cemmap

centre for microdata methods and practice

Knowledge spillovers and patent citations: trends in geographic localization, 1976-2015

Hyuk-Soo Kwon
Jihong Lee
Sokbae Lee
Ryungha Oh

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP55/17

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

KNOWLEDGE SPILLOVERS AND PATENT CITATIONS: TRENDS IN GEOGRAPHIC LOCALIZATION, 1976-2015

HYUK-SOO KWON[1], JIHONG LEE[2], SOKBAE LEE[3,4], AND RYUNGHA OH[2]

ABSTRACT. This paper examines the trends in geographic localization of knowledge spillovers via patent citations, considering US patents from the period of 1976-2015. Despite accelerating globalization and widespread perception of the "death of distance," our multi-cohort "matched-sample" study reveals significant and *growing* localization effects of knowledge spillovers at both intra- and international levels after the 1980s. We also develop a novel network index based on the notion of "farness," which an instrumental variable estimation shows to be a significant and sizable determinant of the observed trends at the state-sector level.

KEYWORDS: Innovation, knowledge spillovers, patent citation, agglomeration, network index, farness

JEL CLASSIFICATION: C36, C81, O33, O34, O51

[1]DEPARTMENT OF ECONOMICS, CORNELL UNIVERSITY, 401 E URIS HALL, ITHACA, NY 14853, USA

[2]DEPARTMENT OF ECONOMICS, SEOUL NATIONAL UNIVERSITY, 1 GWANAK-RO, GWANAK-GU, SEOUL 08826, KOREA

[3]CENTRE FOR MICRODATA METHODS AND PRACTICE, INSTITUTE FOR FISCAL STUDIES, 7 RIDGMOUNT STREET, LONDON, WC1E 7AE, UK

[4]DEPARTMENT OF ECONOMICS, COLUMBIA UNIVERSITY, 1022 INTERNATIONAL AFFAIRS BUILDING, 420 W 118TH STREET, NEW YORK, NY 10027, USA

E-mail addresses: hk882@cornell.edu, jihonglee@snu.ac.kr, sl3841@columbia.edu, ryungha@snu.ac.kr.

Date: November 2017.

The authors wish to thank David Miller and Stephen Redding for helpful comments. This research was supported by Seoul National University Asia Center (#SNUAC-2017-008), the European Research Council (ERC-2014-CoG-646917-ROMIA), and the UK Economic and Social Research Council (ES/P008909/1 via CeMMAP).

1. Introduction

Patents offer a unique source of information on the patterns of economic activity. Through the patent classification system, and from the identity and location of inventors, one can trace the evolution of economic structure across different industries and geographic/institutional boundaries. At the same time, citations data facilitate a measure of the quality of innovation (e.g. Trajtenberg, 1990; Lanjouw and Schankerman, 2004; Hall, Jaffe, and Trajtenberg, 2005), thereby providing some clues about the engine of long-run economic success.

This paper examines another important aspect of information provided by patents, the *geography* of economic activities. Valuable statistics can be extracted not just from simple summation of each piece of information that patents carry. When one approaches an entire pool of patents as a collective, a variety of networks of economic activities emerge. Citations are an explicit source of links between technologies, individuals, and locations. Even when two patents are not linked via citations, a match in technology class or inventor location implies the presence of another, implicit, link.

Jaffe, Trajtenberg, and Henderson (1993), henceforth JTH, develop a pioneering analysis of such network structures of patents to examine the patterns of economic *agglomeration*. Their basic idea is two-fold. On the one hand, patents that are linked by citations *and* by location would reflect the role of geographic distance in how knowledge spreads. When two patents are not connected by a citation link but share the same inventor location, on the other hand, we may be observing a possible outcome of productive agglomeration.

For a particular cohort of US patents, JTH first finds all patents that cite one of the "originating" patents at least once over a sample duration and calculates the frequency of such "citing" patents whose inventor location matches the location of the cited patent. Then, they compare this citation "matching rate" vis-à-vis the frequency of geographic match between the originating patents and a group of "control" patents that are selected to represent technological proximity. Three fixed geographic boundaries–country, state, and metropolitan statistical area–are considered. Their findings display significant *net* localization effects: citations, or knowledge *spillovers*, are more geographically concentrated than the production of knowledge at both intra- and international levels.

The "matched-sample" analysis of JTH suggests that geographic distance matters more for the spread of knowledge than other channels of externalities that shape agglomerative patterns.[1] A potential caveat of JTH, however, is that their results may not be robust to

[1]Numerous papers have found the importance of distance in knowledge spillovers. For a recent survey on innovation and agglomeration, see Carlino and Kerr (2015).

the specification of control patents that would reflect the existing geography of knowledge production. In particular, Thompson and Fox-Kean (2005), henceforth TFK, suggest that, as JTH selects controls only by filtering a technological match between citing and control patents, their method may not guarantee proximity between the pair of control and originating patents. When TFK selects control patents using more stringent criteria, localization effects of knowledge spillovers disappear at *intra*-national levels.[2]

This paper is an attempt to extend this literature in another hitherto unresolved direction. The existing literature on the geography of knowledge spillovers has largely been confined to examination of a particular cohort of patents. TFK, for example, follows citing and control patents that correspond only to patents granted during a single month of 1976. This is, in our view, a serious gap in the literature, since the recent decades have witnessed a meteoric rise in both the number of inventions and the diversity of countries that have joined the global technology ladder (e.g. Kwon, Lee, and Lee, 2017). Moreover, these decades have been associated with the development of internet and other new communication technologies that has even prompted the notion of the "death of distance" (e.g. Cairncross, 2001; Coyle, 1997).

We examine the trends in localization effects of knowledge spillovers considering utility patents granted by the US Patent and Trademark Office (USPTO) during the period of 1976-2015. Our analysis is based on four cohorts of originating patents, each consisting of all corporate and institutional patents granted in 1976, 1986, 1996, and 2006, respectively. The corresponding citing and control patents are found over a fixed 10-year window, and multiple disaggregated criteria are adopted for control selection *à la* TFK. We consider localization effects for each pair of geographic boundary (country, state, or metropolitan statistical area) and industry sector (one of 37 sub-categories defined by NBER).

The first part of our analysis computes localization effects for different cohorts. The results on the 1976 cohort are similar to those obtained by TFK and confirm the lack of intra-national (net) localization effects at the most disaggregated control level that accounts for a technological match between control and originating patents.[3]

In terms of trends, we find that (i) the matching rates between citing and originating patents have grown at all levels of control and spatial boundary since the 1986 cohort; (ii) the matching rates between control and originating patents have increased

[2]In a recent work, Murata, Nakajima, Okamoto, and Tamura (2014) recover intra-national localization effects by employing in the JTH framework the "continuous-distance" metric developed by Duranton and Overman (2005). See also Carlino, Carr, Hunt, and Smith (2012) and Kerr and Kominers (2015).

[3]One important difference is that our cohort consists of patents granted in all of year 1976, as opposed to just one month (January) taken by TFK. This has mitigated the sample size issue pointed out by Henderson, Jaffe, and Trajtenberg (2005).

at intra-national levels but decreased at international level. These findings suggest that concentration of innovation activities has intensified within the US, consistent with other observations on the trends of industrial agglomeration (e.g. Moretti, 2012), but international border effect, or "home bias," has deepened only for diffusion of innovation.

More importantly, our data reveal evidence of highly significant (net) localization effects at every unit of analysis since the 1986 cohort; moreover, the extent of such effects has been *growing*. Surprisingly, spread of ideas has become increasingly more localized than production of ideas, contrary to the common expectation otherwise.

We also compute localization effects across all US states and across six industry categories defined by NBER. It turns out that the rise in localization effects has been accompanied by greater heterogeneity in matching rates at both state and industry levels. We see growing importance of California and few other states as a driving force behind the aggregate trends. This finding is in line with others who have also shown stronger localization effects of innovation activities in certain regions (e.g. Almeida and Kogut, 1999). A noteworthy observation from the industry comparison is that the reduced home bias in knowledge production is associated with certain industries, especially electronics.

The latter part of our study is an attempt to explain the observed trends using variations at the state-sector level. We focus on two particular channels. First, motivated by the leading role played by several states, we ask whether there exist "natural advantages" for the diffusion of ideas. Such factors have been extensively tested in the urban economics literature on agglomeration (e.g. Ellison and Glaeser, 1997, 1999). Second, while knowledge spillovers are understood to be a significant determinant of productive agglomeration (e.g. Marshall, 1890; Rosenthal and Strange, 2001; Ellison, Glaeser, and Kerr, 2010), the causality may also run in the opposite direction.

To capture these effects, we develop a novel network index for each state-sector pair. Our (weighted) "farness index" measures the average distance from each state to another state (i.e. the inverse of "closeness centrality" proposed by Bavelas (1950) and Sabidussi (1966) in social network theory) with the distance divided by the number of patents produced by the other state in the given sector. Our idea is that geographic isolation and relative concentration of innovative activities would make local exchange of ideas more likely. To account for endogeneity, we adapt Moretti (2004) and use the locations of land-grant universities for construction of an instrumental variable. Specifically, our "farness-in-research index" measures each state's average distance to the land-grant universities with every such distance weighted by appropriate patent counts taken before our sample period.

Our IV regression results show significant effects of the farness index. Under the baseline 2SLS estimate, one standard deviation increase in the index leads to a 26 percentage point increase in the localization effect of knowledge spillovers. Controlling for interactions between farness and cohort dummies, as well as lagged dependent variable and own patent share of each state-sector, our estimates indicate that one standard deviation increase in farness generates roughly a 24 percentage point increase in the localization effect from the 1996 cohort to the 2006 cohort. The effects of farness are indeed stronger in recent cohorts, during which greater variations in both citing and control matching rates have appeared across states and industries.

Only few existing papers address how the role of geographic proximity in knowledge spillovers has changed over time. Among these, two papers consider the trends of home bias across national boundaries. Keller (2002) estimates an R&D production function with R&D of other countries as explanatory variable. His results show that the importance of foreign R&D has fallen over the years 1970-1995, suggesting faster diffusion of knowledge across borders. Griffith, Lee, and Van Reenen (2011) examine a panel of USPTO patents granted and citations made to these patents between 1975 and 1999. Using a duration model, they estimate the speed of citations and find evidence of declining "diffusion lag" between domestic and foreign citations.

Regarding intra-national localization trends, Lychagin, Slade, Pinkse, and Van Reenen (2016) examine R&D spillovers into US-based firm productivity over the period 1980-2000 and find no evidence of the "death of distance." Using economics and finance articles published over 1970-2001, Kim, Morse, and Zingales (2009) report evidence of declining local spillover benefits among top US universities.

Our findings stand in sharp contrast to these results. At both intra- and international levels, we observe increasing importance of geographic proximity in knowledge spillovers. One source of the departure may be the measure of diffusion. More importantly, the aforementioned papers (as well as most of the existing literature on knowledge spillovers) are based on datasets that do not include the most recent decades. The surge in patent production during this period makes it particularly important to exploit observations beyond the existing literature.[4]

The rest of the paper is organized as follows. We begin by describing the USPTO data in Section 2 and then our sample patents in Section 3. Our main findings on the aggregate trends of localization effects are presented in Section 4. Section 5 reports the framework

[4]In fact, the estimates of Griffith, Lee, and Van Reenen (2011) show that the declining trend of home bias might have been turning towards the end of their sample period. It would be interesting to conduct their analysis with our dataset.

and estimation results of our regression analysis. Section 6 concludes. Appendices contain materials left out from the main text for expositional reasons.

2. Patent Data

The patent dataset used in this paper is directly extracted from the USPTO bulk data which contain information on all utility patents granted from January 1976 up to, and including, May 2015. The data include patent number, application date, main and additional technology classifications, name of assignee, names and locations of inventors, and patent numbers of cited patents.[5]

Every patent is endowed with a single mandatory "original" (OR) classification and additional "cross-reference" (XR) classifications. The US patent classification (USPC) system is a tree structure consisting of distinct, and mutually exclusive, technology "classes" and "subclasses" that are nested under their parent (sub)classes.[6] For utility patents, classes are identified by a one-, two-, or three-digit integer; each subclass is identified by an additional "indent," indicating its position within a class hierarchy, and a subsequent alphanumeric code. The most disaggregated level of subclasses has nine alphanumeric digits. A group of subclasses are classified as "primary subclasses," and the mandatory original classification must belong to this group.

