Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Testing for localization: A new approach

Yasusada Murata, Ryo Nakajima, Ryuichi Tamura

3 May, 2017
DP2017-017
https://ies.keio.ac.jp/publications/7688/

Keio University

CALAMVS GLADIO FORTIOR
1858

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
3 May, 2017

Testing for localization: A new approach

Yasusada Murata, Ryo Nakajima, Ryuichi Tamura

Keio-IES DP2017-017
3 May, 2017
JEL classification: R12; O31
Keywords: localization; knowledge spillovers; local linear density; boundary bias; micro-geographic data

Abstract

Recent empirical studies document that knowledge spillovers attenuate and industry localization decays with distance. It is thus imperative to detect localization accurately especially at short distances. We propose a new approach to testing for localization that corrects the first-order bias at and near the boundary in existing methods while retaining all desirable properties at interior points. Employing the NBER U.S. Patent Citations Data File, we illustrate the performance of our localization measure based on local linear density estimators. Our results suggest that the existing kernel density methods and regression approaches can be substantially biased at short distances.

Yasusada Murata
Nihon University Population Research Institute (NUPRI); and University Research Center, Nihon University.
2-6 Gobantyou, Chiyoda, Tokyo, 102-8251, Japan.
murata.yasusada@nihon-u.ac.jp

Ryo Nakajima
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
nakajima@econ.keio.ac.jp

Ryuichi Tamura
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ry.tamura@gmail.com

Testing for localization: A new approach*

Yasusada Murata† Ryo Nakajima‡ Ryuichi Tamura§

April 27, 2017

Abstract

Recent empirical studies document that knowledge spillovers attenuate and industry localization decays with distance. It is thus imperative to detect localization accurately especially at short distances. We propose a new approach to testing for localization that corrects the first-order bias at and near the boundary in existing methods while retaining all desirable properties at interior points. Employing the NBER U.S. Patent Citations Data File, we illustrate the performance of our localization measure based on local linear density estimators. Our results suggest that the existing kernel density methods and regression approaches can be substantially biased at short distances.

Keywords: localization; knowledge spillovers; local linear density; boundary bias; micro-geographic data

JEL codes: R12; O31

*The authors thank Gilles Duranton, Hidehiko Ichimura, Bill Kerr, Henry Overman, Will Strange, as well as participants in Conference on Regional Economics and Public Policy (University of Bern), 9th Meeting of the Urban Economics Association (Washington, D.C.), 4th European Meeting of the Urban Economics Association (Saint Petersburg), Third International Conference "Industrial Organization and Spatial Economics" (Saint Petersburg), III Workshop on Urban Economics (Barcelona), 3rd European Meeting of the Urban Economics Association (Palermo) for helpful comments and suggestions. Tamura acknowledges financial support from JST/RISTEX.

†Nihon University Population Research Institute (NUPRI); and University Research Center, Nihon University. E-mail: murata.yasusada@nihon-u.ac.jp

‡Department of Economics, Keio University; and Université Côte d'Azur, CNRS, GREDEG. E-mail: nakajima@econ.keio.ac.jp

§Department of Economics, Keio University. E-mail: ry.tamura@gmail.com

1 Introduction

Localization – concentration over and above some reference distribution of economic activity – is perhaps the most salient feature of the spatial economy. Since the seminal works of Jaffe et al. (1993) and Ellison and Glaeser (1997), the localization of knowledge spillovers and that of industries have been detected largely within administrative units such as states and metropolitan statistical areas (MSAs).[1] It is now widely recognized that localization results differ depending on the chosen spatial units, which is known as the modifiable areal unit problem (Gehlke and Biehl, 1934).

The recent availability of micro-geographic data has spurred the development of distance-based methods that do not rely on administrative units. Duranton and Overman (2005) and Murata et al. (2014) conduct kernel-density (henceforth K-density) estimations to smooth the distribution of distances between establishments or inventors. Belenzon and Schankerman (2013) and Kerr and Kominers (2015) take regression approaches using dummy variables for distance rings, each of which is generated by discretization of distances. These two distance-based methods have shed light on detailed location patterns.

Yet, statistical theory shows that K-density estimations (even when adopting a reflection method) have the first-order bias at short distances (Jones, 1993; Marron and Ruppert, 1994; Simonoff, 1996, Section 3.2),[2] whereas distance-ring dummy-variable (DD) regressions are sensitive to the way of discretizing the micro-geographic data.[3] Since a large number of recent empirical studies document that knowledge spillovers attenuate and industry localization decays with distance,[4] it is imperative to detect localization more accurately especially at short distances.

We thus propose a new approach to testing for localization using micro-geographic data, which allows us to overcome the problems with DD regressions and K-density estimations. The idea is to combine key ingredients of the existing discretizing and smoothing methods as in the following two steps. In the first step, we *discretize* the micro-geographic data into distance

[1]Put differently, the extent of localization is limited within states or MSAs, and the analysis abstracts from the relative position of spatial units by making, for example, the distance from New York to Boston equivalent to that of New York to Los Angeles.

[2]For the interested reader, this statistical property is provided in Appendix A.1.

[3]For example, to detect localization at 4 km, we need to focus on the second bin if the binwidth is 3 km. However, if the binwidth is 5 km, we need to examine the first bin. Hence, the localization result differs depending on the binwidth, which may be viewed as a variant of the modifiable areal unit problem.

[4]Indeed, these results are well documented in the literature (e.g., Rosenthal and Strange, 2003; Duranton and Overman, 2005, 2008; Arzaghi and Henderson, 2008; Belenzon and Schankerman, 2013; Murata et al., 2014; Buzard et al., 2015; Carlino and Kerr, 2015; Kerr and Kominers, 2015).

bins as in DD regressions and count the number of observations in each bin to construct a histogram. In the second step, we *smooth* the histogram to obtain a *local linear density* (L-density) estimator, i.e., using the distance at each bin center and the corresponding height of the histogram, we conduct a local linear regression, which is a smoothing technique as in K-density estimations. Applying the same two-step procedure to a control sample generated by randomization and taking the difference between the two estimates – one capturing the spatial distribution of an observed sample and the other capturing that of a control sample – we construct a new localization measure that enables us to test if there is a significant departure from randomness.

Our new measure derived from such L-density estimations – local linear regressions using the pre-binned data – has advantages over the existing discretizing and smoothing methods for at least two reasons.

First, one may argue that, instead of the L-density estimations, the existing K-density estimations could be used as the latter also rely on a smoothing technique. While the K-density method displays only the second-order bias at interior points, it suffers from the first-order bias at short distances (where localization is most likely to occur).[5] In contrast, the L-density estimators have the second-order bias even at short distances.

Second, one could stop at the first step of the L-density estimations and obtain a localization measure from binning per se. However, such a measure based on histograms is shown to be essentially the same as the localization measure generated by DD regressions, which is sensitive to the way of discretizing the micro-geographic data.[6] We cope with this problem arising from discretization by conducting the local linear regressions in the second step.

In a nutshell, our localization measure, which combines binning and smoothing, resolves the issue of discretization and has a better property at and near the boundary while retaining all desirable properties of the existing methods at interior points.

To illustrate the performance of our L-density measure relative to the K-density measure, we employ the NBER U.S. Patent Citations Data File (Hall et al., 2001). Using the data on observed citation distances for all patent classes, we first show that the estimated K-density with (or without) reflection is far below the estimated L-density at the boundary.

[5] In addition, if the kernel is symmetric and differentiable like the Gaussian kernel that is often used in the literature, the K-density estimator with reflection must have zero derivative at the boundary even when the true density does not (Silverman, 1986, Section 2.10).

[6] Furthermore, the histogram estimators have the first-order bias even at interior points when the derivative of the true density is not zero (Simonoff, 1996, Section 2.1).

Furthermore, this underestimation at the boundary induces overestimation near the boundary since the densities must sum to one over the entire domain of distances (see Figure 1 for these results and Section 4.1 where we discuss them in detail).

Insert Figure 1

The first-order bias caused by the K-density estimations is not offset even when we perform our localization tests by taking the difference between the actual and counterfactual distributions.[7] The reason is that as shown below the actual distribution is more right skewed (i.e., the mass of the actual distribution is concentrated in relatively shorter distances) than the counterfactual ones, and thus suffers more from the first-order bias at short distances.

To see this more clearly, we turn to the disaggregated analysis at the patent class level. The class-specific distance-based tests reveal that the L-density tests detect localization for significantly more patent classes than the K-density tests. We further rank patent classes by the magnitude of the localization indices, and find that the ranking differs between the K- and L-density methods. This result suggests that the localization indices based on the K- and L-densities provide a different piece of information. Our L-density approach will help to revisit the determinants and shapes of agglomeration and coagglomeration, given that the K-density indices have been incorporated into empirical specifications in that literature (see, e.g., Ellison et al., 2010; Kerr and Kominers, 2015).

Since our main focus is on the difference between the K- and L-density methods *at short distances*, we explore at which distance the K-density tests underestimate the number of localized patent classes as compared to the L-density tests. It turns out that 40 out of 61 underestimated classes fall within the domain between 0 km and 100 km, which account for 66 % of the number of underestimated classes. The K-density tests thus underestimate the number of localized classes especially at short distances. Hence, eliminating the first-order bias at short distances is of first-order importance. This is especially so because various empirical studies show that knowledge spillovers attenuate and industry localization decays within short distances. Correcting the K-density tests is crucial for a better understanding of localization of economic activity as they have been widely used for detecting localization ever since Duranton and Overman (2005).

The L-density estimator depends on the binwidth in the first step and on the bandwidth in the second step. We thus examine how sensitive our L-density tests are to their choices.

[7] For the interested reader, this statistical property is provided in Appendix A.2.

We find that neither the localization result *for all distances* nor our main result *for short distances* is affected by different binwidth and bandwidth selections. Hence, the result that the K-density tests underestimate the number of localized patent classes than the L-density tests especially at short distances is fairly robust. All these results remain valid even when we use the common bandwidth for the K- and L-density tests while allowing the binwidth for the L-density tests to vary. Therefore, our results are not driven by binwidth and bandwidth selections.

The rest of the paper is organized as follows. In Section 2 we propose a new approach using L-density estimators and compare it to the existing density-based localization measure. The relation of our L-density measure to alternative regression-based localization measures is relegated to Appendix B. Sections 3 and 4 describe the data and the estimation procedures, respectively. The estimation results and their robustness are reported in Section 4. Section 5 concludes.

2 Density-based localization measures

In this section, we propose a new approach to testing for localization using micro-geographic data. We first present our localization measure based on the L-density estimators. We then compare it to the existing density-based localization measure, namely the K-density measure of localization. We finally introduce possible regression-based localization measures that we discuss in Appendix B.

2.1 A new approach

Following the seminal work by Jaffe et al. (1993), we consider the case of localized knowledge spillovers as evidenced by patent citations for ease of exposition. Our approach, however, can be readily applied to other contexts such as industry localization and co-localization. As in many studies on patent citations and knowledge spillovers, our analysis relies on the premises that patents proxy new ideas, and that patent citations proxy flows of these ideas.

Our localization measure involves the difference between observed and control samples as in Jaffe et al. (1993) and Duranton and Overman (2005). To begin with, we obtain the L-density estimator for an observed sample from the following two steps. In the first step, we discretize the micro-geographic data into distance bins and count the number of observations in each bin to construct a histogram. In the second step, we smooth the histogram to generate an L-density estimator, i.e., using the distance at each bin center and the corresponding height of the histogram, we conduct a local linear regression. Applying the same two-step procedure to a control sample generated by randomization, and taking the difference between the two density estimators – one capturing the spatial distribution of an observed sample and the other capturing that of a control sample, we construct a localization measure to see if there is a significant departure from randomness. The new localization measure based on such L-density estimations – local linear regressions using the pre-binned data – performs better than the existing ones at and near the boundary, while retaining all desirable properties at interior points.

To formally describe the procedure, let $\mathcal{N}^c$ denote the set of cited-citing relationships.[8] We define the distance d_{ij} between an originating patent i and a citing patent j for each citation link $(i, j) \in \mathcal{N}^c$. Let $r = 1, \ldots, R$ denote the index of distance rings. We define the r-th distance ring as $(d_r - \Delta/2, d_r + \Delta/2]$, where d_r is the midpoint of the r-th distance ring and Δ is the binwidth. For example, d_1 corresponds to 5 km, d_2 to 15 km, d_3 to 25 km and so on, in which case the binwidth Δ is 10 km.

The L-density for observed citation distances can be estimated as in the following two steps. In the first step, we classify the micro-geographic data on citation distances $\{d_{ij}\}_{(i,j)\in\mathcal{N}^c}$ into R different rings. To this end, we define for each cited-citing relationship $(i, j) \in \mathcal{N}^c$ an indicator variable as

$$D_{ijr} = \mathbb{I}\left(d_{ij} \in (d_r - \Delta/2, d_r + \Delta/2]\right), \tag{1}$$

which takes one if d_{ij} is in the r-th distance ring and zero otherwise. For each distance ring r, we count the number $N_r^c = \sum_{(i,j)\in\mathcal{N}^c} D_{ijr}$ of citations. Let $N^c = \sum_{r=1}^{R} N_r^c$ be the total number of citations. The height of the histogram for distance ring r is then given by the share N_r^c/N^c of observed citations falling within distance ring r divided by the binwidth Δ as follows:

$$c_r = \frac{N_r^c}{N^c\Delta}. \tag{2}$$

As a result of the first step, the micro-geographic data on citation distances $\{d_{ij}\}_{(i,j)\in\mathcal{N}^c}$, whose sample size is N^c, boils down to the binned data $\{d_r, c_r\}_{r=1}^{R}$ with sample size R, which is much smaller than N^c.

