Matching on the Global Inventor Firm Labor Market

Richard Bräuer

July 6, 2023

Abstract

I analyze the matching of firms and inventors and the patent (citation) arrival rate of the resulting matches as a potential driver of slowing technology growth. I document a global trend towards increased assortative matching and declining inventor mobility to low productivity firms despite a largely constant patent invention function. To arrive at these results, I further develop empirical strategies used in the search and matching labor market literature to account for inventor teams and adapt these estimators to the pecularities of the PATSTAT patent data from 1974-2012, which I use as an employer-employee data set.

1 Introduction

I analyze a potential cause of the productivity growth slowdown in advanced economies, namely a change in the patterns of allocation of research talent across firms. Using the PATSTAT patent data base, I find evidence of an increasing tendency for assortative matching from 1974-2012 on a global scale. Good inventors increasingly match with firms with high quality research departments. I estimate the expected patent and citation arrival rate for matches of inventors or inventor teams and firms and from that I extract inventors' skill and the quality of firms' research departments. I demonstrate that the increased assortative matching has not led to substantial increases in patent invention rates because the patent invention function does not put high premia on assortative matching.

The function itself has remained roughly stable throughout the 40 years under consideration, so its changes cannot explain the changed matching behaviour.

These findings constitute the main contribution of my paper. To reach them, I make several methodological contributions: I extend the literature on search and matching labor markets (Abowd et al., 1999; Card et al., 2013; Hagedorn et al., 2017; Bonhomme et al., 2017) to a setting of inventors and firms, where data on inventors is sparse, output arrives in discrete units (patents) and employment spells are only observed if they are successful and actually produce patents. To do this, I analytically derive an estimator built on the Hagedorn et al. (2017) framework and demonstrate its reliability on simulated data. Additionally, I demonstrate how to use this estimator to rank both single inventors and teams without assuming a specific production function. I plan to include results for inventor teams and team formation in a future version of the paper.

To analyze whether inventor mobility is a potential channel of technology diffusion, I link patents to firm level data (AMADEUS 2000-2010) and find that the share of inventors leaving high revenue productivity firms for low productivity firms declines. Firms' own patents and an inflow of inventors from other firms increase productivity and boost profits. These findings suggest that increased assortative matching of inventors to firms is a plausible driver of the slowdown in technology diffusion. Conversely, since any match a firm secures with the inventors of new technologies is sticky, matches serve as natural imitation protection: Firms themselves report that staff retention and technological lead are their most important strategies for securing the profits from their inventions. Comparatively, patents only play a minor role (Harhoff, 1997). Yet, while some work exists on which inventors match with which firms (Pearce, 2019), matching between skill levels is understudied: To my knowledge, I present the first application of labor market matching estimators that can extract worker skill and firm quality from administrative patent data to inventor-firm relations. In other settings, however, the literature on labor market matching has already documented increased assortative matching (Abowd et al., 1999; Card et al., 2013; Andrews et al., 2008; Hagedorn et al., 2017; Bonhomme et al., 2017).

Rising assortative matching of skilled inventors to high quality firms and slowing technology diffusion might explain a substantial part of the productivity growth slowdown of the last decades: The literature on the economy during the productivity growth slowdown has documented declining labor shares (Autor et al., 2017), increased profit shares (**?**) and increased markups (De Loecker and Eeckhout, 2017). Rising market concentration is often seen as a result of technological leadership by superstar firms (Autor et al., 2017). Theoretically, slowing technology diffusion can explain productivity growth decreases and increased concentration simultaneously (**?**Bresnahan and Trajtenberg, 1995). Empirically, diffusion slowdowns have indeed been linked to productivity growth declines with diverse empirical strategies (Gal, 2017; Comin and Mestieri, 2010). Akcigit and Ates (2019) combine the two approaches and calibrate a standard endogenous growth model to show that declining technology diffusion can fit slowing productivity growth, rising markups and other trends in the US economy in the past decades.

I use the PATSTAT data base provided by the EPO because of its global coverage.[1] PATSTAT contains inventor and firm names and rich information on the content of the patent, up to the original document. PATSTAT can be used as a matched employer-employee data set after extensive data treatment, for which I improve upon Magerman et al. (2006) and Peeters et al. (2010). The final data contains information on the output of each inventor-team-firm match in each year. However, only pairs contributing to at least one patent are observed. I account for this truncation by estimating the Poisson rate of patenting for each employment spell via maximum likelihood estimation. This yields the probability to observe any given employment spell and get an unbiased estimate of its duration and expected number of patents per year.

The literature usually ignores teamwork among workers and what kind of projects teams might undertake. E.g. neither Hagedorn et al. (2017), nor Bonhomme et al. (2017) or Abowd et al. (1999) attempt to account for teams' effects in production. In the following discussion, I show which biases are introduced in the estimation because of this and how to rectify them by generalizing the model

[1] I exclude communist countries before 1988 since both the function and size of firms are not comparable to market economies.

of Hagedorn et al. (2017) to encompass worker teams. Teams are not usually tracked in standard employer-employee data, which might explain this oversight in the literature, yet patent data does not have this limitation.

Transferring these estimators to a patent setting also comes with a major challenge: The output of a match is measured lumpily, with patents and their citations. This causes nonnormal measurement error with an expectation different from zero. The overall strategy to deal with this is first to estimate a distribution of possible true parameters for every observed match via maximum likelihood, second to rank these hypothetical matches from all data points and third to infer the distributions of interest via Bayesian updating.

2 Data

2.1 The PATSTAT Data

Patent data from across the world gathered in the PATSTAT database forms the basis of my empirical strategy. This data contains the filing date of any patent application, a description of the technology and the names of firms and inventors involved. For some participating countries, the data starts in 1850, however, coverage pre-WW2 is generally low. Patents from some countries are only available from a later date onwards: E.g., Japan enters the database in the mid-seventies. Around the same time, coverage rates improve in general and the data can give a reliable picture of worldwide patent activity.

The following graph shows the number of patents over time for selected countries. Note that the stable or shrinking number of national patents for EU countries is offset by a large increase in EU-wide EPO applications.

Peruzzi et al. (2014) provide a PATSTAT-AMADEUS link, which I use to relate my inventor labor market data to actual economic outcomes like profits or firm TFPR. Apart from string matching firm names, they use the other variables in AMADEUS to predict which firms are more likely active in PATSTAT. Their technique allows to merge around 140.000 companies to the PATSTAT database.

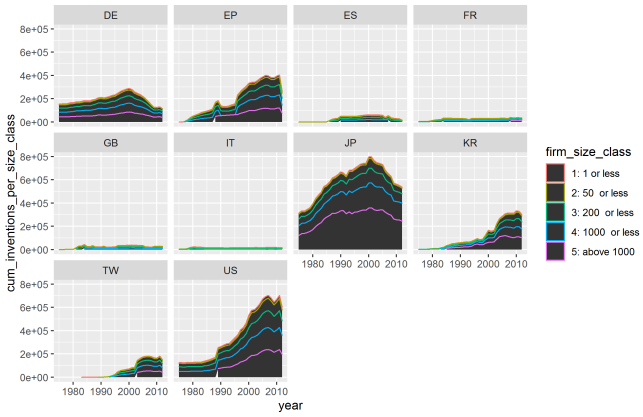

Figure 1: Patent applications per patenting authority; DE = Germany; EP = European Patent Office; ES = Spain; FR = France; GB = Great Britain; IT = Italy; JP = Japan; KR = Republic of Korea; TW = Taiwan; US = USA. Source: PATSTAT

Using PATSTAT as an employer-employee data set entails challenges as well as advantages over commonly used social security data. The following gives a brief overview over the main opportunities and problems when using this data, compared to standard social security employer-employee data sets. A detailed description of the necessary data treatment steps can be found in Appendix A.

The first advantage of PATSTAT is that it is much richer than social security data regarding the type of work that inventors do: Patent applications contain descriptions of the technology and a list of co-inventors. In employer-employee settings, all workers are usually treated as perfect substitutes, only differentiated by the skill with which they produce. I improve upon this treatment by using the IPC 4-digit technology codes assigned to every patent: I contract the technology space into 56 technology clusters, comprised of IPC classes that often appear jointly on patent applications. I use the clustering algorithm of Pons and Latapy (2005). Throughout the rest of the paper, technological clusters will differentiate inventors horizontally, i.e. there will be separate labor markets and rankings in each technology cluster. Inventors are assigned to their main technology cluster. Inventors' patent portfolios are largely within one technology cluster: The most important technology cluster of an inventor covers 88% of his patents on average and 56% of inventors only patent within one cluster.

Second, PATSTAT's patents are a direct measure of the output of a match between firm and inventor, which is usually not available from employer-employee data. Since this data is normally derived from social security declarations, output is approximated using wage information. However, the wage also contains the bargaining position of both parties, which makes it difficult to extract match production.

Third, most workers, including inventors, work in teams. Such worker teams are not observable in standard matched employer-employee data sets. However, PATSTAT's patent applications contain the names of all contributing inventors. I will discuss in section 2.3 which assumptions imply an incentive to form teams and how such teams affect the matching rationale of firms and inventors. I will also provide some evidence on which assumption is supported by the data.

However, PATSTAT data is not originally intended as an employer-employee data set and using it this way also entails some challenges. Importantly, PATSTAT does not contain unique firm or person identifiers. Instead, it contains the names as written into the fields "inventor" and "applicant" on the patent. Thus, an important step whenever using PATSTAT is to identify individuals and firms, for which I improve upon earlier works (Peeters et al., 2010; Magerman et al., 2006) with a multi-step procedure.
This leaves the problem of one name representing multiple inventors. Latest generation disambiguation algorithms use machine learning techniques to estimate the probability that two strings refer to the same inventor (Toole et al., 2021; Li et al., 2014). However, such algorithm results are not published for the PATSTAT data yet. I use name frequency tables, IPC class portfolios of the alleged inventors and the longevity of alleged inventors in an ad hoc criteria list that is similar in spirit to these USPTO algorithms.

After these cleaning procedures, which I detail in Appendix A, I feel confident when interpreting the remaining data as an employer-employee data set which contains both the inventors and the firms involved in any patent.

2.2 Patent Citations

2.3 Patent Contents and Inventor Teams

PATSTAT contains information about the actual content of inventions through patents' technology classes. I use patents' technology classes to extract information about which inventors do similar research and could in principle be substituted. This defines the size of inventors' labor markets. While current search and matching labor market papers treat the whole labor market as one, I split the labor market for inventors into different markets for different technology clusters and propose an algorithm which can be transferred to a more standard setting, should more data on workers' occupation become available. Intuitively, the algorithm clusters technologies between which inventors switch frequently, because this indicates that these inventors are substitutable. Appendix B details the proposed algorithm, which treats each category combination as a potential distinct cluster and then connects such clusters where possible.

PATSTAT also contains information about the team structure in the form of co-signers on patent applications. Inventor teams offer a fundamental challenge: There is neither a theoretical model nor a readily available estimator for a situation where workers are hired and then assigned to teams. The current state of the art estimators treat the output of a worker as a function of his skill and the firm's quality only, assuming that all workers work independently from each other.

In the larger literature, there are two different ways to explain why inventor teams form.
Akcigit et al. (2018) exemplify the first way. Inventor teams create patents according to a Cobb-Douglas production function in the team leader's skill and the number of inventors: $\lambda = (x_i)^{\varsigma} n^{(1-\varsigma)}$. Only the skill of the team leader matters, so mediocre inventors can make a meaningful contribution if they are paired with an excellent inventor. Beyond this one example, this first way of modeling assumes that inventors are in principle substitutable, but that grouping them increases the arrival rate of patents.
The opposite approach is to maintain that inventors are complements: Each inventor possesses unique knowledge. Inventors work together because some re-

search projects require knowledge in multiple areas and one inventor cannot master them all. E.g. Pearce (2019) studies how inventor teams form and how the returns of more depth (teams with deeper expertise of one area) and width (teams with expertise in different fields) have changed over time, relative to inventors performing research alone.

Even before ranking inventors, the raw data can offer some guidance as to which modeling approach is more appropriate for this particular data set.
First, the size of inventor teams is largely independent of firm size: Inventors are organized in teams of 2-4 inventors, no matter how large the firm is (figure 2).

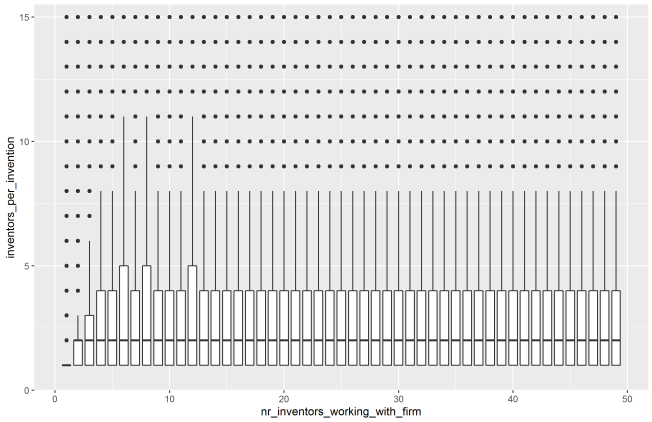

Figure 2: Boxplot of the team size per patent family, sorted by the number of inventors working for the same firm. Firms of basically all sizes opt for teams of 2-4 inventors. Larger firms do not generally assemble larger teams. Even firms too small to form teams of three inventors often do so by cooperating with other firms.

Second, inventors with a high patenting output work in teams with other inventors with many patents. Sorting all inventors by the number of patent families they partook in is of course an imperfect measure of skill, since it ignores the contribution of firms. However, the correlation is strikingly high (figure 3).

Third, patent families created by large teams span more patent classes and teams with more than three inventors span more than two technology clusters on average. One explanation for this pattern is that larger teams form to tackle projects with a broader scope than what any one inventor could cover (Figure 4).