Our dataset, unsurprisingly, reveals substantial growth in technological diversity. Among all the patents granted between 1976-1985 we found 113729 distinct subclasses, and this number increased to 239233 over the entire sample period.[7] Despite this expansion of technological spectrum, the level of specialization has been relatively stable. On average, a patent granted in 1976 received about 3.6 subclass classification codes. It was about 4 for a patent granted in 2006.

For the purpose of our study, it is necessary to assign a geographic location to each patent, based on inventor location. As in JTH and TFK, our analysis is conducted at three different geographic levels: country, state, and CMSA (consolidated metropolitan statistical area). Since patents report inventor location only in terms of country, state and city, each patent is mapped to one of 17 CMSAs,[8] or a "phantom" CMSA created

[5]We obtained the bulk data for the period 1976-2014 from https://www.google.com/googlebooks/uspto-patents-grants-biblio.html and the data for 2015 from https://bulkdata.uspto.gov/.
[6]See "Handbook of Classification" published by USPTO.
[7]After revisions, USPTO was offering around 160000 subclasses as of June 2015.
[8]We follow TFK and use the method provided by the Office of Social and Economic Data Analysis (OSEDA) of the University of Missouri.

for foreign countries and each state.[9] If a patent is produced by a single inventor or by a group of inventors who reside in the same location, the location of the patent is unambiguously determined. For patents with multiple inventor locations, we randomly assigned a unique location, as done also by TFK.[10]

Table 1 breaks down all utility patents granted by USPTO during the sample period according to their locations, defined as domestic or foreign, and as states.

3. Sample Patents

We adopt the experimental design of JTH to document the trends in geographic localization of knowledge spillovers. This is based on constructing three samples of patents: originating, citing, and control patents.[11]

Originating Patents. A sample of "originating patents" consists of a fixed cohort of patents. Two cohorts of such patents (whose application dates were in 1975 and in 1980) were considered by JTH, and one cohort of patents (granted during January 1976) was used by TFK.

In this study, we construct four cohorts of originating patents: all relevant patents granted in 1976, 1986, 1996 and 2006 with at least one US-located inventor. The 1976 cohort is included to re-examine the previous analyses of JTH and TFK. The sample sizes of the two cohorts of originating patents in JTH were 950 and 1450, respectively, while the corresponding sample size in TFK was 2724. The sample sizes of our four cohorts of originating patents are 44016, 38160, 61581, and 80495, respectively.

Citing Patents. A sample of "citing patents" is constructed for each cohort of originating patents by collecting all patents that cite at least one of the originating patents within a fixed window of periods. In JTH, the 1975 and 1980 originating cohorts received 4750 and 5200 citations, respectively, by the end of 1989. TFK obtained 18551 citing patents granted between January 1976 and April 2001.

We use a window of 10 years (including the year in which originating patents were granted) for constructing the samples of citing patents.[12] This ensures that the citing

[9]For a very small number of domestic patents (0.2%), this mapping resulted in two CMSAs. The final CMSA was chosen randomly in these cases.

[10]JTH used a different method based on plurality. Our main results remain unchanged by adopting this rule.

[11]As in TFK, we consider patents assigned to corporation or institution. The detail of our sample selection/culling procedure is provided in Appendix A.

[12]One exception is the 2006 cohort, for which the citing patents were collected only up to, and including, May 2015. From June 2015, USPTO began a new system of patent classification, Cooperative Patent Classification (CPC), in an effort to harmonize its classification system with the European Patent Office

Table 1. Patent Counts

	1976 - 1985		1986 - 1995		1996 - 2005		2006 - 2015		Total	
Number of patents	596983		874190		1444740		2012412		4928325	
US patents	342441		447663		742251		952048		2484403	
Foreign patents	254542		426527		702489		1060364		2443922	
California(CA)	45938	(13.41%)	68384	(15.28%)	159655	(21.51%)	255844	(26.87%)	529821	(21.33%)
New York(NY)	29578	(8.64%)	36753	(8.21%)	52140	(7.02%)	58202	(6.11%)	176673	(7.11%)
Texas(TX)	18853	(5.51%)	30435	(6.80%)	54728	(7.37%)	69549	(7.31%)	173565	(6.99%)
Illinois(IL)	26291	(7.68%)	26302	(5.88%)	33032	(4.45%)	34864	(3.66%)	120489	(4.85%)
Michigan(MI)	19325	(5.64%)	24147	(5.39%)	33502	(4.51%)	35951	(3.78%)	112925	(4.55%)
New Jersey(NJ)	24872	(7.26%)	23496	(5.25%)	28195	(3.80%)	28467	(2.99%)	105030	(4.23%)
Ohio(OH)	21807	(6.37%)	23360	(5.22%)	29066	(3.92%)	27149	(2.85%)	101382	(4.08%)
Pennsylvania(PA)	21653	(6.32%)	22557	(5.04%)	27503	(3.71%)	27005	(2.84%)	98718	(3.97%)
Massachusetts(MA)	11257	(3.29%)	15145	(3.38%)	26790	(3.61%)	38002	(3.99%)	91194	(3.67%)
Minnesota(MN)	8204	(2.40%)	13229	(2.96%)	24734	(3.33%)	32722	(3.44%)	78889	(3.18%)
Washington(WA)	4838	(1.41%)	8329	(1.86%)	18637	(2.51%)	44352	(4.66%)	76156	(3.07%)
Florida(FL)	9019	(2.63%)	15543	(3.47%)	23217	(3.13%)	27593	(2.90%)	75372	(3.03%)
North Carolina(NC)	4396	(1.28%)	7839	(1.75%)	16302	(2.20%)	23336	(2.45%)	51873	(2.09%)
Colorado(CO)	4764	(1.39%)	7840	(1.75%)	17051	(2.30%)	20550	(2.16%)	50205	(2.02%)
Wisconsin(WI)	7202	(2.10%)	10424	(2.33%)	15894	(2.14%)	16010	(1.68%)	49530	(1.99%)
Indiana(IN)	8799	(2.57%)	9429	(2.11%)	12955	(1.75%)	13650	(1.43%)	44833	(1.80%)
Arizona(AZ)	4334	(1.27%)	7721	(1.72%)	14325	(1.93%)	18086	(1.90%)	44466	(1.79%)
Connecticut(CT)	8128	(2.37%)	10210	(2.28%)	12276	(1.65%)	12846	(1.35%)	43460	(1.75%)
Maryland(MD)	6520	(1.90%)	7712	(1.72%)	12459	(1.68%)	13161	(1.38%)	39852	(1.60%)
Oregon(OR)	2805	(0.82%)	5309	(1.19%)	12691	(1.71%)	19012	(2.00%)	39817	(1.60%)
Georgia(GA)	3279	(0.96%)	6118	(1.37%)	12390	(1.67%)	17579	(1.85%)	39366	(1.58%)
Virginia(VA)	5109	(1.49%)	6723	(1.50%)	9522	(1.28%)	12649	(1.33%)	34003	(1.37%)
Missouri(MO)	5395	(1.58%)	6028	(1.35%)	7834	(1.06%)	8378	(0.88%)	27635	(1.11%)
Idaho(ID)	661	(0.19%)	1951	(0.44%)	13144	(1.77%)	10515	(1.10%)	26271	(1.06%)
Tennessee(TN)	3310	(0.97%)	4617	(1.03%)	6999	(0.94%)	7402	(0.78%)	22328	(0.90%)
Oklahoma(OK)	6034	(1.76%)	5674	(1.27%)	4764	(0.64%)	4719	(0.50%)	21191	(0.85%)
Utah(UT)	1876	(0.55%)	3411	(0.76%)	6389	(0.86%)	9068	(0.95%)	20744	(0.83%)
Iowa(IA)	3128	(0.91%)	3621	(0.81%)	5955	(0.80%)	7240	(0.76%)	19944	(0.80%)
South Carolina(SC)	2345	(0.68%)	3630	(0.81%)	4966	(0.67%)	5886	(0.62%)	16827	(0.68%)
Delaware(DE)	3088	(0.90%)	4396	(0.98%)	3994	(0.54%)	3967	(0.42%)	15445	(0.62%)
Louisiana(LA)	3041	(0.89%)	4141	(0.93%)	4234	(0.57%)	3006	(0.32%)	14422	(0.58%)
Kansas(KS)	2082	(0.61%)	2333	(0.52%)	3606	(0.49%)	6366	(0.67%)	14387	(0.58%)
Kentucky(KY)	2437	(0.71%)	2611	(0.58%)	3904	(0.53%)	4525	(0.48%)	13477	(0.54%)
Alabama(AL)	1857	(0.54%)	2677	(0.60%)	3428	(0.46%)	3578	(0.38%)	11540	(0.46%)
New Hampshire(NH)	1032	(0.30%)	2061	(0.46%)	3636	(0.49%)	4226	(0.44%)	10955	(0.44%)
Nevada(NV)	807	(0.24%)	1192	(0.27%)	3001	(0.40%)	5257	(0.55%)	10257	(0.41%)
New Mexico(NM)	1034	(0.30%)	1941	(0.43%)	3180	(0.43%)	3497	(0.37%)	9652	(0.39%)
Vermont(VT)	441	(0.13%)	666	(0.15%)	2704	(0.36%)	3607	(0.38%)	7418	(0.30%)
Nebraska(NE)	759	(0.22%)	1392	(0.31%)	1934	(0.26%)	2268	(0.24%)	6353	(0.26%)
Rhode Island(RI)	843	(0.25%)	1125	(0.25%)	1956	(0.26%)	1929	(0.20%)	5853	(0.24%)
West Virginia(WV)	1297	(0.38%)	1366	(0.31%)	1301	(0.18%)	1040	(0.11%)	5004	(0.20%)
Arkansas(AR)	690	(0.20%)	1001	(0.22%)	1487	(0.20%)	1332	(0.14%)	4510	(0.18%)
Mississippi(MS)	582	(0.17%)	952	(0.21%)	1477	(0.20%)	1293	(0.14%)	4304	(0.17%)
Montana(MT)	432	(0.13%)	747	(0.17%)	1125	(0.15%)	979	(0.10%)	3283	(0.13%)
Maine(ME)	460	(0.13%)	674	(0.15%)	845	(0.11%)	1117	(0.12%)	3096	(0.12%)
Dist. of Columbia(DC)	488	(0.14%)	467	(0.10%)	611	(0.08%)	930	(0.10%)	2496	(0.10%)
North Dakota(ND)	318	(0.09%)	484	(0.11%)	670	(0.09%)	831	(0.09%)	2303	(0.09%)
Hawaii(HI)	296	(0.09%)	539	(0.12%)	594	(0.08%)	783	(0.08%)	2212	(0.09%)
South Dakota(SD)	293	(0.09%)	320	(0.07%)	568	(0.08%)	761	(0.08%)	1942	(0.08%)
Wyoming(WY)	299	(0.09%)	371	(0.08%)	497	(0.07%)	691	(0.07%)	1858	(0.07%)
Alaska(AK)	145	(0.04%)	270	(0.06%)	384	(0.05%)	278	(0.03%)	1077	(0.04%)

Notes: The number in parentheses represents the percentage of patents from the state relative to the total number of US patents.

Table 2. Citation Statistics

year	percent receiving citations	total number of citations	mean citations received
1976	0.79 (0.76)	149843 (131263)	3.40 (2.98)
1986	0.89 (0.87)	253989 (229690)	6.66 (6.02)
1996	0.95 (0.94)	1008675 (928693)	16.38 (15.08)
2006	0.84 (0.80)	810919 (684711)	10.07 (8.51)

Notes: The numbers in parentheses indicate values excluding self-citations.

patents do not overlap across different cohorts. We found 149843 citing patents for the 1976 cohort, 253989 for the 1986 cohort, 1008675 for the 1996 cohort, and 810919 for the 2006 cohort.

Table 2 summarizes some descriptive statistics about citations made to our originating patent cohorts, excluding self-citations.[13] High proportions of patents received citations for all cohorts. The average citation numbers in recent cohorts are substantially larger than in the 1976 cohort.

Control Patents. Key to JTH's experimental design of knowledge spillovers is the construction of a set of "control patents" for each sample of citing patents to mimic the existing geographic distribution of knowledge production. Geographic match of patents may arise as a consequence of agglomeration of research activities in similar fields.

