Working Paper Series

Technology protectionism and the patent system: Evidence from China

Gaétan de Rassenfosse
Emilio Raiteri

September 2020

Innovation and Intellectual Property Policy
Working Paper series no. 11

Available at: https://ideas.repec.org/p/iip/wpaper/11.html

Green open access version. This paper is forthcoming in the *Journal of Industrial Economics*.
DOI: 10.1111/joie.12261

Contact information @gderasse
http://www.gder.info

Technology protectionism and the patent system: Evidence from China

Gaétan de Rassenfosse[a,*], Emilio Raiteri[b]

[a] *College of Management of Technology, École polytechnique fédérale de Lausanne (EPFL), Switzerland.*
[b] *School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands*

Abstract

Governments have strong incentives to allow their inventors to free ride on foreign technologies. They can achieve this result by discriminating against foreigners in the patent system—by refusing to grant foreigners a patent for their inventions. International patent law treaties forbid this practice, which may lower the global innovation incentives and may hurt international trade. Using data on half a million inventions submitted to the Chinese patent office, we find robust evidence of anti-foreign bias in the issuance of patents in 'strategic' technology areas. Foreigners are about 50 percent more likely to be refused a strategic patent than locals.

Keywords: *industrial policy; national treatment principle; patent; technology protectionism; TRIPS*

JEL Classification Codes: O34; K11; F52

*Corresponding author. ODY 201.1, Station 5, CH-1015 Lausanne. gaetan.derassenfosse@epfl.ch

I. Introduction

> "Industry representatives express mixed opinions on whether there is *anti-foreign bias* in the issuance or enforcement of patents in China. However, some non-Chinese firms reportedly find it more difficult to obtain patents in sectors that the Chinese government considers of *strategic importance*."
>
> United States International Trade Commission (ITC) 2010, page xviii (italics added)

Regulation covers virtually all aspects of firm operation, including finance, human resources, production, marketing and trade (e.g., Smith and Grimm, 1987; Shaffer, 1995; Plambeck and Wang, 2009; Goldfarb and Tucker, 2011). Intellectual property (IP) rights form one set of regulations that are critical to technology-based firms. Malfunctioning of the IP system, such as excessive delays or mistakes in the patent examination process, may severely hurt the firm or its competitors. Consequently, a large number of economic, legal, and management scholars have devoted attention to the functioning of the patent system. Scholarly discussions focus mainly on questions related to the *efficiency* of the patent system. Less attention has been paid to the issue of *fairness*.

As with standard trade barriers, IP rights may be awarded and exploited in such a way as to discriminate against foreign interests (Maskus, 2000). Governments typically want the strongest possible protection in foreign countries in order to maximize returns to domestic firms, and the weakest possible protection for foreign firms in their domestic markets to facilitate free-riding on foreign technologies (Scotchmer, 2004).

Such opportunistic behaviors may lower the incentives for global innovation and may hurt international trade of technology-intensive goods (Palangkaraya et al., 2017). To prevent such behaviors, international IP treaties impose the 'national treatment' principle, which states that within each country, foreign applicants must receive treatment equal to that accorded to domestic applicants. This principle also ensures that the patent system is not used as a de facto trade protection policy.

The present paper is concerned with the situation at China's patent office, the National Intellectual Property Administration (CNIPA, formerly SIPO). The claim that the national treatment principle is not being observed at CNIPA has been noted in official reports and is regularly echoed in the press.[1]

This paper tests for anti-foreign bias in the issuance of patents at CNIPA, with a specific focus on patent applications in 'strategic' technology areas. The analysis relies on a sample of about half a million patent applications filed at CNIPA in the period from 2001 to 2009. We identify strategic areas with the help of patent examiners at the World Intellectual Property Organization (WIPO) based in Geneva. We rely on their expertise to link technologies described in the central planners' long-term development plan (SCPRC, 2006) to technology classes listed in the patent documents.

The econometric identification exploits information on the grant outcome of more than 1.6 million exact 'twins' of patent applications in our sample (à la Webster et al., 2014). These twins are filed in other jurisdictions where applicants also sought to protect their inventions. We use them to form an expectation of the probability of grant at CNIPA in order to build our counterfactual.

We find no—or at worst only weak—overall discrimination against foreigners at CNIPA, in contrast with recent evidence produced for Europe and Japan (Webster et al., 2014). However, foreign patent applications in strategic technology areas are about four to seven percentage points less likely to receive a patent grant than similar domestic applications. This figure translates into a risk ratio of about 1.5, meaning that foreign applications in strategic areas are 50 percent more likely to be rejected than domestic applications. Furthermore, we observe a 6.5 percentage-point increase in the probability that strategic patent applications by foreigners will experience a decrease in scope. Thus, not only are foreigners less likely to have their patent applications granted, but they also obtain narrower patents when they do. Given the importance of industrial policy in China and the country's strong focus on indigenous innovation and intellectual property, we argue that the empirical results provide a case of technology protectionism by means of the patent system.

The rest of the paper is organized as follows. Section II provides background information on technological planning and the patent system in China. Section III describes the empirical strategy and Section IV presents the data. Section V discusses the results of the econometric analysis and Section VI presents robustness checks. Section VII concludes.

II. Background

II(i). Central planning of technology development

After Deng Xiaoping became China's leader in 1978, the Chinese government undertook a series of reforms and transitioned to a 'socialist market economy' (e.g., Suliman, 1998).

In this context, the government implemented several innovation-related policies aimed to realize Deng's view that Science and Technology (S&T) should be a primary productive force (OECD, 2008; Liu et al., 2011). From the beginning of the 1980s, science, technology, and related industrial policies were explicitly designed to stimulate the development of advanced technologies to free China from financial obligations for foreign technologies. These policies included the 'Key Technologies R&D Program,' the '863 Program' (the State High-technology Program, started in 1986), the '973 Program' (the State Program for the Support of Basic Research and Development), and the 'Golden Projects' program.

Technology development continues to occupy an important place in the economic planning strategy of the Central Committee of the Communist Party of China. In January 2006, the State Council issued the "National Medium and Long-Term Program for Science and Technology Development 2006–2020" (MLP), whose guiding principle is to make China an innovation-driven nation by "fostering indigenous innovation, leapfrogging in priority fields, and leading the future" (SCPRC, 2006, p. 7).[2]

The MLP identifies priority technology areas and topics that the central planners consider critical for the country's economic and social development. The plan also sets forth a list of ambitious S&T goals to be achieved by 2020, one of which directly concerns IP: gross spending on R&D must meet or exceed to 2.5 percent of GDP; dependence on imported technology must fall below 30 percent; the country must move into the top five countries for the number of invention patents granted to nationals. In fact, IP receives support at the highest level. Hu Jintao, a former president, is reported to have said on many occasions that "competition in the future is competition in IP" (The Economist, 2015).

II(ii). The Chinese patent system and the national treatment principle

China joined the WIPO in 1980 and issued its first patent law in 1984. Since then, Chinese patent law has been revised three times—in 1992, 2000, and 2008—to align it with international standards.[3] The CNIPA is a major international player on the patent scene. It became the world's largest patent office in terms of national applications in 2011 and is the second largest after the U.S. Patent and Trademark Office (USPTO) for international applications (WIPO, 2015).

A landmark change in IP law occurred in 2001 when China joined the World Trade Organization (WTO) and signed the so-called TRIPs agreement (for Trade-Related Aspects of Intellectual Property Rights). Article 3 of the TRIPs agreement affirms the national treatment principle, stating that "each Member shall accord to the nationals of other Members treatment no less favorable than that it accords to its own nationals with regard to the protection of intellectual property." This provision is a key pillar of international patent law. It was mentioned in the 1883 Paris Convention for the Protection of Industrial Property, to which China became a signatory member in 1984.

Discussions about unfair treatment of foreign firms in China have so far focused on *enforcement* of IP rights. Observers note that damages for patent infringement in China are too low to discourage infringement, and discuss potential national preference by Chinese courts (e.g., Love et al., 2016). The U.S. Government and the European Commission are pushing for more aggressive enforcement in China through what is known as TRIPS+ provisions.

Although enforcement is a worthy object of study in its own right, in this matter, it may be thought of as a second-order issue. If foreign applicants are wrongfully denied a patent in the first place, local competitors may legally use, produce, and sell their inventions in the home market. It is thus of primary importance to assess whether the patent system itself, rather than the judicial system, discriminates against foreigners.

II(iii). Concerns in the developed world

Several observers in the United States and Europe have raised concerns that China's policies favoring indigenous innovation are hidden forms of technology protectionism. A report from the U.S. International Trade Commission suggests that this "web of interrelated indigenous innovation policies" may work together to favor domestic over foreign companies in the Chinese market, and that such a discriminatory effect could be especially strong for companies operating in sectors considered strategic by the Chinese government (USITC, 2010). The USITC also reports a close link between these measures and infringement of IP rights in China, stating that through indigenous innovation policies, China "undermines and displaces foreign IP while promoting its own IP" (USITC, 2010, Ch. 5, p. 8).

A USPTO report echoes this view, stating that "numerous commenters articulated the perception that China's patent system, including enforcement mechanisms, benefits Chinese companies at the expense of U.S. and other foreign companies" (USPTO, 2010, p. 5). In a report prepared for the U.S. Chamber of Commerce, McGregor (2010) makes the particularly bold claim that "the [MLP] is considered by many international technology companies to be a blueprint for technology theft on a scale the world has never seen before"(McGregor, 2010, p. 26). More recently, the Trump administration has been particularly vocal on the issue, following the release of a report by the bipartisan Commission on the Theft of American Intellectual Property (The IP Commission, 2018).

In light of these concerns, it is legitimate to ask whether one can find traces of discrimination in the patent system and how important such traces are. As far as we can ascertain, little empirical evidence exists.

II(iv). Empirical evidence

History is rich in examples of developing nations having set up a patent system favoring domestic inventors. For example, until 1836 in the United States, foreigners were not allowed to obtain U.S. patents unless they had resided at least two years in the United States and declared an intent to become U.S. citizens (Scherer, 2004). The 1883 Paris Convention for the Protection of Industrial Property and the TRIPS Agreement were implemented precisely to avoid such behaviors.

To our knowledge, Webster et al. (2014) and de Rassenfosse et al. (2019) are the only studies investigating the practical application of the national treatment principle in the prosecution process in modern IP law. Using a sample of about 50,000 patent applications granted by the USPTO and filed in the early 1990s at the European Patent Office (EPO) and the Japanese Patent Office (JPO), Webster et al. (2014) found that domestic applicants were more likely than foreign applicants to be granted patent protection, all else equal. They take this result as evidence of a violation of the national treatment principle. de Rassenfosse et al. (2019) show that filing international patents under the

Patent Cooperation Treaty can reduce some of the bias.

Three studies provide correlational evidence for the particular case of China. Yang (2008) compares aggregate rates of issuance and pendency between international and domestic patent applications at the USPTO and CNIPA. The author finds no significant difference in average pendency between national and international applications at CNIPA relative to the USPTO but reports evidence of a higher rate of issuance for domestic applications at CNIPA. Liegsalz and Wagner (2013) focus on the pendency of applications filed at CNIPA between 1990 and 2002 and account for patent-level characteristics. They show that Chinese applicants receive patents faster than foreign applicants and that the difference in the grant lag is particularly large in technology fields in which China has a relative technological advantage over other nations. However, the study pre-dates China's accession to the TRIPs agreement. More recent evidence by Tong et al. (2018) confirms that local applicants at CNIPA experience shorter grant delays compared to foreign applicants.

III. Empirical framework

Our empirical analysis seeks to evaluate the extent to which the probability of being granted a patent at CNIPA differs for foreign and Chinese applicants, all other things being equal. We pay close attention to the fate of applications that are in areas of strategic importance to China.

