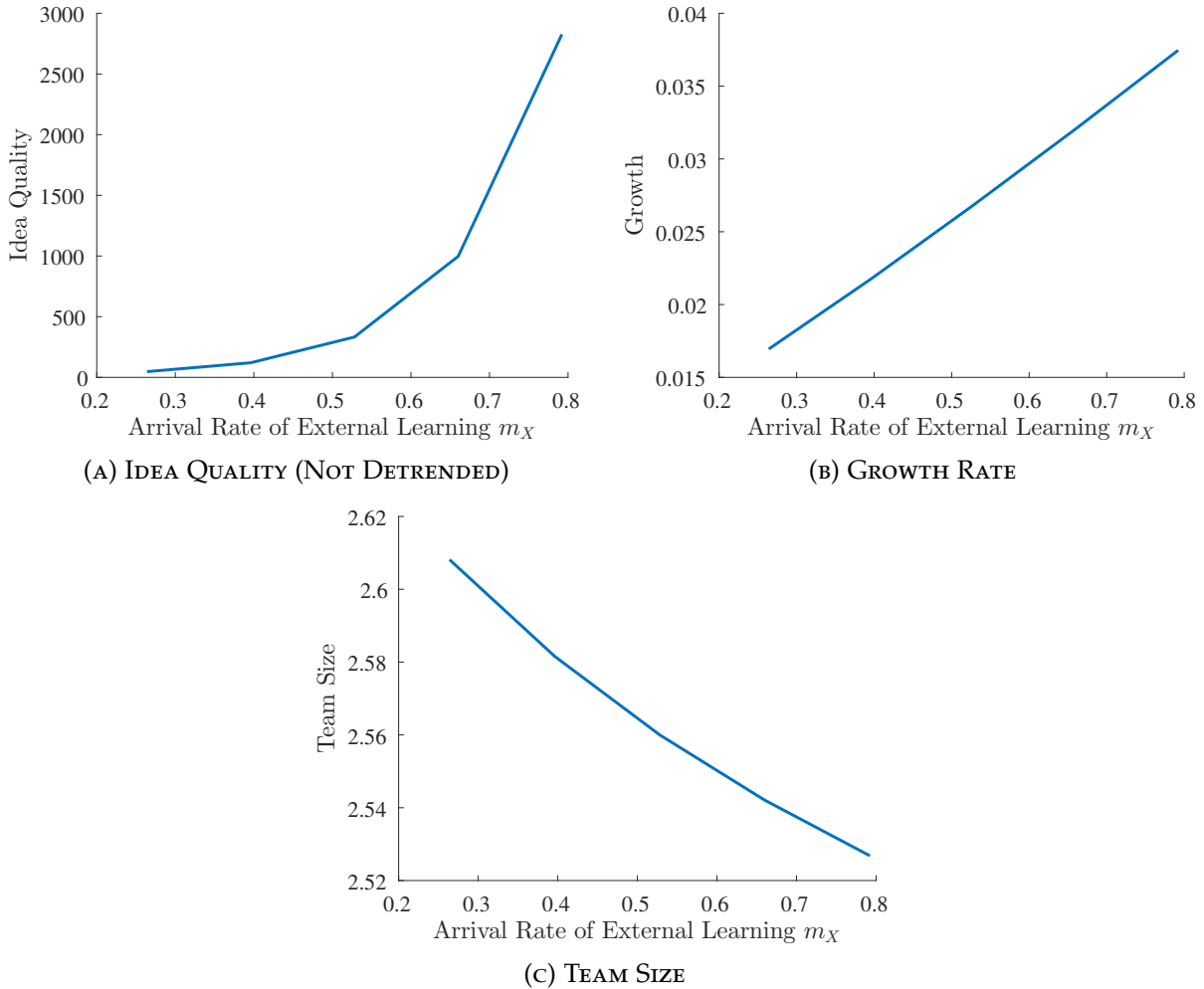
Figure 11 shows that if individuals have less access to external knowledge (m_X decreases, while m_D is held constant), the average idea quality is lower and the productivity of individuals grows slower. Analogous to the case of higher interaction costs, and for similar reasons, a rarer access to external knowledge increases the average team size. See also Appendix Figure A-4 for the full distributions of team size and productivity for different rates of external learning.

Historically, there have been many examples of external knowledge channels improving knowledge and productivity in previously autarkic societies. Roads, railroads, sea routes, transport infrastructure and communication technologies were instrumental in facilitating knowledge “from the outside” and, subsequently, speeding up innovation. The economic history papers presented in the introduction describe many such episodes (see also [Agrawal, Galasso, and Oettl \(2017\)](#)).

³⁷See also Appendix Figure A-3, which shows the effect of κ (for values $\kappa = \{0.9, 1.2, 1.6\}$) on the team size distribution and the de-trended productivity distribution.

³⁸See the papers on trade and technology diffusion discussed in Section 1.1, keeping in mind that in our case, it is *knowledge* used to produce innovation, and not directly production technologies that are diffused.

FIGURE 11: EFFECT OF ACCESS TO EXTERNAL LEARNING AND IDEAS



Panel (A): represents the idea quality $q(z, t)$. Panel (B) depicts the yearly growth rate of product $g_y = \alpha g$ and Panel (C) the average team size. All plots as functions of external ideas meeting rate m_X .

5.4 The U.S. versus Germany: The Effects of Interactions in Different Labor Markets

We can also use our model to understand how interactions affect productivity and growth in different economic environments. It is possible to slice the data in many different ways, and to estimate the model separately by time period, by technology sector, or by country. Here, we estimate our model separately for the U.S. and Germany – two major patenting countries with very different labor markets and education (i.e., human capital acquisition) systems. Related to this, these two countries have very different life cycle profiles of interactions and productivity (as illustrated in Figure 12). Using the same estimation procedure outlined above, but applied to the data from each country separately, we obtain the country-specific parameters summarized in Table 8. The model fits each country’s data remarkably well. The moments in the data and model are summarized in Table 9 and Figure 12.

Panel (A) of Figure 12 shows that team leaders in Germany have, on average, more high-quality interaction than those in the U.S. The endogenous meeting rate is higher for Germany ($m_D = 0.6096$)

TABLE 8: ESTIMATED PARAMETERS FOR THE U.S. AND GERMANY

Parameter	All	U.S.	Germany (DE)
κ	1.1906	1.4766	1.2181
m_X	0.5281	0.5186	0.5023
λ	0.5439	2.4436	0.1531
θ	0.4503	0.4335	0.4555
ν	0.5987	0.2976	0.8537
η	0.2698	0.2975	0.2190
δ	0.2171	0.2918	0.2063
k_0	0.0492	0.1910	0.0097
α	0.1220	0.1667	0.1131

TABLE 9: MOMENTS FOR THE U.S. AND GERMANY

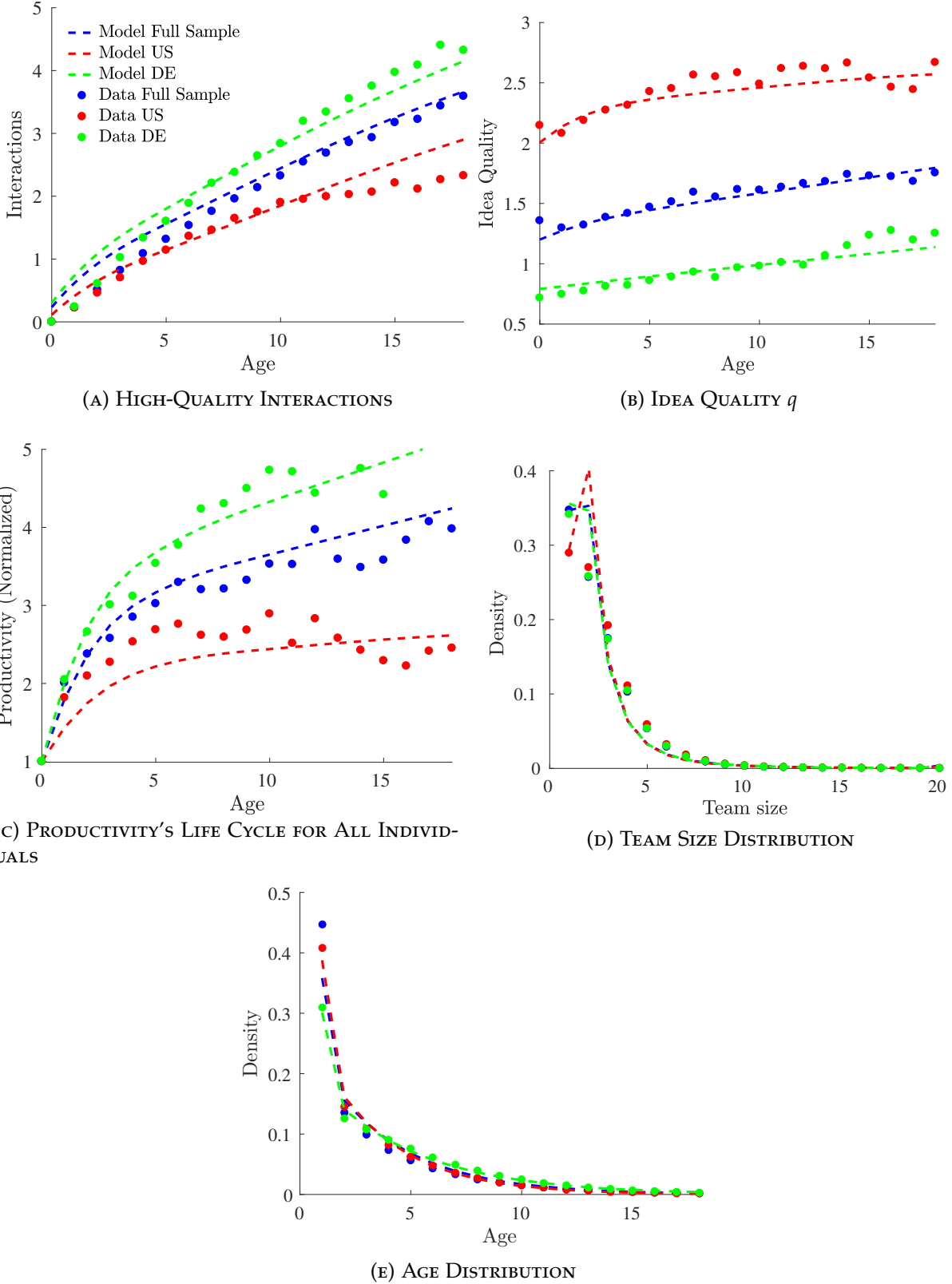
	Model Full Sample	Data Full Sample	Model US	Data US	Model DE	Data DE
Interactions	Figure 12a	Figure 12a	Figure 12a	Figure 12a	Figure 12a	Figure 12a
Patent Quality	Figure 12b	Figure 12b	Figure 12b	Figure 12b	Figure 12b	Figure 12b
Productivity All	Figure 12c	Figure 12c	Figure 12c	Figure 12c	Figure 12c	Figure 12c
Team Size Distribution	Figure 12d	Figure 12d	Figure 12d	Figure 12d	Figure 12d	Figure 12d
Age Distribution	Figure 12e	Figure 12e	Figure 12e	Figure 12e	Figure 12e	Figure 12e
Interactions Coefficient, β_1	0.0561	0.0561	0.0850	0.0850	0.0298	0.0298
Age Coefficient, β_2	0.0389	0.0378	0.0595	0.0595	0.0150	0.0150
Fraction of Team leaders	0.5739	0.5793	0.5130	0.5430	0.5829	0.5806
Growth Rate	2.5%	2.5%	2.74%	2.74%	2.42%	2.42%

