Geographic localization of knowledge spillovers as evidenced by patent citations.

Introduction

What is the question?

The paper compares the geographic location of patent citations with the cited patents to be able to analyze to which extent knowledge spillovers are localized.

How is it motivated?

Productivity of firms or industries is related to the amount they and other industries around them spend in R&D.

Knowledge spillovers among firms in industries is one of the three reasons cited by Krugman and Marshall behind the concentration of industries. Spillovers were not studied because of the thought that they leave no paper trail. But they do leave paper trail in the form of citations in patents. This allows us to use citation patterns to test the extent of spillover localization.

How is spillover-localization measured?

There are other agglomeration effects that could explain the geographic concentration of technologically related activities without resort to localization of knowledge spillovers and that’s why they construct “control” samples of patents and then calculate the geographic matching frequencies between the citations and originating patents and between the controls and originating patent.

The test consists in finding out if the citation matching frequency is greater than the control matching frequency since “control” matching frequency is the result of spillover localization.

Methods

What is a patent and a patent citation?

Patent is a property right in the commercial use of a device. If a patent is granted a public document is created containing extensive information including citations.
Citation serves the legal function of delimiting the scope of the property right conveyed by the patent.

**How “control” patents are constructed?**

We need to compare the probability of a patent matching the originating patent by geographic area with the probability of not matching.

they called this baseline or reference probability the “control frequency”, created because localization fades over time and countries differ in their areas of technological focus.

To derive a control frequency immune to contamination from movements over time or localization based on the pre-existing concentration of technological activity, they went back to patent data base and found a “control patent”. For each citing patent, they identified all patents in the same patent class with the same application year. They then choose from that set a control patent whose grant date was as close as possible to that of the citing patent.

Now they were able to compare the geographic location of the control patent with that of the originating patent cited.

**Which is the relationship between citations and spillovers?**

We can classify the links that might exist into one of three groups:

- Spillovers accompanied by citations
- Citations that occur where there was no spillover
- Spillover that occur without generating a citation

**Data**

Paper writers created Table I showing the number of citations made from university, top corporate or other corporate patents, the lag between the originating application year and the application year of the citing patent and the percentage of self citations.

They also constructed Figure I to show how citations from a patent change over time.

Writers in Table II showed the fraction of patents coming from abroad and from a selection of major U.S SMSAs for several of the datasets.
In Table III the matching proportions for the control samples are shown, as well as a t-statistic testing the equality of the control proportions and the citation proportions.

What key issues need to be taken care of?

Attributes of originating or citing patents affect probability of geographic location

Time: Localization of early citations is more likely than localization of later ones

Technology: Citations that represent research that is technologically similar to the originating research are more likely to be localized.

To explore these, they pooled the citations to university and corporate patents and ran a probit estimation\(^1\) with geographic match / no match between the originating and citing patents as the dependent variable. As independent variables they included the log of the citation lag, dummy variables \(^2\) for top corporate and other corporate originating patents, interactions of the lag and these dummies, a dummy variable studying the citation class and a dummy variable studying the match between control and originating patent

They also include other two variables. The “generality”\(^3\), an attempt to capture the extent to which the technological children of an originating patent are diverse in terms of their own technological location and the fraction of the originating patent’s citations that were self-cites.

Finally, the extent of localization depends fundamentally on the mechanisms by which information flows, and these mechanisms may be different in different technical fields. For this reason, they also include dummy variables for broad technological fields

The results are presented in Table IV showing that the matching-patent-class and generality variables do not work well. The effects are generally insignificant, and show no consistent pattern. The effect of the self-citation fraction, however, is strong.

A probit coefficient does not have an economically meaningful magnitude. However, we can calculate what the coefficient imply and that’s what’s done in Table V. this

---

\(^1\) probit function is the inverse cumulative distribution function (CDF), or quantile function associated with the standard normal distribution

\(^2\) Dummy variable or indicator variable. Takes the value 0 or 1 to indicate the effect that may expect to shift the outcome.

\(^3\) \(G_i = \frac{C_i}{C_0} \) where \(C_i\) is the total number of patents received and \(C_0\) is the number of citations received by patent i from subsequent patents whose primary patent class is k.
table calculates the predicted localization probability evaluating the citation lag at
different values, and evaluating the other independent variables at the mean of the data.

Results

Main findings

Knowledge spillovers do leave paper trail in the form of citations so they found
evidence that they are geographically localized and that their localization fades over
time. Less evidence of technological area on the localization process and of differences
in localization between the citations of university and corporate patents was found.

What is the paper not answering?

They would like to be able to say more about mechanisms of knowledge transfer and
about something resembling social rates of returns at different levels of geographic
aggregation but the limitations of patent and citation data make it difficult.