

CSAE Working Paper WPS/2011-15

IS THE DRAGON LEARNING TO FLY? AN ANALYSIS OF THE CHINESE PATENT EXPLOSION*

Markus EBERHARDT^a

Christian HELMERS^{a,b}

Zhihong YU^c

^a CSAE, University of Oxford

^b Universidad Carlos III de Madrid

^c University of Nottingham

August 2011

ABSTRACT

This paper analyses characteristics and determinants of the recent explosion of patent filings by Chinese firms both in China and the United States. We construct a firm-level dataset by matching USPTO and SIPO patents to Chinese manufacturing census data for the period 1999-2006. Using this integrated firm-level dataset, we show that the patent explosion is accounted for by a tiny, highly select group of Chinese companies in the information & communication technology (ICT) equipment industry. This handful of ICT companies accounts for nearly all Chinese USPTO patent filings as well as the vast majority of domestic SIPO patents despite there being a larger number of Chinese companies distributed across a wider range of industries that seeks patent protection domestically. Our empirical analysis further suggests that firms patenting in both US and China are considerably younger, larger and substantially more export-oriented than firms patenting exclusively in China. Our study contributes to the debate on China's innovative prowess and its potential to transition from an imitator to an innovator economy.

Keywords China, firm, patents.

JEL Classification L25, O12

*The paper was presented at the 'UC Berkeley – Peking University Workshop on the Evolution of Chinese Patenting and its Implications for Research and Innovation' in Beijing, and a workshop at the University of Oxford — we thank seminar/session participants for their useful comments and suggestions. The authors thank Lei Hou and Jiarui Zhang for excellent assistance in the classification of SIPO patents. Eberhardt gratefully acknowledges financial support from the UK Economic and Social Research Council [grant number PTA-026-27-2048].

This paper is dedicated to the memory of Mark Rogers and his important contribution to economic research on innovation and intellectual property.

Corresponding author: christian.helmerts@uc3m.es.

1 Introduction

China's economic success over the past three decades has been widely regarded as the result of its ability to produce manufactured goods at low cost, building on the availability of cheap labour and scale economies, while relying on existing (albeit in part advanced) technologies of production. China's ability to upgrade its technology-base and its moving up the value-chain has been widely regarded as hampered by weak (intellectual) property rights enforcement (Zhao, 2006). More recently, however, there has been increasing evidence to support the argument that China is catching up fast in terms of scientific and technological innovation.¹ The number of domestic invention patents filings with the Chinese patent office has increased at an average annual rate of 35% from around 15,600 to over 122,000 during the period 1999-2006.² This catching-up process is paired with strengthened statutory intellectual property rights protection (Park, 2008) and an increased interest by policymakers in the role of intellectual property in fuelling domestic innovation by increasing foreign technology transfer and providing domestic firms with incentives to invest in R&D. Accordingly, the recently formulated National Patent Development Strategy (2011-2020) envisions an increase in the number of annual patent applications, including invention and utility patents as well as designs, from about 1.2 million in 2010 to 2 million in 2015. The plan also foresees a doubling of the number of patent applications filed by Chinese applicants overseas in the same time-frame. These ambitious targets reflect a boisterous mood and overall positive outlook for Chinese innovation in parts of the academic literature (e.g. Fischer and von Zedtwitz, 2004).

At the same time, there is evidence to suggest that most of the innovation in China is of merely incremental nature and hence the corresponding patents protect 'small inventive steps' rather than substantive new technologies (Puga and Trefler, 2010). While such incremental innovation may still be valuable and in fact account in large part for China's growth success (Breznitz and Murphree, 2011),³ the concern is that the recent strong increase in domestic patent applications is produced overwhelmingly by low quality inventions which are driven by the incentives put in place by the Chinese government to encourage patenting *directly* rather than indirectly through the promotion of innovation (*The Economist*, 14th October 2010).⁴

Our analysis focuses on the recent 'explosion' in the number of patent applications by manufacturing firms registered in China with the domestic State Intellectual Property Office of China (SIPO) as well as the US Patent and Trademark Office (USPTO), and investigates the drivers of this development. Instead of relying on standard measures of patent quality, such as citations, we infer information on underlying inventions by assessing where companies seek patent protection: only domestically with SIPO or also with the USPTO. Not only are the direct and indirect costs associated with obtaining patent protection higher in the US, but inventions are required to overcome a higher novelty hurdle in the patent examination with the USPTO compared to SIPO during our sample period (see Section 3). These differences suggest that a comparison of patents taken out with the USPTO and SIPO reveals information on the underlying invention and the corresponding patentees.

¹According to OECD data, the number of full time equivalent R&D workers has nearly tripled between 1998 and 2008 reaching almost 2 million. Zhou and Leydesdorff (2006) point to China's rapid increase in terms of scientific publications. See also Section 2.1.

²WIPO Statistics Database, January 2011.

³Breznitz and Murphree (2011) argue that the successful *process and logistics innovation* (and not "the fabled creature of true [novel product] 'innovation'") of the last decades has equipped China with "ultra mass-flexibility production" capabilities and thus put it into a uniquely strong position within the global network that represents manufacturing production today.

⁴*The Economist* reports of tax rebates, research awards, and a link to the allocation of government contracts. See also *The New York Times*, 1st January 2011.

To conduct our analysis, we construct a novel firm-level dataset that combines patent data and company financials. We match both SIPO and USPTO patents filed between 1985 and 2006 to a subset of about 20,000 firms contained in China's Annual Survey of Industrial Enterprises (ASIE) compiled by the National Bureau of Statistics of China (NBS) for the period 1999-2006. The period covered by our sample represents perhaps the most interesting period in state innovation and intellectual property policy as well as enterprise innovation activity in China, combining aggressive opening up to FDI, policy commitments related to WTO-entry in 2001, strong increase in exporting, an amendment of the patent law, support for indigenous entrepreneurs, a swelling tide of Chinese returnees equipped with US PhDs and an accelerated pace of privatisation of or quasi-private spin-offs from government research institutions (Fischer and von Zedtwitz, 2004; Naughton, 2007; Hu and Mathews, 2008).

Our results show that a tiny number of private Chinese companies, concentrated in the electronics industry, accounts for the largest share of the dramatic increase in USPTO patents held by Chinese residents. This selection also concerns the underlying technologies which are mostly related to electronics and semiconductors. The highly select group of Chinese firms seeking patent protection in the US also accounts for the overwhelming share of domestic SIPO patents despite there being a relatively larger number of companies across a wider range of industries obtaining patent protection with SIPO.

Our objective is thus to analyse whether this small group of firms responsible for the dramatic increase in Chinese patenting represents the spearhead of a larger group of Chinese companies, poised to lead the Chinese economy to a wider technological take-off; or whether it merely reflects an exceptional, highly select group of firms, potentially supported by public policy, that is unlikely to represent a broader underlying technological leap among Chinese companies. Our analysis, therefore, informs on the broader debate over China's innovative prowess and potential development path, which may be based either on a 'Red Queen' run, which regards Chinese firms' ability to stay close to the world technology frontier and to improve upon and adapt *existing* innovation as key to its continued growth (Breznitz and Murphree, 2011), or the need for the domestic development of genuinely novel product innovation that pushes the global technology frontier to avoid getting caught in a 'middle income trap' (*The Economist*, 25th June 2011).

Analysing the patenting decision and patent productivity for a sub-sample of years we find that there is a significant difference in the firm characteristics associated with patenting an innovation with both SIPO and USPTO in comparison with merely patenting with SIPO. We find that firms which fall into the former category tend to be younger, considerably more export-oriented and larger than their peers which only patent in China. This result is evident in both the analysis of the decision to patent in any one year as well as in that of the patent count. We interpret our findings as evidence for the patent explosion being largely the result of increased patenting by very few companies in the information technology equipment industry. These relatively young companies managed to grow extremely fast in size within few years and succeeded placing themselves into leading positions in global production chains. Our analysis of their patented inventions indicates that these companies are undertaking a myriad of different innovative activities, including product innovation albeit of a relatively low-tech character. These companies also account for the lion's share in domestic patenting with SIPO, despite there being a much broader range of companies in a number of different industries patenting domestically. Our analysis, therefore, suggests that the patent explosion does not reflect a general technological take-off, but the success of an extremely small group of firms within a single industry.

The remainder of this paper is organized as follows. Section 2 briefly reviews the existing literature on government innovation policy and patenting in China and charts recent innovation effort. Section 3 discusses institutional differences of the patent systems in China and the US which have a bearing on firms' decisions to patent in these jurisdictions. Section 4 discusses the construction of our dataset. Section 5 explains our empirical strategy. Sections 6 and 7 discuss some descriptive evidence and our analytical results respectively. Section 8 offers some brief concluding thoughts.

2 Literature & Background

In this section, we cover innovation policy in China, provide some descriptives on broad measures of innovation effort over the past decade and discuss the existing empirical literature on patenting activity by Chinese firms, distinguishing aggregate economy, provincial-level and firm-level data. The latter has thus far been hampered by difficulties arising from matching Chinese firm-level data with official patent data, effectively limiting analysis to self-reported patent applications — a problem we overcome in the present study.

2.1 Innovation Policy

The analysis here is undertaken against the background of a pro-active innovation policy on behalf of the government of a developmental state. As is well-documented (Fischer and von Zedtwitz, 2004; Naughton, 2007; Hu and Jefferson, 2008), China's state-driven innovation policy began with a relatively high 'technology effort' (relatively high R&D to GDP ratio) during the socialist period, albeit one which supported 'episodic' prestige innovations with limited economic relevance (e.g. the development of nuclear weapons or the successful launch of a man-made satellite). Following the pro-market reforms of the late 1970s, it was not until the mid-1980s that a more sophisticated government innovation system, which allocated research funding based on competitive tender, was introduced (the so-called '86-3 program'). Among the stated aims of this policy was the intention "to achieve 'leap-frog' development in key high-tech fields" (MOST, 2011a). Around the same time research institutes were actively encouraged to collaborate with business enterprises and spin-off successful ventures into quasi-private entities — a process which created some of the foremost technology firms in China today (e.g. Founder, Tongfang, Legend). Subsequent technology and innovation policy expanded the focus to cover 'basic research' and reflects the recognition that successful innovation is necessarily built on a substantial talent base (so-called '97-3 program'; MOST, 2011b). The 97-3 program is at least in part responsible for the improvements in the human capital base, which saw the number of university graduates in engineering roughly double between 1999 and 2005, although these courses are not all comparable to undergraduate degrees in the US or Europe. In general terms the innovation *policy* described thus far focused primarily on state-owned entities, and in particular research institutes and universities — business enterprises (state-owned, foreign or private domestic) did not feature prominently (Naughton, 2007; Hu and Mathews, 2008).

As a striking reminder of the distance China's economic development has covered over a relatively short period of time, it was not until the late 1990s that any state support related to technology and innovation was extended to private firms and startups (Naughton, 2007: 361). Thus this period represents the first instance in China's recent economic history when state-owned and private enterprises enjoyed a level playing field in terms of state support for technology and innovation (e.g. tax breaks, subsidised credit). Crucially, the equal treatment of enterprises in innovation policy was ef-

fectively extended to foreign-invested enterprises operating in the Middle Kingdom around the same time (Naughton, 2007: 366). Private property rights for homeowners and businesses were not fully guaranteed by the law until October 2007, on the one hand forcing entrepreneurs into complex private-public or foreign joint ventures in order to conduct business legally while on the other fostering a concentration on short-run profits, to the detriment of time-consuming innovation efforts (Breznitz and Murphree, 2011). Even in the present day the laws surrounding investment are said to be murky and the *The Economist* magazine concludes that “[o]wnership is rarely straightforward in China” (9th July, 2011).

An explicit desire among the central government officials to transform the country from an imitator to an innovator economy was stated in the 2006-2010 5-year plan, which targeted 2015 as the year in which to achieve this goal (Hu and Jefferson, 2008: 288). As a result local governments, which jostle with each other for increased domestic and foreign investment, are argued to act under the impression that “virtually any action they take in support of high-technology industry and technological development will be acceptable” (Naughton, 2007: 368).

2.2 Innovation Effort

Simple aggregates for R&D expenditure by sector indicate how the late 1990s became a game-changer for Chinese innovation effort: in 1997, 42.7% of the national R&D expenditure was spent by business enterprises, in subsequent years this share rose to 44.7% (1998), 49.7% (1999) and 60.2% (2000), reaching 69.1% by 2006 (MOST, *various*).

In Figure 1 we provide some macro evidence for the rapid changes apparent in China’s innovation effort over our sample period. Apart from China, we highlight the US, Japan, Korea as representative OECD countries which China aims to emulate in terms of innovation activity. The top graph for 1999 indicates that among 71 developing and developed countries China’s R&D intensity (around 0.8%) was relatively unremarkable, somewhat above countries at a similar level of development but certainly not an outlier.⁵ The graph for 2007 (for 74 countries) indicates that a mere 8 years later the picture has changed quite substantially, with R&D intensity at 1.5% equivalent to that in countries twice her average per capita income.

A second shift in China’s innovation activity can be seen in Table 1, where we report the above-mentioned share of national R&D expenditures accounted for by business enterprises, again for the years 1999 and 2007. As a means of comparison we provide data for a number of OECD countries as well as emerging economies, but (this is particularly important with regard to the ‘ranking’) the choice of countries covered is not meant to be representative of all OECD and/or emerging economies. In 1999, around half the R&D expenditure in China was accounted for by business enterprises, putting them toward the bottom of this group of 11 countries, alongside with Turkey, Argentina and Mexico. By 2007 the business share had risen to over 70%, thus placing China firmly in line with top OECD innovators such as Japan, Korea, the US or Germany. Note that with the exception of Mexico the figures for other emerging economies remained relatively stable over this time period.

In conclusion, whether or not this effort has resulted in a boost to domestic innovation (of the novel product type) the above numbers plot an impressive sea-change in China’s innovation activity between the late 1990s and the present day.

⁵We stress again that China’s development was unusual in that the R&D to GDP ratio was substantially higher during the socialist period ending in the late 1970s. Although a relatively high level of innovative effort was maintained during the early 1980s the R&D to GDP ratio eventually came down to ‘standard’ developing country levels around the mid-1990s.

2.3 Patenting

The existing literature on patenting in China is surprisingly sparse. While there are some studies at the aggregate (Hu and Mathews, 2008), industry (Hu, 2010) and province level (Sun, 2000; Cheung and Lin, 2004; Yueh, 2009), there is little work at the firm-level (Hu and Jefferson, 2009). Moreover, the existing work at the firm-level relies on firms' self-reported patents due to a lack of actual patent information (Hu and Jefferson, 2009). The existing literature, so far, has focused mostly on the effect of technology transfer on the performance of Chinese firms (Hu et al., 2005) as well as the patenting activity of foreign relative to domestic firms in China (Hu, 2010; Liang and Xue, 2010). The scarcity of empirical evidence on Chinese firms' patenting activities at the firm-level stands in stark contrast to a vast literature on trade and FDI at the firm-level in China (e.g. Hu and Jefferson, 2002; Chuang, Y.-C. and Hsu, 2004; Girma et al., 2009; Hale and Long, 2011).

At the *aggregate economy-level* Hu and Mathews (2008) split the total number of USPTO patents by Chinese residents into different groups of patentees to show that the most dramatic surge in patenting with the USPTO occurred among private domestic companies since 2001 when China joined the WTO. Hu and Mathews (2008) also point to the significant contribution of public research institutes in increasing the number of USPTO patents held by Chinese assignees. The authors provide some evidence that USPTO patents held by private firms receive more forward citations and have more backward cites than those held by other groups, such as state-owned companies. The larger number of forward citations is interpreted as an indicator of a higher value whereas the higher number of backward cites is interpreted to indicate private companies' superior ability to build on existing knowledge.

Hu (2010) analyzes patent data at the *industry-level* to investigate the strong increase in patenting by foreign firms in China, which amounted to over 30% per year between 1995-2004. Since Hu does not have patents matched to firm-level data, he relies on a concordance table produced by the OECD to allocate patents across industries. Using this concordance table, he finds ISIC 29 (machinery), 24 (chemicals), and 32 (telecommunications equipment) to have by far the largest number of patents in China. Patenting by foreign residents in China is found to be driven by import competition, that is, foreign patenting is positively correlated with the amount imported by China from technologically similar industries of other foreign countries. This finding is corroborated by the fact that over 90% of patent applications by foreign firms claim a foreign priority.

An early study on the spatial concentration of patenting by Sun (2000) investigated *province-level* data for 1985-1995 — the period before FDI or private enterprise made substantial impact on Chinese development in general and with regards to innovation activities in particular. It is thus not surprising that the output from collective and state-owned enterprises was found to be an important force in innovation output and overall the relative level of provincial development seeming to drive the empirical results.⁶ All R&D activity measured for this study was limited to universities and 'government R&D institutes', again a reflection of the scarcity of business enterprise R&D at the time. Employing provincial data from 1995 to 2000, Cheung and Lin (2004) find that by the late 1990s the presence of FDI can be found to have a robust positive impact on patenting by 'domestic innovators', with R&D inputs also positive and significant but export volumes insignificant.⁷ Yueh (2009) explores the determinants of aggregate patent counts across 29 Chinese provinces for the period 1991-2003 uncovering substantial differences in patenting activity across provinces, which is

⁶This is reinforced by the finding that export volumes seem to be a good predictor of patenting activity.

⁷At least parts of these stark differences between Sun (2000) and Cheung and Lin (2004) are down to the different empirical methods employed.

not surprising given the heterogeneity in economic progress between the Eastern Seaboard and the central provinces.