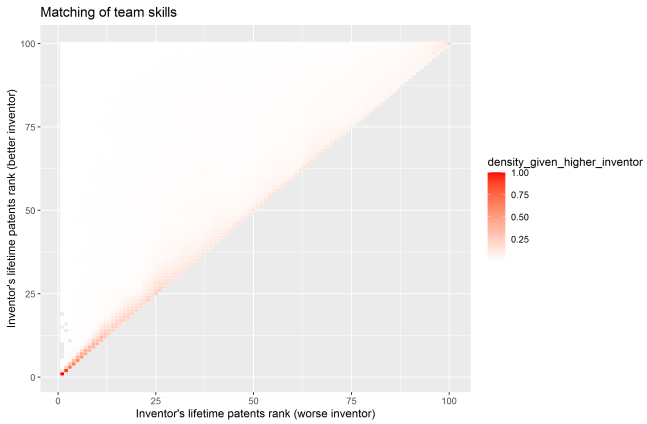

Figure 3: The graph shows the quality of the two inventors in a team. Both axes rank all inventors by the number of patent families they contributed to. 100 is the inventor who participated in most patent families. The x-axis denotes the rank of the less prolific inventor in any two person team, while the y-axis denotes the rank of the better inventor. The density of matches is highest along the diagonal. If the better inventor is in the 100% percentile, his co-inventor is likely to also be in the top percentile. The same holds true across all percentiles: Prolific inventors match with good co-inventors, unproductive inventors match with unproductive co-inventors. Matching a star inventor to a helper seems to be less common.

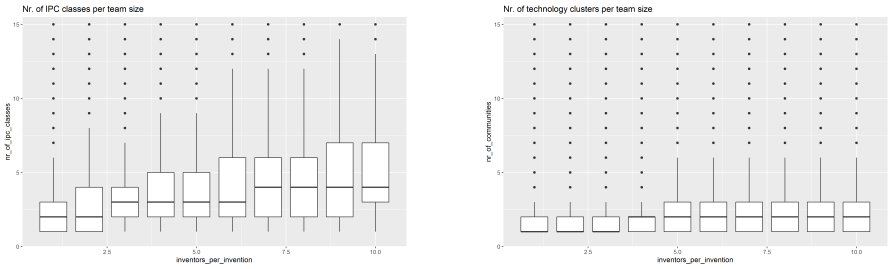

Figure 4: The left graph reports box plots for the number of full length IPC classes (70.000 different classes) per patent family in relation to how many different inventors contributed to the patents of the family. Smaller inventor teams produce patent families with fewer IPC classes. The right graph reports the same statistic, but for technology clusters. Inventors are assigned to one technology cluster, but larger teams produce patents with IPC classes from two or more technology clusters.

These data points are more compatible with some models than with others. First, the fact that firms of all sizes form teams of the same size is not compatible with significant within-firm matching: If firms searched for compatible

inventors to form teams with, larger firms would use their larger pool to find better matches and build larger teams. Instead, marginal productivity seems to decline with team size, just as in Akcigit et al. (2018).
Second, teams consisting of two frequently patenting inventors are not rationalizable with the specific patent invention function of Akcigit et al. (2018): With their production function, high-skilled inventors should form teams with low-skilled helpers. Instead, inventor skill levels within a team seem to be complements, not substitutes.
Third, larger inventor teams produce patent families which span more IPC classes and technology clusters. This is in line with a model in which inventors with different knowledge band together to tackle projects that span multiple areas of expertise.

Section 3 presents a model framework and estimation technique that allows to rank both single inventors and teams. Teams are mostly treated just like another inventor with their own patenting and citation record and get their own ranking based on that and the firms they collaborate with. This method does not require additional assumptions about team formation.

2.4 Estimating the Duration of Matches

An additional challenge when using patent data is that inventors are missing from the data if they do not patent in any given year. The data is thus truncated, since any combinations of inventor and firm not patenting in a certain year are not observed. Even very productive inventors are only observed with a probability of roughly 50-70%.[2] Thus, match productivity and duration have to be estimated.

Estimating an arrival rate for events when the underlying population is not observed is a problem that goes beyond this particular paper. Other use cases are e.g. publications, complaints at government agencies, legal cases at court and trademarks. In all of these applications, only "active" units of observation show up in the data. The arrival rate could be learned from the untruncated data, but this might not exist or be unavailable due to confidentiality issues. The

[2]Estimated patent arrival rates are between 0.2 and 1 (see Appendix C).

methodological contribution of this part of the paper is to demonstrate how to solve the truncation problem present in such data with weak assumptions.

In the patent literature, some studies try to link patent and census data, which faces its own problems and is not always even theoretically possible. Other studies make the ad hoc assumption that inventors work for the same firm between observations. This in itself does not allow to consistently estimate the arrival rate of patents, since the years before the first and after the last patent are still missing.

The estimation I propose exploits the fact that there are many possible patterns of patenting and non-patenting: A spell can produce any number of patents in a given year or none. The estimator infers which distribution of employment lengths and spells would have been most likely to create the observed patterns. It requires two assumptions:

1. Inventors continue to work at the same firm between observations.

2. The arrival rate of patenting events is constant for a given match between firm and inventor.

These assumptions are enough to recover all parameters of interest. Any potential estimator to recover firm and worker ability makes assumption 2 anyway (Hagedorn et al., 2017). Extensions where the arrival rate evolves over time according to a known function follow naturally from the approach presented, but are unnecessary for this specific paper. The weak assumptions necessary to correct for truncations facilitate the transfer of this correction technique to other settings.
The central approach of the estimation is to understand the original, untruncated data as a mixture distribution of different types of employment spells, characterized by their length and the arrival rate of patenting events. This underlying distribution creates a distribution of observable outcomes, like an observed spell of a patent followed by two years of non-patenting, followed by another patent (1001). The estimated underlying distribution of spell types is the one that produces a distribution of observable outcomes close to the one in the data. Given this estimate, it is easy to estimate patent arrival rate and length for every ob-

served spell. The details of how to derive this estimate are described in Appendix C.

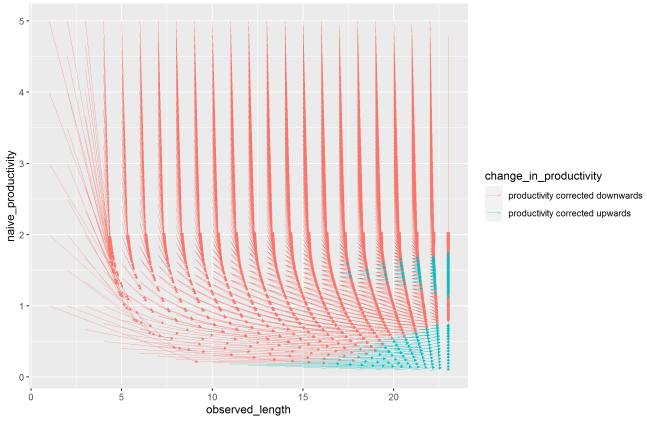

Figure 5: Adjustment of observed employment spell productivity and length. The starting point of each arrow is the productivity and length observed in the data without the correction routine. The end point of each arrow gives the new estimated arrival rate of patents after the routine has concluded. Red highlights spells where the observed productivity was adjusted downwards, blue highlights spells where the observed productivity was adjusted upwards.

Figure (5) shows the results of the correction. It shows the difference between a "naive" treatment of the data, where the truncation issue is just ignored, and the corrected data. Each arrow shows how observed spells were moved in the productivity-length plane: In the bottom left corner, I estimate that spells where only one patent in one year was observed have on average an underlying productivity of just 0.15 patents per year and last roughly 8 years. If one does not take into account the missing zeros, one would overestimate the productivity in these employment spells. For extremely long spells (15 years or more), the estimator even increases the productivity of some spells compared to the naive baseline. This is because it takes into account that long spells with a mediocre productivity would in some cases show up as unproductive. While this is an interesting implication, these estimates do not matter much quantitatively, because few spells are actually that long.

As is intuitive, the GMM routine estimates that productivity is much lower than a naive reading of the data might suggest. This is because it includes

completely unsuccessful years in the productivity estimate. Additionally, the GMM graph is much flatter: The routine concludes that while inventors in longer spells are generally more productive, the difference is much less pronounced. This is because the truncation correction takes into account that long matches with low productivity often generate only one or two patents and thus look just like short spells in the data. Long and short employment spells are less different than one would conclude at first glance.

After this correction technique, the data can be treated as an employer-employee data set which for each match contains

- the estimated length (as average from a distribution of different potential types)
- the estimated start and end date
- the number of patent families the match participated in
- the estimated arrival rate of patenting events (again as average from a distribution of different potential types)
- the number of co-inventors for each realized patenting event
- expected match output $\lambda_{x,y}$ as arrival rate times the inverse of the expected number of co-inventors

While I have to adapt standard labor market matching techniques, they are transferable to this new setting.

2.5 Output Measurement

In the baseline model, matches produce a steady flow of output. However, inventor teams sporadically produce distinct inventions, which again then yield citations, the generally preferred output measure. Thus, once teams have formed, they have to decide how ambitious $a^*(y, \overrightarrow{x})$ they are going to be, given the quality of the firm and all skill levels in the team. High team productivity might manifest either as more difficult projects with higher returns (more citations) or a reliable stream of small, patentable improvements. $a^*(.)$ is an unknown function, i.e. there is no way for the econometrician to know ex ante what the optimal

ambition for a given team is.

This matters

To determine the relationship of patent

which of these scenarios dominates in the data, I use a subset of teams for which all the measurement problems associated with the frequency of patenting are alleviated. These are teams that

- patented for at least three years consecutively, so that the truncation of unsuccessful teams is less of a concern
- contain no inventors that also work in other teams, so there is no question of the time allocation of inventors
- work for firms that patent more than once, so that they are part of a firm's continuous research effort, which the model tries to capture

I assume that I can measure output as expected patent citations per year for these teams with reasonable precision (after the truncation correction described in the data section). I can thus assess the correlation between output and either patenting frequency or citations per patent:

Evidently, both citations and patenting frequency can be used to determine the output of a match. Citations seem to be the better measure, the correlation being roughly twice as high.

Citations (per patent family) have the additional advantage that they are measured without truncation. This reduces the uncertainty of any inventor or team comparison, especially for teams with only a few patents. Since inventors can be part of multiple teams, such teams constitute a substantial share of the data. Thus, I do not compare the output, but citations per patent family in the main specification of the paper to reduce measurement error and increase precision.

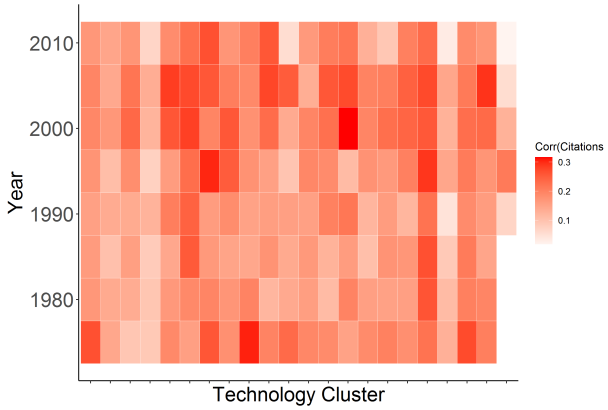

Figure 6: Correlation of expected citations per year (as a measure of output) and patenting frequency. The two are highly correlated throughout all technology sectors (X-axis) and time periods (Y-axis).

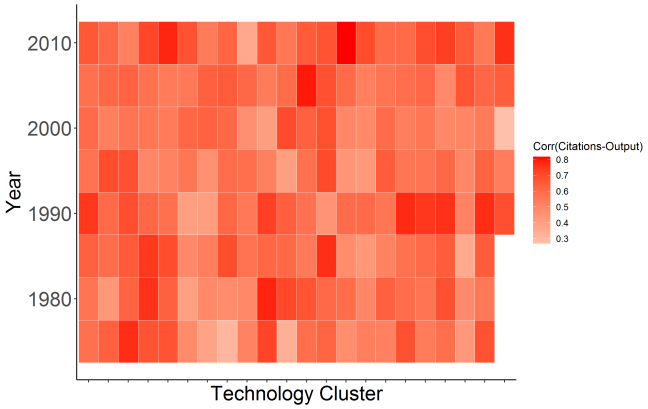

Figure 7: Correlation of expected citations per year (as a measure of output) and average citations per patent family. The two are highly correlated throughout all technology sectors (X-axis) and time periods (Y-axis).

3 Estimation Framework

This section details the framework within which I estimate firm quality and inventor skills. The central assumption in the labor market matching literature is that output is produced by matches of workers and firms. In the context of this paper, this would mean that inventors invent in conjunction with the firm they work for. The arrival rate of new inventions is a function of the inventor's skill and the quality of the firm's research environment, or firm quality for short. Firms' quality y and inventors' skill x are unknown to the econometrician but known to some actors in the economy, as is the patent invention function $\lambda(x, y)$. The goal of the econometrician is to estimate these objects.

This literature usually ignores teamwork among workers. E.g. neither Hagedorn et al. (2017), nor Lamadon et al. (2015) or Abowd et al. (1999) attempt to account for teams' effects in production. In the following discussion, I show which biases are introduced in the estimation because of this and how to rectify them by generalizing the model of Hagedorn et al. (2017) to encompass worker teams. Teams are not usually documented in standard employer-employee data, which might explain this oversight in the literature, yet patent data does not

have this limitation.

Patent data contains not only the team structure, but also the ambition of projects. Patent citations measure the quality of patents. Productive teams might not actually complete more projects, but produce higher quality results. The production function must thus contain quality and quantity of output. Citations can also help determine the ranking of inventors and teams.

