International collaborations and quality of innovations: With whom you partner matters!

Methodological explanations and robustness checks

Dragan Filimonovic* and Christian Rutzer†

May 2020

1 Introduction

In the interactive report of Filimonovic & Rutzer (2021), we analyze how international R&D collaborations affect the quality of innovations. In contrast to the literature, which has so far only examined whether international collaborations have an impact on the quality of innovations Kerr & Kerr (2018), we investigate with which countries partnerships are likely to be particularly valuable. In doing so, we follow the literature and approximate the quality of an invention by the number of received forward citations (see, for example, Kogan et al. 2017, Hall et al. 2005). In this technical note, we describe our estimation strategy in more detail. We start by describing the data. We then set out our empirical model. Finally, we show various estimation results and, in addition to the results presented in Filimonovic & Rutzer (2021), also perform a wide variety of robustness checks. In general, our results remain robust.

2 Data

We use patent data to analyze the effect of different collaboration locations on the quality of innovations. In particular, our data consist of patents registered at the US Patent Office (USPTO). Each patent contains information on inventors' and assignee' (most often one or more firms which are the patent owners) countries of residence. This enables us to geographically classify patents and determine whether they are internationally collaborative in their nature. Given that patents may have more than one assignee, we restrict our sample to those patents that have only one assignee. We refer to assignees as legal owners of patents. Otherwise, it is impossible to properly determine whether there is international collaboration within a (or among) research team(s). By losing a small portion of observations (the majority of patents have only one legal owner), we gain the opportunity to pinpoint a single country as the patent's home nation. We classify a patent as

*Center for International Economics and Business | CIEB, Faculty of Business and Economics, University of Basel; dragan.filimonovic@unibas.ch.

†Center for International Economics and Business | CIEB, Faculty of Business and Economics, University of Basel; christian.rutzer@unibas.ch.

international collaborative if at least one of the inventors is not from the same country as the assignee. If all inventors are from the same country as the assignee, then we classify such a patent as a solely domestic one.

Further, we restrict our sample to patents published between 1990 and 2015. We set 1990 as the start because the internationalization of R&D activities only began to evolve rapidly in the early 90s. The year of 2015 is set as the end period of our sample as it is often shown in the literature that patents receive most of their citations within the first five years of their life (see, e.g., Breschi et al. 2006) and our quality indicator (based on patents' received citations) is calculated on a 5-year basis. Thus, we cannot include years after 2015. Finally, we also drop patents whose legal owner is located in the USA. This is done to prevent a strong bias towards non-collaborative patents driven by US based assignee firms as more than 90% of their inventors have a residency in the USA across all fields. Having in mind that the USA is the world's leading innovation hub and given that the vast majority of its inventions is created domestically by local inventors, leaving the USA in our sample would strongly bias our estimates by giving enormous weight to purely domestic patents. In addition, we use data from the OECD (2020) to consider so-called patent equivalents (i.e., patents which cover the same invention). We then combine all equivalent patents to one artificial patent and calculate the number of received forward citations of the artificial patent.[1]

Our dependent variable is constructed as a quality indicator, meaning that it distinguishes the top-10% most cited patents of each year and broad technological field[2] within the first five years since its publication, from all other patents placed in the lower part of the citation distribution. In contrast to pure citation numbers, such a binary indicator allows us to take into account the patent's relative significance of received forward citations in a particular field and to give more weight to the period where most of patents usually receive the majority of their citations. We also present estimates for different distributional cut-off values – a less restrictive (top-25%) and a more restrictive one (top-1%) – which can be found in Section 4.2. Utilizing other available information about a patent, we also construct additional variables that we used in our regressions. These are: the number of total scientist involved, the number of claims and a so-called originality index. The originality index was constructed by the OECD (for more details, please visit OECD, 2015). It measures the dispersion of a patent's backward citations across different technological fields. More specifically, it rests on the idea that the larger the knowledge pool on which a patent builds upon (i.e., the number of backward citations from different technological fields), the more significant should the innovation codified in a patent be. The index ranges from 0 (lowest originality) to 1 (highest originality).

