

Referee report

Science Parks, knowledge spillovers, and firms' innovative performance. Evidence from Finland

By Mariagrazia Squicciarini

This paper addresses an interesting issue, namely whether there are knowledge spillovers within science parks. For that it uses data of 252 firms belonging to 18 science parks and runs a regression explaining the number of patent applications of firms while located on science parks. Part of the explanatory variables are park specific, the others are firm specific.

Although I do recognize the originality of the data, I have a number of remarks about this paper.

1) My first remark is that the data have a particular type of panel structure since data refer to firms and parks. The park effect is captured by 6 variables. It would have been interesting to control for a park fixed effect instead of the 6 park-specific variables. These six variables might capture other park-specific effects. Another possibility would have been to use hierarchical regression models.

2) My second major remark relates to the measure of the dependent variable. As the author recognizes, the number of patents applied for while on-park will of course increase with the number of years of presence on the park (y_{it}). Why not estimate the number of patents per year of presence in the park to avoid this spurious regression?

3) It is not clear what the main point of interest of the paper is. The title speaks about knowledge spillovers, whereas the abstract emphasizes the location and the length of stay inside a science park. The spillover story is captured by a park-specific variable, with the reservations mentioned before. The location variable is nowhere in the regression, only the number of years in the park is, and again with the reservations expressed above. Hence I am not sure that this study constitutes a perfect framework to be replicated on other data, as claimed by the author in the conclusion.

4) It would be nice to have some more descriptive statistics about all the explanatory variables used in the econometric analysis, at the science park and at the firm level.

5) The issue of self-selection is difficult to tackle with cross-sectional data. The author has controlled for prior patenting activity, which is about the best she could do. But I would not claim that (p.32) "our results ...should also hold in case of selection and self-selection were at stake" or (p. 14) that "endogeneity problems should not impinge upon the results of the study". There is no way to back these assertions from what is presented in this paper.

6) Is the regular tobit model estimated here or a tobit model allowing for count data? Why not use a hurdle or zero-altered Poisson model?

7) It is strange to pick up the preferred specification in table 4 on the basis of a fit indicator that the author herself recognizes as having no particular meaning. Instead, the Poisson model is nested in the negative binomial model and could be formally tested against the latter. Looking at table 4, I have the feeling that the negative binomial would win the race and therefore it would be logical to present the results based on this specification, rather than the Poisson model (unless there is some other reason to prefer to negative binomial model).

8) It would be interesting to compare the results of the effectiveness of science parks found in this study with similar results in either countries from studies mentioned in section 2, at least with the comparable Felsenstein (1994) study. Even if the compared performance outcomes are different, it is still interesting to compare the effects, by relying on some outside estimate of the link between these other measures of performance and the number of patent applications.

9) Uni_j does not imply that the firms in the science park collaborate with these higher education institutes, and hence the argument advanced by Hall et al (2003) to explain the negative effect might not hold here.

10) Could the inter-park and intra-park spillover effects of multiple branches not be dissociated?

11) It was a good idea to examine the time lag phenomenon. It is amazing that the estimated coefficients increase in absolute value (when significant at conventional levels) with the number of lags, except for $ncom$ and $fiem$. Why is that so?

12) In conclusion, I would have liked the author to choose on reasonable grounds her preferred specification. My pick would be the negative binomial with industry dummies in table 4. But there few coefficients are significant. But the main variable of interest, $yyin$, is! This result should be brought forward much more than it is now. It should be made the core of the paper.