III(i). Identification strategy

The gist of the identification strategy is to control for the probability that a Chinese patent application for a focal invention will be granted by looking at the grant outcome of patent applications *for the same invention*—so-called 'twin patents' or 'direct equivalents'—in other patent offices. The grant outcome at other patent offices forms the counterfactual outcome.

Patent systems are national, and inventors seeking international patent protection must apply for a patent in each desired jurisdiction, who take a sovereign decision about patent issuance. We track applications for the same invention across up to seven patent offices, as explained in Section III(iii). Note that the use of twin patents requires restricting the sample to patent applications with at least one international family member, that is, inventions that are filed in at least two countries.[4] Thus, our analysis relates only to inventions with global market potential.

We implement the identification strategy in two ways: using an invention fixed effect specification; and using a control-variable specification. Both approaches have their advantages and disadvantages, which we discuss below.

III(i).1. Fixed-effect approach

The fixed-effect approach follows closely Webster et al. (2014). It is an elegant way of assessing the presence of anti-foreign bias in patent examination outcome. In our context, the invention fixed-effect panel regression model can be written as:

$$(1) \quad y^*_{io} = \beta_0 + \beta_1 F_{io} + \beta_2 S_{io} + \beta_3 (F \times S)_{io} + \gamma_i + \gamma_o + \epsilon_{io}, \quad y_{io} = 1[y^*_{io} > 0]$$

where y^*_{io} is the latent variable underlying the binary grant outcome y_{io} of patent application for invention i in patent office o. The variable F is a dummy that takes value 1 if the applicant is foreign to office o, and 0 if the applicant is domestic. A coefficient $\beta_1 < 0$ suggests anti-foreign bias. The variable S is a dummy that takes value 1 if the patent application is in a strategic area for office o, and 0 otherwise. The variable $(F \times S)_{io}$ is an interaction term that tests for the presence of a specific anti-foreign bias in strategic areas. The terms γ_i and γ_o denote invention and office fixed effects, respectively. We estimate equation (1) as a linear probability model (LPM) to facilitate the treatment of fixed effects. Given that the variables of interest are dummy variables, coefficients of the LPM can be interpreted as differences in group means, which is what we want.

This approach controls for unobserved heterogeneity across inventions. However, it comes with two main disadvantages. First, it assumes implicitly that there is no relevant office-patent application heterogeneity that affects the grant outcome other than being an applicant foreign to office o or in an area strategic to office o. Controlling for such heterogeneity is possible, although extremely expensive in this set-up because one needs to collect data for all patent offices considered. For instance, if one believes that the number of claims listed in the patent application may correlate with foreign origin and affect the grant outcome, one would need to collect claim data for all the patent offices. As the number of control variables grows, data collection becomes increasingly complex and expensive—and sometimes simply impossible because data may not be available for all offices.

Second, this approach assumes that the variable of interest exists for all offices considered. In the present context, the list of strategic technology areas is specific to China and, hence, to CNIPA. We are not aware of similar lists in other countries. Thus, the variable S_{io} only exists at CNIPA, such that $S_{io} = S_{i,China}$. For these two reasons, our preferred specification is an office-specific regression, as explained in the next section.

III(i).2. Control-variable approach

In addition to the fixed effect approach, we model the examination outcome at CNIPA only. The baseline model is

$$(2) \quad y_i^* = \beta_1 F_i + \beta_2 S_i + \beta_3 (F \times S)_i + c_i + \epsilon_i, \quad y_i = 1[y_i^* > 0]$$

where y_i^* is the latent variable underlying the binary grant outcome y_i at CNIPA. We assume that the error term ϵ_i has a standard logistic distribution, but we will also estimate a LPM for comparison purposes. In this set-up, the variable F captures systematic differences in patent applications between foreign and Chinese applicants. The variable S captures systematic differences in patent applications between strategic and non-strategic technological areas. The variable of interest, $F \times S$, captures a specific effect on foreign applicants' strategic patent applications.

The variable c_i plays the same role as the invention fixed effect in equation (1). We call it the 'patentability score:' it captures our best guess for the grant probability of the patent application. As explained further below (Section III(iii)), c_i is the average grant outcome for the twin applications. In the extreme case where all the twin applications were granted a patent, we would expect CNIPA to grant the patent as well. On the other hand, in the extreme case where all the twins were refused, we would expect CNIPA to refuse the patent. In other words, equation (2) models the determinants of the deviation from the expected probability of a grant of the patent application at CNIPA—thus, the coefficient of interest, β_3, is purged from the effect of individual differences in expected grant probability.

The sample includes all patent applications at CNIPA that have at least one foreign equivalent.[5] This restriction is desirable per se. Indeed, patent applications at CNIPA by foreign firms form a selected sample of applications, one for which applicants were willing to incur the substantial cost of international patent protection. By contrast, there is no selection in patent applications by locals, leading to a lower average quality. Imposing that all applications in the sample have a direct equivalent at selected patent authorities puts locals and foreigners on the same level.

The next element of the identification strategy captures features of the patent application that may be specific to strategic technologies by foreign firms. One such potential feature is the choice of the IP law firm, which may affect the grant probability in various ways. For instance, it may be correlated with the quality of the translation service into Chinese or may play a role through informal ties with patent examiners (Tabakovic and Wollmann, 2017). Furthermore, better IP law firms may be more able to identify, and navigate through, local prior art. The final model we estimate is

$$(3) \quad y_i^* = \beta_1 F_i + \beta_2 S_i + \beta_3 (F \times S)_i + c_i + \mathbf{X_i}\gamma + \epsilon_i, \quad y_i = 1[y_1^* > 0]$$

where $\mathbf{X_i}$ includes control variables that may affect the probability of a grant at CNIPA. As explained previously, the possibility to include a large number of controls is an advantage of the present approach over the fixed-effect approach. Section IV(iii) presents the list of covariates.

The last element of the identification strategy in the control-variable approach seeks to account for two potential sources of hidden bias in the construction of the variable c_i.

First, patent offices differ in the stringency of their granting requirements (de Rassenfosse et al., 2016). A grant from a strict office carries more weight than a grant from a lax office, all else equal. In an extension to the analysis, we build the variable c_{wi}, which accounts for the observed grant probability in each office, and use it in lieu of c_i. In concrete terms, we compute the variable c_{wi} as a weighted average of the grant outcomes at other authorities, where the weight is the reciprocal of the overall grant probability at a specific patent office. In this way, we put a higher (lower) weight to successful applications granted by more (less) stringent patent authorities.[6] Second, violation of the national treatment principle may not only occur at CNIPA. For instance, a twin application at the USPTO filed by a U.S. applicant may have a higher probability of receiving a grant than a non-U.S. application, all else equal. If foreign offices discriminate based on the applicant's country of residence, the variable c_i will not capture the true grant probability, leading to a biased estimator.[7] A first solution involves computing the variable as a leave-out-mean, i.e., systematically discarding information from the office of origin of the applicant in the computation of c_i (variable c_{xi}). This solution is quite extreme because it excludes information that may nevertheless be useful. A second solution involves developing an ad hoc test of unobserved heterogeneity inspired by Aakvik (2001) and popularized by Rosenbaum (2002). For this test, we assume that the variable c_i is biased, and we gradually remove bias by altering the grant outcome of the twin applications as explained further in Section III(iii).

In a logit regression model, it is important to note that a negative coefficient for β_3 would not necessarily be evidence of a systematic anti-foreign bias in strategic fields.

Indeed, the marginal effect of a change in both interacted variables in a nonlinear model is not equal to the marginal effect of a change in the interacted variable (Ai and Norton, 2003). We will use the method proposed by Norton et al. (2004) to estimate the magnitude of the marginal effect for the interaction term in an appropriate manner. Besides, the marginal effect of the interacted term may have a different sign for different observations and different values of the covariates. We will also depict the marginal effects of the interaction term over the range of predicted grant-probability scores, as suggested by Hoetker (2007).

III(ii). Identification of strategic technologies

We used the MLP to identify technologies of strategic importance. The plan describes 27 frontier technologies that should constitute the "basis on which future high technologies stem out and emerging industries grow" (SCPRC, 2006, p.33).[8] These frontier technologies fall into eight major technological fields: biotechnology, information technology, advanced materials technology, advanced manufacturing technology, advanced energy technology, marine technology, laser technology, and aerospace technologies.

In order to identify patent applications in these strategic areas, we linked the 27 frontier technologies to specific patent classes. In particular, we worked at the main group level as defined by the International Patent Classification (IPC) taxonomy.[9] The IPC is a hierarchical system for classifying patent applications according to the different areas of technology to which they pertain. The linking of technologies described in the MLP to IPC classes was done in two steps. First, we relied on *IPCCAT*, a tool that allows for automated patent classification based on text analysis.[10] Second, we validated the list with the help of three WIPO experts, which led to some refinements in the classification.[11] In particular, some IPC classes provided by the classification tool were too broad. At the end of the process, we identified 97 strategic main groups out of the 6,812 main groups used to describe the technological content of the patents in our sample.[12]

It is important to emphasize that we do not claim that the MLP is a medium for discrimination. Instead, we use the MLP to infer areas of strategic importance in a consistent manner. As a matter of fact, the MLP is not the first plan designed to support the development of strategic technologies in China. Programs undertaken in the 1990s ('Key technologies R&D Program,' '863 Program,' '973 Program,' 'Golden Projects') already promoted many of the technological areas that are listed in the MLP, notably biotechnology, telecommunications, and energy.

III(iii). On twins, computing c_i and assessing sensitivity to hidden bias

We searched for direct equivalents ('twins') at seven patent authorities for which we have reliable information about patent issuance: USPTO, EPO, JPO, the Canadian Intellectual Property Office (CIPO), the Korean Intellectual Property Office (KIPO), the Russian Federal Service for Intellectual Property (RFSIP), and the Taiwan Intellectual Property Office (TIPO). These seven offices account for the vast majority of total patenting activity outside China.[13] More specifically, we identify one-to-one equivalents: application B is a one-to-one equivalent of application A if B claims A as sole priority (i.e., no merged patent applications) and if A is only claimed by B in B's office (i.e., no split patent applications). In this sense, A and B cover the same technical content and are thus twin applications. We use these data to compute the invention fixed effects in equation (1) and the variable c_i in equation (3). The variable c_i is simply the average grant rate for these equivalent applications.

To assess the sensitivity of the results to hidden bias in c_i, we developed a test inspired by the bounding approach proposed by Aakvik (2001). Aakvik introduced the bounding analysis to evaluate the sensitivity of the results obtained through matching estimators to selection on unobservables. The main idea of this method is to ask how much hidden bias can be present in the selection process before the qualitative conclusions of the study begin to change. The method involves artificially adding increasing levels of hidden bias in the selection and observing when the treatment effect ceases to be significant. A study is highly sensitive to hidden bias if the main results change for small amounts of hidden bias.

We implement two versions of the test. First, we assume that the average grant outcome at other patent authorities is biased upward for home applicants. We then gradually decrease the bias by switching the grant outcome from 1 to 0 for a randomly selected share of twin applications granted by the home office of the foreign applicant, compute again the variable c_i, and re-run the analysis. We progressively change the grant outcome for 1 to 30 percent of the twin applications of the foreign filings at CNIPA that were granted by the home authorities.[14] Second, we also run the test by switching the outcome of granted twin applications in strategic sectors only—even if there is no theoretical ground for believing that the home-bias effect at *other* patent offices should occur only in sectors considered strategic by China.