The patent classification system offers possible channels for selecting control patents. The basic idea is to pick, for each citing patent, another patent that (i) has similar application date and (ii) is classified under the same technology (sub-)class as the citing patent, as well as possibly the originating patent. Such a procedure would generate a sample of patents that mirror the sample of citing patents but do not cite the corresponding originating patents.

We consider the following four measures of technological proximity. The first control measure, which was originally used by JTH, finds a technology match at the level of "three-digit" class; the next three are the *disaggregated* controls introduced by TFK, with increasing level of disaggregation.

(EPO). There were total 67576 patents granted between June and December 2015 that cite the 2006 originating patents. This amounts to only a small fraction of all citing patents since 2006.

[13] A self-citation is defined as a citation from a citing patent whose assignee is the same as that of the corresponding originating patent.

A. [**3-digit**] A control patent has a technology subclass that matches the original classification of the citing patent at the three-digit level.[14]
B. [**Any**] A control patent has a technology subclass that matches the original classification of the citing patent in full.
C. [**Primary**] A control patent has original classification (a primary subclass) that matches the original classification of the citing patent.
D. [**Common**] A control patent has original classification that matches the original classification of the citing patent and a technology subclass that matches any subclass of the corresponding originating patent.

For each measure of technological proximity above, we picked a control patent randomly from all candidate patents whose application dates fell within one-month (30 days) on either side of the application date of the citing patent; if no admissible patent was found, we widened the window to 3 months (90 days) and then to 6 months (180 days). If no control patent was found after three such rounds, a null observation was returned.[15] Our selection procedure was implemented by Python algorithms.

4. Trends in Geographic Localization

4.1. **Methodology.** For each definition of geographic boundary, we test whether the frequency of geographic match (i.e. identical inventor location) between originating and citing patents is equal to or larger than the matching rate between originating and control patents. Formally, for given geographic boundary (country, state, or CMSA) and for given cohort, let p_{ij}^{citing} denote the matching probability between originating and citing patents in state i and industry sector j, and p_{ij}^{control} denote the matching probability between originating and control patents. We consider the 50 US states plus the District of Columbia and the 37 industrial sub-categories under NBER classification.[16]

The overall matching probability can be written as a weighted average of state-sector-level matching rates. This corresponds to

$$p^{\text{ citing}} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}^{\text{citing}} p_{ij}^{\text{citing}} \quad \text{and} \quad p^{\text{ control}} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}^{\text{control}} p_{ij}^{\text{control}},$$

[14]When two patents are said to match at the "three-digit" level, it means that both patents are given a subclass whose parent class (first one-, two-, or three-digit integer of the classification code) is identical.
[15]Appendix B presents the number of control patents found at each round of iteration for each pair of originating and citing patent samples.
[16]We employed NBER's mapping table to match each USPC code with an industrial (sub-)category.

where the weight w_{ij}^{citing} (w_{ij}^{control}) is the number of citing (control) patents in state i and sector j divided by the total number of such patents, and I and J are the total numbers of states and sectors, respectively. We can also define the matching rates for each state i, p_i^{citing} and p_i^{control}, and for each sector j, p_j^{citing} and p_j^{control}.

As in JTH and TFK, we are primarily concerned with the difference $p^{\text{ citing}} - p^{\text{ control}}$ in the two matching rates, which will be referred to simply as the *localization effect* (of knowledge spillovers). We test $H_0 : p^{\text{ citing}} = p^{\text{ control}}$ versus $H_1 : p^{\text{ citing}} > p^{\text{ control}}$ for each cohort and for each definition of geographic boundary. The test statistic used in the paper is

$$t = \frac{\widehat{p}^{\text{ citing}} - \widehat{p}^{\text{ control}}}{[\text{SE}(\widehat{p}^{\text{ citing}})^2 + \text{SE}(\widehat{p}^{\text{ control}})^2]^{1/2}},$$

where

$$\widehat{p}^{\text{ citing}} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}^{\text{citing}}\widehat{p}_{ij}^{\text{citing}},$$

$$\text{SE}(\widehat{p}^{\text{ citing}}) = \left[\sum_{i=1}^{I}\sum_{j=1}^{J}\left(w_{ij}^{\text{citing}}\right)^2\left(\widehat{p}_{ij}^{\text{citing}} - \widehat{p}^{\text{ citing}}\right)^2\right]^{1/2},$$

and $\widehat{p}_{ij}^{\text{citing}}$ is the sample proportion of p_{ij}^{citing}. We similarly define $\widehat{p}^{\text{ control}}$ and $\text{SE}(\widehat{p}^{\text{control}})$.

Our statistical analysis is conducted at the state-sector level, and this differs from JTH and TFK who treat all individual patents as independent and identically distributed. The key advantage of our group level analysis is that, by doing so, we maintain the effective sample size fixed, at $I \times J$, throughout the cohorts. Replicating the individual level analysis over time could potentially suffer from the effects of increasing sample size. The numbers of our sample patents in 1996 and 2006 are far greater than the corresponding number in 1976. Note also that the clustered standard errors allow for arbitrary dependence within each group.

4.2. **Aggregate Trends.** In this section, we report the aggregate citing and control matching rates across cohorts. Table 3 presents these findings, together with t-values for the hypothesis testing.[17]

We begin by summarizing our results for the 1976 cohort of patents.

[17]Notice that the sample sizes for citing patents in Table 3 differ from the corresponding numbers appearing in Table 2. For the calculation of citing matching rates, our sample citing patents are taken to be those that allow us to find corresponding control patents according to the 3-digit criterion.

Table 3. Frequency of Geographic Match

		citing	3-digit	Any	Primary	Common
1976	TOTAL	104127	104127	97356	81090	34059
	country	66.35	57.78	59.84	59.23	61.34
			(15.49)	(10.87)	(11.43)	(6.38)
	state	9.57	4.68	6.55	6.85	8.71
			(9.73)	(5.71)	(4.87)	(1.31)
	CMSA	8.07	3.47	5.27	5.53	7.34
			(11.7)	(6.72)	(6.01)	(1.36)
1986	TOTAL	185213	185213	176372	153062	67993
	country	71.21	56.62	59.02	58.39	58.48
			(22.02)	(17.67)	(17.24)	(14.84)
	state	10.68	4.72	6.41	6.63	7.62
			(8.8)	(5.93)	(5.49)	(3.88)
	CMSA	8.71	3.4	4.95	5.08	5.92
			(12.56)	(8.35)	(8.19)	(6.03)
1996	TOTAL	709662	709662	700537	656061	236091
	country	76.95	55.18	57.92	58.01	58.1
			(24.24)	(19.54)	(18.8)	(13.96)
	state	15.01	6.7	8.59	8.93	10.73
			(4.72)	(3.49)	(3.25)	(2.12)
	CMSA	11.88	4.5	6.23	6.5	8.06
			(7.05)	(5.14)	(4.83)	(3.26)
2006	TOTAL	551994	551994	547432	525909	236784
	country	77.96	52.84	56.07	56	58.08
			(20.52)	(16.97)	(16.56)	(13.64)
	state	18.31	8.05	10.23	10.47	12.53
			(4.41)	(3.4)	(3.29)	(2.36)
	CMSA	14.06	5.37	7.2	7.4	9.35
			(6.18)	(4.77)	(4.63)	(3.14)

Notes: The numbers in the first row of each cohort represent sample sizes. A number in parenthesis is the relevant t-statistic.

Finding 1. *Localization effects of knowledge spillovers in 1976-1985 are sizable at all location and control levels, except at intra-national (state and CMSA) levels under the most disaggregated level of control.*

In the 1976 cohort of patents, the geographic matching rates between originating and citing patents are considerably higher than the corresponding rates between originating and control patents at all geographic levels (country, state, and CMSA) and for all control measures, except at the two intra-national levels under the most disaggregated control group (Common).[18] These results are consistent with the main findings of TFK: even with large samples, using finer selection criteria increases the matching rate of control patents to the extent that the sizable localization effect disappears altogether (its magnitude is less than 1 percentage point) when we control for technological proximity across all originating-citing-control triads.[19]

Next, considering the trends of geographic localization since 1976, we first observe that the sample size of control patents has increased dramatically. The surge took place most notably between 1986 and 1996, with the numbers tailing off somewhat in 2006. Note that TFK had only 2122 control patents to work with in producing their main result; the corresponding figures for our 1996 and 2006 cohorts are, respectively, 236091 and 236784.

The first trend that we observe is on the citing matching rate.

Finding 2. *The frequency of geographic match between originating and citing patents has increased.*

The matching rate of citing patents has increased at every geographic level and from each decade to the next. Between 1976 and 2006, the gain is about 12% at country level and about 6% at CMSA level; at state level, the matching rate almost doubled from 9.57% to 18.31%. This finding contradicts the widespread belief that geographic proximity has been made less important for the flow of ideas by the advent of internet and other new communication technologies. According to our data, distance still matters, and today it matters even more than before, when one considers diffusion of ideas through patent citations.

We next report the trend of control matching rates.

Finding 3. *The frequency of geographic match between originating and control patents has increased at intra-national levels but decreased at international level.*

Within each cohort, and for each definition of geographic boundary, the matching rate of control patents increases with the level of disaggregation. This is consistent with the view that producers with similar technologies are more likely to agglomerate.

Across cohorts, the control matching rates fell in almost all cases between 1976 and 1986, but they then trended upward at the two intra-national levels. For example, under

[18]The matching rates in our sample are generally higher than those reported by TFK. Other than the sample size, one possible reason for this departure is that we consider citations that accrue only for 10 years up to 1985; TFK consider citing patents up to April 2001. The agglomeration effects of both production and diffusion of ideas may decay over time. For related evidence, see Jaffe and Trajtenberg (1999) and Thompson (2006).

[19]The t-statistics are 1.31 (state) and 1.36 (CMSA), which are substantially smaller than those under less disaggregated levels of control. Note that 95% critical value here is 1.645.

the Common criterion, the control matching rate in 1976 was roughly 9% at state level and 7% at CMSA level; the corresponding figures in 2006 were 13% and 9%, respectively.

Interestingly, however, the same trend is not observed at country level: the frequency of control and originating patents simultaneously being domestic dropped monotonically for all measures of control. This suggests that production of knowledge has become increasingly co-located within the US, while the opposite may have been happening across international borders.

Our main results on the trend of localization effects are now summarized.

Finding 4. *Localization effects are substantial and highly significant at all location and control levels in all cohorts of patents since 1986.*

Importantly, we observe significant localization effects in every cohort and for every control measure since the 1986 cohort. This includes even the most disaggregated level of control selection, for which localization effects were not found in the 1976 cohort. The strength of localization effects is also substantial and highly significant (well above the 95% critical value). Despite the intensification of pre-existing geographic distribution of patent production, the increase in localization of citations has indeed been the dominating force.

Finding 5. *Localization effects have strengthened.*

Moreover, localization of knowledge spillovers has strengthened over the decades. Table 4 presents the extent of localization effects in proportional terms. At every geographic level, the difference between citation and control matching rates is greater in 2006 than in 1976, regardless of the selected controls.

This trend appears to be more profound for the cases that had relatively low levels of localization to begin with. Considering the country-level effects, the citing patents were about 13% more localized according to 3-digit controls and 8% more localized according to the most disaggregated controls than the control patents in the 1976 cohort; these figures rose to 32% and 26%, respectively, in the 2006 cohort. When controls were selected under the most stringent criteria, intra-national localization effects leaped from only about 9% in 1976 to over 30% in 2006 at both state and CMSA levels.

4.3. **Comparison by State.** Our previous findings on the patterns of knowledge spillovers treat all locations identically. We next explore possible heterogeneity in localization effects across states. As can be seen from Table 1 above, over 60% of all utility patents granted to domestic inventors across the sample period were concentrated in less

Table 4. The Degree of Localization Effects

		3-digit	Any	Primary	Common
1976	country	12.91%	9.81%	10.72%	7.55%
	state	51.08%	31.53%	28.42%	8.91%
	CMSA	57.05%	34.72%	31.49%	9.04%
1986	country	20.49%	17.12%	18.0%	17.87%
	state	55.81%	39.98%	37.94%	28.66%
	CMSA	60.92%	43.21%	41.65%	32.01%
1996	country	28.29%	24.72%	24.61%	24.5%
	state	55.34%	42.75%	40.49%	28.54%
	CMSA	62.11%	47.57%	45.31%	32.19%
2006	country	32.22%	28.08%	28.17%	25.51%
	state	56.03%	44.15%	42.82%	31.59%
	CMSA	61.81%	48.77%	47.35%	33.49%

than 10 states; furthermore, Californian inventors have been by far the most prolific, and they have actually widened their lead in patent production.[20] This raises the question whether our results are driven by disproportionately large localization effects that have taken place in some states.