[8]In the context of testing for industry localization using micro-geographic data (e.g., Duranton and Overman, 2005), one can reinterpret $\mathcal{N}^c$ as the set of pairs of all possible establishments in the industry in question.

In the second step, we conduct a local linear regression using the pre-binned data, which amounts to solving the minimization problem as follows:

$$\min_{\{\beta_0,\beta_1\}} \sum_{r=1}^{R} \left[c_r - \beta_0 - \beta_1(d_r - d)\right]^2 f((d_r - d)/h), \tag{3}$$

where f is a kernel function and h is a bandwidth. Then, the L-density estimator, which is defined as $\hat{\beta}_0$ at a given distance d, can be expressed as follows (Cheng, 1994, eq.(2.18)):

$$\hat{L}^c(d) = \frac{\sum_{r=1}^{R} \left[\hat{S}_2(d) - \hat{S}_1(d)(d_r - d)\right] f((d_r - d)/h)c_r}{\hat{S}_2(d)\hat{S}_0(d) - \hat{S}_1(d)^2}, \tag{4}$$

where $\hat{S}_q(d) = \sum_{r=1}^{R}(d_r - d)^q f((d_r - d)/h)$.

Having obtained the L-density estimator for the observed sample, we turn to the L-density estimation using a control sample. For each observed cited-citing relationship $(i, j) \in \mathcal{N}^c$, we identify a set of control patents, from which we randomly draw a control patent to generate a counterfactual citation. This allows us to define the set $\tilde{\mathcal{N}}^c$ of hypothetical cited-citing relationships. By construction of counterfactual citations, the total number $\tilde{N}^c$ of counterfactual citations equals that of actual citations, i.e., $\tilde{N}^c = N^c$. Computing the number $\tilde{N}^c_r$ of counterfactual citations in each distance ring, we obtain $\tilde{c}_r = \tilde{N}^c_r/(\tilde{N}^c\tilde{\Delta}) = \tilde{N}^c_r/(N^c\tilde{\Delta})$, which is similar to (2). Using $\{\tilde{d}_r, \tilde{c}_r\}_{r=1}^{\tilde{R}}$ thus generated, we obtain the L-density estimator $\tilde{L}^c(d)$ for counterfactual citations.[9]

We finally define a localization measure based on the L-density estimators by taking the difference between the observed and control samples as in Jaffe et al. (1993) and Duranton and Overman (2005). Our new measure of localization at distance d is thus given by the difference between the two L-density estimators as follows:

$$\gamma_{\mathrm{LD}}(d) \equiv \hat{L}^c(d) - \tilde{L}^c(d). \tag{5}$$

We will describe how this measure can be used for testing for localization in Section 4.2.[10]

[9] See Section 3.3 for more details on the construction of counterfactual citations. In the context of testing for industry localization using micro-geographic data (e.g., Duranton and Overman, 2005), one can reinterpret $\tilde{\mathcal{N}}^c$ as the set of pairs of establishments in a hypothetical industry generated by randomization.

[10] The L-density estimator has become increasingly more popular in recent empirical studies, especially in the context of the regression discontinuity design (e.g., McCrary, 2008). Our novelty lies in constructing the localization test based on the L-density estimators that do not suffer from the first-order bias at short distances, and that are not sensitive to the way of discretizing the micro-geographic data.

2.2 Comparison

We now compare our localization measure γ_{LD} to the existing density-based measure of localization, namely, the K-density measure of localization. We then introduce possible regression-based localization measures that we discuss in Appendix B.

It is easy to see that our localization measure γ_{LD} is similar to the localization measure based on K-density estimations (Duranton and Overman, 2005):

$$\gamma_{\text{KD}}(d) \equiv \hat{K}^c(d) - \tilde{K}^c(d), \tag{6}$$

where

$$\hat{K}^c(d) = \frac{1}{N^c h} \sum_{(i,j) \in \mathcal{N}^c} f((d - d_{ij})/h) \tag{7}$$

is the K-density estimator for actual citations, and the counterfactual estimator $\tilde{K}^c(d)$ is analogously defined by using the set $\tilde{\mathcal{N}}^c$ of counterfactual citations. It is worth emphasizing that both γ_{LD} and γ_{KD} are given by the difference of the density estimators.

However, it is recognized that the K-density estimator, when applied to bounded data (like our distance data, which is bounded from below at zero), can be severely biased at the boundary. To be more precise, even with the reflection method that is often adopted in the localization literature, the bias of the K-density estimator is $O(h)$ at the boundary, while it is $O(h^2)$ at interior points (Jones, 1993; Marron and Ruppert, 1994; Simonoff, 1996, Section 3.2). Thus, in finite samples, the boundary bias can be substantial. As we discuss in Section 4.1, the underestimation at the boundary induces overestimation near the boundary since the densities must sum to one over the entire domain of distances.[11]

Given that knowledge spillovers attenuate and industry localization decays within short distances, the L-density estimator has both theoretical and practical advantages over the existing estimators (Cheng, 1994; Cheng et al., 1997).

Theoretically, the L-density estimator has the same asymptotic rate of convergence at the boundary as in the interior, which is $O(h^2)$, and thus displays a better boundary property than the K-density estimator. One may argue why we combine the existing discretizing method and smoothing technique. Indeed, one could stop at the first step and derive from binning per se

[11] Furthermore, in the special case where the kernel function is symmetric and differentiable like the Gaussian kernel that is widely used in the literature, the K-density estimator must have zero derivative at the boundary even when the true density does not (Silverman, 1986, Section 2.10). See Wand and Jones (1995) for further discussion on boundary bias in nonparametric methods.

the localization measure $\gamma_{\text{HIST}}(d)$ based on the histograms at distance $d \in (d_r - \Delta/2, d_r + \Delta/2]$ as follows:

$$\gamma_{\text{HIST}}(d) \equiv c_r - \tilde{c}_r. \tag{8}$$

We show in Appendix B that γ_{HIST} is essentially the same as the localization measure obtained from the DD regressions (e.g., Belenzon and Schankerman, 2013; Kerr and Kominers, 2015). Without a smoothing parameter h, however, γ_{HIST} is sensitive to the choice of Δ and has the first-order bias even at interior points unless the derivative of the true density is zero (Simonoff, 1996, Section 2.1).

Practically, the implementation of the L-density estimation is simple and fast in computation. The reason is that the sample size for the local linear regression in the second step equals the number R of bins, which is much smaller than the number N^c of citations that is required for the K-density estimation. Note that the kernel function f in (7) relies on d_{ij} for $(i, j) \in \mathcal{N}^c$, rather than on d_r for $r = 1, ..., R$ in the L-density estimator (4).

Alternatively, one could skip binning in the first step and construct a localization measure based only on the local linear regression. Although we discuss this possibility in Appendix B, the analysis requires the information on the set of all possible pairs of patents (not just on the set $\mathcal{N}^c$ of pairs of patents having cited-citing relationships). Thus, the sample size for the local linear regression without pre-binning is much larger than N^c that is required for the K-density estimation.

Our L-density approach, which combines the existing discretizing method and smoothing technique, allows us to address these theoretical and practical problems simultaneously.

3 Data

3.1 Patents and patent citations

Our data are based on the NBER U.S. Patent Citations Data File by Hall et al. (2001), which covers all patent applications between 1963 and 1999 and those granted by 1999, as well as cited-citing relationships for patents granted between 1975 and 1999. For each patent, the list of inventors, the addresses of inventors, and the technological category are recorded, along with other information such as the year of application, assignees, and the type of assignees. The information of patent application month and *patent class* (3-digit) and *subclass* (6-digit) codes is supplemented with the United States Patent and Trademark Office (USPTO) Patent

BIB database.

We begin with $142,245$ U.S. nongovernmental patents granted between January 1975 and December 1979. The sample period is chosen to be comparable to the previous studies. We focus on the "U.S. patents" whose assignees are in the contiguous United States. We observe that $115,905$ (81.5%) of them were cited at least once by other U.S. patents (see Column 1 of Table 1). We call them the *originating patents* and identify the *citing patents* that cited the originating patents by examining all patents granted between January 1975 and December 1999.

Insert Table 1

To focus on knowledge flows between different inventors of different assignees, we exclude "self-citations". A citing patent is classified as self-citing (i) if it had the same assignee as its originating patent; or (ii) if it was invented by the same inventor as its originating patent. To distinguish unique inventors, we use the computerized matching procedure proposed by Trajtenberg et al. (2006). We find that 15.0% of citing patents are classified as self-citations. After excluding self-citations, we obtain $647,983$ citing patents (see Column 1 of Table 1).

3.2 Geographic information

We identify the location of each invention at the census place level. The U.S. Census Bureau defines a place as a concentration of population. There are $23,789$ places in the 1990 census, which we use below. Restricting patent inventors who reside in the contiguous U.S. area, we first match the address of each inventor to its 1990 census place by name. If the name match fails, we locate it via the populated place provided by the U.S. Geographic Names Information System (GNIS). We match the inventor's address with the GNIS populated place, which is more finely delineated than the census place, and then find the census place that is nearest to the identified GNIS populated place by using their spatial coordination information. This procedure identifies $18,139$ census places for 97.0% of all inventors in the sample.

We measure the distance between an originating patent and a citing patent by the great-circle distance between the census place of the originating inventor and that of the citing inventor.[12] When there is more than one inventor of a patent, we compute all possible

[12]Since census places are not spatial points, this definition poses a "zero distance" problem, i.e., even when the actual distance between the originating and citing inventors is not zero, it is measured to be zero if they happen to live in the same census place. As in Murata et al. (2014) we correct this by using the distance between the two randomly chosen points in census place ℓ with area $\mathcal{A}_\ell$, which is given by $[128/(45\pi)]\sqrt{\mathcal{A}_\ell/\pi}$

great-circle distances between the census places of the originating inventors and those of the citing inventors. We then consider two types of distances – minimum and median distances – between the cited and citing patents.

3.3 Control patents and counterfactual citations

To test whether knowledge spillovers are localized, we must control for the existing spatial distribution of technological activities. Following Jaffe et al. (1993), we consider *control patents* satisfying the following two conditions.[13] First, control patents should belong to the same technological area as the citing patent under consideration. In the baseline case we select control patents at the 3-digit level and check the robustness of the result by choosing finer controls at the 6-digit level. We refer to the former as *3-digit controls*, and call the latter *6-digit controls* as in Murata et al. (2014). Second, control patents should be in the same cohort as the citing patents. As in the previous studies, we use one-month and six-month windows for the 3-digit and 6-digit controls, respectively.

The second and third columns of Table 1 report the numbers of originating and citing patents having at least one control. Note that citing patents do not always have controls, and, even if they do, the control is not necessarily unique for each citing patent. As shown, 60.2% of the citing patents have 3-digit controls. The rate of the citing patents having 6-digit controls is lower, at 18.7%. The citing patents with no controls assigned (and their originating patents) are dropped out of the sample. We also drop patent classes in which originating patents are distributed across less than 10 census places to obtain well-behaved estimated density functions. As a result, 92.6% of the originating patents remain "in-sample" for the 3-digit controls, and the corresponding number is 51.0% for the 6-digit controls. We use these in-sample patents in the subsequent analysis.

Once the relevant control patents are identified, we can construct the counterfactual citations, with which we compare the actual citations, as follows. For each observed cited-citing relationship, we define an *admissible patent set* that consists of the citing and control patents at the 3- or 6-digit level, i.e., the patents that either actually cited or could have cited the originating patent. We then allocate a counterfactual citation between the originating patent and a patent that is randomly drawn from the corresponding admissible patent set.

(Kendall and Moran, 1963). The median and average of within-area distances for census places are 1.3 km and 1.7 km, respectively.

[13] The way we control patents in the K- and L-density estimations is also similar to that in the DD regressions in Belenzon and Schankerman (2013) and Kerr and Kominers (2015).

3.4 Actual versus counterfactual citations

Table 2 (a) reports the descriptive statistics for the actual citations, where we use $390,104$ citing patents in Table 1. As shown, both the minimum and median distance distributions defined in Section 3.2 are right skewed, i.e., the mass of each distribution is concentrated in short distances. Furthermore, the distribution of minimum citation distances is more right skewed than that of median citation distances.[14]

Insert Table 2

Table 2 (b) illustrates the descriptive statistics for the counterfactual citations, where we report the averages for 1000 draws from $33,472,826$ control patents. Comparing panels (a) and (b) of Table 2, we find that the counterfactual citation distances tend to be greater than the actual citation distances for both the minimum and median cases. Since the actual distributions are more right skewed than the counterfactual ones, the former suffer more from the first-order downward bias at short distances generated by K-density estimations than the latter. Thus, the boundary bias is not offset even when we take the difference between the actual and counterfactual K-densities for citation distances. We provide this statistical property in Appendix A.2 and illustrate the impact of this boundary bias on the K-density tests in Section 4.

