Technology Polarization*

Minoru Kitahara[1] and Koki Oikawa[†2]

[1]Department of Economics, Osaka City University
[2]School of Social Sciences, Waseda University

April 28, 2016

Preliminary draft. Please don't cite.

Abstract

We construct a new method to describe firm distributions within technology fields and the dynamic behavior of these distributions. To locate firms on a technology space, we apply multidimensional scaling for the inter-firm technological dissimilarity matrices that are computed from patent citation overlaps among firms using the NBER US patent dataset. Our estimated firm distributions show increasing trends in technological distance and *polarization* on average, where we follow Duclos et al. (2004) to measure polarization. We construct a model of inter-group competition in which polarization stimulates R&D incentives. The model fits data before the major patent reform in the United States in 1980s and polarization increases citation-weighted number of patent applications. After about 1990, the impact of polarization is reversed.

*We are grateful to Diego Comin, Ryoji Hiraguchi, Hajime Katayama, Brinja Meiseberg, Atsushi Oyama, Andrew Wait, and seminar and conference participants at GRIPS, Hitotsubashi University, Musashi University, Waseda University, Gakushuin University, JEA, WEAI, and the Workshop on Economic Growth and STI Policies, for their helpful comments and discussions. This work is supported by the Japan Society for the Promotion of Science, Grant-in-Aid C (25380220) and the Research Institute of Science and Technology for Society, Japan Science and Technology Agency. All remaining errors are ours.

†Email: oikawa.koki@gmail.com

Keywords: Patent citation overlaps; polarization; inter-group competition; multi-dimensional scaling
JEL classification: O31, O32, L25

1 Introduction

Since Jaffe (1986) introduced technological distance (or proximity) between firms with using patent data to capture knowledge spillovers, researchers in economics and innovation management have used it to estimate technological relations among firms (Jaffe (1989), Rosenkopf and Almeida (2003), Benner and Waldfogel (2008), Bloom et al. (2013), and so on). They mainly consider how firms' positions in technology spaces and resulting accecible knowledge affect innovation, stock value, productivity, M&A and alliance at the firm level.

This paper investigates the distribution of firm positions in technological spaces and its relation to innovation output at the aggregate level rather than the firm level. Our main question is: Under what type of distributions are innovations stimulated? To see the key factors to answer this question, imagine the following three extreme distributions in a technological space. First, if most firms concentrate on a technological position, they unintentionally help each other through knowledge spillovers because a firm can know and facilitate a new technology developed by another firm relatively easily when they are closely related.[1] Second, if we have a wide-spread firm distribution in a technology space, the spillover effect is relatively small. This is the opposite case to concentrated distributions.[2] Lastly but the most importantly in this paper, consider the middle. Suppose that we have a firm distribution in a technology space such that there are two polars, i.e., concentration points, distant from each other. There are two scenarios to explain this type of distribution. One is segmentation of technology. Even in the same technology space, polars are not technologically related. In this scenario, it is better to consider that the current technological classification should be subdivided. The other scenario is *inter-group competition.* Firms

[1]Since we consider within-category relations between firms, the distance in the product market tends to be low and only weakly depends on technological distance. If we consider broad technology field, we should care about both types of distance as in Bloom et al. (2013). Market competition and R&D incentive do not have a monotonic relation (Aghion et al. (2005)).

[2]Surely, knowledge spillover is not the only impact of technological distance in determining innovations. Closer relation may cause more patent infringements that discourage R&D investment. Diversity could be a virtue of wide-spread distributions (Weitzman (1998)).

around each polar use distinct fundamental technologies and they compete a race for becoming a (de facto) standard in the technology field.[3] In this scenario, they have more incentive to innovate because the winning polar will grab all demands. The level of market competition does not decay in technological distance because the products from the two polars are close in the market.

Several instances of competition for the de factor standard, such as the videotape format war between betamax and VHS, and between producers of operating systems for computers, indicate the historical existence of technology groups and inter-group competition. Open innovation strategies may also induce technology groups. A typical example is that IBM released its software patents in 2005 and induce other firms to develop Linux. Recently, the largest automobile company in Japan made its fuel cell vehicle patents free for use to facilitate entry by other firms, which is an example that a firm tries to generate a technology group (TOYOTA Motor Corporation (2015)).

The degree of *polarization*, developed by the series of papers by Esteban and Ray (Esteban and Ray (1994), Esteban and Ray (2011), and so on), responds such inter-group competition. Intuitively, high polarization results when there are two distinct density masses (polars) with large distance between them, while low polarization is attained when a distribution has only one mass point or if the distribution is equally dispersed like a uniform distribution. We apply the continuous version of the polarization defined in Duclos et al. (2004) (referred to as DER below) to firm distribution in a technology space. This paper is the first work that applies their formalization of polarization to R&D activities.

To obtain the distribution of firms in technology spaces, we use two methods of Stuart and Podolny (1996) with some modifications: patent citation overlaps and multi-dimensional scaling. We choose citation overlaps to examine the technological similarity; this allows us to look at the distribution of firms within technological categories. Other standard methods utilize patent portfolios, within-firm distributions of patents across categories (Jaffe (1986), Jaffe (1989), Benner and Waldfogel (2008), Bar and Leiponen (2012), Bloom et al. (2013), etc.), are not suitable to consider changes inside of the categories.[4] Since the original definition of citation overlap be-

[3]This story may be considered as a race between technological trajectories or paradigms. (cf. Dosi (1982))

[4]Akcigit et al. (2013) define a measure based on patent-level technological distance using overlaps of technology classes among citations to measure missallocation of technology.

tween two firms in Stuart and Podolny (1996) is not independent of a third firm, we modify it to satisfy independence of pair-wise similarity from third firms as illustrated in the next section.[5]

Multidimensional scaling (MDS, hereafter) is a statistical tool to estimate the location of entities by minimizing the sum of squared gaps between dissimilarity and the resulting distance, when dissimilarities among entities are given. Dissimilarity does not have to be a mathematical distance in MDS. MDS is not popular in economics but is one of the typical ways to analyze relational data in behavioral sciences (cf. Cox and Cox (2001)).

Our main findings are as follows. First, we find that the average technological distance and the average polarization have displayed an upward trend in the last three decades in the United States. Second, we estimate the impacts of polarization on the number of citation-weighted patent applications. Our model of inter-group competition implies that polarization raises innovation. But the model fits data only before 1990 and the impact is completely reversed afterwards. We attribute the structural change to the major patent reform in the United States in 1980s that changes institutions from anti-patent to pro-patent.

Data We used the NBER US patent dataset[6]. The dataset provides information on patents granted by the USPTO up to 2006. For patents granted after 1975, the dataset supplies the citation list of each patent (only for those registered in the US patent office, though). The dataset also provides information about changes in patent ownership. Thus, we can specify the original inventors of technologies. In this paper, we basically consider the 2-digit classification defined in Hall et al. (2001), which they call *subcategories*. We omit 6 "miscellaneous" categories out of 37 subcategories because they are not suitable for our purpose. The list of subcategories is summarized in Table 7 in Appendix B. The NBER US patent dataset contains firm identification numbers defined by Compustat for private firms, with which we link the patent data to firm data. The technological distributions are computed for 21 mutually overlapping 5-year windows such as 1976-1980, 1977-1981, ..., 1996-2000.[7]

[5] Stuart and Podolny (1996) are interested in firms' "local search" for a new technology. Firms only have bounded information and tend to look at R&D activities of closely-related firms. Thus, they use the "community matrix" which is developed in social psycology to describe personal familiality.

[6] http://www.nber.org/patents/. A detailed description of the dataset is in Hall et al. (2001).

[7] Stuart and Podolny (1996) also use 5-year windows. Benner and Waldfogel (2008) recommend aggregation of patent data across years into 5- or 10-year periods.

The rest of the paper is organized as follows. Section 2 is devoted to the description of the measurement methodologies including citation overlaps, multi-dimensional scaling, and 2-dimensional kernel density estimation. Section 3 presents a simple model to connect polarization and R&D incentives, and defines the degree of polarization. In Section 4, we investigate the impact of polarization on innovation. Section 5 discusses our results.

2 Technological Distance among Firms in each Technological Category

As mentioned in the introduction, the traditional measures of technological distance are based on the patent portfolio vector of each firm, which contains information about the within-firm distribution of patent holdings over technological categories. For investigating firm distributions within categories, another type of technological distance is needed. In this section, we construct a new measure of technological (dis-)similarity among firms based on patent citation overlaps.

2.1 Citation Overlaps

The first-order citation overlaps between firm i and j are from patent citation lists of the two firms within a period, say P_i and P_j. Since some patents are frequently cited by the same firm, the elements in each list is not unique in general. We consider such a frequently cited patent as important for the firm. When a citation overlap occurs at such an important patent, the overlap contributes to technological closeness more than an overlap that occurred among one-time cited patents.[8] To incorporate this idea, we keep repetition in each citation list. Define $O(P_i, P_j)$ as the patents in P_i that overlap those in P_j with repetition (we do not say "intersection" because the elements are not unique in general). The first-order citation overlap, ω_{ij}^1, is defined as

$$\omega_{ij}^1 \equiv |O(P_i, P_j)| + |O(P_j, P_i)|\,, \tag{1}$$

where $|P|$ is the number of patents in a list, P. Figure 1 illustrates an exapmle with $P_i = \{1, 1, 2, 3, 4\}$ and $P_j = \{1, 3, 5\}$, where each number indicates a patent. ω_{ij}^1

[8] For example, the citation list pair of $\{1, 1, 1, 2\}$ and $\{1\}$ should be more overlapped than a pair like $\{1, 2, 2, 2\}$ and $\{1\}$.

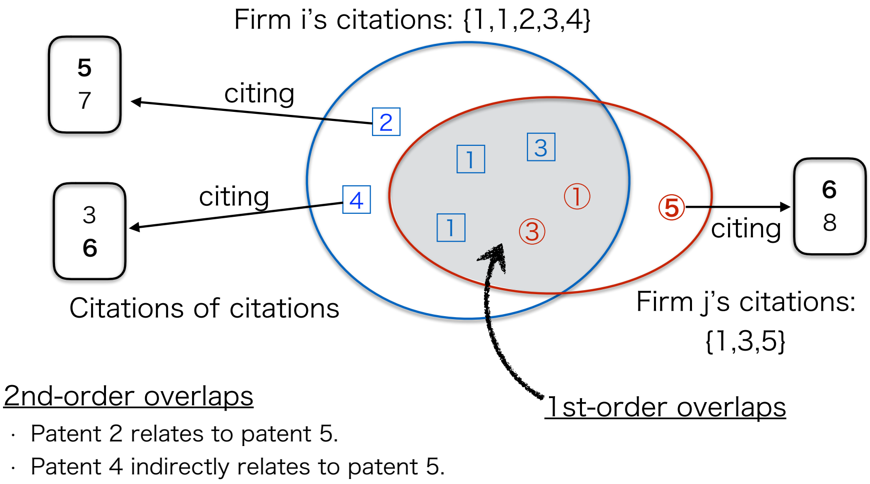

Figure 1: Citation overlaps.

counts the patents in the shaded area.

Even though a citation does not directly overlap, it could be technologically related indirectly. For example, patent 2 cited by firm i could cite patent 5 cited by firm j in Figure 1. Moreover, patent 2 could cite a patent cited by patent 5. To capture these indirect overlaps, we define the second-order overlaps.

Let $\tilde{P}_{ij}$ be the items of P_i that do not overlap P_j. Let $C_i(p)$ be patent p's citations but not included in P_i. This elimination is reasonable to avoid overevaluation of similarity. If we use all citations of p in calculation and if some citations are included in the first-order overlaps, we add relation between firm i's own citations on different levels.

The idea of the second-order overlap is that we put a positive weight on $p_k \in \tilde{P}_{ij}$ if p_k cites a patent in $\tilde{P}_{ji}$ or a patent cited by any patent in $\tilde{P}_{ji}$. Thus, it consists of two components. The first component picks up patents in $\tilde{P}_{ij}$ citing any of $\tilde{P}_{ji}$,

$$\omega_{ij}^{21} = \sum_{k=1}^{n_{ij}} \frac{\left| O\left(C_i(p_k), \tilde{P}_{ji}\right)\right|}{|C_i(p_k)|}. \tag{2}$$

ω_{ji}^{21} is similarly defined.

The second component of the second-order overlaps considers the patents in $\tilde{P}_{ij}$ that do not overlap $\tilde{P}_{ji}$, say $\tilde{P}'_{ij}$. Suppose $\tilde{P}'_{ij}$ contains n'_{ij} items. Then check whether each patent in $\tilde{P}'_{ij}$ cites any patent cited by patents in $\tilde{P}_{ji}$,

$$\omega_{ij}^{22} = \sum_{k=1}^{n'_{ij}} \frac{\left| O\left(C_i(p_k), C_j\left(\tilde{P}_{ji} \right) \right) \right|}{|C_i(p_k)|}, \tag{3}$$

where $C_i(P) \equiv \{C_i(p_k)\}_{k=1}^{n'_{ij}}$, abusing notation. ω_{ji}^{22} is analogous.