The observed localization effects across states are summarized in Table 5. For each cohort, we first report the frequency of patents that cite patents originating from a given state and are themselves from the state; we next report the matching rate of control patents selected according to the most disaggregated procedure (Common). The results are also illustrated in Figure 1. In each graph, a point indicates the pair of matching rates for a given state; also, the points vary in size, reflecting their corresponding sample size (as a proportion of the total). The dotted line in each graph represents equal matching rates so that the vertical distance above this line measures localization effect.

For the 1976 cohort, we do not observe substantial differences between the two matching rates for most of the states, similarly to the state-level findings from the aggregate sample. The 1986 cohort displays stronger localization effects across most states. The differences are large in many states including California, New York, Illinois, Minnesota, and Michigan.

An interesting trend that followed concerns the distribution of observations. Through the 1996 and 2006 cohorts, both citing and control matching rates became substantially

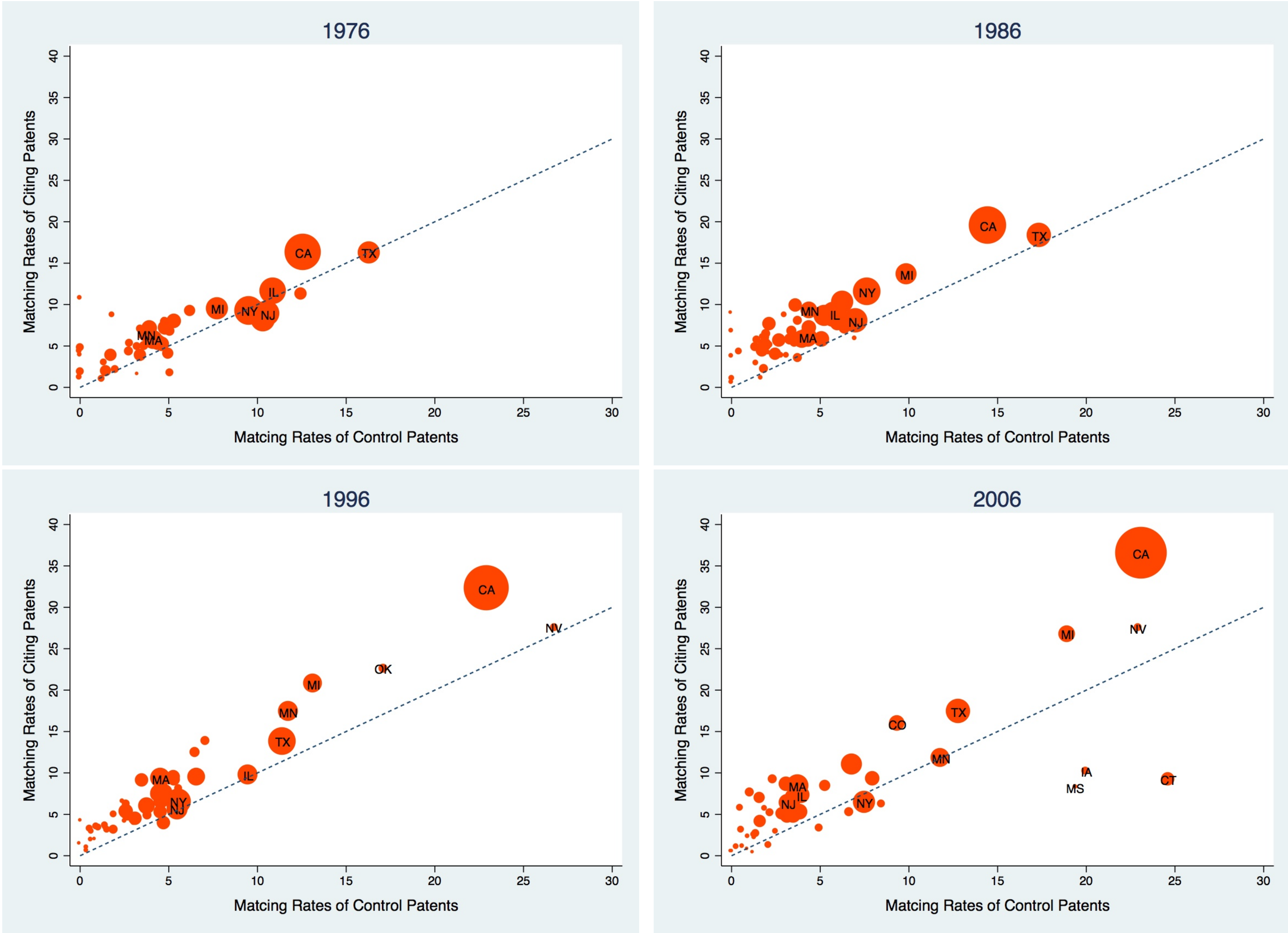
[20]Similar state-wide patterns are observed in the distribution of each of the sample (i.e. originating, citing, and control) patents.

more dispersed across states. This trend was led by a handful of states, including California, Michigan, Nevada, and Texas. In the 2006 cohort, we also observe a small number of states with large control matching rates that far exceed citing matching rates.

TABLE 5. Matching Rates by State

State	1976			1986			1996			2006		
	Citing	Control	t-value	Citing	Control	t-value	Citing	Control	t-value	Citing	Control	t-value
California(CA)	16.24	12.59	(4.03)	19.49	14.47	(3.85)	32.20	22.93	(3.04)	36.52	23.09	(6.81)
New York(NY)	9.23	9.56	(-0.24)	11.53	7.65	(1.81)	6.55	5.54	(1.16)	6.41	7.51	(-0.53)
Texas(TX)	16.27	16.29	(-0.0)	18.33	17.37	(0.15)	13.81	11.43	(0.52)	17.41	12.81	(0.67)
Illinois(IL)	11.56	10.9	(0.34)	8.79	5.82	(2.9)	9.75	9.49	(0.07)	7.23	3.96	(2.16)
Michigan(MI)	9.53	7.76	(0.96)	13.65	9.88	(2.18)	20.71	13.17	(2.3)	26.75	18.95	(1.38)
New Jersey(NJ)	8.8	10.61	(-1.01)	7.95	7.01	(0.61)	5.60	5.48	(0.13)	6.28	3.21	(2.45)
Ohio(OH)	9.23	9.72	(-0.42)	10.28	6.30	(4.66)	9.40	6.61	(2.79)	9.20	7.97	(0.61)
Pennsylvania(PA)	8.15	10.34	(-1.37)	8.56	5.24	(2.81)	7.49	4.47	(2.02)	6.35	3.13	(1.93)
Massachusetts(MA)	5.73	4.14	(1.36)	5.99	4.31	(2.71)	9.27	4.56	(3.75)	8.41	3.73	(3.22)
Minnesota(MN)	6.37	3.74	(1.95)	9.22	4.42	(3.47)	17.35	11.75	(1.52)	11.79	11.82	(-0.01)
Washington(WA)	9.23	6.23	(1.28)	7.57	2.18	(4.75)	6.45	4.62	(1.78)	10.98	6.81	(4.96)
Florida(FL)	7.92	5.35	(1.7)	5.67	4.00	(2.75)	5.97	3.79	(2.49)	5.26	3.88	(1.09)
North Carolina(NC)	7.04	4.91	(0.9)	3.93	2.52	(1.83)	9.03	3.53	(2.7)	4.91	3.49	(1.18)
Colorado(CO)	5.43	4.48	(0.57)	9.90	3.64	(3.14)	7.68	4.78	(2.58)	15.88	9.35	(0.72)
Wisconsin(WI)	7.16	4.81	(2.06)	7.59	6.00	(1.17)	9.12	5.36	(1.95)	7.20	3.43	(2.15)
Indiana(IN)	5.39	4.45	(0.85)	7.20	4.40	(2.13)	7.39	4.36	(3.57)	8.45	5.26	(0.97)
Arizona(AZ)	4.1	5.02	(-0.72)	5.10	1.94	(3.59)	5.26	2.65	(2.22)	5.04	2.88	(1.55)
Connecticut(CT)	7.11	3.92	(1.39)	5.78	5.11	(0.42)	5.31	4.52	(0.38)	9.20	24.64	(-1.28)
Maryland(MD)	5.17	4.66	(0.33)	5.59	2.72	(3.08)	4.41	3.13	(1.31)	4.17	1.61	(3.19)
Oregon(OR)	4.91	3.24	(0.9)	6.67	3.41	(1.57)	3.88	4.72	(-0.5)	4.77	3.18	(1.07)
Georgia(GA)	7.87	4.8	(1.5)	5.77	3.31	(1.6)	9.45	5.25	(2.76)	8.62	3.10	(2.57)
Virginia(VA)	3.77	1.75	(2.16)	4.38	1.73	(2.67)	4.82	2.70	(1.71)	6.89	1.59	(5.56)
Missouri(MO)	3.88	3.43	(0.23)	6.27	4.11	(0.97)	5.51	2.52	(2.65)	7.61	1.06	(4.28)
Idaho(ID)	8.16	4.76	(0.64)	8.71	2.99	(2.42)	13.83	7.10	(1.54)	6.53	3.60	(2.15)
Tennessee(TN)	4.21	2.81	(0.63)	6.34	1.98	(3.3)	4.80	3.81	(0.78)	5.24	6.66	(-0.72)
Oklahoma(OK)	11.2	12.47	(-0.29)	7.02	6.38	(0.35)	22.59	17.09	(0.56)	6.27	8.47	(-0.53)
Utah(UT)	4.81	0	(3.81)	7.94	3.80	(2.06)	12.43	6.51	(2.57)	9.13	2.35	(4.12)
Iowa(IA)	6.75	5.11	(0.92)	4.76	1.34	(2.69)	6.23	2.64	(2.56)	10.15	20.01	(-1.3)
South Carolina(SC)	6.94	3.39	(0.9)	5.84	4.32	(0.78)	8.10	5.55	(1.39)	3.30	4.93	(-1.08)
Delaware(DE)	1.9	1.48	(0.47)	3.54	3.73	(-0.11)	5.60	2.54	(2.69)	7.01	1.74	(1.96)
Louisiana(LA)	4.95	3.66	(0.73)	5.41	3.55	(1.37)	4.90	1.89	(1.99)	5.64	1.86	(1.46)
Kansas(KS)	5.3	2.78	(1.12)	4.31	0.45	(2.87)	3.10	1.55	(2.39)	5.20	2.20	(2.41)
Kentucky(KY)	2.11	2.03	(0.08)	2.25	1.81	(0.48)	3.65	1.43	(1.85)	2.50	1.25	(1.12)
Alabama(AL)	1.77	5.08	(-1.44)	5.97	1.93	(1.93)	3.20	0.55	(3.57)	5.73	0.45	(1.9)
New Hampshire(NH)	0.99	1.2	(-0.19)	5.61	1.45	(1.71)	3.18	1.89	(1.3)	2.67	1.37	(0.91)
Nevada(NV)	8.7	1.79	(1.68)	6.08	1.77	(1.25)	27.60	26.72	(0.06)	27.45	22.93	(0.43)
New Mexico(NM)	3.04	1.33	(0.95)	4.66	1.84	(2.07)	3.50	0.94	(3.01)	3.13	0.52	(2.41)
Vermont(VT)	0	0	(-)	2.83	1.39	(1.07)	3.40	1.03	(1.56)	1.27	2.07	(-1.3)
Nebraska(NE)	-	-	(-)	5.12	2.15	(1.48)	4.04	3.03	(0.55)	2.23	1.32	(0.63)
Rhode Island(RI)	1.21	0	(2.01)	3.88	3.12	(0.32)	2.91	0.63	(1.99)	2.90	2.46	(0.19)
West Virginia(WV)	1.88	0	(1.68)	3.77	2.82	(0.46)	4.12	2.50	(1.28)	61.22	1.02	(2.73)
Arkansas(AR)	3.82	1.75	(0.69)	1.11	0.00	(1.28)	3.42	1.41	(1.27)	2.41	0.89	(0.88)
Mississippi(MS)	1.72	0	(1.55)	6.83	0.00	(2.47)	1.95	0.60	(2.18)	8.22	19.38	(-0.8)
Montana(MT)	4.62	0	(1.59)	5.97	6.98	(-0.19)	6.58	2.43	(0.94)	1.16	0.58	(0.73)
Maine(ME)	3.9	0	(3.03)	4.15	2.08	(0.73)	0.62	0.35	(0.55)	0.39	1.18	(-0.65)
Dist. of Columbia(DC)	4.31	0	(2.06)	0.62	0.00	(1.46)	0.92	0.38	(0.75)	1.14	0.27	(1.43)
North Dakota(ND)	1.59	0	(1.23)	3.73	0.00	(2.66)	2.03	0.81	(0.8)	0.47	0.00	(0.91)
Hawaii(HI)	0	0	(-)	2.30	0.00	(1.56)	4.31	0.00	(2.47)	5.45	0.00	(1.34)
South Dakota(SD)	1.59	3.23	(-0.44)	8.93	0.00	(2.7)	1.44	0.00	(1.68)	0.84	0.90	(-0.07)
Wyoming(WY)	10.74	0	(2.78)	1.17	1.64	(-0.3)	4.20	0.00	(1.95)	0.54	0.00	(1.11)
Alaska(AK)	11.76	0	(1.6)	3.85	0.00	(1.29)	0.88	0.00	(0.91)	2.27	0.00	(1.06)