[1]For more details on how to deal with equivalent patents when using forward citations see Webb et al. (2005).

[2]The IPC classification has been developed by the Schmoch (2008). It covers 5 broad technological groups: **I Electrical engineering:** 1. Electrical machinery, apparatus, energy 2. Audio-visual technology 3. Telecommunications 4. Digital communication 5. Basic communication processes 6. Computer technology 7. IT methods for management 8. Semiconductors **II Instruments:** 9. Optics 10. Measurement11. Analysis of biological materials12. Control13. Medical technology **III Chemistry:** 14. Organic fine chemistry 15. Biotechnology 16. Pharmaceuticals 17. Macromolecular chemistry, polymers 18. Food chemistry 19. Basic materials chemistry 20. Materials, metallurgy 21. Surface technology, coating 22. Micro-structural and nano-technology 23. Chemical engineering 24. Environmental technology **IV Mechanical engineering:** 25. Handling 26. Machine tools 27. Engines, pumps, turbines 28. Textile and paper machines 29. Other special machines 30. Thermal processes and apparatus 31. Mechanical elements 32. Transport **V Other fields:** 33. Furniture, games 34. Other consumer goods 35. Civil engineering

Table 1: Summary Statistics, main sample

Statistic	N	Mean	St. Dev.	Min	Max
% top patents (90th %tile)	4,810,687	0.1	0.3	0	1
No. of inventors	4,810,687	2.7	2.0	1	118
No. of claims	4,810,687	16.1	11.3	1	887
Originality	4,810,687	0.7	0.2	0.0	1.0
% with foreign inventor	4,810,687	0.1	0.3	0	1
% only domestic inventors	4,810,687	0.9	0.3	0	1

Even though our main sample went through harsh restrictions, it still contains a large number of observations. Our main estimates cover more than 4,8 million patents, where the average number of inventors per patent is around 3 and the average number of claims slightly above 16. The share of patents with at least one foreign inventor (inventor that resides in a country different than the patent's home country) is around 10%, meaning that the share of patents with only domestic inventors in our sample is strongly represented. However, the portion of collaborative patents is much larger for industrialized countries, for instance, the share of Swiss patents with at least one inventor from abroad in the period 2010-2015 was 48%.

3 Model

In order to investigate the link between international cooperation and the quality of a patent we estimate a linear probability model of the following form:

$$\begin{aligned} WorldClass_{i,t} = \alpha + TeamSize_i + Claims_i + Originality_i \\ + PartnerCountry_i + FE_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{1}$$

where our dependent variable $WorldClass_{i,t}$ is a dummy variable capturing the quality of an innovation. It equals 1 if a patent is among the top-10% most cited patents and 0 otherwise. $TeamSize_i$ represents the log of total number of inventors who participated in a patent's development, $Claims_i$ is the log number of claims made by a patent and $Originality_i$ is an index ranging from $0-1$ capturing a patent's originality. By controlling for these variables, we aim to exclude the effect of some additional factors that are important for a patent's quality. In particular, we want to remove the impact of "brain power", complexity and scope of an innovation. This brings us a step closer to isolating the pure effect of international collaboration.

$PartnerCountry_i$ is a vector that captures the effect of the country of residence of inventors when they reside outside the legal owner's country (i.e., assignee of the patent). These are our coefficients of interest. It is a set of dummy variables for each possible location of an inventor (major industrialized countries are represented by individual variables while all other partners are classified under the same variable – Rest). A specific dummy variable equals 1 if a patent is collaborative (i.e., if at least one inventor resides abroad) *and* if an inventor resides in that specific country. If a patent has only domestic inventors or if an inventor is from abroad but not from the particular foreign location it will equal 0. To illustrate this, imagine that a Swiss patent (owned by a Swiss company) is a product of collaboration between inventors residing in Switzerland and Germany. In this case, the variable indicating Germany (within the $PartnerCountry_i$ vector) will equal 1 while all other location indicators (e.g. France, Italy, USA, etc.) will be coded with 0 for this particular patent. This approach simplifies the computation of collaboration effects across all partner countries and determines the average impact of a collaboration with a particular country relative to purely domestic patents (those with local inventors only) straightforwardly.