Finally, Hu and Jefferson (2009) conduct their analysis at the *firm-level* exploiting a large dataset assembled by the NBS and focusing on domestic patenting with SIPO during 1995-2001. The main limitation of the analysis is that only firms' self-reported patent counts are available (see Section 4 for more details). Moreover, firms only report an aggregate patent count, not distinguishing between innovation, utility and design patents. Since only innovation patents require an examination by SIPO, utility and design patents are likely to protect innovations characterized by a lower inventive step and thus embody little technological progress. Hu and Jefferson (2009) explain the recent increase in firms' patenting activity in China by the presence of FDI, the change in the patent law and the anticipation of China's WTO accession in 2001. The authors also find the patents-R&D elasticity to be higher for domestic than for foreign-owned companies, which they explain by suggesting that foreign firms conduct R&D in China primarily to adapt existing products and patent existing inventions. In addition, Hu and Jefferson (2009) find the amendment of the patent law in 2000, which became effective only in 2001, to have had a major impact on patenting; yet, given that their sample extends only up to 2001, it remains unclear to which degree the year-dummy variable used to capture the reform reflects firms' (lagged) response to the policy change.⁸

3 USPTO vs SIPO

This section examines differences between the patent systems in the US and China which may have implications for the ability and motivation of Chinese firms to seek patent protection in each country. Since our analysis focuses on invention (SIPO) and utility (USPTO) patents, our discussion here is limited to this type of patents.⁹ Sorell (2002), Yang (2008), and Lin and Connor (2008) offer detailed comparisons of the two systems where Yang (2008) particularly focuses on potential differences in the way domestic and foreign applicants are treated in both countries.

China's first patent law came into force in 1985 and was since amended three times (in 1992, 2001, and 2009). Given that only the second amendment falls into the time period covered by our regression analysis, we omit a discussion of the other two. The second comprehensive amendment of the patent law, adopted on 25 August 2000 and effective from 1 July 2001, was necessary to bring China's patent law in line with the WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which China adopted with WTO entry in November 2001 (Yu, 2001). For our purposes, an important change brought about by the amendment regards the deletion of the requirement to obtain official (government) permission before a Chinese resident is allowed to file for patent protection abroad. Another important change was equal treatment of state- and privately-owned companies as well as the introduction of preliminary injunctions in case of infringement.

Regarding the application process, a fundamental difference is the first-to-file rule applied by SIPO compared to the first-to-invent rule at the USPTO.¹⁰ In addition, the USPTO has a one-year grace period to determine prior art, i.e., an invention may still be patentable despite having been

⁸More generally, their analysis of innovation activity in private enterprises is somewhat hampered by the time period under analysis, 1995-2001: it is argued elsewhere that many of the incentives for technology development and innovation were not extended to private enterprises and start-ups until 1999 (Naughton, 2007: 361), such that their analysis only captures one or two years of the 'level playing field' between state-owned and private enterprises.

⁹USPTO 'utility' patents correspond to SIPO invention patents and must not be confused with SIPO utility patents. A SIPO utility patent is not subject to substantive examination.

¹⁰In the US, whoever demonstrates to have invented the invention first is entitled to the patent whereas in China, the individual that files the patent application first is entitled to the patent.

published for example in a trade journal during the 12 months preceding the patent application, while there is no such provision in China. Moreover, until recently, SIPO granted patents for inventions that were not necessarily ‘new-to-the-world’: before the third amendment to the Chinese Patent Law in 2009, Article 22.2 defined prior art excluding inventions known to the public or in public use outside of China. For example, while a patent publication in the US did represent prior art preventing the granting of a patent in China, if in contrast the invention had been known or used by someone other than the inventor (without obligation of secrecy) in the US, it would still have been patentable in China.

Yang (2008) points out that in the examination process, different emphasis is put on the ‘industrial applicability’ criterion. That is, whereas the USPTO has a broad interpretation of the potential practical purpose an invention might serve, SIPO requires some form of demonstrable industrial applicability. This is related to a broader issue regarding patentable subject matter. The US patent system has been criticized for allowing a broad range of inventions to be patented that may contain only limited technological advance, including software, business models and even DNA segments (Gallini, 2002; van Pottelsberghe, 2010). SIPO in contrast, applies a narrower definition of what constitutes patentable subject matter, more in line with the stance of the European Patent Office (EPO). There are also some differences in the legal provisions in case an application is rejected and the applicant wishes to request re-examination, but these appear to be of less relevance to our analysis and are therefore not discussed here. Similarly, there are considerable differences in patent enforcement and validity disputes between the US and China, which are not discussed here (for more details see Sorell, 2002). Of more direct relevance is the fact that the US does not require foreign applicants to have a patent agent in the US, whereas SIPO requires foreign applicants without residence in China to rely on a domestic patent agent designated by SIPO to handle the application process. Foreign applicants for patents in the US may nevertheless be compelled to rely on the services of a US patent attorney because there usually is a considerable amount of exchange of information between an applicant (or his patent attorney) and the USPTO during the examination process.

From an empirical point of view, Yang (2008) finds no difference in pendency between domestic and foreign applicants in the US and grant times to be shorter in China than the US (on average two as opposed to three years). In addition, she finds the ratio of grants to applications to be higher in the US than China, while the grant rate is considerably higher for domestic applicants in both China and the US.

The fee structure also differs between the USPTO and SIPO. In China, an application costs CNY 900 (The exchange rate between 2002 and 2005 is around US\$ 1=8.27 CNY, so this amounts to around US\$ 110), there is an additional examination fee of CNY 2,500 (US\$ 300) and maintenance fees of CNY 300 (US\$ 35) every five years. At the USPTO the basic application fee is US\$ 330 and examination fees amount to US\$ 220.¹¹ At the USPTO, renewal fees are not payable annually: at 3.5 years, the maintenance fees due amount to US\$ 980, at 7.5 years to US\$ 2,480 and at 11.5 years to US\$ 4,110.¹² Additional costs for Chinese firms arise from the need to translate the patent application into English. If a Chinese applicant relies on the services of a US patent attorney, although not formally required by the USPTO, substantial additional costs arise. Hence, the numbers suggest that obtaining and maintaining patent protection in the US is *considerably* more expensive in the US than China. However, the largest difference in the cost of obtaining patent protection in

¹¹Fees for small firms are half the normal fee. See <http://tinyurl.com/66xb774> (USPTO website).

¹²Small entities pay half the standard fee.

the US for Chinese firms stems from the need to translate the patent application and to potentially hire a US patent attorney who administers the patent application.

4 Data

A number of firm-level datasets have been used in the existing literature on innovation, R&D, and patents in China. Lin et al. (2010) use a cross-section of firms contained in the 2003 World Bank business environment and enterprise performance survey. A more comprehensive dataset for patenting analysis has been assembled by Jefferson et al. (2003) on large and medium sized enterprises for the period 1995-1999 from an annual survey conducted by the NBS and used in Hu et al. (2005) and Hu and Jefferson (2009). Hu et al. (2005) use a subset of about 10,000 companies from this dataset and Hu and Jefferson (2009) extend its coverage to the period 1995-2001, which yields about 22,000 firms.

With regard to patents, to the best of our knowledge, we are not aware of any existing large-scale dataset that contains actual patent holdings at the firm-level. Other studies on patenting in China have, for example, relied on industry-level patenting by employing a correspondence table that maps patents' sectoral classification (IPC) into firms' standard industrial classification (Hu, 2010).¹³ Hu and Jefferson (2009) use firm-level data, but patents are self-reported by firms and no distinction can be made between invention, utility and design patents. It is difficult to gauge how reliable firms' self-reported patent counts are, but experience with similar data suggests very substantial, potentially non-random mis-reporting.¹⁴

The data used for our analysis consists of three components. The first component contains our firm-level information, the second component consists of USPTO and SIPO patent data and the third component is a 'bridge' that links the firm-level data with the patent information. These three components are discussed in turn below.

4.1 Firm-level Data

Our firm-level data come from China's Annual Survey of Industrial Enterprises (ASIE) compiled by the NBS. Firms contained in ASIE include the whole population of state-owned firms as well as all non-state-owned companies with annual sales above CNY 5 million (around \$600,000). On average, more than 200,000 firms are included each year and they account for around 95% of total Chinese industrial output and 98% of industrial exports, covering 39 two-digit industries, of which 30 belong to manufacturing industries, spread across all 31 provinces and municipalities.

The available data cover the period 1999-2006, including more than 2 million observations from about 590,000 firms.¹⁵ The key variables relevant to our study include a unique firm identifier, R&D expenditure (limited coverage), exports, type of ownership registration, the structure of paid-in capital distinguished by investor types (government; foreign investors excluding Hong Kong,

¹³The International Patent Classification (IPC), established by the Strasbourg Agreement 1971, provides for a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain (WIPO, 2011).

¹⁴For example the UK Community Innovation Survey 3 (2001) contains the following question 'How many patents, if any, did your enterprise apply for during the period 1998 to 2000?' The question refers to a three-year range, which makes it difficult to allocate patents into specific years to match annual firm-level data. In any case, cross-checking firms' responses to this question with their actual patent holdings indicates that only about 30% of firms that report to have applied for a patent actually did so.

¹⁵Other previous studies using different versions of this dataset include Brandt et al. (2011), Girma et al. (2008), Girma et al. (2009), Hsieh and Klenow (2009), Lu et al. (2009), and Yu (2011).

Macao, and Taiwan; investors from Hong Kong, Macao, and Taiwan; legal entities; individuals; and collective investors), output, sales, employment, fixed assets, and industry affiliation.¹⁶

4.2 Patent Data

The patent data come from the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT) version October 2010.¹⁷ We extract patents filed by Chinese residents at the US Patent and Trademark Office (USPTO) and those published directly by the State Intellectual Property Office of China (SIPO).¹⁸ Our analysis focuses on the application date of a patent. However, patent data are only visible *after* a patent has been published which implies that although we use the application date, our sample of patents is limited to patents that have already been published. Given the usual 18-month delay between application and publication date at both USPTO and SIPO,¹⁹ this implies that we have patent data up to March 2009 at best. This is not restricting our analysis given that our firm-level data is only available to 2006.²⁰

4.3 Matching

Due to the absence of a unique identifier shared by the firm-level and patent data, the main problem in constructing our dataset consists in matching patents to firms. This is generally challenging for a number of reasons outlined in detail in Helmers et al. (2011). However, in the case of Chinese firms and patents, it is even more difficult because of the different ways in which Chinese firm names can be recorded: (a) using Chinese characters, (b) using pinyin transcription into the Latin alphabet, (c) a translation of the Chinese names into English, and (d) any mix of (a)-(c).

The Chinese census data contain only (a) firm names using Chinese characters whereas PATSTAT contains (b), (c) and (d). In principle, this means that in order to match patents to firms, we would have to either transcribe firms' names contained in the ASIE census or the assignee names contained in PATSTAT. We have opted for an alternative solution: the Oriana database provided by Bureau Van Dijk Electronic Publishing (BvDEP) offers a firm-level dataset reporting balance sheet and profit & loss information for individual companies in countries in the Asia-Pacific region.²¹ The Oriana version available to us contains firm-level data for about 23,000 Chinese firms for the period 2000-2005. The advantage of using Oriana is that it reports firm names using the Latin alphabet as well as firms' unique identification number. This allows us, on the one hand, to link Oriana to the census through the unique identifier and on the other, to use Oriana firm names to match with assignee names contained in the patent documents. Note that we only use Oriana to bridge the census and patent datasets. Oriana's coverage of firm-level variables is far less comprehensive than the ASIE

¹⁶All variables are available for all years except R&D expenditure, which is only reported for the years 2001, 2002, 2005, and 2006.

¹⁷PATSTAT combines patent information from several sources: DocDB (the EPO master bibliographic database containing abstracts and citations), PRS (the patent register for legal data), EPASYS (the database for EP patent grant procedure data), and the EPO patent register as well as the USPTO patent database for names and addresses of applicants and inventors.

¹⁸This includes patents filed with the World Intellectual Property Organization (WIPO) through the PCT route.

¹⁹The USPTO may not automatically publish an application 18 months after its priority date if requested by the applicant provided *the invention disclosed in the application has not and will not be the subject of an application filed in another country or under a multilateral international agreement, that requires publication of applications eighteen months after filing* (35 U.S.C. 122(b) and §1.211). Hence, if a Chinese applicant patents an invention with the USPTO that he also patents with SIPO, the application will be published within 18 months

²⁰Note that we have patent data since 1985, but firm-level information is only available since 1999. The patent data preceding 1999 is used in our descriptive analysis as well as to construct patent stocks.

²¹See <http://www.bvdinfo.com/Products/Company-Information/International/ORIANA.aspx>

data; for example, Oriana does not report firms' R&D expenditure nor ownership structure (in terms of paid-in capital contributed by different types of investors following Guariglia et al., 2011).

While this approach allows us for the first time to match patent data to Chinese firms, it also has some limitations. First, Oriana only contains a subset of the firms contained in the census. However, given that Oriana is a subset of ASIE, we can test for differences between the distributions of our variables of interest between Oriana and the full ASIE sample (see Section 4.4). Second, names in PATSTAT as well Oriana might nevertheless differ according to whether names have simply been transcribed using pinyin or (partly) translated. The main challenge in matching the two datasets, therefore, consists in creating an automated matching algorithm that copes with this difficulty as manual matching is unfeasible due to the large number of Chinese patents. Due to the similarity of firm names,²² probability-based matching does not seem appropriate in the present context due to the relatively higher risk of obtaining a large number of 'false positive' matches, i.e., patents erroneously matched to firms.²³ Instead we clean and standardize firm names in both datasets to a maximum possible to avoid the occurrence of 'false negatives'. We also cross-check matched USPTO and SIPO patents using 'equivalents': we verify whether for example a given matched USPTO patent has a SIPO equivalent for the same innovation to ensure that the SIPO equivalent is allocated to the same company.²⁴ In addition, in the case of USPTO patents, we checked all matched and unmatched firms manually. Due to the dramatically larger number of SIPO patents, we only checked a subsample of matched and unmatched patents.²⁵ In case we found 'false negatives', we amended the corresponding standardisation code of our matching algorithm in order to ensure our match is easily reproducible.²⁶ An Appendix explains the matching algorithm in more detail (see also Table 3).

4.4 Data Cleaning

The merged ASIE-Oriana sample contains 143,458 firm-year observations from 23,915 individual firms spanning the period of 1999-2006. There are 1,467 firms that are contained in Oriana but not in ASIE with most of these firms in non-manufacturing industries. For these firms, we only have information in the period 2000-2005, hence, they enter the sample only in 2000 and drop out in 2006. Given that Oriana provides only a limited a number of variables, for example it does not contain R&D expenditure, we dropped firms not covered by ASIE from our sample. This leaves a sample of 135,086 observations from 20,448 firms. We clean the dataset further by dropping firms in non-manufacturing industries contained in ASIE (two digit GBT code >43 or <13). Given the importance of innovation effort for patenting, our regression analysis is constrained by the R&D expenditure measure which is only available in four years, namely 2001, 2002, 2005 and 2006. The sample used in the regression analysis therefore contains 64,652 firm-year observations from 19,956 firms (see Table 2). All variables employed in the regression analysis are discussed in detail in Section 5 below. Note that for our descriptive analysis of patenting in Section 6, we make use of

²²Chinese firm names frequently contain the city or province in which the firm is located. Furthermore, detailed description of the firm status (co. ltd. is often written out in detail as *youxian gongsi* or even *youxian zeren gongsi*) adds to the sense of similarity. Finally, the Chinese language only contains around 400 syllables which further hinders reliable matching by name.

²³See Thoma et al. (2010) for a recent review of probabilistic matching.

²⁴We verify whether the equivalent found in this way contains the same assignee name as the matched patent.

²⁵In the case of USPTO patents, this meant searching for approximately 1,370 unmatched assignee names manually in Oriana and in the case of SIPO patents, we searched for about 10% of the approximately 22,500 unmatched Oriana names among the approximately 145,000 unmatched SIPO assignee names.

²⁶In case of USPTO and SIPO patents, we found a share of only about 2% of 'false negatives' in the assignee names.

the entire data span for which we have patent data which covers the period 1985 to 2006.

Finally, in order to verify potential sample selection issues due to the use of ASIE companies that are contained in Oriana, we ran some empirical tests to analyse the differences between the wider ASIE sample excluding firms in our sample (around 780,000 observations) and the ASIE-Oriana sample (around 65,000 observations). The results are summarised in Table TA-VI.

5 Empirical Strategy

In the present paper we investigate two main research questions employing our integrated dataset: (i) ‘What drives the recent ‘explosion’ in patent applications from firms registered in China?’ (Q1: patenting versus not patenting), and (ii) ‘Are there any differences in the determinants of Chinese firms’ patent productivity between the US and China?’ (Q2: patenting productivity). We will endeavour to provide support for the results from the descriptive analysis in Section 6 through a number of alternative empirical models.

We begin with the patenting question, Q1, whereby we essentially disregard the number of patents taken out by a firm and focus merely on the prevalence of patenting. We first employ a *multinomial logit model*, which allows us to analyse the ‘discrete choice’ over patenting alternatives (do not patent, in China/US only, in both countries — the former is the omitted alternative in the regression) and to investigate the firm characteristics that influence this choice across the alternatives (Agresti, 1990; Cameron and Trivedi, 2005). Following estimation we evaluate the model by comparing predicted and empirical frequencies across alternatives, by investigating the null hypothesis that coefficients for each covariate are zero across all four alternatives and furthermore by analysing LR tests for combination of alternatives (i.e., asking for instance whether we can combine the ‘patent in China only’ with the ‘patent in both China and the US’ alternative). Finally we interpret the results using percentage changes in the odds across different alternatives.²⁷ In this and all of the following empirical models we use a standard set of characteristics/determinants — see below for a detailed discussion of these covariates.