However, patent data also poses additional challenges when using standard labor market estimators. The first challenge is that standard estimators expect a continuous wage variable, measured with reasonable precision. Instead, patent data contains the discrete number of patents an inventor has applied for. Even taking citations into account, the outcome variable is measured with non-normal error. The second challenge is showing that an estimation procedure accurately ranks inventors and firms when given patent-citations-per-year instead of wages. Specifically, I demonstrate that my estimator yields unbiased inventor and firm rankings even if the patent invention function $\lambda(x, y)$ changes over time or the production function is not logarithmic.

In the next section, I will first lay out a generalization of the Hagedorn et al. (2017) model to accommodate teams. I will discuss under what assumptions the new model reverts to the original and what biases arise from violating these assumptions. I will then detail how to transfer this model to an inventor-firm setting and deal with the issues posed by the different data structure.

3.1 A Theoretical Framework for Matching with Team Building

In a standard theoretical matching model, firms create vacancies and look for workers, while workers search for jobs. The rate at which matches are formed depends on the number of firms and workers searching. Whenever a firm and a worker meet, they reveal their types, decide whether to match or keep searching, and Nash bargain over the wage. The decision which partners to accept for a match is the central decision in the model for both firms and workers.

Whether any worker-firm pair accept a match depends on the value of the match, i.e. a match's expected production $E(f(x,y))$ must be higher than the expected payout from continued search:

$$E(f(x,y)) \geq r(V_u(x) + V_v(y))$$

where $V_u(x)$ denotes the value of being unemployed and continuing to search for jobs for a worker of type x and $V_v(y)$ denotes the analogue for firms. At their meeting, the pair of worker and firm discover each other's type x and y, check whether they can produce more than both outside options and then bargain over a wage that is acceptable to both. In the patent setting, inventors are the and applicants are the firms.

Inventor teams shake up this framework because $E(f(.))$ is no longer just a function of x and y, but also of the other workers the inventor teams up with. This will affect inventors' and firms' matching behavior both in unemployment and while working: Inventors presumably mostly build teams while they are already employed, but there is no way to verify that in the data. In fact, the data occasionally contains teams consisting of inventors from multiple firms and fully formed teams sometimes move to other firms. Thus, it is not clear when bargaining begins and what the outside option of a specific inventor is. Potentially, both workers and firms have to take into account how their chances of building inventor teams change if they accept a match.

Firms' production is the sum of its teams' production. Thus, as far as a firm can affect the team structure of its inventors, it maximizes $f_f(y) = \sum_1^T f_\tau(\overrightarrow{x}, y)$ where τ indexes the individual teams of firm f. The firm knows $f(.)$, the team production function, but it is unknown to the econometrician. The firm is much larger than individual inventors and thus views them as a mass and $f(.)$ is twice differentiable over this mass, just as in the baseline Hagedorn et al. (2017). However, it is not clear how much control the firm actually has to maximize $f_f(y)$: To create optimal teams, someone at the firm needs to verify the skills of all inventors, know the patent invention function and have the authority to order people into the optimal teams even across departmental boundaries or other or-

ganizational frictions.

Inventors themselves can plausibly affect team formation, too. Their incentives are very different from those of the firm: E.g. Akcigit et al. (2018) report that inventors collaborating with better inventors substantially increase their own productivity. Likewise, Dino et al. (2020) highlight the importance of research team leaders and their positive effects on team members. Inventors thus have plausible incentives to join high profile teams even if firm patent citation production decreases as a result.

Equilibrium team formation between inventors has its own literature and since there is no settled result yet, I develop an estimator that is robust to as many different team formation processes as possible. I thus derive my estimator based solely on some weak assumptions about the production function and try to avoid optimal behaviour arguments, which would hinge on specific assumptions about the search and matching process.

Since Abowd et al. (1999), the literature has focused on estimating the skill and quality of individual and firms. Taken to the patent setting, these translate to three parameters of potential interest:

- the percentile of the team in the team skill distribution $\hat{x_\tau}$
- the percentile of the individual inventor in the skill distribution $\hat{x_i}$
- the percentile of the firm in the firm quality distribution $\hat{y_\tau}$

However, these parameters are only intermediate results. The entities of actual interest to econometricians are the joint distribution of skill and quality in the population $N(x, y)$) and the expected patent (citation) output for pairs of a given team skill and firm quality $\lambda_c(x, y) = \lambda_p(x, y) * E(c(x, y))$, where research output is defined as the product of inventions arrival rate $\lambda(x, y)$ and the expected patent citations of these inventions $E(c(x, y))$. Combining the approaches of Hagedorn et al. (2017) and Bonhomme et al. (2019), I aim to directly estimate only these objects of interest.

3.2 Estimation Methodology

The ultimate goal is to estimate the macro objects $\lambda_c(x_{tau}, y_f)$, i.e. the expected output for pairs of an inventor team with a certain combined skill and a firm with a certain quality and $N(x_{tau}, y_f)$, i.e. how often such pairs occur.

Intuitively, I combine the different strands of the literature, by using the Hagedorn et al. (2017) methodology to group similar inventor teams together and then exploiting the fact that the incidental parameter problem (sometimes called limited mobility bias) that usually complicate estimation are reduced when grouping together similar individuals Bonhomme et al. (2017). In principle, I maximize $\mathcal{L}(\lambda_c(x, y), N(x, y))$ by guessing both objects and assigning inventors and firms to skill and quality categories. Then, I improve the likelihood of producing the observed distribution of match outcomes by moving inventors and firms to different categories and changing the aggregate guesses.

Since inventors and firms will be grouped according to their skill, x_{tau} and y_f will consist of a finite number of possible values (100 in this application) and thus the number of their combinations will also be finite (10000). Since the number of patents and their number of citations is also finite, there is a finite number of potential outcomes (indexed by o in the future). Thus, log-likelihood is

$$\mathcal{L}(\lambda_c(x_{tau}, y_f), N(x_{tau}, y_f)) = max_{\overrightarrow{x},\overrightarrow{y}} \sum_{x}^{X} \sum_{y}^{Y} \sum_{o}^{O} p(N(o)|N(x, y), \lambda_c(x, y)) \quad (1)$$

where $p(N_o|N(x, y), \lambda_c(x, y))$ denotes the probability to observe outcome o N_o number of times, given the estimated number of pairs of skill x and quality y and their estimated productivity. Actually maximizing this object is prohibitively expensive computationally: $\overrightarrow{x}$ the maximum likelihood is not linear in observations because the likelihood contribution of e.g. the example outcome $1|-|-|-|0$ for pairs of the best firms and inventor teams ($x = 100, y = 100$) depends not only the estimated productivity of such pairs, but also on how many other observations of the same kind there are for such pairs. Since changing the skill bin of an inventor team changes both the number of observed outcomes and the implied productivity and number of such matches, maximizing $\mathcal{L}$ would require

testing out all different possible combinations of skill and quality assignements, an impossible task. Note that this is true despite the target of the estimation "only" being aggregate objects, not individual matches. The estimation strategy below presents a computationally feasible strategy to approximate the ML estimation and D shows that it converges to the correct values for the sample sizes I encounter. However, the estimator will only yield biased estimates of **match** productivity, skill and quality.

The procedure to execute this strategy contains of three steps. I will discuss these steps using an example match between firm 1 and inventor 1 that produced an invention with one citation and then a second inventions without citations after four years. Thus, in the data, this match shows up as $1|-|-|-|0$ (Fig. 8). Assuming a Poisson distribution for patent arrival, the maximum likelihood estimate for this single match is a patent arrival rate of 0.20, a true length of the employment spell of 10 years and an average of 0.5 citations per patent. However, the probability to create the observed data with these characteristics is below 10%. In a patent context, individual match productivity will always be measured with very low precision.

In the first step, the estimator creates hypothesis about the productivity of each match together with the likelihood of the data given that specific hypothesis. Ideally, one would like to keep track of all possible productivities and their likelihoods of creating the observed data, but this is computationally too expensive. The second step is to rank the inventor teams and firms, given these hypothetical match characteristics, using the Hagedorn et al. (2017) method. I.e. in the above example, this might yield ten different rankings for the same inventor team-firm pair: If the hypothetical match productivity is high, the skill and quality ranking of the participants will both be higher, too. Note that the ranking is based on hypothetical productivity (and the hypothetical productivity of all other matches), not the actually observed data. Section 3.4 details how I adapt Hagedorn et al. (2017) to rank teams.

The likelihood for any of these hypothetical match characteristics is then relatively easy to compute. However, the purpose of the estimation is not to recover the ranking of any individual inventor or team, but to maximize the likelihood

of the hypothesized function $\lambda(\hat{x_\tau}, \hat{y_f})$ and $N(\hat{x_\tau}, \hat{y_f})$, i.e. the number of matches for every skill-quality combination and their patent (citation) arrival rate. The algorithm will accept all the hypothetical rankings from step two at face value, bin similar workers and firms together and compute the initial guesses for λ and N for every match of bins. The estimator then compares the observed distribution of outcomes with the one we would expect given the guess and give higher weights to those rankings of a given inventor that are more plausible. E.g. in the above example, the estimator would assign a high probability of inventor 1 being in the low bin if there are currently fewer $1|-|-|-|0$ patterns observed by inventors in the low bin than he would expect given the current guess. Whenever inventors are "moved" to different bins this way, the aggregate guesses change as well, until the likelihood of the oberved distribution of outcomes in every bin is maximized and the estimator converges. Section 3.5 describes the actual computation in more detail.

Figure 8 gives a good example of how this affects optimization. There are two different hypothetical rankings (2 and 3) that boil down to the same $f(\hat{x_\tau}, \hat{y_f})$ and $N(\hat{x_\tau}, \hat{y_f})$: There is one match between the best inventor and firm F and one match between the second best inventor and firm F. The only difference is who the better inventor actually is, which is not important for the two moments that are estimated. Thus, both rankings contribute to the likelihood of this specific estimation result.

Compared to the original Abowd et al. (1999), this has distinct advantages: In a double fixed effects setting, the incidental parameters problem is well understood to make the estimation of $N(\hat{x_\tau}, \hat{y_f})$ difficult, since error in the estimation of $\hat{x_\tau}$ also affects the estimate for the $\hat{y_f}$ firms for which this team works. There are correction procedures, however, these rely on assumptions about the error terms which are not fulfilled in the patent data setting. Since I group inventor teams and firms into 100 bins, I estimate two parameters for the 10.000 potential skill combinations, well below the millions of data points that I am trying to fit. This makes incidental parameter issues less of a concern. In addition Abowd et al. (1999) type estimators have to make additional assumptions like a log-linear, additively separable production function, which are not necessary for the procedure proposed in this paper. This is especially beneficial in this context,

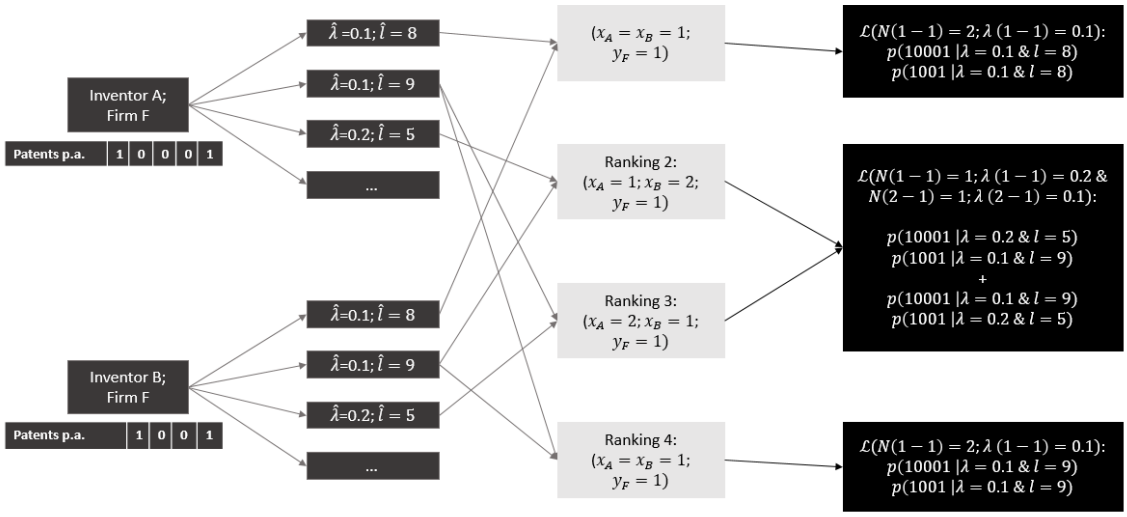

Figure 8: Overview over the estimation technique, from observations on the left to estimated distributions on the right. The technique is illustrated with two inventors (A,B) working for the same firm (F). Since output is measured with such a high measurement error, every observation generates a whole distribution of potential interpretations. These lead to different rankings. The final estimation result does not maximize the likelihood for any specific observed data point individually. Instead it maximizes the chance to observe the joint distribution of ranks. Since ranking 2 and 3 predict the same match distribution (but with different inventors being good or bad), they contribute to the same likelihood.

since the team production function is debated in general.

Bonhomme et al. (2017) offer a different approach that relies on defining firm classes: Since firms are grouped together to classes prior to estimation, fewer parameters need to be estimated and worker mobility (now between firm classes) goes up, alleviating the above concerns. However, this grouping approach forgoes to estimate a worker ranking directly. Instead, "only" the sorting of worker types to firms and the complementarities between workers and firms are evaluated. These group level estimations are in a way a similar strategy compared to the one I am proposing, but since worker rankings and sorting patterns are central to the message of the paper, I follow a different approach.