than for the U.S. ($m_D = 0.4227$). This is a consequence of both lower meeting costs in Germany ($\kappa = 1.2181$ for Germany versus $\kappa = 1.4766$ for the U.S.) and of a higher expected benefit of each meeting. Team leaders in Germany start with ideas of lower quality level relative to team leaders in the U.S., as can be seen from Panel (B) of Figure 12. In the U.S., team leaders who just enter the inventor labor force produce on average patents with 2.1 forward citations in a 3-year window; in Germany, the average patent quality is almost one third less (0.71 forward citations). However, as panel (C) shows, the productivity of all inventors grows faster in Germany than in the U.S. Thus, not only do German inventors interact more often, they also learn more from each interaction. Finally, over the period 1980 to 2010, the U.S. has experienced a 2.74% annual growth in real GDP, which is higher than the 2.42% growth rate in Germany.³⁹ This higher growth rate implies that the efficiency α of transforming innovations into output is higher for the US.

The data also shows that teams are on average larger in the U.S. (see Panel (D)). In line with this, there are fewer team leaders in the U.S. (54% of all inventors) than in Germany (58% of all inventors). In the U.S., superstar team leaders build larger teams – this more rigorous selection is reflected in the span of control η , which is higher in the U.S. (see Table 8).

³⁹The real GDP data used to compute these annual growth rates comes from the expenditure-side real GDP at chained PPPs (in mil. 2011US dollars), from the Penn World Table 9.0 available on www.ggd.net/pwt.

FIGURE 12: TARGETED MOMENTS FOR THE U.S. AND GERMANY



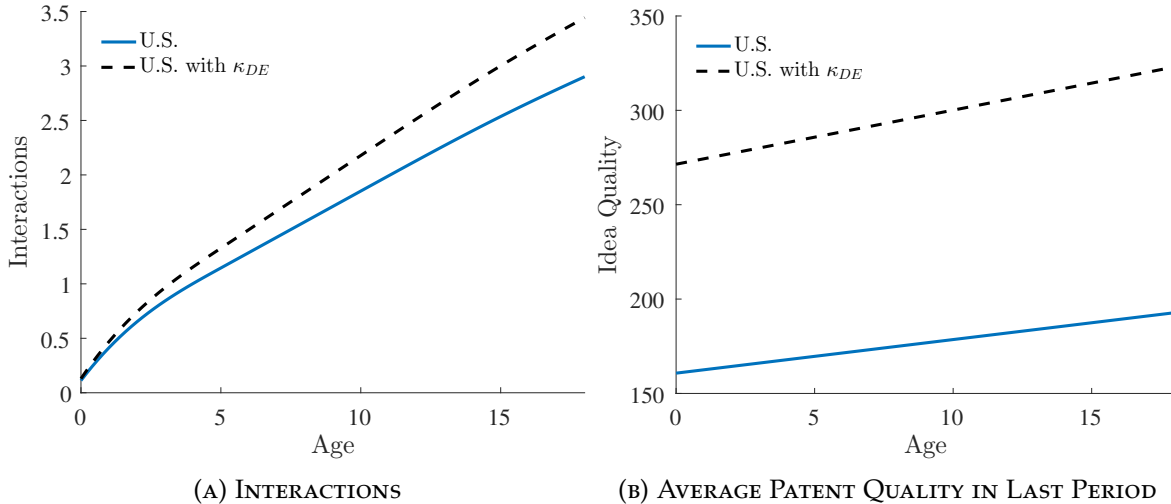
Panel (A) represents the high-quality interactions of a team leader as a function of his age (age 0 is the first year in the patent data). Panel (B): represents the de-trended idea quality $q(z, t)e^{-g_q t} = q(z, t)e^{-g(1-\eta)t}$, where $g_q = g(1-\eta)$ is the growth rate of idea quality, as a function of team leader's age. Panel (C) depicts average inventor productivity z across all inventors at a given age. It is normalized by productivity at age 0. Panel (D) shows the distribution of the team size n and Panel (E) shows the distribution of age (time in the sample). For all panels, the blue figures represents the baseline case with all the sample, the red figures the case using US data and calibration and the green figures using DE data and calibration.

The steeper lifecycle of productivity in Germany is entirely in line with the steeper profile of wages documented in [Lagakos, Moll, Porzio, Qian, and Schoellman \(2017\)](#). These authors argue that the different life cycles are most likely explained by human capital and search frictions. For instance, manual occupations have flatter profiles than high-skilled occupations. Our model provides a micro-foundation for how human capital is built (through interactions) and an explanation of how these human capital acquisition processes differ in the U.S. and in Germany. Also consistent with [Lagakos et al. \(2017\)](#), we found in Section 3 that interactions had a stronger effect on productivity in high-tech sectors.

Counterfactual analysis: lowering interaction costs in the U.S. to their level in Germany

Suppose we reduce the cost of interactions κ_{US} in the U.S. from its actual level of 1.4766 to its level in Germany $\kappa_{DE} = 1.2181$. The reduction in interaction costs increases the optimal meeting rate by 13.8% (from 0.4227 to 0.4812) and increases the growth rate from 2.74% to 2.97%. Figure 13 shows that the average number of interactions per of team leaders increases at all ages as does average patent quality (see Appendix Figure A-5 for additional results). The life cycle profile of productivity in the U.S. steepens as human capital is built more extensively at each age through increased interactions.

FIGURE 13: REDUCING INTERACTION COSTS IN THE U.S. TO THEIR LEVEL IN GERMANY



Panel (A): represents the high-quality interactions of a team leader as a function of his age (age 0 is the first year in the patent data). Panel (B) depicts average patent quality $q(z, T)$ in the last period $T = 25$ years. In each panel, the solid blue line represents the baseline case for the US. The black dashed line corresponds to the case changing the US interaction cost parameter to the lower calibrated value for Germany ($\kappa = \kappa_{DE}$).

6 Conclusion

A key input into the innovation production process is an inventor's own knowledge. That knowledge is built through interactions with others. In this paper, we build an innovation-based endogenous growth model, in which we micro-found the innovation process at the individual inventor and research team levels. An inventor acquires knowledge through interactions with others who are more

knowledgeable than him or through alternative channels such as individual discovery, experience, or learning-by-doing. The most productive (knowledgeable) inventors form research teams of varying sizes and produce ideas of heterogeneous qualities that depend on their own knowledge. Thus, we bring together the diffusion and innovation-based growth literatures; in our paper, the knowledge diffusion model nests within the innovation-based growth model.

One of our key contributions is to bring data to a heretofore largely theoretical literature; we provide a framework that brings micro data to a general equilibrium innovation model. We use brand-new panel data on millions of European inventors matched to their employers and patents. This allows us to give empirical content to typically very hard to measure concepts such as interactions, productivity or knowledge, and research teams, to discipline the key ingredients of the model, and to estimate it. We find that interactions are crucial in explaining individual productivity and aggregate growth.

In future work, our model could be used to study the effects on growth and productivity of many important policies that affect the cost of interactions in different ways. What are the effects of non-compete laws which prevent inventors from easily moving between companies? Do labor market frictions and dynamics indirectly play a role for innovation and productivity because of their impact on the ease of interactions? How do immigration policies affect the inflow of new inventors and ideas? Our theoretical and quantitative framework, coupled with this new large-scale micro data could shed light on these questions.

References

- Acemoglu, D., P. Aghion, and F. Zilibotti (2006). Distance to Frontier, Selection and Economic Growth. *Journal of the European Economic Association* 4(1), 37–74.
- Aghion, P., U. Akcigit, and P. Howitt (2014). What do we learn from schumpeterian growth theory? In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*. Amsterdam: Elsevier.
- Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica* 60(2).
- Agrawal, A., A. Galasso, and A. Oettl (2017). Roads and innovation. *Review of Economics and Statistics* 99, 417–434.
- Akcigit, U., S. Baslandze, and S. Stantcheva (2016). Taxation and the International Mobility of Inventors. *American Economic Review* 106(10), 2930–2981.
- Akcigit, U. and W. R. Kerr (2017). Growth through Heterogeneous Innovations. *Journal of Political Economy*, forthcoming.
- Alvarez, F., F. Buera, and R. Lucas (2008). Models of idea flows. *National Bureau of Economic Research Working Paper* (14135), 1–12.
- Alvarez, F., F. Buera, and R. Lucas (2017). Idea flows, economic growth and trade. *University of Chicago Working Paper*.
- Azoulay, P., C. Fons-Rosen, and J. G. Zivin (2015). Does science advance one funeral at a time? *National Bureau of Economic Research Working Paper* (21788).
- Benhabib, J., J. Perla, and C. Tonetti (2014). Catch-up and fall-back through innovation and imitation. *Journal of Economic Growth* 19, 1–35.

