Title
Isolating Personal Knowledge Spillovers: Coinventor Deaths and Spatial Citation Differentials

Journal
American Economic Review Insights, 5(1)

ISSN
2640-205X

Authors
Balsmeier, Benjamin
Fleming, Lee
Lück, Sonja

Publication Date
2023-03-01

DOI
10.1257/aeri.20210275

AER: Insights 2023, 5(1): 21–34
https://doi.org/10.1257/aeri.20210275

Isolating Personal Knowledge Spillovers: Coinventor Deaths and Spatial Citation Differentials†

By Benjamin Balsmeier, Lee Fleming, and Sonja Lück*

We propose a new method to estimate and isolate the localization of knowledge spillovers due to the physical presence of a person, using after-application but pre-grant deaths of differently located coinventors of the same patent. The approach estimates the differences in local citations between the deceased and still-living inventors at increasingly distant radii. Patents receive 26 percent fewer citations from within a radius of 20 miles around the deceased, relative to still-living coinventors. Differences attenuate with time and distance, are stronger when still-living coinventors live farther from the deceased, and hold for a subsample of possibly premature deaths. (*JEL* O31, O33, O34, R32)

Marshall (1890) offered three (now canonical) explanations for the geographical agglomeration of economic activity: thicker labor markets, scale economies from collocation of production, and localized knowledge spillovers. These theories unfortunately imply similar observable outcomes (Ellison, Glaeser, and Kerr 2010), and empirical work has struggled to disentangle the mechanisms. Krugman (1991, 53) made the classic argument that the last mechanism in particular cannot be estimated as "knowledge flows … are invisible; they leave no paper trail by which they may be measured and tracked." In response, Jaffe, Trajtenberg, and Henderson (1993)—hereafter, JTH—offered a method which matched citing patents, the metric of patent citations as the paper trail, and the result that knowledge flows appear to be very localized.

The critiques of JTH (clearly acknowledged within the original paper) mainly focus on whether geographically localized citations indicate a real knowledge spillover or simply one correlate of the colocation of industrial and technological activity. To address this, JTH took an original sample, matched patents which cited the original sample with similar patents in date of application and technology area, and

* Balsmeier: University of Luxembourg (email: benjamin.balsmeier@uni.lu); Fleming: University of California, Berkeley (email: lfleming@berkeley.edu); Lück: Paderborn University (email: sonja.lueck@uni-paderborn.de).
Amy Finkelstein was coeditor for this article. The authors thank Guan-Cheng Li for invaluable research assistance and feedback from seminar participants at Duke University, UC Berkeley, ETH, and INSEAD, and, in particular, Sharon Belenzon, Katrin Hussinger, Gaétan de Rassenfosse, Maria Roche, Matteo Tranchero, and Martin Wörter. We gratefully acknowledge financial support from the Coleman Fung Institute for Engineering Leadership and the National Science Foundation (1360228). Errors and omissions remain ours.
† Go to https://doi.org/10.1257/aeri.20210275 to visit the article page for additional materials and author disclosure statement(s).

estimated differences in the two citing populations' geographic locations. The main concern with matching, laid out most sharply in Thompson and Fox-Kean (2005)—hereafter, TF-K—is that any result can still be attributed to unobserved differences in technology. A variety of analyses have demonstrated weaker (though often still significant) effects with more sophisticated matching (TF-K; Murata et al. 2014; Arora, Belenzon, and Lee 2018; for a review of the large literature, please see Jaffe and de Rassenfosse 2017).

Here we bypass matching altogether and instead use the death of a collaborative inventor to identify localized knowledge spillovers. Our approach uses a collaborative patent and identifies the difference in citations between the different regions that host the deceased versus still-living coauthors of the same patent. Though we still use patent citations as a measure of knowledge flows, the approach should avoid many of the criticisms of the citing patent matching method and the confounding of other influences with agglomeration (Duranton and Puga 2020). The approach builds upon recent literature that relies on death to identify the mechanisms of invention and scientific discovery (Azoulay, Zivin, and Wang 2010; Jaravel, Petkova, and Bell 2018).

The method complements prior work because it isolates the impact of the physical presence of a single inventor. Most approaches fail to identify the specific individual who is assumedly the actual source of the diffused knowledge. They typically assign one location to an entire patent, either based on one of the inventors' hometowns or a summary location based on multiple inventor locations. This makes inference difficult, as it inhibits controlling for unobservable characteristics of the patented technology. Comparing two or more inventors on the same patent, exactly one of whom has died, solves this issue. Assuming that inventor death remains exogenous to local factors of production and locally pooled labor, it enables cleaner estimation of the third Marshallian mechanism of knowledge flows, and, in particular, it enables us to establish the importance of physical presence for knowledge flows, both locally and at farther distances, and over time.

Findings of lower local citation rates for deceased versus still-living inventors point to the existence of local knowledge spillovers. For example, citations within 20 miles of an inventor are 25.8 percent lower for the deceased inventor, relative to still-living coinventors. The effects attenuate over distance and time, are stronger for more geographically distant coinventors, and hold for a variety of robustness checks, including a sample of possibly premature inventor deaths, examiner-only citations, and linear probability models (baseline estimations are robust Poisson models that exclude self-citations).