Figure 1. Matching Rates by State



Tables A2 and A3 in Appendix C report detailed breakdown of matching rates for California and the rest of the US, respectively. While Californian inventors have been the key driving force behind greater localization of economic activities reflected in patents, our central findings are also observed for the rest of the country. Albeit in smaller scale, localization effects of knowledge spillovers have strengthened across the US without California.

4.4. **Comparison by Industry.** The results from the aggregate sample of Section 4.2 may contain other types of heterogeneity. Since some states have played a particularly important role in reinforcing the localization effects, and since states often specialize in agglomeration of certain types of industries (e.g. Silicon Valley), it is worth checking the geographic patterns of patent citations across different industries. Another reason to break down localization effects by industry is to explore a potential source of divergence in the "home bias" in localization of patent production.

We report the localization trends in terms of NBER's six industrial categories under which the 37 sub-categories are nested: chemical, computer and communication, drugs and medicine, electronic, mechanical, and others. The detailed results (obtained with Common controls and for each geographic level) are given in Table 6 and illustrated in Figures 2-4.

Let us first examine industry-wide localization trends at country level, where our aggregate analysis showed increasing localization of citations but diminishing localization of controls. Figure 2 clearly reveals growing localization effects across all industry categories. In each cohort, the magnitude of localization effects is relatively uniform; also the range of citing matching rates has remained relatively stable. Interestingly, however, the dispersion of control matching rates across industries has steadily widened over the sample decades. The fall in agglomeration of patent production in the "electronic" industry is particularly striking.

At intra-national level, localization effects have grown for all industries in almost all cases. The only exceptions are "mechanical" and "others" in the 2006 cohort at state and CMSA levels, where such effects are statistically insignificant. Again, the distribution of control matching rates has become considerably more scattered. Figures 3 and 4 show that this is mostly due to greater clustering of research activities in "drugs and medicine," "mechanical," and "others."

Table 6. Matching Rates by Industry

Location	Industry	1976 Citing	1976 Control	1976 t-value	1986 Citing	1986 Control	1986 t-value	1996 Citing	1996 Control	1996 t-value	2006 Citing	2006 Control	2006 t-value
country	Chemical	63.51	60.96	(2.07)	69.36	56.42	(10.0)	76.87	54.26	(18.47)	79.24	52.01	(17.96)
	Cmp&Cm	63.22	55.09	(4.39)	68.35	49.99	(8.66)	77.41	54.04	(17.54)	77.20	57.89	(16.26)
	Drgs&Me	74.40	67.04	(2.53)	80.33	70.80	(5.67)	83.61	70.28	(6.3)	87.97	73.18	(8.78)
	Elec	65.19	58.00	(4.8)	65.75	51.49	(14.32)	71.38	47.87	(24.09)	69.40	39.97	(21.02)
	Mech	63.03	56.10	(3.89)	67.73	54.32	(8.02)	71.25	50.67	(10.99)	78.54	60.50	(3.89)
	Others	72.85	69.10	(2.65)	76.17	65.69	(8.48)	78.53	63.57	(9.91)	80.36	63.53	(6.36)
state	Chemical	8.59	9.20	(-0.46)	10.29	7.69	(2.29)	15.01	9.01	(2.1)	17.95	10.06	(1.6)
	Cmp&Cm	8.81	7.83	(0.46)	9.36	6.45	(1.48)	14.32	10.09	(1.15)	16.73	10.97	(1.48)
	Drgs&Me	9.60	9.39	(0.12)	10.97	8.38	(0.91)	17.61	12.97	(0.72)	22.49	16.13	(0.88)
	Elec	9.06	8.21	(0.8)	10.59	7.38	(2.54)	13.36	8.67	(1.88)	17.50	9.11	(1.91)
	Mech	9.72	8.34	(1.47)	11.51	7.27	(4.35)	14.30	9.45	(2.55)	18.65	17.99	(0.14)
	Others	11.15	9.00	(1.13)	11.34	8.27	(1.28)	15.89	12.69	(0.95)	21.37	19.17	(0.31)
CMSA	Chemical	8.53	9.00	(-0.3)	9.23	7.05	(2.36)	12.64	7.18	(3.1)	13.45	7.28	(2.52)
	Cmp&Cm	7.06	5.86	(1.2)	7.49	4.99	(1.7)	11.04	7.62	(1.43)	12.84	8.03	(1.97)
	Drgs&Me	9.35	9.65	(-0.16)	8.84	5.87	(2.34)	13.39	8.96	(1.31)	15.73	11.26	(1.39)
	Elec	7.29	6.29	(1.55)	7.97	5.32	(3.59)	10.73	6.37	(2.25)	14.41	7.12	(1.91)
	Mech	7.48	6.42	(1.54)	9.60	6.03	(4.97)	11.75	7.93	(2.22)	14.88	14.38	(0.11)
	Others	8.86	7.03	(1.91)	9.07	6.02	(2.55)	12.99	10.23	(1.17)	17.35	16.15	(0.23)

Figure 2. Matching Rates by Industry (Country)

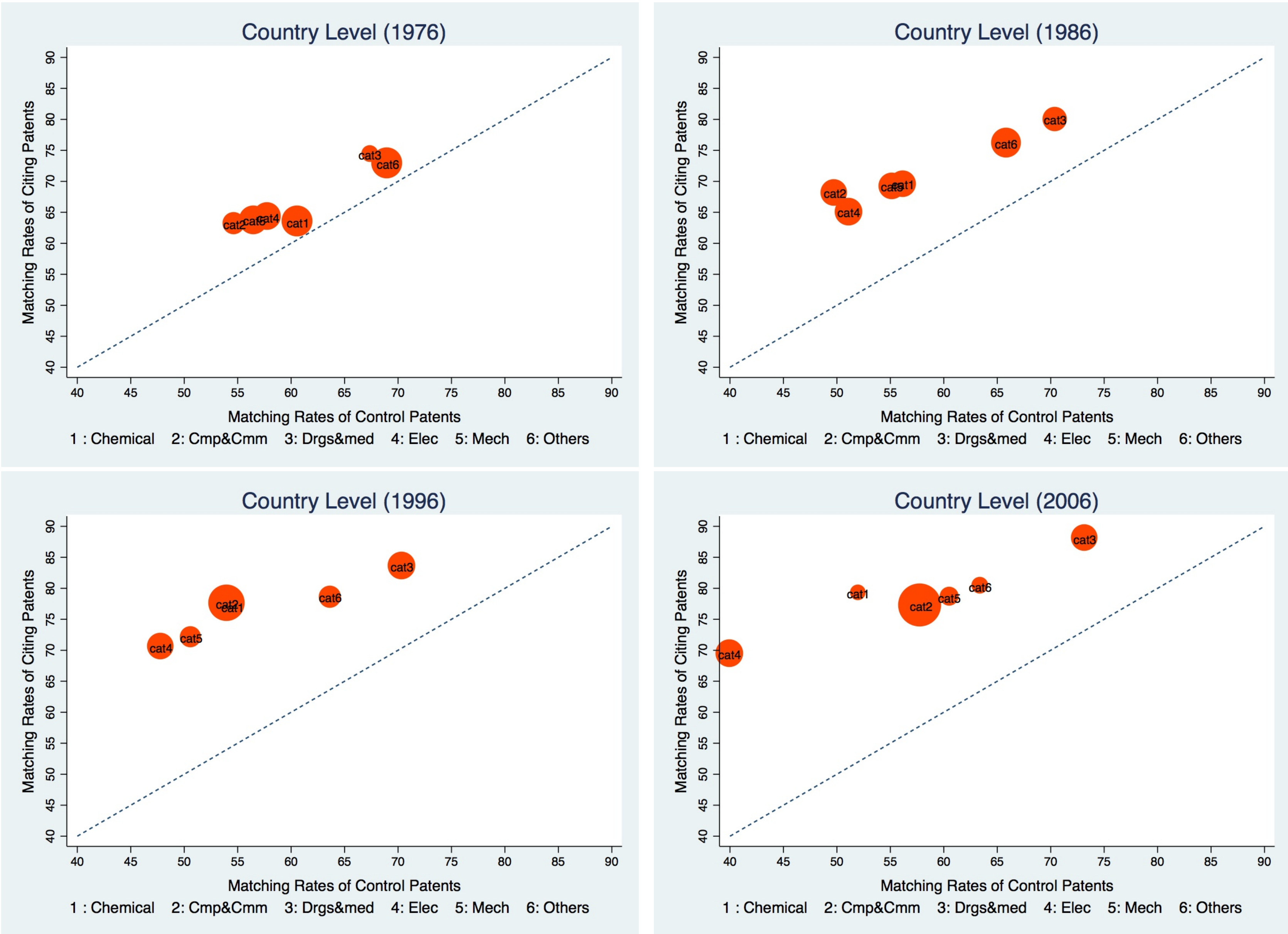


FIGURE 3. Matching Rates by Industry (State)