IV. Data

IV(i). Data sources and sample

We combined data from six offline and online sources. The primary source of data was the EPO Worldwide Patent Statistical Database (PATSTAT, April 2015 edition). PATSTAT contains information on direct equivalents and the grant outcome at the seven selected offices (de Rassenfosse et al., 2014). It also contains most of the patent-level information used in the empirical analysis. Information on the grant outcome at CNIPA comes from the INPADOC legal status table, which is an add-on to PATSTAT. We crawled the Google Patent website to recover the number of independent and dependent claims at CNIPA, and the number of words per claim. We also crawled CNIPA's website to recover data on the IP law firm that handled the applications we analyzed.[15] Finally, we have obtained export data from the WTO to compute indicators of export specialization.[16] We will map these data to IPC codes using the concordance table provided by Lybbert and Zolas (2014).[17]

The sample is composed of applications filed at CNIPA by foreign and domestic firms between 2001 and 2009 and that have at least one unique direct equivalent in one of the following patent offices: CIPO, EPO, JPO, KIPO, RFSIP, TIPO, and USPTO. We expressly excluded utility models and design patents. This selection led to a final sample of 477,854 patent applications. The rationale for constraining our sample to applications filed between 2001 and 2009 was the following. First, including applications filed from 2001 onward ensures that the modifications introduced by the August 2000 amendment to the Chinese patent law to comply with the requirements of the TRIPS agreements were in place and understood by actors involved in the examination process. Second, excluding applications filed after 2009 was necessary in view of grant delays and the resulting data truncation at the time of data collection.

IV(ii). Dependent variable

The variable y_i (y_{io} in equation 1), labeled *Grant*, takes the value 1 if a patent was granted at CNIPA (or at office o in equation 1) and 0 if the application was refused or withdrawn. To mitigate further potential bias related to truncation, we exclude from the sample applications that were still pending at the time of data collection. Such filtering is particularly important because applications by foreigners have longer grant lags (Liegsalz and Wagner, 2013). We also exclude applications for which the applicant never requested an examination because, in such a case, the withdrawal decision was not affected by CNIPA's examination process. Thus, the remaining withdrawn applications in the sample are "quasi-refusals" in the sense of Lazaridis and van Pottelsberghe de la Potterie (2007)—that is, the applicant has withdrawn them after CNIPA has examined them.

IV(iii). Covariates

As far as the variables of interest are concerned, the binary variables S_{io} in equation (1) and S_i in equation (3) take the value 1 if a patent application belongs to any of the strategic IPC main groups recovered from the MLP, and the value 0 otherwise. The variable F_{io} in equation (1) reports whether the country of residence of the applicant recorded in the first priority filing is abroad or in country o. In equation (3), the variable F_i takes the value 1 for applications by Chinese applicants, and the value 0 otherwise.

If a patent application belongs to more than one applicant, we consider it foreign only if none of the applicants resides in China. (Section VI discusses alternative specifications.)

In equation (3), we include a set of control variables whose values may systematically differ between foreigners and locals, and that may correlate with the grant outcome. In addition to the variable c_i, which is the mean grant rate at other offices, the regression model includes the following control variables:

- Patent family size (*family_size*) accounts for the total size of the patent family to which an application belongs. Inventions covering large patent families are particularly valuable, which may affect the probability of grant. In computing the family size, we consider every patent authority for which the information is available in the PATSTAT database and not just the seven patent offices that we consulted in searching for direct equivalents.

- Number of IPC classes (*tot_IPC*) indicates the total number of four-digit IPC classes to which a patent application pertains. Applications covering many IPC classes are supposedly more complex to examine, as they may rely on technologically distinct elements (Lerner, 1994; Harhoff et al., 2003).

- Number of inventors (*nb_inv*) reports the total number of inventors listed in the patent document.

- Number of applicants (*nb_app*) reports the total number of applicants listed in the patent document.

- Revealed Technology Advantage (*RTA*) is a binary variable that reports whether the RTA at CNIPA is strictly larger than one. The RTA is computed as the country's share of patent applications filed at CNIPA within an IPC class (3-digit level of the IPC classification) over the total share of that country's applications at CNIPA. An RTA value above 1 in a specific IPC class indicates that a country is comparatively specialized in the technology sector covered by that IPC class.

- Export specialization (*export_spec*) is a binary variable that indicates whether the patent application belongs to an IPC class associated with a product category in which China had a revealed comparative advantage. We have used WTO data about Chinese and global exports to compute the revealed comparative advantage at the HS6 level and for each year of the period considered in our analysis.[18] We then mapped HS6 codes to IPC codes using the concordance table provided by Lybbert and Zolas (2014). Webster et al. (2014) found that discrimination focused on areas where the home country has a disproportionate share of exports or R&D expenditures, hence the need to control for the variables RTA and *export_spec*.

- Chinese prior art (*prior_art*) is the cumulative sum (up to the focal patent's filing year) of the number of single-child CNIPA patent documents in the IPC main group (6-digit) of the focal patent, starting in 1995. Single-child patents are filed in Chinese at CNIPA and were never extended to any other patent office—they constitute the most 'obscure' patented prior art. Chinese applicants may be more familiar with navigating this prior art than non-Chinese applicants, and it may be denser in strategic areas.

- Priority application lag (*priority_lag*) reports the lag in months between the filing date of the priority patent application and the filing date at CNIPA. The priority date is the closest in time to the actual invention date, and the lag makes it possible to control for the age of the invention when it reaches CNIPA. Logically, the lag is 0 if the Chinese application is the priority filing. The priority date fixes the relevant prior art against which the novelty of the CNIPA application will be assessed, such that the time of submission to CNIPA does not affect the probability of grant. However, applicants may be less keen to push for a grant for older inventions.

- Examination request lag (*exam_request_lag*) reports the lag in months between the date of application at CNIPA and the date of the request for examination. Chinese patent law requires the applicant to submit a request for substantive examination within three years of the filing date. As suggested by Palangkaraya et al. (2008), the applicant's decision to delay examination may correlate with the quality of the application and, therefore, with the probability of grant.

- Number of independent claims (*nb_indep_claims*) reports the number of independent claims listed in the patent application. Independent claims describe the essential features of the invention, and the variable captures the scope of the invention. This datum is missing for less than 1 percent of the patent applications in our sample. In such cases, we rely either on the average number of independent claims included in the equivalent applications filed with other patent authorities, or, if this information is not available, on the number of independent claims included in the granted document.[19]

- Number of dependent claims per independent claim (*dep_claims_ratio*) is the number of dependent claims over the number of independent claims appearing in the patent application. A dependent claim limits the scope of the independent claim to which it refers.

- Number of words per claim (*words_claim*) reports the total number of words per claim included in the patent application. A larger number of words per claim signals narrower claims.

- Experience of the applicant (*experience*) is a binary variable that takes the value 1 if the applicant is in the upper quartile in terms of number of applications filed at CNIPA, and 0 otherwise. It indicates whether the applicant has some level of familiarity with the Chinese patent system. The quartiles translate into at least three patent applications at CNIPA during the study period for foreign applicants and at least two applications for Chinese applicants.

- IP law firm (*law_firm*). China's patent law stipulates that a foreign applicant with no residence in China must appoint a licensed IP law firm to act as its agent to handle the patent application. Chinese applicants may instead appoint any IP law firm. The quality of the IP law firm may affect the probability of grant, especially if there are differences in the quality of attorneys between IP law firms chosen by foreigners and locals. The IP law firm effect for patent i by applicant f is computed as the average grant rate of all but applicant f's patent applications processed by the IP law firm. This implementation ensures that the variable is not endogenous to applicant f or the quality of invention i.[20]

- Dummy variables for the application year at CNIPA (*appln_year*) and the 1-digit level of the IPC class(es) of the application (*IPC_class*).[21]

IV(iv). Descriptive statistics

Table I displays the descriptive statistics by applicant country of residence for the 477,854 applications in the sample.

[Table I about here.]

As the bottom row of Table I shows, applications by Chinese firms represent 4.2 percent of the applications in the sample. This low number attests to the fact that the majority of applications by Chinese firms target the local market. Such filtering is very strict, but it increases comparability between applications by Chinese firms and applications by foreign firms. The table also shows that 73.6 percent of applications by Chinese firms were granted patent protection by CNIPA, against 70.8 percent for applications by foreign firms. Strikingly, strategic IPC subclasses cover about 34.5 percent of applications by Chinese firms and 20.8 percent of applications by foreign firms, even though these classes represent only 97 out of 6,812 total classes.

On average, foreign firms' applications belong to larger families and have more IPC sub-classes assigned to them, but they list a lower number of applicants and inventors. The time lag between the priority date and the application date is shorter for applications by Chinese firms than for foreign firms. Indeed, many applications by Chinese firms are priority filings. Interestingly, applications by Chinese firms are associated with shorter examination-request lags than filings by foreign firms, although the difference is only 1.5 months. On average, applications by foreign firms have the same number of independent claims as applications by Chinese firms (about three). However, the formers have 1.5 more dependent claims per independent claim and 5.3 fewer words per claim.

The 'patentability score' (variable c_i) is statistically significantly higher for foreign firms' applications than for Chinese firms, although the difference is small in magnitude (2.4 percentage points). The average grant rate of the IP law firm is not significantly different between foreign and domestic applications.

[Figure 1 about here.]

Figure I shows the distribution of all applications in the sample by 1-digit IPC code and applicant country of residence. The distribution across IPC codes is roughly similar between Chinese and foreign applicants, except for the H class (electricity), which is more prevalent in the case of Chinese applicants.

Figure II provides a finer overview of the raw data on grant rates. It depicts the proportion of granted patents for various subgroups as a function of the predicted probability of grant. Predictions come from a parsimonious model that regresses the variable grant on year and IPC class fixed effects. We can clearly see that strategic patent applications by Chinese applicants have a higher actual grant rate than non-strategic patent applications by Chinese applicants. They also have a higher grant rate than strategic patent applications by foreigners. However, trends in Figure II could be explained by a variety of factors and we refrain from giving them too much weight. The next section presents results from a series of econometric regression models.

[Figure 2 about here.]

V. Results

V(i). Fixed-effect approach

We start by implementing the fixed effect specification discussed in Section III(i).1. Tables II reports estimates of equation (1) and variations thereof. We use a set of dummy variables to capture office fixed effects (the reference office being the three offices not included). In column (1), we only control for the office fixed effects and the invention fixed effects. Patent offices associated with larger coefficients have a higher issuance rate and are, therefore, less stringent. In column (2), we introduce the variable F_{io}, which captures whether the applicant is foreign to the office. The associated coefficient suggests an overall home bias of 8.8 percentage points. This specification captures the average home bias across all offices and may hide office-specific heterogeneity.

In the next column, we interact the variable F with the office dummies to break down the home bias effect by office. Notice the *positive* coefficient at CNIPA, suggesting a bias in favor of foreigners. However, this result is an artifact of the way we have constructed the sample. Indeed, *all patent families in our sample have a member in China*, and the sample, therefore, misses all families that do not have a Chinese equivalent. As a result, when implementing a fixed effect specification, the coefficient associated with the home bias in China is not comparable with the home bias coefficient obtained for other countries.[22] Thus, our dataset is not perfectly suited to replicate the fixed-effect specification a la Webster et al. (2014). Nevertheless, we can impose restrictions to increase the comparability between applications at CNIPA and at other offices. When we restrict the sample to families with both a Chinese equivalent and a Western equivalent (USPTO or EPO) in column (4), which represents about 90 percent of families, the bias at CNIPA lowers drastically. In contrast, the bias at other offices remains largely unchanged—providing further evidence that sample construction affects the estimates. An additional estimate, in which we impose equivalents at the three largest offices (USPTO, EPO, and JPO), flips the sign of the coefficient: it leads to a home advantage of 7.4 percentage points at CNIPA and, again, leaves the other coefficients largely unchanged (not reported).

In column (5), we consider the possibility that strategic patent applications at CNIPA may have a different grant probability compared to non-strategic applications. We do so by including the interaction term $S\times$ CNIPA. We find that strategic technologies have a higher grant probability of about 2.2 percentage points.[23] Finally, in column (6), we estimate a triple interaction model (office, applicant origin, and strategic area). We find the presence of a home bias at CNIPA in the strategic areas. The coefficient associated with the variable $F \times \ S\times$ CNIPA suggests that locals are 4.8 percentage points more likely to have their patent applications granted compared to foreigners.[24]

[Table II about here.]

