Journal of Urban Economics 89 (2015) 21–31

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail

Octávio Figueiredo [a,*], Paulo Guimarães [a,b], Douglas Woodward [c]

[a] University of Porto, Portugal
[b] Bank of Portugal, Portugal
[c] University of South Carolina, United States

ARTICLE INFO

Article history:
Received 19 February 2014
Revised 25 June 2015
Available online 6 July 2015

JEL classification:
R12
R39
C13
C21

Keywords:
Knowledge spillovers
Agglomeration
Distance decay
Gravity equation
Poisson regression
High-dimensional fixed effects

ABSTRACT

This paper investigates Alfred Marshall's hypothesis that knowledge spillovers increase where industries are localized. At the same time, we take a fresh look at the role of distance in the diffusion of knowledge spillovers. Relying on a cited-citing gravity-like equation with high-dimensional fixed effects that control for multiple sources of observed and non-observed heterogeneity, we implement a Poisson pseudo-maximum-likelihood (PPML) estimator. We find that knowledge spillovers correlate positively with industry localization and that the agglomeration of an industry can offset the adverse effect of distance. The results also corroborate the distance decay effect uncovered in earlier research. Our new approach to estimate the PPML with two high-dimensional fixed effects should prove valuable in applications to a variety of other problems in economics, such as the estimation of gravity equations widely used in modeling migration, trade and other flows among countries and regions.

1. Introduction

Urban scholars have long known that firms in related industries are often concentrated in space. This paper studies Alfred Marshall's premise that the geographical concentration of industries, or localization, is associated with increased knowledge spillovers (Marshall, 1890). The hypothesis is that spatial clustering by firms within a given industry increases knowledge transfer through dense local networks and close interactions. This agglomeration process enhances external scale economies in urban areas and serves as a microfoundation for economic growth models put forth a century after Marshall by Romer (1990) and Grossman and Helpman (1991).

Apart from industry agglomeration, knowledge spillovers may be subject to distance decay. Proximity to knowledge generation is an advantage for firms and individuals seeking to exploit it. Yet today it is believed that knowledge can travel across space at low cost and without barriers. Given ongoing communication innovations, notably the widespread availability of information on the internet, the "death of distance" in knowledge flows has become a popular belief (Cairncross, 1997). Even so, it remains an open empirical question as to the extent to which knowledge spillovers are still determined by distance from knowledge generation.

In short, there are two fundamental forces potentially affecting the geography of knowledge diffusion: industry localization and distance decay. Estimating the effects of these forces is difficult, however, because measuring knowledge transfer across space is challenging. In a comprehensive review of the agglomeration literature, Rosenthal and Strange (2004) noted that knowledge spillovers present significant hurdles in conducting empirical research.[1] Famously, Krugman (1991, pg. 53) alleged that "knowledge flows [...] are invisible; they leave no paper trail by which they may be measured and tracked [...].".

Fortunately, researchers can identify and track patents, a major form of codified scientific and technical knowledge, along with

* Corresponding author at: Faculty of Economics, University of Porto, R. Roberto Frias, 4200-464 Porto, Portugal. Fax: +351 225505050.
E-mail address: octavio@fep.up.pt (O. Figueiredo).

[1] See also the survey of the literature in Puga (2010).

http://dx.doi.org/10.1016/j.jue.2015.06.003
0094-1190/

related citations. In our paper we follow the path of patent citations made available from the U.S. Patent and Trademark Office (USPTO). We posit a regression-based approach as a tractable alternative to estimate the effects of both distance and industry localization on patent citations, while at the same time controlling for a large set of other potentially significant effects. For the dependent variable, our regression takes the number of citations (per industry–county) that each patent receives in the subsequent nine years. This variable is tested as a function of distance between the cited–citing patents and industry localization. To carry out this regression, we constructed a unique database linking USPTO patents to U.S. county-level industry information. Following the suggestion of Santos Silva and Tenreyro (2006) we estimate our gravity-like equation using the Poisson pseudo-maximum likelihood (PPML) estimator and show how to incorporate two high-dimensional fixed effects in the model. For U.S. counties, we find that localization, (measured by industry establishments or employment) has a positive impact on patent citations, after controlling for distance and other factors. Importantly, these results indicate that industry localization can offset distance decay.

To our knowledge, this is the first paper to distinguish clearly the dual effects of localization and distance on knowledge spillovers. Previous research examined the impact of spatial proximity in generating knowledge spillovers without directly controlling for industrial agglomeration. At the crux of this research is Jaffe et al. (1993), who used USPTO citation data and unearthed evidence indicating that knowledge spillovers are highly localized. Based on a case-control approach, they conclude that inventors are much more likely to cite patents by other inventors from the same region.