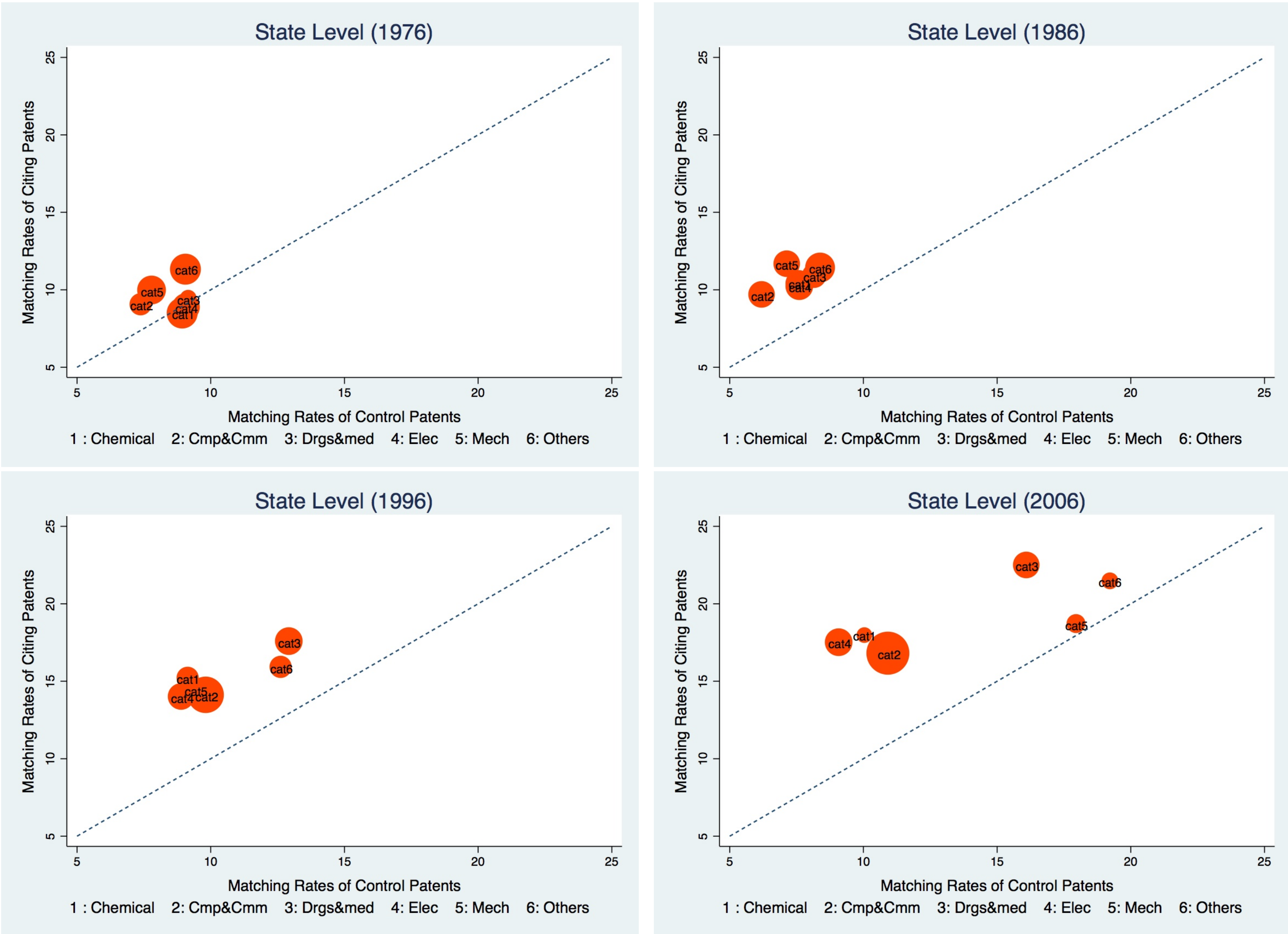
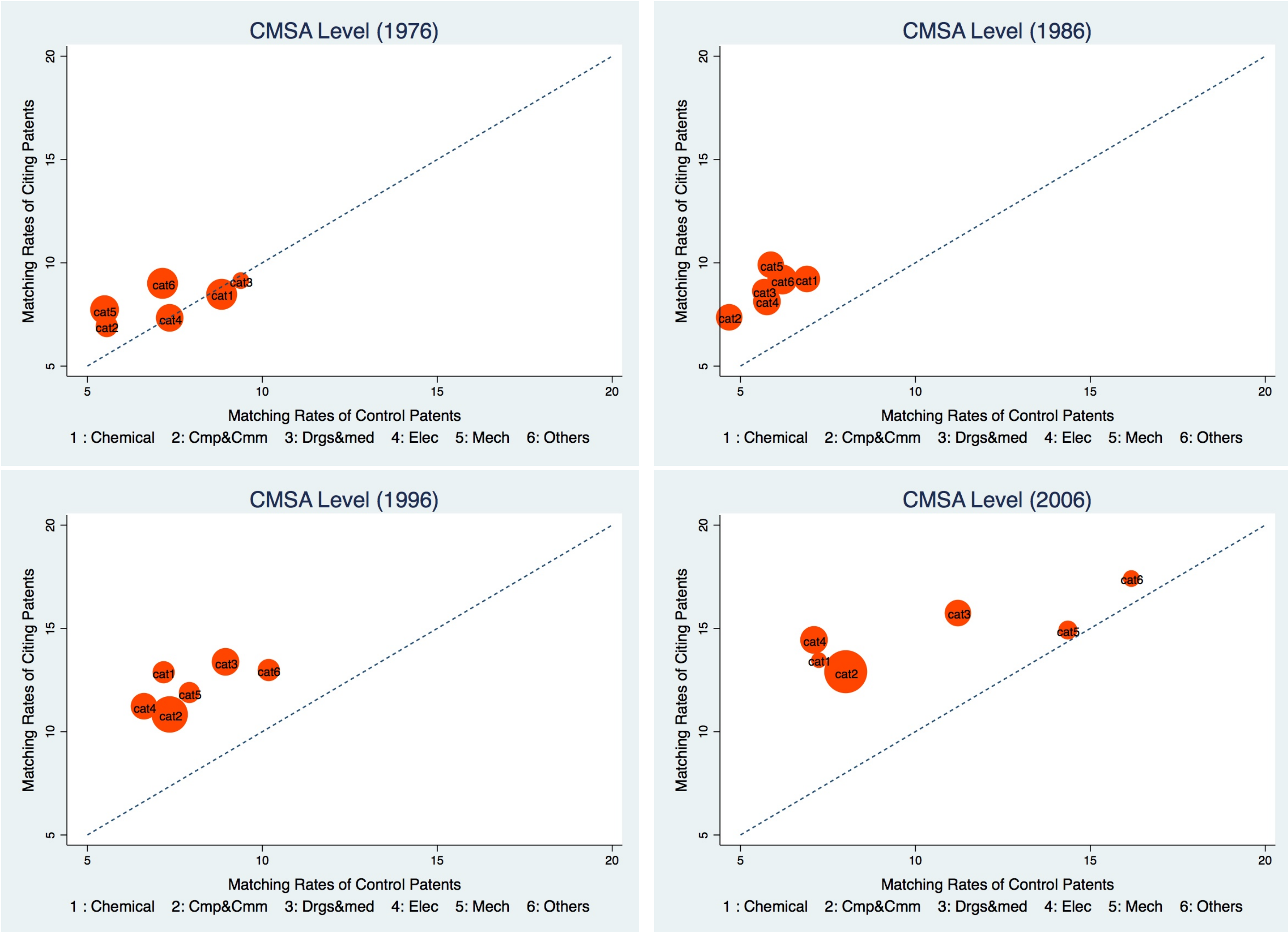


FIGURE 4. Matching Rates by Industry (CMSA)



Given the importance of California, we in addition break down Californian patents by industries and present the results in Figure A1 in Appendix C. The citing patents from California have increasingly become more localized than the corresponding control patents across all industries, except for "others" in 2006.

5. Explaining the Trends: The Role of Farness

5.1. **Farness Index.** Despite numerous claims of the "death of distance," our patent data reveal surprising trends in geographic patterns of innovation. While production of knowledge has become more concentrated, distance has become even more important for knowledge spillovers. How did this happen?

Our findings in Sections 4.3 and 4.4 revealed substantial, and growing, heterogeneity in the components of localization effects across states and industry sectors. We now attempt to explain these state-sector variations in localization effects by focusing on two potential mechanisms.

First, motivated by the leading trends of several key states, especially California, we consider the role of "natural advantages" behind agglomeration in the context of knowledge spillovers.[21] In particular, to think about natural advantages in terms of spatial effects on citation link formation, we view the map of US as a complete (undirected) network with each state as a node and invoke the notion of "closeness centrality" (Bavelas, 1950; Sabidussi, 1966). Closeness centrality of a node measures the average distance between the node and all other nodes.[22] Our idea is that the flow of knowledge may be more localized in states that are more isolated, or less central.

Second, the urban economics literature has emphasized the role of knowledge spillovers behind agglomeration of production (e.g. Marshall, 1890; Rosenthal and Strange, 2001; Ellison, Glaeser, and Kerr, 2010); the opposite forces may also be in play. As more innovators gather in close proximity, for example, there may be less related innovators to cite their work in distance and they themselves may not have enough time to pay attention to patents produced outside of their local networks.[23] There could also be feedback effects that reinforce the two-way channel between agglomeration and diffusion.

[21]For a discussion on natural advantages in agglomeration in general, see Ellison and Glaeser (1997, 1999) for instance.

[22]There are a number of variations to this definition of closeness centrality, as well as the broad notion of centrality itself. A recent paper by Bloch, Jackson, and Tebaldi (2016) offers a formal foundation of centrality measures in network theory.

[23]Lucas and Moll (2014) introduce an explicit time constraint for learning new ideas in an endogenous growth model. They do not however consider the potential effects of distance and network structure among productive individuals.

We develop an index based on the inverse of closeness centrality, or "farness," that incorporates the potential effect of agglomeration of research activities. Specifically, the (weighted) "farness index (F)" for state i, sector j, and cohort c is defined as

$$F_{ijc} := \sum_{k=1, k\neq i}^{I} \frac{d_{i,k,c}}{N_{k,j,c}+1},$$

where

- $d_{i,k,c}$ is the *road* distance between the largest population cities in states i and k in cohort c; and
- $N_{k,j,c}$ is the number of patents granted to state k in sector j and cohort c.

To calculate F for a given state i, sector j, and cohort c, we consider the road distance between i and every other state k which is then weighted by the inverse of the number of patents obtained by state k in sector j and cohort c. We expect the diffusion of ideas from state k to state i to depend, on the one hand, negatively on the geographic distance between the states and, on the other hand, positively on the size of relevant innovation activities in state k. Note that the road distances in our index may differ across cohorts due to demographic changes.

5.2. **Regression Model.** We set up a regression analysis to scrutinize causal implications of our farness index. The baseline model has the following form:

$$\begin{aligned} Y_{ijc} = \beta_0 + \beta_1' X_{ijc} + \sum_{i=1}^{I-1} \alpha_i \mathrm{State}_i + \sum_{j=1}^{J-1} \gamma_j \mathrm{Sector}_j + \sum_{c=1}^{C-1} \delta_j \mathrm{Cohort}_c \\ + \sum_{j=1}^{J-1}\sum_{c=1}^{C-1} \pi_{jc} \mathrm{Sector}_j \times \mathrm{Cohort}_c + U_{ijc}, \end{aligned} \tag{1}$$

where

- $Y_{ijc} = (p_{ijc}^{\text{citing}} - p_{ijc}^{\text{control}})$ is the dependent variable that measures the localization effect for state i, sector j, and cohort c;
- X_{ijc} is a vector of explanatory variables to be explained below;
- State_i, Sector_j, and Cohort_c are state, sector, and cohort dummies, respectively;
- $\mathrm{Sector}_j \times \mathrm{Cohort}_c$ is the interaction term between Sector_j and Cohort_c dummies to allow for cohort-specific sector fixed effects;
- U_{ijc} is the error term; and
- I, J, and C are the numbers of states, sectors, and cohorts, respectively.

In what follows, we run multiple regressions based on (1) with F_{ijc} as the main explanatory variable of interest. In addition, we consider interactions between farness and cohort dummies as well as lagged dependent variable (i.e. localization effect in the preceding cohort) and patent share.

Patent share for state i, sector j, and cohort c is the number of patents granted to state i and sector j divided by the total number of patents granted in cohort c. We use the entire universe of patents, and not just the originating patents, for constructing these measures. Patent shares are included to further reflect the potential link between localization of knowledge production and localization of spillovers.

5.3. **Instrumental Variable.** It is possible that F is endogenous since the unobserved factors affecting localization effect may also influence interstate road constructions, population and the volume of patents produced. To construct an instrument for F_{ijc}, we adapt the idea of Moretti (2004) and use the spatial distribution of land-grant universities established after the Morrill Acts of 1862 and 1890. Specifically, the "farness-in-research index (FIR)" for state i and sector j is defined as follows:

$$FIR_{ij} := \sum_{\ell=1}^{n} \frac{\tilde{d}_{i,\ell}}{N_{i(\ell),j,0} + 1},$$

where

- $\tilde{d}_{i\ell}$ is the *physical* distance between the largest population city in state i and land-grant university ℓ in year 1970;
- $i(\ell)$ refers to the state in which land-grant university ℓ is located; and
- $N_{i(\ell),j,0}$ is the number of patents granted to state $i(\ell)$ in sector j between 1963 and 1975.

Our assumption is that omitted variables from our regression model are independent of the locations of land-grant academic institutions created a century ago. Note that we invoke physical, instead of road, distance to avoid any related endogeneity issue. Road distance nonetheless may be a better measure to capture the spillover effects. Largest population cities and patent weights were chosen also from a pre-1976 period; for the latter, note that 1963 is the first year of NBER's patent data.

5.4. **Data.** Our regression analysis is based on the 48 contiguous US states plus the District of Columbia, thus excluding Alaska and Hawaii, and as before, considers the 37 industrial sub-categories under NBER classification and the four cohorts–1976, 1986, 1996 and 2006.