Following suggestions by Kerr & Kerr (2018), we also include a full set of specific fixed effects ($FE_{i,t}$) for year of publishing, broad technological group and the country of a patent's owner. We also introduce year-technology and year-owner's country interaction terms to account for all year-technology-country specific unobservable factors that may affect the number of forward citations of patents.

4 Estimation results

In the following Section, we present our estimations. We start with the benchmark model and afterwards using alternative quality indicators and triadic patents (instead of USPTO patents) to demonstrate the robustness of our results and justify some of our sampling decisions.

4.1 Standard specification

Let us start with the results of the standard specification as used in our interactive report Filimonovic & Rutzer (2021). In this case, the empirical model (1) is estimated based on the following binary classification of patents: the 10% most cited patents and all others.

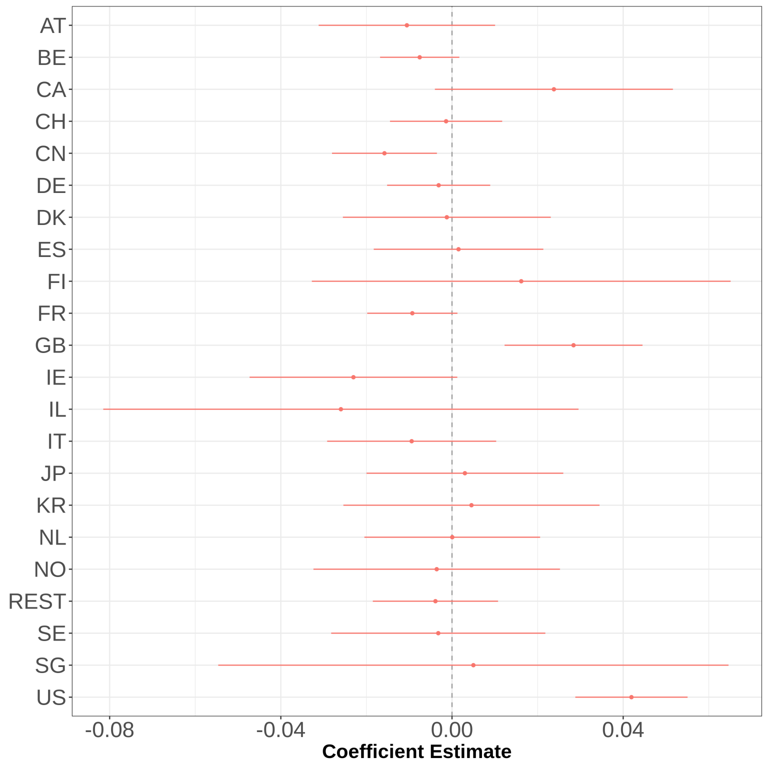
As one can see in Figure 1, only international collaborations with the USA or Great Britain lead on average to innovations of higher quality compared to purely domestic innovations. It is also evident that international research cooperations are in almost all cases at least not worse than solely domestic inventions.

4.2 Alternative world class indicator

In the following, we estimate the model specified by equation (1), but use a different cut-off for classifying our dependent variable $WorldClass_{i,t}$. In particular, we create the indicators $WorldClass75_{i,t}$ and $WorldClass99_{i,t}$, where the first one captures patents that are in the top $25th$ percentile of the forward citation distribution and the latter indicates those patents that are in the top $1st$ percentile of the forward citation distribution, in a given technological field within the first five years since publishing. Therefore, the former is less restrictive and the latter represents a more restrictive classification of top-patents compared to our main indicator.

The estimates in Figure 2 show how the effects of collaborating with different partner countries change when using a less or more restrictive definition of our quality indicator. In the less restrictive case (left panel in Figure 2), the USA and Great Britain still demonstrate the biggest and the most significant effects on the quality of innovations, while the coefficients for some other countries also become significant. When using a more restrictive definition (right panel in Figure 2), almost all coefficients lose their statistical power except for the USA. Even though there are some deviations between the estimates, both

Figure 1: The effect of partnering country on the quality of an innovation: benchmark model



Notes: The estimates show the partnering-country coefficients of model (1) where the dependent variable equals 1 if a patent is among the top-10% most cited patents and 0 otherwise. The horizontal bar graphs show the 95% confidence intervals using clustered standard errors at the level of the legal owner country as in Kerr & Kerr (2018).

