Examination report on Taras HRENDASH's Ph.D. thesis entitled:

Essays on Economics of Innovation

By Prof. Gaétan de Rassenfosse
Ecole polytechnique fédérale de Lausanne

January 26, 2023

Main comments

The thesis offers three papers. All the papers deal with the empirical economics of science and innovation. Each paper is written as an independent, stand-alone piece, and there is no thematic consistency across the papers. The first paper presents a novel clustering algorithm to identify technology clusters by exploiting the characteristics of patent data. The second paper empirically studies the impact of shortened patent pendency on the sales of patents. The third paper provides a demographic model that allows decomposing the factors driving the aging of the U.S. scientific population workforce.

Overall, an impressive amount of work went into the thesis, and the Candidate successfully demonstrated his knowledge of the literature, research skills, and creativity. The research gaps are well identified, attempts at bridging these gaps are credible, and the papers are well written. There is no question that the thesis does advance knowledge in the economics of science/innovation in a meaningful manner. The following section provides detailed comments on each paper.

Specific comments

In this section, I list the main comments and questions I have for each paper.

Paper 1: Cluster Identification in Collaboration Networks of Innovators

- One major concern relates to the fact that collaboration networks, as inferred by patent data, are heavily constrained by firms' boundaries. In contrast to scientists writing papers, inventors producing patents do *not* freely choose who they work with. Besides, they primarily co-invent with colleagues from the same firm, rendering the use of a patent-based collaboration network to identify clusters, at best, questionable. The paper should discuss this point very carefully. In addition, in light of this comment, some interpretations are likely to be incorrect and should be revised accordingly (e.g., "a higher spatial concentration of inventors within such areas presumably leads to a higher frequency of social interactions and increases the chance of new collaborations"; well, not really if inventors are not free to choose who to work with).
- An alternative patent-data-derived network that is less affected by firm-level constraints lies in the network of patent-to-patent citations. Integrating this perspective can potentially improve the paper's relevance and impact significantly. It is not necessary to re-estimate the clustering algorithm on the network of citations as

part of the Ph.D. thesis. However, a conceptual discussion of the pros and cons of using a citation-based network would be a welcome addition to the paper.

- A clustering algorithm that is too restrictive—leading to small clusters and many observations out of the clusters—may become irrelevant. For example, it cannot be used to guide regional innovation policies or smart specialization strategies, which are frequent use cases of clustering algorithms. With that point in mind, it would be helpful to report the number of observations that fall or not into clusters and to compare these numbers with other such numbers reported in the literature. I have the feeling that the clustering algorithm may be too restrictive, leaving too many observations outside of clusters. (I note that a citation-based network might alleviate this concern.)
- Are the optimal parameters for the clustering algorithm robust to excluding self-citations? Since firm boundaries constrain the network, we do not want firms' citation practices to influence quality indicators. Furthermore, are the parameters robust to extending the citation time window? Two years post-grant seems a relatively short window; the literature usually considers 3, 5, or 10 years.

Paper 2: The impact of prioritized examination on the commercialization of patents

- This paper would benefit from a stronger theoretical grounding. Section 2.2.2 provides half a page explanation of the expected effect but builds very little on existing theoretical frameworks and prior work. Alternatively, the Candidate could set up a simple analytical model that explains the firm's decision to fast-track patent applications to guide the empirical analysis.
- In a similar vein, the paper fails to acknowledge some notable contributions in the area, including Kuhn and Teodorescu (2021) and Dechezleprêtre (2013), available at https://doi.org/10.1002/sej.1387 and https://ssrn.com/abstract=2228617, respectively.
- In particular, Kuhn and Teodorescu (2021) find that applications for which applicants elect accelerated examination are of particular strategic and technological importance, as inferred from higher citation rates and litigation probability. Such "strategic importance" is challenging to observe for the econometrician and casts doubt on the causality of the results in the present paper. I do appreciate the efforts that the Candidate puts in trying to obtain causal estimates. However, the paper would be more convincing by toning down claims about causality.
- The variable of interest in equation (2) is the high-intensity treatment group, but some treated patents may fall outside this group. How many treated patents fall outside this group? Similarly, how many patents that are not treated fall in the high-intensity group (in the post-treatment period, of course)? A kind of confusion matrix would be useful here. At the limit, a meaningless grouping will produce null results. It would be helpful to discuss how the quality of the grouping biases the coefficients (building either on the 'attenuation bias' or the bounding approach of Aakvik 2001). It is not necessary to implement this last advice into the thesis.
- Another concern relates to the mechanism. The Candidate implies that a faster grant drives the effect. However, another consequence of a rapid grant is an earlier publication of the patent document, which would otherwise be published after 18 months. If the grant occurs before 18 months, the patent is both disclosed and