Allowing firm quality to change over time also alleviates concerns that peer

effects or agglomeration effects might bias the estimation. Moretti (2019) shows that holding both inventor and firm constant, denser agglomeration of matches can increase output by up to 25%. However, this does not have a large impact on the inventor rankings because most good inventors are in highly agglomerated regions: In the US data of Moretti (2019), ten cities account for between 60 and 75% of all patents in the top technology clusters. Thus, most productive inventors are on equal footing in terms of knowledge spillovers. This carries over to PATSTAT, the data basis for this analysis: While detailed geographical information is only available in PATSTAT from 2000 onwards, the available data shows that many patents come from the NUTS2 region with the most patents in each country. Therefore, agglomeration effects do not significantly change the ranking of inventors and will largely be soaked up in the firm quality measure, which captures both the "pure" research skill of the firm and the knowledge spillovers from other nearby firms. The number of firms with significant research departments in more than two NUTS2 regions per country is small.
To account for changes in cluster size and the possibility that firms' innate quality changes over time non-parametrically, I estimate firm quality for every five years separately. I.e., I effectively treat the same firm after five years as a separate firm and rank it again.

Following the method I sketched above to the letter is prohibitively computationally expensive: Assume that there are only 50 different hypothetical values for every observation (because we only take into account λ-values of 0.1, 0.2, 0.3 etc. and spell lengths of full years). That would still mean that in a small technology cluster with 35.000 employment spells, there are $50^{35.000}$ different combinations of hypothetical spell characteristics, i.e. a completely unmanageable amount of combinations to rank and then compute likelihoods.

I reduce the computational burden in each of the three steps of estimation. First, I estimate a distribution mixture of hypothetical types that can explain the observed data and then draw only five hypothetical types for every observation (section , Appendix C). Second, I rank only within each draw, i.e. I rank all observations using the hypothetical characteristics draw first and then rank all observations again using the second set of drawn characteristics. However, I do not consider rankings between different draws. These five scenarios already give

a reasonable idea of the spread of ranks that are possible for every team and firm. Lastly, I compute the likelihood of observing the actual data for all five draws and pick the hypothetical spell characteristics for every team and firm that yield the highest value. Clearly, this is far from actually testing all possible combinations, but it yields numerically close results even in relatively small samples (Appendix D).

3.3 Step 1: Hypothetical Spell Characteristics

This section details the process though which I select potential characteristics for an observed spell to further consider during the ML-estimation. Only potential characteristics selected in this stage will be considered for likelihood-maximization, which looses some precision, but is necessary to keep the computational costs in check. I consider 10 different hypothetical characteristics, which simulation exercises show to be sufficient to recover aggregate characteristics with high precision in relatively small communities of inventors.

Each match is already associated with an estimated length of the employment spell and the invention arrival rate from 2.4, together with a distribution of lengths and rates from which this estimated is derived. To create hypothetical spell lengths and arrival rates, I draw 10 times from this distribution for every observation. Each match also has an observed average number of citations per invention. The hypothetical citations per invention are randomly assigned from the interval between 0.1 and the observed average number of citations per invention plut the observed standard deviation. E.g. in the above case, $\lambda(c_1^d) \in [0.1, 1.2]$, since the standard deviation of the observed citations per invention is 0.7.[3] Note that these hypothetical distributions are the possibilities further considered by the algorithm, so a wide dispersion of hypotheses avoids presupposing the outcome at the cost of precision once the algorithm has come close to the result.

3.4 Step 2: Aggregating Conflicting Rankings

Hagedorn et al. (2017) propose to not estimate the production function directly,

[3] I set the standard deviation to 3 where it cannot be computed

but instead first rank workers and firms according to their ability. Their method alleviates the identification issues associated with double fixed effects regression. To obtain these rankings, they solve a worker-firm matching model and identify the observable variables that can be used to rank workers (lowest acceptable wage, wages when working for the same firm, etc.). However, their model does not contain teams and their estimator uses wage data, which requires adjustments.

Since patent data effectively measure the output of worker teams instead of wages and theoretical models of team formation are much less developed, the inventors' optimization decisions themselves are less directly linked to our data: Presumably, inventors maximize discounted earnings, not patent output when they make their matching decisions. When doing so, they will also take into account how their matching decisions affect future team formation possibilities, about which we know very little theoretically. To avoid these issues, the estimator proposed here is derived from assumptions about the output function of matches only.

The output $\lambda_c(x_\tau, y_f)$ of a match between inventor team τ and firms f is an unknown function. However, some very weak assumptions will be enough to rank inventor teams. First, the production function is strictly increasing in both skill and firm quality. This assumption alone implies that one can rank all inventor teams within the same firm according to the expected output of their match: Since y_f is the same for all teams, all differences in expected output are due to different team skill.
Second, teams skill x_τ is strictly increasing in additional inventors. While exactly how individual inventor skill x_i features into team skill x_τ remains unspecified, this assumption implies that teams that contain within them other teams will always be better, i.e. a team of two inventors will always be better than the two single inventors. In an important deviation from the original setup, this allows to rank two inventor teams that are in completely different firms and whose output we thus cannot compare. This improves the connectivity of the different inventor team rankings beyond what would be possible in a double fixed effect setting. Note that this assumption does not imply that it is always optimal to form teams: The additional contribution of a second inventor can be so minor

that it would be better to have two separate teams, even if it is positive.
Third, teams skill x_τ is strictly increasing in the skill of every component inventor. Inventors switching between different teams can be ranked using this assumption: If a team performs better after one inventor is switched out for another inventor, this can be attributed to the higher skill of the new inventor. Just like the second assumption, this allows to even rank teams that are never observed in the data: If two inventors switch places, we can now also gauge their individual skill even though they never worked alone: We can rank one of them as better than the other and we can rank both of them as worse than the full teams.

Hagedorn et al. (2017) propose to aggregate all rankings by finding the ranking of all inventor teams that has the fewest disagreements with the data. Specifically, they count disagreements using the Kendall score, i.e. whenever a candidate ranking ranks worker x_1 higher than worker x_2, they count how many of the above rankings have it the other way around. Theoretically, all rankings derived by the above assumptions should be in agreement: If rankings disagree, this indicates that the hypothetical match productivities that led to these ranking are not compatible with the assumptions laid out above. Unfortunately, finding the inventor team ranking that minimizes disagreements is NP hard, i.e. it cannot be solved exactly with current computers. Hagedorn et al. (2017) show that assuming that the ranking distance function has only one local (and global) minimum yields accurate rankings, even for relatively noisy data. Hagedorn et al. (2017) use a Baysian technique to resolve ranking disagreements, but there is no basis for this when using hypothetical productivites without measurement error.

Since the resulting inventor team ranking is only an approximation, I refrain from requiring the final ranking to have zero disagremments: One could conclude in such cases that the drawn hypothetical productivities cannot be correct and redraw some of them to resolve ranking disagremments. However, this is very expensive computationally and seems to not be necessary to obtain convergence in even small samples D.

The outcome of this disagreement minimization are 10 hypothetical rankings that can be understood intuitively as statements like "If solo-inventor i had 0.8 citations per patent on his first spell and 1.2 citations per patent on his second

spell, his skill rank would be X".

3.5 Step 3: Maximum Likelihood Estimation

The last step is to move from the 10 hypothetical rankings above to the actual estimate of aggregate objects of interest (and a biased estimate for the individual team). To group inventor teams and firms into groups, I create percentiles of the respective ranking. Assuming all 10 hypothetical rankings would be correct, one could just compute average productivity and the number of matches in each bin. However, this will lead to systematic deviations between the expected and the observed distribution of outcomes within each bin: Consider e.g. the matches between the best teams and the best firms. Inventor teams have only been sorted into this bin if their hypothetical productivity (the basis for the rankings) was particularly high. However, their observed productivity will on average be lower. Thus, the likelihood of creating the observed distribution is very low, given that the hypothetical productivities are correct. This allows one to pick the most likely hypothesis for inventors and converge on the actual distribution of matches.

Maximizing the However, when comparing the outcomes obtained by this

The problem resulting from large measurement error becomes apparent when looking at an example: Consider four inventors A, B, C and D. They work for two different firms, X and Z. Figure 9 shows the matching between inventors and firms and the resulting patent invention rates. C and D have the same patent invention rate of 50% a year. However, they will likely not have the same outcome in the data: If the firm produces the expected number of patents, only one of them will be successful. Thus, a naive ranking according to their outcome would produce much higher skill diversion within the firm than is actually the case.

Importantly, even ideal data would not alleviate measurement error in the rate of patent inventions. Ideal data would contain the employment biographies of inventors, a designation that marks when they are assigned to research activities and the patenting outcomes. However, even such data would only imperfectly measure the patenting productivity of inventors. Because patenting is a rare, discreet event, any data will contain enough measurement error to make rankings suspect. Any study of the patent invention function has to solve this problem,

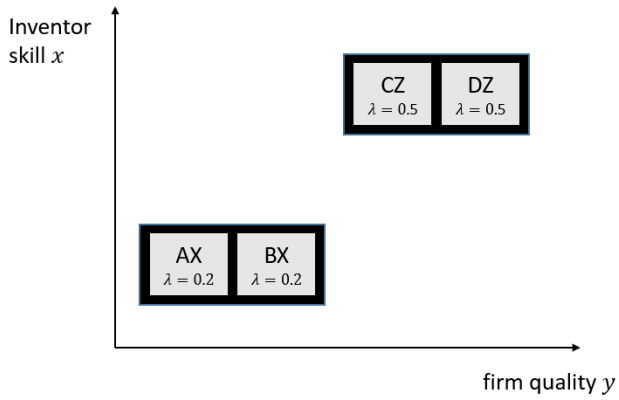

Figure 9: Example of inventor-firm matching. Inventors A and B have both matched with firm X and both produce 0.2 patents per year. Hence, they both have the same skill. Inventors C and D have both matched with firm Z and both produce 0.5 patents per year. This is due to both inventors having higher skill than A and B and firm Z being of higher quality than firm X.

regardless of estimation technique and data sources.

Measurement error in spell productivity will affect the results in two main ways. First, it will bias the estimate for assortative matching towards 0: Even if a high quality firm only works with high quality inventors, some matches will be unlucky and look unproductive. Likewise, unproductive matches of low quality firms with low quality inventors will sometimes seem really productive.

Second, measurement error will make inventor skill look more important than it really is: Because inventors usually have few spells, their patent arrival rate is measured with even more error than the average arrival rate of (large) firms. Thus, any estimator will pick up on the fact that inventors at the same firm have widely different outcomes and conclude that inventor skill is an important driver of patenting.

The size of the bias can be substantial: In the simulation exercise described in Appendix D with an unadjusted Hagedorn et al. (2017) estimator, measurement error in λ reduced estimated assortative matching by half (0.4 instead of 0.8) and twisted the production function from $\lambda_{y_f,x_i} = y_f * x_i$ to $\lambda_{y_f,x_i} \approx x_i$, i.e. the estimator was unable to detect any significant effect of firm quality.

In the example of the four inventors above, consider a potential observed outcome for the matches in figure 10. On average, half of the inventors in both firms will

produce more and half will produce less patents than expected. The econometrician observes these $\hat{\lambda}$ and would conclude that assortative matching is weak and inventors' skills are an important driver of match productivity differences.

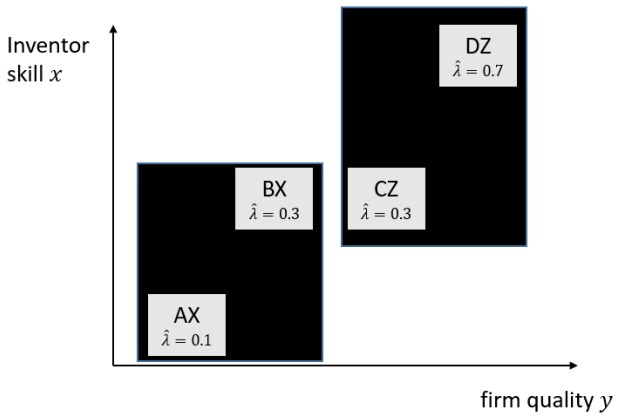

Figure 10: Example of inventor-firm matching. The econometrician does not observe the original patent arrival rate λ, but $\hat{\lambda}$, which is measured with error. Half of the inventors in both firms produced more patents than expected, the other half produced less. If taken at face value, the econometrician would overestimate the difference between the inventors in both firms and underestimate the degree of sorting.

I use a maximum likelihood argument to correct for this problem: I search for the distribution of match productivities that is most likely to produce the observed data. To compute this probability, one needs the global joint distribution of employment length and patenting probabilities, i.e. how many spells of which type there are in the data set. I estimate this distribution anyway, in order to correct for the truncation problem of only patenting matches being observed (Section 2.4), but the reasoning is flexible enough to incorporate any estimation technique that yields this joint distribution. In the best case, with untruncated data, the econometrician can just observe this distribution. In the example, the truncation correction procedure would conclude that two employment spells had a patent arrival rate of 0.2 and two spells had a patent arrival rate of 0.5. With this additional information, there are two possible scenarios that could have produced the observed distribution (Figure 11).

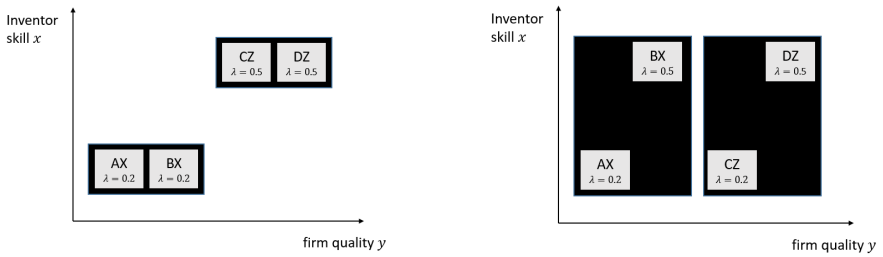

Figure 11: Two different scenarios that both conform to the overall distribution of spells (two with 0.2 and two with 0.5), but have drastically different implications for the estimated production function and assortative matching. The likelihood of the observed data is lower if the right scenario is correct, so it is discarded. If the right scenario was correct, all four matches had to draw exactly the productivities observed. In contrast, in the left case, either match at one of the two firms could have over- or underperformed and the result would have been indistinguishable from the observed data.