- Benhabib, J., J. Perla, and C. Tonetti (2017). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *NBER* (23095).
- Bloom, N., R. Sadun, and J. V. Reenen (2012). Americans do i.t. better: Us multinationals and the productivity miracle. *American Economic Review* 102(1), 167–201.
- Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica* 81(4), 1347–1393.
- Breschi, S., F. Lissoni, and G. Tarasconi (2016). Inventor data for research on migration and innovation: The ethnic-inv pilot database. In C. Fink and E. Miguelez (Eds.), *The International Mobility of Talent and Innovation: New Evidence and Policy Implications*. Cambridge Univ. Press.
- Buera, F. and E. Oberfield (2016). The global diffusion of ideas. *National Bureau of Economic Research Working Paper* (21844).
- Coffano, M. and G. Tarasconi (2014). Crios-patstat database: Sources, contents and access rules. *Center for Research on Innovation, Organization and Strategy, CRIOS Working Paper* (1).
- Goolsbee, A. (1998). Does r&d policy primarily benefit scientists and engineers? *American Economic Review (Papers and Proceedings)* 88(2), 298–302.
- Goolsbee, A. (2003). Investment tax subsidies and the wages of capital goods workers: To the workers go the spoils? *National Tax Journal* LVI(1), 153–166.
- Grossman, G. and E. Helpman (1991). Quality ladders in the theory of growth. *The Review of Economic Studies* 58(1), 43–61.
- Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The nber patent citation data file: Lessons, insights and methodological tools. National Bureau of Economic Research WP 8498.
- Hoisl, K. (2007). Tracing mobile inventors: The causality between inventor mobility and inventor productivity. *Research Policy* 36(5), 619–636.
- Jaravel, X., N. Petkova, and A. Bell (2017). Team-specific capital and innovation. *American Economic Review*, forthcoming.
- Jones, B. F. (2009). The burden of knowledge and the death of the renaissance man: Is innovation getting harder? *Review of Economic Studies* 76(1).
- Jones, B. F. (2010). Age and great invention. *Review of Economics and Statistics* 92(1), 1–14.
- Jones, B. F. and B. Weinberg (2011). Age Dynamics in Scientific Creativity. *Proceedings of the National Academy of Sciences* 108(47), 18855–19096.
- Jones, B. F., S. Wuchty, and B. Uzzi (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* 322(5905), 1259–1262.
- Jones, C. I. (1995). R&d-based models of economic growth. *Journal of Political Economy*, 759–784.
- Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy* 112(5), 986–1018.
- König, M. D., J. Lorenz, and F. Zilibotti (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics* 11(3), 1053–1102.
- Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica* 65(6), 1389–1419.
- Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2017). Life-Cycle wage Growth Across Countries. *Journal of Political Economy*, forthcoming.

- Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018). Life-Cycle Human Capital Accumulation Across Countries: Lessons From U.S. Immigrants. *Journal of Human Capital*, forthcoming.
- Lemieux, T. (2006). The mincer equation thirty years after schooling, experience, and earnings. In S. Grossbard-Schechtman (Ed.), *Jacob Mincer, A Pioneer of Modern Labor Economics*, pp. 127–145. Springer.
- 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. and B. Moll (2014). Knowledge growth and the allocation of time. *Journal of Political Economy* 122(1), 1–51.
- Lucas, R. E. J. (1978). On the size distribution of business firms. *Bell Journal of Economics*, 508–523.
- Lucas, R. J. (2009). Ideas and growth. *Economica* 76, 1–19.
- Luttmer, E. G. (2007). Selection, Growth, and the Size Distribution of Firms. *Quarterly Journal of Economics* 122(3), 1103–1144.
- Luttmer, E. G. J. (2012). Eventually, noise and imitation implies balanced growth. *Federal Reserve Bank of Minneapolis Working Paper* (699).
- Luttmer, E. G. J. (2015). An assignment model of knowledge diffusion and income inequality. *Federal Reserve Bank of Minneapolis Working Paper* (509).
- Martinez, C. (2010). Insight into different types of patent families.
- Nakajima, R., R. Tamura, and N. Hanaki (2010). The effect of collaboration network on inventors' job match, productivity and tenure. *Labour Economics* 17(4), 723–734.
- Pakes, A. S. (1986). Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica* 54(4), 755–84.
- Perla, J. and C. Tonetti (2014). Equilibrium imitation and growth. *Journal of Political Economy* 122(1), 52–76.
- Perla, J., C. Tonetti, and M. E. Waugh (2015, January). Equilibrium technology diffusion, trade, and growth. Working Paper 20881, National Bureau of Economic Research.
- Romer, P. (1990). Endogenous technological change. *Journal of Political Economy* 98(5), 71–102.
- Schankerman, M. and A. Pakes (1986). Estimates of the value of patent rights in european countries during the post-1950 period. *Economic Journal* 96(384), 1052–1076.
- Staley, M. (2011). Growth and the diffusion of ideas. *Journal of Mathematical Economics* 47, 470–478.
- Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *Rand Journal of Economics*, 172–187.
- Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in the production of knowledge. *Science* 316(5827), 1036–1039.

Appendix

A-1 Proofs from Section 2

Value function of final good producers

We here show that the value function of any final good producer is linear in A_i . With some probability x (which we do not need to specify here), the final good producer will purchase an idea of quality q for a total price $P = pq$ from the research teams. The value of a final good producer satisfies:

$$rV_i(t) - \dot{V}_i(t) = \Pi_i(t) + x\mathbb{E}(V(A_i + q) - V(A_i) - P) \quad (\text{A-1})$$

where the expectation is over the possible values of q . Assume that on a balanced growth path, the value V grows at rate g_V so that $\dot{V}_i(t) = g_V V_i(t)$. The wage $w_u(t)$ grows at rate g_w and \bar{A} grows at rate $g_{\bar{A}}$.

Profits are: $\Pi_i(t) = \alpha(1 - \alpha)^{\frac{1-\alpha}{\alpha}} w_u(t)^{\frac{\alpha-1}{\alpha}} A_i(t)$.

Let us conjecture that the value function takes the form:

$$V(A_i) = \hat{v} A_i w_u^{\frac{\alpha-1}{\alpha}} + \tilde{v} \bar{A}^\alpha \quad (\text{A-2})$$

for some coefficients \hat{v} and \tilde{v} .

Under this conjecture, the left hand side of A-1 becomes:

$$r\hat{v} A_i w_u^{\frac{\alpha-1}{\alpha}} + r\tilde{v} \bar{A}^\alpha - \left(\frac{\alpha-1}{\alpha} g_w \right) \hat{v} A_i w_u^{\frac{\alpha-1}{\alpha}} - \alpha g_{\bar{A}} \tilde{v} \bar{A}^\alpha$$

Under the assumptions in the main text, the total price for an idea of quality q is $P = \beta \frac{vq}{\bar{A}(t)^{1-\alpha}}$, or, using the formula for the wage, $P = \tilde{\Omega} q w_u^{\frac{\alpha-1}{\alpha}}$ where $\tilde{\Omega}$ is a constant. The right hand side of (A-1) is then:

$$\alpha(1 - \alpha)^{\frac{1-\alpha}{\alpha}} w_u(t)^{\frac{\alpha-1}{\alpha}} A_i(t) + x\mathbb{E}_q \left[\hat{v} q w_u^{\frac{\alpha-1}{\alpha}} - \tilde{\Omega} q w_u^{\frac{\alpha-1}{\alpha}} \right]$$

Equating the coefficients on all terms in $A_i(t)$ on the left hand side and right hand side we get:

$$\hat{v} = \frac{\alpha(1 - \alpha)^{\frac{1-\alpha}{\alpha}}}{r - \left(\frac{\alpha-1}{\alpha} g_w \right)}$$

Note that:

$$x\mathbb{E}_q \left[\hat{v} q w_u^{\frac{\alpha-1}{\alpha}} - \tilde{\Omega} q w_u^{\frac{\alpha-1}{\alpha}} \right] = x [\hat{v} - \tilde{\Omega}] \bar{q} w_u^{\frac{\alpha-1}{\alpha}}$$

Note that from the formula for the wage: $\frac{1}{1-\alpha} w_u(t) = \bar{A}(t)^\alpha$ for some constant. In addition, $\bar{q} g_{\bar{A}} \bar{A}$.

Equating all remaining terms and substituting for \bar{A} yields:

$$\begin{aligned} r\tilde{v}\frac{1}{1-\alpha}w_u - \alpha g_{\bar{A}}\tilde{v}\frac{1}{1-\alpha}w_u &= x[\hat{v} - \tilde{\Omega}]g_{\bar{A}}\left(\frac{1}{1-\alpha}\right)^{\frac{1}{\alpha}}w_u^{\frac{1}{\alpha}}w_u^{\frac{\alpha-1}{\alpha}} \\ \tilde{v} &= \frac{x[\hat{v} - \tilde{\Omega}]g_{\bar{A}}}{(r - \alpha g_{\bar{A}})(1-\alpha)^{\frac{1-\alpha}{\alpha}}} \end{aligned}$$

The value function thus takes the shape conjectured in (A-2) and is linear in $A_i(t)$.