4 Estimation

We now describe the estimation procedure in more detail and provide the estimation results. We start with the K- and L-density estimations using the aggregate data and illustrate the first-order bias at short distances generated by the K-density estimation. We then turn to the disaggregate analysis at the patent class level. We construct class-specific localization tests based on K- and L-density estimators and summarize the localization results. We first report the localization results *for all distances* for the sake of completeness, and then illustrate our main results *for short distances.* We finally examine the robustness of the results by considering different binwidths and bandwidths.

[14] We find that the median citation distance is larger than the minimum citation distance at every percentile point. Table 2 (a) reports the distances at several lower percentile, namely, 1, 5, 10, 25, 50 percentile, points.

4.1 K-density and L-density estimations

It is well known that the K-density estimators can be severely biased when the domain of the data is bounded. As mentioned above and as shown in Appendix A.1, the first-order boundary bias caused by the K-density estimation still remains even when adopting the reflection method. In contrast, the L-density estimators have the second-order bias only. We begin with the analysis by illustrating the performance of these density estimators for all distances. We then focus on the behavior at and near the boundary.

4.1.1 Density estimations for all distances

Figure 1 (a) presents the density estimation results — the K-density estimates with and without reflection, and the L-density estimate — for the actual citation distances. For comparison purposes, we superimpose the histogram generated by the first step of the L-density estimation. To obtain these results, we stick to the standard practice in the existing studies. More precisely, as to the K-density estimation, we follow Duranton and Overman (2005) who use the Gaussian kernel function and the rule-of-thumb (ROT) bandwidth h_{S} in Silverman (1986, Section 3.4.2). With regard to the first step of the L-density estimation, we use the binwidth $\Delta = 2\sigma(N^c)^{-1/2}$ for binning as in McCrary (2008). For the LL regression in the second step, we choose the triangular kernel function and employ the bandwidth h_{CV} obtained from the cross-validation method as in Fan and Gijbels (1996).[15,16] We retain them when conducting localization tests in Section 4.3, and relegate a number of robustness checks on the binwidth and bandwidth selections to Section 4.4.

Comparison of the density estimates reveals that the K-density estimates, with and without reflection, and the L-density estimate are quite similar at distances greater than about 300 km. This result is in accordance with the theoretical prediction that the order of biases differs only at and near the boundary between the two estimators. In what follows, we thus focus on the behaviors of the estimated densities within shorter distances between 0 km and 300 km in Figure 1 (b), which is an enlargement of Figure 1 (a).

[15]It is widely recognized that the choice of the kernel function is not so important for the performance of the resulting LL regression (see Fan and Gijbels, 1996, p.76). Yet, Cheng et al. (1997) prove that the triangular kernel, defined as $f_T(x) = 1 - |x|$ for $x \in (-1, 1]$ and $f_T(x) = 0$ otherwise, is optimal at the boundary. We thus use it in the subsequent analysis.

[16]More concretely, let $\hat{L}^c_h(d)$ denote the L-density estimator in equation (4), involving a specific value of the bandwidth parameter h. For each $r = 1, \cdots, R$, we use data $\{d_s, c_s\}_{s\neq r}$ to build a "leave-one-out" L-density estimator $\hat{L}^c_{h,-r}(d)$ and then determine the bandwidth parameter h_{CV} so as to minimize $(1/R)\sum_{r=1}^{R}[c_r - \hat{L}^c_{h,-r}(d_r)]^2 w(d_r)$, where $w(d_r)$ is a weight function.

4.1.2 Density estimations at and near the boundary

Figure 1 (b) illustrates the estimated densities at and near the boundary. Several remarks are in order. First, the K-density estimate without reflection is substantially lower than that with reflection and the L-density estimate at the boundary. This can be explained by the fact that the K-density estimator, if not corrected by a reflection method, is inconsistent at short distances (Jones, 1993, p.136; Simonoff, 1996, Section 3.2; See also Appendix A.1).

Second, the K-density estimate with reflection is located under the L-density estimate at the boundary. The discrepancy between the two density estimates can be ascribed to the difference in the order of boundary biases, which is the first order for the K-density with reflection (Jones, 1993, p.137; See also Appendix A.1) and the second order for the L-density (Cheng, 1994, Theorem 3). Interestingly, the underestimation at the boundary induces overestimation near the boundary since the densities must sum to one over the entire domain of distances. Indeed, the K-density estimate even with reflection underestimates severely at the boundary up to around 45 km. This underestimation at the boundary is compensated by the overestimation near the boundary, which is at distances between 45 km and 300 km in Figure 1 (b).

Third, the slope of the density at the boundary varies across estimators. It is positive, zero and negative for the K-density without reflection, the K-density with reflection and the L-density, respectively. Indeed, without reflection, the derivative of the K-density estimate at the boundary must be theoretically positive since there is no data with negative distances, while with reflection, the estimate using the symmetric and differentiable Gaussian kernel is subject to the constraint that the derivative at the boundary is zero. By contrast, the L-density estimation does not impose such restrictions on the derivative.

Last, although the L-density estimation requires pre-binning, the estimated L-density does not perfectly mimic the histogram, especially at and near the boundary. The difference arises because the L-density estimator at distance d, which by construction smooths the histogram obtained from binning in the first step, puts some non-negative weights on distances away from d in the second step, as seen from equations (3) and (4).

4.2 Localization tests and indices

We now address whether knowledge spillovers as evidenced by patent citations are localized. To this end, we start with the localization measures γ_{KD} and γ_{LD}, which are based on the K-

and L-density estimations, respectively. Those measures allow us to compare the estimated density of actual citation distances to that of counterfactual citation distances. However, testing for localization formally requires a *confidence band* of the estimated counterfactual densities. We thus sample counterfactual citations repeatedly from the admissible patent set, and regard the deviation of the actual density from the upper bound of the confidence band as evidence of localization.

To be more specific, let us denote the upper *local* 5% confidence intervals by $\overline{K^c}(d)$ and $\overline{L^c}(d)$ for the K- and L-density estimators, respectively. We run 1000 Monte Carlo simulations, and estimate the counterfactual K- and L-densities for each simulation. Then, $\overline{K^c}(d)$ and $\overline{L^c}(d)$ are given by the 950th of the estimated counterfactual K- and L-densities that are ranked in ascending order. In turn, the upper *global* 5% confidence intervals, denoted by $\overline{\overline{K^c}}(d)$ and $\overline{\overline{L^c}}(d)$, are defined as the identical local confidence intervals such that, when we consider them across all distances between 0 km and the threshold distance $\bar{d}$ km, only 5% hit them. Analogous to equation (6) (resp., (5)), we can say that knowledge spillovers exhibit global localization at 5% significance level for the K-density (resp., L-density) estimators if $\hat{K}^c(d) - \overline{\overline{K^c}}(d) > 0$ (resp., if $\hat{L}^c(d) - \overline{\overline{L^c}}(d) > 0$) for at least one $d \in [0, \bar{d})$. Thus, using the indices of global localization at distance d:

$$\Gamma_{\mathrm{KD}}(d) = \max\{\hat{K}^c(d) - \overline{\overline{K^c}}(d), 0\}$$
$$\Gamma_{\mathrm{LD}}(d) = \max\{\hat{L}^c(d) - \overline{\overline{L^c}}(d), 0\},$$

we define the *cross-distance* global localization indices as $\Gamma_{\mathrm{KD}} \equiv \sum_d \Gamma_{\mathrm{KD}}(d)$ and $\Gamma_{\mathrm{LD}} \equiv \sum_d \Gamma_{\mathrm{LD}}(d)$, where the summations are taken up to the median distance $\bar{d}$ of all possible actual and counterfactual citation distances. The K-density (resp., L-density) test detects localization if $\Gamma_{\mathrm{KD}} > 0$ (resp., if $\Gamma_{\mathrm{LD}} > 0$).

As shown in Murata et al. (2014), the extent of knowledge spillovers differs by technology. To take this heterogeneity into account, we detect localization of knowledge spillovers *originating from each patent class*.[17] To this end, we classify all originating patents into different patent classes by their primary class, and examine whether each patent class – to which originating patents belong – displays localization.

Let $\mathcal{A}$ be the set of all patent classes. For each patent class $A \in \mathcal{A}$, we estimate the K- and

[17] Note that citing patents, which cite an originating patent in one patent class, may or may not belong to that patent class. Put differently, we allow for intra- and inter-class knowledge spillovers.

L-densities for the actual and counterfactual citation distances. The Monte Carlo simulation procedure described above provides the class-specific local and global confidence intervals by which localization is tested. Finally, for each patent class A, we denote by $\Gamma^A_{\mathrm{KD}}(d)$ and $\Gamma^A_{\mathrm{LD}}(d)$ the class-specific global localization indices at distance d for the K- and L-density estimators, respectively. The *cross-distance* global localization indices for patent class A are then given by $\Gamma^A_{\mathrm{KD}} \equiv \sum_d \Gamma^A_{\mathrm{KD}}(d)$ and $\Gamma^A_{\mathrm{LD}} \equiv \sum_d \Gamma^A_{\mathrm{LD}}(d)$, where the summations are taken up to the class-specific median distance $\overline{d}^A$. Hence, the K-density (resp., L-density) test detects localization of knowledge spillovers originating from patent class A if $\Gamma^A_{\mathrm{KD}} > 0$ (resp., if $\Gamma^A_{\mathrm{LD}} > 0$).[18]

4.3 Localization results

We now summarize the localization results obtained from the K- and L-density methods. We first report the localization results for all distances for the sake of completeness, and then focus on the main results for short distances.

4.3.1 Localization results for all distances

Table 3 summarizes the numbers and the percentages of localized patent classes. Columns 1 and 2 (resp., 3 and 4) report the case with the 3-digit (resp., 6-digit) controls. In each case we consider the K-density with reflection and the L-density. Panels (a) and (b) show the results for the minimum and median citation distances, respectively. In both panels, the K-density tests with the 3-digit (resp., 6-digit) controls detect localization for about 70% (resp., 30%) of feasible patent classes.[19] In the L-density tests, the numbers of localized patent classes increase systematically for both the 3- and 6-digit controls. We may thus conclude that the K-density tests underestimate localized patent classes than the L-density tests.

Insert Tables 3 and 4

In what follows, we report the localization results using the minimum citation distances and the 3-digit controls as the benchmark.[20] Table 4 further presents the top 20 patent classes

[18] To perform the localization tests for each patent class, we adopt a *class-specific bandwidth* and denote it by h^A_{S} or h^A_{CV}.

[19] These results are the same as in Murata et al. (2014). As in the previous studies, we drop patent classes in which originating patents are distributed across less than 10 census places to obtain well-behaved estimated density functions. Thus, the numbers of feasible patent classes differ between the 3- and 6-digit cases.

[20] As shown in Appendix F, the results for the median citation distances are quite similar to our main results for the minimum citation distances given in Section 4.3.2. We also obtain similar results even with the 6-digit controls. However, Henderson et al. (2005) argue that the localization tests using the 6-digit controls are subject to sample selection biases. The sensitivity analysis in Murata et al. (2014), which encompasses

with the highest degrees of localization, measured by Γ^A_{KD} and Γ^A_{LD}. Since about a half of them are not overlapped, the localization indices based on the K- and L-densities provide a different piece of information. Our L-density approach will help to revisit the determinants and shapes of agglomeration and coagglomeration, given that the K-density indices have been incorporated into empirical specifications in that literature (see, e.g., Ellison et al., 2010; Kerr and Kominers, 2015).

To see the difference in the number of localized patent classes between the two tests, let $\mathcal{A}_{\mathrm{KD}}$ and $\mathcal{A}_{\mathrm{LD}}$ be the sets of localized patent classes by the K- and L-density tests, respectively. Then, we take the following three disjoint sets: $\mathcal{A}_{\mathrm{KD}\cap\mathrm{LD}}$, which consists of localized classes detected by both the K- and L-density tests; and $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$ (resp., $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$), which consists of localized classes detected only by the K-density (resp., L-density) test. Given these disjoint sets, we can decompose the sets of localized patent classes as $\mathcal{A}_{\mathrm{KD}} = \mathcal{A}_{\mathrm{KD}\cap\mathrm{LD}} \cup \mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$ and $\mathcal{A}_{\mathrm{LD}} = \mathcal{A}_{\mathrm{KD}\cap\mathrm{LD}} \cup \mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$. Thus, the increase in localized classes arises due to the difference in size between $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ and $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$. For the localization results in Table 3, we find that 61 patent classes belong to $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ while 5 patent classes belong to $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$. Hence, the number of localized classes increases by 56 ($= 61 - 5$).

4.3.2 Localization results at and near the boundary

Turning to our main results, Figure 2 illustrates the difference between the K- and L-density tests using an example of a patent class belonging to $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ (patent class 283: Printed Matter).[21] Panel (a) (resp., panel (b)) depicts the K-density (resp., L-density) test. As can be seen from panel (a) (resp., panel (b)), the actual density does not (resp., does) exceed the upper bound of the global 5% confidence interval. Thus, the K-density test does not detect localization, whereas the L-density test does for this patent class. It is worth emphasizing that the difference in the localization result comes from the difference in the boundary behavior between the two panels.