The total citation overlap index is the ratio of the sum of the above overlaps to the total number of citations of both firms.

$$\omega_{ij} = \frac{\omega_{ij}^{1} + \eta(\omega_{ij}^{21} + \omega_{ji}^{21}) + \eta^2(\omega_{ij}^{22} + \omega_{ji}^{22})}{|P_i| + |P_j|}, \tag{4}$$

where $\eta \in (0, 1)$. We interpret η as the discount factor of technological relevance as generations go back. If a new technology is a child of citations, parent level relations are more significant than relations among grand parents. Note that $\omega_{ii} = 1$, $\omega_{ij} \in [0, 1]$ and, $\omega_{ij} = \omega_{ji}$.

Example Suppose that $P_i = \{1, 1, 2, 3, 4\}$ and $P_j = \{1, 3, 5\}$ as in Figure 1. The first-order overlap is the number of common patents, namely $\omega_{ij}^{1} = |\{1, 1, 3\}| + |\{1, 3\}| = 5$. Next, look at the patents in P_i that do not overlap P_j, $\tilde{P}_{ij} = \{2, 4\}$. We put some weights for the patents in $\tilde{P}_{ij}$ according to the relations with $\tilde{P}_{ji} = \{5\}$.

Suppose that patent 2 cites patent 5 and 7, patent 4 cites patents 3 and 6, and patent 5 cites patents 6 and 8. Since we eliminate patents overlapped on the first-order stage, $C_i(2) = \{5, 7\}$, $C_i(4) = \{6\}$, and $C_j(5) = \{6, 8\}$. Then,

$$\omega_{ij}^{21} = \frac{\left| O\left(C_i(2), \tilde{P}_{ji} \right) \right|}{|C_i(2)|} + \frac{\left| O\left(C_i(4), \tilde{P}_{ji} \right) \right|}{|C_i(4)|} = \frac{1}{2}.$$

Similarly, $\omega_{ji}^{21} = 0$. The second component of the second-order overlap is defined for patents with zero weight in calculation of ω^{21}. In the current example, $\tilde{P}'_{ij} = \{4\}$ and $\tilde{P}'_{ji} = \{5\}$. Look at whether patent 4's citations overlap citations of patent 5 (excluding overlapped patents at the first-order level). Here,

$$\omega_{ij}^{22} = \frac{|O\left(C_i(4), C_j(5)\right)|}{|C_i(4)|} = 1 \qquad \text{and} \qquad \omega_{ji}^{22} = \frac{|O\left(C_j(5), C_i(4)\right)|}{|C_j(5)|} = \frac{1}{2}$$

Finally, the total overlap index is defined as

$$\omega_{ab} = \frac{5 + \eta(\frac{1}{2} + 0) + \eta^2(1 + \frac{1}{2})}{3 + 5},$$

with some constant $\eta \in (0, 1)$.

Surely, we can define third- or higher-order overlaps but they require tedious computations. On the other hand, first-order overlaps do not give us much information about similarity of firms because there are not many direct citation overlaps, especially before the mid-1980s.[9] Hence, we use the first- and second-order overlaps.

The citation overlap index defined in (4) is an index of technological similarity. We transform the citation overlap index such that

$$d_{ij} = -\log(\omega_{ij}), \tag{5}$$

as long as $\omega_{ij} > 0$. d_{ij} is nonnegative, symmetric, $d_{ii} = 0$ but the triangle inequality does not hold. Thus we call it technological dissimilarity rather than distance. We will explain how to deal with pairs with $\omega_{ij} = 0$ in the next subsection.

Technological dissimilarities are defined in each subcategory and in each period (5-year window), τ. D_τ is the matrix of d_{ij}, where firm i and j applied for at least one patent (which is granted later) in the current subcategory during period τ. We omit the subscript indicating subcategories for notational simplicity.

We also calculate dynamic citation overlaps in each subcategory. A firm in period $\tau-1$ has a citation overlap index with firms in period τ (often including the same firm). We calculate the citation overlaps in the same way, derive $d_{ij}^{\tau-1,\tau}$ as the dissimilarity between firm i in period $\tau-1$ and firm j in period τ, and define $\hat{D}_{\tau-1,\tau}$ as the dynamic dissimilarity matrix.

The overall dissimilarity matrix, $\mathcal{D}_\tau$, for $\tau \geq 2$ is

$$\mathcal{D}_\tau \equiv \begin{bmatrix} D_{\tau-1} & 0 \\ \hat{D}_{\tau-1,\tau} & D_\tau \end{bmatrix}, \tag{6}$$

[9]The average number of citations in a patent dramatically increases during the 1980s. See Hall et al. (2001).

where

$$D_\tau = \begin{bmatrix} 0 & 0 & \dots & 0 \\ d^\tau_{2,1} & 0 & \dots & 0 \\ d^\tau_{3,1} & d^\tau_{3,2} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d^\tau_{n_\tau,1} & d^\tau_{n_\tau,2} & \dots & 0 \end{bmatrix}, \quad \hat{D}_{\tau-1,\tau} = \begin{bmatrix} d^{\tau-1,\tau}_{1,1} & d^{\tau-1,\tau}_{1,2} & \dots & d^{\tau-1,\tau}_{1,n_{\tau-1}} \\ d^{\tau-1,\tau}_{2,1} & d^{\tau-1,\tau}_{2,2} & \dots & d^{\tau-1,\tau}_{2,n_{\tau-1}} \\ \vdots & \vdots & \ddots & \vdots \\ d^{\tau-1,\tau}_{n_\tau,1} & d^{\tau-1,\tau}_{n_\tau,2} & \dots & d^{\tau-1,\tau}_{n_\tau,n_{\tau-1}} \end{bmatrix},$$

where n_τ is the number of firms. We define $\mathcal{D}_\tau$ as a lower triangular matrix because it has full information from symmetry.

2.2 Mapping Firm Locations: Multi-dimensional Scaling

Now we estimate the distribution of firms in technological spaces by using the dissimilarity matrix defined in the previous subsection. As in Stuart and Podolny (1996), we estimate firm locations by multi-dimensional scaling (MDS) with 2 dimensions.[10]

MDS estimates a distribution of firms such that the pairwise distances among firms are consistent with the original dissimilarities. More precisely, it estimates the distribution to minimize the *stress*, S, defined as

$$S = \left[\frac{\sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij} (\delta_{ij} - d_{ij})^2}{\sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij} d_{ij}^2} \right]^{\frac{1}{2}}, \tag{7}$$

where δ_{ij} is the Euclidean distance between estimated positions of firm i and j in a 2-dimensional space and w_{ij} is a weight.

We estimate the distribution of firms in a technological space by a dynamic procedure. First, we run MDS over the first 5-year window, $\tau = 1$ (1976-1980 except subcategory 33 (biotechnology), which starts with 1986-90 because only a few firms apply patents in this subcategory until the late 1980s.), with the dissimilarity matrix, D_1.[11] Let X_1 be the resultant distribution of firms. Next, to find the locations of firms

[10] When we apply one-dimensional MDS to our dissimilarity matrixes, we obtain the average stress of 0.55 whereas two-dimensional MDS returns 0.37 on average. This is a large gain of accuracy.

[11] Since MDS is sensitive to initial distributions, we repeated the MDS procedure 100 times with random initial distributions and selected the outcome with the smallest stress. The initial distri-

within the second 5-year window, we consider a dissimilarity matrix, $\mathcal{D}_2$ defined in (6) and run MDS under the constraint that the locations of firms within the previous 5-year window, X_1, are fixed. The initial distribution of the MDS procedure at this stage consists of X_1, which is predetermined, and a random distribution of firms in $\tau = 2$. Since the outcome contains both X_1 and X_2, we omit X_1 to get X_2. This process is repeated until the final 5-year window.

Since infinite dissimilarity (or $\omega_{ij} = 0$) cannot be processed by the MDS procedure, the standard code for MDS[12] ignores such information and allocates random distance without any restriction. In our procedure, we dropped firm i if $\omega_{ij} = 0$ for any j.[13] Even after we dropped all firms that do not have any technological relatives, it is not rare to have some ω_{ij} equals zero. For such pairs of firms, we impose the following constraint in our MDS procedure:

$$(w_{ij}, d_{ij}) = \begin{cases} (1, -\log \bar{\omega}_{ij}) & \text{if } \delta_{ij} < -\log \bar{\omega}_{ij}, \\ (0, \text{not defined}) & \text{otherwise,} \end{cases} \tag{8}$$

where

$$\bar{\omega}_{ij} \equiv \frac{2}{|C_i| + |C_j|}. \tag{9}$$

In words, the weight on d_{ij} is zero and relative locations of firm i and j are randomly determined as long as the resulting distance δ_{ij} is not shorter than the threshold level, $-\log \bar{\omega}_{ij}$, where $\bar{\omega}_{ij}$ is the first-order overlap as if they have just one direct citation overlap. But once δ_{ij} is closer than the threshold level, the weight is set at 1 and a positive value is added to the stress, (7), according to the gap from the threshold.[14]

Figure 2 is an example of firm distribution estimated by MDS for mutually ex-

butions are generated by a bivariate normal distribution with mean $(0, 0)$ and the same standard deviation vector as D_τ. We also used 100 random distributions for MDS in the later stages.

[12]Our code is based on mdscale.m contained in Matlab Statistics Toolbox.

[13]The ratio of firms dropped is about 5-6% on average. It varies across categories and decreases over time.

[14]One may consider that $d_{ij} = 1 - \omega_{ij}$ is a natural definition of the technological dissimilarity without constraints like (8). However, it is too restrictive in that a pair of firms without overlaps has a constant dissimilarity of 1. Since more than 70% of pairs of firms have $\omega = 0$ in our sample, attaching an arbitrary constant dissimilarity to those pairs results in firm distribution that almost ignores observed positive ω's. Instead, we assumed that patent citation overlaps only provide partial information about technological dissimilarity. This is partly because only granted patents are recorded in the dataset, not all technology is patented, and technological relationships are not observed up to second-order overlaps. The constraint (8) introduces varying thresholds and randomness to take into account unobserved technological relations.

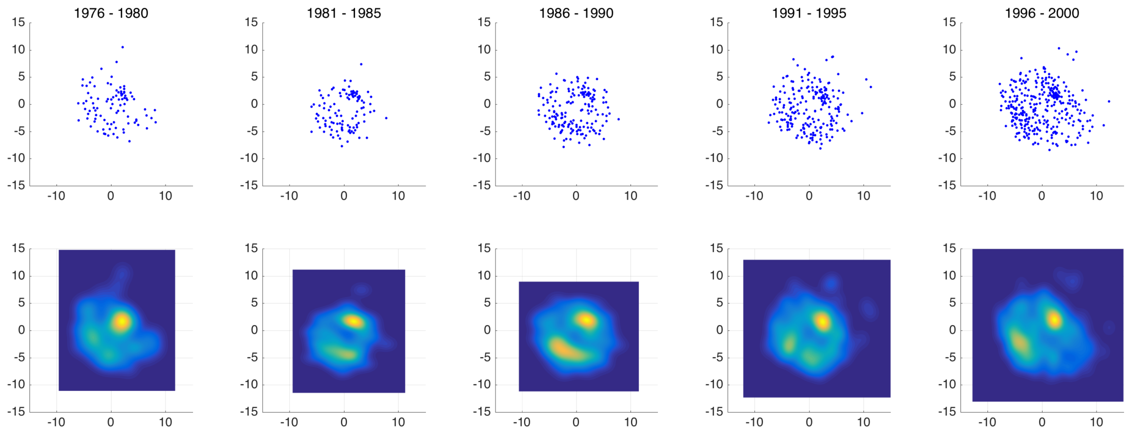

Figure 2: Example of post-MDS firm distribution. Technological subcategory 46 (Semiconductor devices). The top row is the MDS result and the second row is the estimated density by kernel density estimation, viewed from above.

clusive 5-year windows. The five panels in the top row are firm distributions in each 5-year window. The bottom row draws estimated densities of those firm distributions by using kernel density estimation, viewed from above (lighter color indicates greater density). We estimated these distributions for all 2-digit subcategories and all 5-year windows in the sample.

MDS estimates distances among entities and generates a map satisfying these distances. The stress defined in (7) is neutral for rotation and inversion of the whole map. Since we consider the dynamic dissimilarity matrix to bridge different 5-year windows, the orientations are anchored by distributions in the previous 5-year windows. However, notice that the axes in Figure 2 do not have any meaning. Firms are just distributed with the estimated relative positioning.[15]

2.3 Average Dissimilarity and Distances after MDS

Figure 3 shows the average dissimilarity and the average post-MDS distance among subcategories. The weighted dissimilarity/distance are a weighted average of those with weights of a number of firms in each category. The figure tells us that the average technological distance on technology fields has been getting larger over time.

This fact does not specify changes in the distribution of firm location on technological maps. Figure 4 draws two examples of distributional changes when the

[15]The figures for other subcategories are available upon request.