We can further rewrite the value function as a function of own productivity $A_i(t)$ and aggregate productivity $\bar{A}(t)$ only. Using that $w_u(t)^{\frac{\alpha-1}{\alpha}} = \frac{1}{\bar{A}^{1-\alpha}}(1-\alpha)^{\frac{\alpha-1}{\alpha}}$, we get that: $V(A_i) = \frac{\alpha(1-\alpha)^{\frac{1-\alpha}{\alpha}}}{r - (\frac{\alpha-1}{\alpha}g_w)}A_i\frac{1}{\bar{A}^{1-\alpha}}(1-\alpha)^{\frac{\alpha-1}{\alpha}} + \tilde{v}\bar{A}^\alpha$, which simplifies to:

$$V(A_i) = \frac{\alpha}{r + (\frac{1-\alpha}{\alpha}g_w)}\frac{A_i}{\bar{A}^{1-\alpha}} + \tilde{v}\bar{A}^\alpha = v\frac{A_i}{\bar{A}^{1-\alpha}} + \tilde{v}\bar{A}^\alpha \quad \text{with:} \quad v = \frac{\alpha}{r + \frac{1-\alpha}{\alpha}g_w}$$

This means that the marginal return to purchasing more quality ($\frac{dV(A_i)}{A_i}$) is the same across all final good producers, regardless of their starting level of A_i . This confirms the claim in the main text.

Proof of proposition 1

Proof. First lets show

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1/\theta}} = k > 0$$

Recall that the system of BGP equations consists of equations (13) and (14). Using equation (14),

$$\frac{1 - \Phi(x)}{x^{-1/\theta}} = \frac{\int_0^\infty (1 - \Gamma(x, s)) d\Omega(s)}{x^{-1/\theta}}.$$

This equation implies it is sufficient to show, $1 - \Gamma(x, s) > 1 - \Gamma(xe^{gs}, 0)$, $\forall s$, since taking the limit on both sides,

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1/\theta}} = \lim_{x \rightarrow \infty} \frac{\int_0^\infty (1 - \Gamma(x, s)) d\Omega(s)}{x^{-1/\theta}} > \lim_{x \rightarrow \infty} \frac{1 - \Gamma(xe^{gs}, 0)}{x^{-1/\theta}} = k_0 e^{-\frac{g}{\theta}s} > 0,$$

The last inequality directly from Assumption 4 and $e^{-\frac{g}{\theta}s} > 0$, $\forall s$.

We also have that $1 - \Gamma(x, s) > 1 - \Gamma(xe^{gs}, 0) \iff \Gamma(xe^{gs}, 0) > \Gamma(x, s) \forall s$. Taking logs on both sides of this inequality,

$$\begin{aligned} \log \Gamma(xe^{gs}, 0) > \log \Gamma(x, s) &= -m_D \int_0^s (1 - \Phi(xe^{g(s-\tau)})) d\tau - m_X \int_0^s (1 - \Psi(xe^{g(s-\tau)}, \tau)) d\tau + \log \Gamma(xe^{gs}, 0) \\ \iff 0 > -m_D \int_0^s (1 - \Phi(xe^{g(s-\tau)})) d\tau &- m_X \int_0^s (1 - \Psi(xe^{g(s-\tau)}, \tau)) d\tau, \end{aligned}$$

This inequality holds whenever either $m_D > 0$ or $m_X > 0$ and the distributions are not degenerate.

Therefore we have,

$$k := \lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1/\theta}} > k_0 e^{-\frac{g}{\theta}s} > 0.$$

We now show that the growth rate is given by the formula in the proposition. Combining equations (13) and (14), we get:

$$\Phi(x) = \int_0^\infty \Gamma(xe^{gs}, 0) \left[e^{-m_D \int_0^s (1 - \Phi(xe^{g(s-\tau)})) d\tau - m_X \int_0^s (1 - \Psi(xe^{g(s-\tau)}, \tau)) d\tau} \right] d\Omega(s).$$

Taking the derivative with respect to x ,

$$\begin{aligned} \phi(x) &= \int_0^\infty \frac{d\Gamma(xe^{gs}, 0)}{dx} e^{gs} e^{-m_D \int_0^s (1 - \Phi(xe^{g(s-\tau)})) d\tau - m_X \int_0^s (1 - \Psi(xe^{g(s-\tau)}, \tau)) d\tau} d\Omega(s) \\ &+ \int_0^\infty \Gamma(xe^{gs}, 0) e^{-m_D \int_0^s (1 - \Phi(xe^{g(s-\tau)})) d\tau - m_X \int_0^s (1 - \Psi(xe^{g(s-\tau)}, \tau)) d\tau} \left[m_D \int_0^s \phi(xe^{g(s-\tau)}) e^{g(s-\tau)} d\tau \right. \\ &\left. + m_X \int_0^s \psi(xe^{g(s-\tau)}, \tau) e^{g(s-\tau)} d\tau \right] d\Omega(s). \end{aligned} \quad (\text{A-3})$$

Under assumption 4:

$$\lim_{x \rightarrow \infty} \frac{\psi(xe^{g(s-\tau)}, s)}{x^{-1/\theta-1}} = \frac{\rho(s)}{\theta} e^{-\frac{g}{\theta}(s-\tau) - g(s-\tau)}, \quad \lim_{x \rightarrow \infty} \frac{\frac{d\Gamma(xe^{gs}, 0)}{dx} e^{gs}}{x^{-1/\theta-1}} = \frac{k_0 e^{-gs/\theta}}{\theta}.$$

We already showed that $\lim_{x \rightarrow \infty} \frac{\phi(x)}{x^{-1/\theta-1}} = \frac{k}{\theta}$, which implies that $\lim_{x \rightarrow \infty} \frac{\phi(xe^{g(s-\tau)})}{x^{-1/\theta-1}} = \frac{k}{\theta} e^{-\frac{g}{\theta}(s-\tau) - g(s-\tau)}$. Hence, for x large, the growth rate has to satisfy,

$$\frac{kx^{-1/\theta-1}}{\theta} = \int_0^\infty \left[\frac{k_0 e^{-\frac{g}{\theta}s} x^{-1/\theta-1}}{\theta} + m_D \int_0^s \frac{kx^{-1/\theta-1}}{\theta} e^{-\frac{g}{\theta}(s-\tau)} d\tau + m_X \int_0^s \frac{\rho(\tau) x^{-1/\theta-1}}{\theta} e^{-\frac{g}{\theta}(s-\tau)} d\tau \right] d\Omega(s)$$

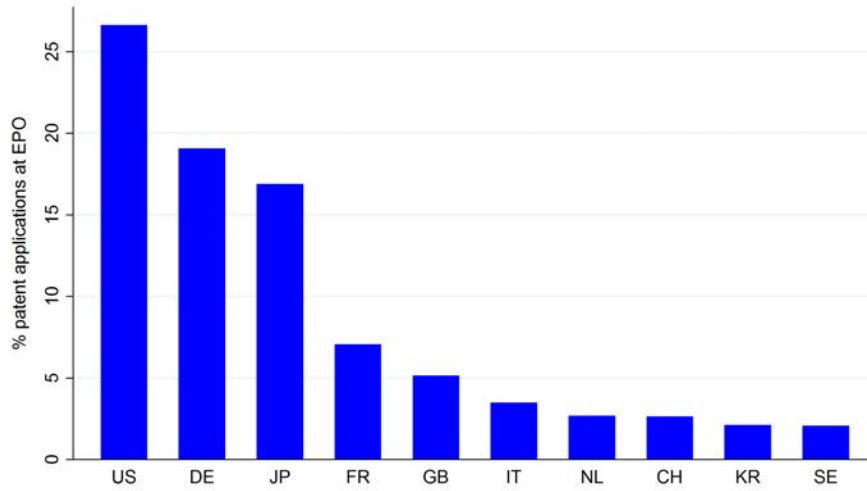
which yields the growth rate formula in the main text:

$$\frac{g}{\theta} = \frac{m_D \int_0^\infty (1 - e^{-\frac{g}{\theta}s}) d\Omega(s) + \frac{m_X g}{k\theta} \int_0^\infty \int_0^s \rho(\tau) e^{-\frac{g}{\theta}(s-\tau)} d\tau d\Omega(s)}{1 - \frac{k_0}{k} \int_0^\infty e^{-\frac{g}{\theta}s} d\Omega(s)}.$$