Jaffe et al. (1993) acknowledged the need to control for industry clustering. As the authors illustrated, even though many citations to Stanford University patents come from its surrounding region, the Silicon Valley, we cannot directly attribute this tethering of patents and citations to proximity effects alone. Stanford patents are related to semiconductors and a disproportionate fraction of the semiconductor industry's research and development in the United States is located in the Silicon Valley. Attempting to control for the preexisting clustering of industries, the case-control method matches each citing patent to a non-citing patent. Purportedly, by matching each citing patent to a non-citing patent belonging to the same USPTO technological class, Jaffe et al. (1993) are able to indirectly control for the clustering of industries and related technological activities.

Thompson and Fox-Kean (2005) argued that Jaffe et al.'s (1993) method for selecting the control group may induce spurious evidence because matched patents may not belong to the same industry. Thompson and Fox-Kean (2005) reassessed Jaffe et al.'s (1993) finding by using the same methodology, but with a more restrictive way of selecting the control group. They find that when the matching is done based on USPTO technological subclasses, the evidence for a distance decay effect is much smaller and in some cases it even disappears. At the same time, these authors question the usefulness of a case-control approach to analyze this problem. As Thompson and Fox-Kean (2005, pp. 451) stressed, "controlling for unobservables using matching methods is invariably a dangerous exercise because one can rarely be confident that the controls are doing their job." Both Thompson and Fox-Kean (2005) and (in a reply) Henderson et al. (2005) agreed that one cannot really trust evidence about localization effects obtained after selecting control patents according to technology classification.[2] As a solution, these authors suggest that future research in this area should merge USPTO citation data with industrial data from other sources. They also advocated regression-based methods rather than the received case-control approach.

Nevertheless, other researchers have used the case-control methodology to obtain evidence that mostly favors the distance decay effect. Agrawal et al. (2008) refined the methodology by selecting the controls from the set that matches on the highest possible number of USPTO primary and secondary subclasses and confirm Jaffe et al.'s (1993) basic result. Using European Patent Office (EPO) data and the narrower technological matching between citing and control patents available from the EPO, Breschi and Lissoni (2009) found support for Thompson and Fox-Kean's (2005) findings. Murata et al. (2014) combined the case-control approach with the point pattern grounded methodology proposed in Duranton and Overman (2005). Using sensitivity analysis methods on USPTO data and relaxing the assumption of perfect controls, they found evidence in favor of localized spillovers, except in cases where the hidden biases induced by imperfect technologically based controls are extremely large. Thompson (2006) also uncovered a negative relation between distance and spillovers, although the localization effects he found were modest. His work is particularly interesting because the author followed a different strategy to select the controls. Since 2001, the USPTO has indicated whether each citation in a patent was added by the inventor or by the examiner. Thompson (2006) used the within-patent geographic distribution of examiner citations to control for the clustering of technological activities across space. The idea is that the examiners—who work in a single campus located in Alexandria, Virginia—cannot be learning about prior art because of geographic proximity. Thus, their added citations reflect instead the relevant distribution of industries and related technological activities across space.

Overall, the literature finds varying, inconsistent evidence on the role of proximity as a determinant of knowledge spillovers. The case-control approach to tracking knowledge spillovers suffers from a major drawback; namely, that industry localization effects are only partially captured by an implicit, unobservable term that is handled differently across studies using the procedure.

We argue that our model offers a way to overcome this perennial problem in empirical research on knowledge spillovers. Both the fundamental forces of agglomeration and distance are explicitly accounted for. We start our investigation with the approach to patents found in Caballero and Jaffe (1993) and Jaffe and Trajtenberg (1996), which model citations at the individual level. This allows us to motivate our econometric choice and to establish the link with the case-control approach. Then, using gravity-like regression and PPML estimation, we find that both distance and localization determine the flow of patent citations. There appears to be a tradeoff between these determinants; for example, we show that a 10% increase in the distance variable could be compensated by a 47% increase of establishments in the same industry.

Note that with our approach, we are drawing from the same tool kit that researchers have found to be advantageous in fields outside of urban economics. Related gravity models have been remarkably successful in international trade, finance, and other fields—so much so that deeper theoretical development has been stimulated in the wake of consistently solid empirical findings (Head and Mayer, 2014; Fally, forthcoming). Establishing a full theoretical foundation for our empirical model is beyond the scope of our paper. Even so, it should be recognized that fixed-effects, gravity-based PPML estimation, unlike previous methods used to study knowledge flows, has been shown to be consistent with microeconomic theory. Moreover, while our paper is focused on testing localization, the PPML estimator allows us to compare the elasticity for knowledge spillovers with respect to distance with gravity-based estimates for trade and other economic activities.

[2] These authors note that both the USPTO technological classes and subclasses have a low correlation with industrial activity, which brings into question its usefulness as a matching variable.