Volume Title: The Rate and Direction of Inventive Activity Revisited

Volume Author/Editor: Josh Lerner and Scott Stern, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-47303-1; 978-0-226-47303-1 (cloth)

Volume URL: http://www.nber.org/books/lern11-1

Conference Date: September 30 - October 2, 2010

Publication Date: March 2012

Chapter Title: Comment on "The Diffusion of Scientific Knowledge across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine"

Chapter Authors: Adam B. Jaffe

Chapter URL: http://www.nber.org/chapters/c12351

Chapter pages in book: (p. 156 - 160)

Comment Adam B. Jaffe

I am reminded of President Kennedy's quip about how when he had accumulated his brain trust, it was the largest accumulation of brain power in the White House since Jefferson dined alone.

I feel like this is the largest concentration of knowledge and insight about technical change since Zvi Griliches opened his mail at the lunch seminar.

So quickly, what did they do? They built this new data set, 10,000 so-called superstar scientists, and then they tested this issue of the geographic localization of knowledge flows as proxied by these citations from articles and from patents, to both articles and patents. For those of you who are small-minded and wondering where is the fourth element of the pair, there are almost no citations from articles to patents. There are a few, but not very many, so we don't bother.

In looking at this, there is this identification problem, which he used the quote from my paper with Manuel and Rebecca to illustrate. We do not know whether the apparent localization is due to the fact that proximity facilitates communication or whether it is just due to the fact that there is already a geographic concentration of interest in a given topic. And so that is something we would like to tease out. So, there are solutions you could think of in this language of difference-in-difference estimation. We are going to look at scientists who move. That is the difference. But, we are not just going to look at scientists that move.

We are going to compare them to scientists who did not move. So, that is the difference-in-difference structure to this project. For the citations from articles, they actually do some regressions that look at interaction effects with article and scientist attributes to try to see if they can tease out more about what is going on here. And then the last thing they did very briefly is they have some speculative normative inferences about scientists moving, and I'll comment on that.

Figure 2C.1 is my summary of the chapter. There are three kinds of citations—article-to-article, patent-to-article, and patent-to-patent. After the move, we are only looking at citations to the output of the scientists from before the move. We are not looking at the dynamic of the new work that the scientist is doing after the move, because again, that is a more complicated question about exactly what is going on there. We are really trying to do this very clean test about just the localization through communication. We are looking at citations from the time period after the move to the work, the output, be it articles or patents, from before the move. We look at citations

Adam B. Jaffe is the Fred C. Hecht Professor in Economics and dean of the College of Arts and Sciences at Brandeis University and a research associate of the National Bureau of Economic Research.

	From Old Location	From New Location
Article to article	No change	UP
Patent to article	Weakly Down	Weakly Up
Patent to patent	Down	Weakly Up

Fig. 2C.1 Effects of move on citations to premove output

coming from the old location and from the new location. And basically what we find is for these article-to-article citations, they are up. For the patent-to-patent citations there is pretty clear evidence at the old location of what they call forgetting. And for the others, there are some weaker effects where weaker basically means it goes in that direction but it is not statistically significant in the difference-in-differences formulation.

So, what I like about this chapter. First of all, it is a very important and interesting problem, in my view, but of course my saying that communicates absolutely nothing new because I have worked on this problem a lot myself; obviously, I think it is important and interesting. You can make your own judgment about whether it actually is important and interesting, but I think it is. Second, this is an incredible data construction effort, one that as a dean, I would say is probably foolhardy for junior faculty to undertake. But I mean that as a compliment. This is the kind of work that our profession underrewards. It really should reward it more. That is why it might be foolhardy, but it really is incredibly important work. Where they were faced with a choice about how to do something, they always chose the very labor-intensive but better approach over the easier but not as good approach. So this is really an incredible data set, and I think there are going to be enormous spillovers as we go, as this field evolves to other work.

The chapter very clearly explains everything they did. It is very well done. And this difference-in-difference approach is about as clean a causal test as you can get. There was some discussion of this the other day, about how one of the problems in this area generally is that everything is always correlated with everything and it is very hard to test causality. This is about as good as you can get.

It is the job of a discussant to make some additional suggestions. I am not going to worry about which of these things you might actually do in this chapter and which of these things you might do some other time, but these are just thoughts I have about things you might do. First of all, the first two are very small points. The word superstar seems inappropriate to me. We have 10,000 people here, so I would call them productive scientists. If you want to call them stars, I would probably buy that. But they are not superstars.

There is another small thing. There is a paper by Almeida and Kogut, that is conceptually similar. They look at the semiconductor industry and the effect of mobility of engineers on citations. So you should connect to that.

Another obvious suggestion is that you can look at other things beside just the total number of citations. Manuel and Rebecca and I had this measure of "generality," which captures the extent to which citations are broadly distributed rather than concentrated technologically. That would be interesting. You might conjecture that the less closely technologically related people are less geographically sensitive, but that is something in your data that you could actually look at.