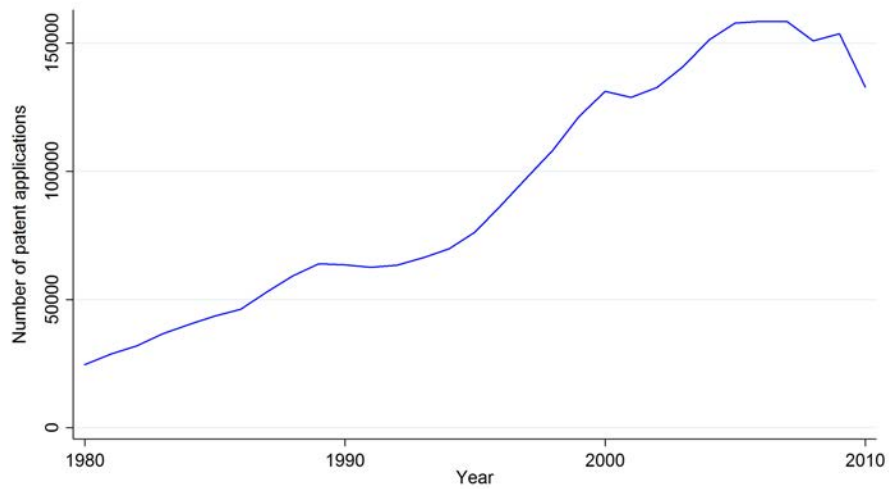
□

A-2 Additional Data Information

FIGURE A-1: APPLICATIONS AT THE EPO BY COUNTRY AND YEAR



(A) APPLICATIONS ACROSS COUNTRIES

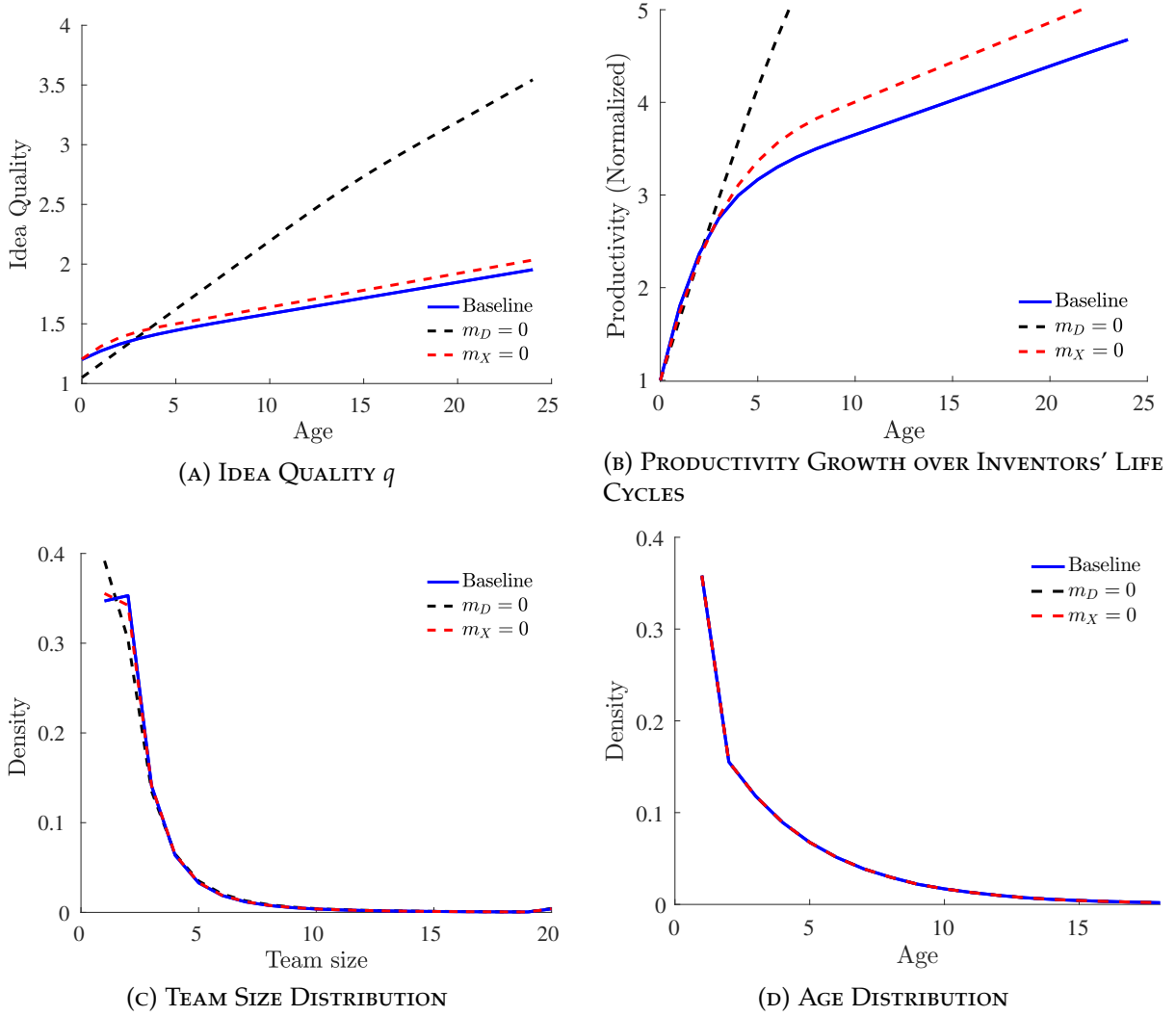


(B) APPLICATIONS BY YEAR

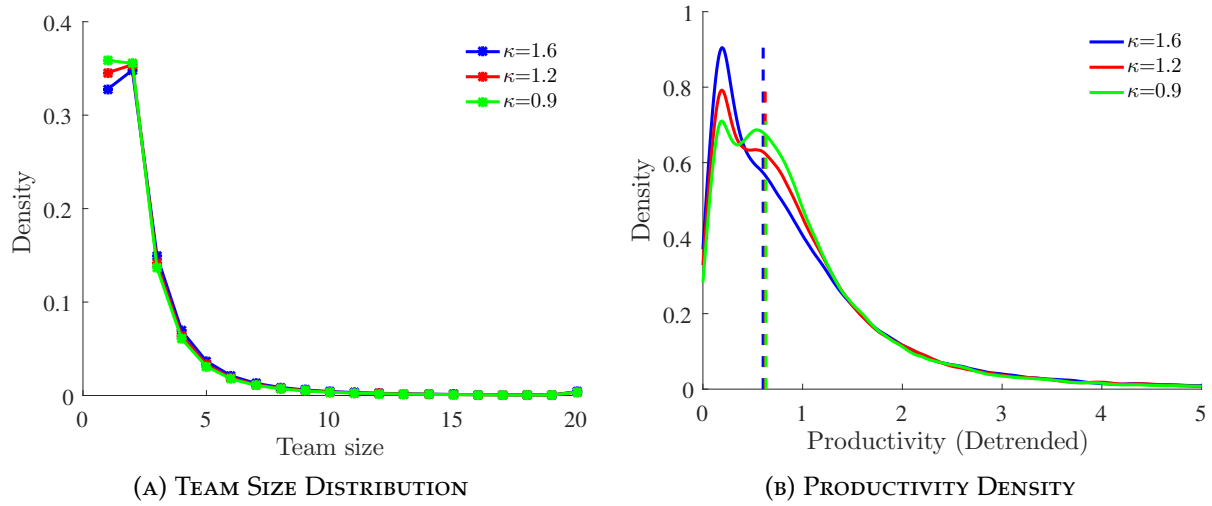
Notes: Panel (A) shows the percentage of all applications at the EPO by country. Panel (B) shows total applications at the EPO by year.

A-3 Additional Figures from Section 5

FIGURE A-2: EFFECT OF INTERACTIONS AND EXTERNAL LEARNING ON GROWTH: ADDITIONAL RESULTS

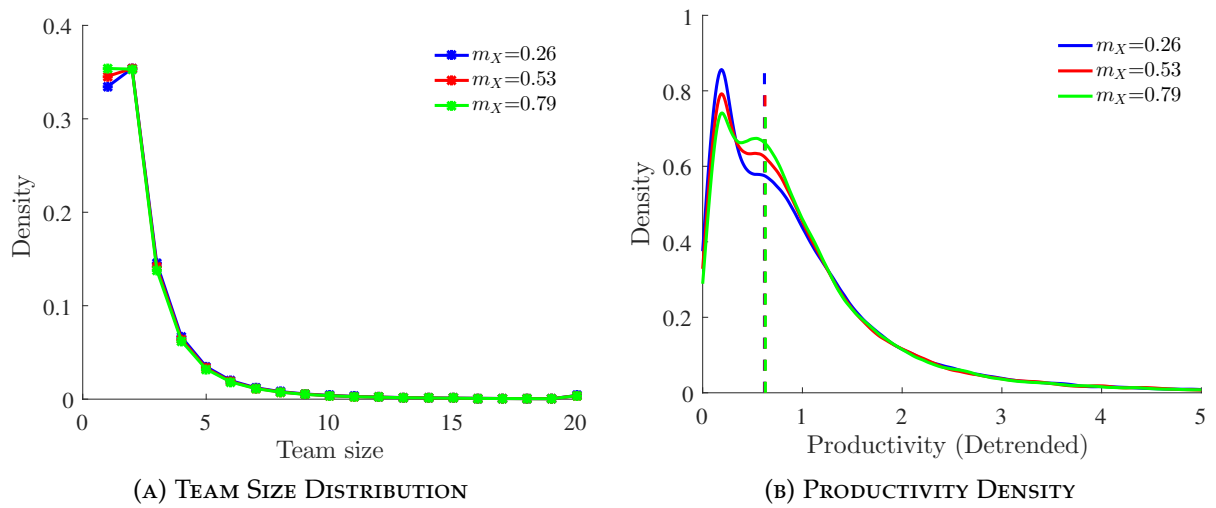


Panel (A): represents the de-trended idea quality $q(z, t)e^{-g_q t} = q(z, t)e^{-g(1-\eta)t}$, where $g_q = g(1 - \eta)$ is the growth rate of idea quality, as a function of team leader's age. Panel (B) depicts average inventor productivity z across all inventors at a given age. It is normalized by productivity at age 0. Panel (C) shows the distribution of the team size n and Panel (D) shows the distribution of age (time in the sample). In each panel, the solid blue line represents the baseline case, with both the endogenous interaction channel and the external learning channel active. The black dashed line corresponds to the case with no endogenous interactions ($m_D = 0$) and the red dashed line to the case with no external learning ($m_X = 0$).