I. Data

The approach relies most crucially upon data for the geographical location of the following: an inventor who dies after applying for a patent but before that patent is granted, the still-living coinventors of that same patent, and inventors of the patents that cite that same patent.

Data collection starts with the population of all US patent inventors who appear on at least one patent issued by the US Patent and Trademark Office (USPTO), with application dates between 1976 and 2005, during which time inventor deaths appeared on the front page of the patent grant document. US inventors who died after application but before grant are often missing in secondary patent data sources but appear as originally published in the USPTO HTML files. The front page also provides the city and state for each inventor. As the original location data suffers from inconsistencies in location names and misspellings, we disambiguated all city-state combinations and used the Google Maps algorithm to identify remaining cases (e.g., some inventors list a neighborhood or unincorporated township). SimpleMaps (2020) provides latitude and longitude data for towns and locations.

We scraped all HTML data as described in Balsmeier et al. (2018) and kept only patents with (1) exactly one deceased inventor, (2) at least one living coinventor, (3) all living coinventors residing in a different city than the deceased inventor, (4) all inventors (deceased and coinventors) living in the United States, (5) an inventor identification number for all living and deceased inventors from the disambiguated USPTO PatentsView (2021) database, (6) calculable latitude and longitude locations, and (7) an application date that falls between January 1, 1976, and December 31, 2005. This created the full analysis sample of 5,491 (3,870 living and 1,621 deceased) inventors from a total of 1,621 patents with exactly one deceased inventor.

The number of inventors per patent (including the deceased) is skewed with most patents having two (41 percent), three (26 percent), or four inventors (14 percent) and the maximum of one patent with 18 inventors. Coinventors tend to live relatively close to the deceased inventor at a median distance of 25 miles and an average of 284 miles, though some inventors (13.2 percent) live more than 500 miles apart from the deceased. The number of patents applied for and granted per year ranges between 1 and 100, with higher numbers in the 1990s.

We then identified all citations from future granted US patents to each analysis sample patent up through 2020, as provided by the USPTO's PatentsView database. We identified a total of 34,749 citations to all patents in the analysis sample, implying 21.4 cites on average. Thirty-one percent of citations arise within 5 years, 59 percent within 10 years, and 80 percent within 15 years since patent grant. Since the last observed year of patent grant in the analysis sample is 2008, we observe at least a ten-year citation window for every patent (while the last application date in the analysis sample is 2005, there is typically a delay or "pendency" for applications to be granted as patents by the USPTO; hence, the last observed patent in the analysis sample was granted in 2008).

The econometric analyses rely on the geographic distances between the citing inventors and each of the deceased and still-living inventors of the deceased patent. Locations of all inventors on the citing patents were again disambiguated and longitude/latitude information added from SimpleMaps. For each inventor on a patent in the analysis sample, we take the distance to the closest inventor on the citing patent as the relevant distance of knowledge flow. From this we observe 15 percent of citations occurring within 10 miles, 19 percent within 20 miles, and 28 percent within 150 miles.

As detailed in Table 1, the dependent variable is the number of future cites to an analysis sample patent that occur from within a given distance radius around each inventor, for all available citing data. The unit of observation is an inventor-patent pair. We consider cumulative radii starting from $r = 10$ miles, as data remain too sparse within shorter radii, and stop at $r = 150$ miles, as we did not find any

Table 1—Full Sample Descriptive Statistics of Geographic Distances of Citations

Variable	Obs.	Median	Mean	SD	Min.	Max.	Share of zero cites	Share of patents with zero cites
No. cites within 10 miles	5,491	0	2.17	9.61	0	182	71.74	61.38
No. cites within 20 miles	5,491	0	3.41	14.18	0	246	62.56	54.60
No. cites within 30 miles	5,491	0	4.02	15.64	0	265	57.75	51.02
No. cites within 40 miles	5,491	0	4.34	16.76	0	273	55.22	49.23
No. cites within 50 miles	5,491	0	4.49	17.06	0	273	53.49	47.56
No. cites within 60 miles	5,491	0	4.59	17.20	0	273	52.19	46.33
No. cites within 70 miles	5,491	0	4.70	17.27	0	273	50.57	45.03
No. cites within 80 miles	5,491	1	4.80	17.35	0	273	49.12	43.62
No. cites within 90 miles	5,491	1	4.91	17.45	0	273	48.24	43.06
No. cites within 100 miles	5,491	1	5.04	17.65	0	273	46.97	41.89
No. cites within 110 miles	5,491	1	5.17	17.84	0	273	45.66	40.65
No. cites within 120 miles	5,491	1	5.24	17.90	0	273	44.95	40.04
No. cites within 130 miles	5,491	1	5.38	18.08	0	273	44.11	39.36
No. cites within 140 miles	5,491	1	5.47	18.17	0	273	43.23	38.43
No. cites within 150 miles	5,491	1	5.55	18.26	0	273	42.54	37.63