The next section reports the results of the control-variable approach. This approach allows us to account for a large number of possible reasons for the effect at CNIPA, such as a difference in patent attorney quality or in the amount of prior art available.

V(ii). Control-variable approach

Table III presents estimates of equation (3) and variations thereof. We log-transform the variables *family_size*, *tot_IPC*, *nb_indep_claims*, *dep_claims_ratio* and *word_claims* to account for their skewness.

[Table III about here.]

Odd-numbered columns in Table III report the coefficients obtained using the linear probability model, whereas even-numbered columns display the marginal effects at sample means obtained using the logit regression model. The first two columns display the results when the control variables are not included in the regressions. Columns (3) and (4) report the results for the full model, which includes the control variables, the mean grant rate, and the law firm effect. Columns (5) and (6) report the results for the regression model estimated on a matched sample of applications to increase further comparability between groups. We matched applications by Chinese firms to applications by foreign firms using the propensity score matching method (Rosenbaum and Rubin, 1985). Applications were paired based on the predicted probability of an applicant being from a foreign applicant. We computed this probability by estimating a probit regression model of the variable F on the relevant application-specific characteristics described in Section IV. Given the abundance of foreign applications in our original sample, we matched each Chinese application to up to two control foreign applications. Appendix A discusses the matching procedure in greater detail.

Results in columns (1) and (2) provide evidence of low levels of overall discrimination against foreigners (in the range 2.1–5.1 percentage points), and a greater level of discrimination in technology domains that the Chinese government deems strategic (in the range 5.7–6.7 percentage points). Controlling for additional confounding factors in columns (3) and (4) reduces overall discrimination to negligible levels (1.3–3.0 percentage points) but leaves the effect of discrimination in strategic areas essentially unchanged (6.0–6.4 percentage points). Note that the results are robust to the inclusion of interactions terms $RTA \times F$, $export\_spec \times F$, and $prior\_art \times F$ (not reported).[25] Estimating the full model on a matched sample of applications in columns (5) and (6) considerably reduces the sample size but confirms the finding of discrimination in strategic areas (3.6–4.1 percentage points).

Overall, the probability of grant for applications from foreigners in strategic areas is between 3.6 and 6.7 percentage points lower than what it should be in the absence of discrimination. To put this figure in perspective, consider the unconditional probability of grant for Chinese applications at CNIPA, which is 0.736 (Table I). The risk ratio of being refused a patent for foreigners in strategic areas compared to locals is thus about 1.5, implying a 50 percent higher rejection risk for foreigners.[26]

As discussed in Section III, the marginal effect for the interaction term in a non-linear model may have a different sign for different observations and different values of the covariates (Ai and Norton, 2003). The left side of Figure III displays the median spline plot of the interaction effect as a function of the predicted probability of grant for the full-sample regression (upper panel) and for the matched-sampled regression (lower panel). The effect is always negative and is the largest in magnitude for applications that have a lower predicted probability of being granted. Dividing the interaction effect by the probability of rejection leads to the right-hand graphs. The relative impact of the interaction term is larger for applications with a higher predicted probability of being granted.

[Figure 3 about here.]

V(iii). Time-specific estimates

Although we use the 2006 MLP to identify strategic technologies, we have explained that many of these technologies already appeared in previous long-term development plans. To the extent that the 2006 MLP reflects growing interest over these technologies, discrimination that favors them should be rising over the sample period. To test this hypothesis, we split the sample into two parts: applications filed before 2006 and applications filed after. Table IV presents estimates of equation (3) using both the linear probability and the logit models. It seems that discrimination is particularly acute after 2006, reaching 7.7–7.8 percentage points. The coefficient associated with the interaction term $F \times S$ is still negative in the pre-2006 period, but it is smaller in magnitude than in the baseline estimates (2.1–2.6 percentage points). These findings provide additional evidence that the MLP is associated with bias against foreigners.

[Table IV about here.]

V(iv). Technology-specific estimates

Table V provides evidence that the main effect is robust to the exclusion of any strategic area. Columns (1)–(8) report the results of a set of separate regressions for each of the eight strategic areas. These regressions include all the controls considered in the main analysis and two interaction terms. The first interaction term is a dummy variable capturing whether the application belongs to a specific strategic area (e.g., biotech in column 1) multiplied by the variable F. The second interaction term is a dummy variable capturing whether the application belongs to a strategic area other than the focal area (e.g., strategic but not biotech in column 1) multiplied by the variable F. The coefficients associated with the second interaction term is always negative and significant, suggesting that the main effect is not sensitive to the exclusion of any of the strategic areas. It ranges between 3.9 and 9.5 percentage points.

[Table V about here.]

In Table VI, we estimate the regression model used in columns (4) and (5) of Table III on separate subsamples composed only of patent applications for each of the eight strategic technology areas. Thus, we are adopting a split-sample approach, and our main variable of interest is simply the variable F. In contrast to Table V, the effect of the control variables explicitly relates to strategic patent applications. Discrimination is strongest in the fields of biotechnology (12–23 percentage points), followed by energy (11–12 percentage points) and ICT (7–10 percentage points). We do not find evidence of discrimination in the other strategic technology areas (the coefficients are not statistically significant). Thus, discrimination at CNIPA seems to concentrate on a set of particular technologies. However, notice that the small sample sizes could drive the lack of statistical significance for some of the coefficients.

[Table VI about here.]

VI. Robustness checks

VI(i). Sensitivity to hidden bias

The control-variable approach uses the mean grant outcome c_i of direct equivalent applications at different patent authorities as a benchmark for the grant probability of the invention at CNIPA. However, the variable c_i may lead to biased estimates of the coefficients associated with the variables of interest if foreigners face positive discrimination in their home office.

Table VII reports the coefficient associated with the interaction term $F \times S$ estimated using the linear probability model for different levels of hidden bias. As the table shows, when the bias is introduced for all foreign applications, the result is robust to a very high level of hidden bias. When we selectively introduce the bias exclusively for foreign twin applications in strategic sectors, the result appears to be robust to a large amount of hidden bias. The coefficient is still negative and significant for a level of hidden bias over 25 percent.

[Table VII about here.]

Finally, we also run the analysis using three alternative strategies to compute c_i. First, we compute it as a 'leave out mean' by discarding the grant outcome at the home patent authority (c_{xi}). Second, we compute it by taking into account the stringency of the pertinent patent authorities (c_{wi}), as explained in Section III(i).2. Third, we compute it by discarding information from the Taiwan Intellectual Property Office, which may align too closely with the outcome at CNIPA (c_{ni}).

Table VIII displays the coefficient retrieved from the linear probability model for the main variables of interest for the three modified versions of c_i. As the table shows, the effect of the interaction term is always negative and significantly different from 0 at the 0.001 probability threshold. It ranges between 4.3 and 6.8 percentage points, which is quantitatively similar to the baseline specification.

[Table VIII about here.]

VI(ii). Sensitivity to applicants' country of origin

The variable *Foreign* takes the value 1 when the country of residence of the applicant is abroad and the value 0 if the country of residence is China. Sometimes, however, large multinational corporations that have subsidiaries in several countries may decide to file a patent application from a local office and not from headquarters. As a result, patent applications by the same corporation could sometimes be categorized as Chinese and sometimes as foreign. In addition, the variable F takes the value 1 only if all applicants reside outside China. This strict definition of foreignness may represent a confounding factor if an application by Chinese and foreign co-applicants is treated as a foreign application.

To ensure that our results are robust to these issues, we run three different specifications. First, we assign applicants to a country only if at least 80 percent of their patent applications list that country as their residence. If the 80 percent threshold is not met, we exclude all the applications by that applicant from the sample. Second, we manually remove from the sample (i) all the applications filed by well known non-Chinese multinational corporations that are listed as Chinese applications and (ii) all non-Chinese applications filed by established Chinese multinationals.[27] Third, we exclude from the original sample all applications co-filed by a foreign and a Chinese applicant.

Table IX reports the coefficient for the main variables of interest recovered from the linear probability model run on the three subsamples. Column (1) displays the results for the sample obtained under the 80-percent rule; column (2) the results obtained by excluding known multinationals; and column (3) the results obtained by excluding Chinese-foreign co-filings. As the table shows, the main finding is robust to these alternative definitions of foreignness.

[Table IX about here.]

VI(iii). Additional considerations

We perform two additional exercises to shed more light on the phenomenon. First, one might suspect that the likelihood of discrimination is higher for foreign applications that have a competing local application. We used the Google Big Query patent data platform to identify a competing Chinese application filed in a 12-month time window around the focal patent and in the same IPC4 class as the focal patent. The data platform provides a similarity measure that comes from a model that has learned a set-of-words embedding of the patent full text to the technology classes (CPCs) of that patent using the WSABIE embedding algorithm (Weston et al., 2011).[28] We consider that an application is competing if the cosine similarity score with the focal patent is greater than 0.7 (or 0.8 in an additional test). We create two dummy variables: $compete\_CN$, which takes value 1 if there is a competing application filed by a Chinese applicant and 0 otherwise, and $compete\_FOR$, which takes value 1 if there is a competing application filed by a foreign applicant and 0 otherwise. We then re-run the same model as in the main analysis, first on the sample of applications that have a competing application filed by a Chinese applicant, and then on the sample of applications that have a competing application filed by a foreign applicant. Table X reports the results of the regressions on these two subsamples. As the table shows, the coefficients associated with the variable $F \times S$ are quite similar in magnitude and always negative and significant for both samples.

[Table X about here.]

Second, we consider another outcome measure, namely the change in scope for granted patents. The original outcome variable is quite sharp: either the patent gets granted or it does not. However, even if a patent application gets granted, it may be reduced in scope during the examination. The variable $Scope\_reduction$ is a dummy variable that takes value 1 if the patent granted by CNIPA has a lower number of independent claims than the patent application, and 0 otherwise. (Clearly, this variable is available only for the sample of granted patents.) Columns (1) to (3) of Table XI report the results of the standard model, including the full set of controls. Column (4), our preferred specification, reports the results for the sample of patent applications that initially had at least two independent claims. (There is no way to reduce the number of independent claims and still grant a patent if the original application had only one independent claim.) The result suggests a 6.5 percentage-point increase in the probability that scope reduction will occur for strategic patent applications by foreigners. Thus, not only are foreigners less likely to have their patent applications granted, but, when they do, they obtain narrower patents.

[Table XI about here.]

VII. Concluding remarks

The empirical analysis provides no clear evidence of a general violation of the national treatment principle at CNIPA. However, foreign applicants in strategic technology areas are significantly more likely than Chinese applicants to have their patent applications refused. When their applications are granted, they also obtain narrower patents than Chinese applicants.

The results presented in this paper seem to confirm the view that the patent system works as a barrier to entry in sectors that the Chinese government considers strategic. The analysis rules out many potential explanations for the effect, notably by controlling for differences in the baseline grant probability. We also took the conservative approach of controlling for the quality of IP law firm in order to rule out the possibility that results may be driven by foreign companies, e.g., relying on lower-quality patent attorneys or having a poor translation of their patent documents into Chinese. In fact, discussions with several heads of IP at western companies in these strategic sectors suggest that IP managers work with high-quality IP law firms. Thus, the reasons for the discrimination we observe are not found among the applicants or the patent attorneys. What remains is the patent office.