FIGURE A-3: REDUCING INTERACTION COSTS κ : ADDITIONAL RESULTS


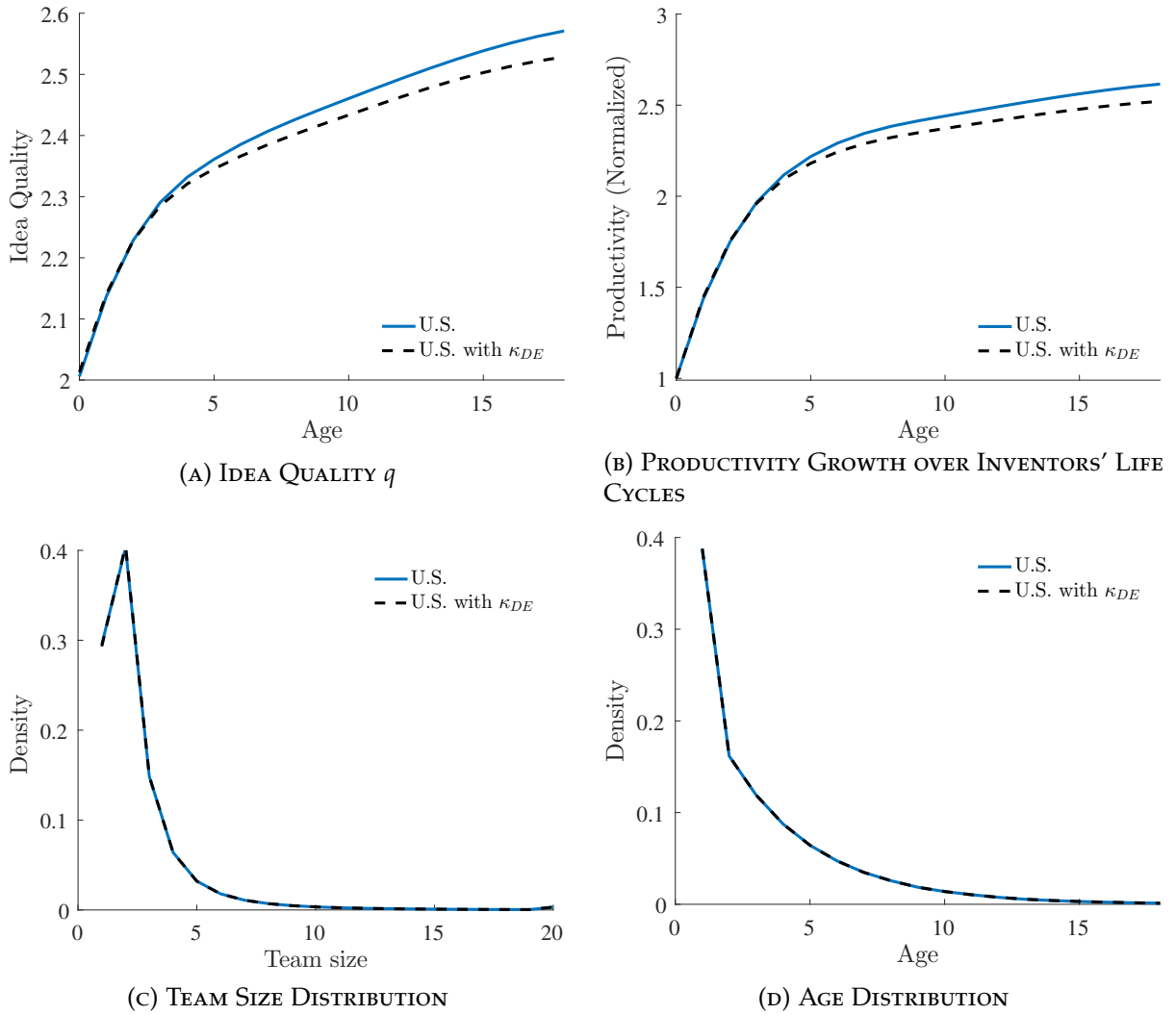
Panel (A) shows the distribution of the team size n and Panel (B): represents the distribution of de-trended idea quality $q(x) = q(z, t)e^{-g(1-\eta)t}$, for three levels of interaction costs, $\kappa = 0.9, 1.2, 1.6$.

FIGURE A-4: REDUCING THE EXTERNAL LEARNING RATE: ADDITIONAL RESULTS



Panel (A) shows the distribution of the team size n and Panel (B): represents the distribution of de-trended idea quality $q(x) = q(z, t)e^{-g(1-\eta)t}$, for three levels of external learning rate, $m_X = 0.26, 0.53, 0.79$.

FIGURE A-5: REDUCING INTERACTION COSTS κ_{US} IN THE U.S. TO GERMANY'S LEVEL



Panel (A): represents the de-trended idea quality $q(z, t)e^{-g_q t} = q(z, t)e^{-g(1-\eta)t}$, where $g_q = g(1 - \eta)$ is the growth rate of idea quality, as a function of team leader's age. Panel (B) depicts average inventor productivity z across all inventors at a given age. It is normalized by productivity at age 0. Panel (C) shows the distribution of the team size n and Panel (D) shows the distribution of age (time in the sample). In each panel, the solid blue line represents the baseline case for the US. The black dashed line corresponds to the case changing the US interaction cost parameter to the lower calibrated value for Germany ($\kappa = \kappa_{DE}$).

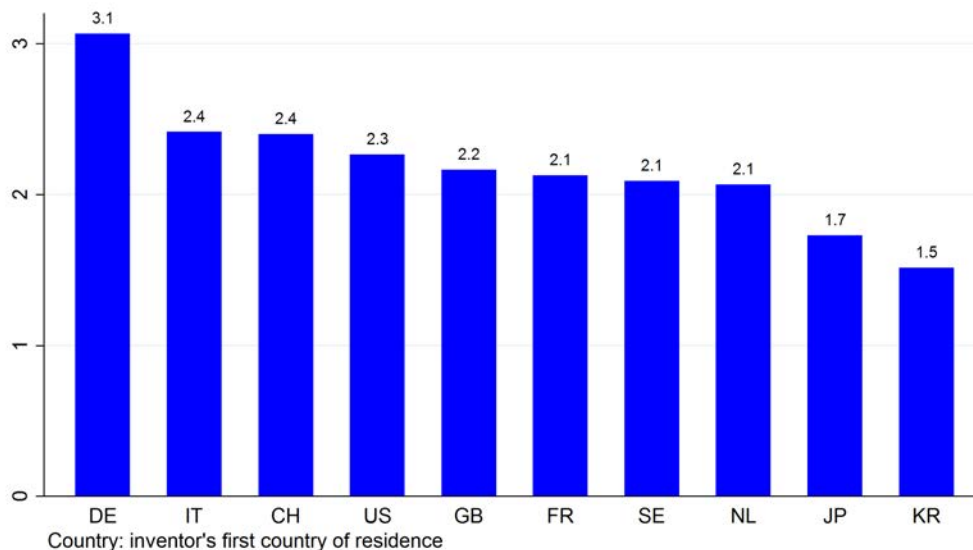
Online Appendix for “Dancing with the Stars: Innovation through Interactions”

Ufuk Akcigit Santiago Caicedo Ernest Miguez
Stefanie Stantcheva Valerio Sterzi

March 23, 2018

OA-1 Data and Variable Construction

FIGURE OA-1: AVERAGE NUMBER OF PATENTS PER INVENTOR: TOP 10 COUNTRIES



OA-1.1 The Disambiguation Algorithm

For disambiguation, we rely on the Massacrator 2.0 (Pezzoni et al. (2014)). In a nutshell, Massacrator 2.0 is a disambiguation algorithm based on a 3-step procedure (see Raffo and Lhuillery (2009), for a detailed overview of inventor disambiguation approaches): first, *cleaning/parsing* of the relevant text string (inventors' name and surname, plus addresses). Second, the *matching* stage, in where the algorithm selects pairs of inventors, from different patents, who are likely candidates to be the same person, due to homonymy or quasi-homonymy of their names. Finally, the *filtering* stage, where all the pairs of inventors' names who are likely to be the same person are compared according to additional information retrieved either from the patent documentation or external sources. This includes

TABLE OA-1: NUMBER OF PATENT APPLICATIONS WORLDWIDE AND BY SELECTED PATENT OFFICES (1995-2010)

<i>Year</i>	<i>Applications worldwide</i>	<i>Patent Families</i>	<i>Applications at EPO</i>	<i>Share EPO over applications</i>	<i>Share EPO over families</i>	<i>Applications at USPTO</i>	<i>Share USPTO over applications</i>	<i>Share USPTO over families</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1995	1,047,400	572,565	71,861	6.86%	12.55%	192,638	18.39%	33.64%
1996	1,088,400	592,413	79,104	7.27%	13.35%	194,631	17.88%	32.85%
1997	1,163,200	627,835	90,047	7.74%	14.34%	226,298	19.45%	36.04%
1998	1,214,800	665,622	102,159	8.41%	15.35%	236,149	19.44%	35.48%
1999	1,268,400	695,482	112,525	8.87%	16.18%	267,705	21.11%	38.49%
2000	1,377,400	772,496	126,947	9.22%	16.43%	321,306	23.33%	41.59%
2001	1,456,900	803,126	135,919	9.33%	16.92%	375,503	25.77%	46.76%
2002	1,443,600	806,431	134,700	9.33%	16.70%	385,722	26.72%	47.83%
2003	1,490,300	841,558	140,760	9.45%	16.73%	402,947	27.04%	47.88%
2004	1,574,400	863,716	149,644	9.50%	17.33%	453,647	28.81%	52.52%
2005	1,702,900	905,780	159,748	9.38%	17.64%	495,523	29.10%	54.71%
2006	1,791,200	929,842	166,677	9.31%	17.93%	490,291	27.37%	52.73%
2007	1,876,900	956,422	168,116	8.96%	17.58%	500,638	26.67%	52.34%
2008	1,929,200	984,022	168,860	8.75%	17.16%	488,789	25.34%	49.67%
2009	1,861,700	955,880	158,985	8.54%	16.63%	457,481	24.57%	47.86%
2010	1,996,800	1,009,086	164,431	8.23%	16.30%	478,533	23.96%	47.42%

Source: Adapted from Miguelez (2016). World Intellectual Property Indicators – 2013 edition (Wipo, 2013) and author’s calculations from PatStat.

information of common co-inventors, inventors' addresses geographical location, information on common applicants of their patents, common technological classes (according to the International Patent Classification – IPC), and the common cited prior art of the two inventors. The algorithm is calibrated against two benchmark datasets. The first one is a group of 530 academic French inventors, for whom all their patents were retrieved manually from PatStat and other sources. The second is a list of 342 faculty members of the EPFL (“Ecole Polytechnique Federale de Lausanne”), for whom again their patenting record is manually verified (see [Lissoni et al. \(2010\)](#)).