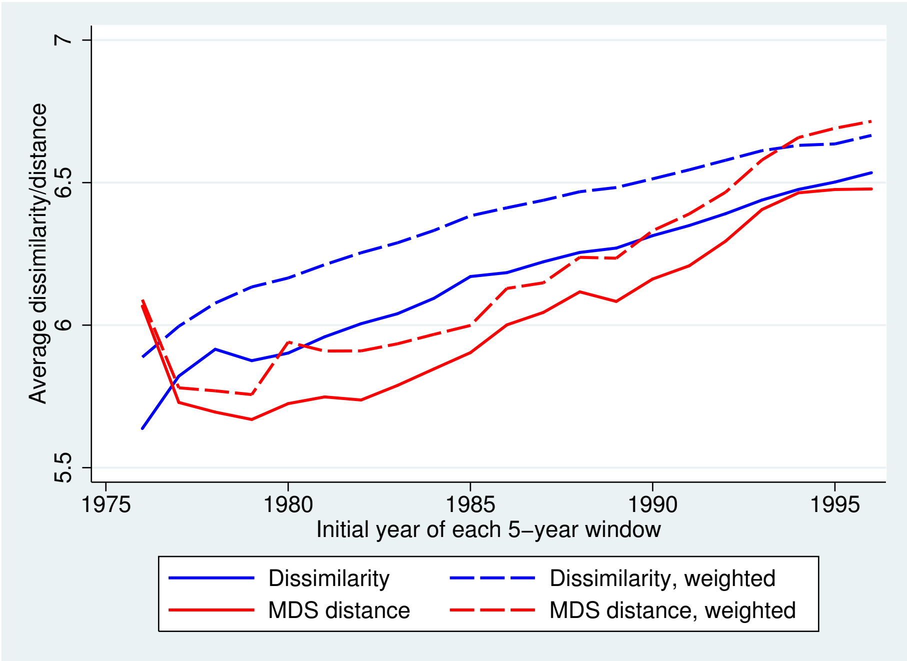

Figure 3: Average dissimilarity/distance.

average distance increases. The distribution may simply become more fragmented with a higher standard deviation. Alternatively, the original distribution is split into two humps, that is, there are two polars and technological groups emerge around those polars.

3 Polarization

3.1 A Simple Model for Inter-group Competition

Esteban and Ray (1994) presents the fundamental idea of polarization. Their definition of the measure of polarization on one-dimensional distribution is

$$P^{\alpha} = K \sum_{i}^{m} \sum_{j}^{m} n_i^{1+\alpha} n_j \delta_{ij}, \qquad K > 0,\ \alpha \in [0.25, 1], \tag{10}$$

where $i, j = 1, 2, ..., m$ are groups, n_i is the share of group i, and δ_{ij} is the distance between groups. To capture inter-group competition, both homogeneity within a group and heterogeneity across groups should be accentuated because a conflict tends

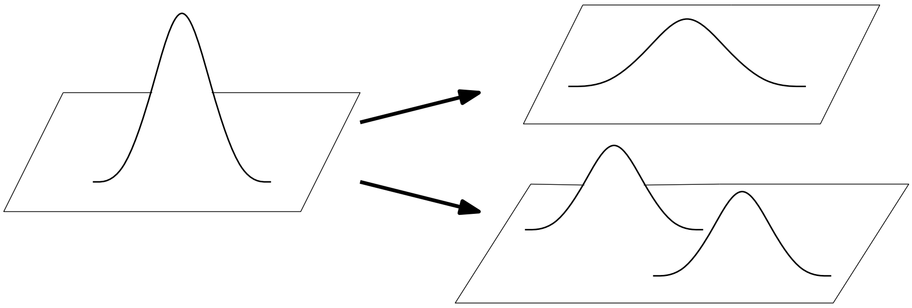

Figure 4: Examples of changes in distribution, given an increase in the average distance.

to be harsh when there are two large distant groups. Polarization defined in (10) satisfies these requirements, whereas inequality measures such as the Gini coefficient and the HHI accentuate only one of those aspects. Esteban and Ray (2011) construct a model of group contests and describe how the overall efforts depends on the degree of inter-group competition and show that the combination of Gini, HHI, and polarization explain severity of conflicts well.

We apply this polarization measure to the distribution of firms by interpreting group contests as R&D races for becoming a dominant technology. Suppose that each technological category corresponds to an industry. Intra-group homogeneity matters because the probability of winning a race is higher and, moreover, more knowledge spillovers are likely in the future if more applications are created by firms with the same fundamental technology. At the same time, inter-group heterogeneity matters because firms in one technological group have more incentive to make R&D efforts to win the race when they have rivals which are technologically distant because losing firms will pay greater cost to catch up the winners' technology to survive.

Suppose there are $m \geq 1$ technology groups that compete a race for standard in a technology field. The number of firms in group i is N_i. Denote $N = \sum_{i=1}^{m} N_i$ and $n_i = \frac{N_i}{N}$. Let r_{ih} be the R&D effort by firm h in group i and $\frac{1}{2}r_{ih}^2$ is its cost. Probability of winning the race for group i is

$$p_i = \frac{R_i}{R}, \tag{11}$$

where

$$R_i \equiv \left[\sum_{h=1}^{N_i} r_{ih}^{\frac{1}{\epsilon}}\right]^{\epsilon} - \psi \sum_{h=1}^{N_i} r_{ih}, \quad \epsilon \geq 1\, \psi \in (0,1), \tag{12}$$

$$R \equiv \sum_{i=1}^{m} R_i.$$

R_i is the group-level aggregation of R&D. We consider a complementarity among R&D activities within groups, which is represented by $\epsilon \geq 1$. However, the complementarity effect is weaken by duplication of research. Thus, some portion of research efforts do not contribute to the aggregate R&D. Parameter ψ stands for the degree of duplication.

The expected payoff function for firm h in group i has three component. The first component is the profit when the fundamental technology of group i becomes the standard in the industry. We assume that the winning group grabs the whole demands in the market, so that firms in losing groups earn no profit. We assume the profit of firms in the winning group is $\frac{\bar{\pi}}{n_i}$. The second component comes from catch-up cost when a rival group wins. If group j wins, firms in group i switch their own fundamental technology to the winning technology to survive. The catch-up cost depends on how different their technologies are. Let δ_{ij} be the technological distance between the two group and $S(\delta_{ij})$ be the catch-up cost. S is strictly increasing and $S(0) = 0$. The third component is the cost of R&D. In sum,

$$\pi_{ih}(r_{ih}) = p_i \frac{\bar{\pi}}{n_i} - \sum_{j=1}^{m} p_j S(\delta_{ij}) - \frac{1}{2} r_{ih}^2, \tag{13}$$

$$= \frac{\bar{\pi}}{n_i} - \sum_{j \neq i} p_j \left[\frac{\bar{\pi}}{n_i} + S(\delta_{ij})\right] - \frac{1}{2} r_{ih}^2. \tag{14}$$

Define

$$\Delta_{ij} = \begin{cases} 0, & \text{for } j = i, \\ \frac{\bar{\pi}}{n_i} + S(\delta_{ij}), & \text{for } j \neq i. \end{cases} \tag{15}$$

Then, we write the maximization problem for firm h in group i as

$$\max - \sum_{j=1}^{m} p_j \Delta_{ij} - \frac{1}{2} r_{ih}^2. \tag{16}$$

At any interior solution, we have

$$\frac{1}{R}\left(-\psi + r_{ih}^{\frac{1}{\epsilon}-1}\left(\sum_{l=1}^{N} {}_i r_{il}^{\frac{1}{\epsilon}}\right)^{\epsilon-1}\right)\sum_{j=1}^{m} p_j \Delta_{ij} = r_{ih}. \tag{17}$$

The optimal choice of r_{ih} is unique and does not depend on h under the current assumptions, the equilibrium is symmetric in each group. Thus, we write $r_i = r_{ih}$. Then condition (17) becomes

$$\frac{\sigma_i}{R}\sum_{j=1}^{m} p_j \Delta_{ij} = r_i \tag{18}$$

where

$$\sigma_i \equiv -\psi + N_i^{\epsilon-1} \tag{19}$$

in equilibrium. σ_i is the marginal contribution of individual R&D in a symmetric equilibrium.

Connection to Polarization Let $\rho \equiv \frac{R}{N}$ and $\mu_i \equiv \frac{p_i}{n_i}$. ρ is the per-firm aggregate R&D output. μ_i is the ratio of intra-group per-firm R&D outputs to the aggregate per-firm R&D outputs since

$$\mu_i = \frac{R_i/R}{N_i/N} = \frac{R_i/N_i}{R/N} = \frac{\sigma_i r_i}{\rho}. \tag{20}$$

Multiply both sides of equation (18) by ρp_i,

$$\begin{aligned}
\rho p_i \frac{\sigma_i}{R}\sum_{j=1}^{m} p_j \Delta_{ij} &= \rho p_i r_i \\
\Leftrightarrow \quad \sum_{j=1}^{m} p_i p_j \frac{\sigma_i \Delta_{ij}}{N} &= \rho p_i r_i \\
\Leftrightarrow \quad \sum_{j=1}^{m} \mu_i \mu_j n_i n_j \frac{\sigma_i \Delta_{ij}}{N} &= \rho p_i r_i
\end{aligned} \tag{21}$$

Multiply bothe sides of (21) by $\frac{\rho}{r_i}$ and define

$$\phi(\mu_i, \mu_j, r_i, \rho) \equiv \frac{\mu_i \mu_j \rho}{r_i}, \tag{22}$$

we have

$$\sum_{j=1}^{m} \phi(\mu_i, \mu_j, r_i, \rho) n_i n_j \frac{\sigma_i \Delta_{ij}}{N} = \rho^2 p_i. \tag{23}$$

Take the summation over i,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \phi(\mu_i, \mu_j, r_i, \rho) n_i n_j \frac{\sigma_i \Delta_{ij}}{N} = \rho^2. \tag{24}$$

Now consider the situation in which $\phi(\mu_i, \mu_j, r_i, \rho) = \bar{\sigma} \equiv \psi + \bar{N}^{\epsilon-1}$, where $\bar{N} = \frac{N}{m}$. This condition holds when everything is symmetric among groups such that $N_i = N_j$ and $r_i = r_j$ for all i and j.[16] Let $\hat{\rho}$ satisfy

$$\hat{\rho}^2 \equiv \sum_{i=1}^{m} \sum_{j=1}^{m} n_i n_j \frac{\bar{\sigma} \sigma_i \Delta_{ij}}{N}. \tag{25}$$

We consider $\hat{\rho}$ as a proxy for per-capita aggregate R&D. Developing the right-hand side,

$$\begin{aligned} &\sum_{i=1}^{m} \sum_{j=1}^{m} n_i n_j \frac{\bar{\sigma} \sigma_i \Delta_{ij}}{N} \\ &\quad = -\frac{\bar{\sigma}\psi}{N} \sum_{i=1}^{m} \sum_{j=1}^{m} n_i n_j \Delta_{ij} + \frac{\bar{\sigma}}{N} \sum_{i=1}^{m} \sum_{j=1}^{m} n_i n_j N_i^{\epsilon-1} \Delta_{ij} \end{aligned} \tag{26}$$

The first term of (26) is

$$\begin{aligned} -\frac{\bar{\sigma}\psi}{N} \sum_{i=1}^{m} \sum_{j \neq i} n_i n_j \left(\frac{\bar{\pi}}{n_i} + S(\delta_{ij}) \right) &= -\frac{\bar{\sigma}\psi}{N} \left[\bar{\pi} \sum_{i=1}^{m} (1 - n_i) + \sum_{i=1}^{m} \sum_{j \neq i} n_i n_j S(\delta_{ij}) \right] \\ &= -\frac{\bar{\sigma}\psi}{N} \left[(m-1)\bar{\pi} + \sum_{i=1}^{m} \sum_{j \neq i} n_i n_j S(\delta_{ij}) \right] \end{aligned}$$

[16] From equation (20), $\phi(\mu_i, \mu_j, r_i, \rho) = \mu_j \sigma_i$. When all variables are symmetric, $\mu_i = 1$ for any i and $\phi(\mu_i, \mu_j, r_i, \rho) = \bar{\sigma}$, which equals $\frac{\rho}{\bar{r}}$, or the ratio fo per-capita aggregate R&D to individual R&D.

