RESEARCH PROPOSAL

Sifan Zhou
University at Albany

Localized knowledge diffusion is one of the principal reasons why industries[1] and innovation[2] agglomerate. Yet, we observe firms belonging to the agglomerated industries spread their establishments across clusters in different regions and even in different countries. Existing literature has shown that (1) knowledge diffusion decreases in distance[3], (2) knowledge diffusion is easier within firm than between firms[4]. A natural inference from these two patterns is that a firm can have an establishment in one cluster, absorb external knowledge locally, and transmit the absorbed external knowledge to the firm's inventive activities at other locations. Empirical works on within-firm knowledge diffusion has solely focused on the diffusion of a firm's own knowledge from the firm's one plant to another, as measured by the firm's self-citations in patents. I propose to study the role of within-firm knowledge diffusion in transmitting external knowledge. It will not only contribute to understanding multi-location firms' management strategies for innovation, but also contribute to understanding the connection between individual innovation clusters. In this study,

(1) I will test whether a firm absorbs external knowledge more from a location where the firm has some establishment than from a location where the firm has no establishment. Test whether a firm absorbs more external knowledge from a location after the firm opens an establishment there than before. Test whether an R&D lab boosts the firm's absorption of local external knowledge more than a manufacturing plant or a sales branch does. [5]
(2) I will examine how distance mediates the within-firm diffusion of external knowledge. Examine whether within-firm diffusion decreases in distance faster or more slowly than cross-firm diffusion does.

I will measure the knowledge diffusion using patent citations. A patent cites the predecessor technology and knowledge upon which the new invention builds. Patent citation is regarded as the "paper trail" left behind by knowledge diffusion. My basic OLS regression[6] takes the form of:

$$CITE_{ftij} = \beta_1 DIST_{ij} + \beta_2 EST_{ftj} + \beta_3 EST_{ftj} * DIST_{ij} + X'_{ftij}\gamma + \epsilon_{ftij}. \qquad (1)$$

[1] For the role localized knowledge diffusion has to play in industry agglomeration, see Marshall (1920), Ellison, Glaeser and Kerr (2010).
[2] Audretsch and Feldman (1996)
[3] See Jaffe et al. (1993), Almeida and Kogut (1999), Thompson (2006).
[4] See Kogut and Zander (1992) for theoretical development and Singh (2005) for empirical evidence.
[5] See Cohen and Levinthal (1990) on absorptive capacity.
[6] Since citation is a count variable, I will also use Poisson regression or Negative Binomial regression to test robustness.

$CITE_{ftij}$ denotes the number of citations a firm f gives from its new patents invented at "citing" location i in year t to the patent stock not owned by the same firm at "cited" location j. It measures the knowledge flow from outside the firm at location j into the firm's inventive activities at location i. A location here is a geographic unit, such as a state, an Economic Area[7], or a county. I will test whether using geographic units of different sizes affect the results. For a given firm in a given year, each of its patenting locations is a "citing" location. For each firm-year-citing location combination, every location with relevant patent stock is a "cited" location. $Dist_{ij}$ denotes the logged distance between the "citing" location and the "cited" location. Localized knowledge diffusion predicts $\beta_1<0$.

EST_{ftj} is a dummy variable, equal to one if the firm has at least one establishment in the "cited" location j. If having an establishment at a location helps the firm's absorption to the external knowledge from that location, β_2 will be positive. Controlling firm-cited location fixed effects, I can test whether a firm will cite a location more after opening an establishment there than before. EST_{ftj} can be replaced with a dummy indicating whether the "cited" location has the firm's R&D lab, or its manufacturing plant, or its sales branch to examine how these types of establishments differ in absorbing and diffusing knowledge. EST_{ftj} can also be replaced with a continuous measure such as the firm's employment at the "cited" location to capture the effects at the extensive margin. β_3 is the coefficient of the interaction term between $Dist_{ij}$ and EST_{ftj}. If within-firm knowledge diffusion decreases in distance, $\beta_1 + \beta_3 < 0$. A positive β_3 indicates within-firm diffusion decreases in distance more slowly than cross-firm diffusion does.

X_{ftij} is a set of control variables. It includes the number of new patents firm f apply for from the "citing" location i in year t that will give out citations, PAT_{fti}. Note that locations differ by their abundance of patent stock relevant to a specific firm's technology classes. X_{ftij} will also include the patent stock from the "cited" location j that will potentially receive citations, $PAT_{(-f)j}$. Depending on the specifications, X_{ftij} also includes year fixed effects, firm-cited location fixed effects, cited-location-citing-location fixed effects, and cited-location-citing-location linear trends.