Discrimination against foreigners can be either intentional or unintentional. Intentional discrimination relates to disparate treatment of a specific group of applicants, whereas unintentional discrimination arises when policies, practices, and rules have disparate impacts on a specific group of applicants. This paper does not provide evidence of intentional discrimination since we do not observe what is happening inside CNIPA. For example, our results could be driven by the possibility that foreign patent applications in strategic areas receive a large number of third-party observations (i.e., prior art submitted by external parties). As far as we know, however, third-party observations are very rarely used at CNIPA. Our results could also be a consequence of the assignment of more-experienced examiners to strategic applications by foreigners, causing such applications to receive more exacting treatment.[29] Nevertheless, this explanation, if confirmed, would still imply a systematic difference in the treatment of locals and foreigners in apparent violation of the national treatment principle.

The empirical analysis conducted in this paper may be a starting point for future research. First, the use of twin patents provides a promising identification strategy that researchers could use to study other facets of the patent system. While this approach has limits, we hope that we have convinced the reader that carefully-thought sensitivity tests and robustness tests may alleviate the most pressing concerns. Pushed by a need to understand better the real effect of IP, clever identification strategies have been emerging recently (e.g., Sampat and Williams, 2015; Galasso and Schankerman, 2015; Kovács, 2017). The twin approach seems to us a useful addition to existing methods. Second, we need to understand better the consequences of the anti-foreign bias for foreign and Chinese firms. Standard theories in industrial organization and trade predict that IP protection helps R&D investments and sustains the production and commercialization of innovative products. Future research could investigate the extent to which discrimination helps local firms—and hurts foreign firms.

The results reported here have direct implications for the practice of IP. Foreign firms may not have sufficient insight or cases to realize that they are being discriminated against. The phenomenon uncovered in the empirical analysis suggests that firms in strategic areas should adapt their patenting strategy—for example, by selectively foregoing patenting in China or by filing more and narrower applications covering the same underlying invention.

Our finding also calls for action at the policy level. Most of the political efforts in international IP law as it relates to China has been geared toward harmonizing the legal framework and ensuring better enforcement of registered rights. Governments have indeed made considerable progress on these fronts, but subtler barriers may remain. The patent prosecution process may be one such barrier: Patent offices have vast discretionary power, and, where patents are concerned, policymakers do not verify the observance of the national treatment principle as they do for trade. If discrimination is unintentional, it is local policymakers' duty to identify the sources of such disparate impact and correct them. If discrimination turns out to be intentional, it is the duty of policymakers in other countries to report and condemn them. In any event, the formation of a WTO committee monitoring the international patent prosecution process may be warranted.

Appendix A. Matching procedure

As described in Section V, we have adopted the propensity score matching model to increase the comparability between the groups of foreign and Chinese patent applications (Rosenbaum and Rubin, 1985). The propensity score is the predicted probability that a given application has been filed by a foreign applicant. To compute that probability (our 'propensity score') we run a probit regression of the variable F on the set of observable application-specific characteristics described in Section IV.[30] Given the abundance of foreign applications in our sample, we match each Chinese filing with up to two foreign applications. To ensure that we do not introduce any additional bias by including more than a single control unit, we also set a tolerance threshold for the maximum distance, in terms of propensity score, between matched units.[31]

For the matching procedure to be successful, the empirical distribution of the relevant covariates should be balanced, and no significant differences in the covariates' means should remain after the pairing (Caliendo and Kopeinig, 2008). Table XII reports descriptive statistics by country of residence for the matched sample and the t-test for differences in the covariates' means between the two groups. As the table shows, there is no significant difference for the majority of covariates. Of particular importance is the fact that there is no significant difference between the Chinese and foreign applications regarding the $patentability\_score$ and the law_firm variable in the matched sample.

The difference is still significant for the *family_size*, the *nb_applicant*, the *RTA* and the *prior_art* variables, although the matching procedure has been able to increase drastically comparability between groups (cf. Table I). For instance, the difference in the average family size between foreign and domestic applications is 2.3 for the full sample and only 0.18 for the matched sample. The same holds for the difference in the average number of applicants, which goes from 0.12 in the full sample to 0.03 for the matched sample. The difference also shrinks for the *prior_art* and the RTA variables, but less remarkably. Note that we include the matching covariates as control variables in the main regression models.

[Table XII about here.]

Appendix B. Strategic IPC classes

Table XIII reports the list of strategic IPC classes (main group level) identified as described in Section III(ii).

[Table XIII about here.]

References

Aakvik, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford bulletin of economics and statistics*, 63(1):115–143.

Abrami, R. and Manty, T. (2010). Real blue? Viagra and intellectual property law in China. *Harvard Business School Case 9-910-409.*

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics letters*, 80(1):123–129.

Benzineb, K. and Guyot, J. (2011). Automated patent classification. In Lupu, M., Mayer, K., Tait, J., and Trippe, A. J., editors, *Current challenges in patent information retrieval*, pages 239–261. Springer.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

de Rassenfosse, G., Dernis, H., and Boedt, G. (2014). An introduction to the patstat database with example queries. *Australian Economic Review*, 47(3):395–408.

de Rassenfosse, G., Griffiths, W. E., Jaffe, A. B., and Webster, E. (2016). Low-quality patents in the eye of the beholder: evidence from multiple examiners. Technical report, National Bureau of Economic Research Working Paper 22244.

de Rassenfosse, G., Jensen, P. H., Julius, T., Palangkaraya, A., and Webster, E. (2019). Are foreigners treated equally under the trade-related aspects of intellectual property rights agreement? *The Journal of Law and Economics*, 62(4):663–685.

Galasso, A. and Schankerman, M. (2015). Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics*, 130(1):317–369.

Goldfarb, A. and Tucker, C. E. (2011). Privacy regulation and online advertising. *Management science*, 57(1):57–71.

Harhoff, D., Scherer, F. M., and Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, 28(4):331–343.

Kovács, B. (2017). Too hot to reject: The effect of weather variations on the patent examination process at the United States Patent and Trademark Office. *Research Policy*, 46(10):1824–1835.

Lazaridis, G. and van Pottelsberghe de la Potterie, B. (2007). The rigour of EPO's patentability criteria: An insight into the 'induced withdrawals'. *World Patent Information*, 29(4):317–326.

Lemley, M. A. and Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics*, 94(3):817–827.

Lerner, J. (1994). The importance of patent scope: an empirical analysis. *The RAND Journal of Economics*, 25(2):319–333.

Leuven, E. and Sianesi, B. (2015). Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components.*

Liegsalz, J. and Wagner, S. (2013). Patent examination at the state intellectual property office in China. *Research Policy*, 42(2):552–563.

Liu, F.-C., Simon, D. F., Sun, Y.-t., and Cao, C. (2011). China's innovation policies: Evolution, institutional structure, and trajectory. *Research Policy*, 40(7):917–931.

Love, B. J., Helmers, C., and Eberhardt, M. (2016). Patent litigation in China: Protecting rights or the local economy? *Vanderbilt Journal of Entertainment & Technology Law, Forthcoming.*

Lybbert, T. J. and Zolas, N. J. (2014). Getting patents and economic data to speak to each other: an 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, 43(3):530–542.

Maskus, K. E. (2000). Lessons from studying the international economics of intellectual property rights. *Vanderbit Law Review*, 53:2219.

McGregor, J. (2010). China's drive for 'indigenous innovation': A web of industrial policies. *US Chamber of Commerce.*

Norton, E. C., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, 4:154–167.

OECD (2008). *OECD Reviews of Innovation Policy - China*. Organisation for Economic Co-operation and Development, Paris.

Palangkaraya, A., Jensen, P. ., and Webster, E. (2008). Applicant behaviour in patent examination request lags. *Economics Letters*, 101(3):243 – 245.

Palangkaraya, A., Jensen, P., and Webster, E. (2017). The effect of patents on trade. *Journal of International Economics*, 105:1–9.

Plambeck, E. and Wang, Q. (2009). Effects of e-waste regulation on new product introduction. *Management Science*, 55(3):333–347.

Rosenbaum, P. R. (2002). *Observational studies*. Springer.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Sampat, B. and Williams, H. L. (2015). How do patents affect follow-on innovation? Evidence from the human genome. Technical report, National Bureau of Economic Research Working Paper 21666.

Scherer, F. M. (2004). A note on global welfare in pharmaceutical patenting. *The World Economy*, 27(7):1127–1142.

Scotchmer, S. (2004). The political economy of intellectual property treaties. *Journal of Law, Economics, and Organization*, 20(2):415–437.

SCPRC (2006). *Outline of the Long–Term National Plan for the Development of Science and Technology (2006–2020)*. State Council of the People's Republic of China (SCPRC), Beijing.

Shaffer, B. (1995). Firm-level responses to government regulation: Theoretical and research approaches. *Journal of Management*, 21(3):495–514.

Smith, Ken G, . and Grimm, C. M. (1987). Environmental variation, strategic change and firm performance: A study of railroad deregulation. *Strategic Management Journal*, 8(4):363–376.

Suliman, M. O. (1998). *China's transition to a socialist market economy.* Greenwood Publishing Group.

Tabakovic, H. and Wollmann, T. (2017). Effects of regulatory capture: Evidence from patent examiners. *Paper presented at the NBER Summer Institute.*

The Economist (2015). Letters to the editor. August 29th, 2015.

The IP Commission (2018). *Recommendations Regarding the Trump Administration's Section 301 Investigation.* The Commission on the Theft of American Intellectual Property, Washington, DC.

Tong, T. W., Zhang, K., He, Z.-L., and Zhang, Y. (2018). What determines the duration of patent examination in China? An outcome-specific duration analysis of invention patent applications at SIPO. *Research Policy*, 47(3):583–591.

USITC (2010). *China: Intellectual Property Infringement, Indigenous Innovation Policies, and Frameworks for Measuring the Effects on the US Economy.* USITC Publication 4199, Washington, DC.

USPTO (2010). *Report on Patent Enforcement in China*. U.S. Patent and Trademark Office, Arlington, VA.

Webster, E., Jensen, P. H., and Palangkaraya, A. (2014). Patent examination outcomes and the national treatment principle. *The RAND Journal of Economics*, 45(2):449–469.

Weston, J., Bengio, S., and Usunier, N. (2011). Wsabie: Scaling up to large vocabulary image annotation. In *Twenty-Second International Joint Conference on Artificial Intelligence*.

WIPO (2015). World intellectual property indicators. Technical report, World Intellectual Property Organization.

Yang, D. (2008). Pendency and grant ratios of invention patents: A comparative study of the United States and China. *Research Policy*, 37(6):1035–1046.

Yueh, L. (2009). Patent laws and innovation in China. *International Review of Law and Economics*, 29(4):304–313.

Notes

[1]For a recent example see "U.S. firm alleges China's government colluded with local competitor", Washington Post, September 13 2015.

[2]The MLP was implemented first through the Chinese Communist Party's 11$^{\text{th}}$ Five-Year Guideline (2006–2010) and then through the 12$^{\text{th}}$ Five-Year Guideline (2011–2015).

[3]See Yueh (2009) and Liegsalz and Wagner (2013) for in-depth analyses of the Chinese patent system and its evolution over time. The case study on the Pfizer Viagra in China provides useful background information on the political forces that led to a strengthening of IP rights in the country (Abrami and Manty, 2010).

[4]A 'patent family' refers to a group of patent applications that are all related to each other by way of one or several common priority filings. (A priority filing is the first patent application that was filed to protect an invention.)

[5]In an extension to the analysis, we also estimate the regression model on a matched sample of applications.

[6]For instance, the value of c_{wi} effect for an application with two granted international twins at the USPTO and at the EPO, instead of being equal to one ($= c_i$), will be equal to *(1/.68 + 1/.41)/2*. Where .68 and .41 are the average grant probabilities respectively at the USPTO and the EPO, computed based on all the twin applications in our sample at each of the offices.

[7]Note that the fixed-effect specification (equation 1) explicitly addresses this concern.

[8] Frontier technologies are selected in China per the following principles: (i) representing the development direction of world high-tech frontiers; (ii) having a pioneering role in shaping and developing new industries in the future; (iii) being conducive to industrial technology upgrading and to achieving leapfrogging in development; (iv) possessing a strong team of talented personnel and a sound R&D basis (SCPRC, 2006, p.33).