Insert Figures 2 and 3

How often do we observe this pattern? To obtain our main results at short distances, we now illustrate the spatial pattern of the difference in the number of localized patent classes

the 3- and 6-digit controls, shows that the 6-digit controls correspond to the case where the biases induced by imperfect matching between citing and control patents due to unobserved heterogeneity are infinitely large. See Carlino and Kerr (2015) for more discussion.

[21] As mentioned above, we consider the minimum citation distances here. The result for the median citation distances is similar and can be found in Figure F1 in Appendix F.

between the K- and L-density tests. Figure 3 gives the frequency distribution of patent classes for the distances at which the members in $\mathcal{A}_{\mathrm{LD}\backslash\mathrm{KD}}$ initially localize. This allows us to explore at which distance the K-density tests underestimate the number of localized patent classes as compared to the L-density tests. The top panel of Figure 3 shows that the patent classes in $\mathcal{A}_{\mathrm{LD}\backslash\mathrm{KD}}$ are most frequently observed at shorter distances. Indeed, 40 out of 61 patent classes in $\mathcal{A}_{\mathrm{LD}\backslash\mathrm{KD}}$ are localized within 100 km, which account for 66 % of the number of underestimated patent classes (see the first row of Table 5). We also illustrate in the bottom panel of Figure 3 the spatial distribution of the members in $\mathcal{A}_{\mathrm{KD}\backslash\mathrm{LD}}$, although it does not display a clear pattern.[22]

Insert Table 5

We may thus conclude that the K-density tests underestimate the number of localized patent classes than the L-density tests especially at short distances. Hence, eliminating the first-order bias at short distances is of first-order importance. This is especially so because a large number of recent studies document that knowledge spillovers attenuate and industry localization decays within short distances. Thus, correcting the K-density tests is crucial for a better understanding of localization of economic activity as they have been widely used for detecting localization ever since Duranton and Overman (2005).

4.4 Robustness check

We check the robustness of our main result that the K-density tests underestimate the number of localized patent classes than the L-density tests, especially at short distances.

First, we stick to the baseline cross-validation method and consider different values of h^A_{CV}/Δ. In particular, we impose $h^A_{\mathrm{CV}}/\Delta > \tau$ with $\tau = 10, 20, 30$, which is in line with McCrary (2008) who provides the condition in terms of the ratio between the bandwidth and binwidth to ensure the robustness to different binwidths.

Second, we replace h^A_{CV} with Fan and Gijbels' (1996) ROT bandwidth h^A_{FG} while imposing the same thresholds $\tau = 10, 20, 30$.[23]

[22]Again, we consider the minimum citation distances here. See Figure F2 in Appendix F for the result using the median citation distances. Both results are very similar.

[23]The ROT bandwidth in Fan and Gijbels (1996) is more involved as it requires estimating derivatives nonparametrically. Let g be the estimated fourth-order polynomial for the pre-binned data $\{d_r, c_r\}_{r=1}^R$. Denote by g'' and σ^2 the second derivative and the mean squared error of the polynomial regression. Then, according to Fan and Gijbels (1996), the bandwidth is given by $h_{\mathrm{FG}} = \kappa[\sigma^2 \int w_0(\nu)\mathrm{d}\nu / \sum_{r=1}^R \{[g''(d_r)]^2 w_0(d_r)\}]^{1/5}$, where $w_0(\nu)$ is an indicator function on the interval $[0, d_R]$ with d_R being the midpoint of the R-th distance ring, and the constant κ depends on the kernel function.

Last, to show that our results are not driven by different bandwidth selections between the K- and L-density estimations, we consider Silverman's (1986) ROT bandwidth h_{S}^{A} not only for the K-density tests but also for the L-density tests while allowing the binwidth Δ for the L-density tests to vary to satisfy $h_{\mathrm{S}}^{A}/\Delta > \tau$ with $\tau = 10, 20, 30$.

Note that $h_{\mathrm{CV}}^{A}/\Delta > \tau$ with $\tau = 10, 20, 30$ is not always satisfied. If the condition is not met, we take the following heuristic iterative procedure: Starting from the baseline binwidth Δ_0, we compute the corresponding bandwidth $h_{\mathrm{CV},0}^{A}$ for each patent class A. While the condition is not satisfied for at least one patent class $A \in \mathcal{A}$, we decrease the binwidth by 10 percent iteratively, i.e., $\Delta_{t+1} = 0.9\Delta_t$, to compute the updated bandwidth $h_{\mathrm{CV},t+1}^{A}$. We do the same for $h_{\mathrm{FG}}^{A}/\Delta > \tau$ with $\tau = 10, 20, 30$ if violated.

Insert Table 6

Table 6 presents the robustness results and reports the percentages of patent classes for which the iterative procedure is implemented at least once. With the ROT bandwidth given by Fan and Gijbels (1996), all patent classes satisfy $h_{\mathrm{FG}}^{A}/\Delta > 10$, so that no iteration is needed. When the threshold is set at $\tau = 20, 30$, some patent classes violate $h_{\mathrm{FG}}^{A}/\Delta > \tau$ in the case of the 6-digit controls. Furthermore, when the cross-validation method is used, some patent classes do not meet the condition even with the smallest threshold $\tau = 10$. In all cases, however, once the iterative procedure is adopted, we can show that the K-density tests underestimate the percentage of localized patent classes than the L-density tests. Finally, when we use Silverman's (1986) ROT bandwidth not only for the K-density tests but also for the L-density tests, we obtain fairly similar results. Hence, our results for all distances do not seem to be affected by the choice of binwidth and bandwidth parameters.

Insert Figure 4

Turning to the robustness of our main result at short distances, Figures 4 (a), (b), and (c) replicate Figure 3 using the cross-validation method with $h_{\mathrm{CV}}^{A}/\Delta > 10$, the ROT bandwidth by Fan and Gijbels (1996) with $h_{\mathrm{FG}}^{A}/\Delta > 10$, and the ROT bandwidth by Silverman (1986) with $h_{\mathrm{S}}^{A}/\Delta > 10$, respectively.[24] Again, we can see clear evidence on the downward biases of the K-density tests within short distances: We find that (a) 66 %, (b) 71 %, and (c) 76 % of the patent classes underestimated by the K-density tests fall within 100 km (see Table 5).[25] Since these numbers are similar to the one shown in the benchmark, our main result that the

[24]We obtain quite similar results for different values of τ. The results are available upon request.

[25]The results for the median citation distances can be found in the bottom half of Table 5 and Figure F3.

K-density tests underestimate the number of localized patent classes than the L-density tests *especially at short distances* is not sensitive to the binwidth and bandwidth selections.

5 Concluding remarks

Combining the existing discretizing method and smoothing technique, we propose a new approach to testing for localization using micro-geographic data. Our localization measure is based on local linear density estimators that have a better boundary property, and thus suitable for analyzing localization that generally occurs within short distances.

Our L-density approach can be applied to flows of people, goods, and ideas. In this paper we illustrate the case of knowledge spillovers using the data on citation flows. The analysis can be readily used for commuting and commodity flows, as well as for industry localization and co-localization.

When applied to industry localization, our localization test satisfies all five criteria summarized in Duranton and Overman (2005), namely it (i) is comparable across industries; (ii) controls for the overall agglomeration of manufacturing; (iii) controls for industrial concentration; (iv) is unbiased with respect to scale and aggregation; and (v) gives an indication of the significance of the results.

While having these desirable properties, the L-density approach allows us to address both the first-order biases at short distances and the issue that localization results differ depending on the way of discretizing the micro-geographic data. Our results suggest that coping with these problems is crucial for a better understanding of to what extent people, goods, and ideas are localized in the spatial economy.

References

[1] Arzaghi, M. and J.V. Henderson (2008), "Networking off Madison Avenue", *Review of Economic Studies* 75, 1011-1038.

[2] Belenzon, S. and M. Schankerman (2013) "Spreading the word: Geography, policy, and knowledge spillovers", *Review of Economics and Statistics* 95, 884-903.

[3] Buzard, K., G.A. Carlino, R.M. Hunt, J.K. Carr and T.E. Smith (2015), "Localized knowledge spillovers: Evidence from the agglomeration of American R&D Labs and patent data", *Working Paper* #15-03, Federal Reserve Bank of Philadelphia.

[4] Carlino, G.A. and W.R. Kerr (2015), "Agglomeration and innovation", *Handbook of Regional and Urban Economics* 5, 349-404.

[5] Cheng, M.-Y. (1994), "On boundary effects of smooth curve estimators (dissertation)", *Mimeo Series* #2319, Institute of Statistics, University of North Carolina.

[6] Cheng, M.-Y., J. Fan, and J.S. Marron (1997), "On automatic boundary corrections", *The Annals of Statistics* 25, 1691-1708.

[7] Davidson, R. and J.G. MacKinnon (2004), *Econometric Theory and Methods* (Oxford: Oxford University Press).

[8] Duranton, G. and H. Overman (2005), "Testing for localization using micro-geographic data", *Review of Economic Studies*, 72: 1077-1106.

[9] Duranton, G. and H. Overman (2008), "Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data", *Journal of Regional Science*, 48: 213-243.

[10] Ellison, G.D. and E.L. Glaeser (1997), "Geographic concentration of in US manufacturing industries: A dartboard approach", *Journal of Political Economy*, 105: 889-927.

[11] Ellison, G.D., E.L. Glaeser, and W.R. Kerr (2010), "What causes industry agglomeration? Evidence from coagglomeration patterns", *American Economic Review*, 100: 1195-1213.

[12] Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications* (London: Chapman & Hall).

[13] Gehlke, C.E. and K. Biehl (1934), "Certain effects of grouping upon the size of the correlation coefficient in census tract material", *Journal of the American Statistical Association* 29: 169-170.

[14] Hall, B., A. Jaffe, and M. Trajtenberg (2001), "The NBER patent citation data file: Lessons, insights and methodological tools", *NBER Working Paper* #8498.

[15] Henderson, R., A. Jaffe, and M. Trajtenberg (2005) "Patent citations and the geography of knowledge spillovers: A reassessment: Comment." *American Economic Review* 95, 461-464.

[16] Jaffe, A., M. Trajtenberg, and R. Henderson (1993), "Geographic localization of knowledge spillovers as evidenced by patent citations", *Quarterly Journal of Economics* 108: 577-598.

[17] Jones, M.C. (1993), "Simple boundary correction for density estimation", *Statistics and Computing* 3: 135-146.

[18] Kendall, M. and P. Moran (1963), *Geometrical Probability* (London: Charles Griffin & Company Limited).

[19] Kerr, W.R. and S.D. Kominers (2015), "Agglomerative forces and cluster shapes", *Review of Economics and Statistics* 97, 877-899.

[20] Marron, J.S. and D. Ruppert (1994), "Transformations to reduce boundary bias in kernel density estimation", *Journal of the Royal Statistical Society. Series B* 56: 653-671.

[21] McCrary, J. (2008), "Manipulation of the running variable in the regression discontinuity design: A density test", *Journal of Econometrics* 142, 698-714.

[22] 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: 967-985.

[23] Nadaraya, E.A. (1964), "On estimating regression", *Theory of Probability and its Applications* 9: 141-142.

[24] Rosenthal, S. and W. Strange (2003), "Geography, industrial organization, and agglomeration", *Review of Economics and Statistics* 85, 377-393.

[25] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis* (New York: Chapman and Hall).

[26] Simonoff, J.S. (1996), *Smoothing Methods in Statistics* (New York: Springer-Verlag).

[27] Trajtenberg, M., G. Shiff, and R. Melamed (2006), "The NAMES GAME: Harnessing inventors' patent data for economic research", *NBER Working Paper* #12479.

[28] Wand, M.P. and M.C. Jones (1995), *Kernel Smoothing* (London: Chapman & Hall).

[29] Watson, G.S. (1964), "Smooth regression analysis", *Sankhyā: The Indian Journal of Statistics, Series A* 26: 359-372.

Table 1: Sample patent sizes

	Total	3-digit	6-digit
Originating patents	115,905	107,561	59,168
(Percent)	(100.00)	(92.64)	(51.04)
Citing patents	647,983	390,104	120,876
(Percent)	(100.00)	(60.20)	(18.65)
Control patents	—	33,472,826	941,532

Note: Column 1 reports the number of originating patents (that were cited at least once by other U.S. patents) and the number of citing patents (that cited the originating patents). Self-citations are excluded. Columns 2 and 3 report the numbers of originating and citing patents having at least one control, as well as the numbers of control patents. The citing patents with no controls assigned (and their originating patents) are dropped out of the sample.

Table 2: Descriptive statistics of minimum and median citation distances for actual and counterfactual citations

Citation			Moments			Percentiles				
distances	Min	Max	Mean	Std Dev	Skewness	1st	5th	10th	25th	50th
(a) Actual citations										
Minimum	0.47	4513.48	1373.67	1240.81	0.89	4.00	19.10	75.72	375.81	973.77
Median	0.49	4517.25	1483.43	1264.69	0.79	11.22	37.02	123.05	441.82	1098.28
(b) Counterfactual citations										
Minimum	0.42	4516.11	1472.08	1241.77	0.79	6.53	42.78	149.88	453.12	1089.60
Median	0.52	4539.75	1577.99	1261.29	0.69	15.96	70.03	197.76	528.93	1187.97

Note: To compute the actual citation distances, we use 390,104 citing patents in Table 1. For the counterfactual citation distances, we report the averages for 1000 draws from 33,472,826 control patents in Table 1. We measure the distance between an originating patent and an actual (or counterfactual) citing patent by the great-circle distance between the census place of the originating inventor and that of the citing inventor. When there is more than one inventor of a patent, we compute all possible great-circle distances between the census places of the originating inventors and those of the citing inventors. We then consider two types of distances – minimum and median distances – between the cited and citing patents. All distances are in kilometers.