For dependent variables, we choose localization effects based on Primary controls (i.e. the differences between citing matching rates and control matching rates under Primary, which make up the corresponding numbers in Table 3). Our previous findings on the localization trends of knowledge spillovers do not depend on the control selection method. The Common criterion, while imposing the most stringent requirements, generates considerably fewer sample patents than the other criteria.[24]

In compiling farness index, the largest population city in each state i and cohort c is extracted from the National Historical Geographic Information System (NHGIS) for the year $c-6$ (e.g. 1970 for the 1976 cohort), while the largest population cities in FIR are taken to be the same as those for the 1976 cohort.[25] Land-grant university information is obtained from the United States Department of Agriculture (USDA), and both the road and physical distances via Google Maps. NBER provides patent information used for constructing FIR. The indices are first studentized.

In all our regressions below, we drop cells with fewer than 20 patents, cluster standard errors at the state-sector level, and weight each model by the share of the originating patents across state-sector cells in each cohort. Also, all regressions include a constant and a dummy for state, sector, cohort, and cohort-specific sector, as specified in (1).

5.5. **Estimation Results.** We begin by demonstrating how our instrument, FIR_{ij}, is related to the main explanatory variable, F_{ijc}, and the dependent variable, Y_{ijc}, respectively. Figure 5 shows plots of regressing F_{ijc} on FIR_{ij} (Panel A) and regressing Y_{ijc} on FIR_{ij} (Panel B).

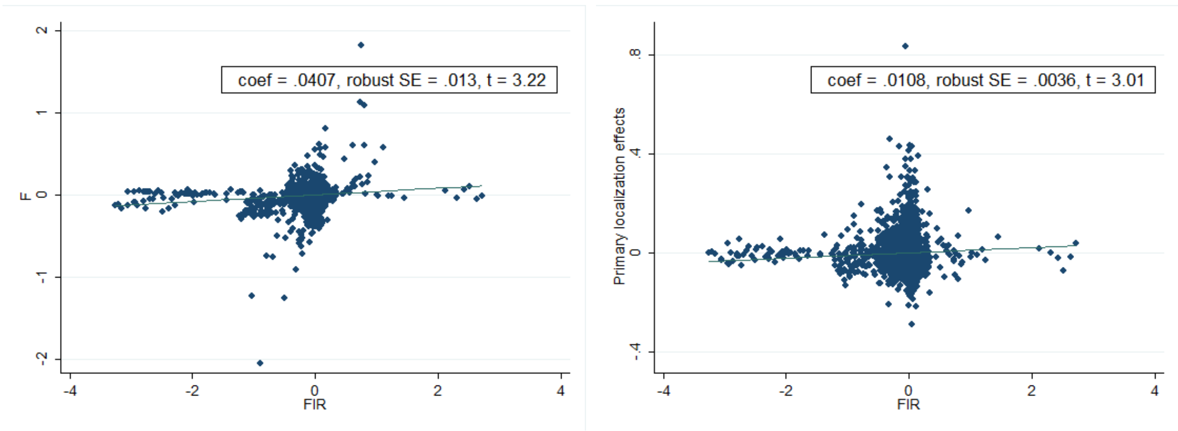
The instrument (FIR_{ij}) is positively associated with both the measure of farness (F_{ijc}) and localization effect (Y_{ijc}). The first-stage regression depicted in panel A shows that the variations in the farness measure are well captured by the instrument. An increase of one standard deviation in FIR corresponds to roughly 0.04 standard deviation increase in F (on average across cohorts), and an increase of one standard deviation in FIR increases localization effect by roughly 1 percentage point. These estimates are statistically highly significant.

Table 7 presents our main estimation results. In columns (1) to (4), we report our baseline OLS regression results, including additional controls on the interactions between

[24]All the regressions below are replicated with alternative measure of dependent variables using the Common criterion. Our central messages remain valid. See Appendix D.

[25]Metropolitan population information became available only from latter parts of the 20th century.

FIGURE 5. The Effects of Instrumental Variable



Panel A. 2SLS first-stage regression   Panel B. OLS reduced-form regression

Notes: The number of observations is 3660.

farness indices and cohorts. Columns (5)-(8) contain 2SLS regression estimates corresponding to columns (1)-(4). Here, when F_{ijc} is interacted with cohort dummies, we also include interactions between FIR_{ij} and cohort dummies.

Let us first summarize the results from OLS regressions. Comparing columns (1) with (2), we see that significant positive correlation between farness and localization effect appears only when we include lagged localization effect as explanatory variable. Moreover, the magnitude of correlation is increasing as we move closer to the present time. Introducing patent share does not alter these observations. The additional variable is in fact a strong predictor of localization effects, supporting our view that concentration of knowledge production may have served to strengthen localization effects of knowledge spread.

The overall picture remains much the same when we conduct 2SLS regressions to identify causal relationships but the coefficient estimates on farness variables are substantially greater. One standard deviation increase in F would induce an increase of 26 percentage points in the localization effect, which is now significant at the 1% level (column (5)); accounting for the effects of lagged dependent variable, the corresponding estimate is over 50 percentage point (column (6)). These are sizable impacts.

Breaking down the trending impact of farness, we confirm its growing importance over the sample period in columns (7) and (8). The final column reports estimates that indicate one standard deviation increase in F associated with roughly a 24 percentage point

Table 7. Estimation Results

	I. OLS and 2SLS estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
F	0.00159	0.0377*	-0.0411*	-0.0166	0.264***	0.505***	0.00472	0.0162
	(0.0137)	(0.0196)	(0.0228)	(0.0221)	(0.0574)	(0.161)	(0.0600)	(0.0575)
F x 1996			0.0601**	0.0601**			0.168***	0.130***
			(0.0274)	(0.0273)			(0.0396)	(0.0374)
F x 2006			0.191***	0.177***			0.483***	0.373***
			(0.0408)	(0.0400)			(0.123)	(0.109)
Lagged localization		0.123***	0.113***	0.101***		0.103***	0.0890***	0.0828***
		(0.0283)	(0.0270)	(0.0267)		(0.0335)	(0.0296)	(0.0286)
Patent share				1.087***				1.292***
				(0.318)				(0.370)
IV estimation	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3660	2848	2848	2848	3660	2848	2848	2848
R^2	0.395	0.425	0.438	0.445	0.285	0.196	0.378	0.418
	II. 2SLS first-stage F-statistics							
F					10.4	4.742	6.147	6.306
F x 1996							29.591	28.918
F x 2006							21.336	19.234

Notes: Robust standard errors are in parentheses. First-stage F-statistics refer to the significance of IV on each second stage explanatory variable under each of columns (5)-(8).

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

increase in the localization effect from the 1996 cohort to the 2006 cohort. These findings are consistent with the growing state- and industry-wide heterogeneity in matching rates observed and illustrated in Figures 1 and 3. IV estimation reduces the effects of lagged localization but strengthens those of patent share, relative to OLS results.

In sum, our regression analysis suggests that a state's spatial network characteristic and its relative volume of research activities in each sector together form an important determinant of the corresponding localization effect of knowledge spillovers. In terms of natural advantages, the index is meant to capture the role of distance in knowledge spillovers. From this perspective, our regression results reassert the surprising pattern: spatial proximity has become even more important for the flow of ideas just when IT revolution has reduced the costs of communication.

6. Concluding Remarks

This paper reports strong evidence of significant and *growing* localization effects of knowledge spillovers vis-à-vis knowledge production. Our results are surprising given the rapid globalization and development of communication technologies witnessed in recent decades. There is no doubt that information now travels at an unprecedented level of precision and speed. Patents and other scholarly publications are digitized and alerted around the world immediately upon publication. Why then has the "death of distance" not materialized?

To shed light on this latter question, we conduct regression analysis at the state-sector level to identify the effects of (i) natural advantages in localization of knowledge spillovers and (ii) relative concentration of knowledge production. In particular, we construct a novel network index based on the concept of farness, which turns out to be a significant and sizable determinant of the observed patterns of localization effects.

Still, there could be other reasons behind the trends. One potential explanation might be that the exponential growth in knowledge production has been accompanied by greater specialization, and the improvement in remote communication is simply not enough to convey the full scope and sophistication of latest research outputs. Researchers may be in need of close personal contact more than before. Another possibility is that distance matters less for higher quality research but this kind of ideas have become harder to come by (e.g. Bloom, Jones, Van Reenen, and Webb, 2017). Identifying the sources of greater localization of knowledge spillovers is the major outstanding question from the current study.

We wrap up by mentioning several directions to potentially enrich our analysis. First, instead of using discrete geographic boundaries, we could examine localization effects using continuous-distance metrics, as in Murata, Nakajima, Okamoto, and Tamura (2014). A related issue is to consider more sophisticated methods of constructing the control distribution of localization. Our regression analysis may also benefit from adopting the techniques of Murata, Nakajima, Okamoto, and Tamura (2014) since the aggregation issues apply to the farness index as well.

Appendix A. Sample Patents: Basic Selection Criteria

The USPTO bulk data contain some patents with typographical errors as well as missing information (e.g. grant and application date). We remove such patents in obtaining our sample patents. The following criteria are imposed on the sample selection procedure.

- Originating patent:
  (1) Has at least one US inventor, based on the location data before the CMSA mapping.
  (2) Has corporation or institution assignee distinct from inventor.
  (3) Is granted in 1976,1986,1996, or 2006.
- Citing patent:
  (1) Cites one of the originating patents defined above and is not self-citation.
  (2) Has application date within 10 years of each cohort (except for the 2006 cohort, for which citing patents granted up to May 2015 are included).
- Control patent:
  (1) Has corporation or institution assignee and CMSA information.
  (2) The corresponding citing patent cites an originating patent that has CMSA information, at least one US inventor, is assigned to a corporation or an institution, and has NBER class information.
  (3) The corresponding citing patent has corporation or institution assignee, CMSA information, and USPC class information.

Appendix B. Iteration Results for Control Selection

The table below shows the percentage of control patents selected in each round of iteration for each cohort and each technological match criterion. The final row in each cohort reports the proportions of citing patents for which control patents could not be found within our time frame.

Table A1. Iteration Results for Control Selection

	Class	3-digit	Any	Primary	Common
1976	1-month	99.93	66.46	41.32	16.33
	3-month	0.05	19.70	22.99	9.70
	6-month	0.01	7.16	13.22	6.48
	missing	0.01	6.68	22.47	67.49
1986	1-month	99.87	88.34	50.22	19.10
	3-month	0.11	5.90	21.37	10.54
	6-month	0.01	0.00	10.81	7.02
	missing	0.01	5.76	17.60	63.34
1996	1-month	99.98	95.59	69.87	19.30
	3-month	0.02	2.40	15.34	8.40
	6-month	0.00	0.00	6.43	5.56
	missing	0.00	2.01	8.36	66.74
2006	1-month	99.97	96.06	77.47	26.05
	3-month	0.02	2.18	12.07	9.66
	6-month	0.00	0.12	4.61	6.49
	missing	0.00	1.64	5.85	57.80

Appendix C. Comparison by State and Industry: California

Table A2. Frequency of Geographic Match: California

		citing	3-digit	Any	Primary	Common
1976	TOTAL	16190	16190	15136	12823	5202
	country	68.29	57.64	60.79	59.92	63.23
			(6.98)	(4.91)	(5.68)	(2.78)
	state	16.24	9.14	11.3	11.64	12.59
			(9.37)	(5.87)	(5.14)	(4.03)
	CMSA	9.38	3.68	5.34	5.69	6.59
			(11.61)	(7.48)	(6.99)	(4.58)
1986	TOTAL	31352	31352	29890	26512	11995
	country	72.49	57	59.55	59.26	59.72
			(6.98)	(5.7)	(5.2)	(4.28)
	state	19.49	10.61	13.08	13.53	14.47
			(7.74)	(5.09)	(4.67)	(3.85)
	CMSA	10.98	4.54	6.5	6.92	7.89
			(10.07)	(5.85)	(5.44)	(4.19)
1996	TOTAL	176073	176073	174567	166264	57989
	country	78.12	56.63	59.59	59.6	59.68
			(8.08)	(6.37)	(6.21)	(4.23)
	state	32.2	16.99	20.17	20.7	22.93
			(7.68)	(5.64)	(5.29)	(3.04)
	CMSA	21.58	9.51	12.26	12.64	13.98
			(8.15)	(6.13)	(5.91)	(3.95)
2006	TOTAL	177003	177003	176642	171202	77966
	country	78.16	52.6	55.44	55.48	56.92
			(7.64)	(6.46)	(6.34)	(5.57)
	state	36.52	17.67	20.91	21.02	23.09
			(10.14)	(8.41)	(8.44)	(6.81)
	CMSA	24.96	10.15	12.74	12.81	14.83
			(9.15)	(7.52)	(7.55)	(6.07)