Another thing to do would be to look at the citations made by the scientists who move. In the chapter, you say you threw them out. I hope you did not really throw them out. You just meant you were not using them in this chapter. I am assuming that MIT recruited Scott back to Cambridge for the benefit that his work would have on Pierre, but presumably Scott moved back in part for the benefit that Pierre is going to have on him. So that is just as interesting. It is a different issue, but it is just as interesting.

This next point is probably the biggest one I want to make. I started to say that you threw the baby out with the bathwater, and then when I thought about it, realized that is not quite the right metaphor. I think the metaphor here is you did not throw the baby out, but the bathwater itself is actually pretty interesting. What I mean by that is I would not have just reported difference-in-difference results. I would have actually reported some of the results before you do the difference-in-difference. So for example, I would have liked to have seen the picture represented hypothetically in figure 2C.2.

There is a solid line and a dotted line. I made this up. This is not data. But I am guessing that something like this is going on, that if we look at the whole pattern of citations over time from a work at the new location, they are getting more citations than they were getting at the old location. But we don't know that it looks like this. It might look like figure 2C.3. It may just be that you get a very rapid diffusion, which then fades out. And there are lots of other issues that one would like to understand. Now the causal interpretation of this is more complicated than the difference-in-difference, but before we go trying to get a really pure causal story, I think we should try to have a better sense of what it is we are explaining. What is actually

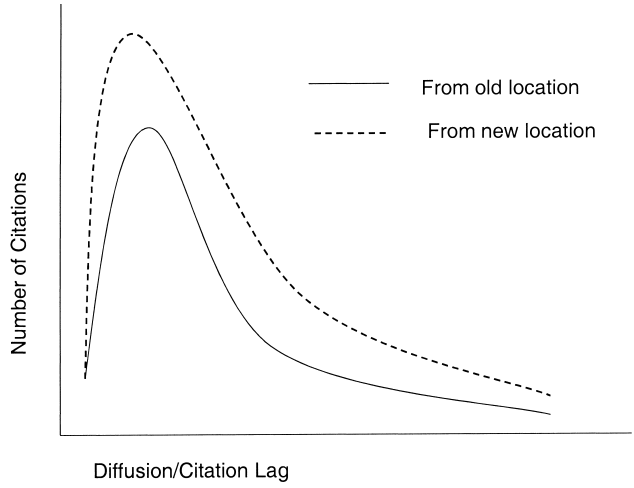

Fig. 2C.2 Postmove citations to premove output

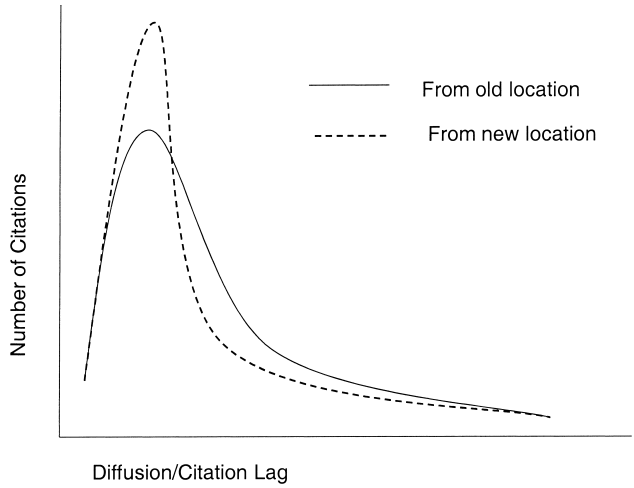

Fig. 2C.3 Postmove citations to premove output—alternative time path

going on here? And that is lost when we start by looking at the difference-in-difference result. So I think the chapter and the whole line of research would be much more interesting if you let us see the first order effects before we go to the comparison, because after all, it is already one difference. You do have something changing in terms of the scientist moving and I think that that is worth knowing something about.

So just a couple of final points. I think the normative speculations about mobility are really pushing it, because in some sense, what you are looking

at are second order effects of what actually is affected both privately and publicly when somebody moves. So yes, you said they were speculative, so in that sense I cannot really convict you. But still, I found that very uncompelling. And then the last thing I would like to suggest is that you should put these data up on the web analogously to the NBER patent data set. And since one of the themes of this conference is a little bit of kind of history of the development of this field, I will tell you when Manuel and Rebecca and Bronwyn and I were first thinking about accumulating these data in the early 1990s, we actually had some extended discussions from a private return perspective in terms of our own careers. Should we hold on to this and write as many papers as we could or should we make it public so that other people could use it? I don't remember who took which side in the debate.

But the fact of the matter is, we put these data up, and I think this was before we, at least, were thinking about open source versus walled gardens or any of that stuff. We put all of the data on the web. I looked last night on Google Scholar. The paper that is the sort of handbook for using these data has 997 citations. So clearly from a private perspective, if our objective function is sort of prestige and so forth as measured by citations, we benefited professionally by making these data public. And so I would urge you to do the same. Google Scholar is doing some very similar work about massaging these various data. It seems to me it is socially wasteful for people to do that twice. So in some sense, maybe you guys should talk to them about some kind of joint venture where you would share the efforts to clean these data, to organize them in a way that would be useful for research.