Next we employ a *bivariate probit model* to analyse two dichotomous outcomes, namely patenting with USPTO and patenting with SIPO. The joint outcomes can then be analysed using a simple correlation coefficient (ρ) between the two outcomes; the statistical significance of ρ may either reflect the jointness of the decision to patent in both countries or unobserved factors common to both decisions (Anand and Khanna, 2000), i.e., model misspecification. Furthermore we test the hypothesis that separate probit models fit the data better than the joint model and conduct parameter homogeneity testing across the two equations, which can potentially indicate differences in the impact of covariates on the patenting decision. Perhaps the most interesting element of the bivariate probit analysis is the visual comparison of predicted probabilities: here we first compute the predicted probabilities for each of the four possible outcomes²⁸ and then plot them against various covariates of interest (firm size, export-sales ratio, firm age) using fractional polynomial regression to indicate the general tendency. Our focus here is on the firms which *only* patent in China con-

²⁷The interpretation of these results is as follows: by how many percent will the odds of having a patent alter when increasing the x variable by one unit. The comparison with the omitted base outcome ‘no patent’ can just be interpreted as simple changes in the odds for a unit increase in the covariate, whereas in the comparison of the alternative outcomes the percent changes are relative. This multinomial logit analysis is conducted making extensive use of the routines and examples provided in Long and Freese (2006).

²⁸1: no patents, 2: patents in China only, 3: patents in the US only, 4: patents in both the US and China.

trasted with those which *also* patent in the US.²⁹ We further provide a set of descriptive statistics relating to firm characteristics and sectoral characteristics of the observations making up the four different outcomes.

A final set of results then moves on to analyse patenting productivity, Q_2 , using *count data models* by estimating standard patent production functions which relate the patent *count* (the ‘product’) to a vector of firm-level characteristics (the ‘inputs’). In a single *cross-section* the work by Bound et al. (1984) provides a comparison of multiple empirical implementations using OLS, Poisson, Negative Binomial and Nonlinear Least Squares. In empirical practice the choice between these different approaches is primarily driven by the well-known ‘overdispersion’ problem for the Poisson estimator (e.g. Cameron and Trivedi, 2006; Hilbe, 2011), which represents a violation of the assumed equality between mean and variance of the count variable.³⁰

More recent applications have further concerned themselves with the issue of ‘excess zeros’ in count data, although this phenomenon is intimately linked to overdispersion given that both may arise from unobserved heterogeneity (Cameron and Trivedi, 2006). In order to distinguish the ‘innovating firms’ which chose not to patent in year t from firms which never innovate and therefore never patent (these are labelled ‘certain zeros’) ‘zero-inflated’ versions of the Poisson (ZIP) and Negative Binomial (ZINB) estimators (Lambert, 1992; Winkelmann, 2003) first estimate a nonlinear (logit or probit) model for ‘certain zeros’ and then analyse those observations which are predicted *not* to be ‘certain zeros’ in a count regression model (Poisson or Negative Binomial).³¹ Since the estimation approach is maximum likelihood-based each of the two steps is iterated until the respective ‘best fit’ is achieved. In the present data context the large number of zero patents and thus the use of the ‘zero-inflated’ versions of count data estimators is arguably more important than in the standard approaches using US or OECD data (Hu and Jefferson, 2009): for the USPTO data we have a mere 68 non-zero observations (0.11% of the sample), whereas for the SIPO data the figure is 922 (1.43%). Given the uncertainty over any differences in the determinants of the patenting decision (‘inflation’ equation) and the patenting productivity (‘count data’ equation) we follow the standard in the literature by including *the same covariates* in both equations (Lambert, 1992; Winkelmann, 2003).

In a panel context the standard count data modelling approach is to follow Hausman et al. (1984) by adopting a fixed or random effects Poisson regression, rather than the above ‘cross-section’ estimators, since the former allow for unobserved time-invariant heterogeneity. In the present case of China, however, the use of dedicated panel estimators accounting for unobserved heterogeneity at the firm-level would dramatically reduce the sample size, since only firms with at least one patent over the sample period could be considered. We therefore treat our panel as repeated cross-sections, in the spirit of previous work on China by Hu and Jefferson (2009).

We follow the suggestion in Trivedi and Munkin (2010) and employ a number of diagnostic tools to compare ‘fit’ across different empirical count models. Prime amongst these is a frequency comparison for 0 to 9 patents between the various count regression models and the observed empirical frequency. Furthermore we employ a number of formal ‘goodness-of-fit’ tests and statistics, including Greene’s (1994) variant of the Vuong (1989) model comparison test (Poisson vs. ZIP, Neg-

²⁹Note that using data for firms that *only* patent in the US reduces the sample to a handful of firms.

³⁰Overdispersion is manifest in many count data models whereby the variance of the count variable is often substantially larger than its mean.

³¹Note that the inflation model defines the logit regression in reverse (‘certain zero’=1, not ‘certain zero’=0) to a standard logit model (no patent=0, some patents=1), such that we can expect the reverse signs in the former compared with the latter.

Bin vs. ZINB), as well as various information criteria discussed in Long and Freese (2006). Patent production function results for SIPO and USPTO patents are relegated to a technical appendix — in the main section we present our preferred estimates from the Negative Binomial model, where we estimate the equations for USPTO and SIPO in a seemingly unrelated regression (SUR) framework. This allows us to conduct parameter homogeneity tests across models akin to the analysis in the bivariate probit model (we refer to this as ‘correlated NegBin’).

In the following we briefly discuss the choice of covariates employed in the empirical analysis. The firm-level characteristics considered in this analysis follow the suggestions in Hall and Ziedonis (2001), namely measures for R&D expenditure (innovation effort), firm size, firm age, as well as some characteristics with particular relevance for China, namely firm ownership type and export orientation (export-sales ratio). We employ R&D expenditure deflated by employment (in logs), adding a dummy variable for firms with no data on R&D expenditure (about 1.7% of observations) and a dummy for firms with zero R&D expenditure (about 70% of observations); in both cases we set *log* R&D per worker to zero. We chose this specification to avoid confounding the R&D effect with that of the size of the firm (Hall and Ziedonis, 2001), which is measured by employment (in logs) and meant to capture possible economies of scale in the production of patents. Log R&D expenditure per worker is entered as linear and squared terms to allow for firms at different tails of the distribution to impact patenting decisions and patent count differentially. Firm age is computed from data on the year the company was founded. In an OECD country context this variable is intended to capture the experience of older firms in the management of the patent application process (Hall and Ziedonis, 2001), however in a China emerging from a planned economy, this is an indicator for state-owned enterprises founded during the socialist period prior to 1978 and this variable therefore indicates a socialist period legacy. Ownership types include private enterprises, as well as two types of foreign-invested enterprises (FIEs) in a distinction which is commonly made between those from Hong Kong, Macao and Taiwan (HMT) and from elsewhere (other). We further distinguish Private, State-Owned (SOEs), Collective and Other firms, whereby our designation is based on the proportion of paid-in capital in excess of 50%, following Guariglia et al. (2011). Given China’s strong reliance on manufacturing for export we include export orientation constructed as annual export value over total sales in our regressions. With the exception of dummy variables all of the above are in logarithms. Finally, we add year dummies to all our models which will allow us to chart the increase in patenting both with USPTO and SIPO, accounting for any unobserved effects that impact all firms in our sample such as the 2001 amendment to the Chinese patent law.

Note further that our empirics concentrate primarily on ‘consistency’ (in the plain English sense of the word) in the sign and statistical significance of these covariates across the different empirical models tested. In all empirical models we cluster standard errors at the firm-level.³²

6 Descriptive Evidence

Tables 4 and 5 list the Top 10 companies patenting with the USPTO and SIPO respectively. These tables are constructed using the entire integrated dataset to provide as complete a picture as possible and are thus not restricted to firms and industries included in the regression sample.

Table 4 illustrates the strong concentration of USPTO patents in the hands of very few companies: the top 10 assignees account for over 85% of all USPTO patents. Interestingly, three companies,

³²Note that most of the diagnostic tests carried out, however, have to be constructed from the standard regression residuals.

Hongfujin (1), Fuzhun (3) and Futaihong (6), are subsidiaries of the same Taiwan-owned multinational company Foxconn Technology Group that is the world's largest contract manufacturer in 3C (Computer, Communication, Consumer electronics) products. For example, Hongfujin was founded by Foxconn in Shenzhen in 1998, mainly producing high-tech key components and system products for companies such as Apple, HP, Dell, Cisco, IBM, and Acer. Some of the most renowned products they make/assemble include Apple's iconic ipod, iphone and ipad. In 2006, Hongfujin earned nearly US\$ 20 billion and employed around 20,000 workers in Shenzhen. Similarly, Fuzhun and Futaihong were also established by Foxconn in Shenzhen in 1998 and 2002, respectively. Fuzhun's main products include computer equipments such as heat sinks and coolers, supplying consumer electronics giants such as Sony, Fujitsu, Siemens, and Samsung. Futaihong specialises in wireless communication equipment for clients such as Nokia and Motorola. These three Foxconn subsidiaries account for around 40% of total USPTO patents, and with Huawei these four companies alone account for over 60% of USPTO patents in our matched dataset. As shown in the last column of Table 4, with the exception of Sinopec, China International Marine and BYD, all Top 10 USPTO companies are in the 3C industry (with 2 digit GB/T industry code 40).³³ Table 5, shows a similar picture for SIPO patent holdings with the top 10 companies accounting for about 75% of all patents. In the case of SIPO patents, the dominant player is Huawei, which holds more than a third of SIPO patents, whereas only one Foxconn company, Hongfujin, is among the Top 10. Similarly to the USPTO Top 10, with the exception of Sinopec, BYD and Baoshan Iron & Steel, all companies listed in Table 5 are in the electronics/communications equipment industry. Note that there is a significant overlap of companies in Tables 4 and 5; six companies appear in both lists, with four of these in 3C industries.

It is also noteworthy that these top patenting companies are very large and most of them highly export oriented. Going through the profiles of the companies listed in Tables 4 and 5, we find that, apart from two companies (LG Shanghai and Inventec) in Table 5, all companies employ more than 3,000 workers, with some gigantic companies such as Sinopec, Baoshan Iron & Steel and Huawei employing more than 100,000 workers. It is also not very surprising that most of the firms turn out to be heavily engaged in exporting. According to the Ministry of Commerce, except for three companies (BYD, LG Shanghai and Fuzhun), all of the companies in Tables 4 and 5 are among the top 100 Chinese exporters in 2006.³⁴ Indeed, two Foxconn companies, Hongfujin and Futaihong are ranked 1st and 8th largest exporter respectively in China, with export revenues around US\$ 18.4 billion and US\$ 4.7 billion in 2006. Huawei is ranked 17th with an export revenue of nearly US\$ 3 billion, whilst Lenovo and Sinopec are ranked 12th and 15th with export revenues of US\$ 3.9 and US\$ 3.1 billion, respectively.

To summarise, Chinese companies' patents in China and especially in the US are highly concentrated in a handful of very large, highly export-oriented firms in the computer, communication and consumer electronics industry. In particular, Taiwan-owned contract manufacturer Foxconn emerges as a major player in patenting with the USPTO, whilst two indigenous firms Huawei and ZTE play a leading role in domestic patenting.

Apart from asking who patents, the question of what is patented is equally important. We classify USPTO and SIPO patents according to the type of innovation that they protect:³⁵ product

³³ Apart from the three Foxconn subsidiaries, the other four companies in the 3C industries are Huawei, ZTE, SMI, and Lenovo.

³⁴ See "China's top 200 exporters 2006" available at <http://zhs.mofcom.gov.cn/aa/aa.html>.

³⁵ We read each patent's claims to determine whether it protects a product or process innovation or both. We thank Max Ernicke for his advice on the identification of innovation types in patent claims.

or process innovation or a combination of both.³⁶ In the case of USPTO patents, we read all 1,912 patent documents, whereas in the case of SIPO patents we rely on a subsample of 980 patents due to the fact that claims are not readily available but must be retrieved from the original patent documents which are available only in Chinese. Table 6 shows a breakdown of USPTO and SIPO patents according to the innovation type that they protect. For USPTO patents, Table 6 shows that nearly half of all patents filed by Chinese residents cover product innovations. Only about 20% of patents cover process innovations. The pattern looks different in the case of SIPO patents (excluding USPTO equivalents): nearly 37% of patents protect process innovations and slightly less than 30% cover product innovations. The share of patents protecting both product and process innovations remains approximately the same as for USPTO patents. The last two columns of Table 6 show the distribution of innovation type when we add USPTO equivalents to the SIPO patents.³⁷ We use sample weights in computing shares to reflect the small proportion of SIPO patents with USPTO equivalents (2.5%) among all SIPO patents. The results change very little relative to the distribution when looking only at SIPO patents. In brief, Table 6 suggests that inventions that are patented in China but not simultaneously in the US are more likely to protect process innovations. With regard to USPTO patents, the results indicate that the share of patents protecting product innovations is in fact substantial. The list of such USPTO ‘product patents’ included in our sample covers a wide range of innovations. It includes low-tech products such as a ‘computer enclosure’ (i.e., a metal box) or a ‘refrigerator with a foldable table’ (i.e., an almost standard refrigerator that can be (mis)used as a table); the list also includes a relatively limited number of potentially more valuable product innovations such as a hydrocarbon conversion catalyst, fuel cells, or a foldable mobile telephone (this was in fact novel in 2004 which is the patent’s priority year). Nevertheless, in our reading, product innovations covered by USPTO patents are mostly related to mechanical innovations related to the ICT equipment industry embodying relatively modest innovative content. In contrast, USPTO patents covering processes/methods and the combination of process & product innovations appear to be of more innovative character. The overwhelming share of these patents protect ICT-related innovations, such as central process units or data processing/transmission methods. When we look at the inventive content of SIPO patents, a somewhat similar picture emerges despite the difference between USPTO and SIPO patents in their relative shares of product and process innovations. SIPO patents protecting product innovations are also mostly related to mechanical inventions in the ICT equipment industry, covering inventions such as a ‘turbofan for air conditioner’ (i.e., a fan). At the same time, SIPO ‘product patents’ appear to cover a wider range of products including steel and rubber products, pharmaceuticals (especially related to traditional Chinese medicine), and foodstuff. Similar to USPTO patents, it appears that process patents and those protecting both products and processes related to the ICT sector are of more innovative character.

Finally, Tables 7 and 8 contain information on the sample of firms used in our regression analysis. As previously discussed, due to the limited availability of R&D data, the sample collapses to about 64,500 firm-year observations. Table 7 provides some basic descriptive statistics regarding firms’ patenting activities. The table shows the patent explosion for both USPTO and SIPO patents. The number of USPTO patents increased from 18 to 716 whereas SIPO patent filings increased from 780 to over 11,300 between 2001 and 2006. The table underscores our findings from Tables 4 and 5,

³⁶An example for a product innovation is a chemical compound whereas a process innovation protects for instance the method to produce the compound. This example also illustrates how a patent can protect both a product and process innovation: it covers both the compound and its production method.

³⁷This means we add those USPTO patents that have a SIPO equivalent for which we have not examined the claims. This assumes, that equivalents protect the same innovation type, which may not necessarily hold in all cases.

that is, the overwhelming share of patents is filed by companies in the ICT equipment industry where few companies account for most patents as indicated by the standard deviation and maximum. Table 8 offers descriptive statistics for the variables used in our regression analysis which is discussed next in Section 7.

7 Results

It should be noted from the onset that the present data poses fundamental challenges to the empirical estimators typically employed in patent decision and patenting productivity analysis: in particular for the analysis of USPTO patents, where we only have a minute share of firm-years which have non-zero observations (amounting to 0.11% of observations). We rely in parts on the predictive power of the models (compared with the true outcomes observed) and employ various tests and robustness checks. Specifically in order to address the low share of non-zero observations we employ ‘rare events’ methods (King and Zeng, 2001a,b) in a technical appendix and find no evidence to suggest that our models investigating the patenting decision (Q1) are substantially distorted by this data property. In order to address the same issue in the patenting productivity analysis (Q2) we employ a number of ‘zero-deflated’ count regression models which in terms of diagnostics however seem to be somewhat inferior in to the Negative Binomial model presented below (in the ZINB case this is inferiority is arguably marginal). All of our empirical analysis is, therefore, in the spirit of trying to paint a broad brush picture of the differences in the patenting decision and factors that seem to correlate with high patenting count arising from all the below regressions whilst not relying too heavily on any single empirical method employed.

7.1 Patenting decision

We begin our discussion with the empirical results for the patenting decision, Q1, for which we provide analysis from a multinomial logit model in Table 9 and from a bivariate probit model in Table 10, together with plots for predicted probabilities and various descriptive statistics in Figure 2 and Tables 11 and 12.

Table 9 shows estimates for patenting in China only (column [1]), in the US only ([2]) and in both countries ([3]) in comparison to the omitted reference alternative not to patent at all. The predicted probabilities for each of the three alternatives, presented at the bottom of the table, are close to the observed probabilities while the tests for combination of any of the above categories/alternatives, also presented toward the bottom of the table, rejects in all cases at the 1% level of significance — this provides some indication that we can distinguish these alternatives empirically and that the empirical model has a certain goodness of fit. The independence tests presented in the column to the right of the three equation results suggest that with the exception of firm-ownership none of the covariates are zero across all three equations, thus indicating our set of covariates has explanatory power in the patent decision analysis. In the right half of the table we present the percentage changes in the odds by moving between groups — here we only present results for differences between alternatives which are statistically significant at the 10% level. As we will show below, the group of firms patenting with ‘USPTO only’ is very small whilst merely accounting for a handful of patents. We therefore focus primarily on the results comparing the groups of firms patenting in China only and patenting in both China and the US ([3] vs. [1]).