The second term of (26) is

$$\frac{\bar{\sigma}}{N^{2-\epsilon}}\sum_{i=1}^{m}\sum_{j\neq i} n_i^{\epsilon} n_j \left(\frac{\bar{\pi}}{n_i} + S(\delta_{ij})\right) = \frac{\bar{\sigma}}{N^{2-\epsilon}}\sum_{i=1}^{m}\left[\bar{\pi} n_i^{\epsilon-1}\sum_{j\neq i} n_j + \sum_{j\neq i} n_i^{\epsilon} n_j S(\delta_{ij})\right]$$
$$= \frac{\bar{\sigma}}{N^{2-\epsilon}}\left[\bar{\pi}\left(\sum_{i=1}^{m} n_i^{\epsilon-1} - \sum_{i=1}^{m} n_i^{\epsilon}\right) + \sum_{i=1}^{m}\sum_{j=1}^{m} n_i^{\epsilon} n_j S(\delta_{ij})\right]$$

Summing up those terms and assume $S(\delta) = a\delta$ $(a > 0)$ for simplicity,

$$\hat{\rho}^2 = -\frac{\bar{\sigma}\psi(m-1)\bar{\pi}}{N} - \frac{a\bar{\sigma}\psi}{N}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j \delta_{ij}}_{\text{average distance}}$$
$$+ \frac{\bar{\sigma}\bar{\pi}}{N^{2-\epsilon}}\underbrace{\left(\sum_{i=1}^{m} n_i^{\epsilon-1} - \sum_{i=1}^{m} n_i^{\epsilon}\right)}_{\text{fragmentation}} + \frac{a\bar{\sigma}}{N^{2-\epsilon}}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} n_i^{\epsilon} n_j \delta_{ij}}_{\text{polarization}} \tag{27}$$

The average individual R&D is related to three distributional statistics: the average distance (the summation in the second term); fragmentation or negative of concentration, the parenthesis in the third term, which is equivalent to 1 minus HHI when $\epsilon = 2$; and the polarization (when ϵ is in the appropriate region), the summation in the forth term.

In the current model, technological distance stimulate individual R&D because a losing firm must pay higher cost to catch up the new mainstream technology if the winning group is further away in the technology space. This aspect is captured by polarization and thus the coefficient is positive. But at the same time, more efforts imply more duplications in research within groups. Hence, the R&D incentive stimulated by distance is weaken by degree of duplication, which is represented in the negative sign on the average distance. The degree of fragmentation has a positive coefficient because each group has significant probability to win.

Keeping the current model in mind, we move to continuous technology space and introduce the extended version for continuous distributions which is developed by DER in the next subsection.

3.2 Polarization Measure on Two-dimensional Spaces

DER extend the measure of polarization in (10) to be applicable for continuous distributions. Our polarization measure follows DER,

$$P^{\alpha}(f) \equiv \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(x)^{1+\alpha} f(y) \delta(x, y) dy dx, \tag{28}$$

where f is the density of firms, $\delta(x, y)$ is the Euclidean distance, and α is a positive parameter in between $[0.2, 0.5]$. Only the difference from DER is that our polarization is defined over distributions with 2-dimensional domains (one-dimension in DER), which makes the valid range of α narrow. We can easily show that the upper bound of α is the inverse of the number of dimension (proof is in Appendix A). The lower bound is complicated. We describe how to get the lower bound of valid α also in Appendix A. In the regressions in the following section, we report the results for both bounds of α, $\{0.2, 0.5\}$.[17] Below, We report the estimates with both bounds of α.

The average distance, G, and concentration, H, of density f are defined as follows.

$$G(f) \equiv \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(x) f(y) \delta(x, y) dy dx, \tag{29}$$

$$H(f) \equiv \int_{\mathbb{R}^2} f(x)^2 dx, \tag{30}$$

Note that G is equivalent to polarization with $\alpha = 0$.

We estimate $f_{k\tau}$, the density of firms in category k and period τ, by the 2-dimensional kernel density estimation (2D-KDE).[18] Let $\hat{f}_{k\tau}$ be the estimated distribution. The estimates of (28)-(30) over firm locations on technological fields obtained

[17] When $\alpha > 0.2$, squeezing both humps of a distribution with two humps (like the bottom part in the right side of Figure 4) in a symmetric way increases polarization. When $\alpha < 0.5$, squeezing the whole distribution reduces polarization.

[18] For 2-dimensional kernel density estimation, we used the code described in Botev et al. (2010).

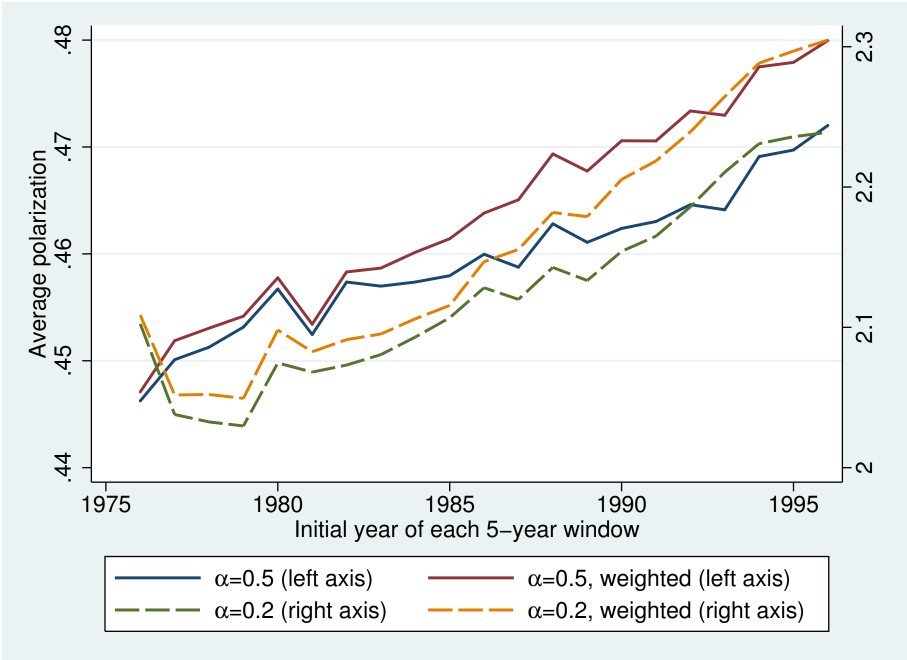

Figure 5: Average polarization indices.

by the procedure in Section 2, $X_{k\tau} = \{x_1, x_2, ..., x_{n_{k\tau}}\}$, are

$$\hat{P}^{\alpha}_{k\tau} \equiv \frac{1}{n^2_{k\tau}} \sum_{i=1}^{n_{k\tau}} \sum_{j=1}^{n_{k\tau}} \hat{f}_{k\tau}(x_i)^{\alpha} \delta(x_i, x_j), \tag{31}$$

$$\hat{H}_{k\tau} \equiv \frac{1}{n_{k\tau}} \sum_{i=1}^{n_{k\tau}} \hat{f}_{k\tau}(x_i). \tag{32}$$

$\hat{G}_{k\tau}$ is the special case with $\alpha = 0$ in (31), thus

$$\hat{G}_{k\tau} \equiv \frac{1}{n^2_{k\tau}} \sum_{i=1}^{n_{k\tau}} \sum_{j=1}^{n_{k\tau}} \delta(x_i, x_j) \tag{33}$$

Figure 5 represents the estimated polarizations averaged over subcategories. The weighted average is computed by setting the share of the number of firms within each period (5-year window) as weights for the subcategories. Clearly, the average polarizations have upward trends regardless of the values of α.

Table 1 shows the descriptive statistics that are used in regressions in the next section.

Table 1: Summary statistics.

	Obs.	Mean	SD	Min	Max
Average dissimilarity	649	6.160758	.6917609	2.632249	7.368289
$\hat{G}$	647	6.032016	.6695385	2.353421	7.570821
$\hat{P}^{0.2}$	639	2.12773	.1579996	1.472612	2.530501
$\hat{P}^{0.5}$	639	.4595821	.0249231	.3863933	.5175242
$\hat{H}$	639	.0070822	.0014643	.0042083	.014769
Num. of firms	651	246.4101	141.7449	0	1019
Stress	647	.3729808	.0617119	5.09e-09	.4728863
Num.patent app.	648	3812.71	4782.57	1	40835
CW.patent app.	648	43875.55	56416.79	2	335869

4 The Impact of Polarization on Innovation

4.1 Basic Results

In this section, we investigate the empirical relationship between polarization and innovation. For the measure of innovation, we use citation-weighted number of patent applications, a_{kt}, where k is subcategory and t is year of application. Note that we denote t as a year and $\tau(t)$ as the 5-year window from $t-5$ to $t-1$. Below, we estimate the impact of the distribution properties during $\tau(t)$ on innovations in t.

It is important to note that the citation-weighted patent applications after the late 1990s are less informative in our sample and, thus, we drop the citation-weighted patent application in and after 1998 from our estimation. Figure 6 illustrates the citation-weighted patent applications over the sample periods. The citation-weighted patent applications hit a peak around 1995 and declined sharply after 1997, whereas the unweighted patent applications continue to increase. This is simply because of the time lags between application and citation. Since the NBER US Patent dataset contains only granted patents, citation-weights are highly affected by this time-lag problem.[19]

Since a_{kt} is count data, we apply the Poisson regression model and the negative

[19]Hall et al. (2001) introduced weights for dealing with this problem but the current problem is not resolved because zero citation multiplied by any weight is zero.

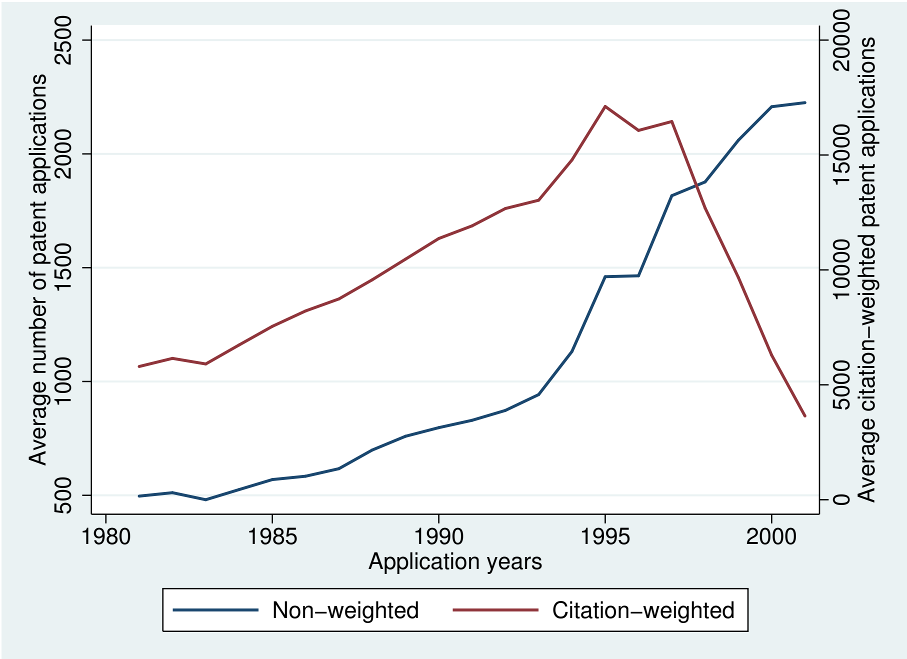

Figure 6: Non- and citation-weighted number of patent applications.

binomial regression model. The equation to be estimated with these models is

$$a_{kt} = \exp\left\{\beta_0^a + \beta_1^a \log \hat{P}^{\alpha}_{k,\tau(t)} + \beta_2^a \log \hat{H}_{k,\tau(t)} + \beta_3^a \log \hat{G}_{k,\tau(t)} \right.$$
$$\left. + \beta_4^a \log A_{k,\tau(t)} + \beta_5^a \log Y_{k,\tau(t)} + \beta_6^a \log L_{k,\tau(t)} + \beta_7^a \nu_k + \beta_8^a \nu_t + \epsilon^a_{kt} \right\}, \quad (34)$$

where $A_{k,\tau(t)}$ is the citation-weighted stock of patents at the beginning of $\tau(t)$ as the proxy of knowledge stock in the subcategory (described in detail in Appendix B), $Y_{k,\tau(t)}$ is the average of total sales of all related firms during $\tau(t)$, $L_{k,\tau(t)}$ is the average of total employment of those firms, and ν_k and ν_t are dummy variables for subcategories and years. This regression evaluates the impact of distribution properties during the past 5 years on the amount of new innovations.

We also consider the impact of distribution properties on growth rate of citation-weighted applications over 5-year windows, $\gamma_{kt} \equiv \frac{a_{kt}}{a_{kt-5}} - 1$. The equation to be estimated by panel regression with subcategory fixed effects is

$$\gamma_{kt} = \beta_0^{\gamma} + \beta_1^{\gamma} \log \hat{P}^{\alpha}_{k,\tau(t)} + \beta_2^{\gamma} \log \hat{H}_{k,\tau(t)} + \beta_3^{\gamma} \log \hat{G}_{k,\tau(t)}$$
$$+ \beta_4^{\gamma} \log A_{k,\tau(t)} + \beta_5^{\gamma} \log Y_{k,\tau(t)} + \beta_6^{\gamma} \log L_{k,\tau(t)} + \beta_7^{\gamma} \nu_k + \beta_8^{\gamma} \nu_t + \epsilon^{\gamma}_{kt}. \quad (35)$$

Table 2 shows the regression results. Columns (1-6) omitted explanatory variables about knowledge stock and business size (all regressions include subcategory and year dummies). Because citation-weighted patents are count data and highly skewed, the negative binomial regression is appropriate. Columns (3) and (4) report significant positive coefficient on polarization with both the upper and lower boundaries of α, and significantly negative coefficients for concentration and the average distance. These signs are consistent with the model in Section 3.1. The change rate of citation-weighted patent applications, γ_{kt}, also have similar results (Columns (5) and (6), linear regressions). However, this result is not robust. Looking at Columns (1) and (2), which conduct Poisson regressions, we find the opposite signs of coefficients for polarization. Since the estimates of Poisson regression are consistent regardless of the distributional assumption, we need to change the model specification.