Table 3: Localization results for all distances

	3-digit control		6-digit control	
	K-density	L-density	K-density	L-density
	(1)	(2)	(3)	(4)
All classes*	384	384	360	360
(a) Minimum citation distance				
Localized patent classes	273	329	109	149
Percentage of localized patent classes	71.09%	85.68%	30.28%	41.39%
(b) Median citation distance				
Localized patent classes	275	334	109	165
Percentage of localized patent classes	71.61%	86.98%	30.28%	45.83%

Note: For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method and McCrary's (2008) binwidth. (*) The numbers of feasible patent classes differ between the 3- and 6-digit control cases because the patent classes in which originating patents are distributed across less than 10 census places are dropped out of the sample.

Table 4: Top 20 localized patent classes by Γ^A_{KD} and Γ^A_{LD}

K-density case (Γ^A_{KD})

Patent class	3-digit code	Rank by Γ^A_{KD}	Rank by Γ^A_{LD}
Hydraulic and earth engineering	405	1	1
Boots, shoes, and leggings	36	2	3
Butchering	452	3	14
Surgery	606	4	4
Apparel apparatus	223	5	6
Communications, electrical: acoustic wave systems and devices	367	6	37
Land vehicles: bodies and top	296	7	2
Sugar, starch, and carbohydrates	127	8	23
Acoustics	181	9	26
Pipe joints or couplings	285	10	7
Roll or roller	492	11	8
Cutlery	30	12	91
Compositions: ceramic	501	13	29
Implements or apparatus for applying pushing or pulling force	254	14	62
Fences	256	15	5
Fluid sprinkling, spraying, and diffusing	239	16	38
Expanded, threaded, driven, headed, tool-deformed, or locked-threaded fastener	411	17	22
Special receptacle or package	206	18	28
Article dispensing	221	19	65
Fluid handling	137	20	71

L-density case (Γ^A_{LD})

Patent class	3-digit code	Rank by Γ^A_{LD}	Rank by Γ^A_{KD}
Hydraulic and earth engineering	405	1	1
Land vehicles: bodies and tops	296	2	7
Boots, shoes, and leggings	36	3	2
Surgery	606	4	4
Fences	256	5	15
Apparel apparatus	223	6	5
Pipe joints or couplings	285	7	10
Roll or roller	492	8	11
Sewing	112	9	35
Chemistry: electrical current producing apparatus, product, and process	429	10	85
Railways	104	11	190
Resilient tires and wheels	152	12	111
Safes, bank protection, or a related device	109	13	89
Butchering	452	14	3
Imperforate bowl: centrifugal separators	494	15	198
Metal deforming	72	16	26
Textiles: fluid treating apparatus	68	17	24
Winding, tensioning, or guiding	242	18	22
Prosthesis, parts thereof, or aids and accessories therefor	623	19	71
Perfume compositions	512	20	49

Note: For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method and McCrary's (2008) binwidth.

Table 5: Localization results for short distances

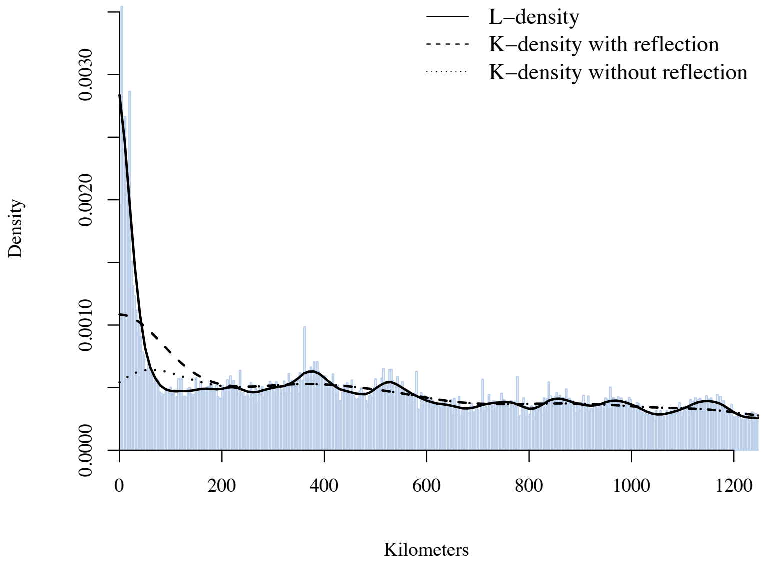
		$\lvert\mathcal{A}_{\text{LD}\setminus\text{KD}}\rvert$ for all d (1)	$\lvert\mathcal{A}_{\text{LD}\setminus\text{KD}}\rvert$ for $d \in [0, 100]$ (2)	% short distances (3)
(a) Minimum citation distance				
Cross-validation (no restriction)	Fig 3	61	40	66%
Cross-validation ($h^A_{\text{CV}}/\Delta > 10$)	Fig 4(a)	61	40	66%
Fan and Gijbels' ROT ($h^A_{\text{FG}}/\Delta > 10$)	Fig 4(b)	49	35	71%
Silverman's ROT ($h^A_{\text{S}}/\Delta > 10$)	Fig 4(c)	46	35	76%
(b) Median citation distance				
Cross-validation (no restriction)	Fig F2	63	40	63%
Cross-validation ($h^A_{\text{CV}}/\Delta > 10$)	Fig F3(a)	62	39	63%
Fan and Gijbels' ROT ($h^A_{\text{FG}}/\Delta > 10$)	Fig F3(b)	48	29	60%
Silverman's ROT ($h^A_{\text{S}}/\Delta > 10$)	Fig F3(c)	49	30	61%

Note: Column (1) (resp., Column (2)) reports the number $\lvert\mathcal{A}_{\text{LD}\setminus\text{KD}}\rvert$ of localized patent classes detected only by L-density tests for all distances (resp., for short distances between 0 km and 100 km). Column (3) provides the percentage of underestimated patent classes that fall within the domain between 0 km and 100 km.

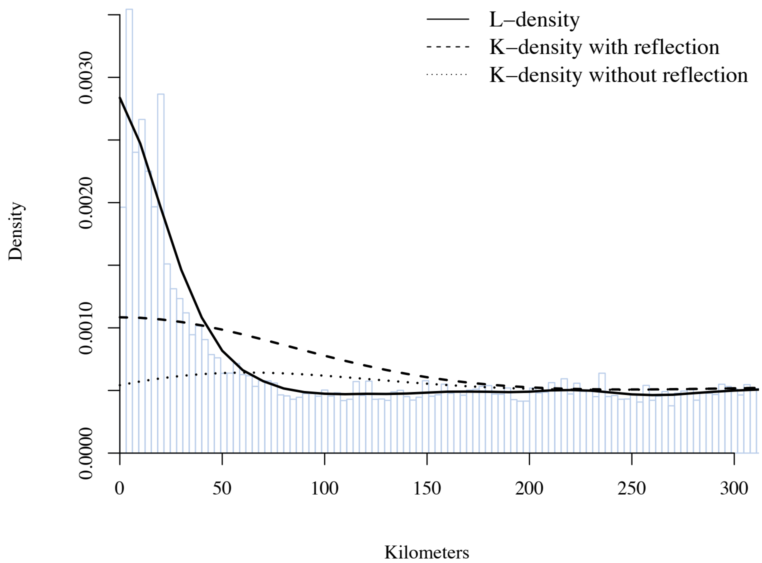
Table 6: Robustness check: Localization results for all distances

	3-digit control		6-digit control	
	% Localized	% Iterated†	% Localized	% Iterated†
K-density	71.09%	–	30.28%	–
L-density				
(a) Cross-validation:				
$h^A_{\mathrm{CV}}/\Delta > 10$	85.68%	8.33%	40.83%	24.17%
$h^A_{\mathrm{CV}}/\Delta > 20$	85.42%	44.27%	40.56%	44.72%
$h^A_{\mathrm{CV}}/\Delta > 30$	85.42%	59.38%	40.28%	51.39%
(b) Fan and Gijbels' ROT:				
$h^A_{\mathrm{FG}}/\Delta > 10$	82.03%	0.00%	38.61%	0.00%
$h^A_{\mathrm{FG}}/\Delta > 20$	82.03%	0.00%	38.61%	1.11%
$h^A_{\mathrm{FG}}/\Delta > 30$	82.03%	0.00%	38.33%	25.83%
(c) Silverman's ROT:				
$h^A_{\mathrm{S}}/\Delta > 10$	81.25%	–	37.22%	–
$h^A_{\mathrm{S}}/\Delta > 20$	80.73%	–	37.78%	–
$h^A_{\mathrm{S}}/\Delta > 30$	81.00%	–	37.78%	–

Note: For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt (a) the cross-validation method, (b) Fan and Gijbels' (1996) rule-of-thumb bandwidth, and (c) Silverman's (1986) rule-of-thumb bandwidth. In cases (a) and (b), $h^A_{\mathrm{CV}}/\Delta > 10, 20, 30$ and $h^A_{\mathrm{FG}}/\Delta > 10, 20, 30$ hold for all patent classes $A \in \mathcal{A}$. Note that while the inequality condition is not satisfied for all patent classes, the binwidth, Δ, is iteratively decreased by 10 percent for that condition to be satisfied. The percentage of patent classes for which the iterative procedure is implemented at least once is reported. In case (c), we choose the binwidth, Δ, such that $h^A_{\mathrm{S}}/\Delta > 10, 20, 30$ holds for all patent classes $A \in \mathcal{A}$.



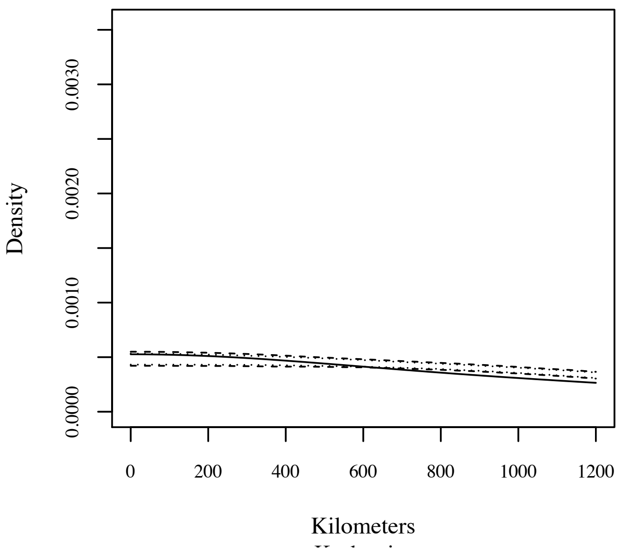
(a) Density estimation results for 0 km to 1200 km



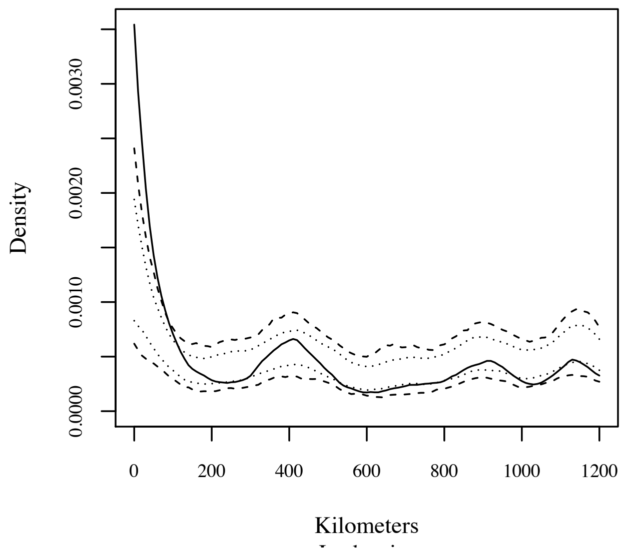
(b) Density estimation results for 0 km to 300 km

Figure 1: Density estimation results at the aggregate level.

Panels (a) and (b) present the same density estimation results for different domains of distances. For the K-density estimation, we use Silverman's (1986) rule-of-thumb bandwidth. For the second step of the L-density estimation, we adopt the cross-validation method. To depict the histogram we use McCrary's (2008) binwidth, which constitutes the first step of the L-density estimation.



(a) K-density test



(b) L-density test

Figure 2: K-density and L-density tests using patent class 283 (Printed Matter).

The solid curves are the estimated densities for the actual citations. The dashed (resp., dotted) curves correspond to the global (resp., local) confidence intervals obtained from the counterfactual citations. For the K-density test, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density test, we adopt the cross-validation method and McCrary's (2008) binwidth.