Investigating the empirical estimates for the three equations more closely we can see that R&D

expenditure has a consistent positive effect on patenting, convex with regards to the ‘SIPO only’ and ‘both’ alternatives, thus suggesting that firms with very large R&D intensity populate both of these groups. The relationship seems much stronger for the group of firms that patent in both China and the US, as is also confirmed by our ‘Changes in the Odds’ analysis. Firm size seems to boost patenting propensity in all three groups. Export intensity has a particularly divisive effect, whereby firms with higher export-sales ratio have a higher propensity to patent in the US — whether they also patent in China or not. A doubling of the export-sales ratio implies a 75% increase in the odds to patent in both countries relative to in China alone. Focusing just on the firms which exclusively patent in China, a doubling of the export-sales ratio will *reduce* the odds to patent by 11%. With regards to firm age it can be seen that older firms are less likely to patent in the US. Finally, while the patent explosion is clearly discernable for the ‘China only’ group the difference to the bulk of the remaining firms which *do* patent (most of which are contained in the ‘both’ group) is not found to be statistically significant at the 10% level except for the final year.

A first tentative conclusion from this analysis is that differences in the R&D intensity, firm age and in particular export intensity seem to drive firm’s decisions to patent abroad vis-à-vis exclusively in China. Surprisingly, firm ownership as defined here does not seem to have any discernable impact on the patenting choice.

Table 10 reports results for a bivariate probit regression, analysing the patenting decisions of firms (simple dummy variable) in a joint empirical model. For the USPTO equation we have around 0.11% of observations with patents, for the SIPO equation around 1.43%. With regard to the regression results, we limit ourselves to discussing the signs and statistical significance of covariates, as well as the result from cross-equation parameter homogeneity tests. The latter are carried out both for individual variables (column marked ‘Indiv.’) and groups of variables (‘Joint’, with the symbols indicating variable groupings). We furthermore compute predicted probabilities and correlate these with some of the key covariates of interest, using nonlinear (fractional polynomial) regression (95% confidence intervals are indicated): this analysis is presented in Figure 2, where the left plot always refers to the predicted probability of patenting *only* in China, whereas the right plot refers to the predicted probability of *also* patenting in the US. In all plots we limit the analysis to firms which do have a patent in either China or the US — results are much more precise if we used the entire sample of firms but we prefer to compare firms that *do* patent to highlight the difference in their characteristics.

Analysing the results in some more detail, a test for independence between the two probit equations rejects emphatically, indicating a high level of correlation between the decision to patent in the US and in China. Parameter homogeneity tests indicate a marginally significant difference in R&D intensity (R&D expenditure per worker, in logs, and its square term) between the USPTO and SIPO patenting decisions, with both indicating a statistically significant positive and convex relationship. As the first row of plots in Figure 2 indicates this is a minor difference in the shape of the innovation effort-patenting relationship between the ‘SIPO only’ and ‘both countries’ groups. For firm size (workers, in logs), we find an increased propensity to patent with size in both equations, although more so for the USPTO equation, hence the rejection in the parameter homogeneity tests — as our second set of plots in Figure 2 indicates this is again a minor difference in degrees, rather than a fundamental difference between the two patenting decisions. The results with regard to export orientation are much more pronounced: the bivariate probit coefficients are negative significant for SIPO and positive significant for USPTO. The suggested importance of export orientation for the patenting decision is illustrated rather starkly when we compare the predicted probabilities

in the third row of Figure 2: with increasing export-intensity the propensity of patenting only in China decreases markedly, whereas the propensity of patenting in both countries rises steadily before a turning point at around 60% export intensity. Related to this, the dummy variable indicating zero exports (around 43% of observations) is positive significant (albeit marginally so) in the SIPO equation. The homogeneity test for firm age (in logs) rejects at the 1% level, with the variable insignificant for the SIPO and negative significant for the USPTO equations, respectively. Although the plots in the bottom row of Figure 2 show some noise in the tails of the age distribution of firms, it is nevertheless indicative that age seems to be positively related with patenting exclusively at SIPO, whereas the relationship appears unrelated or even negative for the USPTO patenting predictions. Interestingly, firm ownership does not differentiate the patenting decision, whether analysed jointly or for individual ownership types. Finally, the comparison of the year dummies quite clearly shows the ‘explosion’ in patenting at SIPO over the sample period, whereas this effect is much less marked in the USPTO equation.

In Tables 11 and 12 we present descriptive statistics and details on sector of operation for each of the four patenting ‘groups’ (SIPO only, USPTO only, no patents, both SIPO and USPTO) respectively. We discuss group medians except for the binary variables where we use the mean. From the top right panel of firm characteristics in Table 11 it is quite clear that the ‘USPTO only’ group is a very marginal one, made up of only 32 firm-year observations during our 4 sample years; the ‘patent stock’ statistics indicate that these firms previously had submitted patents to SIPO, just not in our four sample years. We therefore concentrate on the comparison of the three other groups, beginning with the ‘China only’ versus the ‘China and US’ group in the top left and bottom right panels of the table respectively. The latter group is made up of 150 firm-year observations, whereas the former amounts to 2,200. Firm-level characteristics for the ‘China and US’ group are *noticeably* different: R&D per worker is double that for the ‘China only’ group, export intensity a staggering 7 times, while the firm size is around 70% larger; median age is 14 years for ‘China only’ versus 9 years for ‘China and US’ groups. Foreign as well as state-owned enterprises are represented more prominently in the ‘China and US’ group. Comparing the ‘China only’ with the ‘never patent’ group in the bottom left panel, it can be seen that the latter group’s median R&D per worker is zero while that for the former is CNY 700 per worker, with median firm size the only other characteristic which differs substantially: ‘China only’ firms are roughly double the size of ‘never patent’ firms.

Table 12 lists the Top-10 sectors for each of the four groups (Top-6 for ‘US only’), based on the share of overall SIPO patents (for ‘China only’ group) and/or USPTO patents (for ‘US only’ and ‘China and US’ groups). The picture emerging is rather stark: amongst the firms patenting only in China the ‘Communications equipment and Computers’ (SIC 40) and ‘Instruments and Office Machinery’ (SIC 41) sectors dominate, accounting for around 15% of all SIPO patents. The same two industrial sectors in the group of firms patenting in both countries account for almost 77% of SIPO patents. For USPTO patents the picture is even more extreme³⁸ with the same two ICT equipment sectors mentioned accounting for almost the entire balance of USPTO patents. This reiterates the suggestion above that the patent ‘explosion’ of Chinese firms is driven by a very small number of ICT equipment firms — Huawei and various subsidiaries/affiliates of Foxconn — which account for 97% of all USPTO patents and almost 80% of all SIPO patents.

In summary, this analysis confirms our previous finding that patenting with USPTO relative to SIPO in our regression sample is primarily associated with younger, more export-oriented firms and

³⁸As suggested before the ‘US only’ group is rather insubstantial (less than 2% of all USPTO patents).

to a lesser extent with firm size. While R&D investment clearly matters for patenting, there does not seem to be a substantive difference here between the two patenting decisions, while (as in the previous analysis) ownership type does not seem to impact the patenting choice.

7.2 Patent productivity

We now turn to the empirical analysis of patenting productivity, Q2, which is analysed using count regression models. Detailed results for USPTO and SIPO patenting from a number of alternative estimation methods are presented in Tables TA-II to TA-III in a technical appendix with predicted probabilities for all count data models detailed in Table TA-V. For our discussion here we focus on results for the Negative Binomial model which received favourable diagnostics in comparison to rival approaches.³⁹ Table 13 details results for the USPTO and SIPO equations, which were estimated as seemingly unrelated regression equations so as to construct a joint variance covariance matrix. In the right part of the table we then present parameter homogeneity tests for individual or groups of variables across the two equations. We begin our discussion of the results by noting the LR-test results with regard to the θ parameter which allows for different means and variances in the count data — as can be seen the test emphatically rejects insignificance for both the USPTO and SIPO equations. Note that the sum of absolute deviations of observed and predicted values for the Negative Binomial model is around one tenth and one fiftieth of that of a Poisson model for USPTO and SIPO equations respectively, thus indicating the substantially improved fit of the former over the Poisson alternatives.

Turning to the covariates employed in both models, parameter homogeneity tests suggest that the R&D effort is not associated differentially with patenting with USPTO and SIPO — in both cases the relationship between patenting and R&D intensity is suggested to be convex. Firm size is positive significant in both equations, however the parameter homogeneity test suggests that size matters more for USPTO patenting — this is not surprising, given that our above discussion has revealed that the vast majority of USPTO patents are taken out by a small number of very large global ICT equipment manufacturers. The results for export intensity follow the same pattern as our analysis of the patenting decision: USPTO patent productivity is positively associated with export intensity whereas the correlation is negative for SIPO patenting, with a parameter homogeneity test rejecting emphatically. Firm age is negatively correlated with patent count in both equations, but the association is stronger in the USPTO case, leading to a rejection of parameter homogeneity at the 5% level. In contrast to the previous analysis of the patenting decision firm ownership type is now found to be statistically different in its impact on patent productivity with USPTO and SIPO. Closer analysis however establishes that this result is primarily driven by the ‘Other’ category, such that we do not attach too much attention to this finding. It is, however, notable that foreign-invested firms, both from Hong Kong, Taiwan and Macao as well as from other countries, are significantly positively associated with the patent count, indicating that foreign firms may be more ‘productive’ conditional on all other factors. An alternative interpretation would suggest that these firms benefit from additional R&D facilities owned by the multinational and can thus not be compared like for like with indigenous firms. Another simpler explanation would be that the HMT foreign-ownership variable captures a ‘Foxconn effect’. Finally, the joint analysis of parameters on the year dummies suggests no statistically significant difference between SIPO and USPTO patenting ‘explosions’ — as can be seen, the failure to reject any difference is primarily driven by the large standard errors of

³⁹Table TA-IV presents results for the ‘correlated’ Zero-Inflated Negative Binomial (ZINB) regression model.

the USPTO equation estimates, with homogeneity tests for individual years marginally statistically significant for 2002 and 2005.

In summary, this second analysis focused on patent productivity has revealed very similar patterns to our earlier study of the patenting decision. In broad terms, we have found a number of standard covariates statistically significantly associated with higher patent counts. In terms of our comparison of patenting in China and patenting in the US we found the export intensity, firm age and firm size of particular importance in distinguishing USPTO patent count from SIPO patent count. This analysis suggests that firms associated with USPTO patenting are larger, younger and more export-oriented than their peers associated with SIPO patenting. Caution is however needed in arguing for the robustness of these results: while broadly speaking qualitatively similar to the results just discussed (in terms of signs and magnitudes) the ‘correlated ZINB’ count regression equation could not establish statistical significance in the difference of many of the covariates between the USPTO and SIPO equations, due to the imprecision of estimates in the former.

8 Conclusion

What is behind the recent Chinese patent explosion? Is China transitioning fast from imitating technology to producing genuine innovation? What impact does the patent explosion have on the Chinese economy and on the rest of the world?

While answers to these questions are of immediate concern to policy makers in China and beyond, their empirical investigation has to date been severely hampered by data limitations: there were no data available for Chinese firms that included companies’ actual patents filings. We overcome this constraint and construct a dataset that contains domestic (SIPO) as well as US (USPTO) patent filings by about 20,000 manufacturing firms registered in China. We employ the data to investigate the factors behind the Chinese patent explosion during the period 1999-2006.

Our answer to the first question is unambiguous: a handful of companies in the ICT equipment sector account for the overwhelming share of patents, where this high concentration is even more pronounced in USPTO filings. These companies are very large, relatively young, more R&D intensive, and strongly export oriented, in short, true global players. For these companies, a substantial share of patents covers product innovation albeit of relatively low-tech character. Process innovations and combinations of product and process innovation covered by patents held by these companies appear to be technologically more innovative and potentially valuable. Hence, our results tentatively suggest that these few, highly patent-active companies are not merely ‘innovation castles in the air’ inflated by Chinese public policy directed at increased patenting, but (at least to some extent) innovative companies highly integrated into the global economy.

Does this imply, there is evidence for wider technological take-off among Chinese companies? Our analysis suggests that the answer is most likely negative; patenting is concentrated in very few industries and even within these few industries undertaken by very few albeit highly active companies. Yet, this conclusion is subject to the caveat that our sample covers only about 20,000 manufacturing companies. Referring back to our introductory remarks on the ‘Red Queen Run’ vs. the ‘Middle Income Trap’ arguments, our analysis suggests that reality most likely lies between these two extremes. Contrary to a genuine ‘Red-Queen Run,’ a number of Chinese companies appear to be truly innovative, potentially even pushing the technology frontier beyond China. At the same time, there are very few such companies, and some of the most active among them are foreign-invested.

Most companies are thus likely to concentrate on incremental process innovation rather than the generation of 'new-to-the-world' innovation.

What is the likely impact of the patent explosion? In our view, it points to China becoming an economy that competes not only on cheap labour and sheer scale, but also in terms of innovation. However, not unlike other successful Asian economies, there are very few such companies driving this development.

References

- [1] Agresti, Alan (1990): *Categorical Data Analysis*, New York & Chichester: Wiley.
- [2] Aitken, Brian J., and Ann E. Harrison, (1999): 'Do domestic firms benefit from direct foreign investment? Evidence from Venezuela'. *American Economic Review*, Vol. 89, No. 3, pp. 605-618.
- [3] Blalock, Garrick and Paul J. Gertler (2008): 'Welfare gains from Foreign Direct Investment through technology transfer to local suppliers', *Journal of International Economics*, Vol. 74, No. 2, pp. 402-421.
- [4] Blomström, Magnus and Ari Kokko (1998): 'Multinational corporations and spillovers'. *Journal of Economic Surveys*, Vol. 12, No. 3, pp. 247-77.
- [5] Bound, John, Clint Cummins, Zvi Griliches, Bronwyn H. Hall and Adam B. Jaffe (1984): 'Who Does R&D and Who Patents?', in: National Bureau of Economic Research (ed) *R&D, Patents, and Productivity*, Chapter 2, pp. 21-54.
- [6] Brandt, Loren, Johannes Van Biesebroeck and Yifan Zhang (2011): 'Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing.' NBER Working Paper #15152.
- [7] Breznitz, Dan and Michael Murphree (2011): *Run of the Red Queen*, Yale University Press.
- [8] Cameron, Colin and Pravin K. Trivedi (2005): *Microeconometrics: Methods and Applications*, Cambridge University Press.
- [9] Chang, Sea Jin, Jaiho Chung and Dean Xu (2007): 'FDI and Technology Spillovers in China,' CEI Working Paper Series 2007-7, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.
- [10] Cheung, Kui-yin and Ping Lin (2004): 'Spillover effects of FDI on innovation in China: Evidence from the provincial data', *China Economic Review*, Vol. 15, No. 1, pp. 25-44.
- [11] Chuang, Y.-C. and Hsu, P.-F. (2004): 'FDI, trade, and spillover efficiency: evidence from China's manufacturing sector,' *Applied Economics*, Vol. 36, No. 10, pp. 1103-1115.
- [12] Fischer, William A. and Maximilian von Zedtwitz (2004): 'Chinese R&D: Naissance, Renaissance or Mirage?', *R&D Management* Vol. 34, No. 4, pp. 349-365.
- [13] Gallini, Nancy (2002): 'The economics of patents: Lessons from recent U.S. patent reform,' *Journal of Economic Perspectives*, Vol. 16(2), pp. 131-154.
- [14] Girma, Sourafel, Yundan Gong and Holger Görg (2009): 'What Determines Innovation Activity in Chinese State-owned Enterprises? The Role of Foreign Direct Investment,' *World Development*, Vol. 37, No. 4, pp. 866-873
- [15] Girma, Sourafel, Yundan Gong, Holger Görg and Zhihong Yu (2009): 'Can Production Subsidies Explain China's Export Performance? Evidence from Firm-level Data,' *Scandinavian Journal of Economics*, Vol. 111, No. 4, pp. 863-891