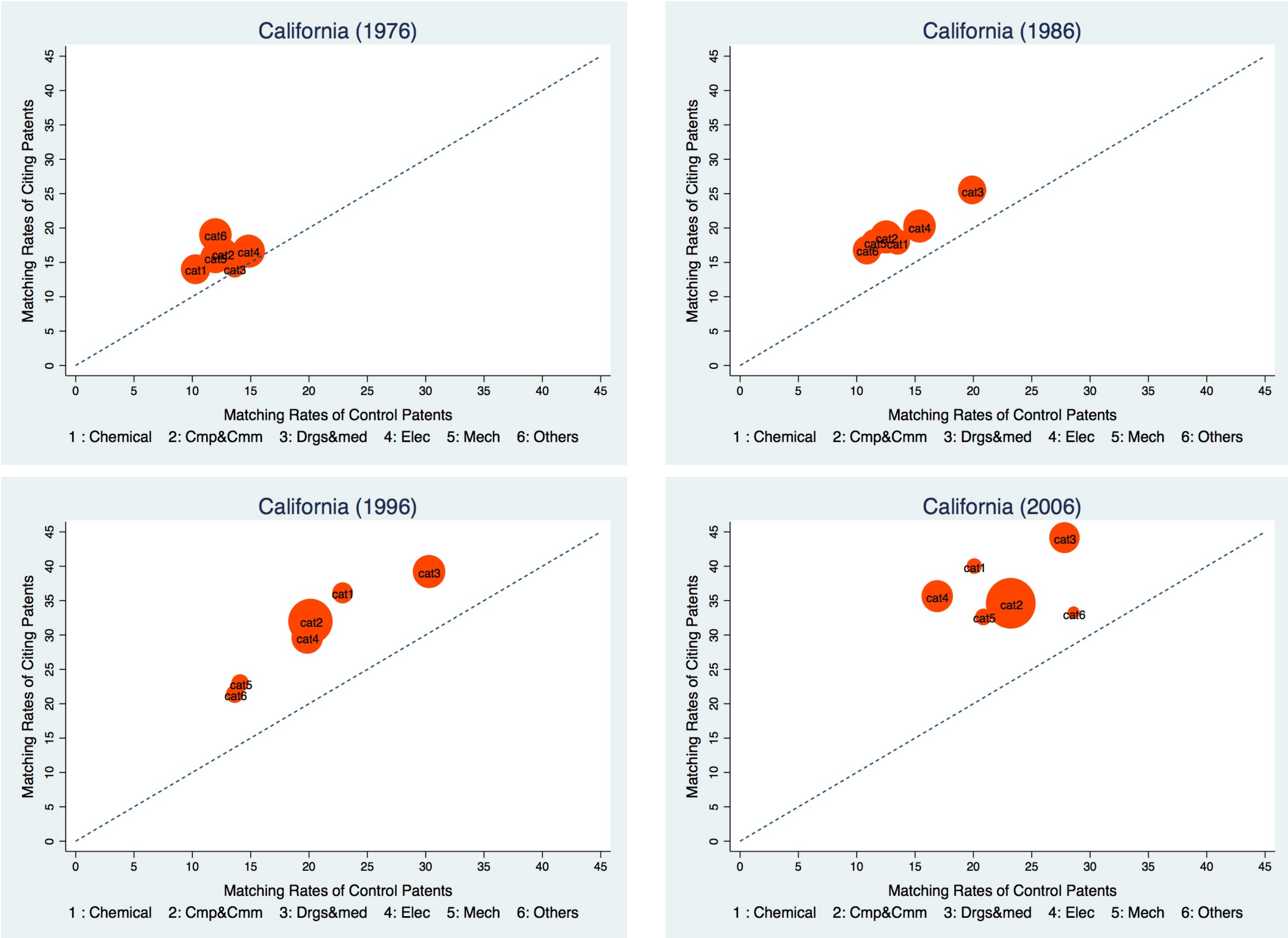
Notes: The numbers in the first row of each cohort represent sample sizes. A number in parenthesis is the relevant t-statistic.

TABLE A3. Frequency of Geographic Match: Without California

		citing	3-digit	Any	Primary	Common
1976	TOTAL	87937	87937	82220	68267	28857
	country	65.99	57.81	59.66	59.1	61
			(13.91)	(9.76)	(10.08)	(5.79)
	state	8.34	3.86	5.68	5.95	8.02
			(9.25)	(5.15)	(4.37)	(0.46)
	CMSA	7.83	3.43	5.26	5.5	7.48
			(9.57)	(5.29)	(4.71)	(0.57)
1986	TOTAL	153861	153861	146482	126550	55998
	country	70.95	56.54	58.91	58.21	58.22
			(22.36)	(17.82)	(17.98)	(15.84)
	state	8.88	3.52	5.05	5.18	6.15
			(8.99)	(5.98)	(5.72)	(3.72)
	CMSA	8.25	3.17	4.63	4.7	5.5
			(10.46)	(7.06)	(7.08)	(5.22)
1996	TOTAL	533589	533589	525970	489797	178102
	country	76.56	54.7	57.37	57.47	57.58
			(31.09)	(25.83)	(24.62)	(19.72)
	state	9.34	3.31	4.75	4.94	6.75
			(9.76)	(6.9)	(6.56)	(3.17)
	CMSA	8.68	2.85	4.23	4.41	6.13
			(11.82)	(8.35)	(7.93)	(3.79)
2006	TOTAL	374991	374991	370790	354707	158818
	country	77.87	52.95	56.38	56.26	58.64
			(28.47)	(23.07)	(22.22)	(16.63)
	state	9.72	3.51	5.14	5.38	7.34
			(6.68)	(4.58)	(4.33)	(1.94)
	CMSA	8.92	3.11	4.57	4.79	6.66
			(7.79)	(5.36)	(5.05)	(2.22)

Notes: The numbers in the first row of each cohort represent sample sizes. A number in parenthesis is the relevant t-statistic.

Figure A1. Matching Rates by Industry (California)



Appendix D. Additional Estimation Results

In this section, we replicate all the regressions conducted above with dependent variables defined with Common, instead of Primary, control selection. The results are presented in TableA4. The findings are broadly consistent with the corresponding estimation results under Primary controls but with somewhat lower magnitudes and/or statistical significance. Note that the number of observations here is lower due to the additional demand imposed by the Common criterion.

Table A4. Estimation with Common Controls

	I. OLS and 2SLS estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
F	0.00321	0.0417**	-0.0246	0.000205	0.277***	0.375***	-0.0525	-0.0425
	(0.0187)	(0.0196)	(0.0302)	(0.0297)	(0.0743)	(0.137)	(0.0716)	(0.0690)
F x 1996			0.0302	0.0292			0.101**	0.0730*
			(0.0329)	(0.0332)			(0.0420)	(0.0413)
F x 2006			0.168***	0.154***			0.417***	0.336***
			(0.0529)	(0.0518)			(0.130)	(0.124)
Lagged localization		0.0656**	0.0536**	0.0407		0.0547*	0.0340	0.0275
		(0.0258)	(0.0261)	(0.0263)		(0.0284)	(0.0295)	(0.0296)
Patent share				0.966***				0.924***
				(0.322)				(0.346)
IV estimation	No	No	No	No	Yes	Yes	Yes	Yes
Observations	2637	2110	2110	2110	2637	2110	2110	2110
R^2	0.388	0.387	0.397	0.402	0.301	0.284	0.372	0.391
	II. 2SLS first-stage F-statistics							
F					10.122	4.5	5.763	5.866
F x 1996							28.192	27.849
F x 2006							19.517	17.359

Notes: Robust standard errors are in parentheses. First-stage F-statistics refer to the significance of IV on each second stage explanatory variable under each of columns (5)-(8).

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

References

Almeida, P., and B. Kogut (1999): "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science*, 45(7), 905–917.

Bavelas, A. (1950): "Communication Patterns in Task–Oriented Groups," *Journal of the Acoustical Society of America*, 22(6), 725–730.

Bloch, F., M. Jackson, and P. Tebaldi (2016): "Centrality Measures in Networks," Working Paper.

Bloom, N., C. Jones, J. Van Reenen, and M. Webb (2017): "Are Ideas Getting Harder to Find?," Working Paper, Standford University.

Cairncross, F. (2001): *The Death of Distance: How the Communications Revolution Will Change Our Lives.* Boston: Harvard Business School Press.

Carlino, G., J. Carr, R. Hunt, and T. Smith (2012): "The Agglomeration of R&D Labs," Federal Reserve Bank of Philadelphia Working Paper, 12–22.

Carlino, G., and W. Kerr (2015): "Agglomeration and Innovation," *Handbook of Regional and Urban Economcis 5*, 349–404.

Coyle, D. (1997): "Economics: The Weightless Economy," *Critical Quarterly*, 39(4), 92–98.

Duranton, G., and H. Overman (2005): "Testing for Localization Using Micro-geographic Data," *Review of Economic Studies*, 72(4), 1077–1106.

Ellison, G., and E. Glaeser (1997): "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

——— (1999): "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *American Economic Review*, 89(2), 311–316.

Ellison, G., E. Glaeser, and W. Kerr (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195–1213.

Griffith, R., S. Lee, and J. Van Reenen (2011): "Is Distance Dying At Last? Falling Home Bias in Fixed-effects Models of Patent Citations," *Quantitative Economics*, 2(2), 211–249.

Hall, B., A. Jaffe, and M. Trajtenberg (2005): "Market Value and Patent Citations," *RAND Journal of Economics*, 36(1), 16–38.

Henderson, R., A. Jaffe, and M. Trajtenberg (2005): "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment," *American Economic*

Review, 95(1), 461–464.
Jaffe, A., and M. Trajtenberg (1999): "International Knowledge Flows: Evidence from Patent Citations," *Economics of Innovation and New Technology*, 8(1-2), 105–136.
Jaffe, A., M. Trajtenberg, and R. Henderson (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108(3), 577–598.
Keller, W. (2002): "Geographic Localization of International Technology Diffusion," *American Economic Review*, 92(1), 120–142.
Kerr, W., and S. Kominers (2015): "Agglomerative Forces and Cluster Shapes," *Review of Economics and Statistics*, 97(4), 877–899.
Kim, E. H., A. Morse, and L. Zingales (2009): "Are Elite Universities Losing Their Competitive Edge?," *Journal of Financial Economics*, 93(3), 353–381.
Kwon, S., J. Lee, and S. Lee (2017): "International Trends in Technological Progress: Evidence from Patent Citations, 1980–2011," *Economic Journal*, 127(605), F50–F70.
Lanjouw, J., and M. Schankerman (2004): "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *Economic Journal*, 114(495), 441–465.
Lucas, R., and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1), 1–51.
Lychagin, S., M. Slade, J. Pinkse, and J. Van Reenen (2016): "Spillovers in Space: Does Geography Matter?," *Journal of Industrial Economics*, 64(2), 295–335.
Marshall, A. (1890): *Principles of Economics*. London: Macmillan.
Moretti, E. (2004): "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data," *Journal of Econometrics*, 121(1–2), 175–212.
——— (2012): *The New Geography of Jobs*. New York: Houghton Mifflin Harcourt.
Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014): "Localized Knowledge Spillovers and Patent Citations: A Distance-based Approach," *Review of Economics and Statistics*, 96(5), 967–985.
Rosenthal, S., and W. Strange (2001): "The Determinants of Agglomeration," *Journal of Urban Economics*, 50(2), 191–229.
Sabidussi, G. (1966): "The Centrality Index of a Graph," *Psychometrika*, 31(4), 581–603.
Thompson, P. (2006): "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor-and Examiner-added Citations," *Review of Economics and*

Statistics, 88(2), 383–388.

THOMPSON, P., AND M. FOX-KEAN (2005): "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review*, 95(1), 450–460.

TRAJTENBERG, M. (1990): "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21(1), 172–187.