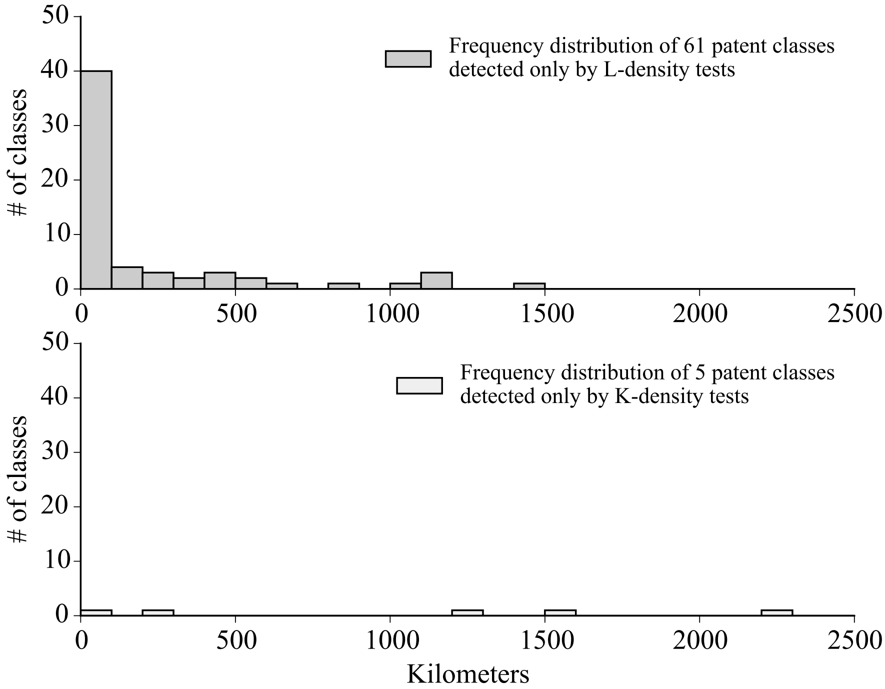
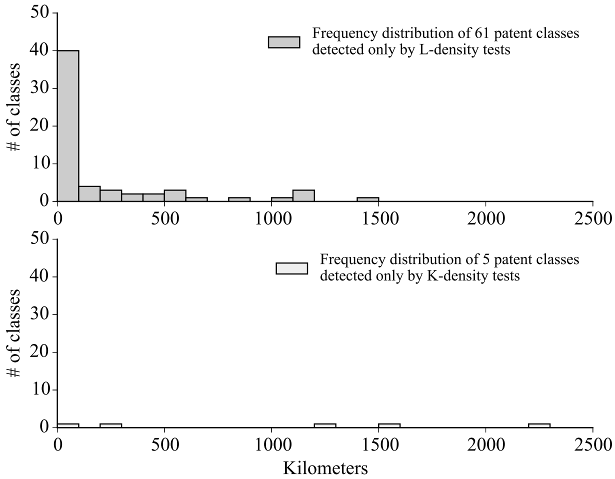
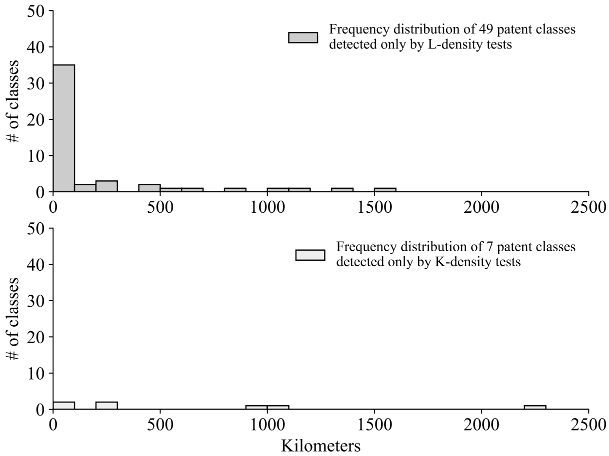


Figure 3: Localization results at and near the boundary.

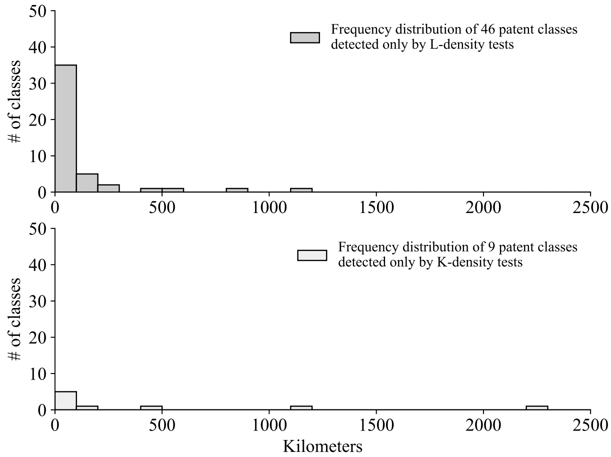
The top (resp., bottom) panel depicts the frequency distributions of patent classes for the distances at which the members in $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ (resp., $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$) initially display localization. The bars in dark grey (resp., light grey) are the frequencies of localized patent classes detected only by the L-density (resp., K-density) tests. For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method and McCrary's (2008) binwidth.



(a) Cross-validation ($h^A_{\mathrm{CV}}/\Delta > 10$)



(b) Fan and Gijbels' ROT ($h^A_{\mathrm{FG}}/\Delta > 10$)



(c) Silverman's ROT ($h^A_{\mathrm{S}}/\Delta > 10$)

Figure 4: Robustness check: Localization results at and near the boundary.

In all cases (a), (b), and (c), the top (resp., bottom) panel depicts the frequency distributions of patent classes for the distances at which the members in $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ (resp., $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$) initially display localization. The bars in dark grey (resp., light grey) are the frequencies of localized patent classes detected only by the L-density (resp., K-density) tests. For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method in panel (a), Fan and Gijbels' (1996) rule-of-thumb bandwidth in panel (b), and Silverman's (1986) rule-of-thumb bandwidth in panel (c). In cases (a) and (b), $h^A_{\mathrm{CV}}/\Delta > 10$ and $h^A_{\mathrm{FG}}/\Delta > 10$ hold for all patent classes $A \in \mathcal{A}$. Note that while the inequality condition is not satisfied for all patent classes, the binwidth, Δ, is iteratively decreased by 10 percent for that condition to be satisfied. In case (c), we choose the binwidth, Δ, such that $h^A_{\mathrm{S}}/\Delta > 10$ holds for all patent classes $A \in \mathcal{A}$.

Appendix A: Some boundary properties of K-density estimators

For illustrative purposes, we consider a kernel function $f((d - d_{ij})/h)$ defined on $[-1, 1]$. However, the following results can be extended to other kernel functions. Assume that the density φ is being estimated at some distance $d = \delta h$ either within the boundary region $(0 \leq \delta < 1)$ or outside the boundary region $(\delta \geq 1)$.

A.1. K-density estimators (even when adopting the reflection method) have the first-order bias at short distances.

Without reflection. The expectation of the K-density estimator has the Taylor series representation (e.g., Jones, 1993, p.136; Simonoff, 1996, Section 3.2):

$$\mathrm{E}\left[\hat{K}(d)\right] = a_0(\delta)\varphi(d) - ha_1(\delta)\varphi'(d) + O(h^2), \tag{9}$$

where $a_l(\delta) = \int_{-1}^{\min\{1,\delta\}} u^l f(u)du$.

If the distance is away from the boundary $(\delta \geq 1)$, the bias is of the second order $O(h^2)$ because $a_0(\delta) = 1$ and $a_1(\delta) = 0$ in (9).

However, in the boundary region $(0 \leq \delta < 1)$, the K-density estimator is inconsistent due to $a_0(\delta) < 1$ and has the first-order bias $O(h)$ since $a_1(\delta) \neq 0$.

To obtain a consistent estimate, one could renormalize the K-density estimator by dividing $\hat{K}(d)$ by $a_0(\delta)$. However, the first-order bias $O(h)$ in the boundary region remains because $a_1(\delta)/a_0(\delta) \neq 0$ for $0 \leq \delta < 1$.

With reflection. Jones (1993, p.137) shows the following result for the K-density with the reflection method by Silverman (1986):

$$\mathrm{E}\left[\hat{K}^R(d)\right] = \varphi(d) - 2h\left\{a_1(\delta) + \delta[1 - a_0(\delta)]\right\}\varphi'(d) + O(h^2). \tag{10}$$

Thus, the K-density estimator with reflection is consistent but still has the first-order bias $O(h)$ in the boundary region because $\{a_1(\delta) + \delta[1 - a_0(\delta)]\} \neq 0$ for $0 \leq \delta < 1$.

Since $\lim_{\delta \to 0}\{a_1(\delta) + \delta[1 - a_0(\delta)]\} = a_1(\delta) < 0$, the second term of (10) generates the first-order downward bias at the boundary $d = 0$ by reducing $\mathrm{E}\left[\hat{K}^R(d)\right]$ when $\varphi'(d) < 0$. By continuity, the first-order downward bias remains near the boundary. Furthermore, when $\varphi'(d) < 0$, the larger the absolute value $|\varphi'(d)|$, the greater the first-order downward bias.

A.2. The first-order bias caused by the K-density estimations with reflection is not offset even when we take the difference between the actual and counterfactual distributions.

We use (10) and take the difference between the actual and counterfactual K-density estimators, which yields

$$\mathrm{E}\left[\hat{K}^R(d) - \tilde{K}^R(d)\right] = [\varphi(d) - \tilde{\varphi}(d)] - 2h\left\{a_1(\delta) + \delta[1 - a_0(\delta)]\right\}[\varphi'(d) - \tilde{\varphi}'(d)] + O(h^2). \quad (11)$$

The above expression shows that the first-order bias remains as long as the derivative of actual density $\varphi'(d)$ differs from that of counterfactual density $\tilde{\varphi}'(d)$.

As before, since $\lim_{\delta \to 0}\{a_1(\delta) + \delta[1 - a_0(\delta)]\} = a_1(\delta) < 0$, the second term of (11) generates the first-order downward bias at the boundary $d = 0$ by reducing $\mathrm{E}\left[\hat{K}^R(d) - \tilde{K}^R(d)\right]$ when $[\varphi'(d) - \tilde{\varphi}'(d)] < 0$. By continuity, the first-order downward bias remains near the boundary. Furthermore, when $[\varphi'(d) - \tilde{\varphi}'(d)] < 0$, the larger the absolute value $|\varphi'(d) - \tilde{\varphi}'(d)|$, the greater the first-order downward bias. Although we consider the case where the actual and counterfactual bandwidths are the same in (11), this assumption is made for expositional purposes only, and a similar argument applies to the case with different bandwidths.

Appendix B: Regression-based localization measures

Our localization measure γ_{LD} requires pre-binning and the local linear regression. Therefore, it is natural to think that alternative measures can be derived from nonparametric regressions. Indeed, one could construct a localization measure based on the DD regression. We will show that such a localization measure is essentially the same as γ_{HIST} obtained from binning in the first step of the L-density estimations. Alternatively, one could conduct the local linear regression without pre-binning to obtain a localization measure. In what follows, we explore these possibilities building on the nonparametric regression literature and argue that these approaches have difficulties from either a theoretical or a practical point of view.

Let $\mathcal{N}$ denote the set of all possible pairs of patents. It is worth emphasizing that unlike $\mathcal{N}^c$, the set $\mathcal{N}$ includes pairs of patents that do not have a cited-citing relationship. Thus, the number N of all possible pairs of patents is much larger than the number N^c of citations. Denote by c_{ij} the citation dummy variable which takes one if patent i is cited by patent j, i.e., if $(i, j) \in \mathcal{N}^c$, and zero otherwise, i.e., if $(i, j) \in \mathcal{N} \backslash \mathcal{N}^c$. We start with the regression model as follows:

$$c_{ij} = m(d_{ij}) + \varepsilon_{ij}, \tag{12}$$

where ε_{ij} is an idiosyncratic error term with zero mean.[26] The function m can be viewed as the likelihood of citation between two patents with distance d_{ij} since $m(d_{ij}) = \mathrm{E}(c_{ij}|d_{ij}) = \mathrm{Prob}(c_{ij} = 1|d_{ij})$.[27]

B.1. A special case: Distance-ring dummy-variable regression

We first analyze a special case in which a functional form for m in the regression model (12) is known. A popular choice of m is the sum of dummy variables for distance rings (e.g., Belenzon and Schankerman, 2013; Kerr and Kominers, 2015). Replacing $m(d_{ij})$ in (12) with the sum of the dummy variables, the DD regression model is given by

$$c_{ij} = \sum_{r=1}^{R} \mu_r D_{ijr} + \varepsilon_{ij}, \tag{13}$$

where D_{ijr} is the dummy variable defined as (1) and μ_r is the coefficient on D_{ijr}.[28] Let $N_r = \sum_{(i,j)\in\mathcal{N}} D_{ijr}$ be the number of all possible pairs of patents in distance ring r. The OLS estimator for the DD regression model can then be expressed as follows (see Appendix D)

$$\hat{\mu}_r = \frac{N_r^c}{N_r}, \tag{14}$$

where $N_r^c = \sum_{(i,j)\in\mathcal{N}^c} D_{ijr}$ is the number of citations in distance ring r as in Section 2.1.