Columns (7-12) are results controlled by knowledge stock and business sizes. The knowledge stock has a positive impact on levels of innovation a_{kt} and negative but insignificant impact on growth of innovation. These are natural results in the knowledge accumulation process. The sales volume always has a positive impact because it represents size of demands for subcategories. The coefficients of employment are negative most probably because the combination of sales and employment represents average productivity. The inconsistency between Poisson and negative binomial regressions seen before is now resolved. However, the significance levels in negative binomial regressions (Columns (9) and (10)) becomes low and the sign of coefficients are inconsistent with the model. Only the regressions (11) and (12) about γ_{kt} weakly keep the consistency with the model and previous simple model specification.

So far, our hypothesis of inter-group competition does not seem to work well. And the characteristics of firm distributions in technology spaces are not related to aggregate innovations. But this result drastically changes if we split the sample in time series. We consider a structural shift in the next subsection.

Table 2: Citation-weighted patent application vs polarization measures.

	(1)Poisson	(2)Poisson	(3)NegBin	(4)NegBin	(5)	(6)	(7)Poisson	(8)Poisson	(9)NegBin	(10)NegBin	(11)	(12)
	a_{kt}	a_{kt}	a_{kt}	a_{kt}	γ_{kt}	γ_{kt}	a_{kt}	a_{kt}	a_{kt}	a_{kt}	γ_{kt}	γ_{kt}
$\log \hat{P}^{0.2}_{k\tau(t)}$	-2.726***		4.843***		6.148***		-3.492***		-2.222		5.248***	
	(-38.78)		(2.82)		(3.27)		(-48.89)		(-1.58)		(2.64)	
$\log \hat{P}^{0.5}_{k\tau(t)}$		-2.111***		2.049**		2.340**		-2.154***		-1.154		1.718
		(-57.41)		(2.18)		(2.26)		(-57.87)		(-1.53)		(1.60)
$\log \hat{H}_{k\tau(t)}$	-0.370***	0.0230*	-0.488**	-0.665**	-0.480*	-0.633*	0.173***	0.506***	0.0395	0.182	-0.434*	-0.500
	(-35.79)	(1.66)	(-2.07)	(-2.05)	(-1.86)	(-1.78)	(16.40)	(36.15)	(0.22)	(0.73)	(-1.66)	(-1.39)
$\log \hat{G}_{k\tau(t)}$	1.199***	0.413***	-4.109***	-1.519**	-6.291***	-2.834***	1.927***	0.523***	0.873	-0.166	-5.633***	-2.550***
	(19.23)	(12.58)	(-2.80)	(-1.97)	(-3.92)	(-3.35)	(30.62)	(15.83)	(0.75)	(-0.28)	(-3.40)	(-3.00)
$\log A_{k\tau(t)}$							0.851***	0.846***	0.636***	0.638***	-0.0668	-0.0745
							(203.35)	(202.16)	(8.47)	(8.49)	(-0.62)	(-0.69)
$\log Y_{k\tau(t)}$							0.791***	0.794***	0.681***	0.677***	0.426***	0.460***
							(111.48)	(112.09)	(6.73)	(6.71)	(2.94)	(3.17)
$\log L_{k\tau(t)}$							-0.517***	-0.520***	-0.404***	-0.405***	-0.476***	-0.487***
							(-62.07)	(-62.38)	(-3.22)	(-3.22)	(-2.68)	(-2.73)
Const.	6.507***	6.163***	9.689***	9.399***	4.596***	4.065***	-7.538***	-7.644***	-4.448***	-4.394***	3.379***	2.454*
	(191.02)	(183.85)	(12.20)	(11.56)	(5.34)	(4.62)	(-136.47)	(-140.45)	(-4.47)	(-4.45)	(2.59)	(1.90)
Overdispersion			-2.686***	-2.680***					-3.221***	-3.221***		
			(-43.40)	(-43.31)					(-51.47)	(-51.46)		
N	516	516	516	516	516	516	516	516	516	516	516	516
adj. R^2					0.143	0.133					0.154	0.146
pseudo R^2	0.947	0.947	0.146	0.146			0.978	0.978	0.172	0.172		

t-statistics in parentheses. All regressions consider fixed effects of technological categories and year dummies.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 A Structural Change through 1980s

In response to "productivity slow down" from 1970s, industries in the United States had experienced major institutional changes. The biggest issue related to the current context is the patent reform. The patent policy in the United States had dramatically shifted from anti-patent to pro-patent through the early 1980s. Patents had become much more valuable than before and patenting a technology or idea has become one of the most important strategies for firms. At the same time, Jaffe and Lerner (2004) point out that the reform significantly reduced the quality of patent examination because of the flood of patent applications.[20]

This structural shift might affect the relationship between polarization and innovation. Our interpretation of polarization as the degree of inter-group competition may not hold if a bunch of useless patents clustering around some main technologies. More importantly, under the pro-patent system, inter-group competition may stimulate patent litigation rather than R&D investment, which discourages innovation (Lanjouw and Schankerman (2004)).

Thus, to see whether any structural shift exists, we conducted the Chow test on the above negative binomial regression. Figure 7 illustrate the log-likelihood ratio test statistics for each cut-off year. We can see there is a highly significant structural change between former periods and later periods and the peak of significance is 1990. Since the estimations tell the impact of polarization in the preceding 5 years on the patent applications in the current year, the threshold of 1990 implies the polarization of distribution within 1985-1989, around that time the patent reform prevailed.

Table 3 reports the regression results using equations (34) and (35) for samples divided into periods before 1990 and after. We can see a clear structural change between the results. For $t \leq 1990$ (Columns (1)-(4)), the estimates of polarization is significantly positive. Further, the coefficients of HHI and the average distance have signs that are consistent with the model in the previous section and they are mainly significant. To the contrary, the results for $t > 1990$ (Columns (5)-(8)) are totally different. In regressions (5) and (6), all coefficients are highly significant but the signs of the coefficients for distributional characteristics are reversed. The story of inter-group competition cannot be applied.

[20]Kortum and Lerner (1998) explain the surge in patent application during 1980s by the change in R&D management rather than the patent reform. Hall and Ziedonis (2001) attribute the change to patent management.

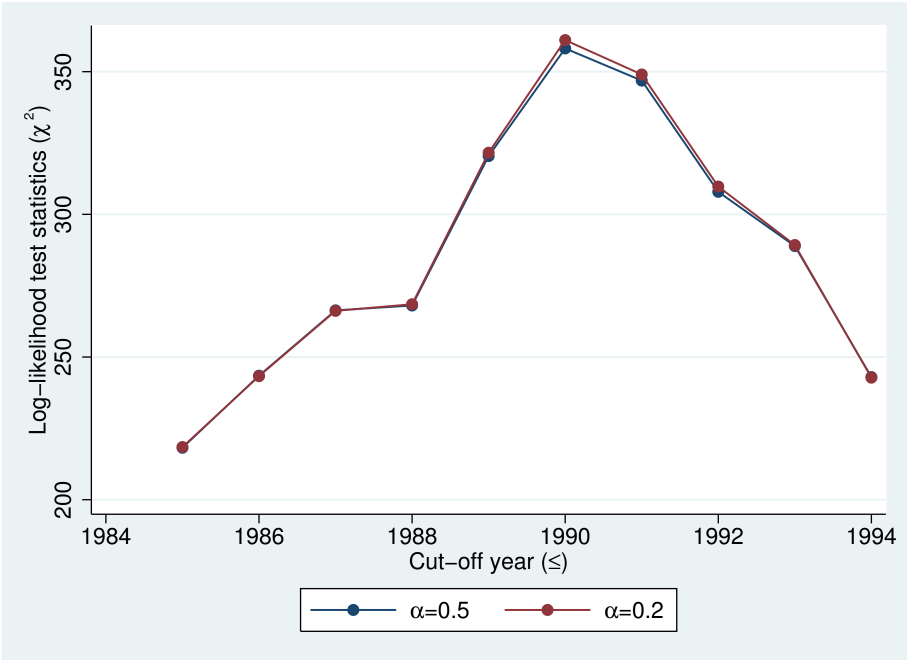

Figure 7: Log-likelihood ratio statistics. All cutoff years reject the null hypothesis that former and latter periods separated the cut-offs are nested in the full sample regression.

Figure 8 illustrates the relationship between polarization (for $\alpha = 0.5$) and the number of patent applications after controlled by the other variables for samples before and in 1990, after 1990, and for the full sample.

The impacts of polarization are not small quantitatively. The non-weighted average polarizations illustrated in Figure 5 change by 3.8% and 2.6% in the former period for $\alpha = 0.2$ and 0.5, respectively, and by 5.2% and 2.6% in the later period for $\alpha = 0.2$ and 0.5, respectively. Hence, the occurrence of innovations is increased by 4.1%-11.6% through the surge in polarization in the former period and it is decreased by 9.1%-35.9% through the surge in polarization in the later period.

One reason of the drastic change of the regression results can be seen by the following regression about quality of patents. Table 4 reports the regression result where we take the average quality of the patent (defined as the per-patent citations) as the dependent variable in regression equation (35). All regressions include both subcategory and year dummies. As seen in regressions (3)-(4) in Table 4, polarization reduces patent quality on average. Moreover, the estimated coefficients for distributional characteristics are quite similar to those in regressions (5)-(6) in Table 3. It

Table 3: Regressions by year groups.

	$t \leq 1990$				$t > 1990$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	a_{kt}	a_{kt}	γ_{kt}	γ_{kt}	a_{kt}	a_{kt}	γ_{kt}	γ_{kt}
$\log \hat{P}^{0.2}_{k\tau(t)}$	3.053**		7.873***		-6.885***		4.864	
	(2.25)		(3.05)		(-3.29)		(1.36)	
$\log \hat{P}^{0.5}_{k\tau(t)}$		1.564**		4.050***		-3.479***		1.865
		(2.16)		(2.94)		(-2.84)		(0.89)
$\log \hat{H}_{k\tau(t)}$	-0.283	-0.472*	-0.425	-0.918*	0.922***	1.332***	-0.169	-0.299
	(-1.54)	(-1.88)	(-1.21)	(-1.91)	(3.22)	(3.24)	(-0.35)	(-0.43)
$\log \hat{G}_{k\tau(t)}$	-2.816**	-1.355**	-6.065***	-2.310**	6.349***	3.121***	-5.926**	-3.275*
	(-2.44)	(-2.32)	(-2.76)	(-2.07)	(3.62)	(3.16)	(-1.98)	(-1.96)
$\log A_{k\tau(t)}$	0.334***	0.335***	-0.236	-0.232	0.438***	0.437***	0.140	0.134
	(3.61)	(3.63)	(-1.33)	(-1.31)	(3.16)	(3.13)	(0.59)	(0.57)
$\log Y_{k\tau(t)}$	1.145***	1.153***	1.001***	1.021***	1.189***	1.179***	1.147***	1.177***
	(8.07)	(8.15)	(3.73)	(3.81)	(4.77)	(4.69)	(2.74)	(2.79)
$\log L_{k\tau(t)}$	-0.791***	-0.797***	-1.190***	-1.205***	-1.052***	-1.085***	-0.932**	-0.924**
	(-4.82)	(-4.87)	(-3.85)	(-3.90)	(-4.09)	(-4.17)	(-2.17)	(-2.13)
Const.	-2.785**	-2.901**	2.503	2.202	-6.188***	-5.756***	-3.842	-4.659
	(-2.12)	(-2.23)	(0.99)	(0.88)	(-3.57)	(-3.30)	(-1.28)	(-1.56)
Overdispersion	-4.027***	-4.026***			-3.790***	-3.778***		
	(-48.09)	(-48.08)			(-38.72)	(-38.58)		
N	300	300	300	300	216	216	216	216
adj. R^2			0.240	0.238			0.098	0.093
pseudo R^2	0.198	0.198			0.207	0.206		

t-statistics in parentheses. All regressions consider fixed effects of technological categories and year dummies.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

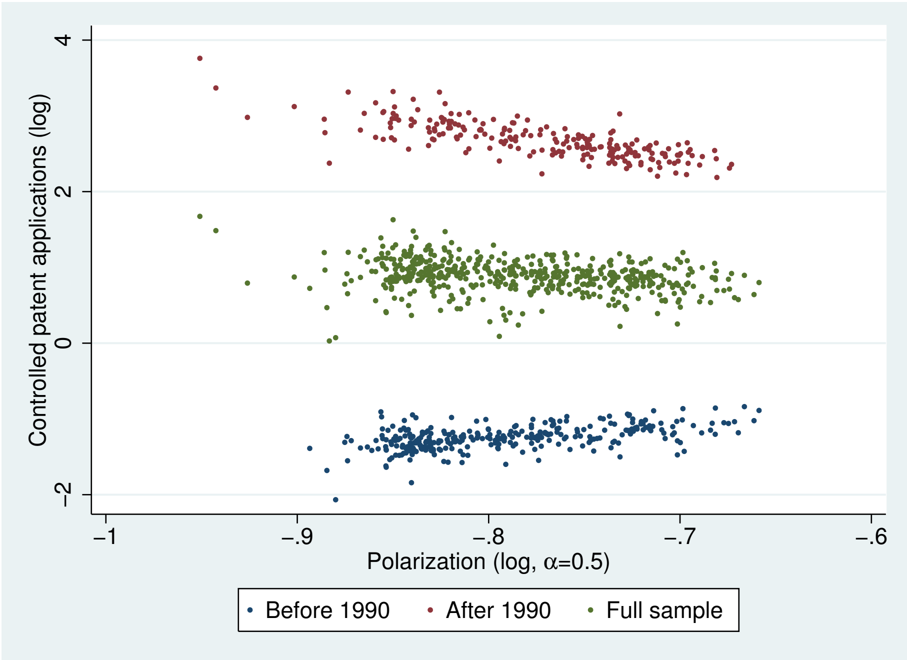

Figure 8: The relationship between polarization and the number of patent applications after controlled by other variables. $\alpha = 0.5$.

suggests that inter-group competition negatively affects patent quality after patent reform and thus the relation between innovation and inter-group competition had been reversed.