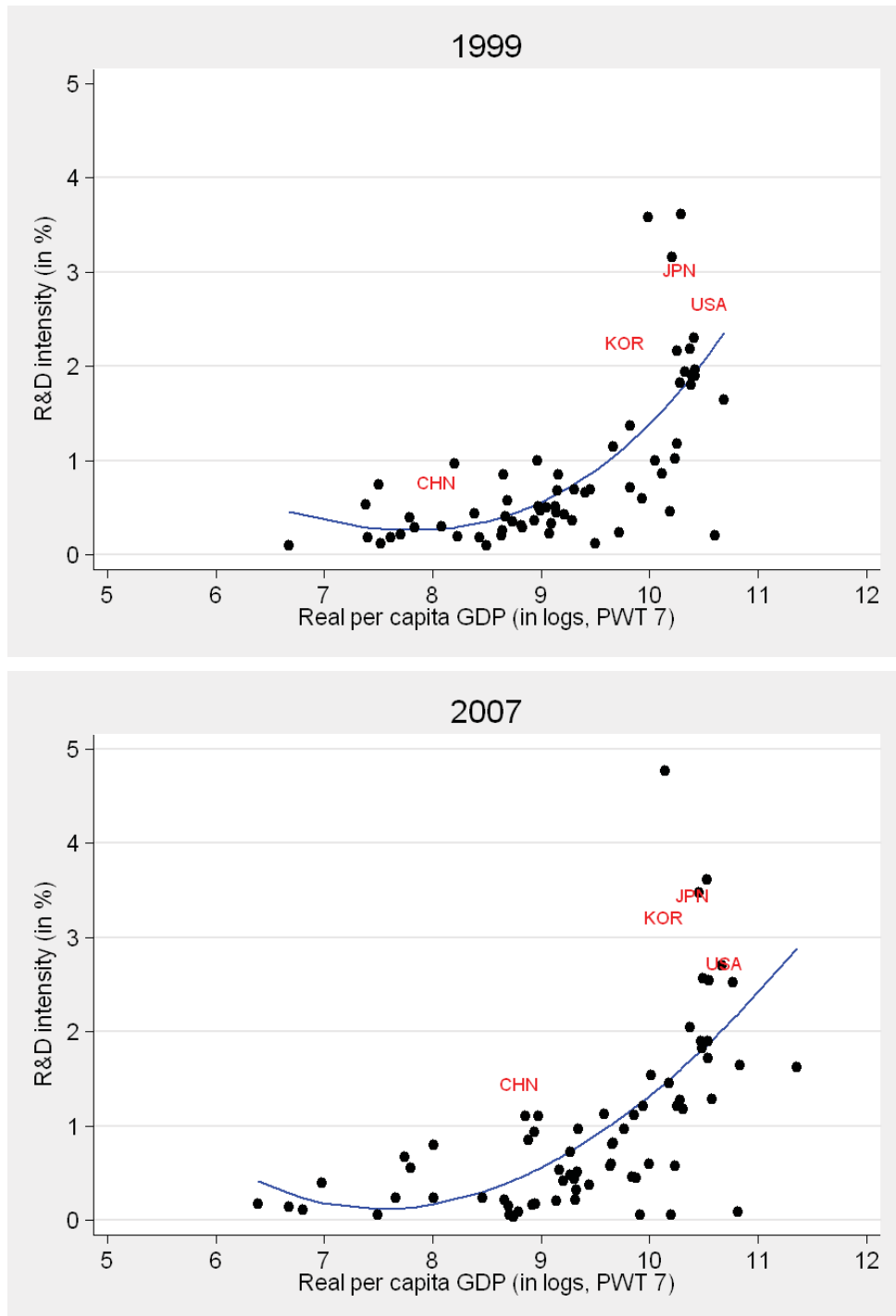
- [16] Görg, Holger and David Greenaway (2004): 'Much ado about nothing? Do domestic firms really benefit from Foreign Direct Investment?' *World Bank Research Observer*, Vol. 19, No. 2, pp. 171-197.
- [17] Greene, William H. (1994): 'Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models,' Working Papers 94-10, New York University, Leonard N. Stern School of Business, Department of Economics.
- [18] Guariglia, Alessandra, Xiaoxuan Liu and Lina Song (2011): 'Internal Finance and Growth: Microeconomic Evidence on Chinese Firms', *Journal of Development Economics*, Vol. 96, No. 1, pp. 79-94.
- [19] Hale, Galina and Cheryl Long (2007): 'Are there productivity spillovers from Foreign Direct Investment in China?', *Pacific Economic Review*, Vol. 16, No. 2, pp. 135-153.
- [20] Haskel, Jonathan E., Sonia C. Pereira and Matthew J. Slaughter (2007): 'Does inward foreign direct investment boost the productivity of domestic firms?' *The Review of Economics and Statistics*, Vol. 89, No. 3, pp. 482-496.
- [21] Hall, Bronwyn H., Zvi Griliches and Jerry A. Hausman (1986): 'Patents and R and D: Is There a Lag?', *International Economic Review*, Vol. 27, No. 2, pp. 265-83.
- [22] Hall, Bronwyn H. and Rose Marie Ham (2001): 'The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry, 1979-95,' *RAND Journal of Economics*, Vol. 32, No. 1, pp.101-128.
- [23] Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches (1984): 'Econometric Models for Count Data with an Application to the Patents-R&D Relationship,' *Econometrica*, Vol. 52, No. 4, pp. 909-38.
- [24] Helmers, Christian, Mark Rogers and Philipp Schautschick (2011): 'Intellectual Property at the Firm-Level in the UK: The Oxford Firm-Level Intellectual Property Database,' University of Oxford, Department of Economics Discussion Paper No. 546.
- [25] Heston, Alan, Robert Summers and Bettina Aten (2011): 'Penn World Table Version 7.0', Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- [26] Hilbe, Joseph M. (2011): 'Negative binomial regression,' 2nd edition, Cambridge University Press.
- [27] Hsieh, Chang-Tai and Peter Klenow (2009): 'Misallocation and Manufacturing TFP in China and India,' *The Quarterly Journal of Economics*, Vol. 124(4), pp. 1403-1448
- [28] Hu, Guangzhou Albert (2010): 'Propensity to patent, competition and China's foreign patenting surge,' *Research Policy*, Vol. 39, pp. 985-993.
- [29] Hu, Guangzhou Albert and Gary Jefferson (2002): 'FDI impact and spillover: Evidence from China's electronic and textile industries', *World Economy*, Vol. 25, No. 8, pp. 1063-1076.

- [30] Hu, Guangzhou Albert and Gary Jefferson (2008) 'Science and Technology in China,' In: *China's Great Economic Transformation*. Brandt, Loren and Thomas Rawski (eds), New York City: Cambridge University Press.
- [31] Hu, Guangzhou Albert and Gary Jefferson (2009): 'A great wall of patents: What is behind China's recent patent explosion?', *Journal of Development Economics*, Vol. 90, No. 1, pp. 57-68.
- [32] Hu, Guangzhou Albert, Gary Jefferson, and Qian Jinchang (2005): 'R&D and Technology Transfer: Firm-Level Evidence From Chinese Industry,' *Review of Economics and Statistics*, Vol. 87, No. 4, pp. 780-786.
- [33] Hu, Mei-Chih and John A. Mathews (2008): 'China's national innovative capacity,' *Research Policy*, Vol. 37, No. 9, pp. 1465-1479.
- [34] Javorcik, Beata S. (2004): 'Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages,' *American Economic Review*, Vol. 94, No. 3, pp. 605-627.
- [35] Jefferson, Gary, Albert Guangzhou Hu, Xiaojin Guan and Xiaoyun Yu (2003): 'Ownership, performance, and innovation in China's large- and medium-size industrial enterprise sector,' *China Economic Review*, Vol. 14, No. 1, pp. 89-113.
- [36] King, Gary, and Langche Zeng (2001a): 'Logistic Regression in Rare Events Data,' *Political Analysis*, Vol. 9, No. 2, pp. 137-163.
- [37] King, Gary, and Langche Zeng (2001b): 'Explaining Rare Events in International Relations,' *International Organization*, Vol. 55, No. 3, pp. 693-715.
- [38] Lambert, Diane (1992): 'Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing,' *Technometrics*, Vol. 34, No. 1, pp. 1-14.
- [39] Liang Zheng and Lan Xue (2010): 'The evolution of China's IPR system and its impact on the patenting behaviours and strategies of multinationals in China,' *International Journal of Technology Management*, Vol. 51, No.2-4, pp. 469-496.
- [40] Lin, Chen, Ping Lin, and Frank Song (2010): 'Property rights protection and corporate R&D: Evidence from China,' *Journal of Development Economics*, Vol. 93, No. 1, pp. 49-62.
- [41] Lin, Yasong and Marco T. Connor (2008): 'An overview of the judicial protection of patents in China,' *Journal of Intellectual Property Law & Practice*, Vol. 3(3), pp. 163-173.
- [42] Lipsey, Robert and Fredrik Sjöholm (2004): 'Host Country Impacts Of Inward FDI: Why Such Different Answers?,' EIJS Working Paper Series 192, The European Institute of Japanese Studies.
- [43] Long, J. Scott and Jeremy Freese (2006): *Regression Models for Categorical Dependent Variables Using Stata*, 2nd edition, College Station, TX: Stata Press.
- [44] Lu, Jiangyong, Yi Lu and Zhigang Tao (2010): 'Exporting behavior of foreign affiliates: Theory and evidence,' *Journal of International Economics*, Vol. 81, No. 2, pp. 197-205.

- [45] Moran, Theodore H. (2007): 'How to Investigate the Impact of Foreign Direct Investment on Development and Use the Results to Guide Policy,' *Brookings Trade Forum*, Foreign Direct Investment, pp. 1-60.
- [46] MOST [Ministry of Science and Technology of the People's Republic of China] (*various years*): 'S&T Statistics Data Book' (1998-2008), MOST website <http://www.most.gov.cn>, accessed 26th April 2011.
- [47] MOST [Ministry of Science and Technology of the People's Republic of China] (2011a): 'National High-tech R&D Program (863 Program)', MOST website <http://www.most.gov.cn>, accessed 20th February 2011.
- [48] MOST [Ministry of Science and Technology of the People's Republic of China] (2011b): 'National Basic Research Program of China (973 Program)', MOST website <http://www.most.gov.cn>, accessed 20th February 2011.
- [49] Naughton, Barry (2007): *The Chinese economy: transitions and growth*, Boston/MA: MIT Press.
- [50] Pakes, Ariel and Zvi Griliches (1980): 'Patents and R&D at the firm level: A first report', *Economics Letters*, Vol. 5, No. 4, pp. 377-381.
- [51] Park, Walter (2008): 'International patent protection: 1960-2005', *Research Policy*, Vol. 37(4), pp. 761-766.
- [52] Puga, Diego and Trefler, Daniel (2010): 'Wake up and smell the ginseng: International trade and the rise of incremental innovation in low-wage countries', *Journal of Development Economics*, Vol. 91(1), pp. 64-76.
- [53] Sorell, Louis (2002): 'A Comparative Analysis of Selected Aspects of Patent Law in China and the United States,' *Pacific Rim Law & Policy Journal*, Vol. 11, pp. 319-338.
- [54] Sun, Yifei (2000): 'Spatial Distribution of Patents in China', *Regional Studies*, Vol. 34, No. 5, pp. 441-454.
- [55] Thoma, G., S. Torrisi, A. Gambardella, D. Guellec, B. Hall, and D. Harhoff (2010): 'Harmonizing and Combining Large Datasets — An Application to Firm-Level Patent and Accounting Data,' NBER Working Paper No. 15851.
- [56] Trivedi, Pravin and Murat Munkin (2010): 'Recent Developments in Cross Section and Panel Count Models', Chapter 4 in A. Ullah and D. Giles (ed.), *Handbook of Empirical Economics and Finance*, pp. 87-131.
- [57] van Pottelsberghe de la Potterie, Bruno (2010): 'The Quality Factor in Patent Systems,' ECARES Working Paper 2010-27.
- [58] Wei, Yingqi and Xiaming Liu (2006): 'Productivity spillovers from R&D, exports and FDI in China's manufacturing sector', *Journal of International Business Studies* Vol. 37, No. 4, pp. 544-557.
- [59] Winkelmann Rainer (2003): 'Econometric Analysis of Count Data,' Heidelberg, 4th edition.

- [60] WIPO (2011): International Patent Classification (IPC), World Intellectual Property Organisation of the United Nations, website accessed 31st May 2011 at <http://www.wipo.int/classifications/ipc/en/>.
- [61] Yang, Deli (2008): 'Pendency and grant ratios of invention patents: A comparative study of the US and China,' *Research Policy*, Vol. 37, No. 6-7, pp. 1035-1046.
- [62] Yu, Xiang (2001): 'The Second Amendment of the Chinese Patent Law and the Comparison between the New Patent Law and Trips,' *The Journal of World Intellectual Property*, Vol. 4, No. 1, pp. 137-155.
- [63] Yu, Miaojie (2011): 'Processing Trade, Firm Productivity, and Tariff Reductions: Evidence from Chinese Products,' Peking University, CCER , Working paper.
- [64] Yueh, Linda (2009): 'Patent Laws and Innovation in China,' *International Review of Law and Economics*, Vol. 29, No. 4, pp. 304-313.
- [65] Zhao, Minyuan (2006): 'Conducting R&D in Countries with Weak Intellectual Property Rights Protection,' *Management Science*, Vol. 52 (8), pp. 1185-1199.
- [66] Zhou, Ping and Loet Leydesdorff (2006): 'The emergence of China as a leading nation in science,' *Research Policy*, Vol. 35, No. 1, pp. 83-104.

Figure 1: R&D intensity (in % of GDP) across countries (1999, 2007)



Notes: We plot gross domestic expenditure on R&D (in percent) against the level of development (proxied by GDP pc, in logarithms). These data come from the UNESCO UIS and the Penn World Table 7.0 (Heston et al, 2011), respectively. For the latter we employed 'China Version 2', a 'non-official' version of China's economic size. The plots use data from 71 (1999) and 74 (2007) countries.

Table 1: Business Share (in %) of national R&D expenditures

	1999	Rank	2007	Rank	Change
Japan	70.7	3	77.9	1	10.2
Republic of Korea	71.4	2	76.2	2	6.7
China	49.6	8	72.3	3	45.8
United States	74.2	1	72.2	4	-2.7
Germany	69.8	4	70.0	5	0.3
United Kingdom	66.8	5	63.4	6	-5.1
France	63.2	6	63.2	7	0.0
Czech Republic	62.9	7	61.9	8	-1.6
Mexico	25.5	11	47.4	9	85.9
Turkey	38.0	9	41.3	10	8.7
Argentina	28.3	10	30.3	11	7.1

Notes: These data come from the UNESCO UIS. The selection of countries is meant to represent a number of OECD economies as well as emerging economies. Change indicates the relative (percentage) change (taking base year values into account).

Table 2: Oriana-ASIE dataset

Year	Oriana		Oriana-ASIE		Share	Sample	
	# observations	%	# observations	%	%	# observations	%
1999	12,542	8.74	11,694	8.66	93.24		
2000	14,017	9.77	13,078	9.68	93.30		
2001	16,969	11.83	15,857	11.74	93.45	14,295	22.11
2002	18,972	13.22	17,618	13.11	92.86	15,929	24.64
2003	20,066	13.99	18,674	13.83	93.06		
2004	21,713	15.14	20,290	15.03	93.45		
2005	20,729	14.45	19,320	14.30	93.20	17,618	27.25
2006	18,450	12.86	18,450	13.66	100	16,810	26.00
Total	143,458	100.00	135,086	100.00		64,652	100.00

Table 3: Benchmarking the matching outcome (1985-2006)

	Assignee names					Patents			
	Raw [‡] Data	Cleaned [§] Data	# Matched	# Equivalents Corrected	Match Success	# Patents	# Matched	# Equivalents Corrected	Match Success
SIPO	168,359	146,333	1,219	1,220	0.83%	405,180	44,344	44,968	11.09%
USPTO	3,580	1,475	112	117	7.93%	4,541	1,880	1,912	42.11%
Oriana	23,858	23,714	1,229	1,229	5.18%				

Notes: SIPO and USPTO patents extracted from PATSTAT version October 2010.

[‡] The data contain patent applications between 1985 and 2006.

[§] Only for-profit companies are kept in the sample.

Table 4: USPTO — Top 10 companies (1985-2006)

Rank	Company	#Patents	Share	Main Industry affiliation [‡]
1	Hongfujin Precision Industry (Foxconn)	513	26.42%	Electronic computer (404)
2	Huawei Technology	399	20.55%	Communications equipment (401)
3	Fuzhun Precision Industry (Foxconn)	215	11.07%	Electronic computer (404)
4	China Petroleum Chemical (Sinopec)	161	8.29%	Crude Petroleum and Natural Gas Exploration (079)
5	Semiconductor Manufacturing International	126	6.49%	Electronic apparatus (405)
6	Futaihong Precision Industry (Foxconn)	100	5.15%	Communications equipment (401)
7	ZTE	61	3.14%	Communications equipment (401)
8	Lenovo	38	1.96%	Electronic computer (404)
9	BYD	33	1.70%	Automobiles (372)
10	China International Marine Containers	18	0.93%	Containers and metallic packages (343)
	Other	278	14.32%	
	Total	1,942	100.00%	

Note: ‡ Chinese GB/T 3 digit industry code in brackets.

Table 5: SIPO — Top 10 companies (1985-2006)

Rank	Company	#Patents	Share	Main Industry affiliation [‡]
1	Huawei Technologies	15,603	34.09%	Communications equipment (401)
2	ZTE	4,594	10.04%	Communications equipment (401)
3	LG Electronics Appliances Tianjin	4,244	9.27%	Household electrical apparatus (395)
4	Hongfujin Precision Industry (Foxconn)	3,710	8.11%	Electronic computer (404)
5	China Petroleum Chemical (Sinopec)	1,977	4.32%	Crude Petroleum and Natural Gas Exploration (079)
6	Lenovo	1,137	2.48%	Electronic computer (404)
7	BYD	835	1.82%	Automobiles (372)
8	LG Electronics Shanghai	775	1.69%	Communications, computers & other electronic equipment (400)
9	Baoshan Iron & Steel	756	1.65%	Ferrous metal smelting and rolling (320)
10	Inventec Shanghai	711	1.55%	Communications, computers & other electronic equipment (409)
	Other	11,423	24.96%	
	Total	45,765	100.00%	

Note: ‡ Chinese GB/T 3 digit industry code in brackets.