Notes: Unit of observation is an inventor-patent pair. N = 5,491 from a total of 1,621 patents with exactly one deceased inventor and 3,870 living coinventors. Application dates fall between January 1, 1976, and December 31, 2005. Distance is defined as the minimal distance between the city center of the deceased or still-living inventor of the cited patent and the city center of the closest inventor of the citing patent, measured in miles. All citations are from US patents granted through 2020.

significant effect beyond. This implies increasing concentric rings of the distance centered on the hometowns of the inventors in the analysis sample (both deceased and still living). The average number of cites that occur within 10 miles of a sampled inventor is 2.17 and increases to 5.55 within 150 miles. The number of cites is right skewed, with a median of zero or one, a maximum of 273, and a high share of zeros ranging between 43 percent and 72 percent for the full analysis sample, over the entire available citation data.

Perhaps the most concerning issue in using death for identification is the possibility of correlation between age, mortality, and outcomes. The USPTO-reported inventor deaths do not provide the reason or precise date of death or other personal characteristics; however, based on recently released data (Kaltenberg, Jaffe, and Lachman 2021), we will illustrate statistically indistinguishable estimates for more likely "premature" deaths (those under 60) as compared to the full analysis sample. As expected, descriptive statistics show a higher average age of deceased inventors of 52 years versus 45 years for still-living coinventors.

We follow Jaravel, Petkova, and Bell (2018) and create the alternative sample of possibly premature deaths by keeping only the sample analysis patents where the deceased inventor died at or before the age of 60. The remaining age-adjusted sample consists of 2,247 (1,525 living and 722 deceased) inventors and a total of 722 patents with exactly one deceased inventor. The average age at patent application drops to 43.7 years for the deceased and 44.0 years for the living coinventors. The difference in ages between both groups is small in magnitude and statistically insignificant (p-value: 0.57, two-sided t-test) in this adjusted sample. Similarly, the average number of patents each individual inventor applied for over the last five years before the year of application of the patent on which death is observed is also insignificant (avg. 3.3 versus avg. 3.1, p-value: 0.41, two-sided t-test). Neither inventor age nor the number of prior patent applications predicts death in the premature age sample. Finally, deceased and still-living coinventors do not appear to live in different areas; in particular, the US geographic centroid is 18 miles apart for the two groups (please see online Appendices A1–A10 for further descriptive statistics and details on data sources).

II. Empirical Strategy

Our identification strategy relies on the differences in citations that occur from within a given radius to the deceased as compared to citations that occur within a same-sized radius to the still-living coinventors of the same patent. Since we hold the cited (deceased) patent constant, any measurable difference should only arise from differences in the geographic distance of citations to the deceased versus still-living inventors—and not from any characteristic of the deceased patent. In other words, we identify the effect from the relative difference in citations within the immediate vicinities of the deceased inventor, relative to the citations within the immediate vicinities of the still-living coinventors.

As an example, consider two coinventors on the same patent who live in Berkeley, California, and Stanford, California. If the Stanford inventor dies and the Berkeley inventor does not, then future prior art citations to the patent from Silicon Valley (which is closer to Stanford than to Berkeley) will decrease more, relative to citations from San Francisco (which is closer to Berkeley than to Stanford).

Since the invention itself remains the same, we can attribute these changes to the death of the Stanford inventor. Not needing to compare two different inventions or relying on similar but differently codified, prosecuted, or assigned versions of a certain type of invention is the key strength of this approach. By estimating effects *within* patents, our approach effectively rules out any observable or unobservable patent characteristic that might influence the results. The other strength is that we can attribute a decrease in knowledge spillovers to individual inventors, not needing to worry about other sources of knowledge flows that might contribute to the localization of knowledge flows—for example, colocation of companies that work in the same industry, local labor market conditions, colocation of universities, general clustering of inventive activity, and so forth.

The unit of observation is a patent-inventor pair. The dependent variable is the number of cites arising within a certain radius and time window. As the number of cites is a nonnegative integer variable, we estimate a Poisson model (with standard errors clustered at the patent level):

$$(1) \qquad E\left[Cites_{prit} \middle| X_{prit}\right] = e^{\left(\alpha_0 + \beta_1 Deceased_{ip} + \pi_p\right)},$$

where $Cites_{prit}$ is the number of cites that occur within a radius r of the location of inventor i for the same multiauthored patent p within a time window of t since grant of p. $Deceased_{ip}$ indicates the inventor who died after application but before the grant of patent p, and π_p is an indicator for patent fixed effects.

We run separate and assumedly independent regressions for cumulative radii ranging from $r = 10$ miles to $r = 150$ miles, at ten-mile increments. This implies increasing concentric rings of the distances centered on the hometowns of the deceased and still-living inventors. The baseline specification uses all citations to a given patent that we can observe (excluding self-citations).