Between these two possibilities, the left scenario in Figure 11 is much more likely: In it, each firm has on average the expected number of patents. In the other case, one firm consistently overperformed and the other firm consistently underperformed. Ideally, I would like to compute the maximum likelihood for every possible distribution of "true" productivities among the observed spells in this way. However, with millions of spells, this is not feasible. Instead, I use a pruning algorithm that randomly draws possible true productivities for each spell and then consecutively eliminates the most unlikely draws.

To check the performance and robustness of my empirical findings, I undertook a simulation exercise described in more detail in Appendix D. This exercise corroborates that the algorithm is able to identify the correct distribution of match productivity and the correlation between firm and inventor skill. However, the estimate for any individual inventor is still subject to substantial error, especially if they match only with few firms. In the above example, the maximum likelihood technique would conclude that both firms matched with inventors of the same skill level and could thus provide a correct estimate for every inventor. However, if firms match with inventors with different skill levels who do not move to other firms, the individual skill estimates are subject to substantial error.

3.6 Patents and Productivity

To relate the patent data to economic outcomes, I follow the approach of Doraszelski and Jamandreu (2013). They jointly estimate firm level productivity and the effect of endogenously chosen R&D investment on productivity. For R&D investment, I substitute the observed size of the firm's research department, patenting outcomes and the firm's estimated quality. This has both econometric and theoretical advantages. Doraszelski and Jamandreu (2013) themselves note that it is unclear how much of the time variation of R&D investment is due to accounting practices and how much is economically relevant. Additionally, there is presumably a time lag of unclear length between investment in R&D and actual productivity improvement. Thus, I use the measured quality adjusted size of the research department as an endogenous choice variable. I use realized patenting counts to narrow down when the investments into research paid off.
Specifically, I assume that (log) revenue is a function of (log) inputs and (log) productivity

$$y_{it} = \beta_0 + \beta_k * k_{it} + \beta_m * m_{it} + \beta_l * l_{it} + \omega_{it} \tag{2}$$

where k denotes the log of capital in the books, m denotes the log of intermediate inputs and l denotes the log of employees. I also assume that productivity follows a Markov process of the form

$$\omega_{it} = g(\omega_{it-1}; p_{it-1}; \Lambda_{it-1}) + u_{it-1} \tag{3}$$

where p_{it-1} denotes the number of patents a firm has filed in the last year, Λ_{it-1} denotes the quality weighted size of the firm's research department and ω_{it-1} is lagged productivity. As is common in the productivity estimation literature, I will approximate the function $g(.)$ by a third order polynomial of all its terms. Including both the researchers of the firm and their output allows for a positive effect of this highly skilled personnel even before they produce patentable research.
The equations are identified by the timing assumptions prevalent in this literature: It is assumed that the firm has to decide on investment and thus k at the end of the previous year, before knowing its productivity. Thus, capital is by assumption uncorrelated with ω_{it} in equation (2). In contrast, l and m are optimally chosen, given the productivity the firm expects. The law of motion

yields that productivity is predicted by ω_{it-1},p_{it-1} and Λ_{it-1}. Thus, l and m are exogenous when controlling for the productivity the firm could expect. Thus, I estimate

$$\omega_{it} = \beta_0 + \beta_k * k_{it} + \beta_m * m_{it} + \beta_l * l_{it} + g(\omega_{it-1}; p_{it-1}; \Lambda_{it-1}) + u_{it-1} \qquad (4)$$

which yields unbiased estimates of $\beta_k, \beta_l \& \beta_m$. For a detailed and general discussion of this control function approach to production function estimation, see **?**. Since I cannot control for prices with the data at hand, I follow **?** to compute the markups implied by firm behavior: In static equilibrium, firms will equate revenue productivity of a flexible factor with this factor's costs. This can be used to back out the markup implied by that firm's input choice.

4 Results and Stylized Facts

This section presents the results from the above section (3) and distills them into stylized facts. I explore staples of employer-employee matching estimation (assortative matching, patent invention), but also the concentration of technological capabilities in firms. Ultimately, the empirical analysis alone cannot decide on the welfare implications of the observed changes. A model or additional information is needed to differentiate between welfare enhancing and decreasing developments.

4.1 Matching of Inventors and Firms

Firms and inventors generally match assortatively, i.e. good inventors move to good firms. To present the results from all patenting authorities and technology clusters succinctly, figure 12 pools technology clusters and time periods and shows how often the respective combination of firm and inventor quality is observed.

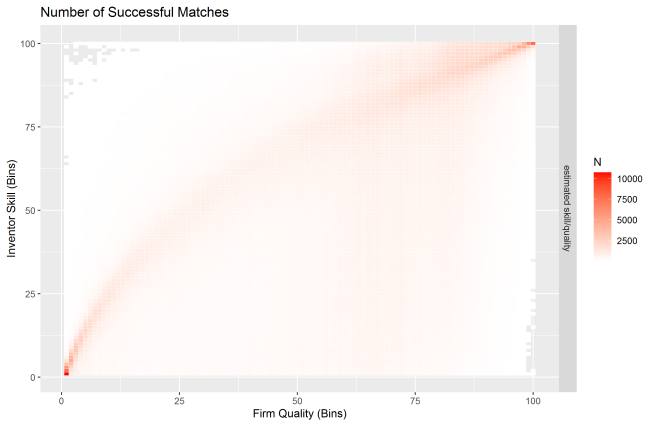

Figure 12: This graph shows the matching of inventors to firms, pooling all time periods (1974-2012) and technology clusters. Red areas are densely populated with spells, while blue areas are largely empty. Matching is assortative, i.e. better inventors go to better firms. Grey areas of the plane have fewer than 50 matches.

Evidently, in general, highly skilled inventors seek out high quality firms. Inventors seem to be less picky than firms, so the matching area is curved upwards: An inventor in the 50% skill percentile will only match with firms in the 25% percentile of quality. In general, the core matching area is quite narrow, with the rest of all matches dispersed relatively evenly across the plane.

Figure 12 does not take into account the different lengths of employment spells. However, there is no strong pattern regarding the duration of matches: Matches within every cell of the plane are estimated to last between 7 and 10 years on average. Hence, the number of expended hours in every cell largely follows the number of matches.

Figure 13: This graph shows the labor input provided by matches of inventors of a certain skill and firms of a certain quality, pooling all time periods and communities. Matching is assortative, i.e. better inventors go to better firms.

Assortative matching not only differs between technology clusters, but also evolves over time. Thus, the correlation of inventor skill and firm quality changes over time. The correlation captures linear relationships, yet the pattern in figure 12 is still close enough to linear to be captured this way. Figure 14 documents the development of the correlation over time for the five biggest technology clusters and the three largest patenting authorities.

Figure 14: Evolution of the correlation between inventor skill and firm quality within matches for the biggest technology clusters in the largest patenting authorities. Assortative matching increases in most technology clusters and over all.

The correlation is increasing over time and in most communities. The overall

increase from 1974 to 2010 in the US is from 0.45 to 0.6, or 33% (top panel). This amounts to about 0.004 per year. The rise is not monotone: Assortative matching peaked between 1985 and 2000 at around 0.65. It has been decreasing slightly from since then.

The outlier with respect to the overall trend towards assortative matching is computing: After 1985, assortative matching is continuously sliding downwards. Combustion maintains its high level of assortative matching, while chemistry, foodstuffs and semiconducturs are rising.

More mature technologies in concentrated industries seem to experience rising assortative matching. Semiconductors is a prime example of a technology focused solely on a specific technological problem: increasing the number of transistors in integrated circuits. Bloom et al. (2017) cite semiconductors as one of their prime examples for decreasing technology growth, as it becomes harder and harder to double transistor numbers.

To determine whether patenting rates decline in semiconductors and the economy overall, one has to turn to the patent invention function λ. Overall productivity will be determined by how many inventor years are invested in each cell and how the productivity of these cells changes over time.

4.2 Patent Invention Function

The estimated patent invention functions are highly stable over time and put more weight on inventor rather than firm quality. I estimate the patent invention function non-parametrically on the same grid as assortative matching: I group workers and firms into 100 percentiles according to their ranking and estimate the labor input weighted average within each combination.

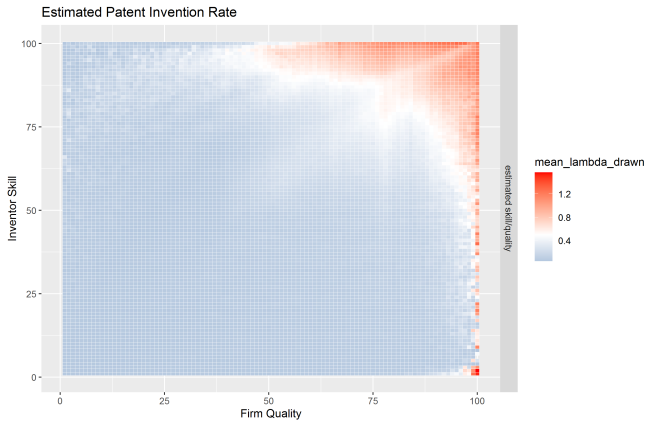

Figure 15: Pooled Patent Invention Function: Expected number of patent author shares as a function of firm quality and inventor skill. Matches of highly skilled inventors with high quality firms have much higher patent arrival rates. Inventor skill is more important than firm quality.

Inventor skill and firm quality are both important drivers of patent inventions, however, inventor skill is slightly more important: E.g. an inventor in the top 1% matched with a firm in the middle of the distribution will create more than one patent per year, while the reverse combination is less productive.

A large debate in the literature is whether inventions have become harder to find, i.e. whether the rate of patenting λ has slowed down. More inventions can conceptually be the result of more efficient matching of inventors and firms, of matches of a given quality becoming more productive, or of more inventors.

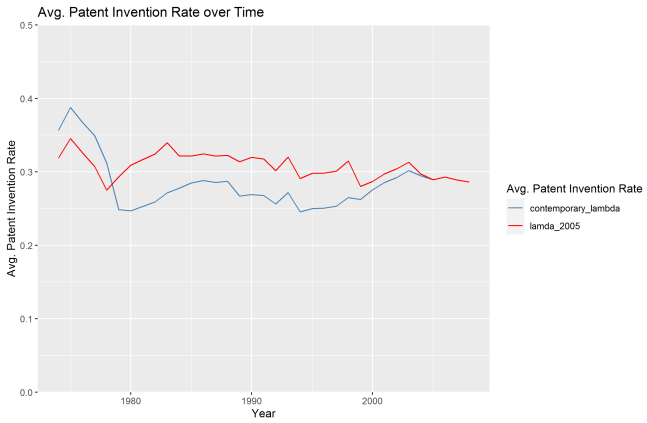

Figure 16: Average productivity of matches over time. The blue line gives the average estimated productivity of all matches formed in any one year. The blue line gives the same, but using the patent invention function of 2005-2010 for the whole data set. If matches could have used the patent invention function of 2005-2010, they would have produced more patents than they did. The only exception is at the very beginning of the data set (1975-1908), which is also the production function estimated with the lowest precision due to relatively few observations in many cells.

In general, the patent invention rate changes very little. However, if there is a trend at all, patents were slightly easier to produce with the patent invention function of 2005-2010 than they were before (figure 16). The patent invention rate of matches started between 1974-1979 is the highest overall, but it is the least precisely estimated rate, due to few matches in many of the bin combinations.

4.3 Concentration of Technological Competences

Patenting is a highly concentrated activity, even among those few firms who patent at all (Figure 1). Within patenting authorities and technology clusters, patents are still highly concentrated among the top 5% of firms (figure 17). Patenting at all major patenting authorities is also becoming more concentrated over time (figure 18). Patenting by small firms is declining the fastest.

Figure 19 shows that large technology clusters (with many patents) are more concentrated than smaller ones. Technology clusters where more inventions are patented each year have a higher share of innovative contributions by the top 5% of firms. The fitted relationship is positive even though the largest and most

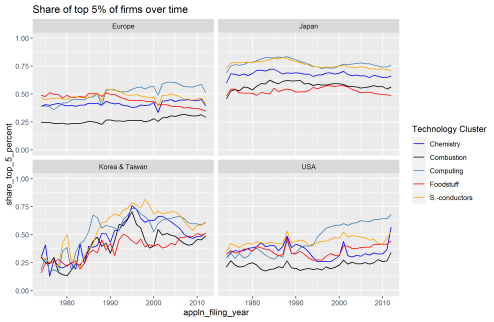

Figure 17: Concentration in the largest technology clusters over time. Concentration is measured as the share of inventions made by inventors working with the top 5% of firms (by patent output). If two inventors from different firms are listed under the same invention, both firms receive half an invention.

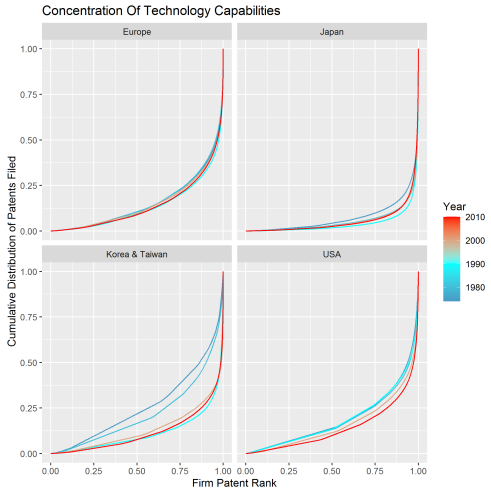

Figure 18: Cumulative distribution of patents per firm. Every firm is assigned its share of patents in each technology cluster and time. Concentration is continuously rising in the US and Korea & Taiwan, rising slowly in Europe and rising and then slightly falling in Japan.

concentrated technology clusters are beyond the right edge of the graph. Because concentrated fields are larger, the overall concentration is even higher than the average concentration within fields or IPC classes.