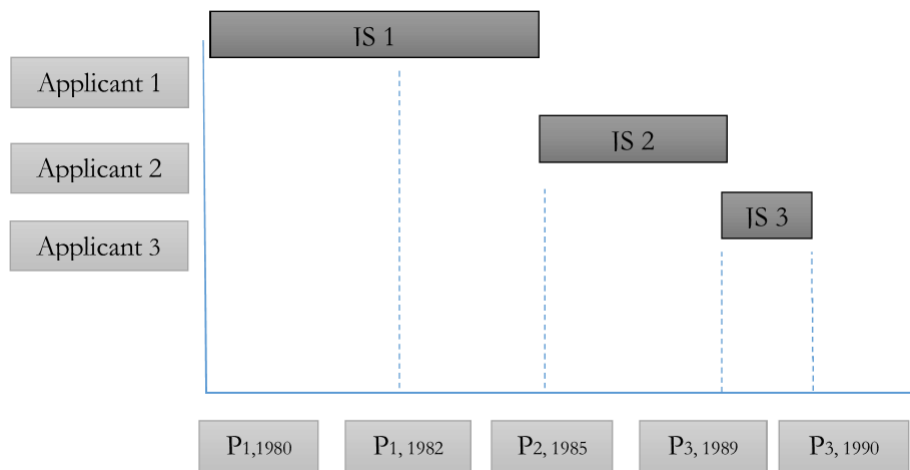
OA-1.2 Technology Fields and the International Patent Classification

The International Patent Classification (IPC) provides a hierarchical system of codes for the classification of patents and utility models according to the different areas of technology to which they pertain. This large list of technology codes can be grouped into meaningful technological fields. Here we used the classification built by the French “Observatoire des Sciences et des Techniques” (OST), which groups IPC codes into 30 meaningful fields, which are further regrouped into 7 board fields (OST, 2008). Table OA-2 shows the OST Technological classification used in the paper, as well as the distribution of patent application across fields, for the whole period under analysis.

OA-1.3 Inventors' Employment Histories

We assume that a given inventor i is active in a firm f_1 from the year of his first patent in firm f_1 to either (1) the year before the first patent invented in firm f_2 or (2) to the year of the last patent in firm f_1 in case he does not have any further patent afterwards. Figure OA-2 depicts a hypothetical example of employment history construction. In that figure the inventor is listed in five patent applications from three different applicants and has thus three job spells (JS). JS 1 lasts from 1980, the year of the first patent in firm 1 to 1982, the year before the first patent in firm 2. JS 2 lasts from 1983, the year for the first patent in firm 2, to 1988, the year before the first patent in firm 3. Finally, in 1989 and 1990 the inventor is in firm 3.

FIGURE OA-2: EXAMPLE OF EMPLOYMENT HISTORY CONSTRUCTION



We make use of the HAN database (OECD HAN database, January 2014) of Harmonized Applicants' Names in order to track assignees over time.

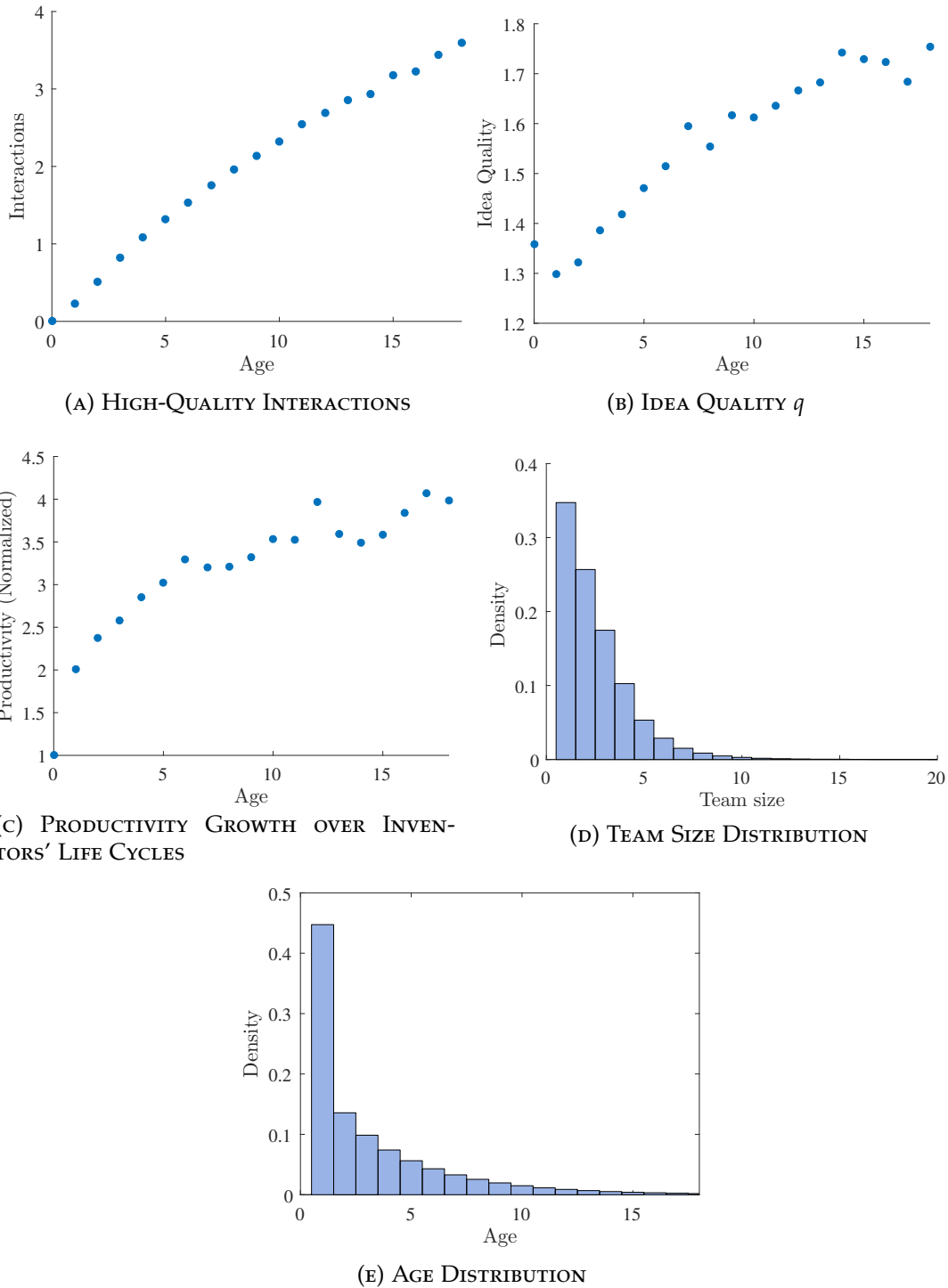
TABLE OA-2: OST TECHNOLOGICAL RECLASSIFICATION

30-code	30-Name	# patents	%	7-Code	7-Name
1	Electrical engineering	251,175	5.88	1	Electrical engineering; Electronics
2	Audiovisual technology	165,344	3.87	1	Electrical engineering; Electronics
3	Telecommunications	256,220	6.00	1	Electrical engineering; Electronics
4	Information technology	220,204	5.15	1	Electrical engineering; Electronics
5	Semiconductors	107,972	2.53	1	Electrical engineering; Electronics
6	Optics	136,551	3.20	2	Instruments
7	Control technology	304,393	7.12	2	Instruments
8	Medical engineering	195,726	4.58	2	Instruments
9	Nuclear technology	18,774	0.44	2	Instruments
10	Organic chemistry	195,457	4.58	3	Chemicals; Materials
11	Macromolecular chemistry	161,291	3.78	3	Chemicals; Materials
12	Basic chemistry	146,652	3.43	3	Chemicals; Materials
13	Surface technology	95,226	2.23	3	Chemicals; Materials
14	Materials; Metallurgy	101,127	2.37	3	Chemicals; Materials
15	Biotechnologies	151,143	3.54	4	Pharmaceuticals; Biotechnology
16	Pharmaceuticals; Cosmetics	252,034	5.90	4	Pharmaceuticals; Biotechnology
17	Agricultural and food products	45,034	1.05	4	Pharmaceuticals; Biotechnology
18	Mechanical engineering (excl. Transport)	159,877	3.74	5	Industrial processes
19	Handling; Printing	178,495	4.18	5	Industrial processes
20	Agricultural and food apparatuses	168,358	3.94	5	Industrial processes
21	Materials processing	38,158	0.89	5	Industrial processes
22	Environmental technologies	42,269	0.99	5	Industrial processes
23	Machine tools	94,044	2.20	6	Mechanical eng.; Machines; Transport
24	Engines; Pumps; Turbines	109,656	2.57	6	Mechanical eng.; Machines; Transport
25	Thermal processes	63,544	1.49	6	Mechanical eng.; Machines; Transport
26	Mechanical elements	144,345	3.38	6	Mechanical eng.; Machines; Transport
27	Transport technology	175,345	4.10	6	Mechanical eng.; Machines; Transport
28	Space technology; Weapons	15,523	0.36	6	Mechanical eng.; Machines; Transport
29	Consumer goods	166,739	3.90	7	Consumer goods; Civil engineering
30	Civil engineering	111,541	2.61	7	Consumer goods; Civil engineering

Notes: Patents assigned to more than one category are counted more than once.

The OECD HAN database, January 2014 edition, combines information on 317,927 applicants' names listed in 2,776,774 patent applications. Of them, 2,519,069 (90.7%) have only one applicant, while 9.3% patents list more than one. For the remaining 178,281 patent applications with no link to the HAN database, harmonized applicants' names are extracted from the CRIOS-PatStat database (Coffano and Tarasconi (2014)).

FIGURE OA-3: CALIBRATED MOMENTS



Panel (A) represents the high-quality interactions of a team leader as a function of his age (age 0 is the first year in the patent data). Panel (B): represents the idea quality measured as the 3-year citations of patents produced by a team, as a function of team leader's age. Panel (C) depicts average inventor productivity (his citation weighted patent stock, $P_{i,t}$ on the main text) across all inventors at a given age. It is normalized by productivity at age 0. Panel (D) shows the distribution of the team size n and Panel (E) shows the distribution of age (time in the sample).