A key challenge to this study is the endogeneity issue of establishment locations. Firms do not locate its establishments randomly, but will consider the benefits of knowledge spillovers from universities and other firms when making location decisions. Once a firm has an establishment at a location, the establishment may shape the direction of the innovative efforts by universities and other firms at the location, especially when the firm's establishment is big and R&D intensive. This will make the local knowledge "external" to the firm more relevant to the focal firm's

[7] Economic Areas are geographic units defined by the Bureau of Economic Analysis in an effort to identify geographic areas that represent the relevant regional markets for labor, products, and information. There are 179 Economic Areas in the U.S.

business lines, and thus more "citable" by the firm's new patents. These will all lead to an overestimation of β_2.

To solve the endogeneity issue, I will focus on the effects of new establishment openings on the parent firm's patent citations to local patent stock three years ago or five years ago. This will alleviate the concern that locations with recent technology breakthroughs attract new establishments and also have patents worth more citations. My previous work[8] on firms' location decisions for R&D labs finds that generous state level R&D tax credits attract new R&D labs even after the external innovation environment, the parent firm's pre-existing spatial distribution of activities, and location fixed effects are all controlled for. Since firms in this study perform R&D and apply for patents, they are likely to care about local R&D incentives and intellectual property protection policies in deciding where to open new establishments. I will use the changes in state level R&D tax credits[9], corporate tax rates, non-compete enforcement, and trade secrets enforcement to instrument the establishment openings.

In this study, the data for firms' locations will come from the Longitudinal Business Database (LBD)[10]. The LBD provides establishment level data on the employment, location, and corporate affiliation. Its establishment-level industry codes allow me to distinguish R&D labs from manufacturing plants and sales branches. Its time-consistent establishment level identifiers allow me to identify establishment births. The data for patents come from the United States Patent and Trademark Office (USPTO). I will use inventor addresses to locate patents and use application date to proxy the invention date. Kerr and Fu (2006) created a linkage between patent assignee to LBD. I will use this linkage to tell which patents belong to a focal firm and which patents form the knowledge stock external to the firm. This study will cover all the multi-location firms that applied for patents in more than one year during the period of 1977-2001.[11]

If I join the fellowship program, I would be excited to participate in NBER's projects on productivity, innovation and entrepreneurship programs. The NBER location also allows me convenient access to the Boston RDC to conduct this research. After completing the fellowship program, I plan to apply for Assistant Professor Positions in Economics Departments. I believe the fellowship program at the NBER will equip me with the valuable research experiences, and help me develop and publish high-quality papers.

---

[8] Zhou (2014)

[9] Wilson (2009)

[10] I am a Special Sworn Status researcher of the U.S. Census Bureau with active access to the confidential LBD and Census data. The NBER Postdoc Fellowship will enable me to conduct this project at the Boston Census Research Data Center.

[11] While LBD is available during 1976-2012 and USPTO patent data is available since 1975, the patent assignee to firm linkage only covers firms in LBD for the period of 1977-2001. (Kerr and Fu 2006) Since the patent-firm linkage is essential to this study, I will limit my sample period to 1977-2001. I will try to extend the patent-firm linkage to more recent years in order to extend my period of study.

References

Almeida, P. and B. Kogut (1999). "Localization of knowledge and the mobility of engineers in regional networks." Management Science **45**(7): 905-917.

Audretsch, D. B. and M. P. Feldman (1996). "R&D spillovers and the geography of innovation and production." The American Economic Review: 630-640.

Cohen, W. M. and D. A. Levinthal (1990). "Absorptive capacity: a new perspective on learning and innovation." Administrative science quarterly: 128-152.

Ellison, G., et al. (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." The American Economic Review **100**(3): 1195-1213.

Jaffe, A. B., et al. (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." The Quarterly Journal of Economics **108**(3): 577-598.

Kerr, W. R. and S. Fu (2008). "The survey of industrial R&D—patent database link project." The Journal of Technology Transfer **33**(2): 173-186.

Kogut, B. and U. Zander (1992). "Knowledge of the firm, combinative capabilities, and the replication of technology." Organization science **3**(3): 383-397.

Singh, J. (2005). "Collaborative networks as determinants of knowledge diffusion patterns." Management Science **51**(5): 756-770.

Marshall, A. (1920). Principles of Economics, London: Macmillan.

Sifan Zhou. (2014). Where do firms open R&D labs in the U.S.? University at Albany.

Thompson, P. (2006). "Patent citations and the geography of knowledge spillovers: evidence from inventor-and examiner-added citations." The Review of Economics and Statistics **88**(2): 383-388.

Wilson, D. J. (2009). "Beggar thy neighbor? The in-state, out-of-state, and aggregate effects of R&D tax credits." The Review of Economics and Statistics **91**(2): 431-436.