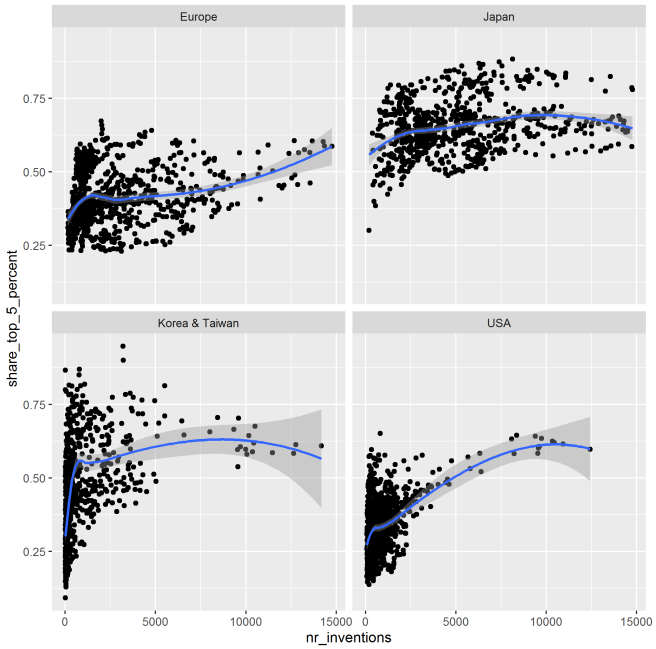

Figure 19: Concentration as a function of the number of inventions within a sector. Each observation marks a different technology cluster at one point in time. The x axis denotes how many patents were filed in one year, the y axis gives the share of the top 5% of firms. There is a positive correlation between the two: The bigger the technology cluster, the larger is the share of patenting done by the top 5% of firms.

All in all, innovation has been highly concentrated among few firms throughout the time period. Active technology clusters are also the more concentrated ones. Outside the absolute top, well established firms (with more than 50 patents every year) produce a large share of all patents. These firms are a tiny minority of all firms in the economy. The ability to regularly produce more than 50 patents represents a sizable investment from the firm, one that most other firms seem unable to make. Additionally, good inventors are overwhelmingly concentrated in large established firms, a trend that has increased over the time covered. Chapter **??** will explore potential causal forces behind this correlation.

4.4 Knowledge Production and Profits

I find that patenting increases profits faster than productivity. This is troubling in itself, since patents are granted to incentivize firms to produce public goods, not help them appropriate private profits. The results hold in a simple setting with an exogenous production function and when using a combination of the Doraszelski and Jamandreu (2013) and the **?** estimator to jointly estimate markups and the law of motion of productivity.

Table 1: Knowledge Production and Profits

	(1) $ln(EBIT)$	(2) $ln(\omega_{simple})$	(3) $ln(\omega)$	(4) μ	(5) $ln(EBIT)$
log(number of patents)	0.0536***	0.0046	0.0030	-0.0284	0.0566
	(0.0081)	(0.0042)	(0.0026)	(0.1269)	(0.0327)
Inflow Movers	-	-	0.0048*	-0.0007	0.1986***
			(0.0036)	(0.0602)	(0.0454)
Rank Firm	-	-	-0.0004	0.0024	-0.0031
			(0.0002)	(0.0046)	(0.0023)
Control Function	NO	NO	YES	YES	YES
Firm Fixed Effects	YES	YES	YES	YES	YES
Observations	27332	27332	14771	14771	14771

$ln(\omega_{simple})$, ω & $ln(EBIT)$ denote log productivity with a $\frac{2}{3}$; $\frac{1}{3}$ production function, log productivity with an estimated production function and profits. μ denotes firms' markups. Production function as in De Loecker & Warzynski 2012; effect of patents as in Doraszelski and Jamandreu (2013). Bootstrapped Standard Errors (n=500) in parentheses.

Table 1 reports the results of my estimation. Columns 1 and 2 report simple FE estimates, to alleviate concerns about the validity of production function estimation. According to these results, a firm can expect to increase profits by about 5% when they double their patenting output.

Columns 3 - 5 report the results from a joint semiparametric estimation of the production function and the law of motion of productivity. In the semiparametric estimation, I include the inflow of inventors moving to firms and the rank of the firm in the research quality rating as potential variables into the law of motion. This tests the hypotheses that moving inventors bring technological knowledge

with them or that a firm's research quality proxies for its absorptive capacities. Given the patents a firm is already applying for, research quality does not offer additional benefits. The number of inventors that move to the firm does however increase both productivity and profits. This further supports the hypothesis that moving inventors diffuse technology.

4.5 Inventor Mobility and Technology Diffusion

I find that between 2000 and 2010, inventors who are leaving top firms increasingly move to other top firms, instead of transferring their knowledge to less productive firms. Since technology personnel movement is an important driver of technology diffusion, this might explain the increasing gap between "The Best and the Rest" in a large number of countries (Andrews et al., 2016; Gal, 2017). This dispersion might also hurt overall productivity growth (Akcigit and Ates, 2019). Firms themselves rank retaining knowledgeable employees as one of their most important strategies for protecting intellectual property (Harhoff, 1997). To measure technology diffusion through moving inventors, I turn to the sample of patent data matched with firm productivity estimates and rank all firms within a region and five year time period. Splitting the resulting ranking into 50 productivity classes, I count movements of inventors from the top 10% of firms to the rest. I analyze the top 10% of firms to synchronize with the productivity dispersion literature, which considers the top 10% of firms "frontier firms" worthy of special attention. The matched sample between PATSTAT and AMADEUS is only large enough between 2000 and 2010. During this time frame, the trend of moving only between top firms is stable and persistent.

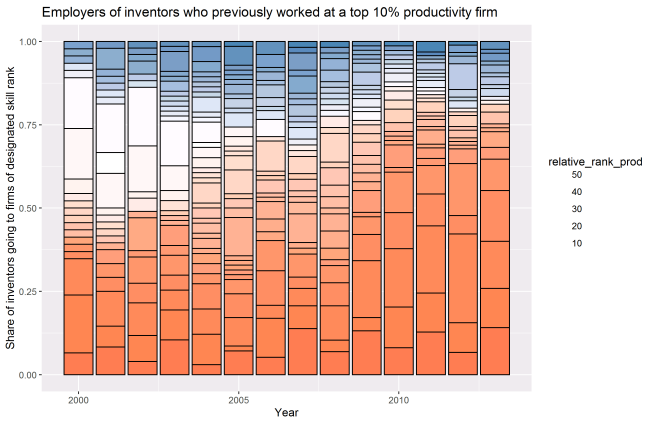

Figure 20: The graph shows the subsequent employers of inventors who have left a firm in the top 10% productivity decile. Firms are ranked according to their productivity and grouped into 50 different skill classes, 50 designating the most productive 2% of firms.

Figure 20 shows which firms inventors move to after having worked for a top firm. Movements from top firms to laggard firms are becoming less frequent over time. This decline is not simply driven by overall rising concentration, as the output share of the top firms has only increased moderately. Instead, the matching behavior of productive and unproductive firms is driving this change.
Documenting inventor rankings between firms is in the spirit of a branch of endogenous growth models focused on technology diffusion (Arkolakis et al., 2018). These models focus less on firms' R&D decisions and more on the random meetings of inventor-entrepreneurs and the resulting exchange of ideas. In these models, equilibrium productivity growth is determined both by how many new ideas are created and how fast these diffuse. While fast diffusion leads to faster growth, it also diminishes the incentives to invent new technologies, since the associated technological edge is lost quickly.

4.6 Summary of Empirical Findings

A major result of the paper is that inventors and firms have matched assortatively since 1974 and that this has increased over time. With the only notable exception of computing, most technology clusters exhibit this trend. Throughout the time period, the higher quality firms increased their quality weighted share

of researchers.
The estimated patent invention functions are largely stable over time and put more weight on inventor rather than firm quality. Because the contributions of firms to patenting probability are not large, rematching inventors and firms can only lead to small increases in production. In this, the paper comes to a similar conclusion as Hagedorn et al. (2017), although the method has been altered substantially and the context is different.
The stability of the patent invention function seems to contradict popular explanations of technology stagnation: It does not seem like inventions are getting harder to find. However, the arrival rate of patent families is also influenced by which projects firms decide to undertake in the first place: There is strong evidence that firms attempt more incremental and applied research projects (Arora et al., 2019), while patents with more scientific content are more valuable (**?**). If firms are racing each other to the same ideas more often, firms start smaller projects in equilibrium (Silipo, 2005).

5 Conclusion

I analyze the matching of firms and inventors and the productivity of the resulting matches as a potential driver of slowing technology growth. I document which matches are formed and how much each party contributes to patent invention. To answer these questions, I transfer empirical strategies used in the search and matching labor market literature to the PATSTAT patent data from 1974-2010, which I use as an employer-employee data set.
Assortative matching has risen over time in nearly all technology clusters. Highly skilled inventors increasingly match with firms with a high research quality. High quality firms produce more patents and are larger on average. Decreasing movement of inventors might cause a decrease in knowledge diffusion documented by Akcigit and Ates (2019); Andrews et al. (2016); Gal (2017). I merge conventional firm production productivity estimates to my data and find that less and less inventors move from top productivity firms to firms with lower productivity. I find that such movements are – when they still happen – associated with productivity and size increases for the receiving firm. The same pattern holds true

for newly invented patents between 2000 and 2010 (there is no matched data for earlier years).

It is often hypothesized that ideas are getting harder to find. This would mean that matches between inventors and firms of a given quality produce less patents today than they did previously. Yet, matches' estimated productivity is not declining: If matches from the 80s or 90s had used the patent invention function of 2005-2010, they would have produced slightly more patents, not less.

These results open up interesting avenues for future research: First, the paper already lays out a matching labor market estimator that can estimate the combined skill of teams and rank them together with single inventors. The next logical step is to incorporate these results and compare them to single-inventor results. E.g. a very supermodular within team patent invention function might explain (increasing) assortative matching, even if there is little gain from matching with highly productive firms per se: These might still serve as marketplaces where good inventors meet each other.

Second, narrowing the scope of my analysis to a specific country with high quality firm data would allow to assign specific product lines to specific technologies and measure markups with a higher degree of certainty. More complete data could be used to more precisely measure the contribution of assortative matching to productivity dispersion, markups and profits using state of the art markup estimators. I committed to a global scope for this paper to understand patenting behaviour across the developed world, since most contributions on the technology growth slowdown also assume that it is a global phenomenon. However, diving deeper into specific countries with a higher quality data set could enhance our understanding of firm level responses.

References

J. M. Abowd, F. Kramarz, D. N. Margolis, High Wage Workers and High Wage Firms, Econometrica 67 (1999) 251–333.

D. Card, J. Heining, P. Kline, Workplace Heterogeneity and the Rise of West German Wage Inequality, The Quarterly Journal of Economics 128 (2013) 967–1015.

M. Hagedorn, T. H. Law, I. Manovskii, Identifying Equilibrium Models of Labor Market Sorting, Econometrica 85 (2017) 29–65.

S. Bonhomme, T. Lamadon, E. Manresa, A Distributional Framework for Matched Employer Employee Data, SSRN Scholarly Paper ID 3333421, Social Science Research Network, Rochester, NY, 2017.

D. Harhoff, Incentives to Innovate: A Structural Model of Oligopoly (Available Only in German!), Technical Report FS IV 97-09, Wissenschaftszentrum Berlin (WZB), Research Unit: Competition and Innovation (CIG), 1997.

J. Pearce, Idea Production and Team Structure, Working Paper (2019) 84.

M. J. Andrews, L. Gill, T. Schank, R. Upward, High wage workers and low wage firms: Negative assortative matching or limited mobility bias?, Journal of the Royal Statistical Society: Series A (Statistics in Society) 171 (2008) 673–697.

D. Autor, D. Dorn, L. F. Katz, C. Patterson, J. Van Reenen, The Fall of the Labor Share and the Rise of Superstar Firms, Working Paper 23396, National Bureau of Economic Research, 2017.

J. De Loecker, J. Eeckhout, The Rise of Market Power and the Macroeconomic Implications, Working Paper 23687, National Bureau of Economic Research, 2017.

T. F. Bresnahan, M. Trajtenberg, General purpose technologies 'Engines of growth'?, Journal of Econometrics 65 (1995) 83–108.

P. Gal, Essays on the Role of Frictions for Firms, Sectors and the Macroeconomy, Tinbergen Institute, 2017.

D. A. Comin, M. Mestieri, An Intensive Exploration of Technology Diffusion, Technical Report 16379, National Bureau of Economic Research, Inc, 2010.

U. Akcigit, S. T. Ates, What Happened to U.S. Business Dynamism?, Technical Report, 2019.

T. Magerman, B. Van Looy, X. Song, Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization, SSRN Scholarly Paper ID 944470, Social Science Research Network, Rochester, NY, 2006.

B. Peeters, X. Song, J. Caellert, J. Grouwels, B. Van Looy, Harmonizing harmonized patentee names: An exploratory assessment of top patentees, Eurostat Working Paper (2010).

M. Peruzzi, G. Zachmann, R. Veugelers, Remerge: Regression Based Record Linkage With An Application To PATSTAT, Bruegel Working Paper 10 (2014).

P. Pons, M. Latapy, Computing Communities in Large Networks Using Random Walks, in: p. Yolum, T. Güngör, F. Gürgen, C. Özturan (Eds.), Computer and Information Sciences - ISCIS 2005, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2005, pp. 284–293.

A. Toole, C. Jones, S. Madhavan, PatentsView: An Open Data Platform to Advance Science and Technology Policy, SSRN Scholarly Paper ID 3874213, Social Science Research Network, Rochester, NY, 2021.

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

U. Akcigit, S. Caicedo, E. Miguelez, S. Stantcheva, V. Sterzi, Dancing with the Stars: Innovation Through Interactions, Working Paper 24466, National Bureau of Economic Research, 2018.