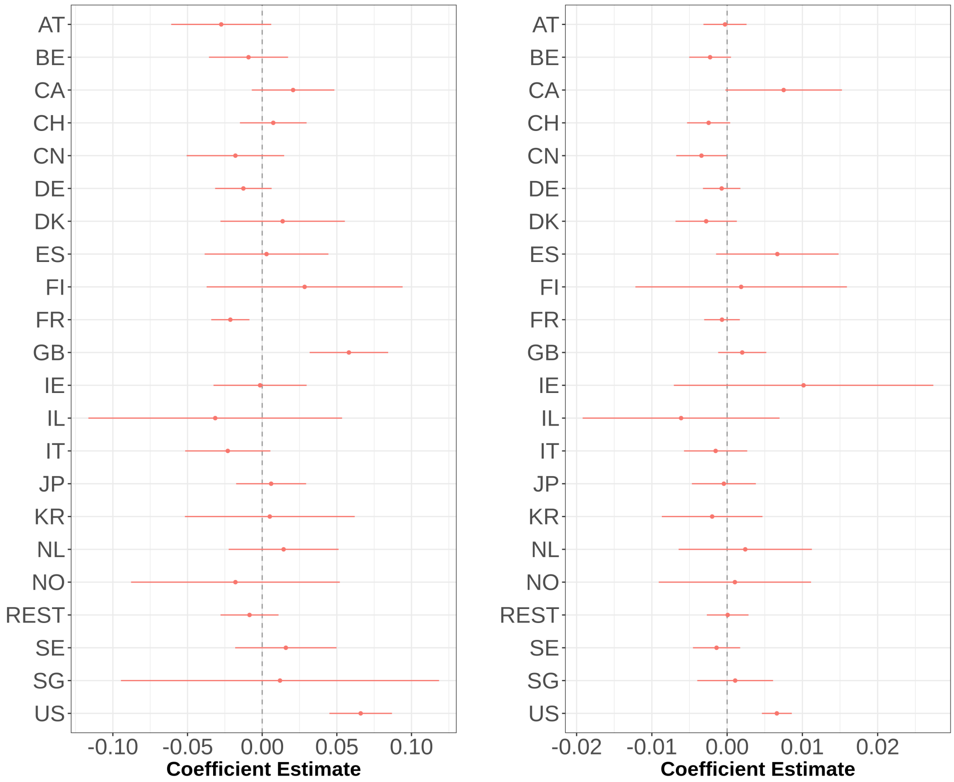
alternative definitions converge with our benchmark estimations suggesting that collaboration with the USA, primarily, but also with Great Britain potentially has the largest benefit regarding the quality of innovations. Taking the top-10% most cited patents, thus, represents a reasonable solution to the trade-off between the number of identified world class patents and the ability to truly approximate the best innovations.

4.3 Sample with triadic patents

A possible concern may arise due to the fact that we use only patents registered at the USPTO. If better patents are more often developed and registered in the USA or if the chances that a patent receives citations are substantially higher when it is registered with a particular patent office, in this case the USPTO, our estimates would be biased. In other words, if any type of strong "self-selection" among patent applications exist, our estimates could capturing this rather than the effects of collaboration. Therefore, we present our estimates of model (1), using a sub-sample of triadic patents only.

The term triadic means that a patent is registered at all three major patent offices: the European Patent Office (EPO), the Japan Patent Office (JPO) and the United States Patent and Trademark Office (USPTO). By taking into account only triadic patents, we should significantly reduce the possibility of bias coming from decisions on where a patent is registered, but on the other hand, we reduce the sample size. This would additionally reduce the statistical power of our estimates, especially for specific partner-technology samples. Thus, we do not make this restriction in the main sample.