The approach makes three identifying assumptions. First, inventor death remains orthogonal to any location characteristic that may lead to the localization of knowledge flows; for example, inventors are not more or less likely to die where companies of the same industry colocate, local labor market conditions are particularly good or bad, or universities are in close proximity. To stay in the example, we assume that, all else equal, dying in Stanford is equally likely as dying in Berkeley; even if there would be differences in the specific example, we have enough cases in the larger sample such that the average place of death in the sample is not correlated with regional characteristics that would influence knowledge spillovers (facilitating factors would technically also violate our identification assumption but should work against us—i.e., lead to an underestimation of the true effect).

The second assumption is that inventor death has no direct effect on coinventors' likelihood of citation within a certain radius, as might arise, for example, if inventor death impacted the future productivity of proximal coinventors (Jaravel, Petkova, and Bell 2018; Azoulay, Zivin, and Wang 2010). For example, death could decrease productivity of coinventors through loss of knowledge or management expertise, or increase productivity if coinventors were freed from constraints (Azoulay, Fons-Rosen, and Graff Zivin 2019). Inventors might also continue to work on related ideas after the deceased patent is granted, and this related activity might generate spillovers (though by law every inventor who contributes to a patent must be listed, even if deceased[1]). To eliminate either possibility, all analyses exclude self-citations (results remain robust to inclusion), as this is the most plausible path through which the death of a coinventor might influence the estimations.[2] That means that even if inventors might file fewer, more, or different patents once their coinventor dies, those differences should not influence the estimates, as any self-citations would be discarded.

The third assumption is that inventor death is not correlated with some unobserved personal inventor characteristic that is itself correlated with the number of citations that come from within a certain radius. As an example and as discussed earlier, inventor age correlates with death in the full sample. This opens the possibility that deaths might have been anticipated and potential recipients of knowledge flows may have changed their inventive and citing behavior. Older inventors may also generate fewer or more numerous spillovers than younger inventors; for example, they may have weaker influences on the inventors around them (if they were less aware of newer technologies) or stronger influences (if their social networks were larger and they were experienced and respected contributors). Another related concern could be that old inventors have assembled larger research teams and budgets, which may lead to smaller drop in citations after their death, as there are more

[1] From USPTO Statue 2109 Inventorship [R-10.2019] (https://www.uspto.gov/web/offices/pac/mpep/s2109.html): II. AN INVENTOR MUST CONTRIBUTE TO THE CONCEPTION OF THE INVENTION. See also USPTO Statute 409 Death, Legal Incapacity, or Unavailability of Inventor [R-11.2013] (https://www.uspto.gov/web/offices/pac/mpep/s409.html): Pre-AIA 37 CFR 1.47(a) and pre-AIA 35 USC. 116, second paragraph, requires all available joint inventors to file an application "on behalf of" themselves and on behalf of a joint inventor who "cannot be found or reached after diligent effort" or who refuses to "join in an application."

[2] This requires removing all citations where any of the cited patent's inventors also appear as an inventor on the citing patent.

people around who would still work on follow-on projects (though if any of those people were coauthors of the deceased, their self-cites would be dropped).

It is worth noting two things though. First, under the null hypothesis of no local knowledge spillovers, any differential influence of deceased and living inventors (be it due to death per se or age which is correlated with death) should be constant with respect to distance. Second, any bias could go both ways, depending on whether older inventors have a stronger or weaker impact on their peers; that is, the coefficient of death might be larger or smaller than the baseline. To lessen these concerns, we look for significantly different coefficients in the possibly premature death subsample.

III. Results

Figure 1, panel A illustrates the results based on the analysis sample for each separate estimation of equation (1), where the dependent variable is all observable cites that occurred within the specified radii around an inventor's home city center. Patents receive 25.8 percent fewer cites $(1 - e^{(-0.299)})$ from within a radius of 20 miles around the deceased, relative to a radius of 20 miles around the still-living coinventors (Table 2, panel A). As can be seen, and confirming a large number of results around agglomeration economies (Duranton and Puga 2020), the impact of physical inventor presence on knowledge flows attenuates quickly; the (negative) effect is strongest within small radii and weakens almost monotonically as the concentric rings grow larger, becoming insignificant after 60 miles.

Figure 1, panel B illustrates both the full and premature death sample; though less precise (not surprisingly, given the smaller sample size), the premature death sample differs little in magnitude as compared to the full sample. Table 2, panel B shows tabular estimates. These results suggest that the correlation between age and death in the full sample is not biasing the results. It also implies that older inventors might be similarly important for the diffusion of knowledge, relative to their younger counterparts. Note that since identification relies on deaths that occurred between patent application and grant, the observed deaths in our sample will by construction oversample active inventors (more "typical" inventor deaths probably occur after a retirement from patenting and are less likely to occur during the application process).