T. Lamadon, E. Manresa, S. Bonhomme, A Distributional Framework for Matched Employer Employee Data, Technical Report 1399, Society for Economic Dynamics, 2015.

H. Dino, S. Yu, L. Wan, M. Wang, K. Zhang, H. Guo, I. Hussain, Detecting leaders and key members of scientific teams in co-authorship networks, Computers & Electrical Engineering 85 (2020) 106703.

S. Bonhomme, T. Lamadon, E. Manresa, A Distributional Framework for Matched Employer Employee Data, Econometrica 87 (2019) 699–739.

E. Moretti, The Effect of High-Tech Clusters on the Productivity of Top Inventors, Working Paper 26270, National Bureau of Economic Research, 2019.

E. Moreno-Centeno, A. R. Escobedo, Axiomatic aggregation of incomplete rankings, IIE Transactions 48 (2016) 475–488.

U. Doraszelski, J. Jamandreu, R&D and Productivity: Estimating Endogenous Productivity, The Review of Economic Studies 80 (2013) 1338–1383.

N. Bloom, C. I. Jones, J. Van Reenen, M. Webb, Are Ideas Getting Harder to Find?, Working Paper 23782, National Bureau of Economic Research, 2017.

D. Andrews, C. Criscuolo, P. Gal, The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy, OECD Productivity Working Papers 5 (2016).

C. Arkolakis, T. Papageorgiou, O. A. Timoshenko, Firm learning and growth, Review of Economic Dynamics 27 (2018) 146–168.

A. Arora, S. Belenzon, A. Patacconi, J. Suh, The Changing Structure of American Innovation: Some Cautionary Remarks for Economic Growth, Working Paper 25893, National Bureau of Economic Research, 2019.

D. B. Silipo, The Evolution of Cooperation in Patent Races:Theory and Experimental Evidence, J Econ 85 (2005) 1–38.

U. Akcigit, W. R. Kerr, Growth through Heterogeneous Innovations, Journal of Political Economy 126 (2018) 1374–1443.

D. Karlis, An EM algorithm for multivariate Poisson distribution and related models, Journal of Applied Statistics 30 (2003) 63–77.

Appendices

A Using Patent Data as an Employer-Employee Data Set

Since PATSTAT does not contain IDs, only string names, I consolidate spelling mistakes and disambiguate entities with the same name before using the data. This appendix describes the procedure.

First, Magerman et al. (2006) have already constructed consolidated identifiers by correcting spelling mistakes, omitting titles and reading out abbreviations

like "Ltd.". They have also constructed a sector variable, which assigns names in the database to categories like "company", "individual", "university" etc. After fusing such different spellings of the same name, they find an additional 30% of patents for the top 450 applicants, compared to the raw HAN identifiers provided by PATSTAT.

Second, Peeters et al. (2010) have manually checked the record of the top 450 applicants and searched for additional possible variants in the data. They can assign another 30% of patents to these applicants. However, since some of these applicants have over 100.000 patents in different countries, different spellings and mistakes play a much larger role than in the general population.

To disambiguate additional names both on the inventor and firm side, I clean names similarly to Magerman et al. (2006) and then sort all words alphabetically. This equates reversed spellings of names like "Erik van Houten" and "van Houten, Erik". This reduces the number of unique inventor identifiers by another 25%. I additionally clean firm names of addresses that are sporadically entered in the field "name", e.g. "Intel Corporation, Santa Clara, CA". This fuses around 3% of the remaining firm identifiers.

To gauge the quality of the resulting ID, I draw a list of prominent inventors from Wikipedia and link them to our data. Just as Peeters et al. (2010) for the firm side, I find that these highly active individuals are split over multiple IDs due to spelling mistakes, different name formats etc. However, the automated correction of Magerman et al. (2006) already does a decent job of aggregating them: After manual search, I e.g. link 38 PATSTAT person IDs to the most prolific inventor in the world (Dr. Shunpai Yamazaki). Magerman et al. (2006) already linked the most important 30, so I can only marginally improve upon their results. My 38 IDs participate in 5585 patent families across the world while the 30 IDs of Magerman et al. (2006) participate in 5581. The newly discovered name variants are clearly errors that only show up once. In addition, such spelling variants often show up within a patent family where the inventor is also cited on other patents. The patent family is the relevant unit of observation. Thus, even if undetected spelling variants exist, they are largely irrelevant to my productivity measures. I thus have confidence that the IDs provided by

Magerman et al. (2006) capture the large majority of an inventor's patents.

However, this still leaves the problem that some names might belong to more than one inventor. Combining such inventors into one person would create the impression of a prolific inventor frequently moving between firms.

First, I collect the frequency with which words occur in the inventor names submitted on patents in each country. I then eliminate inventor names that do not contain two infrequent words: E.g., "Erik van Houten" contains two words common in Dutch names ("Erik" and "van") and only one uncommon word "Houten". Thus, I will not consider this inventor in the sample.

Second, PATSTAT contains the IPC classes associated with each inventor's patents. Inventors will typically not master a variety of technical fields and thus names with more diverse portfolios are more likely to stand for more than one inventor. Specifically, I exclude workers whose most common IPC 4-digit category accounts for 20% or less of their patents, whose top technology field accounts for 50% or less of their patents and whose top two technology fields account for 80% or less of their patents. I check these numbers against the statistics for inventors crosschecked with Wikipedia to guarantee that these criteria are not too strict.

Third, I exclude inventors from the sample who were active for more than 40 years, on the basis that these are likely overlapping inventors of the same name.

The observed time span, the diversity of IPC classes and technology communities and the number of distinct names are conceptually different criteria. Nonetheless, they are reasonably correlated (0.15-0.6), which suggests that the criteria identify suspect inventors reliably.

B Constructing Technology Clusters from IPC Classes

This appendix details how I extract sub-labor markets from the IPC classes of inventors' patents. Each patent is assigned to one or more IPC classes that de-

scribe its technological contents. Not all inventors are interchangeable. Not all inventors can work on all research projects. I.e., there is horizontal as well as vertical differentiation between inventors. To sort all inventors into one ranking would thus be misguided. The goal of the algorithm below is to separate inventors into groups: These groups have to be so small that every inventor in the group can contribute to the work of the other members of the group. However, they have to be so large as to include every inventor who could substitute for the members of the group.

I reduce the sample to all patents with only one inventor, so that the assignation of IPC classes to inventors is unambiguous. I observe the succession of combinations of IPC 4-digit classes every inventor patents in, sorting IPC classes from the same patent or from patents in the same year in a random order. From this I compute the conditional probability of an inventor whose last patent was in one combination to move to another one. E.g.: If only one inventor ever applies for a patent with the IPC class "A01P" and then patents in the class "B06P", I would conclude that the two combinations are very similar, since 100% of inventors moved from one to the other.

I find that even IPC 8-digit classes form a network that is only sparsely connected by moving inventors: Most inventors only patent in very few classes and mobility between classes is rare.

This sparseness of the network also determines my strategy for defining clusters: Since the network has many nodes, few strong edges and the number of final clusters and their geometrical forms are unknown, density based clustering will efficiently yield the network structure.

First, I assign every IPC class with less than 1% of patents to the nearest class with more than 1%, to avoid many small clusters with few inventors. I then use density clustering among the large IPC classes to determine which classes to combine into clusters. The "knee" in the 3-nearest neighbor cdf is at roughly 0.11, which I take as the ϵ for density based clustering: All connections with a movement probability of 11% or higher are selected into the same cluster.

The result of this procedure is a stable assignment of IPC classes to technol-

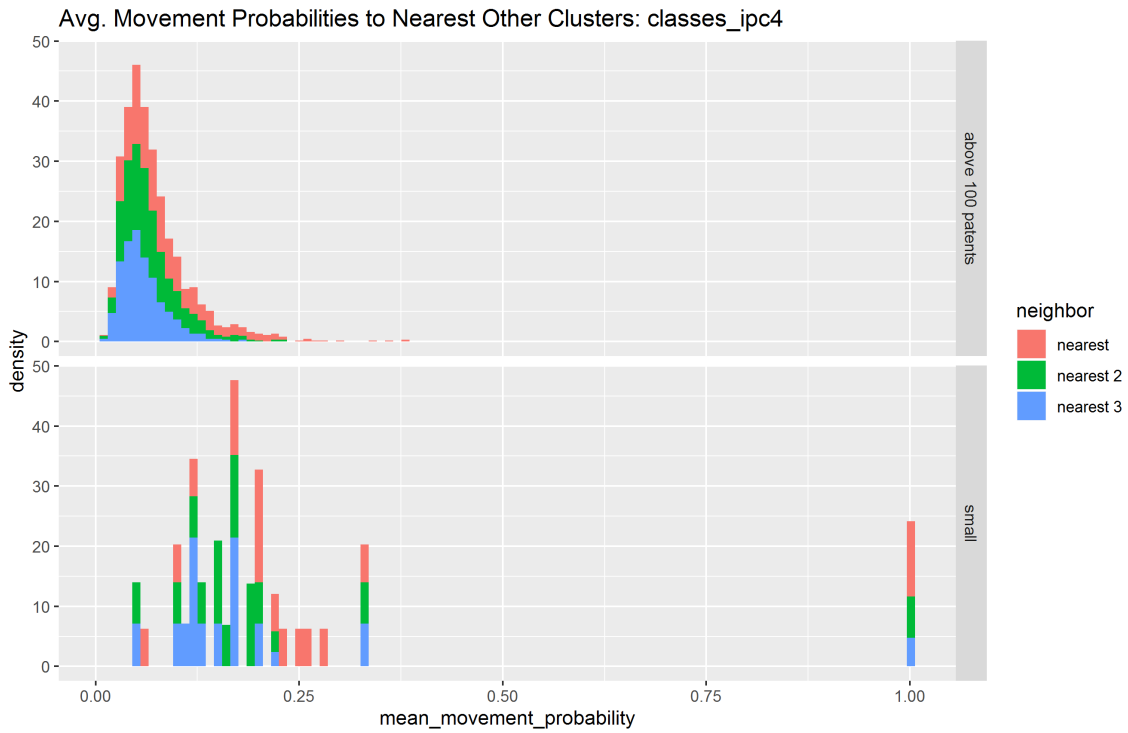

Figure 21: The figure shows the distribution of movement probabilities between IPC class combinations. The top figure shows the distribution for common IPC classes (with more than 100 patents). The bottom figure shows the same distribution for the less common classes. Evidently, less frequent classes are often strongly connected. In contrast, the frequent classes stand more alone. Red, green and blue describe the movement probabilities to the nearest, nearest two and nearest three other classes.

ogy clusters. Around 20% of patents are part of the largest cluster. Figure 23 shows the technology field assignment of the biggest IPC groups and the strength of their relations with each other. For comparison, figure 24 also reports the community assignments of an alternative community finding algorithm for large data sets (Pons and Latapy, 2005). It results in significantly larger communities because small nodes with connections to two large communities often are enough to connect the two large communities. This happens even though these small notes represent a negligible amount of patents. Hence, the algorithm is highly sensible to which small nodes and weak edges are considered. If neither are included, the largest community contains roughly 50% of all patents, which is not plausible as a sub-labor market: Mobility within this large community is low.

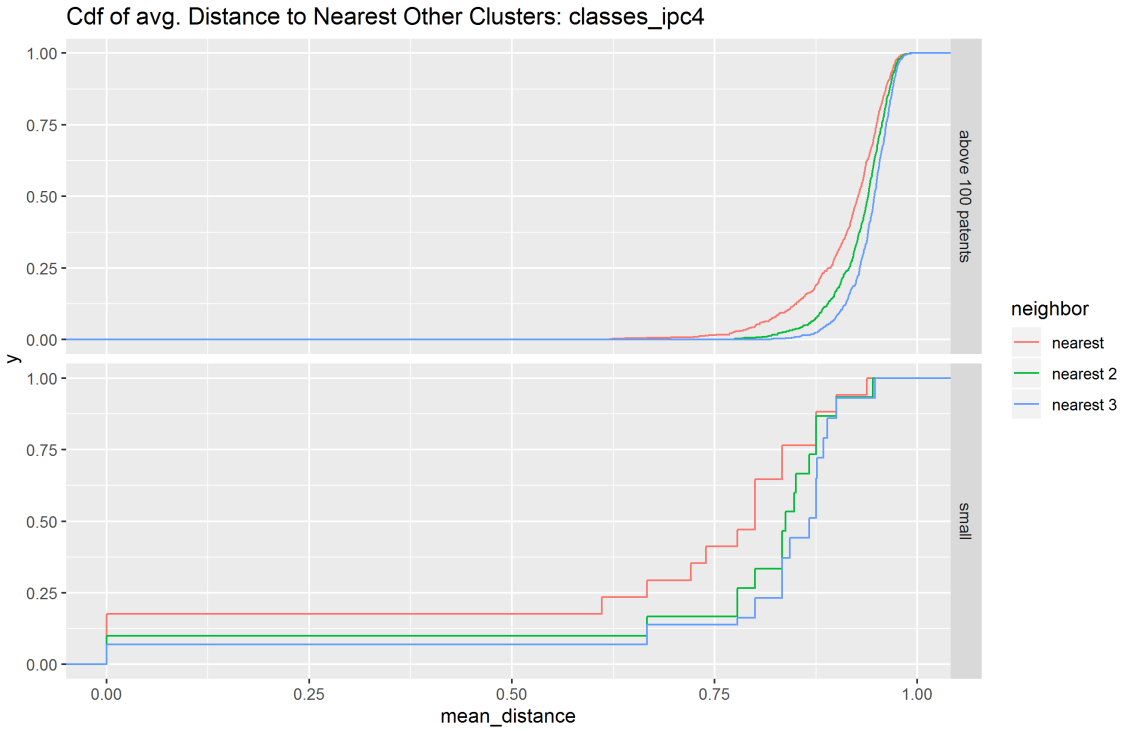

Figure 22: knn distance plot of inventor movement probabilities. Distance is defined as 1- movement probability. Note the uptake of the cdf for big clusters at roughly 0.11. I will take this as the ϵ parameter in density based clustering. Thus, most small IPC class combinations will be fused into clusters.