[9] For instance, *H04L 1/02* is a complete classification symbol. The section symbol *H* indicates that the patent application belongs to the Electricity section; the class symbol *H04* identifies the Electric communication technique class; the subclass symbol *H04L* specifies the Transmission of digital information field; the main-group level symbol *H04L 1* (formally *H04L 1/00*) narrows down the technological field of the application to Arrangements for detecting or preventing errors in the information received; the last two digits in the complete symbol further limit the domain to technologies detecting errors by diversity reception.

[10] Available at `https://www3.wipo.int/ipccat/`. IPCCAT's typical precision scores for English patents are about 90 percent at Class level, 85 percent at Sub-Class level

and 75 percent at Main Group level (Benzineb and Guyot, 2011).

[11] We are grateful to Gabriel Berlicki, Zhou Hao, and Lutz Mailaender for having agreed to help us.

[12]The list of strategic IPC main groups is available in Appendix B.

[13] Our own computation based on PATSTAT data suggests that these offices account for more than 80 percent of total patenting activity outside China.

[14] We run the test 30 times for every level of bias considered. We then report the empirical mean for β_3 over the 30 trials.

[15]See `https://patents.google.com/` and `http://english.CNIPA.gov.cn/`.

[16]The raw data are available at `https://timeseries.wto.org/`.

[17]The concordance table is available at `https://sites.google.com/site/nikolaszolas/PatentCrosswalk`.

[18]Harmonized System (HS) codes define product categories, and HS6 is the most detailed level available.

[19] In this way, we recover the number of independent claims for 91 percent of the applications for which the claim information was missing. In the remaining cases, we impute the number of independent claims for the missing observations through a Poisson regression on a set of relevant patent-specific characteristics, including IPC (3-digit), application year, number of applicants, number of inventors, the total number of IPC codes assigned to the patent application, and the country of residence of the first-listed applicant.

[20] We have also run the linear probability model with a fixed effect for the IP law firm. The results from this specification are in line with the baseline specification presented in Table III (available upon request from the authors).

[21]To ensure that the granularity of the IPC classification does not drive our results, we have also run the LPM with IPC classes measured at more fine-grained levels. The results are similar to the main specification reported in Table III (available upon request from the authors).

[22]For instance, European applicants seeking international protection tend to file first at the USPTO. They will file at CNIPA in rarer instances, and only for their most important inventions. Applications by foreigners at CNIPA are thus not directly comparable with applications by locals. This difference is reflected in the average family size, which reaches 3.02 for Chinese applicants and 5.33 for foreign applicants, see Table I. Note that this feature is not an issue in the control-variable regression: i) it stacks the odds against finding a bias against foreigners; and ii) one can easily control for the family size

in such regression.

[23]This finding does not depend on sample composition (not reported).

[24]Restricting the sample to families with both a Chinese equivalent and a Western equivalent leads to a 3.3 percentage point bias (not reported).

[25]Controlling for the interaction term $RTA \times F$ further decreases the coefficient associated with the interaction term $F \times S$ to -7.2 percentage points. Controlling for the interaction term $export\_spec \times F$ leaves the coefficient of interest largely unchanged. Controlling for the interaction term $prior\_art \times F$ decreases the coefficient of interest to -7.7 percentage points. Finally, controlling for all three interaction terms decreases the coefficient of interest to -8.1 percentage points. Results are available upon request from the authors.

[26]Obtained using results presented in column (4) of Table III. The risk of a foreign strategic patent application to be rejected = 1-0.736+0.030-0.098+0.06=0.256. The corresponding figure for locals is 0.166, leading to a risk ratio of 1.54.

[27]For instance, Taiwanese companies frequently file applications through their Chinese subsidiaries.

[28]More details on the similarity algorithm are available at `https://media.epo.org/play/gsgoogle2017`.

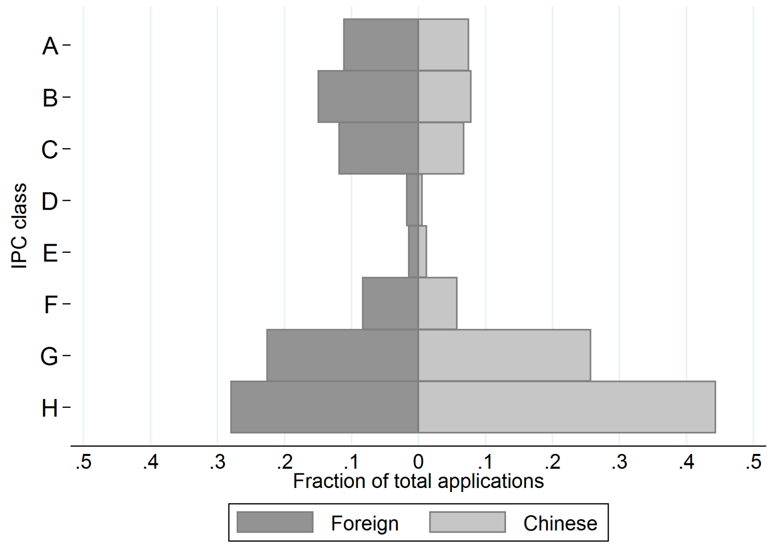
[29]Although this explanation would contradict U.S. evidence that experienced examiners are more lenient than less experienced ones (Lemley and Sampat, 2012). Note that we are not in a position to determine whether foreign applications are being unduly denied or, conversely, whether Chinese applications are being unduly granted.

[30] All else equal, for Chinese applications it is much more likely that the filing at CNIPA coincides with the first priority application. Therefore, we do not consider the variable *priority_lag* in the matching procedure, as the country of residence directly affects the probability of the application being filed at CNIPA first.

[31] To perform the matching procedure we use the Stata module PSMATCH2, developed by Leuven and Sianesi (2015). The caliper option that determines the maximum distance threshold is set to 0.25 of the standard deviation of the propensity score as recommended by Rosenbaum and Rubin (1985).

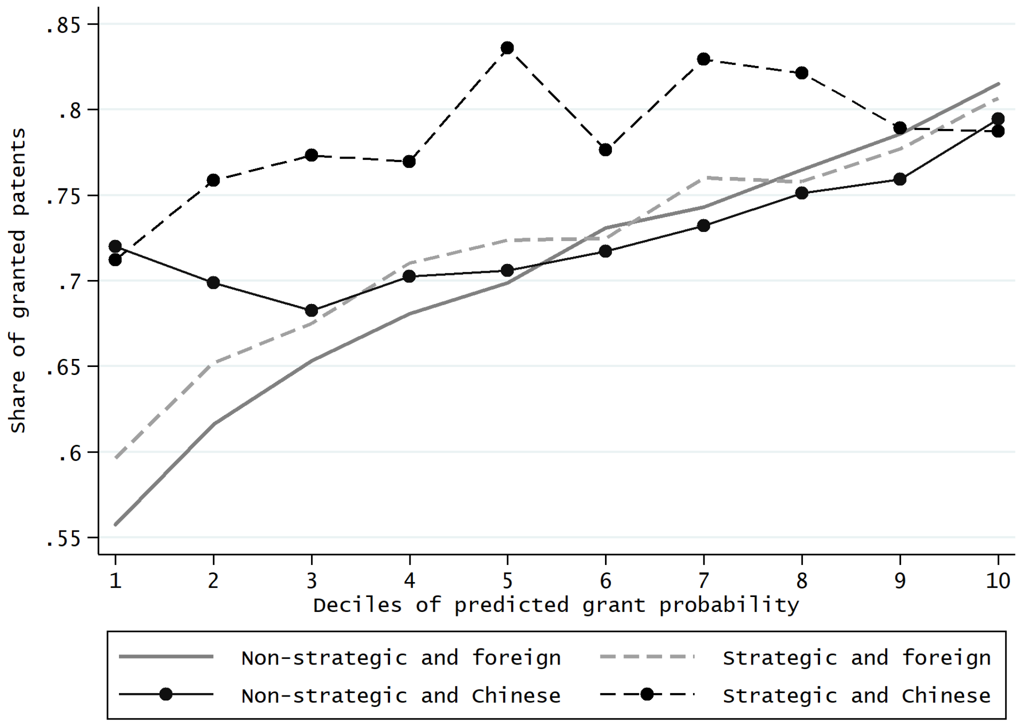
List of Figures

Figure I: Distribution of CNIPA applications by 1-digit IPC code and country of residence



IPC classes correspond to: A: Human Activities; B: Performing Operations; C: Chemistry/Metallurgy; D: Textiles/Paper; E: Fixed Constructions;F: Mechanical Engineering;G: Physics; and H: Electricity. The labels "Foreign" and "Chinese" indicate the country of residence of patent applicants.

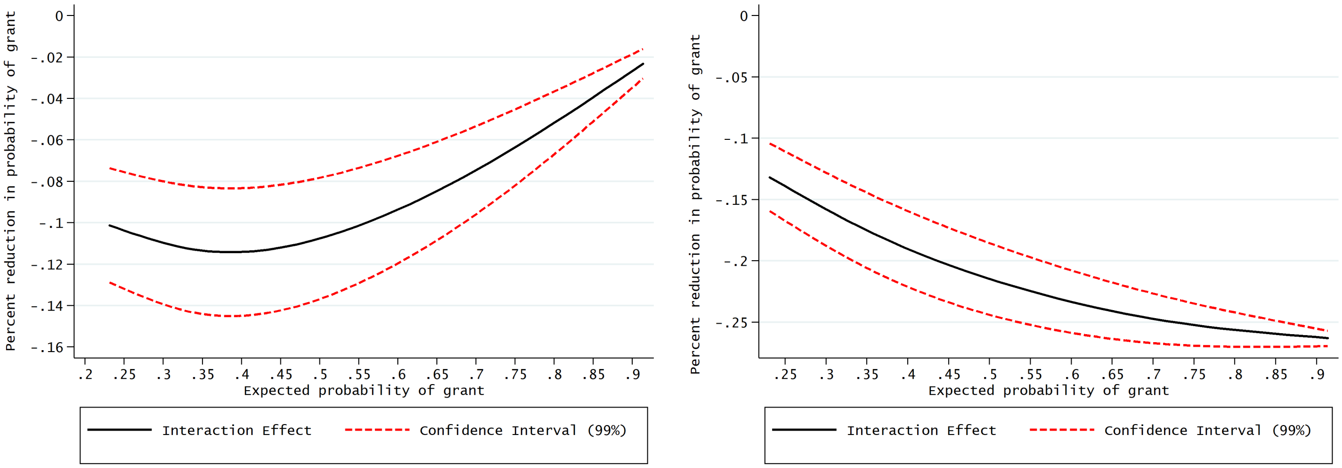
Figure II: Grant rate by subgroups



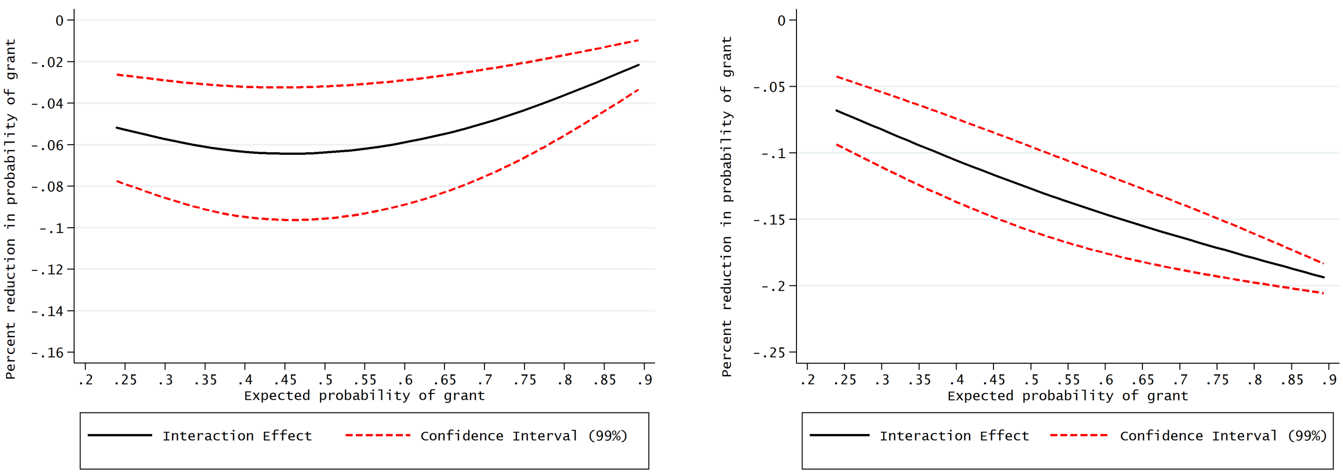
The predicted grant probability is obtained from a LPM regression with IPC class and year fixed effects. The labels "foreign" and "Chinese" indicate the country of residence of patent applicants.