Table 6: Product vs. Process Innovation (1985-2006)

Innovation Type	USPTO		SIPO			
	Share	#Patents	Share Excl. US Equivalents	#Patents Incl. US Equivalents [‡]	Share	#Patents
Product	46.81%	895	29.90%	293	29.89%	634
Process	20.35%	389	36.94%	362	36.71%	697
Product & Process	32.85%	628	33.16%	325	33.40%	799
Total	100%	1,912	100%	980	100%	2,130

Notes: Patent classified by reading patent claims.

‡ Equivalents with the USPTO and SIPO patents are weighted with the respective sample share.

Table 7: R&D Sample — Chinese and US Patents (2001/2, 2005/6)

USPTO									
year	ALL SECTORS		ICT EQUIPMENT		ALL SECTORS				
	firms	patents	firms	patents	mean	median	sd	min	max
2001	14,295	18	2,152	13	0.001	0	0.066	0	5
2002	15,929	31	2,393	25	0.002	0	0.098	0	10
2005	17,618	401	2,559	391	0.023	0	1.439	0	146
2006	16,810	716	2,448	710	0.043	0	2.696	0	263
Total	64,652	1,166	9,552	1,139	0.018	0	1.568	0	263
SIPO									
year	ALL SECTORS		ICT EQUIPMENT		ALL SECTORS				
	firms	patents	firms	patents	mean	median	sd	min	max
2001	14,295	780	2,152	641	0.055	0	3.497	0	394
2002	15,929	2,604	2,393	2,397	0.163	0	10.081	0	1,015
2005	17,618	7,761	2,559	7,194	0.441	0	30.362	0	3,781
2006	16,810	11,341	2,448	10,509	0.675	0	51.855	0	6,570
Total	64,652	22,486	9,552	20,741	0.348	0	31.275	0	6,570

Notes: $n = 64,652$ observations for $N = 19,956$ firms, average $T = 3.3$ with minimum $T = 1$ and maximum $T = 4$. The columns marked 'ICT equipment' refer to the subsample of 2,831 firms ($n = 9,552$ observations) in Sectors 40 and 41 (Communications Equipment & Computers; Instruments & Office Machinery) which dominates the USPTO patent count. The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006).

Table 8: R&D Sample — Descriptive Statistics (2001/2, 2005/6)

variable	type	obs	mean	median	sd	min	max
<i>Patents</i>							
USPTO patents	count	64,652	0.018	0	1.568	0	263
SIPO patents	count	64,652	0.348	0	31.275	0	6,570
USPTO Stocks	count	64,652	0.031	0	2.680	0	461
SIPO Stocks	count	64,652	0.705	0	61.466	0	13,003
<i>Innovation effort</i>							
R&D per worker	continuous	63,553	1.386	0	17.377	0	2,434
Missing R&D	dummy	64,652	0.017	0		0	1
Zero R&D	dummy	64,652	0.698	1		0	1
<i>Firm characteristics</i>							
Export/Sales	continuous	64,576	0.319	0.040	0.406	0	1
US Export/Sales	continuous	33,470	14.759	0.002	54.291	0	2,837
Zero Exports	dummy	64,652	0.430	0		0	1
Labour	continuous	64,652	1,154	580	2,759	1	147,722
Age	continuous	64,652	18	12	16	1	179
<i>Ownership type</i>							
FIE (Other)	dummy	64,652	0.166	0	0.372	0	1
FIE (HMT)	dummy	64,652	0.153	0	0.360	0	1
Private	dummy	64,652	0.464	0	0.499	0	1
SOE	dummy	64,652	0.125	0	0.330	0	1
Collective	dummy	64,652	0.076	0	0.265	0	1
Other	dummy	64,652	0.024	0	0.153	0	1

Notes: $n = 64,652$ observations for $N = 19,956$ firms, average $T = 3.2$ with minimum $T = 1$ and maximum $T = 4$. pw — per worker. The 1.7% missing observations for 'R&D pw', amounting to 1,099, make up the balance for the full sample with 63,553 observations for which 'R&D pw' is zero or positive (we set 'ln R&D pw' to zero and add the 'R&D missing' dummy to the regressions below). R&D pw is reported in thousands of real RMB 2000 values. Ownership type uses majority paid-in capital, not official registration, following Guariglia, Liu and Song (2011).

Table 9: R&D Sample — Multinomial Logit regression (2001/2, 2005/6)

	SIPO only	USPTO only	both	Ind‡	Changes in the Odds of Patenting†					
					Raw coefficient, %age change for unit increase in x					
					[1]v.[0]	[2]v.[0]	[3]v.[0]	[2]v.[1]	[3]v.[1]	[3]v.[2]
log R&D pw	0.326 [0.027]***	0.589 [0.175]***	0.639 [0.098]***	242.70 (.00)	38.5%	80.3%	89.5%		36.8%	
(log R&D pw) ²	0.019 [0.008]*	-0.047 [0.062]	0.098 [0.016]***	32.59 (.00)	1.9%		10.3%		8.3%	15.6%
log Workers	0.588 [0.048]***	0.627 [0.223]**	1.567 [0.161]***	426.10 (.00)	80.0%	87.2%	379.1%		166.3%	156.0%
log Exp/Sales	-0.118 [0.025]***	1.696 [0.494]***	0.444 [0.157]**	40.98 (.00)	-11.1%	445.4%	55.8%	513.7%	75.3%	-71.4%
log Firm age	-0.061 [0.055]	-0.711 [0.271]**	-0.593 [0.234]*	12.81 (.01)		-50.9%	-44.8%	-47.8%	-41.3%	
FIE (other)	0.041 [0.166]	-0.072 [1.147]	1.213 [0.450]**	4.43 (.22)			236.3%		222.6%	
FIE (HMT)	-0.016 [0.182]	0.162 [0.983]	1.244 [0.572]*	3.85 (.28)			246.9%		252.6%	
Private	0.073 [0.122]	-0.066 [0.889]	0.338 [0.454]	0.88 (.83)						
Collective	-0.472 [0.286]	0.473 [1.294]	-0.785 [0.661]	4.65 (.20)	-37.7%					
Other	0.143 [0.246]	0.818 [1.294]	1.099 [0.781]	2.40 (.49)						
R&D missing	0.472 [0.296]	-28.672 [0.677]***	1.240 [1.089]	8.83 (.03)	-	-	-	-	-	-
R&D zero	-1.290 [0.101]***	-1.686 [0.671]*	-0.164 [0.420]	81.73 (.00)	-72.5%	-81.5%			208.4%	385.4%
Exports zero	0.171 [0.110]	-1.676 [0.949]	0.178 [0.525]	102.71 (.00)		-81.3%		-84.2%		538.4%
2002	0.437 [0.144]**	0.272 [0.730]	0.455 [0.522]	4.90 (.18)	54.8%					
2005	1.234 [0.148]***	0.449 [0.714]	0.704 [0.522]	207.23 (.00)	243.4%					
2006	1.382 [0.147]***	-0.153 [0.745]	0.428 [0.545]	10.00 (.02)	298.5%			-78.5%	-61.5%	
Constant	-8.772 [0.428]***	-8.880 [2.572]***	-18.665 [1.545]***							
Combine χ^2 ‡ p-value					44.0 (.00)	133.1 (.00)	1584.9 (.00)	35.7 (.00)	51.6 (.00)	320.7 (.00)
Pred Prob (%)	1.350	0.029	0.076							
Obs Prob (%)	1.350	0.029	0.076							

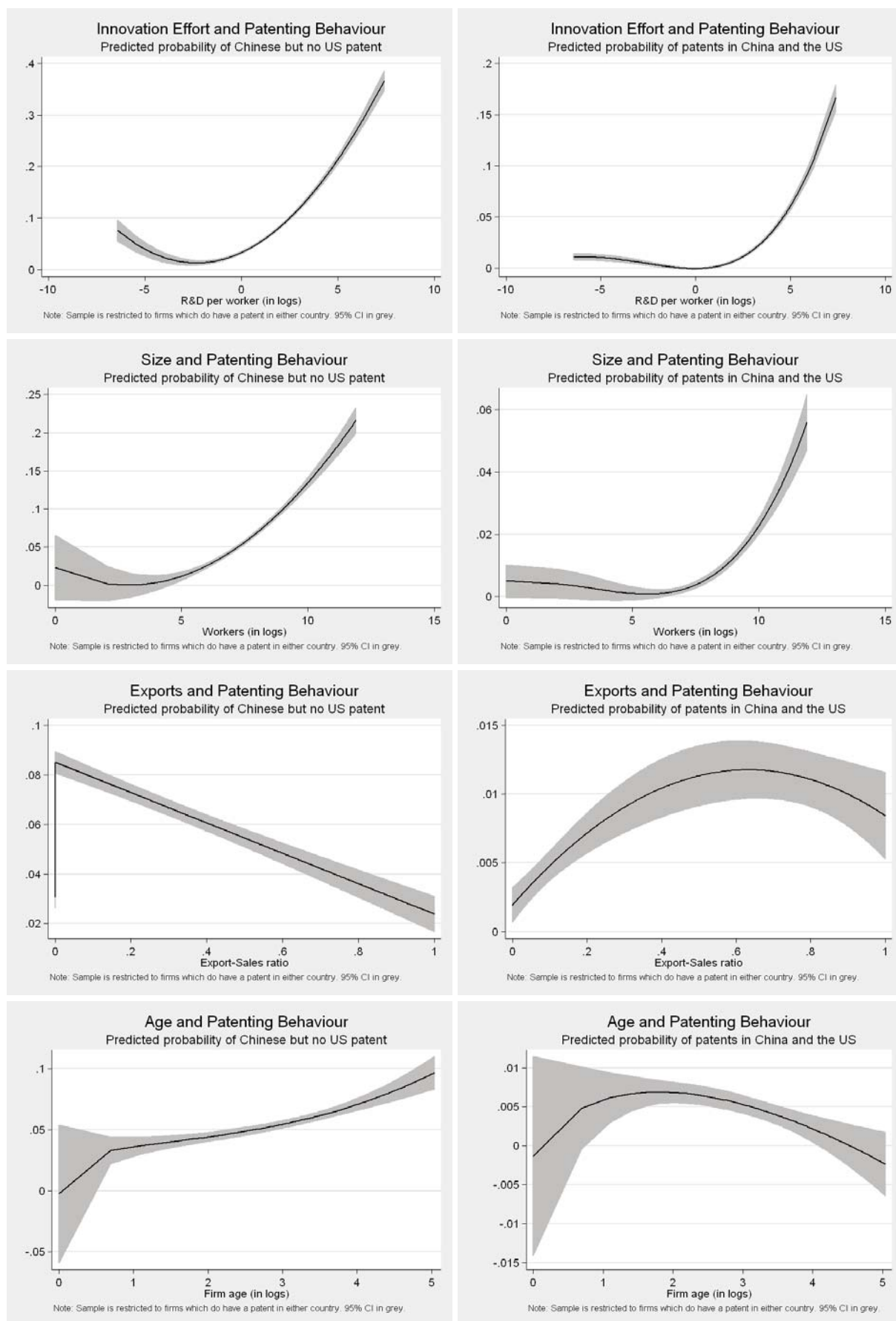
Notes: $n = 64,652$ observations for $N = 19,956$ firms. The omitted category is [0] — no patents (98.53% predicted probability, 98.53% true probability). † We only report the differences between alternatives which are statistically significant at the 10% level. Each column head indicates which two alternatives are tested. We report the %age change for a unit increase (for dummy variables discrete change) in the independent variable (%age change from moving between alternatives, e.g. [1]v.[0] from group without patents to group with patents at SIPO). Note that we do not report the results for missing R&D. ‡ Wald test for the effect of independent variables, which has the H_0 that all coefficients associated with a given variable are zero. Here we also tested (in separate tests) the R&D expenditure, the ownership type and year dummy variables jointly, which rejected the null for all four tests at the 1% level (ownership type at the 5% level). ‡ LR test for combining alternatives, which has the H_0 that all coefficients (except for the intercept) associated with a pair of alternatives are 0 (i.e. the two alternatives could be combined). The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006).

Table 10: R&D Sample — Bivariate Probit regression (2001/2, 2005/6)

dep. var.	[1]	[2]	Homogeneity Testing		
	1_{USPTO}	1_{SIPO}	Individual	Group	Joint
log R&D pw	0.190 [0.027]**	0.159 [0.011]**	1.33 (.25)	×	5.10 (.08)
(log R&D pw) ²	0.025 [0.007]**	0.014 [0.003]**	3.24 (.07)	×	
log Workers	0.397 [0.041]**	0.271 [0.022]**	11.26 (.00)		
log Exp/Sales	0.286 [0.063]**	-0.052 [0.011]**	29.28 (.00)		
log Firm age	-0.213 [0.060]**	-0.040 [0.023]	8.74 (.00)		
FIE (other)	0.204 [0.182]	0.055 [0.067]	0.68 (.41)	◇	2.27 (.81)
FIE (HMT)	0.244 [0.168]	0.041 [0.074]	1.57 (.21)	◇	
Private	0.123 [0.136]	0.043 [0.051]	0.34 (.56)	◇	
Collective	0.071 [0.247]	-0.171 [0.111]	0.94 (.33)	◇	
Other	0.280 [0.217]	0.111 [0.104]	0.59 (.44)	◇	
‘Missing R&D’	0.020 [0.276]	0.137 [0.116]	0.18 (.67)		
‘Zero R&D’	-0.303 [0.117]**	-0.479 [0.039]**	2.28 (.13)		
‘Zero Exports’	-0.231 [0.162]	0.092 [0.045]*	4.29 (.04)		
2002	0.082 [0.128]	0.150 [0.053]**	0.26 (.61)	▷	11.85 (.01)
2005	0.211 [0.131]	0.463 [0.053]**	3.44 (.06)	▷	
2006	0.092 [0.140]	0.525 [0.054]**	9.05 (.00)	▷	
Constant	-5.375 [0.422]**	-4.201 [0.188]**			
Obs	64,652	64,652			
Non-zero obs	0.11%	1.43%			
ρ (st.error)	.733 (.034)				
Wald ($\rho = 0$)	165.2 (.00)				
LL	-4257				

Notes: $n = 64,652$ observations for $N = 19,956$ firms. *, ** indicate statistical significance at the 1% and 5% level respectively. ρ indicates the interrelatedness of the two probit models estimated jointly. Note that the bivariate probit model (SUR) assumes that the model is correctly specified — a model misspecification (omitted variable in both equations) may cause the interrelatedness evident in a statistically significant ρ . We test the null of parameter homogeneity across the two probit regressions, with results presented in the column marked ‘individual’. Joint tests for groups of variables are indicated by the same symbol with results reported in the final column. The omitted base year is 2001, the omitted ownership category ‘state-owned’. The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006).

Figure 2: Predicted Patenting Probabilities



Notes: Plots in the left column graph the predicted probability of patenting in China but not in the US against the variable indicated; plots in the right column are for firms patenting in both China and the US. 95% confidence intervals of a fractional polynomial regression line are indicated in gray. We only use data for firms which do have a patent in the US and/or in China (post-1986, thus overwhelmingly SIPO patents) — these amount to 1,937 observations from 849 firms (around 3% of all firms in our sample). The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006).

Table 11: Firm Characteristics — by patenting behaviour (2001/2, 2005/6)

	FIRM ALWAYS PATENTS IN CHINA						FIRM ONLY PATENTS IN THE US					
	obs	mean	median	sd	min	max	obs	mean	median	sd	min	max
<i>Patents</i>												
USPTO	2,198	0	0	0	0	0	32	0.500	0	1.136	0	6
SIPO	2,198	2.322	0	24.526	0	686	32	0	0	0.000	0	0
USPTO Stock	2,198	0.014	0	0.121	0	1.622	32	1.004	0.850	1.201	0	6,000
SIPO Stock	2,198	5.967	0.850	87	0	3,018	32	0.610	0	0.952	0	2.782
<i>Innovation Effort</i>												
R&D pw	2,129	5.889	0.723	16.249	0	245.309	32	2.551	0.458	4.380	0	14.426
R&D missing	2,198	0.031	0		0	1	32	0.000	0		0	0
R&D zero	2,198	0.294	0		0	1	32	0.375	0		0	1
<i>Firm characteristics</i>												
Exp/Sales	2,197	0.197	0.041	0.298	0	1	32	0.795	0.918	0.255	0	1
Exports zero	2,198	0.313	0		0	1	32	0.031	0		0	1
Labour	2,198	3,130	1,063	8,687	8	147,722	32	1,028	630	1,083	59	4,948
Firm Age	2,198	24.487	14	23	1	154	32	8.688	8.5	4	3	18
<i>Ownership type</i>												
FIE (other)	2,198	0.144	0		0	1	32	0.156	0		0	1
FIE (HMT)	2,198	0.091	0		0	1	32	0.219	0		0	1
Private	2,198	0.507	1		0	1	32	0.594	1		0	1
State	2,198	0.199	0		0	1	32	0.000	0		0	0
Collective	2,198	0.039	0		0	1	32	0.000	0		0	0
Other	2,198	0.034	0		0	1	32	0.031	0		0	1
	FIRM NEVER PATENTS						FIRM PATENTS IN BOTH CHINA AND THE US					
	obs	mean	median	sd	min	max	obs	mean	median	sd	min	max
<i>Patents</i>												
USPTO	62,272	0	0	0	0	0	150	8	0	32	0	263
SIPO	62,272	0	0	0	0	0	150	116	1	634	0	6,570
USPTO Stock	62,272	0	0	0.019	0	3	150	13	0.850	54	0	461
SIPO Stock	62,272	0	0	0.083	0	5	150	214	2.831	1,217	0	13,003
<i>Innovation Effort</i>												
R&D pw	61,246	1.169	0	17.110	0	2,434	146	26.468	1.496	62.115	0	540
R&D missing	62,272	0.016	0		0	1	150	0.027	0		0	1
R&D zero	62,272	0.713	1		0	1	150	0.313	0		0	1
<i>Firm characteristics</i>												
Exp/Sales	62,197	0.323	0.040	0.409	0	1	150	0.428	0.272	0.404	0	1
Exports zero	62,272	0.435	0		0	1	150	0.267	0		0	1
Labour	62,272	1,067	566	2,058	1	135,438	150	8,235	1,783	17,563	43	131,864
Firm Age	62,272	17.466	12	16	1	179	150	15.240	9	16	1	81
<i>Ownership type</i>												
FIE (other)	62,272	0.167	0		0	1	150	0.287	0		0	1
FIE (HMT)	62,272	0.155	0		0	1	150	0.207	0		0	1
Private	62,272	0.463	0		0	1	150	0.320	0		0	1
State	62,272	0.122	0		0	1	150	0.087	0		0	1
Collective	62,272	0.077	0		0	1	150	0.047	0		0	1
Other	62,272	0.024	0		0	1	150	0.053	0		0	1

Notes: Descriptive statistics are reported for each of the four groups of firms, where group affiliation is determined by patenting in the four years that make up the sample period for our regressions above. The patent stock variables indicate that some of these firms had patents in previous/other years (other than 2001, 2002, 2005, 2006). Around 3.7% of observations ($n = 2,380$) in our sample are for the 810 firms which took out a patent in the sample period, whereas 96.3% are for the 19,238 firms that did not patent. The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006).