The structural shift in the estimations suggests that inter-group competition affect R&D incentives in different ways before and after the pro-patent policy shift. As the patent reform prevails, we have observed a rapid increase in patent cases and patenting has become much more strategic than before. Firms devote more resources into protecting own patents and attacking others', rather than conducting new R&D projects, to win R&D races. Hence, harsh inter-group competition decreases new innovations.

Table 4: Regression on average quality of patents.

	$t \leq 1990$		$t > 1990$	
	(1)	(2)	(3)	(4)
	$\ln q_{kt}$	$\ln q_{kt}$	$\ln q_{kt}$	$\ln q_{kt}$
$\log \hat{P}^{0.2}_{k\tau(t)}$	0.632		-6.498***	
	(0.55)		(-4.43)	
$\log \hat{P}^{0.5}_{k\tau(t)}$		0.414		-3.273***
		(0.67)		(-3.79)
$\log \hat{H}_{k\tau(t)}$	-0.154	-0.218	0.768***	1.156***
	(-0.98)	(-1.02)	(3.96)	(4.07)
$\log \hat{G}_{k\tau(t)}$	-0.742	-0.498	5.649***	2.592***
	(-0.75)	(-1.00)	(4.58)	(3.74)
$\log A_{k\tau(t)}$	-0.0255	-0.0251	-0.368***	-0.360***
	(-0.32)	(-0.32)	(-3.41)	(-3.29)
$\log Y_{k\tau(t)}$	0.144	0.142	0.589***	0.570***
	(1.20)	(1.19)	(3.34)	(3.18)
$\log L_{k\tau(t)}$	-0.293**	-0.291**	-0.702***	-0.721***
	(-2.12)	(-2.11)	(-3.86)	(-3.88)
Const.	3.717***	3.759***	2.859**	3.280***
	(3.28)	(3.35)	(2.42)	(2.75)
N	300	300	215	215
adj. R^2	0.050	0.051	0.755	0.748

t statistics in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Discussions

5.1 Truncation Problem of Forward Citation for Quality-adjusted Patents

There exist long forward citation lags as reported by Hall et al. (2001), quality-adjusted numbers of patents is exposed to the truncation problem: recent patents tend to be undervalued because they do not have sufficient time lags for subsequent citations. To deal with this problem, we multiply citation-weighted patent applications by weights derived from the distribution of forward citation lags, introduced in Hall et al. (2001). We call this HJT weights. For consistency, we also re-estimate knowledge stock with using the HJT weights.

Table 5 shows the same estimations as before with HJT-adjusted patent applications. a_{kt}^{HJT} is HJT-adjusted patent applications in category k and year t. q_{kt}^{HJT} are per-patent quality with the HJT weights.

Table 5: Estimations with HJT weights.

	Whole sample				$t \leq 1990$				$t > 1990$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	a_{kt}^{HJT}	a_{kt}^{HJT}	$\ln q_{kt}^{HJT}$	$\ln q_{kt}^{HJT}$	a_{kt}^{HJT}	a_{kt}^{HJT}	$\ln q_{kt}^{HJT}$	$\ln q_{kt}^{HJT}$	a_{kt}^{HJT}	a_{kt}^{HJT}	$\ln q_{kt}^{HJT}$	$\ln q_{kt}^{HJT}$
$\log \hat{P}_{k\tau(t)}^{0.2}$	-1.918		-1.305		3.102**		0.641		-6.663***		-6.385***	
	(-1.37)		(-1.30)		(2.29)		(0.55)		(-3.14)		(-4.35)	
$\log \hat{P}_{k\tau(t)}^{0.5}$		-1.025		-0.550		1.590**		0.417		-3.283***		-3.165***
		(-1.36)		(-1.03)		(2.20)		(0.68)		(-2.63)		(-3.66)
$\log \hat{H}_{k\tau(t)}$	0.0397	0.171	0.0711	0.121	-0.283	-0.475*	-0.156	-0.221	0.917***	1.290***	0.761***	1.128***
	(0.22)	(0.68)	(0.54)	(0.67)	(-1.54)	(-1.89)	(-0.99)	(-1.03)	(3.18)	(3.10)	(3.93)	(3.97)
$\log \hat{G}_{k\tau(t)}$	0.644	-0.234	0.707	0.0172	-2.833**	-1.349**	-0.743	-0.495	6.261***	3.083***	5.662***	2.626***
	(0.55)	(-0.40)	(0.85)	(0.04)	(-2.46)	(-2.31)	(-0.75)	(-0.99)	(3.51)	(3.07)	(4.59)	(3.78)
$\log A_{k\tau(t)}^{HJT}$	0.692***	0.694***	-0.0429	-0.0411	0.342***	0.344***	-0.0287	-0.0281	0.489***	0.494***	-0.332***	-0.324***
	(9.31)	(9.34)	(-0.80)	(-0.77)	(3.73)	(3.74)	(-0.36)	(-0.36)	(3.26)	(3.26)	(-3.23)	(-3.11)
$\log Y_{k\tau(t)}$	0.657***	0.655***	0.252***	0.246***	1.139***	1.147***	0.143	0.141	1.130***	1.112***	0.616***	0.594***
	(6.50)	(6.50)	(3.47)	(3.40)	(8.04)	(8.12)	(1.19)	(1.18)	(4.27)	(4.16)	(3.47)	(3.29)
$\log L_{k\tau(t)}$	-0.398***	-0.399***	-0.423***	-0.422***	-0.785***	-0.791***	-0.285**	-0.284**	-1.009***	-1.034***	-0.744***	-0.760***
	(-3.18)	(-3.18)	(-4.75)	(-4.73)	(-4.80)	(-4.84)	(-2.06)	(-2.05)	(-3.68)	(-3.72)	(-4.07)	(-4.07)
Const.	-3.852***	-3.830***	3.586***	3.718***	-2.629**	-2.748**	3.906***	3.947***	-5.562***	-5.091***	2.668**	3.123***
	(-3.90)	(-3.91)	(5.54)	(5.81)	(-2.02)	(-2.13)	(3.46)	(3.54)	(-3.19)	(-2.90)	(2.27)	(2.64)
Overdispersion	-3.223***	-3.223***			-4.028***	-4.027***			-3.770***	-3.757***		
	(-51.62)	(-51.62)			(-48.21)	(-48.20)			(-38.77)	(-38.62)		
N	515	515	515	515	300	300	300	300	215	215	215	215
adj. R^2			0.472	0.472			0.369	0.369			0.101	0.075
pseudo R^2	0.175	0.175			0.197	0.197			0.199	0.198		

t-statistics in parentheses. All regressions consider fixed effects of technological categories and year dummies.

Regressions (1)-(2), (5)-(6), and (9)-(10) are negative binomial regressions. The others are linear regressions.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.2 Does polarization just respond to more detailed technological categories?

One possible explanation of increasing polarization is that technology has become more segmented. A new polar that emerged in a 2-digit technological category could be a new field or a new product. If so, observed increasing polarization does not imply inter-group competition. To evaluate this possibility, we check finer primary classifications (3-digit class defined by USPTO) of patents of each firm and observe the distribution of the classifications on each technology space. More concretely, given technology maps created in Section 2, we put 3-digit class lists for firms in each 2-digit subcategory and 5-year window. Then take the average distances only among firms associated with each 3-digit class.

If 3-digit classes are randomly distributed, the average distances in a coarser classification is almost the same as that in finer classifications. The difference between them imply a bias from segmentation. If the average distance among 3-digit classes has a decreasing trend, it implies that finer classes have concentrated on polars and thus the segmentation effect mainly explains polarization.

Figure 9 illustrates the time-series of those average distances. We draw two types of class distance. One is described above (shown as "Class" in the figure). The other is that we focus on the most important 3-digit class for each firm ("Top class only"), where the top class of a firm is defined as the class in which the firm applied patents most frequently in each 5-year window (we include both in a tie). Naturally, the average distance within 3-digit classes tends to be lower than that within subcategories. The important fact here is that the average distance among 3-digit classes also have an upward trend. Since the trend is relatively weak so that relative distance among classes to among subcategories have been decreasing, some part of polarization should be attributed to the segmentation effect. However, it does not seem the main source of polarization.

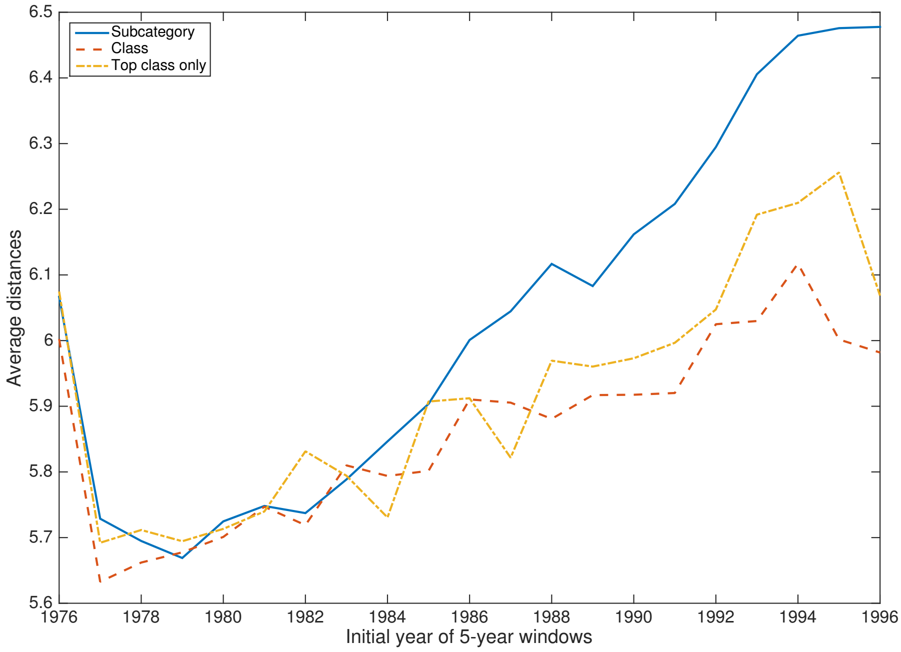

Figure 9: The average distances within 3-digit classes.

5.3 Relation between the Increasing Dissimilarity and Polarization

5.3.1 Decomposition of Polarization

As shown in DER, the measure of polarization can be decomposed into three components. such as

$$P^{\alpha} = G\bar{\iota}_{\alpha}(1+\rho), \tag{36}$$

where

$$g(i) \equiv \frac{1}{n}\sum_{j=1}^{n}\delta(x_i,x_j), \qquad G = \frac{1}{n}\sum_{i=1}^{n} g(i),$$

$$\iota_{\alpha}(i) \equiv f(x_i)^{\alpha}, \qquad \bar{\iota}_{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\iota_{\alpha}(i),$$

$$\rho \equiv \frac{\mathrm{Cov}(\iota_{\alpha}, g)}{\bar{\iota}_{\alpha} G},$$

In other words, polarization equals the product of average distance, concentration with polarization parameter which DER call *identification*, and their normalized covariance. Figure 10 depicts the time-series behaviors of unweighted-average identification and normalized covariance for $\alpha \in \{0.2, 0.5\}$ across technological categories.

When we apply the growth accounting on (36) with these averaged variables, the degree of contribution of the average distance, $\hat{G}$, is about 246% whereas −158% from the change in identification and 12% from the change in normalized covariance if we set the initial year window as 1977-1981.[21] [22] Therefore, the increasing polarization is mainly driven by increasing average technological distance, which is from increasing average dissimilarity. In the next subsection, we consider whether there exists a mechanism to derive an increasing dissimilarity in our methodology.