Figure 2 illustrates how the lower local citation rate for the deceased relative to still-living coinventors is even more pronounced for geographically dispersed inventor teams, where the nearest still-living coinventor is at least 500 miles away from the deceased. For these more distant inventor teams, local citations within 20 miles around the deceased inventor drop by 70.6 percent $(1 - e^{(-1.225)}$; Table 2, panel C), rather than the 25.8 percent for the baseline result. It indicates that the importance of personal knowledge spillovers intensifies when collaborative inventors live in distant parts of the country, arguably because the alternate source of information is much farther away. In other words, a nearby and still-living coinventor can more easily "substitute" for the deceased inventor, relative to a distant and still-living coinventor.

Figure 3 illustrates how the magnitude of the effect for within 20 miles tends to attenuate with time. Results come from separate estimations of equation (1),

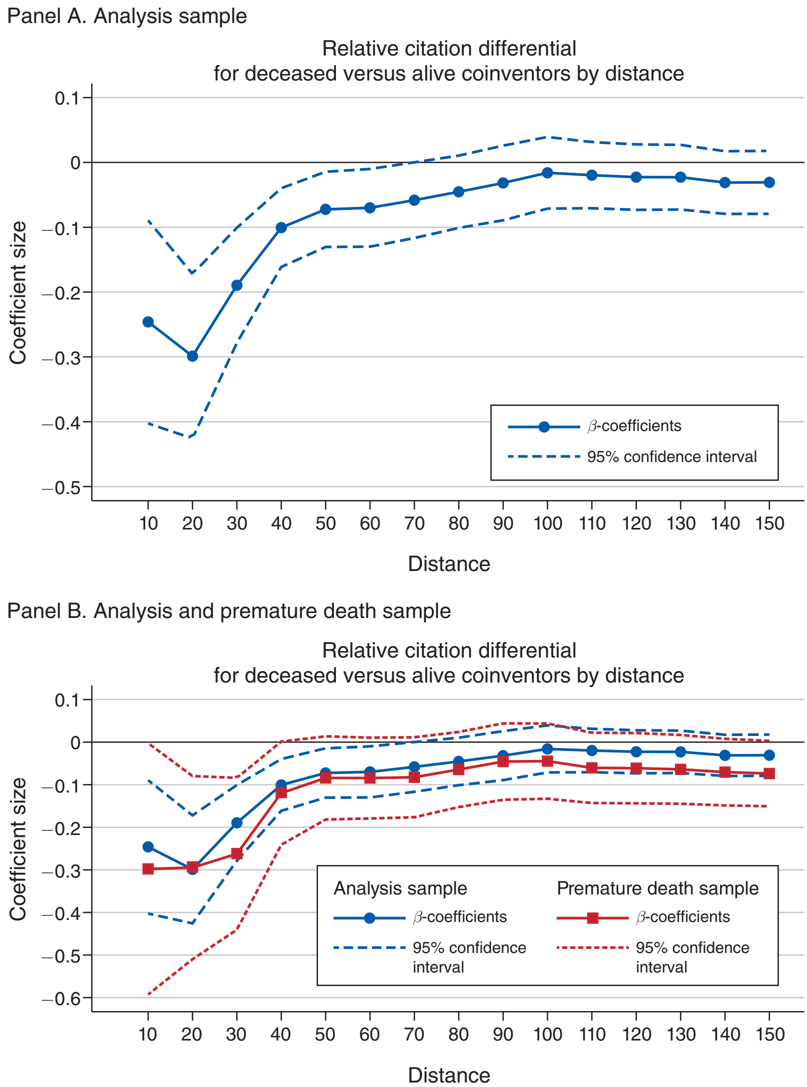

Figure 1. Citation Penalty for Deceased Relative to Living Coinventors by Distance

Notes: Panel A plots results (β_1) of separate Poisson models as specified in equation (1), where the dependent variable is the number of cites to a patent p that occur within a radius of X miles of inventor i for the same multiauthored patent p. Panel B compares analysis sample estimates of Panel A (blue) with estimates of the same models based on the premature death sample (death at age ≤ 60, colored red). All models are estimated with patent fixed effects. Distance is defined as the minimal distance between the city center of the deceased or still-living inventor of the cited patent and the city center of the closest inventor of the citing patent, measured in miles, considering all citations from US patents granted through 2020. Panel A (blue): 3,870 living and 1,621 deceased inventors; panel B (red): 1,525 living and 722 deceased inventors. Confidence bands are computed based on standard errors clustered at the patent level and assume independence of regressions.

with the radius fixed at 20 miles (where the data provide the strongest support) but count the citations that occur within this radius separately for each year since patent grant. Online Appendices A11–A13 support these findings by showing full estimates for varying radii with varying time windows t over which citations since grant of patent p are observed, ranging from all to 15, 10, and 5 years. The strongest