C Truncation Correction

In this framework, the expected number of patents per year λ_x^y is constant within one match. Specifically, I treat λ_x^y as the Poisson arrival rate of new patents, given x and y. Each match exists for a given number of years (l_{true}). Let type j denote all employment spells with the same l_{true} and λ_{true}. I understand the untruncated data as generated by a mixture distribution of different types of employment spells. I define $s_{l;\lambda}$ as the share of type $l;\lambda$ in the overall mix of types. E.g., all employment spells lasting 5 years and producing 0.3 patents per year would be considered a type, with $s_{5;0.3}$ giving the share of such employment spells in all spells.

$\boldsymbol{s}$, the vector of the individual type's population shares, has to be recovered from the observed minimum length of employment spells l_{ob} (the time between the first and the last patent) and the distribution of patents during these years. The only

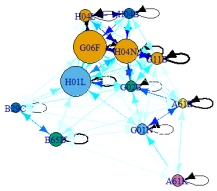

Figure 23: Technology fields of and connections between the largest IPC classes. Classes were grouped into fields using a density based algorithm that groups together all classes connected through an inventor movement probability above 11%. Thicker and darker arrows denote more movements of inventors from one technology class to the other. The size of the classes denotes the number of patents assigned to each class.

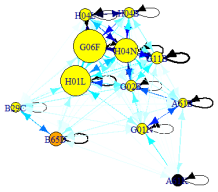

Figure 24: Technology fields of and connections between the largest IPC classes. Classes were grouped into fields using the walktrap algorithm of Pons and Latapy (2005). Thicker and darker arrows denote more movements of inventors from one technology class to the other. The size of the classes denotes the number of patents assigned to each class.

additional assumption necessary is that workers do not leave a firm between two observed patents, so that I can recover a minimum length of each spell from the data. I will recover $\hat{\boldsymbol{s}}$ and from this construct unbiased estimates $\hat{\lambda_x^y}$ and $\hat{l_x^y}$ for each observation from the estimated distribution of true types.
This procedure is necessary since the estimates $\hat{\lambda_x^y}$ and $\hat{l}$ given P_{ob} & l_{ob} cannot be computed for each match in isolation: Consider a match for which I observe one patent in one year. This observation could either be an unproductive match that lasted for a long time or a productive, but short lived one. The one data point itself is not informative on the matter. However, if unproductive and long matches were common in the data, I would also observe some of them. Thus, the whole observed distribution informs my estimate for one specific observation. Therefore, it is necessary to analyze the whole distribution jointly.
Given that the above setup already assumes a Poisson distribution for patents, a maximum likelihood estimator does not require additional assumptions, but is more efficient. Unfortunately, it requires the optimization of a non-linear log likelihood function over several 1000 parameters, so it is only feasible when making additional simplifying assumptions. Therefore, I estimate the spell distribution using GMM.
Given the above mixture distribution

$$N_{ob} = \sum_{j}^{J} p(ob|j) * N(j) \tag{5}$$

N_{ob} denotes the expected number of times a specific outcome (like 2 patents interrupted by a year of inactivity) is observed. It equals the sum of the expected number of occurrences given each of the specific types. $p(ob|j)$ is a constant number: E.g. the chance to observe the above two patents interrupted by one year of silence for type $5; 0.3$ is about 2%. Treating $N(j)$ as the coefficients to be estimated, one has a data set with several million different possible outcomes and how often they have occured in the data, which one can use to estimate the several thousand $N(j)$s. Note that since $N(j)$ has to be greater than 0, this is not strictly linear. However, it is still computable in a very reasonable time frame:
Because of the positivity constraint on all coefficients, there is no analytical solution and several possible numerical techniques exist. Aside from estimating the whole system of equations jointly, splitting the data is a possibility, too: Since

observations with e.g. 20 observed years can only be created by spells of at least length 20, one could estimate longer employment lengths first and then "cascade" down the spell lengths. Additionally, it is numerically hard to recover the distribution for very short spells because there are comparatively few different observable outcomes. I test specifications where I restrict the underlying true employment spells to be either at least 2 or at least 4 years long. I either use the whole distribution in the estimation (including the very short outcomes), or exclude the shortest observed outcomes from the estimation as well. This leads to 8 different numerical techniques.

Since their qualitative results are very similar and additional assumptions do not seem to yield more stable results, I opt to estimate using all available data and assuming that no employment spell is shorter than 4 years. While this method leads to a slightly better fit, all different strategies yield very comparable estimates:

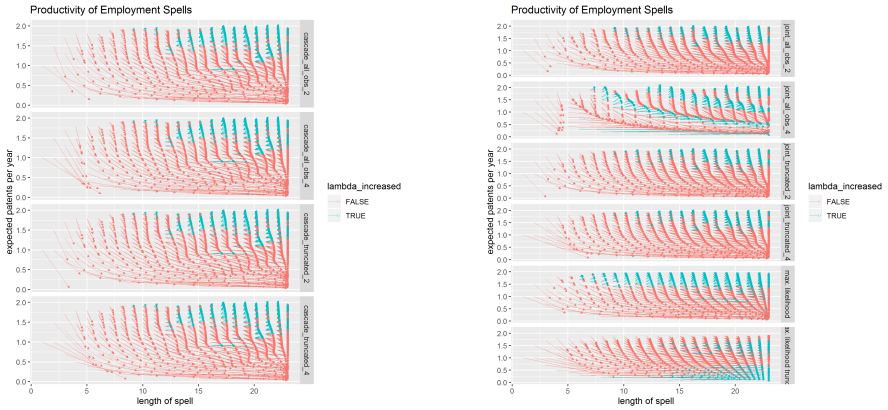

Figure 25: Adjustment of observed employment spell productivity and length. The starting point of each arrow is the productivity and length observed in the data without the correction routine. The end point of each arrow gives the new estimated arrival rate of patents after the routine has concluded. Red highlights spells where the observed productivity was adjusted downward, blue highlights spells where the observed productivity was adjusted upwards. Each panel is the result of a slightly different numerical technique. The left table contains results when imputing for each spell length separately, the right table contains the results when fitting to the whole data at once. Within each table, the top two panels report the result when assuming that employment spells last at least 4 or 2 years respectively. The lower two panels report results when making the same assumption, but also only targeting the part of the data that contains at least 4 or 2 consecutive years. All methods come to broadly similar conclusions. Estimating for each spell length separately is less efficient, but much faster.

The Poisson distribution underlies all of the above estimations. This distribution is used both in theoretical models (Akcigit and Kerr, 2018) and in empirical applications throughout a vast range of scenarios, even including sport scores (Karlis, 2003). The central assumption of the Poisson distribution is that the arrival rate of events is constant, which seems suspect in many circumstances including patents: It seems reasonable that after a successfully completed project, the arrival rate of success falls and then rises again after some time has passed. However, in practice, it seems that inventors work on different projects simultaneously so that a constant arrival rate is a good fit for the data. The only systematic forecasting error of the Poisson model is for very successful spells: The model assumes that inventors with multiple patents per year can uphold their performance, which seems to not always be the case. However, this concerns a negligible number of inventors. 26 reports systematic mismatch of the Poisson model of patent invention for each numerical technique. Evidently, the fit is very good for all outcomes except rare and high productivity outcomes.

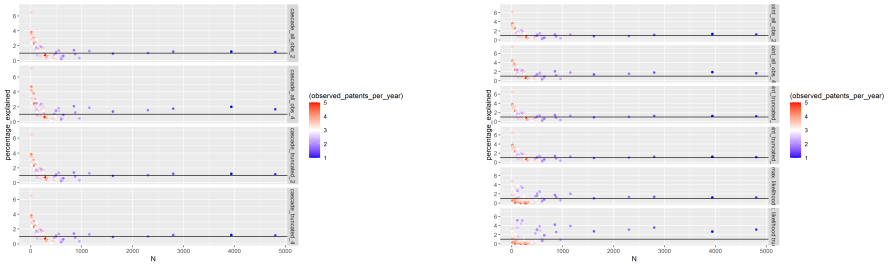

Figure 26: The ratio of the predicted times each outcome should occur and the actual number of occurrences. The fit is very good for results occurring more often, there is a slight upward bias for very rare and very productive spells, whose frequency of occurrences is overestimated. Each panel is the result of a slightly different numerical technique. The left table contains results when imputing for each spell length separately, the right table contains the results when fitting to the whole data at once. Within each table, the top two panels report the result when assuming that employment spells last at least 4 or 2 years respectively. The lower two panels report results when making the same assumption, but also only targeting the part of the data that contains at least 4 or 2 consecutive years. All methods come to broadly similar conclusions. Estimating for each spell length separately is less efficient, but much faster.

Once I have estimated a distribution of λ_i^f and (l) with either GMM or ML, I compute the implied shares of all types, given any realization of P_{ob} & l_{ob}. I

can thus derive my final estimates $\hat{p}$ & $\hat{l}$ for any observation. Based on the same technique, I can also obtain an estimate of how likely it is to observe the underlying spell at all.

D Simulation: Description and Results

While the Hagedorn et al. (2017) method is a consistent estimator, its performance in the actual data is more unclear: Since there are usually only a few observations for every inventor, a consistent estimator might not perform well in practice. A Monte Carlo simulation will reveal the estimator's performance in more realistic samples.

The simulated data covers a 40 year stretch of a technology cluster, just like the actual data. Every year, inventors enter the technology cluster. However, not all matches produce a patent and thus not all matches are observed. With around 50.000 observed inventors the simulated data is as large as the smallest actual technology cluster.

Inventors have a constant chance ρ to match with a firm. They then compare the firm's quality to the quality of their current firms and decide whether to switch or not. Whenever inventors have to decide between two firms, they will pick the one with the higher quality: Since the higher quality firm will produce more patents, it can offer a higher wage and secure the inventor.

In the main specification, the parameters of the model are as follows:

- The patent invention function is $\lambda_{x_i,y_f} = x_i * y_f$
- Inventors match with a new random firm with a 5% probability every year
- Inventors leave the economy with a 10% probability every year
- Inventors and firms will reject any matches that do not at least produce $\frac{skill^2+quality^2}{2.05}$

The algorithm to recover the estimates contains four steps:
First, using the estimated overall distribution of employment spells (appendix C), the algorithm computes the unconditional distribution of potential types for every observed spell: It computes that e.g. 5% of all observed spells with one patent are produced by employment spells with a patent arrival rate of 0.5 and a length of 7.
Second, using this underlying distribution, the algorithm draws 20 potential underlying productivities for every observed spell. These serve as hypotheses about the actual "true" productivity that led to the observed patent outcome.
Third, the algorithm prunes these hypotheses: It computes how many observed spells of a certain type we expect to see, given the drawn productivities. In the case of a firm with 100 employees, all of which have only one patent, it would e.g. conclude that a patent arrival rate of 0.5 and a length of 7 is an unreasonable hypothesis for these spells: The firm would also have to have more successful inventors. The "only one patent"-outcome is overrepresented in the data. The algorithm sequentially prunes these hypotheses, recomputing the expected distribution after each discarded draw. The algorithm stops after only 5 hypotheses are remaining.
Lastly, the algorithm runs the whole ranking estimator five times, once for each set of drawn productivities. This allows to estimate how sensitive an estimated rank is to plausible variations in productivities and thus how large the estimation error for each ranking is.

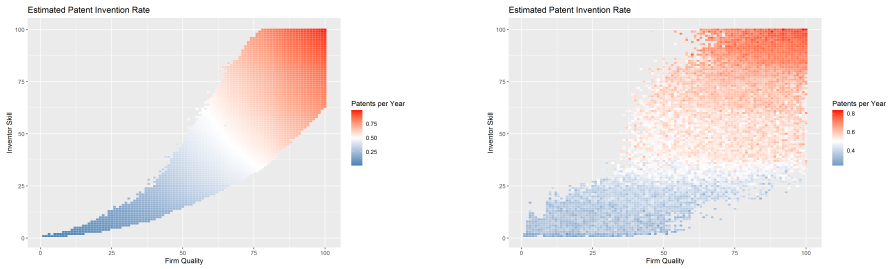

Figure 27: The left graph reports the patent invention function using the true parameters of the model: Combinations of inventors and firms with higher skill produce more patents according to $\lambda_{x_i,y_f} = x_i * y_f$. Grey areas of the matching plane have no matches in them, so the patent arrival rate is not reported for these.

Just using global rankings, the estimator recovers these parameters reliably: Figure 27 compares the estimated and the true production function, which are by and large identical. Figure 28 shows the number of matches for each skill and quality level. The estimator also recovers the core region with match support reliably. Only some single matches are estimated to be in actually empty regions of the matching plane.

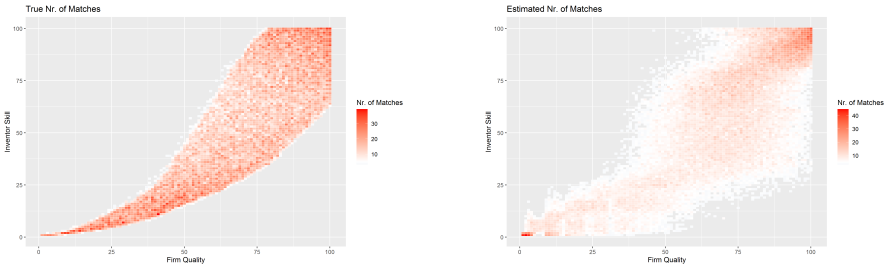

Figure 28: The left graph shows the matching behavior of inventors and firms, using their true skill and quality bins. The grey area has no matches, because inventors and firms reject matches that do not produce a positive matching surplus. The right panel shows the estimated distribution of matches. Apart from some single matches in the empty regions, the estimate recovers the true distribution well.