Table 12: Sector distribution — by patenting behaviour (2001/2, 2005/6)

<i>Firm Always Patents in China</i>							
Sector	Obs	Share	ISIC2	SIPO	Share	USPTO	Share
Communications Equipment, Computers	234	0.36%	40	2,349	10.45%		
Instruments and Office Machinery	284	0.44%	41	1,141	5.07%		
Transport Equipment	189	0.29%	37	230	1.02%		
Pharmaceuticals	291	0.45%	27	208	0.93%		
Chemical Materials and Chemical Products	194	0.30%	26	182	0.81%		
Ferrous metal smelting and pressing	96	0.15%	32	177	0.79%		
General-purpose Equipment	118	0.18%	35	123	0.55%		
Non-ferrous metal smelting and pressing	68	0.11%	33	120	0.53%		
Special Equipment	100	0.15%	36	85	0.38%		
Manufacture of Chemical Fibers	51	0.08%	28	62	0.28%		
Other	573	0.89%		427	1.90%		

<i>Firm Always Patents in United States</i>							
Sector	Obs	Share	ISIC2	SIPO	Share	USPTO	Share
Instruments and Office Machinery	8	0.01%	41			8	0.69%
Plastic products	6	0.01%	30			2	0.17%
Fabricated metal products	6	0.01%	34			2	0.17%
Communications Equipment, Computers	6	0.01%	40			2	0.17%
Pharmaceuticals	2	0.00%	27			1	0.09%
Handicraft	4	0.01%	42			1	0.09%
Other industries	0	0.00%				0	0.00%

<i>Firm Never Patents</i>							
Sector	Obs	Share	ISIC2	SIPO	Share	USPTO	Share
Textiles	5,397	8.35%	17				
Instruments and Office Machinery	4,488	6.94%	41				
Communications equipment, Computers	4,446	6.88%	40				
Textile and garment, footwear	4,383	6.78%	18				
Transport Equipment	4,024	6.22%	37				
Non-metallic mineral products	3,851	5.96%	31				
Chemical Materials and Chemical Products	3,629	5.61%	26				
General-purpose Equipment	2,853	4.41%	35				
Leather, fur	2,476	3.83%	19				
Food processing	2,375	3.67%	13				
Other industries	24,350	37.66%					

<i>Firm Patents in China and the United States</i>							
Sector	Obs	Share	ISIC2	SIPO	Share	USPTO	Share
Instruments and Office Machinery	61	0.09%	41	16,898	75.15%	1,033	88.59%
Communications equipment, Computers	25	0.04%	40	353	1.57%	96	8.23%
Educational and Sports	8	0.01%	24	6	0.03%	5	0.43%
Pharmaceuticals	16	0.02%	27	21	0.09%	5	0.43%
Fabricated metal products	8	0.01%	34	4	0.02%	3	0.26%
non-ferrous metal smelting and pressing	8	0.01%	33	68	0.30%	2	0.17%
Food Manufacturing	4	0.01%	14	3	0.01%	1	0.09%
Beverage	4	0.01%	15	3	0.01%	1	0.09%
Textile and garment, footwear	4	0.01%	18	15	0.07%	1	0.09%
Ferrous metal smelting and pressing	4	0.01%	32	3	0.01%	1	0.09%
Other industries	8	0.01%		8	0.04%	2	0.17%

Notes: We report the top (ten) industrial sectors for each of the four groups defined by their patenting behaviour. In total we have 64,652 firm-years (3.2 per 19,956 firms), 1,166 USPTO patents and 22,486 SIPO patents in the sample years (2001/2, 2005/6). Each 'Share' is expressed with reference to these overall counts.

Table 13: R&D Sample — Correlated NegBin estimation (2001/2, 2005/6)

	Negative Binomial Model				
	USPTO	SIPO	Homogeneity Testing		
			Individual	Group	Joint
log R&D pw	0.730 [0.160]***	0.605 [0.046]***	0.60 (.44)	×	0.60 (.74)
(log R&D pw) ²	0.034 [0.038]	0.046 [0.018]***	0.08(.77)	×	
log Workers	1.906 [0.219]***	1.064 [0.160]***	15.04 (.00)		
log Exp/Sales	1.019 [0.250]***	-0.144 [0.071]**	21.55 (.00)		
log Firm age	-0.971 [0.235]***	-0.414 [0.137]***	4.79 (.03)		
FIE (other)	1.988 [0.915]**	1.445 [0.427]***	0.32 (.57)	◇	34.32 (.00)
FIE (HMT)	2.101 [0.753]***	1.160 [0.537]**	1.81 (.18)	◇	
Private	0.352 [0.671]	0.083 [0.263]	0.14 (.70)	◇	
Collective	-0.626 [1.011]	-0.174 [0.593]	0.15 (.70)	◇	
Other	3.671 [0.898]***	1.475 [0.610]**	9.56 (.00)	◇	
R&D missing	0.341 [1.024]	0.056 [0.382]			
R&D zero	-1.220 [0.430]***	-1.118 [0.261]***	0.04 (.84)		
Exports zero	0.331 [0.566]	0.672 [0.300]**	0.53 (.47)		
2002	-0.341 [0.585]	0.616 [0.174]***	2.62 (.11)	▷	3.74 (.29)
2005	0.503 [0.645]	1.541 [0.246]***	2.46 (.12)	▷	
2006	0.926 [0.672]	1.621 [0.229]***	1.05 (.31)	▷	
Constant	-18.441 [1.897]***	-10.975 [1.328]***			
lnθ	5.285 [0.322]***	4.290 [0.128]***			
Obs	64,652	64,652			
Non-Zero Obs	0.11%	1.43%			
Firms	19,956	19,956			
LL	-560	-6261			
Vuong (p)					
LR-Test $\theta = 0$ (p)	3000.44 (.00)	59,000 (.00)			
$\Sigma \hat{y}_i - y_i $	0.061	0.12			
AIC	0.018	0.194			
BIC _R	-714,816	-703,414			
McFadden R ² _{adj}	0.191	0.130			

Notes: NegBin regressions for SIPO and USPTO patents were combined using seemingly unrelated regression, which yields the same parameter coefficients as in individual regressions but a joint variance-covariance matrix (clustering at the firm-level). We then test parameter homogeneity in count equations indicated, reporting p -values for the H_0 of parameter *homogeneity* across models. We also tested groups of variables (grouping indicated by the symbols) jointly, with results for the H_0 of joint homogeneity reported. See Table TA-II for further details on estimation and diagnostics.

APPENDIX

A Matching of patent data

The Oriana dataset contains 23,714 unique firm names which are matched with the assignee names of SIPO and USPTO patents. The SIPO and USPTO patent files contain 168,359 and 3,580 unique assignee names respectively. The assignees contains a large range of different assignee types, including private individuals, police, military, universities, and public research institutes (e.g., Chinese Academy of Sciences). As a first step in the matching process, we attempt to keep only private and state-owned companies (or some hybrid form) because none of the other assignee types is contained in Oriana. After dropping any assignees that are not private or state-owned companies, cleaning/standardizing assignee names, and keeping only patents applied for between 1985-2006, we obtain 146,333 and 1,475 unique names in the SIPO and USPTO patent files respectively. These two files are then matched with the 23,714 names contained in Oriana. Table 3 shows the resulting matching rates. In a third step, we define equivalent groups.^a We then verify whether the matched sample contains the corresponding equivalents; for example, if a SIPO patent was matched and we find it to have a USPTO equivalent, we check whether the USPTO patent was also matched. If it was not matched, we verify the USPTO patent's assignee name and add it to the matched sample if it coincides with the assignee name of the SIPO patent. This third step ensures consistency between the USPTO and SIPO matches and adds a number of patents to our matched sample (see Table 3).

^aWe apply a definition that assigns patents into the same equivalent group if patents share the same priority documents.

TECHNICAL APPENDIX

Table TA-I: R&D Sample — Rare Events Logit (2001/2, 2005/6)

	Bivariate Probit		Logit	RELogit	Logit	RELogit
	[1]	[2]	[3]	[4]	[5]	[6]
	1_{USPTO}	1_{SIPO}	1_{USPTO}		1_{SIPO}	
log R&D pw	0.190 [0.027]**	0.159 [0.011]**	0.554 [0.086]**	0.509 [0.086]**	0.340 [0.026]**	0.338 [0.026]**
(log R&D pw) ²	0.025 [0.007]**	0.014 [0.003]**	0.071 [0.017]**	0.079 [0.017]**	0.025 [0.007]**	0.026 [0.007]**
log Workers	0.397 [0.041]**	0.271 [0.022]**	1.247 [0.141]**	1.235 [0.141]**	0.641 [0.048]**	0.640 [0.048]**
log Exp/Sales	0.286 [0.063]**	-0.052 [0.011]**	0.647 [0.167]**	0.610 [0.167]**	-0.105 [0.026]**	-0.105 [0.026]**
log Firm age	-0.213 [0.060]**	-0.040 [0.023]	-0.635 [0.194]**	-0.633 [0.194]**	-0.085 [0.055]	-0.085 [0.055]
FIE (other)	0.204 [0.182]	0.055 [0.067]	0.864 [0.473]	0.827 [0.473]	0.119 [0.161]	0.119 [0.161]
FIE (HMT)	0.244 [0.168]	0.041 [0.074]	0.960 [0.490]	0.926 [0.490]	0.061 [0.184]	0.063 [0.184]
Private	0.123 [0.136]	0.043 [0.051]	0.290 [0.406]	0.247 [0.406]	0.091 [0.120]	0.090 [0.120]
Collective	0.071 [0.247]	-0.171 [0.111]	-0.105 [0.758]	0.077 [0.758]	-0.462 [0.279]	-0.445 [0.279]
Other	0.280 [0.217]	0.111 [0.104]	1.004 [0.664]	1.099 [0.664]	0.174 [0.243]	0.184 [0.243]
Missing R&D	0.020 [0.276]	0.137 [0.116]	0.250 [1.009]	0.673 [1.009]	0.508 [0.288]	0.526 [0.288]
Zero R&D	-0.303 [0.117]**	-0.479 [0.039]**	-0.607 [0.375]	-0.624 [0.374]	-1.233 [0.100]**	-1.231 [0.100]**
Zero Exports	-0.231 [0.162]	0.092 [0.045]*	-0.400 [0.515]	-0.355 [0.515]	0.176 [0.112]	0.177 [0.112]
2002	0.082 [0.128]	0.150 [0.053]**	0.350 [0.397]	0.322 [0.396]	0.435 [0.137]**	0.432 [0.137]**
2005	0.211 [0.131]	0.463 [0.053]**	0.508 [0.423]	0.463 [0.423]	1.210 [0.142]**	1.205 [0.142]**
2006	0.092 [0.140]	0.525 [0.054]**	0.109 [0.459]	0.068 [0.458]	1.342 [0.141]**	1.336 [0.141]**
Constant	-5.375 [0.422]**	-4.201 [0.188]**	-14.645 [1.384]**	-14.386 [1.384]**	-9.079 [0.426]**	-9.062 [0.426]**
Obs	64,652	64,652	64,652	64,652	64,652	64,652
Non-zero obs	0.11%	1.43%	0.11%	1.43%	0.11%	1.43%

Notes: $n = 64,652$ observations for 19,956 firms. *, ** indicate statistical significance at the 1% and 5% level respectively. We reprint the Bivariate (SUR) probit estimates from Table 10 in the main section of the paper and add standard logit and ‘rare events’ corrected logit (following King and Zeng, 2001a,b) for USPTO and SIPO patenting, respectively. Focusing exclusively on sign and statistical significance of covariates, we conclude that the rare events correction confirms the results of the standard logit approach.

Table TA-II: R&D Sample — US Patent Production Functions (2001/2, 2005/6)

dep var	Poisson [1]	NegBin [2]	ZIP [3]		ZINB [4]	
	USPTO	USPTO	logit USPTO	Poisson USPTO	logit USPTO	Neg Bin USPTO
log R&D pw	0.685 [0.105]***	0.730 [0.160]***	-0.550 [0.117]***	-0.020 [0.257]	-0.634 [0.216]***	-0.085 [0.451]
(log R&D pw) ²	0.101 [0.030]***	0.034 [0.038]	-0.052 [0.026]*	0.053 [0.039]	-0.071 [0.052]	0.019 [0.127]
log Workers	2.151 [0.166]***	1.906 [0.219]***	-0.957 [0.187]***	0.935 [0.088]***	-0.510 [0.499]	1.682 [0.654]**
log Exports/Sales	1.518 [0.293]***	1.019 [0.250]***	-0.038 [0.517]	1.443 [0.794]*	-0.327 [0.642]	0.837 [0.691]
log Firm age	-1.077 [0.322]***	-0.971 [0.235]***	0.645 [0.258]**	-0.121 [0.360]	-0.172 [0.763]	-1.765 [1.195]
FIE (other)	0.927 [0.568]	1.988 [0.915]**	-1.435 [1.723]	-1.223 [3.816]	0.631 [2.068]	2.659 [3.320]
FIE (HMT)	1.354 [0.521]***	2.101 [0.753]***	-0.958 [1.655]	0.630 [4.119]	1.228 [2.329]	3.892 [3.388]
Private	-0.463 [0.643]	0.352 [0.671]	-0.428 [1.504]	-0.181 [3.697]	1.404 [1.777]	2.930 [2.696]
Collective	-0.088 [0.889]	-0.626 [1.011]	-0.615 [1.790]	-0.949 [3.912]	1.708 [3.021]	2.916 [5.173]
Other	3.309 [1.071]***	3.671 [0.898]***	-0.753 [1.575]	1.267 [3.951]	1.466 [2.625]	5.057 [4.871]
R&D missing	0.349 [1.164]	0.341 [1.024]	0.004 [1.239]	0.362 [1.194]	-0.123 [1.498]	0.382 [2.428]
R&D zero	0.271 [0.372]	-1.220 [0.430]***	0.537 [0.415]	-0.263 [0.371]	-0.557 [1.233]	-2.465 [2.079]
Exports zero	-0.693 [0.871]	0.331 [0.566]	-0.083 [0.543]	-1.167 [0.825]	1.579 [1.594]	2.361 [2.397]
2002	0.291 [0.405]	-0.341 [0.585]	-0.243 [0.816]	0.157 [1.074]	-0.908 [0.926]	-0.991 [1.418]
2005	1.566 [0.821]*	0.503 [0.645]	0.189 [0.822]	1.659 [1.150]	-0.814 [0.967]	-0.310 [1.353]
2006	1.521 [0.806]*	0.926 [0.672]	0.776 [0.814]	2.617 [1.173]**	0.752 [1.383]	2.154 [2.401]
Constant	-20.915 [2.123]***	-18.441 [1.897]***	11.861 [2.482]***	-6.571 [4.171]	7.537 [4.449]*	-11.884 [4.837]**
lnθ		5.285 [0.322]***				1.705 [1.287]
Non-Zero Obs	0.11%	0.11%	0.11%	0.11%	0.11%	0.11%
LL	-2060	-560		-700		-511
Vuong (<i>p</i>)				2.32 (.01)		4.36 (.00)
LR-Test $\theta = 0$ (<i>p</i>)		3000.44 (.00)				377.31 (.00)
$\Sigma \hat{y}_i - y_i $	0.61	0.06		0.09		0.07
AIC	0.064	0.018		0.023		0.017
BIC _R	-711827	-714816		-714359		-714726
McFadden R _{adj} ²	0.792	0.191		0.712		0.236

Notes: $n = 64,652$ observations for 19,956 firms. Dependent variable in all count models is USPTO patent count. Standard errors (in parentheses) clustered at the firm-level. ‘Non-Zero Obs’ indicates the share of firm-years for which the dependent variable is not zero. The omitted base year is 2001, the omitted ownership category ‘state-owned’. The data covers the years for which R&D expenditure is available (2001, 2002, 2005, 2006). The Vuong test compares the Poisson/NegBin models (H_0) with the ZIP/ZINB versions — this test is conducted using unclustered standard errors. The LR-test compares the Poisson/ZIP (H_0) with the NegBin/ZINB model. $\Sigma|\hat{y}_i - y_i|$ reports the summed absolute deviations for average count-predictions of each model relative to the observed frequencies (0 to 9 patents separately, above 9 patents as single category — see Table TA-V. For ZIP and ZINB we also report the ‘inflation’ equation which predicts being in the ‘always zero’ group (vis-a-vis patenting) via logit regression. AIC is $1/n$ that reported by Stata, see Long and Freese (2006: 112). Other measures of fit from countfit routine (Freese & Long, 2006) are also reported.