Table 2—Localization of Knowledge Flows for Baseline Poisson Model

	Cites from within X miles						
	10	20	30	40	50	100	150
Panel A. Analysis sample ($N = 5{,}491$)							
$Deceased_{ip}$	−0.246	−0.299	−0.190	−0.101	−0.072	−0.016	−0.031
	(0.080)	(0.065)	(0.045)	(0.031)	(0.030)	(0.028)	(0.025)
Panel B. Premature death sample ($N = 2{,}247$, *age at death* ≤ 60)							
$Deceased_{ip}$	−0.298	−0.295	−0.262	−0.120	−0.084	−0.045	−0.074
	(0.150)	(0.109)	(0.091)	(0.062)	(0.050)	(0.045)	(0.039)
Panel C. Large distance sample ($N = 749$, *distance to all coinventors at least 500 miles*)							
$Deceased_{ip}$	−1.391	−1.225	−0.997	−0.954	−0.804	−0.604	−0.512
	(0.287)	(0.257)	(0.234)	(0.234)	(0.218)	(0.210)	(0.208)

Notes: This table presents results of separate Poisson models as specified in equation (1), where the dependent variable is the number of cites to a patent p that occur within a radius of X miles of inventor i for the same multiauthored patent p. All models are estimated with patent fixed effects. Distance is defined as the minimal distance between the city center of the deceased or still-living inventor of the cited patent and the city center of the closest inventor of the citing patent, measured in miles, considering all citations from US patents granted through 2020. The unit of observation is an inventor-patent pair. $Deceased_{ip}$ is a dummy that indicates the inventor who died after application but before grant of patent p. Panel A: 3,870 living and 1,621 deceased inventors; panel B: 1,525 living and 722 deceased inventors; panel C: 535 living and 214 deceased inventors. N includes patents with zero future cites as reported in Table 1. Standard errors clustered at the patent level are reported in parentheses.

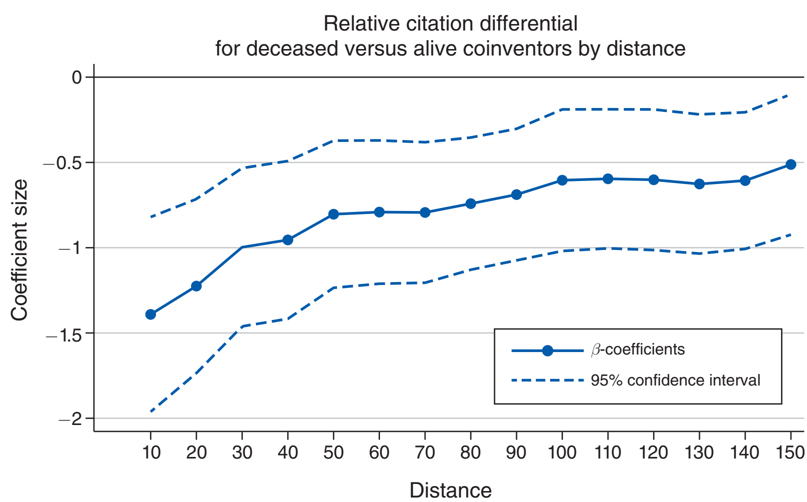

Figure 2. Estimates for Patents Where Next Living Coinventor Lives at Least 500 Miles Away from Hometown of Deceased Inventor

Notes: This figure plots results (β_1) of separate Poisson models as specified in equation (1), where the dependent variable is the number of cites to a patent p that occur within a radius of X miles of inventor i for the same multiauthored patent p. All models are estimated with patent fixed effects. Distance is defined as the minimal distance between the city center of the deceased or still-living inventor of the cited patent and the city center of the closest inventor of the citing patent, measured in miles, considering all citations from US patents granted through 2020. N = 749, including 214 deceased inventors and 535 living inventors who live at least 500 miles away from the deceased inventor on the same patent. Confidence bands are computed based on standard errors clustered at the patent level and assume independence of regressions.

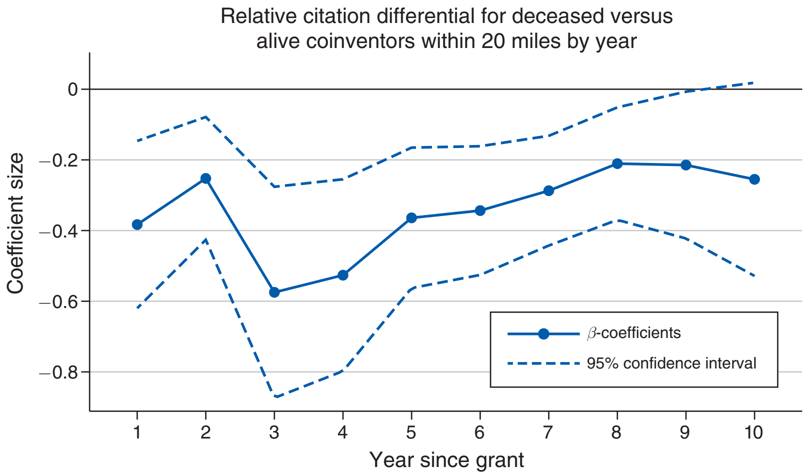

FIGURE 3. CITATION PENALTY FOR DECEASED RELATIVE TO LIVING COINVENTORS WITHIN 20 MILES IN YEARS 1 TO 10 SINCE GRANT

Notes: This figure plots results (β_1) of separate Poisson models as specified in equation (1), where the dependent variable is the number of cites to a patent p that occur within a radius of 20 miles of inventor i in year X after grant of p for the same multiauthored patent p. All models are estimated with patent fixed effects. Distance is defined as the minimal distance between the city center of the deceased or still-living inventor of the cited patent and the city center of the closest inventor of the citing patent, measured in miles, considering all citations from US patents granted through 2020. N = 5,491 from 3,870 living and 1,621 deceased inventors. Confidence bands are computed based on standard errors clustered at the patent level and assume independence of regressions.