OA-2 Computational Algorithm

Computational algorithm

In this appendix we describe the details of the simulation of the model. We simulate the model for $n_{sim} = 1,000,000$ individuals for $T = 25$ years, and subdivide each period using a step size of $\Delta t = 0.2$.

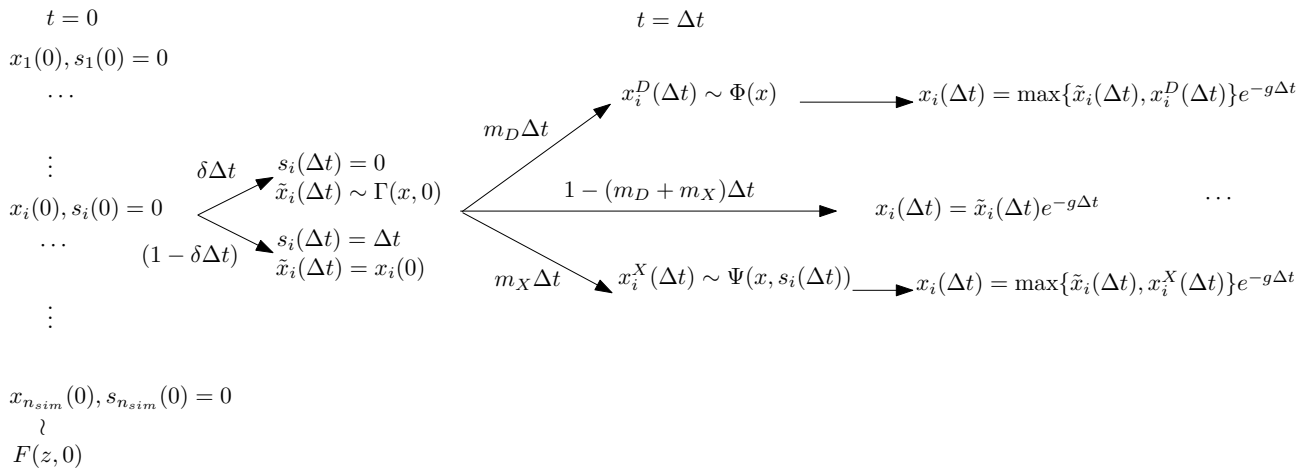
We assume individuals initially draw from a Log-logistic distribution $\Gamma(x, 0) = \frac{1}{1+k_0x^{-1/\theta}}$. To do this, we first simulate n_{sim} draws from a uniform distribution $(U_1, \dots, U_{n_{sim}})$ and then apply the inverse distribution $F^{-1}(z, 0)(U)$ to produce a simulated vector for the initial period $x(0) = (x_1(0), \dots, x_{n_{sim}}(0))$, where $F(z, 0) = \frac{1}{1+\lambda z^{-1/\theta}}$. We use this same procedure for all simulations involving a known distribution.

Given the initial vector $x(0)$ we simulate the paths for each of the n_{sim} individuals. With probability $m_D \Delta t$ an individual meets another. We simulate these meetings using the empirical distribution, fitted with a Pareto tail.¹ Individuals who meet with others, choose the maximum between their productivity $x_i(t-1)$ and the one they draw from the simulated distribution $x_i^D(t) \sim \Phi(x)$. So their new productivity in period t is $z_i(t) := \max\{x_i(t-1), x_i^D(t)\}$.

Each period, individuals also draw from the age-dependent external distributions $\Psi(x, s)$ with probability $m_X \Delta t$. We assume the external distributions are Log-logistic with location parameter $\rho(s) = \rho s^\nu$, $\Psi(x, s) = \frac{1}{1+\rho s^\nu x^{-1/\theta}}$. Individuals who draw from the external source distribution obtain the maximum productivity from either their last period productivity or the one they draw from the external distribution $x_i^X(t) \sim \Psi(x, s)$.

Hence, in each period, individuals are characterized by two states: their age $s_i(t)$ and their productivity $z_i(t)$. We compute the growth in productivity using the formula: $g_z(t) = \frac{\mathbb{E}[z(t)] - \mathbb{E}[z(t-\Delta)]}{\Delta t}$. We de-trend the gross productivity vectors using this growth rate, $x_i(t) = e^{-g_z \Delta t} z_i(t) \forall i = \{1, \dots, n_{sim}\}$. Figure OA-4 illustrates the simulation procedure.

FIGURE OA-4: SIMULATION PROCEDURE



We save the information of whether an individual interacted with another individual or drew an idea from the external source of ideas and whether the draw was successful (i.e., whether the individual learned something) or not.

Once we have the simulated vectors of productivity $x(t)$ and age $s(t) = [s_1(t), \dots, s_{n_{sim}}(t)]$, and have registered all endogenous and external interactions, we compute each of the targeted moments

¹We use MATLAB program *paretotails.m*. For more information see <https://www.mathworks.com/help/stats/examples/nonparametric-estimates-of-cumulative-distribution-functions-and-their-inverses.html>.

and minimize the score function:

$$\mathcal{L}(\chi) := \min \sum_{i=1}^{86} \omega_i \frac{|\text{model}(i) - \text{data}(i)|}{\frac{1}{2}|\text{model}(i)| + \frac{1}{2}|\text{data}(i)|}.$$

Modeling R&D subsidies

One of the things we check in the main text is the goodness of fit of our model for non-targeted moments, two of which are elasticities of wages and team sizes to R&R tax credits. We now explain how we can compute these in the model.

In the model, R&D subsidies or tax credits can be modeled as a subsidy on the teams' research wage costs. Specifically a fraction ς of the wages paid to team members is covered by the government. The optimization problem of a team leader with productivity z is now,

$$\max_{n \geq 0} p(t)z^{1-\eta}n^\eta - (1-\varsigma)w(t)n.$$

The number of team members hired is $n(z, t) = \left(\frac{p(t)\eta}{(1-\varsigma)w(t)}\right)^{\frac{1}{1-\eta}} z$, the idea quality produced is $q(z, t) = \left(\frac{p(t)\eta}{(1-\varsigma)w(t)}\right)^{\frac{\eta}{1-\eta}} z$, and profits are $\pi(z, t) = p(t) \left(\frac{p(t)\eta}{(1-\varsigma)w(t)}\right)^{\frac{\eta}{1-\eta}} (1-\eta)z$. Note for a given $w(t)$, higher subsidies increase team size, quality of patents and team leaders profits.

The occupational choice is also affected and in turn changes equilibrium wages and the fraction of individuals that become team leaders. As before, an individual of productivity z chooses to be a team leader if the profits he makes are larger than the wage he receives as a worker. In this case, the cutoff $\bar{z}(t)$, determined by the individual who is indifferent between becoming a team leader and a worker, is given by:

$$w(t) = \pi(\bar{z}(t), t) = \left(\frac{\eta}{(1-\varsigma)w(t)}\right)^{\frac{\eta}{1-\eta}} (1-\eta)\bar{z}(t)p(t)^{1/(1-\eta)} \Rightarrow \frac{w(t)}{p(t)} = \frac{(1-\eta)^{1-\eta}\eta^\eta}{(1-\varsigma)^\eta} \bar{z}(t)^{1-\eta}. \quad (\text{OA-1})$$

The effect of the subsidy on wages depends on the team member share in the idea production. If η is higher, output is more elastic with respect to team members, and team members would accrue a larger share of the subsidy.

We solve for the cutoff $\bar{z}(t)$ and wages $w(t)$ using equation (OA-1) and the labor market clearing condition, to yield:

$$(1-\varsigma)(1-\eta)\bar{z}(t)F(\bar{z}(t), t) = \eta \int_{\bar{z}(t)}^{\infty} zf(z, t)dz. \quad (\text{OA-2})$$

A subsidy $\varsigma > 0$ increases the aggregate demand for team members, and, thus the share of inventors who become team members. We can show this formally, by implicitly differentiating equation (OA-2), which yields:

$$\frac{d\bar{z}}{d\varsigma} = \frac{(1-\varsigma)(1-\eta)(F(\bar{z}) + \bar{z}f(\bar{z})) + \eta\bar{z}f(\bar{z})}{(1-\eta)\bar{z}F(\bar{z})} > 0$$

and which is positive.

Using the characterization from above we show average team size is increasing in the subsidy ς .

Proposition 1. *Average team size is increasing in ς*

Proof. We can rewrite the optimal team size using the market clearing condition as:

$$n(z) = \frac{F(\bar{z}, t)}{\int_{\bar{z}}^{\infty} z f(z, t) dz} z.$$

Using this expression we compute the average team size,

$$\bar{n}(t) := \frac{\int_{\bar{z}(t)}^{\infty} n(z, t) f(z, t)}{1 - F(\bar{z}, t)} = \frac{F(\bar{z}, t)}{1 - F(\bar{z}, t)}$$

Since $F(\cdot, t)$ is monotone increasing, the average team size is increasing in \bar{z} , which, as we just saw, is itself increasing in ζ .

□

References

- Coffano, Monica and Gianluca Tarasconi**, “CRIOS-Patstat Database: Sources, Contents and Access Rules,” *Center for Research on Innovation, Organization and Strategy, CRIOS Working Paper*, 2014, (1).
- Lissoni, Francesco, Monica Coffano, Andrea Maurino, Michele Pezzoni, and Gianluca Tarasconi**, “APE-INV’s “Name Game” Algorithm Challenge: A Guideline for Benchmark Data Analysis & Reporting,” *Version*, 2010, 1, 875–889.
- Migueluez, Ernest**, “Inventor Diasporas and the Internationalization of Technology,” *World Bank Economic Review*, 2016. forthcoming.
- Pezzoni, Michele, Francesco Lissoni, and Gianluca Tarasconi**, “How to Kill Inventors: Testing the Massacrator© Algorithm for Inventor Disambiguation,” *Scientometrics*, 2014, 101 (1), 477–504.
- Raffo, Julio and Stéphane Lhuillery**, “How to Play the “Names Game”: Patent Retrieval Comparing Different Heuristics,” *Research Policy*, 2009, 38 (10), 1617–1627.