Figure 2: The effect of the partnering country by different forward citation cut-offs



Top 25% most cited patents   Top 1% most cited patents

Notes: The left panel shows estimates of $PartnerCountry_i$ coefficients of model (1) *but where the dependent variable equals* 1 *if a patent is among the top-*25% *most cited patents and* 0 *otherwise. The right panel shows estimates of $PartnerCountry_i$ coefficients in model 1 but where the dependent variable equals 1 if a patent is among the top-1% most cited patents and 0 otherwise. The horizontal bar graphs show the 95% confidence intervals using clustered standard errors at the level of the legal owner country as in Kerr & Kerr (2018).*

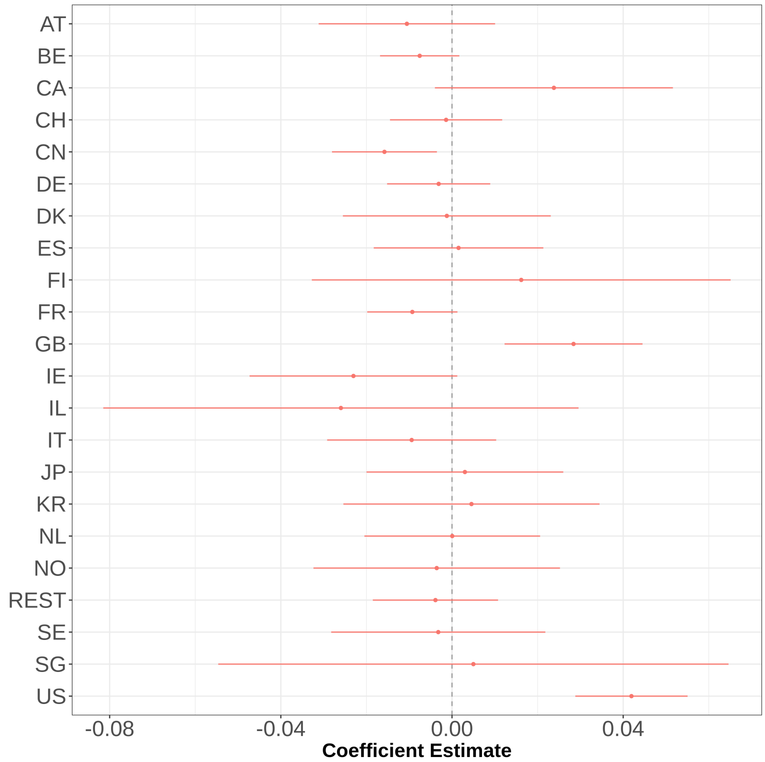
The findings from the triadic patent sample (Figure 3) are very similar to our main estimates presented in the report and in this technical note in Figure 1. This suggests that a potential bias linked to registrations at the USPTO, and for that matter any patent registration office, is negligible and that it should not have any substantial biasing effects on our main results. Our estimates show again the USA and Great Britain as the two partner locations with the most significant impact on the quality of an innovation.

4.4 Patent claims as the dependent variable

The number of forward citation is often used as a proxy of the quality of an innovation codified by a patent but it is not the only indicator employed for this purpose. Many scholars rely on several indicators or even employ composite systems of individual measures in analyzing the quality of innovations. Thus, we will check our estimates of collaboration effects using a completely different indicator - the number of claims.

For example, Song & Li (2014) suggest that both forward citations and claims are important components of the quality of an innovation. Hence, they use both to develop a composite quality index. It is argued that claims are the final goal of an inventor when

Figure 3: The effect of the partnering country on the quality of an innovation (triadic patents)



Notes: The estimates are based on the subset of triadic patents. The estimates show the partnering-country coefficients of model (1) where the dependent variable equals 1 if a patent is among the top-10% most cited patents and 0 otherwise. The horizontal bar graphs show the 95% confidence intervals using clustered standard errors at the level of the legal owner country as in Kerr & Kerr (2018).