Having analyzed the observed sample, we turn to the DD regression using a control sample. For each observed cited-citing relationship $(i, j) \in \mathcal{N}^c$, we identify a set of control patents, from which we randomly draw a control patent to generate a counterfactual citation. As in the

[26]In the context of industry localization, one may reinterpret $\mathcal{N}$ as the set of all possible pairs of establishments in all manufacturing industries and d_{ij} as the distance between establishments i and j. Setting $c_{ij} = 1$ if establishments i and j are in the same industry and $c_{ij} = 0$ otherwise, then yields essentially the same localization test as in Duranton and Overman (2005) once the regression model (12) is estimated via the Nadaraya-Watson regression (Nadaraya, 1964; Watson, 1964). We will come back to this point in Appendices B.2 and E.

[27]There is another variant of the regression model based on different observation units. However, as shown in Appendix C, this alternative approach leads to the same function m as in the regression model (12). Hence, without loss of generality, we focus on the regression model (12).

[28]A few remarks are in order. First, $\sum_{r=1}^{R} D_{ijr} = 1$ holds for any patent pair (i, j). We thus exclude a constant term from (13). Second, the dummy variables are orthogonal. Letting D_r be a column vector of the r-th distance ring with a typical element D_{ijr}, we have $D_r' D_s = 0$ for distance rings r and $s \neq r$. Finally, the DD regression model is a linear probability model in the context of binary dependent variable model.

main text, let $\tilde{\mathcal{N}}^c$ be the set of counterfactual cited-citing relationships. By construction, the total number of counterfactual citations equals that of actual citations, i.e., $\tilde{N}^c = N^c$. Denote by $\tilde{c}_{ij}$ a counterfactual citation dummy, which takes one if $(i,j) \in \tilde{\mathcal{N}}^c$ and zero otherwise. Then, as shown in Appendix D, the OLS estimator for the r-th distance ring is given by $\tilde{\mu}_r = \tilde{N}_r^c/\tilde{N}_r = \tilde{N}_r^c/N_r$, where the counterfactual number $\tilde{N}_r$ of all possible pairs of patents in the r-th distance ring is the same as the denominator N_r in (14) because the set $\mathcal{N}$ of all possible patent pairs that we use to estimate $\hat{\mu}_r$ and $\tilde{\mu}_r$ is common.

Finally, we define the measure of localization based on the DD regressions at distance $d \in (d_r - \Delta/2, d_r + \Delta/2]$ as the difference between the actual and counterfactual coefficients:

$$\gamma_{\mathrm{DD}}(d) \equiv \hat{\mu}_r - \tilde{\mu}_r = w_{\mathrm{DD}} \times \left[N_r^c - \tilde{N}_r^c\right], \tag{15}$$

where $w_{\mathrm{DD}} \equiv 1/N_r$. Note that, for $\hat{\mu}_r$ and $\tilde{\mu}_r$ to be comparable, the binwidths for the observed and control samples are assumed to be the same here, i.e., $\Delta = \tilde{\Delta}$. Under this assumption, the localization measure (8) based on the histograms at distance $d \in (d_r - \Delta/2, d_r + \Delta/2]$ can be rewritten as $\gamma_{\mathrm{HIST}}(d) \equiv c_r - \tilde{c}_r = w_{\mathrm{HIST}} \times [N_r^c - \tilde{N}_r^c] = (w_{\mathrm{HIST}}/w_{\mathrm{DD}}) \times \gamma_{\mathrm{DD}}(d)$, where $w_{\mathrm{HIST}} \equiv 1/(N^c\Delta)$. Thus, the localization measure (15) obtained from the DD regressions is proportional to the localization measure (8) derived from binning in the first step of the L-density estimations. Hence, these two localization measures provide qualitatively the same localization results: Knowledge spillovers are localized at distance $d \in (d_r - \Delta/2, d_r + \Delta/2]$ if the difference $N_r^c - \tilde{N}_r^c$ is positive and significantly different from zero.

B.2. A general case: Local linear regression

We consider a general case of the regression model (12) without imposing a specific functional form on m. A typical estimator is then the local linear (LL) regression estimator.[29] Formally, the LL regression estimator $\hat{m}(d)$ is given by $\hat{\beta}_0$ that solves the minimization problem:

$$\min_{\{\beta_0,\beta_1\}} \sum_{(i,j)\in\mathcal{N}} \left[c_{ij} - \beta_0 - \beta_1(d_{ij} - d)\right]^2 f((d_{ij} - d)/h) \tag{16}$$

[29] Another estimator that is often used in the literature is the Nadaraya-Watson (NW) regression estimator. We show in Appendix E that the localization measure obtained from the NW regressions is essentially the same as γ_{KD} that is derived from the K-density estimations.

as follows:

$$\hat{m}(d) = \frac{\hat{\mathbb{S}}(d)}{\hat{\mathbb{S}}_2(d)\hat{\mathbb{S}}_0(d) - \hat{\mathbb{S}}_1(d)^2}, \tag{17}$$

where

$$\hat{\mathbb{S}}(d) = \sum_{(i,j)\in\mathcal{N}^c} \left[\hat{\mathbb{S}}_2(d) - \hat{\mathbb{S}}_1(d)(d_{ij} - d)\right] f((d_{ij} - d)/h) \tag{18}$$

$$\hat{\mathbb{S}}_q(d) = \sum_{(i,j)\in\mathcal{N}} (d_{ij} - d)^q f((d_{ij} - d)/h) \tag{19}$$

(see, e.g., Wand and Jones, 1995). As in the previous cases, we consider the estimator $\tilde{m}(d)$ for counterfactual citations. Taking the difference between $\hat{m}(d)$ and $\tilde{m}(d)$, the localization measure based on the LL regressions can be defined as follows:

$$\gamma_{\mathrm{LL}}(d) \equiv \hat{m}(d) - \tilde{m}(d) = w_{\mathrm{LL}} \times \left[\hat{\mathbb{S}}(d) - \tilde{\mathbb{S}}(d)\right], \tag{20}$$

where $w_{\mathrm{LL}} \equiv [\hat{\mathbb{S}}_2(d)\hat{\mathbb{S}}_0(d) - \hat{\mathbb{S}}_1(d)^2]^{-1}$. Note that the actual and counterfactual bandwidths are assumed to be the same here. One may think of the difference $\hat{\mathbb{S}}(d) - \tilde{\mathbb{S}}(d)$ as a localization measure. We will discuss such a possibility in Appendix B.3.

B.3. Difficulties with regression-based localization measures

In Section 2 we have compared the two different localization measures, γ_{LD} and γ_{KD}, both of which are based on density estimators. We now discuss the relationship between our localization measure γ_{LD} and the localization measures based on nonparametric regressions.

Our localization measure γ_{LD} obtained from the L-density estimations shares a similarity with γ_{HIST} in (8) and γ_{DD} in (15) since they rely on the binwidth Δ. However, the difference is that γ_{LD} depends also on the smoothing parameter h. This allows us to cope with the problem that γ_{HIST} and γ_{DD} are sensitive to the choice of Δ. Furthermore, our measure is similar to γ_{LL} in (20) because both measures are obtained from local linear regressions. Notwithstanding the similarities, these alternative approaches to γ_{LD} have difficulties from either a theoretical or a practical point of view.

Theoretically, the localization measure obtained from DD regressions is essentially the same as that derived from the histograms. However, the histogram estimator has the first-order bias even at interior points unless the derivative of the true density is zero (Simonoff, 1996, Section 2.1). Thus, the localization tests using γ_{HIST} and γ_{DD} tend to underperform those using γ_{KD}. In contrast, the LL regression estimator has the $O(h^2)$ bias at the boundary as well as at interior points (Fan and Gijbels, 1996, Section 3.2). Therefore, the localization measure γ_{LL} could perform better than γ_{KD} that is subject to the $O(h)$ bias at the boundary.

Despite such a desirable boundary property, the LL regression estimator has a drawback in terms of computational time. As shown in equations (17)–(19), the implementation of the LL regression requires the summation $\hat{\mathbb{S}}_q(d)$ with $q = 0, 1, 2$, each of which involves the kernel function evaluation as many times as the sample size N. Thus, when testing for localization using micro-geographic data, the required sample size is often quite large. For example, in the case of the NBER U.S. Patent Citations Data File that we use in this paper, the number of patents eligible for the LL regression analysis is $937,469$, so that the number of observations, which is the number of all possible pairs of patents, can be up to 8.8×10^{11}. Such a large number of observations make estimation infeasible in practice.[30]

It should be noted that a computational burden is much less severe for the K- and L-density estimations. Indeed, the localization measures γ_{KD} and γ_{LD} can be computed without much difficulty. The reason is that the K-density estimator requires the kernel function evaluation only for N^c times rather than N times as shown in (7), where the summation is taken over the set $\mathcal{N}^c$ of patent pairs having a cited-citing relationship, not the set $\mathcal{N}$ of all possible patent pairs. Similarly, as discussed in Section 2, the L-density estimator requires only R ($< N^c$) times, where R is the number of bins. One may think of the difference $\hat{\mathbb{S}}(d) - \tilde{\mathbb{S}}(d)$ in (20) as an alternative localization measure using the LL regressions. However, the computational burden remains heavy since, as shown in (18), the computation of $\hat{\mathbb{S}}(d)$ still involves the summation of $\hat{\mathbb{S}}_q(d)$ with $q = 1, 2$, each of which requires the kernel function evaluation for N times rather than N^c or R times.

Appendix C: An alternative regression model

In this appendix, we consider an alternative regression model with different observation units. Let z and z' be geographic sites (e.g., zip code area or county), and consider

$$\frac{c_{zz'}}{n_{zz'}} = m_s(d_{zz'}) + \varepsilon_{zz'}, \tag{21}$$

[30] Another example can be taken from Duranton and Overman (2005). Since there are $176,106$ establishments in their data, the sample size for the LL regression, i.e., the number of all possible pairs of establishments, will be more than 1.5×10^{10}, where we use $(c_{ij}, d_{ij}) = (c_{ji}, d_{ji})$ because their data is undirectional.

where $c_{zz'}$ and $n_{zz'}$ are the number of citations and the number of patent pairs between z and z', respectively; and $d_{zz'}$ is the distance between sites z and z'. Thus, for each pair of sites (z, z'), the citation rate $c_{zz'}/n_{zz'}$ is regressed on the distance $d_{zz'}$ between the two sites.

It can be verified that the function m_s in (21) is identical to the function m of the regression model (12). To see this, let z_i and z_j be the geographic sites to which patents i and j belong, respectively. Noting that $c_{zz'} = \sum_{\{i|z_i=z\}} \sum_{\{j|z_j=z'\}} c_{ij}$, we readily obtain the result as follows:

$$m_s(d_{zz'}) = \mathrm{E}\left(\frac{c_{zz'}}{n_{zz'}}\middle| d_{zz'}\right) = \frac{1}{n_{zz'}} \sum_{\{i|z_i=z\}} \sum_{\{j|z_j=z'\}} \mathrm{E}(c_{ij}|d_{z_i z_j}) = \frac{1}{n_{zz'}} \sum_{\{i|z_i=z\}} \sum_{\{j|z_j=z'\}} m(d_{z_i z_j}).$$

Since the value of $m(d_{z_i z_j})$ is constant for any i and j such that $z_i = z$ and $z_j = z'$, we obtain $m_s(d_{zz'}) = m(d_{z_i z_j})$. Hence, without loss of generality, we focus on the regression model (12).

Appendix D: The OLS estimators for the DD regressions

Since the regressors are orthogonal in (13), the OLS estimator for distance ring r is given by[31]

$$\hat{\mu}_r = \frac{\sum_{(i,j)\in\mathcal{N}} D_{ijr} c_{ij}}{\sum_{(i,j)\in\mathcal{N}} D_{ijr}^2} = \frac{\sum_{(i,j)\in\mathcal{N}^c} D_{ijr}}{\sum_{(i,j)\in\mathcal{N}} D_{ijr}},$$

where we use $\sum_{(i,j)\in\mathcal{N}} D_{ijr} c_{ij} = \sum_{(i,j)\in\mathcal{N}^c} D_{ijr}$ and $D_{ijr}^2 = D_{ijr}$. Using the definitions of N_r^c and N_r, we obtain (14). Similarly, using the counterfactual citation dummy $\tilde{c}_{ij}$ and the set $\tilde{\mathcal{N}}^c$ of counterfactual citations, the OLS estimator for the r-th distance ring is given by

$$\tilde{\mu}_r = \frac{\sum_{(i,j)\in\mathcal{N}} D_{ijr} \tilde{c}_{ij}}{\sum_{(i,j)\in\mathcal{N}} D_{ijr}^2} = \frac{\sum_{(i,j)\in\tilde{\mathcal{N}}^c} D_{ijr}}{\sum_{(i,j)\in\mathcal{N}} D_{ijr}},$$

where we use $\sum_{(i,j)\in\mathcal{N}} D_{ijr} \tilde{c}_{ij} = \sum_{(i,j)\in\tilde{\mathcal{N}}^c} D_{ijr}$ and $D_{ijr}^2 = D_{ijr}$. Using the properties of $\tilde{N}_r^c$ and $\tilde{N}_r$, we obtain $\tilde{\mu}_r = \tilde{N}_r^c/\tilde{N}_r$.

[31] This result comes from the Frisch-Waugh-Lovell Theorem (see, e.g., Davidson and MacKinnon, 2004, pp.65-66). When all regressors are orthogonal, the coefficient of any single variable in the multiple regression can be expressed as the coefficient of the variable in the simple regression after dropping out all the other independent variables.