Figure III: Average interaction effect by grant probability

Full Sample



Matched Sample



The left hand side graph displays the median spline plot of the interaction effect as a function of the predicted probability of grant. The right hand side graphs reports the interaction effect divided by the predicted probability of an application being rejected. The top graphs are produced using the results from the full sample (column 4 in table III). The bottom graphs are produced using the results from the matched sample (column 6 in table III).

List of Tables

Table I: Descriptive statistics by applicant origin

	Chinese applicants				Foreign applicants				t-test	
	min	mean	max	sd	min	mean	max	sd	Diff.	sd
grant	0.0	0.736	1		0.0	0.708	1		0.028	.003
S	0.0	0.345	1		0.0	0.208	1		0.137	.003
family_size	2.0	3.017	21	1.659	2.0	5.325	65	3.143	-2.308	.022
tot_IPC	1.0	2.479	15	1.312	1.0	2.667	21	1.531	-0.188	.011
nb_inv	0.0	2.436	20	1.861	0.0	2.223	33	1.638	0.213	.012
nb_app	0.0	1.145	7	0.370	1.0	1.029	13	0.200	0.116	.002
RTA	0.0	0.662	1		0.0	0.636	1		0.026	.004
export_spec	0.0	0.781	1		0.0	0.594	1		0.188	.004
prior_art	0.0	7.686	11	1.666	0.0	6.744	11	1.852	.942	.014
priority_lag	0.0	0.995	20	3.120	0.0	10.719	63	2.964	-9.725	.022
exam_request_lag	3.0	21.363	61	9.438	0.0	22.810	127	8.763	-1.447	.065
nb_indep_claims	1.0	3.129	67	3.626	1.0	3.057	557	3.972	0.0716	.029
dep_claims_ratio	0.0	5.273	71	4.420	0.0	6.792	288	6.254	-1.515	.045
words_claim	11.7	73.793	2497	57.219	11.0	68.533	22141	60.957	5.260	.448
experience	0.0	0.830	1		0.0	0.903	1		-0.073	.002
patentability_score	0.0	0.553	1	0.455	0.0	0.578	1	0.385	-0.024	.003
law firm	0.0	0.710	1	0.097	0.0	0.709	1	0.038	0.001	.000
N	19,119				458,735					
	(4.2%)				(95.8%)					

The column t-test reports the difference between the averages of the two groups and the standard error of that difference.

Table II: Estimates of the fixed effect model

Sample:	All	All	All	US or EP	All	All
	(1)	(2)	(3)	(4)	(5)	(6)
Office effects:						
$o =$ CNIPA	0.128	0.128	0.077	0.119	0.069	0.054
	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)
$o =$ EPO	-0.159	-0.180	-0.061	-0.058	-0.062	-0.62
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
$o =$ JPO	-0.068	-0.088	-0.030	-0.020	-0.030	-0.030
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
$o =$ KIPO	0.012	-0.011	0.193	0.202	0.193	0.193
	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
$o =$ USPTO	0.090	0.078	0.171	0.176	0.172	0.172
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign bias:						
F		-0.088				
		(0.001)				
$F\times$ CNIPA			0.056	0.018	0.059	0.075
			(0.004)	(0.004)	(0.004)	(0.004)
$F\times$ EPO			-0.133	-0.132	-0.132	-0.132
			(0.002)	(0.002)	(0.002)	(0.002)
$F\times$ JPO			-0.037	-0.046	-0.037	-0.037
			(0.002)	(0.002)	(0.002)	(0.002)
$F\times$ KIPO			-0.233	-0.242	-0.233	-0.233
			(0.003)	(0.003)	(0.003)	(0.003)
$F\times$ USPTO			-0.097	-0.098	-0.098	-0.098
			(0.002)	(0.002)	(0.002)	(0.002)
Effect in strategic areas:						
$S\times$ CNIPA					0.022	0.067
					(0.001)	(0.007)
$F \times S\times$ CNIPA						-0.048
						(0.007)
Constant	1.025	0.901	0.793	0.818	0.793	0.818
	(0.140)	(0.143)	(0.146)	(0.146)	(0.146)	(0.146)
N	1,827,957	1,827,957	1,827,957	1,682,877	1,827,957	1,827,957
N families	477,784	477,784	477,784	425,170	477,784	477,784
R^2	0.083	0.088	0.092	0.095	0.092	0.92

Econometric method is LPM.

Standard errors clustered at the family level in parentheses.

All regression models include invention fixed effects.

Table III: Estimates of the control-variable model

Estimator:	LPM (1)	Logit (2)	LPM (3)	Logit (4)	LPM (5)	Logit (6)
F	-0.021	-0.051	-0.013	-0.030	0.001	-0.009
	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)	(0.008)
S	0.080	0.093	0.080	0.098	0.048	0.060
	(0.007)	(0.008)	(0.006)	(0.009)	(0.006)	(0.008)
$F \times S$	-0.067	-0.057	-0.064	-0.060	-0.036	-0.041
	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
Control variables:						
patentability_score	0.568	0.568	0.547	0.545	0.336	0.333
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
log_fam_size			0.049	0.035	0.091	0.101
			(0.001)	(0.002)	(0.005)	(0.006)
log_tot_ipc			-0.004	-0.006	0.005	0.004
			(0.001)	(0.001)	(0.004)	(0.004)
nb_inventors			0.025	0.031	0.030	0.042
			(0.000)	(0.001)	(0.001)	(0.002)
nb_applicants			0.077	0.153	0.113	0.208
			(0.003)	(0.006)	(0.005)	(0.010)
RTA			-0.010	-0.008	0.000	-0.001
			(0.001)	(0.001)	(0.004)	(0.005)
export_spec			-0.019	-0.021	-0.029	-0.033
			(0.001)	(0.001)	(0.004)	(0.005)
prior_art			-0.002	-0.002	-0.003	-0.004
			(0.000)	(0.000)	(0.001)	(0.001)
priority_lag			-0.002	-0.003	-0.007	-0.007
			(0.000)	(0.000)	(0.001)	(0.001)
exam_request			-0.001	-0.002	-0.004	-0.004
			(0.000)	(0.000)	(0.000)	(0.000)
log_nb_indep_claims			-0.009	-0.010	0.030	0.033
			(0.001)	(0.001)	(0.003)	(0.004)
log_words_claim			0.020	0.024	0.080	0.095
			(0.001)	(0.002)	(0.004)	(0.005)
log_dep_claims_ratio			-0.000	0.000	0.029	0.033
			(0.001)	(0.001)	(0.002)	(0.002)
experience			0.047	0.045	0.055	0.043
			(0.002)	(0.002)	(0.006)	(0.006)
law firm			0.311	0.349	0.245	0.278
			(0.015)	(0.018)	(0.027)	(0.030)
Constant	0.425		-0.052		-0.176	
	(0.005)		(0.014)		(0.033)	
Fixed effects:						
Application Year	Yes	Yes	Yes	Yes	Yes	Yes
1-digit IPC	Yes	Yes	Yes	Yes	Yes	Yes
N	477,854	477,854	477,854	477,854	49,386	49,386
R^2	0.245	0.215	0.260	0.231	0.196	0.177

Robust standard errors in parentheses.
Use of a matched sample of similar patents in columns (5) and (6).
The coefficient of the interaction F x S is significant at the 0.001 probability threshold.
For the logit models the R^2 row reports the pseudo R^2.
The average marginal effect for the interaction term is computed using the method proposed by Ai and Norton (2003).

Table IV: Time-specific results

Estimator:	LPM		Logit	
	(1)	(2)	(3)	(4)
	Pre-2006	Post-2006	Pre-2006	Post-2006
F	-0.088	0.013	-0.127	0.007
	(0.007)	(0.005)	(0.010)	(0.007)
S	0.041	0.095	0.054	0.118
	(0.011)	(0.008)	(0.015)	(0.010)
$F \times S$	-0.026	-0.078	-0.021	-0.077
	(0.011)	(0.008)	(0.016)	(0.008)
Control variables:				
patentability_score	0.556	0.540	0.537	0.540
	(0.002)	(0.002)	(0.003)	(0.002)
log_fam_size	0.027	0.073	0.015	0.058
	(0.002)	(0.002)	(0.002)	(0.002)
log_tot_ipc	-0.002	-0.007	-0.002	-0.010
	(0.001)	(0.001)	(0.002)	(0.002)
nb_inventors	0.033	0.019	0.042	0.024
	(0.001)	(0.000)	(0.001)	(0.001)
nb_applicants	0.132	0.072	0.754	0.132
	(0.006)	(0.003)	(0.069)	(0.006)
RTA	-0.006	-0.015	-0.002	-0.013
	(0.002)	(0.002)	(0.002)	(0.002)
export_spec	-0.014	-0.022	-0.015	-0.025
	(0.002)	(0.002)	(0.002)	(0.002)
prior_art	-0.000	-0.003	0.000	-0.004
	(0.001)	(0.001)	(0.001)	(0.001)
priority_lag	-0.002	-0.003	-0.001	-0.004
	(0.000)	(0.000)	(0.000)	(0.000)
exam_request	-0.003	0.000	-0.003	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
log_nb_indep_claims	-0.015	-0.002	-0.016	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)
log_words_claim	0.012	0.030	0.016	0.035
	(0.002)	(0.002)	(0.002)	(0.002)
log_dep_claims_ratio	0.001	0.002	0.002	0.003
	(0.001)	(0.001)	(0.001)	(0.001)
experience	0.051	0.044	0.050	0.041
	(0.003)	(0.003)	(0.003)	(0.003)
law firm	0.341	0.287	0.388	0.287
	(0.028)	(0.018)	(0.033)	(0.018)
Constant	0.021	-0.088		
	(0.025)	(0.018)		
Fixed effects:				
Application Year	Yes	Yes	Yes	Yes
1-digit IPC	Yes	Yes	Yes	Yes
Observations	228,863	248,991	228,863	248,991
R^2	0.273	0.254	0.244	0.226

Robust standard errors in parentheses
For the logit models the R^2 row reports the pseudo R^2.
The average marginal effect for the interaction term is computed using the methodology proposed by Ai and Norton (2003).

