

Science Parks, Knowledge Spillovers, and Firms' Innovative Performance. Evidence from Finland.

by Mariagrazia Squicciarini

Referee Report for the *Economics-ejournal*

The goal of this paper is to study whether location in a science park (SP) relates to firms' innovative performance. The empirical test uses data on 252 tenant firms in Finland and employs the number of patents that they produce as a measure of innovative performance. Firms and SPs' features are the key covariates. The paper concludes that SPs' characteristics are correlated with firms' inventive performance. This correlation is then interpreted as a result of knowledge spillovers.

The issue of spillovers and the role of regions and locations in fostering them are largely explored in the economic literature. This paper contributes to it by using data on Finnish firms and SPs. However, as also noted by the author, empirical research on these topics is hard to perform as it needs to cope with a number of econometric issues.

As a matter of fact, my main concern about the present paper is that, the author has neither solved the potential econometric/conceptual problems of the empirical analysis, nor she has provided convincing analytical arguments that these problems do not affect the estimated results. For example, given that only firms located inside SPs are taken into account, the paper lacks a term of comparison and the empirical analysis selects on the dependent variable (page 4 “[F]irms normally have to apply to join a science park...”). Moreover, the choice to locate in a SP might be endogenously determined. The author starts by acknowledging that selection and endogeneity might bias the empirical results, but then she provides some arguments why they are not major drawbacks of the paper (e.g. page 14). This strategy is not rigorous enough for a scientific paper: either the author tries to address the econometric concerns, or she provides stronger and more rigorous/analytical arguments supporting the view that they do not bias the estimated results.

One way to tackle the problem of endogeneity would be with panel data and fixed effects for firms, which I thought the author had done when I started reading Section 3. Also, – unless the author has done it in a companion paper – since there are only 252 firms in the sample, one might think of collecting data on a control sample of similar firms in terms of size, sectors, etc. located outside the SPs.

However, let us assume that the paper does not suffer from selection/endogeneity problems and that there is correlation between firms' patenting activity and the characteristics of the SPs. The author interprets this as the result of knowledge spillovers. Why? Knowledge spillovers are only one possible – yet, not demonstrated by the author – reason why location in a SP affects research output. There are relevant contributions in the economic literature that use sophisticated techniques, indicators and empirical frameworks to assess the existence and the geographical breadth of knowledge spillovers (Jaffe, Trajtenberg and Henderson, 1993; Thompson and Fox-Kean, 2005; Thompson, 2006). Did I miss something in the present paper that tests the role of knowledge spillovers – as unintended knowledge flows between people or

institutions – in affecting innovative performance? If this is not the case, is there any reason why the author makes this hypothesis in the first place? Unless the author wants to directly and properly test the existence and importance of knowledge spillovers, she could simply discuss all the possible explanations for the estimated results, including knowledge spillovers.

If this is the direction in which the paper wants to go, I would then recommend to look at the work by Alcacer and Chung (2007) – and related literature – on competitive concerns and co-location according to which firms care about the net spillovers effect (i.e. the balance between inward and outward spillovers) when they endogenously decide their location. Then, for example, the statement in page 14 “[i]nnovation-oriented firms would be comparatively more willing to locate inside science park” could easily go the opposite direction, as the expected outward spillovers might be higher than inward spillovers.

I also have some concerns about the variables used in the regressions and their interpretation. First, the “input indicators” (firms and SPs’ characteristics) are measured as of 2002 while the output indicator is measured in the preceding period. If anything, input indicators should be measured a few months/years before the output indicator. Second, among the firm-level controls there are important omissions like firms’ R&D intensity (possibly their absorptive capacity) and the fact that firms have branches also in locations other than the SPs.

Third, the output indicator is the firm’s sum of patents and utility models throughout the period in which the firm locates in a SP. One of the covariates is the number of years that the firm has spent in the SP. The author interprets the positive coefficient of this variable as suggesting that the longer the period spent in the SP, the higher the benefits from co-location. Actually, it may simply be that the larger the number of years, the larger the number of patents that a firm can apply for, with no role of co-location, spillovers or advantages given by the SP (as the author also acknowledges in a footnote). I would therefore use this covariate as a simple control, with no link to the benefits of being in a SP, unless the author demonstrates it analytically.

Fourth, the variable $Nout_i$ is always positive and statistically significant. This could signal a problem of unobserved heterogeneity: if a firm is good at doing research before joining the SP, it then continues to perform well also after joining the SP – as opposed to the argument of “persistence”.

Finally, the inclusion of sectoral dummies – Table 4 and 5 – makes the statistical significance of most of the SPs’ variables either disappear (as in the case of *Uni*) or diminish (as in the case of *ncom* and *big*). I would then be careful in reading the results: some sectors benefit more than others from being located in the SPs (as mentioned by the author in page 25), and these cross-sectoral differences erode the effects of the SPs’ variables. Why? This would be interesting to explain, probably by performing separate regressions at the sectoral level (though the number of observations would be low in each regression). The use of interacted variables also goes in this direction, though one would need to interact the other sectoral dummies with the key SPs’ characteristics.

Additional comments:

Covariates: 1. For the spillover argument it would be useful to add the area (km²) of the science parks or the density of the local firms and institutions; 2. Why does the author use a 0/1 dummy variable for Uni_j ? If spillovers exist, then, the higher the number of Uni_j given the area of the SP (or the higher their density), the larger the spillovers. 3. In general, if the author has a strong theory about the sign and statistical significance of the covariates, then, there should be a discussion of the theory. Differently, if the variables are used as controls and they can take any sign, no need to spend pages on their description.

Data: The source of data and the method used to construct the variables are sometimes unclear: e.g. how was the matching between firms and patent data performed? What use has the author made of survey data?

Causation: the author states a couple of times that the paper is not a test of causation between location and innovative performance. Throughout the paper, however, she says that the aim of the paper is to investigate whether location inside the SPs matters for firms' inventive performance due to the effect of knowledge spillovers. Well, if the paper is not a test of causation, than the results cannot be read as suggesting that location affects productivity via spillovers.

Model: Section 4 is entitled "The Model", but then there is no model in the traditional sense of the word. I would suggest to use a more appropriate title like "Econometric specification".

Length: The paper is too long. Section 2 and 3 can be reduced considerably.

Results: Page 25: "[S]ince the three model specifications used (P, NB and T) yield similar results...". Is the author sure about this statement? If I read the tables correctly, the P regressions give quite different results compared to the other estimation models. Which one is then correct/credible?

Dependent variable: Why mix up patents and utility models? I would either limit the analysis to patents only, or run two different specifications: one with patents only, and the other one with both patents and utility models.

Interaction terms: What is the theory behind the inclusion of the two interaction terms? Why these and not others?

In sum, I think this is a potentially interesting paper, but at this stage it needs work in order to be of published in this Journal.