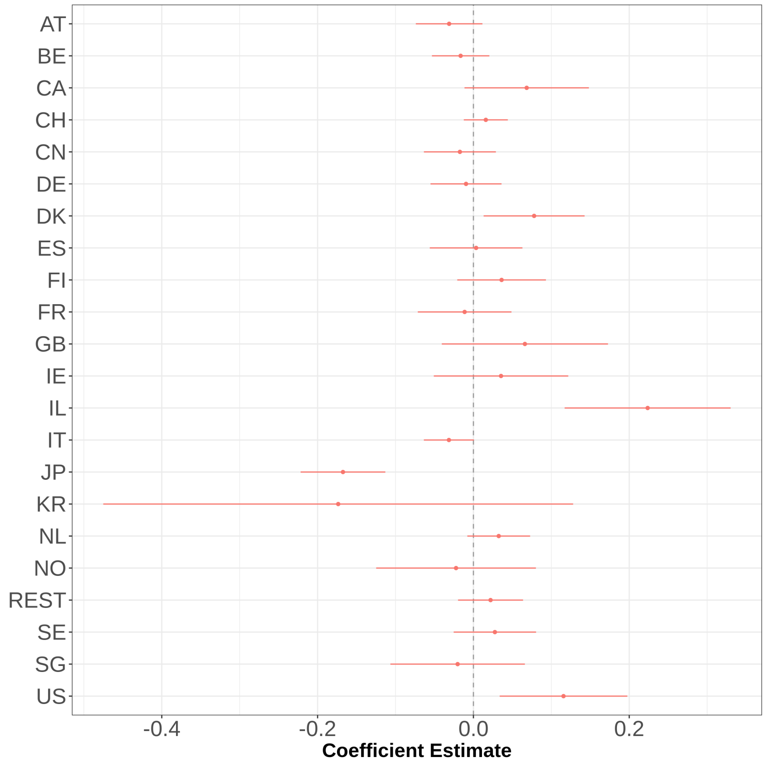
applying for a patent meaning that the larger this number is, the more innovative solutions across different fields are protected. In addition, one could argue that the likelihood of getting cited increases if inventors come from English speaking countries. Given that all our estimates suggest positive and significant effects for the USA and Great Britain, this may imply that our results are capturing to some extent the quality of the patent document as well. Moving away from citations and using claims as a quality indicator can help us in estimating the extent of such bias.

The results presented in Figure 5 suggest that collaborations with the USA also tends to produce the patents with the most claims relative to patents with home inventors only. The coefficient for Great Britain losses its statistical power but its tendency remains positive. We also see that cooperation with some other countries tends to produce significant effects on the number of patent claims such as Israel, Canada, Sweden and the Netherlands. This also suggests that these two measures are not perfectly correlated and do not measure exactly the same dimension of patent quality. This is in line with the work of Song & Li (2014). According to them, the number of claims is more related to the quality of a patent document than to the quality of the innovation itself as the number of claims hampers the examination procedure.

4.5 Swiss patents only

As a final robustness check, we present our estimates for the sample of Swiss patents. This means that we leave only those patents whose legal owner is a company or an