5.3.2 Citation overlaps with random citations

As the number of patents have been drastically increasing in the recent decades, the expansion of the pool of citable patents may decrease citation overlaps. This is one possible explanation of the observed upward trend in technological dissimilarity. In this subsection, we examine how plausible this explanation is by experimentation.

Suppose that two firms independently apply p patents, each of them cites m patents out of N patents at random. Let p be the average patent application per firm in each year, m be the average number of patent citations of those patent applications, and N be the number of patents previously granted. We consider 10 years and 25 years lag for backward citations.[23] We obtain the average first-order citation overlap from 5000 random draws of the lists of citations for each category and year from 1976 to 2000. Then we take the average of technological distances for each 5-year windows.

The left panel of Figure 11 shows the average technological distances of random citation firms for 10 and 25 years backward citation lags, and the actual technological distance obtained in Section 2. The right panel is the actual distance relative to the average distances with random citations. Since the average distance with random citation is considered as the baseline distance, the relative distances shown in the right panel tell us the real similarity or dissimilarity between firms. As seen in the

[21] If we use 1976-1980 as the first 5-year window, the numbers are 131%, −34%, and 3%, respectively. It is because identification in 1976-1980 year window is extremely low.

[22] Equation (36) do not exactly hold with sample statistics. Thus, we rescaled the numbers. The non-rescaled percentages sum up to 106%.

[23] Hall et al. (2001) report that about 50% of citations occur within 10 years after patent grant, about 75% within 25 years, and about 95% within 50 years. The result for 50 years lag for backward citations is almost the same as the result with 25 years lag in the current experimentation.

figure, the distances with random citations are not increasing from the early 1980s. Thus, the relative distances also have increasing trends in those periods. While the citation pool have been expanded since the late 1980s, firms apply more patents and citations of each patent have been also increasing. Hence, we conclude that the observed upward trend in technological distances is not from the expansion of citation pool.

5.4 Technology Group?

The polarization measure is a continuous statistic and we do not identify exact polars and boundaries of technology groups. It is not clear if closely distant firms tend to be in the same technology group in our distributions. To see the relationship among firms, we examine how distances in post-MDS distributions affect patent citation activity. Table 6 shows the results. Column (1) shows the logistic regression in which the dependent variable is whether citation occurs. Column (2) is the negative binomial regression with the number of citations as the dependent variable. Columns (3) and (4) are modifications of Column (2), where the number of citations are counted only within 5 years after granted, between 5 and 10 years after granted, respectively. The post-MDS distances negatively and nonlinearly affect citation activities, which is consistent with the idea of technology groups. All estimations include category and year dummies.

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: An Inverted-U Relationship. *The Quarterly Journal of Economics 120*(2), 701–728.

Akcigit, U., M. A. Celik, and J. Greenwood (2013). Buy, Keep or Sell: Economic Growth and the Market for Ideas. NBER Working Papers 19763, National Bureau of Economic Research, Inc.

Bar, T. and A. Leiponen (2012). A measure of technological distance. *Economics Letters 116*(3), 457–459.

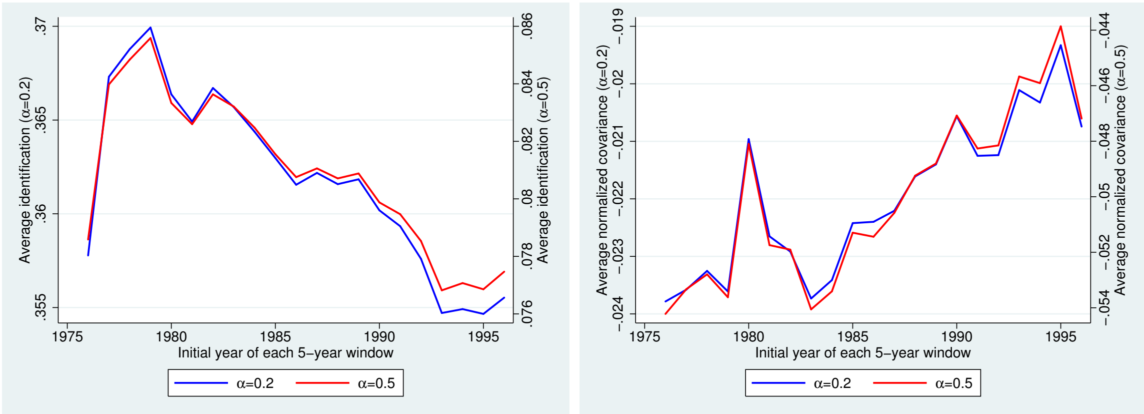

Figure 10: The average identification (left panel) and the average normalized covariance between alienation and identification (right panel) across technological categories
.

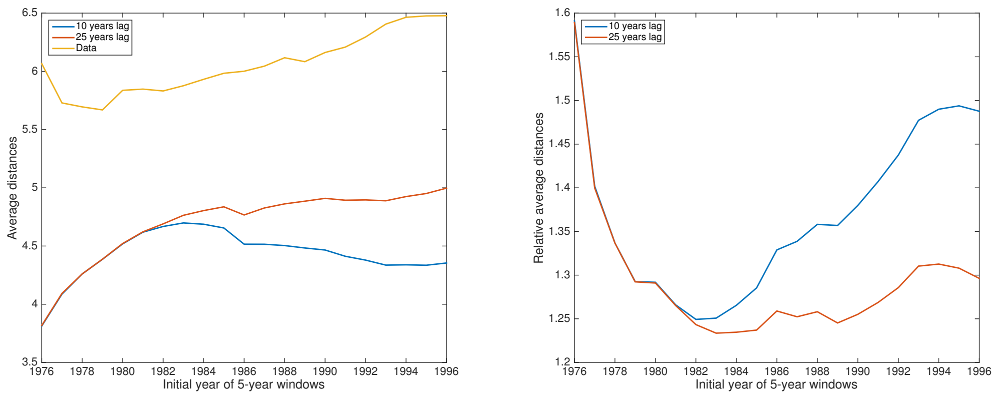

Figure 11: Experimented average distances with random citation (the left panel) and the average distance in data relative to the distances with random citations (the right panel).

Table 6: Distance vs. Citation

	(1)	(2)	(3)	(4)
	Logit	Neg.Bin	Neg.Bin	Neg.Bin
	Cite	Num. Cite	Num. Cite (< 5 years)	Num. Cite (< 10 years
δ_{ij}	-0.0170***	-0.453***	-0.472***	-0.455***
	(-15.87)	(-270.16)	(-257.14)	(-224.72)
δ_{ij}^2	-0.00755***	0.0200***	0.0200***	0.0199***
	(-87.68)	(155.15)	(140.47)	(126.92)
Const.	-3.798***	-1.288***	-2.241***	-2.391***
	(-510.88)	(-126.09)	(-187.27)	(-184.13)
Overdispersion		4.117***	4.228***	4.426***
		(3990.36)	(3206.80)	(3000.77)
N	33321080	33321080	33321080	33321080
pseudo R^2	0.054	0.038	0.051	0.044

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include fixed effects of technological categories and year dummies.

Benner, M. and J. Waldfogel (2008). Close to you? bias and precision in patent-based measures of technological proximity. *Research Policy 37*(9), 1556–1567.

Bloom, N., M. Schankerman, and J. van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica 81*(4), 1347–1393.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese (2010). Kernel density estimation via diffusion. *The Annals of Statistics 38*(5), 2916–2957.

Cox, T. F. and M. A. Cox (2001). *Multidimensional Scaling* (2nd ed.). Chapman & Hall.

Dosi, G. (1982). Technological paradigms and technological trajectories : A suggested interpretation of the determinants and directions of technical change. *Research Policy 11*(3), 147–162.

Duclos, J.-Y., J. Esteban, and D. Ray (2004). Polarization: Concepts, measurement, estimation. *Econometrica 72*(6), 1737–1772.

Esteban, J. and D. Ray (1994). On the Measurement of Polarization. *Econometrica 62*(4), 819–51.

Esteban, J. and D. Ray (2011). Linking Conflict to Inequality and Polarization. *American Economic Review 101*(4), 1345–74.

Griliches, Z. (1958). Research Costs and Social Returns: Hybrid Corn and Related Innovations. *Journal of Political Economy 66*, 419.

Hall, B. H. (2007). Measuring the Returns to R&D: The Depreciation Problem. NBER Working Papers 13473, National Bureau of Economic Research, Inc.

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

Hall, B. H. and R. H. Ziedonis (2001, Spring). The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995. *RAND Journal of Economics 32*(1), 101–28.

Jaffe, A. B. (1986). Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value. *American Economic Review 76*(5), 984–1001.

Jaffe, A. B. (1989). Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy 18*(2), 87–97.

Jaffe, A. B. and J. Lerner (2004). *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It.* Princeton University Press.

Kortum, S. and J. Lerner (1998, June). Stronger protection or technological revolution: what is behind the recent surge in patenting? *Carnegie-Rochester Conference Series on Public Policy 48*(1), 247–304.

Lanjouw, J. O. and M. Schankerman (2004, April). Protecting Intellectual Property Rights: Are Small Firms Handicapped? *Journal of Law and Economics 47*(1), 45–74.

Li, W. C. Y. (2012). Depreciation of Business R&D Capital. Technical report, Bureau of Economic Analysis.

Rosenkopf, L. and P. Almeida (2003). Overcoming local search through alliances and mobility. *Management Science 49*(6), 751–766.

Stuart, T. E. and J. M. Podolny (1996). Local search and the evolution of technological capabilities. *Strategic Management Journal 17*, 21–38.

TOYOTA Motor Corporation (2015). Toyota opens its fuel cell vehicle patents for free use. News Release, January 5th.

Weitzman, M. L. (1998). Recombinant Growth. *Quarterly Journal of Economics 113*(2), 331–360.

A The Valid Range of α

The coefficient of polarization, α, has the upper and lower bound to satisfy the axioms introduced in Duclos et al. (2004).

A.1 The Upper Bound of α

The definition of polarization measurement:

$$P(f) \equiv k \int \int f(\boldsymbol{x})^{1+\alpha} f(\boldsymbol{y}) \delta(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y} \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2. \tag{37}$$

Suppose f is symmetric and has a finite support, C_m. C_m is a "circle" with the center vector of $\boldsymbol{m} \equiv (m, m)$. δ is the Euclidean distance. Symmetry is in the sense that

$$\delta(\boldsymbol{x}, \boldsymbol{m}) = \delta(\boldsymbol{y}, \boldsymbol{m}) \quad \Rightarrow \quad f(\boldsymbol{x}) = f(\boldsymbol{y}).$$

Let λ-squeezed density be denoted by:

$$f^{\lambda}(\boldsymbol{x}) = \frac{1}{\lambda^2} f\left(\frac{\boldsymbol{x} - (1-\lambda)\boldsymbol{m}}{\lambda}\right). \tag{38}$$

Then,

$$\begin{aligned} P(f^{\lambda}) &= k \int f^{\lambda}(\boldsymbol{x})^{1+\alpha} \int f^{\lambda}(\boldsymbol{y}) \delta(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x} \\ &= k\lambda^{-4-2\alpha} \int f\left(\frac{\boldsymbol{x} - (1-\lambda)\boldsymbol{m}}{\lambda}\right)^{1+\alpha} \int f\left(\frac{\boldsymbol{y} - (1-\lambda)\boldsymbol{m}}{\lambda}\right) \delta(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x}. \\ &= k\lambda^{1-2\alpha} \int f(\boldsymbol{x}')^{1+\alpha} \int f(\boldsymbol{y}') \delta(\boldsymbol{x}', \boldsymbol{y}') d\boldsymbol{y}' d\boldsymbol{x}', \end{aligned}$$

where we use $\delta(\boldsymbol{x}, \boldsymbol{y}) = \lambda \delta(\boldsymbol{x}', \boldsymbol{y}')$ and changes of variables with

$$\boldsymbol{x}' = \frac{\boldsymbol{x} - (1-\lambda)\boldsymbol{m}}{\lambda}, \quad \boldsymbol{y}' = \frac{\boldsymbol{y} - (1-\lambda)\boldsymbol{m}}{\lambda}.$$

Therefore, $P(f^{\lambda})$ is nondecreasing in λ if and only if $\alpha \leq 0.5$.

A.2 The Lower Bound of α

First, we quote Axiom 2 in DER:

Axiom 2 (DER) *If a symmetric distribution is composed of three basic densities with the same root and mutually disjoint supports, then a symmetric squeeze of the side densities cannot reduce polarization.*

In this axiom, a “basic density” is a symmetric and unimodal density with a compact support. A “root” is a normalized basic density. A “squeeze” is of density f is defined in (38).

We modify Axiom 2 (DER) to the following Axiom 2’ so that we apply it to 2-dimensional distributions.