granted. We know from Hegde and Luo (2018) that publication matters (but see also de Rassenfosse et al. 2016, available at https://doi.org/10.1016/j.respol.2016.03.017). Tat the minimum, the thesis should discuss this point. At best, there are ways to tease out the publication vs. grant effects, for instance, by looking at fast-track patents granted after 18 months (if there are any) or looking at patents that have a published international equivalent before the U.S. grant (and for which information on the content of the patent is already public).

- In Table 2.3, it is surprising that the triadic variable has a negative impact, whereas the patent family size has a positive effect. What family size is used? The geographic family size (counting the number of distinct countries in the INPADOC family) or the size of the INPADOC family? The former count is more relevant. What is the correlation coefficient between the triadic status and the patent family size? Wouldn't it make more size to drop the triadic variable and add a square term for family size to capture potential nonlinearities?
- The Candidate should provide a more convincing explanation as to why the study is limited to VC-backed firms. Do the results hold if we extend the sample to all firms? For instance, to universities, non-VC-backed small and micro entities, and individual inventors? If the results do not hold, how do the conclusions of the paper change? Google is an avid user of the fast-track program. Do the results hold when including Google's patents? If not, why? At minimum, the thesis should discuss the external validity of the results.

Paper 3: Why has science become an Old Man's game?

- This paper is written jointly with two senior coauthors and lacks a CRediT (Contributor Roles Taxonomy) statement. Such a statement would be valuable to accurately assess the Candidate's contribution to the paper.
- It would be helpful, though not necessary, if the Candidate could use to model to explain what would happen should the system be in a steady state (e.g., assuming a 2-percent growth rate). This would help put the present results in perspective.
- I appreciate the impressive data collection effort. And I would encourage the Candidate to provide more information on the directories. Do they really offer comprehensive coverage of the population of interest? A total of 3464 scientists in 1961 does not seem a large number. That said, comprehensive coverage is not necessary for the present study. What matters is that the coverage is consistent across years (i.e., covering the same universities). Otherwise, some of the entries and exits would simply result from a change in coverage. As this comment suggests, more information on the directories would be helpful.
- It might be possible to better infer entries and exits by matching scientists to publication data (and, therefore, affiliation). My hunch is that it would not change the substantive conclusions of the paper, but a short discussion might be helpful. I am certainly not suggesting matching scientists to publication data.
- In the note to Figure 3.4, why would the mean age of scientists "have stayed close to that of the 1960s" and not exactly the same? A clarification would be helpful.
- The conclusion mentions that "increasing the hiring of faculty members is not often discussed in policy circles despite its obvious potential to increase scientific productivity." This does not strike me as the most relevant policy implication.

Increasing the pool of scientists is a one-off action (unless we keep growing to unsustainable levels) that is costly and will only have a transitory effect. A more obvious policy choice would be to force old professors to retire. From this perspective, it might be helpful for the paper to consider a gradual (or radical) limit of 75 years (and 80, 70, and 65 for comparison purposes).

Recommendation

Upon careful inspection of the thesis, I am happy to confirm that it satisfies formal and content requirements for a PhD thesis in Economics. Furthermore, I recommend the dissertation for a defense.