Figure 4: The effect of the partner country on the number of claims

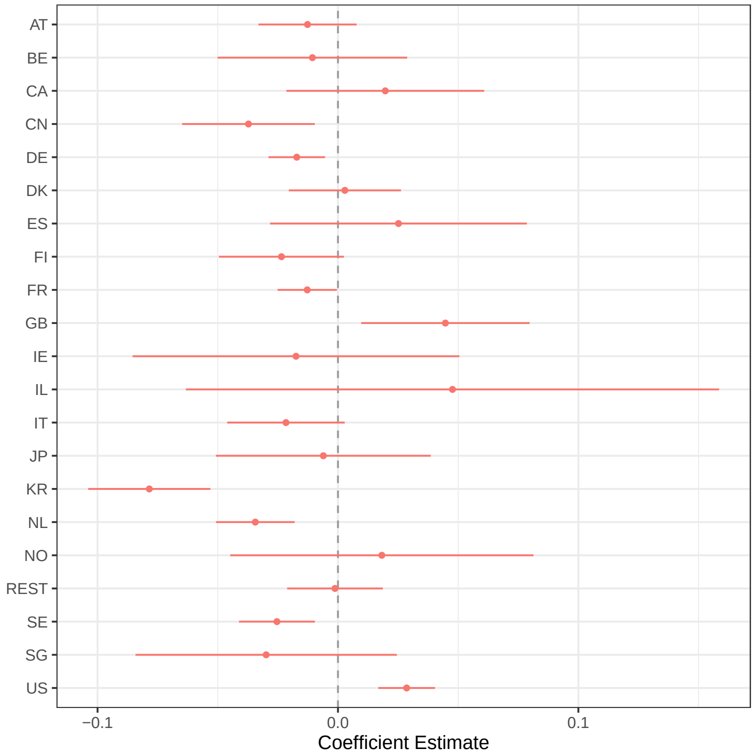


Notes: The horizontal bar graphs show the 95% confidence intervals using clustered standard errors at the level of the legal owner country as in Kerr & Kerr (2018).

academic institution residing in Switzerland. We employ the model in (1) with some small adjustments. More specifically, as we now have only one legal owner country, we lose the possibility to control for legal owner fixed effects and, thus, we also cluster standard errors on the technology level. All other regression components remain unchanged. This estimation could be especially relevant given that our entire Innoscape project (to which this report belongs) intends to investigate the Swiss innovation landscape. Therefore, obtaining a direct estimation on which countries are the most beneficial partners from a perspective of Switzerland provides a good insight how the general results apply to the Swiss case. However, we do not include these estimates in the main report as once we restrict the patent sample to one country we get a substantial decrease in the number of observations that could significantly impact the results of such estimations and question its reliability.

Nonetheless, we see that the general pattern found in our main sample also holds for the Swiss case. Again, the USA and Great Britain are the only two partner countries with a positive and significant impact on the probability that a patent becomes a world class innovation, relative to purely domestic Swiss patents (i.e., those which have all inventors residing in Switzerland). This means that collaborating with firms and research locations from the USA and Great Britain on RD projects is very important for Swiss firms when it comes to the quality of their innovations. The emergence of some negative coefficients should not be interpreted as an indication that Swiss firms should cooperate less with some partners. Given that Switzerland is a highly-innovative environment – a world leading research hub in many areas – it is not surprising that many collaborative patents do not have any impact or have a slight negative effect on the probability of developing a high quality invention relative to non-collaborative, purely domestic Swiss patents.

Figure 5: The effect of the partner country on the Swiss patent quality



Notes: The horizontal bar graphs show the 95% confidence intervals using clustered standard errors at the level of the technological field.

References

Breschi, S., Tarasconi, G., Catalini, C., Novella, L., Guatta, P. & Johnson, H. (2006), Highly cited patents, highly cited publications, and research networks, Technical report.

Filimonovic, D. & Rutzer, C. (2021), International collaborations and quality of innovations: With whom you partner matters!, Interactive report. The report can be accessed at: `https://innoscape.ch/en/publications/international-rd`.

Hall, B. H., Jaffe, A. & Trajtenberg, M. (2005), 'Market value and patent citations', *The RAND Journal of Economics* **36**(1), 16–38.

Kerr, S. P. & Kerr, W. R. (2018), 'Global collaborative patents', *The Economic Journal* **128**(612), 235–272.

Kogan, L., Papanikolaou, D., Seru, A. & Stoffman, N. (2017), 'Technological innovation, resource allocation, and growth', *The Quarterly Journal of Economics* **132**(2), 665–712.

OECD (2020), Sti microdata lab, Raw patent data. Url: `https://survey2018.oecd.org/Survey.aspx?s=85071fe20881410ebed2258aa9f93561`.

Schmoch, U. (2008), Concept of a technology classification for country comparisons, Report, World Intellectual Property Organisation (WIPO).

Song, H. & Li, Z. (2014), 'Patent quality and the measuring indicator system: Comparison among china provinces and key countries', *Berkeley center for law & technology IP scholars conference* .

Webb, C., Dernis, H., Harhoff, D. & Hoisl, K. (2005), 'Analysing european and international patent citations', *OECD Working Paper* .