relative difference in citations consistently occurs within five years following inventor death, both temporally and geographically. It indicates that effects from inventor death appear to wane as time progresses (similar to findings of JTH and TF-K), though estimates remain too noisy for establishing significantly different effects across time.

Robustness Checks and Extensions.—Results proved robust to including more granular distances and a variety of citation time windows (the pattern of weakening effects over time and distance emerges through all the robustness checks), inverse frequency weighting by the number of inventors on a patent, and CEM matching (Iacus, King, and Porro 2012) for an age-balanced sample (age also does not predict death in the balanced sample). Although the baseline Poisson model is preferred, results remain robust to alternatively estimating Poisson models without the top 1 percent most highly cited patents in the analysis sample, a LPM with $\Pr(Cites_{pri} > 0)$ as the dependent variable, and a $\log(\mathrm{Y}+1)$ equation using OLS, with Y the number of citations within a given radius. Results also remain robust to using the subsample of citations that were added by examiners (which consistently measure distances between the deceased and still-living inventors, and future citing inventors), which should lessen concerns of social or strategic bias in citation patterns (please see online Appendices A11–A26 for tabular results and graphical illustrations).

Comparison to Prior Results.—One way to benchmark our estimates is to consider what proportion of knowledge spillovers are lost when an inventor dies. Our data indicate that 19 percent of all patent cites occur from within 20 miles of a given inventor location. If we apply the baseline estimate of a 25.8 percent reduction in local cites within 20 miles, we can calculate that an inventor death reduces all knowledge spillovers by $0.258 \times 19 = 4.9\%$. This would be slightly higher if one added the significant effects at radii of 30 to 60 miles, but given that our baseline models indicate rapidly decreasing and no significant effects beyond 60 miles, 4.9 percent provides a conservative estimate.

Panel C of Table 2 indicates that the loss of local spillovers is sensitive to the distance between a deceased inventor and his or her still-living coinventors; the estimate increases when a deceased inventor's coinventors live at least 500 miles away. From panel C, the 70.6 percent reduction in local knowledge spillovers within 20 miles implies a reduction in knowledge flows of $0.706 \times 19 = 13.4\%$. This might provide an upper-bound estimate for a world where all inventors lived apart from one another—that is, there were no local spillovers generated by physical presence. These estimates might also inform assessments of the impact from a less than fatal event—for example, if an inventor simply moved out of the local region or a pandemic forced inventors to work remotely. In that case, she or he would still be accessible via email, phone, or video and the loss of local spillovers would probably be less.

Rough comparisons to prior research can be made (all comparisons exclude self-citations). JTH took an originating sample, found citing patents to that originating sample, and matched each citing patent with a patent from the same technology class and application date. They then compared the proportions of the two citing patent populations that were invented in the same region as the originating patents. The approach found that "citations are five to ten times as likely to come from the same SMSA [standard metropolitan statistical area] as control patents; two to six times as likely excluding self-citations." (JTH, 591). The critique of TF-K recreated the original JTH case matching method; however, it matched at a much finer granularity of technology classifications and found no localized spillovers.

JTH and TF-K observe whether a citing or control patent is located within a SMSA; given the typical size of SMSAs, the closest analog in our estimations is the 10 or 20 mile radius citation models (we use the 20 mile coefficient for consistency in discussion). Keeping the differences in approaches in mind, JTH observe a 100–500 percent greater fraction of locally generated citations, TF-K observe no difference in the fraction of locally generated citations, and we observe decreases between 25.8 percent and 70.6 percent of locally generated citations, for a deceased inventor versus still-living coinventors of the same patent. While the magnitude of our estimates are far less than the 100–500 percent implied by the original JTH estimate, they remain highly significant, unlike the insignificant results estimated by TF-K.

All these estimates can be justly criticized for a number of reasons, including knowledge spillovers that are not captured by a citation, personal versus nonpersonal knowledge spillovers, and fundamental differences in methods. Of primary importance, our method focuses on the interpersonal and local impact of one inventor, as opposed to JTH and most of the literature that seeks to measure all localized spillovers from one patent.