Axiom 2’ *If a line-symmetric distribution is composed of two basic densities with the same root and mutually disjoint circles as supports, then a symmetric squeeze of the densities cannot reduce polarization.*

In other words, a double squeeze of two identical balls with the same symmetric and unimodal density cannot lower polarization.

Before moving on the proof, define u as the uniform basic density such as

$$u_{m,r}(x) = \begin{cases} 1/(\pi r^n) & \text{if } \|x-m\| \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

And consider a distribution that consists of $u_{(-a,0),r}$, where$a > 0$, and a stretched uniform basic density of $u_{(a,0),1}$ by $s_1 > 0$ along the horizontal axis and by $s_2 > 0$ along the vertical axis. The average distance of pairs on any two points on the support of this distribution can be written as

$$A_{r,(s_1,s_2)}(a) = \int\int u_{(-a,0),r}(x) u_{(a,0),1}(y) \|x - (s_1(y_1 - a) + a, s_2 y_2)\| dx dy. \tag{40}$$

Lemma 1 *For $r, s_1, s_2 > 0$ and $a > (r + s_1)/2$, $A_{r,(s_1,s_2)}(a)$ and $A_{r,(s_1,s_2)}(0)$ are increasing in both s_1 and s_2, and $A'_{r,(s_1,s_2)}(a)$ is decreasing in both s_1 and s_2.*

Proof. First, consider $A_{r,(s_1,s_2)}(0)$. Note that for $y_1 > 0$, $\int_{-r}^{r} \|(x_1, 0) - (y_1, y_2)\| dx_1$ is increasing in y_1 since $\|(r, 0) - (y_1, y_2)\| < \|(-r, 0) - (y_1, y_2)\|$ when $y_1 > 0$. Thus, for $s > 0$, $\int_{-s}^{s} \int_{-r}^{r} \|(x_1, 0) - (y_1, y_2)\| dx_1 dy_1/s$ is increasing in s. Therefore, $A_{r,(s_1,s_2)}(0)$ is increasing in s_1. Similarly, $A_{r,(s_1,s_2)}(0)$ and $A_{r,(s_1,s_2)}(a)$ are increasing in s_2. Next, suppose $y_1 > x_1$ and $y_2 > 0$. Since $\partial\|(x_1 - a, x_2) - (y_1, y_2)\|/\partial a = 1/\sqrt{1 + (x_2 - y_2)^2/(x_1 - a - y_1)^2}$ is larger at $x_2 = r$ than at $x_2 = -r$, $A'_{r,(s_1,s_2)}(a)$ is decreasing in s_2.

Since for $y_1 > x_1$, $\partial\|(x_1, x_2) - (y_1, y_2)\|/\partial y_1 = 1/\sqrt{1 + (x_2 - y_2)^2/(x_1 - y_1)^2}$ is increasing in y_1, $(\|(x_1, x_2) - (a + \Delta, y_2)\| + \|(x_1, x_2) - (a - \Delta, y_2)\|)/2$ is increasing in $\Delta \in (0, a - x_1)$. Thus, $A_{r,(s_1,s_2)}(a)$ is increasing in s_1. Similarly, since $\partial(\partial\|(x_1 - a, x_2) - (y_1, y_2)\|/\partial a)/\partial y_1 = (x_2 - y_2)^2/((x_1 - a - y_1)^2 + (x_2 - y_2)^2)^{3/2}$ is decreasing in y_1, $A'_{r,(s_1,s_2)}(a)$ is decreasing in s_1. ■

Now we search the lowest α to satisfy Axiom 2'. First, notice the fact that a double squeeze is decomposed into an outward slide (so as to multiple the distance between the means by $1/\lambda$) and a global squeeze (so as to restore the distance),

$$0.5f^{\lambda} + 0.5g^{\lambda} = \overbrace{(\underbrace{0.5f_{(1/\lambda-1)(\mu_f-(\mu_f+\mu_g)/2)} + 0.5g_{(1/\lambda-1)(\mu_g-(\mu_f+\mu_g)/2)}}_{\text{outward slide with distance multiplied by } 1/\lambda})^{\lambda}}^{\text{distance restored, while each squeezed by } \lambda}, \tag{41}$$

where μ_h denotes the mean of density h, and h_d denotes the slide of h by d such that $h_d(x + d) = h(x)$. So is the growth rate of polarization by double squeeze. But the growth rate of polarization by a global squeeze only depends on the squeeze rate,

$$\frac{P_\alpha((0.5f_{(1/\lambda-1)(\mu_f-(\mu_f+\mu_g)/2)} + 0.5g_{(1/\lambda-1)(\mu_g-(\mu_f+\mu_g)/2)})^{\lambda})}{P_\alpha(0.5f_{(1/\lambda-1)(\mu_f-(\mu_f+\mu_g)/2)} + 0.5g_{(1/\lambda-1)(\mu_g-(\mu_f+\mu_g)/2)})} = \lambda^{1-n\alpha}. \tag{42}$$

Thus, the growth rate of polarization of density (41) by a double squeeze is

$$\frac{d\ln P_\alpha(0.5f^{\lambda} + 0.5g^{\lambda})}{d\ln(1/\lambda)} = \frac{d\ln P_\alpha(0.5f_{(1/\lambda-1)(\mu_f-(\mu_f+\mu_g)/2)} + 0.5g_{(1/\lambda-1)(\mu_g-(\mu_f+\mu_g)/2)})}{d\ln(1/\lambda)} + n\alpha - 1, \tag{43}$$

The first term in the right-hand side is the growth rate by the outward slide, which we focus below.

A basic density is decomposed into uniform basic densities with the same mean,

$$f = \int u_{\mu_f,r} dW_f(r) \tag{44}$$

for some distribution function W_f.[24]

Now consider a distribution that consists of two disjoint symmetric basic densities, $f_{-a}/2$ and $f_a/2$, where $(-a, 0)$ and $(a, 0)$ are the mean of each basic density,

[24] In particular, suppose that f is differentiable. Then,

$$\frac{\partial f((r, 0) + \mu_f)}{\partial r} = -\frac{1}{\pi r^2} W'_f(r). \tag{45}$$

respectively ($a > 0$). According to (44), polarization of this density is decomposed into average distances between decomposed uniform densities with different levels of double squeezes,

$$P_\alpha((f_{-a}+f_a)/2) = K_{f,\alpha} \int\int\int \frac{u_{-a,r}(x) + u_{a,r}(x)}{2} \frac{u_{-a,s}(y) + u_{a,s}(y)}{2} \|x-y\| dxdydV_f(r,s) \tag{46}$$

for some $K_{f,\alpha} > 0$ and distribution function V_f. The constant $K_{f,\alpha}$ in (46) cab be derived as follows. Suppose that f is differentiable. Then,[25]

$$K_{f,\alpha} dV'_{f,\alpha}(r,s) = (1+\alpha) \left(\frac{f((r,0)+\mu_f)}{2} \right)^\alpha W'_f(r)drW'_f(s)ds \tag{47}$$

with

$$K_{f,\alpha} = \int\int (1+\alpha) \left(\frac{f((r,0)+\mu_f)}{2} \right)^\alpha W'_f(r)drW'_f(s)ds. \tag{48}$$

Given (46), we can focus on the lower bound of the growth rate of such distances,

$$\frac{d\ln P_\alpha((f_{-a}+f_a)/2)}{d\ln\|a\|} \geq \inf_{r,s\leq r_f} \frac{d\ln\int\int (u_{-a,r}(x)+u_{a,r}(x))(u_{-a,s}(y)+u_{a,s}(y))\|x-y\|dxdy}{d\ln\|a\|}. \tag{49}$$

where r_f denotes the radius of the support of f.

Lemma 1 implies that the growth rate decreases as each radius increases, i.e., for $r' \geq r$,

$$\begin{aligned} &\frac{d\ln\int\int (u_{-a,r}(x)+u_{a,r}(x))(u_{-a,s}(y)+u_{a,s}(y))\|x-y\|dxdy}{d\ln\|a\|} \\ \geq &\frac{d\ln\int\int (u_{-a,r'}(x)+u_{a,r'}(x))(u_{-a,s}(y)+u_{a,s}(y))\|x-y\|dxdy}{d\ln\|a\|}. \end{aligned} \tag{50}$$

Thus, it suffices to consider the growth rate of the average distance within uniform

[25] Suppose a density g can be decomposed as

$$g(x) = \int w(r)h_r(x)dr.$$

Letting $(1+\alpha)w(r)g^\alpha(x) = v(r)$, we have

$$\left(\int w(r)h_r(x)dr\right)^{1+\alpha} = \int v(r)h_r(x)dr.$$

identical balls touching each other,

$$\frac{d\ln P_\alpha((f_{-a}+f_a)/2)}{d\ln\|a\|} \geq \lim_{r\uparrow\|a\|} \frac{d\ln\int\int(u_{-a,r}(x)+u_{a,r}(x))(u_{-a,r}(y)+u_{a,r}(y))\|x-y\|dxdy}{d\ln\|a\|}$$
$$= \lim_{r\uparrow 1} \left.\frac{d\ln P_\alpha((u_{(-1-e,0),r}+u_{(1+e,0),r})/2)}{d\ln(1+e)}\right|_{e=0}, \tag{51}$$

which is also equal to the growth rate of polarization by the outward slide. Therefore, the lower bound of α is attained by solving the equation at such density,

$$\lim_{\lambda\uparrow 1} \frac{d\ln P_\alpha((u^\lambda_{(-1,0),1}+u^\lambda_{(1,0),1})/2)}{d\ln(1/\lambda)} = \lim_{r\uparrow 1} \left.\frac{d\ln P_0((u_{(-1-e,0),r}+u_{(1+e,0),r})/2)}{d\ln(1+e)}\right|_{e=0} + n\alpha - 1 = 0. \tag{52}$$

We can search numerically, which α makes (52) satisfy equality. We obtain $\underline{\alpha} \approx 0.202$.[26]

B Technological Categories and Estimation of Knowledge Capital Stock

Table 7 is the list of the 2-digit technological categories defined by USPTO.

Knowledge stock, $A_{k,\tau(t)}$, is estimated by the cumulative number of citation-weighted patents applied till the beginning of $\tau(t)$, namely period $t-5$. In other words,

$$A_{k,\tau(t)} = \sum_{s=0}^{t-5}(1-\zeta)^{t-5-s}a_{k,s}, \tag{53}$$

where ζ_k is the depreciation rate of R&D in technological category k. Since the dataset contains patents from 1951 and the initial state is not significant, we simply assume the initial knowledge stock is zero. To calculate $A_{k,\tau(t)}$, we apply the R&D depreciation rates estimated in Li (2012). Table 8 summarizes the result reported in Li (2012) with zero gestation lag of R&D.

By matching technological categories defined in USPTO with the list of industries in Table 8, we use depreciation rates in Table B. We assign a depreciation rate of 15% to technological categories not listed in in Table B, which is the traditional number assumed in Griliches (1958) (cf. Hall (2007) for details).

[26] For the one-dimensional case, this procedure exactly yields 0.25, the same lower bound in DER.

Table 7: Technological categories.

1-digit categories	2-digit categories
1. Chemical	Agriculture, Food, Textiles (11); Coating (12); Gas (13); Organic Compound (14); Resins (15)
2. Computers & Communications	Communications (21); Computer Hardware & Software (22); Computer Peripherals (23); Information Storage (24)
3. Drugs & Medical	Drugs (31); Surgery & Medical Instruments (32); Biotechnology (33)
4. Electrical & Electric	Electrical Devices(41); Electrical Lighting(42); Measuring & Testing (43); Nuclear & X-rays (44); Power Systems (45); Semiconductor Devices(46)
5. Mechanical	Mat. Proc & Handling (51); Metal Working (52); Motors & Engines + Parts (53); Optics (54); Transportation (55)
6. Others	Agriculture, Husbandry, Food (61); Amusement Devices (62); Apparel & Textile (63); Earth Working & Wells (64); Furniture,House Fixtures (65); Heating (66); Pipes & Joints (67); Receptacles (68)

Table 8: Summary of Depreciation Rates of Business R&D Assets Based on BEA-NSF Dataset

Industry	Depreciation rate
a. Computers & peripheral equipment	40%
b. Software	22%
c. Pharmaceutical	10%
d. Semiconductor	25%
e. Aerospace	22%
f. Communication equipment	27%
g. Computer system design	36%
h. Motor vehicles, bodies & trailers, & parts	31%
i. Navigational, measuring, electromedical, & control instruments	29%
j. Scientific research & development	16%

Table 9: Summary of Depreciation Rates in the Current Study

Technological category	Depreciation rate
Communication (21)	27% (f)
Computer Peripherals (23)	40% (a)
Other computers & communications (22,24)	33% (mean of a, b, g)
Drugs & Medical (31-33)	10% (c)
Measuring & Testing (43)	29% (i)
Semiconductor Devices (46)	25% (d)
Motors & Engines + Parts (53)	31% (mean of e and h)
Transportation (55)	27%

The numbers in the parentheses in the left column indicate 2-digit technological categories. The alphabets in the parentheses in the right column indicate industries described in Table 8. Depreciation rates for other categories are 15%.