Table V: Technology-specific results

	(1) Biotech	(2) Adv. Mtl	(3) ICT	(4) Adv. Mfg.	(5) Energy	(6) Marine	(7) Laser	(8) Aerospace
Estimator: LPM								
F	-0.032	-0.031	-0.030	-0.030	-0.030	-0.030	-0.030	-0.030
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
S_area	0.112	0.036	0.107	0.013	0.128	0.002	0.163	-0.108
	(0.025)	(0.015)	(0.007)	(0.018)	(0.032)	(0.021)	(0.084)	(0.277)
$F \times S\_area$	-0.153	0.010	-0.085	-0.002	-0.087	0.016	-0.137	0.182
	(0.025)	(0.015)	(0.010)	(0.019)	(0.033)	(0.022)	(0.084)	(0.279)
S_other	0.080	0.094	0.051	0.093	0.079	0.089	0.079	0.080
	(0.007)	(0.007)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$F \times S\_other$	-0.054	-0.095	-0.039	-0.076	-0.067	-0.074	-0.064	-0.065
	(0.007)	(0.007)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Estimator: Logit								
F	-0.032	-0.031	-0.030	-0.030	-0.030	-0.030	-0.030	-0.030
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
S_area	0.112	0.036	0.107	0.013	0.128	0.002	0.163	-0.108
	(0.025)	(0.015)	(0.009)	(0.018)	(0.032)	(0.021)	(0.084)	(0.277)
$F \times S\_area$	-0.123	0.010	-.0675	0.015	-0.063	0.017	-0.106	0.028
	(0.017)	(0.053)	(0.018)	(0.068)	(0.022)	(0.077)	(0.056)	(0.028)
S_other	0.099	0.118	0.069	0.115	0.097	0.111	0.098	0.099
	(0.009)	(0.009)	(0.015)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$F \times S\_other$	-0.069	-0.117	-0.055	-0.096	-0.083	-0.093	-0.081	-0.081
	(0.009)	(0.009)	(0.015)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
N (S_area)	13,538	38,592	51,695	17,382	11,801	13,248	2,082	288
N	477,854	477,854	477,854	477,854	477,854	477,854	477,854	477,854

All control variables included but not reported.

Estimates performed on the full sample of 477,854 observations.

N (S_area) reports the number of observations that fall in the considered strategic area.

Robust standard errors in parentheses.

Table VI: Technology specific results with split-sample approach

	(1) Biotech	(2) Adv. Mtl	(3) ICT	(4) Adv. Mfg.	(5) Energy	(6) Marine	(7) Laser	(8) Aerospace
Estimator: LPM								
F	-0.124	-0.013	-0.069	-0.031	-0.108	-0.023	-0.119	-0.279
	(0.026)	(0.013)	(0.009)	(0.017)	(0.029)	(0.021)	(0.062)	(0.209)
Estimator: Logit								
F	-0.230	-0.023	-0.101	-0.048	-0.121	-0.036	-0.175	-0.210
	(0.043)	(0.014)	(0.011)	(0.022)	(0.032)	(0.026)	(0.088)	(0.114)
Observations	13,538	38,592	51,695	17,382	11,801	13,248	2,082	288

All control variables included but not reported.

Robust standard errors in parentheses.

Table VII: Sensitivity to hidden bias

	Foreign		Foreign and Strategic	
% bias	Interaction effect	t-stat	Interaction effect	t-stat
1 %	-0.064	-9.65	-0.062	-9.41
5 %	-0.063	-9.62	-0.054	-8.23
10 %	-0.064	-9.68	-0.047	-7.06
15 %	-0.063	-9.65	-0.039	-5.92
20 %	-0.063	-9.69	-0.031	-4.73
25 %	-0.063	-9.68	-0.024	-3.58
30 %	-0.063	-9.60	-0.016	-2.45
N	477,853		477,853	

To recover the interaction effects $F \times S$ we run the test 30 times for every level of bias considered. We then report the empirical mean over the 30 trials.

Table VIII: Results recovered through different computations of c_i

	No_Home_Auth	Auth_Stringency	No_TW_Auth
F	0.007	-0.018	-0.012
	(0.004)	(0.004)	(0.004)
S	0.084	0.054	0.082
	(0.006)	(0.006)	(0.006)
$F \times S$	-0.066	-0.043	-0.068
	(0.007)	(0.007)	(0.007)
Patentability score:			
c_{xi}	0.503		
	(0.002)		
c_{wi}		0.295	
		(0.001)	
c_{ni}			0.533
			(0.002)
N	447,285	477,854	475,994
R^2	0.251	0.251	0.256

Column *No_Home_Auth* shows results for the patentability score (c_{xi}) computed by discarding the grant outcome at the home patent authority; Column *Auth_Stringency* shows results for the patentability score (c_{wi}) computed by taking into account the stringency of each patent authority; Column *No_TW_Auth* shows results for the patentability score (c_{ni}) computed by discarding the grant outcome at the Taiwanese patent authority.
Robust standard errors in parentheses.
All control variables included but not reported.

Table IX: Results recovered by excluding applicants of not certain origin

	(1)	(2)	(3)
F	-0.058	-0.054	-0.004
	(0.006)	(0.005)	(0.005)
S	0.077	0.077	0.075
	(0.008)	(0.007)	(0.007)
$F \times S$	-0.064	-0.062	-0.060
	(0.008)	(0.007)	(0.007)
N	430,407	473,252	475,218
R^2	0.266	0.262	0.262

Robust standard errors in parentheses.
All control variables included but not reported.

Table X: Results for applications with local and foreign competing inventions

Estimator:	LPM		Logit	
	(1)	(2)	(3)	(4)
Sample:	compete_CN	compete_FOR	compete_CN	compete_FOR
	70 percent similarity threshold			
F	-0.016	-0.018	-0.031	-0.035
	(0.013)	(0.006)	(0.017)	(0.007)
S	0.059	0.058	0.080	0.070
	(0.015)	(0.009)	(0.020)	(0.011)
$F \times S$	-0.043	-0.047	-0.048	-0.042
	(0.016)	(0.009)	(0.017)	(0.008)
Observations	20,009	371,530	20,009	371,530
R^2	0.229	0.263	0.199	0.232
	80 percent similarity threshold			
F	-0.014	-0.018	-0.027	-0.034
	(0.016)	(0.007)	(0.021)	(0.009)
S	0.070	0.067	0.095	0.083
	(0.017)	(0.010)	(0.023)	(0.013)
$F \times S$	-0.049	-0.052	-0.054	-0.049
	(0.018)	(0.010)	(0.019)	(0.013)
Observations	12,570	240,356	12,570	240,356
R^2	0.222	0.266	0.194	0.233

Robust standard errors in parentheses

Control variables included but not reported

The average marginal effect for the interaction term is computed using the methodology proposed by Ai and Norton (2003).

Table XI: Results for scope reduction

	(1)	(2)	(3)	(4)
F	0.041	0.026	-0.013	-0.017
	(0.005)	(0.006)	(0.004)	(0.007)
S	-0.012	-0.006	-0.043	-0.056
	(0.008)	(0.008)	(0.006)	(0.009)
$F \times S$	0.092	0.075	0.048	0.065
	(0.008)	(0.008)	(0.006)	(0.009)
Control variables:				
patentability_score	-0.072	-0.058	-0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.003)
log_fam_size		-0.013	0.022	0.035
		(0.002)	(0.001)	(0.002)
log_tot_ipc		0.020	-0.003	-0.006
		(0.001)	(0.001)	(0.002)
priority_lag		0.003	0.002	0.002
		(0.000)	(0.000)	(0.000)
exam_request		0.004	0.001	0.002
		(0.000)	(0.000)	(0.000)
log_nb_indep_claims			0.363	0.348
			(0.001)	(0.002)
log_words_claim			-0.042	-0.059
			(0.002)	(0.002)
log_dep_claims_ratio			-0.027	-0.034
			(0.001)	(0.001)
nb_applicants		0.007	0.011	0.016
		(0.003)	(0.003)	(0.004)
nb_inventors		0.007	0.002	0.002
		(0.000)	(0.000)	(0.001)
experience		-0.022	-0.020	-0.031
		(0.003)	(0.002)	(0.004)
RTA		-0.022	0.000	-0.005
		(0.002)	(0.001)	(0.002)
export_spec		0.025	0.007	0.017
		(0.002)	(0.001)	(0.002)
prior_art		0.008	0.002	0.003
		(0.001)	(0.000)	(0.001)
law firm		-0.023	0.111	0.190
		(0.019)	(0.014)	(0.024)
Constant	0.278	0.140	0.110	0.151
	(0.006)	(0.016)	(0.014)	(0.023)
Fixed effects:				
Application Year	Yes	Yes	Yes	Yes
1-digit IPC	Yes	Yes	Yes	Yes
Observations	335,430	335,430	335,430	215,912
R^2	0.026	0.034	0.383	0.223

Robust standard errors in parentheses.
The sample used in column (4) is limited to patent applications that had at least two independent claims at filing.

Table XII: Descriptive statistics by applicant origin for the matched sample

	Chinese applicants				Foreign applicants				t-test	
	min	mean	max	sd	min	mean	max	sd	Diff.	sd
S	0.0	0.345	1	0.475	0.0	0.339	1	0.473	0.005	0.004
patentability_score	0.0	0.553	1	0.455	0.0	0.558	1	0.437	-0.005	0.004
law firm	0.0	0.710	1	0.097	0.0	0.709	1	0.052	0.001	0.001
family_size	2.0	3.017	21	1.659	2.0	3.202	28	1.623	-0.185	0.015
tot_IPC	1.0	2.479	15	1.312	1.0	2.512	17	1.422	-0.032	0.013
exam_request	3.0	21.363	61	9.438	2.0	21.480	87	9.129	-0.117	0.085
nb_indep_claims	1.0	3.137	67	3.658	1.0	3.364	130	5.122	-0.227	0.042
dep_ind_ratio	0.0	5.273	71	4.420	0.0	5.482	88	5.180	-0.209	0.045
words_per_claim	11.7	73.793	2497	57.219	13.3	73.467	2637	52.659	0.325	0.503
nb_inventors	0.0	2.436	20	1.861	0.0	2.414	33	1.851	0.022	0.017
nb_applicants	0.0	1.145	7	0.370	1.0	1.114	9	0.407	0.031	0.003
experience	0.0	0.830	1	0.376	0.0	0.818	1	0.386	0.012	0.030
RTA	0.0	0.662	1	0.473	0.0	0.643	1	0.479	0.019	0.004
export_spec	0.0	0.781	1	0.413	0.0	0.682	1	0.466	.010	0.004
prior_art	0.0	7.686	11	1.666	0.0	7.294	11	1.750	0.391	0.015
App_Year Effects		Y				Y			-	
1-digit IPC Effects		Y				Y			-	
N	19119				30267					

The column t-test reports the difference between the averages of the two groups and the standard error of that difference.

Table XIII: List of strategic IPC main groups

Frontier Technologies (MLP)	IPCs (Main group)
Biotechnology	A61B5 A61K48 B01J37 C07K1 C07K14 C12N15 C12N5 C12N9 C12P17 C12P7 C12Q1
Information Technology	A63F13 G02B27 G06F13 G06F15 G06F17 G06F21 G06F3 G06T15 G06T19 G10L15 H04L12 H04L29 H04L9 H04Q3 H04Q9 H04W12 H04W4 H04W88
Advanced Materials	F24J2 G01N G05B13 G05B19 H01B12 H01F36 H01F6 H01G11 H01G9 H01L27 H01L31 H02J13
Advanced Manufacturing	B25J13 B25J9 B81C1 B81C3 B82B1 B82B3 B82Y10 B82Y40 G01M13 G05D1 G06F11 G06F19 G06N3
Advanced Energy	C01B3 C10G3 C10G45 C10L1 E21B43 F02C1 G21B1 G21B3 G21C1 G21C3 H01M10 H01M4 H01M8 H05H1
Marine Technology	B63B22 B63B3 B63B35 C09K5 C09K8 E21C45 E21C50 F03B1 G01H1 G01H11 G01H3 G01H5 G01H7 G01H9 G01S13 G01S15 G01S5 G01S7 G01V1 G01V3 G01V7 G01V9
Lasers Technology	H01S3 H01S5
Aerospace	B64B1 B64C1 B64G1 B64G3 B64G4