IV. Conclusion

This work contributed a novel approach to empirical estimations of the localization of knowledge spillovers, based on the relative difference in local citations to the same patent, between deceased versus still-living coinventors. It used inventor death between the application and grant of a coauthored patent to estimate the personal impact of an inventor upon the diffusion of knowledge from their unique geographic location, providing an arguably causal method that isolates knowledge spillovers that can be attributed to the physical presence of one inventor. It estimated decreases of between 25.8 percent and 70.6 percent in local citations within 20 miles, following the death of an inventor, relative to his or her coinventors living in other locations. The work added to the aggregation of evidence that personal spillovers are mostly local (JTH) and a growing number of empirical innovations (TF-K; Murata et al. 2014; Ganguli, Lin, and Reynolds 2020) that support the use of citations as a measure of knowledge flow. The localization appeared to attenuate with time, was stronger when still-living coinventors lived farther away from the deceased, and provided an ironic illustration of how personal interaction matters, even for a canonical and legally stipulated example of codified knowledge—that is, a published patent (Williams 2017). The approach should hopefully improve our ability to empirically disentangle knowledge spillovers from the other two Marshallian agglomeration mechanisms of colocation and pooled labor.

Perhaps the biggest drawback of the approach taken here are the demands made on the data; many patents are not highly cited to begin with and this makes it difficult to identify differences in precisely defined distances (i.e., within radii, there often exists no citation, let alone a sharp geographic difference in the distance from citing inventors to collaborative inventors living in different locations). Recent efforts to link mortality and inventor databases offer much promise, though as described in Kaltenberg, Jaffe, and Lachman (2021), the effort is difficult and few links can be as certain as those indicated on the front page of patents. Nonetheless, if the data can be linked, the approach taken here offers a quasi-experimental and arguably causal method to investigating spillovers in a variety of contexts, including geography, time, organizational boundaries, social networks, and technology or knowledge space. The approach might be extended to any bibliometric and citation database where the careers of collaborative creators can be located across a space, thus allowing the investigation of knowledge spillovers generated by a collaborative author, be they an inventor, humanist, or any type of publishing scientist.

REFERENCES

Alcácer, Juan, and Michelle Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88 (4): 774–79.

Arora, Ashish, Sharon Belenzon, and Honggi Lee. 2018. "Reversed Citations and the Localization of Knowledge Spillovers." *Journal of Economic Geography* 18 (3): 495–521.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review* 109 (8): 2889–920.

Azoulay, Pierre, Joshua G. Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics* 125 (2): 549–89.

Balsmeier, Benjamin, Mohamad Assaf, Tyler Chesebro, Gabe Fierro, Kevin Johnson, Scott Johnson, Guan-Cheng Li, et al. 2018. "Machine Learning and Natural Language Processing on the Patent

Corpus: Data, Tools, and New Measures." *Journal of Economics and Management Strategy* 27 (3): 535–53.

Balsmeier, Benjamin, Lee Fleming, and Sonja Lück. 2023. "Replication data for: Isolating Personal Knowledge Spillovers: Coinventor Deaths and Spatial Citation Differentials." Harvard Dataverse. https://doi.org/10.7910/DVN/S5BBFY.

Duranton, Gilles, and Diego Puga. 2020. "The Economics of Urban Density." *Journal of Economic Perspectives* 34 (3): 3–26.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–213.

Ganguli, Ina, Jeffrey Lin, and Nicholas Reynolds. 2020. "The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences." *American Economic Journal: Applied Economics* 12 (2): 278–302.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20 (1): 1–24.

Jaffe, Adam B., and Gaétan de Rassenfosse. 2017. "Patent Citation Data in Social Science Research: Overview and Best Practices." *Journal of the Association for Information Science and Technology* 68 (6): 1360–74.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-Specific Capital and Innovation." *American Economic Review* 108 (4–5): 1034–73.

Kaltenberg, Mary, Adam Jaffe, and Margie E. Lachman. 2021. "Matched Inventor Ages from Patents, Based on Web Scraped Sources." Harvard Dataverse. https://doi.org/10.7910/DVN/YRLSKU (accessed May 6, 2021).

Krugman, Paul. 1991. *Geography and Trade*. Cambridge, MA: MIT Press.

Li, Guan-Cheng. 2019. "USPTO Inventor, Assginee and Geography Disambiguation Dataset, 1976–2018." Harvard Dataverse, V1. https://doi.org/10.7910/DVN/KPMMPV (accessed November 12, 2019).

Marshall, Alfred. 1890. *Principles of Economics*. London: MacMillan.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96 (5): 967–85.

PatentsView. 2021. "Data Download Tables." https://patentsview.org/download/data-download-tables (application.zip downloaded on April 7, 2021; inventor.zip downloaded on March 13, 2021; location.zip downloaded on March 13, 2021; patent.zip downloaded on April 7, 2021; patent_inventor.zip downloaded on March 13, 2021; persistent_inventor_disambig.zip downloaded on June 26, 2021; rawinventor.zip downloaded on March 13, 2021; uspatentcitation.zip downloaded on April 6, 2021).

SimpleMaps. 2020. "United States Cities Database." https://simplemaps.com (accessed November 26, 2020).

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95 (1): 450–60.

Williams, Heidi L. 2017. "How Do Patents Affect Research Investments?" *Annual Review of Economics* 9: 441–69.