Appendix E: On the relationship between K-density estimation and Nadaraya-Watson regression

In general the regression model (12) can be estimated via local polynomial regressions. Let p denote the degree of the polynomial being fit. At distance d, the local polynomial regression estimator $\hat{m}(d)$ is obtained by fitting the p-th order polynomial to the data using weighted least squares. More specifically, $\hat{m}(d)$ is given by $\hat{\beta}_0$ that solves the minimization problem:

$$\min_{\{\beta_0,\cdots,\beta_p\}} \sum_{(i,j)\in\mathcal{N}} \left[c_{ij} - \beta_0 - \beta_1(d_{ij}-d) - \cdots - \beta_p(d_{ij}-d)^p\right]^2 f((d_{ij}-d)/h). \tag{22}$$

The commonly used values for the order of the polynomial are $p = 0$ and 1. In such cases, we have simple explicit formulae (see, e.g., Wand and Jones, 1995). In Appendix B.2, we deal with the case with $p = 1$. When $p = 0$, the value $\hat{\beta}_0$ is called the Nadaraya-Watson (NW) regression estimator and is given by

$$\hat{m}_{\mathrm{NW}}(d) = \frac{\sum_{(i,j)\in\mathcal{N}} f((d_{ij}-d)/h)c_{ij}}{\sum_{(i,j)\in\mathcal{N}} f((d_{ij}-d)/h)}.$$

Noting that $\sum_{(i,j)\in\mathcal{N}} f((d_{ij}-d)/h)c_{ij} = \sum_{(i,j)\in\mathcal{N}^c} f((d_{ij}-d)/h)$, and that $f((d_{ij}-d)/h) = f((d-d_{ij})/h)$ by symmetry of the kernel function f, the NW regression estimator can be expressed in term of density estimators as follows:

$$\hat{m}_{\mathrm{NW}}(d) = \frac{N^c}{N}\frac{\hat{K}^c(d)}{\hat{K}(d)},$$

where

$$\hat{K}^c(d) = \frac{1}{N^c h}\sum_{(i,j)\in\mathcal{N}^c} f((d-d_{ij})/h)$$
$$\hat{K}(d) = \frac{1}{Nh}\sum_{(i,j)\in\mathcal{N}} f((d-d_{ij})/h)$$

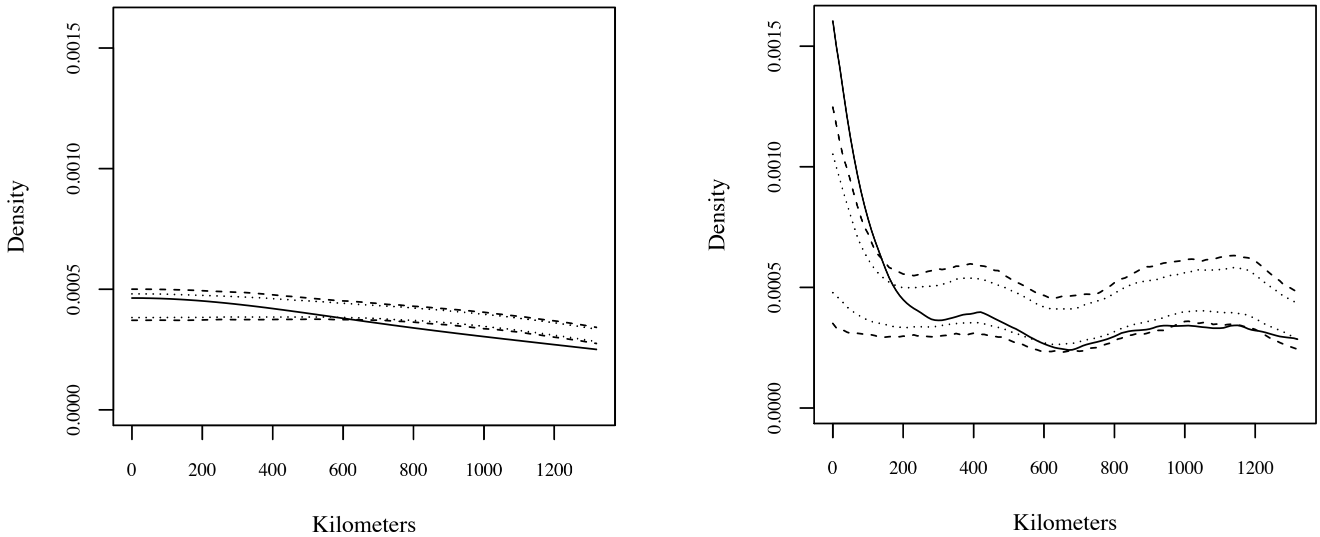
are K-density estimators. Note that the former expression, $\hat{K}^c(d)$, is the same as (7).

We denote by $\tilde{m}_{\mathrm{NW}}(d)$ the NW regression estimator when the dependent variable is a counterfactual citation dummy $\tilde{c}_{ij}$. Then, the localization measure is given by the difference between the two estimators as follows:

$$\gamma_{\mathrm{NW}}(d) \equiv \hat{m}_{\mathrm{NW}}(d) - \tilde{m}_{\mathrm{NW}}(d) = w_{\mathrm{NW}} \times \left[\hat{K}^c(d) - \tilde{K}^c(d)\right]$$

where $w_{\mathrm{NW}} \equiv (N^c/N)[\hat{K}(d)]^{-1} > 0$. Note that the actual and counterfactual bandwidths are assumed to be the same here. Under this assumption, the localization measure based on the NW regression estimators can be rewritten in terms of the localization measure (6) derived from the K-density estimators as $\gamma_{\mathrm{NW}}(d) = w_{\mathrm{NW}} \times \gamma_{\mathrm{KD}}(d)$. Since $w_{\mathrm{NW}} > 0$ for any distance, $\gamma_{\mathrm{NW}}(d)$ and $\gamma_{\mathrm{KD}}(d)$ have the same sign. Hence, they provide qualitatively the same localization results: Knowledge spillovers are localized at distance d if $\gamma_{\mathrm{NW}}(d)$ and $\gamma_{\mathrm{KD}}(d)$ are positive and significantly different from zero.

Appendix F: The results for the median citation distances



(a) K-density test

(b) L-density test

Figure F1: K-density and L-density tests using patent class 283 (median citation distances).

The solid curves are the estimated densities for the actual citations. The dashed (resp., dotted) curves correspond to the global (resp., local) confidence intervals obtained from the counterfactual citations. For the K-density test, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density test, we adopt the cross-validation method and McCrary's (2008) binwidth.

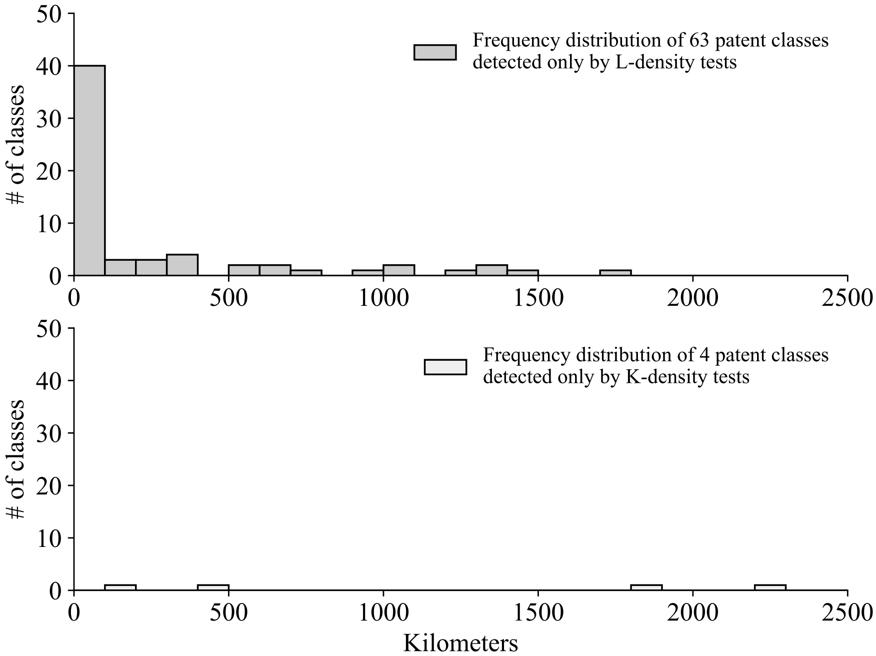
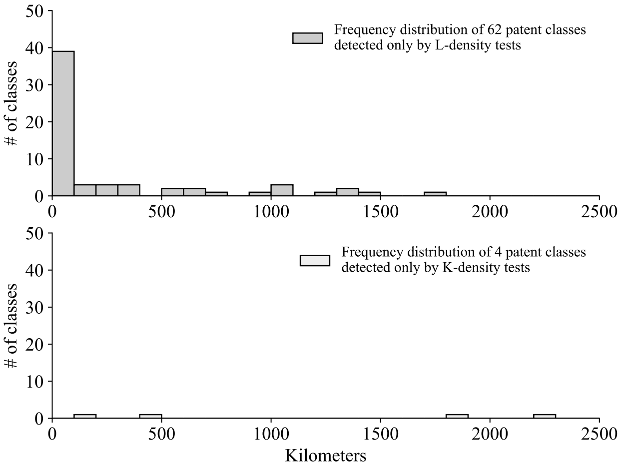
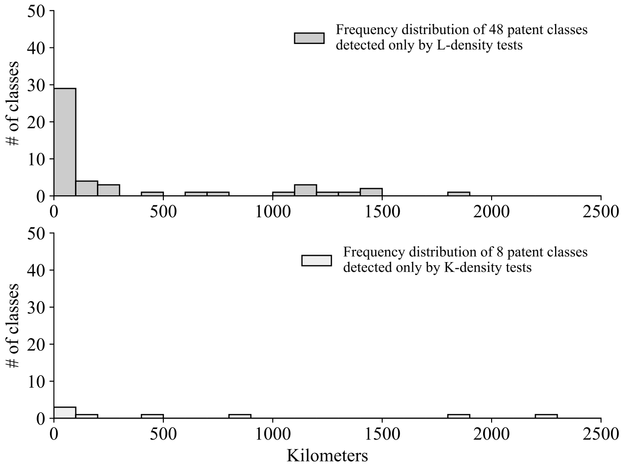


Figure F2: Localization results at and near the boundary (median citation distances).

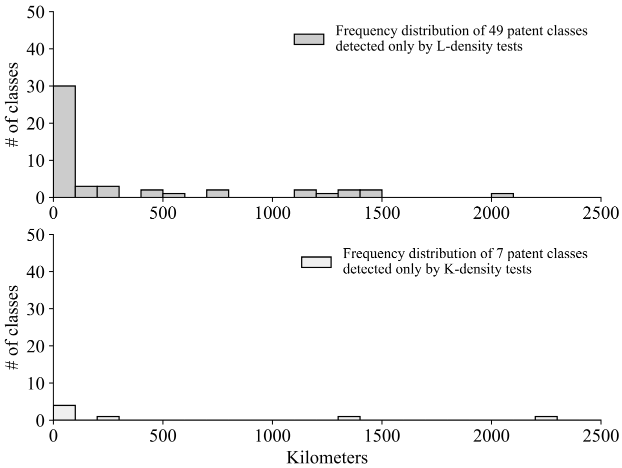
The top (resp., bottom) panel depicts the frequency distributions of patent classes for the distances at which the members in $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ (resp., $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$) initially display localization. The bars in dark grey (resp., light grey) are the frequencies of localized patent classes detected only by the L-density (resp., K-density) tests. For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method and McCrary's (2008) binwidth.



(a) Cross-validation ($h^A_{\mathrm{CV}}/\Delta > 10$)



(b) Fan and Gijbels' ROT ($h^A_{\mathrm{FG}}/\Delta > 10$)



(c) Silverman's ROT ($h^A_{\mathrm{S}}/\Delta > 10$)

Figure F3: Robustness check: Localization results at and near the boundary (median citation distances).

In all cases (a), (b), and (c), the top (resp., bottom) panel depicts the frequency distributions of patent classes for the distances at which the members in $\mathcal{A}_{\mathrm{LD}\setminus\mathrm{KD}}$ (resp., $\mathcal{A}_{\mathrm{KD}\setminus\mathrm{LD}}$) initially display localization. The bars in dark grey (resp., light grey) are the frequencies of localized patent classes detected only by the L-density (resp., K-density) tests. For the K-density tests, we use Silverman's (1986) rule-of-thumb bandwidth. For the L-density tests, we adopt the cross-validation method in panel (a), Fan and Gijbels' (1996) rule-of-thumb bandwidth in panel (b), and Silverman's (1986) rule-of-thumb bandwidth in panel (c). In cases (a) and (b), $h^A_{\mathrm{CV}}/\Delta > 10$ and $h^A_{\mathrm{FG}}/\Delta > 10$ hold for all patent classes $A \in \mathcal{A}$. Note that while the inequality condition is not satisfied for all patent classes, the binwidth, Δ, is iteratively decreased by 10 percent for that condition to be satisfied. In case (c), we choose the binwidth, Δ, such that $h^A_{\mathrm{S}}/\Delta > 10$ holds for all patent classes $A \in \mathcal{A}$.