Table TA-III: R&D Sample — Chinese Patent Production Functions (2001/2, 2005/6)

dep var	Poisson [1]	NegBin [2]	ZIP [3]		ZINB [4]	
	SIPO	SIPO	logit SIPO	Poisson SIPO	logit SIPO	Neg Bin SIPO
log R&D pw	0.724 [0.068]***	0.605 [0.046]***	-0.209 [0.061]***	0.401 [0.116]***	-0.313 [0.067]***	0.329 [0.062]***
(log R&D pw) ²	0.096 [0.015]***	0.046 [0.018]***	-0.008 [0.014]	0.074 [0.028]***	-0.003 [0.016]	0.079 [0.026]***
log Workers	1.686 [0.128]***	1.064 [0.160]***	-0.290 [0.103]***	1.050 [0.123]***	-0.438 [0.067]***	0.756 [0.089]***
log Exports/Sales	0.607 [0.124]***	-0.144 [0.071]**	0.641 [0.138]***	0.869 [0.168]***	0.308 [0.073]***	0.188 [0.074]**
log Firm age	-0.377 [0.218]*	-0.414 [0.137]***	-0.012 [0.122]	-0.336 [0.225]	-0.097 [0.136]	-0.386 [0.195]**
FIE (other)	1.074 [0.468]**	1.445 [0.427]***	-0.253 [0.285]	0.200 [0.564]	0.678 [0.384]*	1.998 [0.705]***
FIE (HMT)	1.323 [0.453]***	1.160 [0.537]**	0.020 [0.276]	0.943 [0.503]*	0.585 [0.361]	1.723 [0.664]***
Private	0.051 [0.384]	0.083 [0.263]	-0.115 [0.228]	0.070 [0.381]	-0.100 [0.248]	0.047 [0.299]
Collective	0.910 [0.418]**	-0.174 [0.593]	0.807 [0.361]**	1.131 [0.577]*	1.064 [0.418]**	0.965 [0.447]**
Other	0.680 [0.826]	1.475 [0.610]**	-0.055 [0.451]	0.575 [0.949]	-0.286 [0.701]	0.386 [0.749]
R&D missing	-0.250 [0.511]	0.056 [0.382]	-0.513 [0.381]	-0.238 [0.534]	0.383 [0.436]	1.203 [0.559]**
R&D zero	-0.610 [0.624]	-1.118 [0.261]***	1.235 [0.283]***	0.122 [0.632]	1.784 [0.231]***	0.785 [0.284]***
Exports zero	-0.124 [0.445]	0.672 [0.300]**	-0.648 [0.206]***	-0.720 [0.458]	-0.499 [0.224]**	-0.307 [0.297]
2002	0.958 [0.447]**	0.616 [0.174]***	-0.228 [0.207]	0.401 [0.348]	-0.401 [0.264]	0.062 [0.300]
2005	0.838 [0.418]**	1.541 [0.246]***	-1.193 [0.231]***	0.057 [0.408]	-1.138 [0.253]***	0.489 [0.321]
2006	0.448 [0.562]	1.621 [0.229]***	-1.529 [0.296]***	-0.404 [0.578]	-1.223 [0.264]***	0.714 [0.340]**
Constant	-14.935 [1.517]***	-10.975 [1.328]***	6.962 [0.867]***	-5.448 [1.520]***	5.666 [0.899]***	-6.249 [1.294]***
lnθ		4.290 [0.128]***				2.376 [0.384]***
Non-Zero Obs	1.43%	1.43%	1.43%	1.43%	1.43%	1.43%
LL	-35763	-6261		-19232		-5855
Vuong (<i>p</i>)				7.14 (.00)		9.89 (.00)
LR-Test $\theta = 0$ (<i>p</i>)		59,000 (.00)				26,754 (.00)
$\Sigma \hat{y}_i - y_i $	6.21	0.12		0.43		0.38
AIC	1.107	0.194		0.596		0.182
BIC _R	-644,422	-703,414		-677,295		-704,038
McFadden R ² _{adj}	0.797	0.130		0.774		0.184

Notes: $n = 64,652$ observations for 19,956 firms. Dependent variable in all count models is SIPO patent count. See Table TA-II for further details on estimation and diagnostics.

Table TA-IV: R&D Sample — Correlated ZINB estimation (2001/2, 2005/6)

	Zero-Inflated Negative Binomial Model									
	logit		Homogeneity Testing			NegBin		Homogeneity Testing		
	USPTO	SIPO	Indiv.	Group	Joint	USPTO	SIPO	Indiv.	Group	Joint
log R&D pw	-0.634 [0.216]***	-0.313 [0.067]***	(.14)	×	(.01)	-0.085 [0.451]	0.329 [0.062]***	(.36)	×	(.00)
(log R&D pw) ²	-0.071 [0.052]	-0.003 [0.016]	(.20)	×		0.019 [0.127]	0.079 [0.026]***	(.64)	×	
log Workers	-0.510 [0.499]	-0.438 [0.067]***	(.88)			1.682 [0.654]**	0.756 [0.089]***	(.17)		
log Exp/Sales	-0.327 [0.642]	0.308 [0.073]***	(.32)			0.837 [0.691]	0.188 [0.074]**	(.34)		
log Firm age	-0.172 [0.763]	-0.097 [0.136]	(.92)			-1.765 [1.195]	-0.386 [0.195]**	(.26)		
FIE (other)	0.631 [2.068]	0.678 [0.384]*	(.98)	◇	(.20)	2.659 [3.320]	1.998 [0.705]***	(.84)	◇	(.03)
FIE (HMT)	1.228 [2.329]	0.585 [0.361]	(.78)	◇		3.892 [3.388]	1.723 [0.664]***	(.52)	◇	
Private	1.404 [1.777]	-0.100 [0.248]	(.40)	◇		2.930 [2.696]	0.047 [0.299]	(.28)	◇	
Collective	1.708 [3.021]	1.064 [0.418]**	(.83)	◇		2.916 [5.173]	0.965 [0.447]**	(.71)	◇	
Other	1.466 [2.625]	-0.286 [0.701]	(.51)	◇		5.057 [4.871]	0.386 [0.749]	(.33)	◇	
R&D missing	-0.123 [1.498]	0.383 [0.436]				0.382 [2.428]	1.203 [0.559]**			
R&D zero	-0.557 [1.233]	1.784 [0.231]***	(.06)			-2.465 [2.079]	0.785 [0.284]***	(.12)		
Exports zero	1.579 [1.594]	-0.499 [0.224]**	(.19)			2.361 [2.397]	-0.307 [0.297]	(.27)		
2002	-0.908 [0.926]	-0.401 [0.264]	(.60)	▷	(.07)	-0.991 [1.418]	0.062 [0.300]	(.47)	▷	(.27)
2005	-0.814 [0.967]	-1.138 [0.253]***	(.75)	▷		-0.310 [1.353]	0.489 [0.321]	(.57)	▷	
2006	0.752 [1.383]	-1.223 [0.264]***	(.17)	▷		2.154 [2.401]	0.714 [0.340]**	(.55)	▷	
Constant	7.537 [4.449]*	5.666 [0.899]***				-11.884 [4.837]**	-6.249 [1.294]***			
lnθ						1.705 [1.287]	2.376 [0.384]***			
Obs	64,652	64,652				64,652	64,652			
Non-Zero Obs	0.11%	1.43%				0.11%	1.43%			
Firms	19,956	19,956				19,956	19,956			
LL						-511	-5855			
Vuong (p)						4.36 (.00)	9.89 (.00)			
LR-Test θ = 0 (p)						377.31 (.00)	26,754 (.00)			
Σ ŷ _i - y _i						0.074	0.380			
AIC						0.017	0.182			
BIC _R						-714726	-704,038			
McFadden R ² _{adj}						0.236	0.184			

Notes: ZINB regressions for SIPO and USPTO patents were combined using seemingly unrelated regression, which yields the same parameter coefficients as in individual regressions but a joint variance-covariance matrix (clustering at the firm-level). We then test parameter homogeneity in count equations (and the excess zero/inflation equation in the ZINB case) indicated, reporting p -values for the H_0 of parameter homogeneity across models. We also tested groups of variables (grouping indicated by the symbols) jointly, with results for the H_0 of joint homogeneity reported.

Table TA-V: R&D Sample — Predictions for USPTO and SIPO models

PANEL A: USPTO models												
Patents	0	1	2	3	4	5	6	7	8	9	>9	$\Sigma \hat{y}_i - y_i $
Observed	99.890	0.060	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	
Poisson	99.601	0.286	0.047	0.020	0.011	0.006	0.003	0.002	0.001	0.001	0.022	
$ \hat{y}_i - y_i $	0.289	0.226	0.037	0.020	0.011	0.006	0.003	0.002	0.001	0.001	0.018	0.614
NegBin	99.895	0.062	0.015	0.007	0.004	0.003	0.002	0.001	0.001	0.001	0.010	
$ \hat{y}_i - y_i $	0.005	0.002	0.005	0.007	0.004	0.003	0.002	0.001	0.001	0.001	0.030	0.061
ZIP	99.896	0.033	0.016	0.010	0.007	0.005	0.004	0.003	0.002	0.002	0.022	
$ \hat{y}_i - y_i $	0.006	0.027	0.006	0.010	0.007	0.005	0.004	0.003	0.002	0.002	0.018	0.090
ZINB	99.898	0.045	0.015	0.008	0.005	0.004	0.003	0.002	0.002	0.001	0.018	
$ \hat{y}_i - y_i $	0.008	0.015	0.005	0.008	0.005	0.004	0.003	0.002	0.002	0.001	0.022	0.074
PANEL B: SIPO models												
Patents	0	1	2	3	4	5	6	7	8	9	>9	$\Sigma \hat{y}_i - y_i $
Observed	98.570	0.720	0.270	0.120	0.050	0.050	0.030	0.030	0.020	0.020	0.120	
Poisson	95.464	3.130	0.580	0.234	0.128	0.080	0.054	0.038	0.029	0.022	0.241	
$ \hat{y}_i - y_i $	3.106	2.410	0.310	0.114	0.078	0.030	0.024	0.008	0.009	0.002	0.121	6.21
NegBin	98.574	0.707	0.237	0.118	0.071	0.047	0.034	0.025	0.020	0.016	0.152	
$ \hat{y}_i - y_i $	0.004	0.013	0.033	0.002	0.021	0.003	0.004	0.005	0.000	0.004	0.032	0.12
ZIP	98.602	0.522	0.250	0.145	0.093	0.064	0.046	0.035	0.027	0.022	0.195	
$ \hat{y}_i - y_i $	0.032	0.198	0.020	0.025	0.043	0.014	0.016	0.005	0.007	0.002	0.075	0.43
ZINB	98.605	0.566	0.234	0.130	0.083	0.058	0.042	0.032	0.026	0.021	0.204	
$ \hat{y}_i - y_i $	0.035	0.154	0.036	0.010	0.033	0.008	0.012	0.002	0.006	0.001	0.084	0.38

Notes: For each estimator we compute the average predicted outcome/frequency for zero to nine patents and over nine patents (reported in %). The deviations from the observed frequencies (at the top of each panel) are reported in the lines marked $|\hat{y}_i - y_i|$, ‘Sum’ simply adds up these absolute deviations.

Table TA-VI: Testing Sample Representativeness

Sample	Full [1]		Full [2]		ICT sector only [3]		Full: Random sample [4]		Full [5]	
Equality test	Conditional		Conditional		Conditional		Conditional		Unconditional	
	coeff	×sample	coeff	×sample	coeff	×sample	coeff	×sample	coeff	×sample
log R&D pw	0.125 [0.002]***	0.073 [0.004]***	0.115 [0.002]***	0.056 [0.004]***	0.118 [0.004]***	0.046 [0.009]***	0.130 [0.002]***	0.007 [0.006]	-0.033 [0.001]***	-0.085 [0.003]***
(log R&D pw) ²	0.012 [0.001]***	0.007 [0.001]***								
log Workers	0.668 [0.001]***	0.114 [0.004]***	0.668 [0.001]***	0.113 [0.004]***	0.737 [0.004]***	0.108 [0.012]***	0.752 [0.001]***	-0.005 [0.004]	4.592 [0.001]***	1.879 [0.004]***
log Exp/Sales	-0.161 [0.002]***	-0.039 [0.004]***	-0.162 [0.002]***	-0.040 [0.004]***	-0.082 [0.005]***	-0.046 [0.011]***	-0.183 [0.002]***	-0.001 [0.006]	-0.245 [0.001]***	-0.464 [0.004]***
log Firm age	-0.058 [0.001]***	-0.037 [0.005]***	-0.058 [0.001]***	-0.039 [0.005]***	-0.095 [0.006]***	-0.075 [0.017]***	-0.056 [0.001]***	0.002 [0.005]	2.071 [0.001]***	0.453 [0.004]***
FIE (other)	0.427 [0.009]***	-0.025 [0.025]	0.427 [0.009]***	-0.014 [0.025]	0.254 [0.030]***	0.184 [0.071]***	0.464 [0.009]***	-0.045 [0.031]	0.068 [0.000]***	0.098 [0.001]***
FIE (HMT)	0.174 [0.009]***	-0.361 [0.025]***	0.173 [0.009]***	-0.355 [0.025]***	-0.104 [0.030]***	-0.150 [0.072]**	0.137 [0.009]***	-0.068 [0.031]**	0.072 [0.000]***	0.081 [0.001]***
Private	-0.718 [0.009]***	0.324 [0.026]***	-0.716 [0.009]***	0.329 [0.026]***	-0.860 [0.031]***	0.298 [0.078]***	-0.678 [0.009]***	-0.061 [0.031]**	0.672 [0.001]***	-0.208 [0.002]***
Collective	0.128 [0.008]***	-0.287 [0.023]***	0.127 [0.008]***	-0.283 [0.023]***	-0.088 [0.028]***	-0.063 [0.069]	0.087 [0.008]***	-0.039 [0.028]	0.098 [0.000]***	-0.022 [0.001]***
Other	0.108 [0.008]***	-0.267 [0.027]***	0.107 [0.008]***	-0.261 [0.027]***	-0.014 [0.030]	-0.368 [0.083]***	0.056 [0.008]***	-0.048 [0.030]	0.078 [0.000]***	0.046 [0.001]***
R&D missing	-0.297 [0.012]***	-0.069 [0.036]*	-0.343 [0.011]***	-0.100 [0.036]***	-0.260 [0.039]***	-0.109 [0.099]	-0.405 [0.011]***	-0.012 [0.041]	0.891 [0.000]***	-0.193 [0.001]***
R&D zero	-0.312 [0.005]***	-0.105 [0.011]***	-0.358 [0.004]***	-0.139 [0.010]***	-0.263 [0.011]***	-0.170 [0.025]***	-0.431 [0.004]***	-0.014 [0.013]	0.741 [0.001]***	-0.312 [0.002]***
Exports zero	0.156 [0.003]***	0.285 [0.010]***	0.157 [0.003]***	0.287 [0.010]***	0.085 [0.012]***	0.214 [0.031]***	0.182 [0.003]***	-0.011 [0.012]	0.011 [0.000]***	0.003 [0.000]***
y4	0.070 [0.004]***	0.009 [0.012]	0.070 [0.004]***	0.009 [0.012]	0.040 [0.013]***	0.018 [0.033]	0.079 [0.004]***	-0.015 [0.014]		
y7	0.469 [0.003]***	-0.190 [0.012]***	0.469 [0.003]***	-0.187 [0.012]***	0.365 [0.012]***	-0.174 [0.033]***	0.453 [0.004]***	-0.009 [0.013]		
y8	0.609 [0.003]***	-0.233 [0.012]***	0.610 [0.003]***	-0.228 [0.012]***	0.503 [0.012]***	-0.224 [0.033]***	0.588 [0.003]***	-0.007 [0.013]		
Constant	6.519 [0.012]***	0.361 [0.041]***	6.563 [0.011]***	0.393 [0.041]***	6.631 [0.040]***	0.406 [0.115]***	6.300 [0.011]***	0.088 [0.041]**		
Obs	841,394		841,394		68,189		841,394		841,394 (each)	

Notes: Results presented in columns [1]-[4] represent the regression coefficients from pooled regressions of log sales on all covariates listed. In each case (with exception of column [4], see below) the result in the left column marked ‘coeff’ represents the coefficient for the 780,000 ASIE observations *not* contained in our matched ASIE-Oriana data, while the result in the right column marked ‘×sample’ represents the coefficient for the 65,000 observations in our ASIE-Oriana matched sample (implemented via an interaction term × the covariate). It can be seen that for the vast majority of covariates the latter term is statistically significant: the ASIE-Oriana matched sample is *not* representative of the wider ASIE database. In column [4] we conduct a counterfactual experiment, at random selecting 65,000 observations to make up the ‘pseudo’ ASIE-Oriana matched sample. As can be seen for the vast majority of covariates the interaction terms are statistically insignificant, as one would expect from random sampling. Finally, in column [5] we simply regress each variable on an intercept (result in left column) and a dummy for the ASIE-Oriana matched sample (right column) — thus each reported line represents a separate regression of 841,394 observations. We term these regressions as ‘unconditional’ equality tests, whereas the results in the previous columns represent conditional equality tests.