Is the Dragon Learning to Fly?
An Analysis of the Chinese Patent Explosion

Markus EBERHARDT[a,c*] Christian HELMERS[b,c] Zhihong YU[a]

[a] *University of Nottingham*

[b] *Universidad Carlos III de Madrid*

[c] *Centre for the Study of African Economies, University of Oxford*

January 2012

Abstract: This paper analyses characteristics of the recent explosion of patent filings by Chinese firms both domestically and in the United States. We construct a firm-level dataset by matching USPTO and SIPO patents to Chinese manufacturing census data for 1999-2006. Using this integrated 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. Our analysis further suggests that firms patenting in both the US and China are younger, larger and 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, firms, patents

JEL classification: L25, O12

*Corresponding author: christian.helmers@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 patent filings with the Chinese patent office has increased at an average annual rate of 32 percent 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 total 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 timeframe. These ambitious targets reflect an overall positive outlook for Chinese development in general (e.g. Subramanian, 2011) and Chinese innovation in particular (e.g. Fischer and von Zedtwitz, 2004) in parts of the literature.

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 inventions embodying little technological progress and driven mainly by the incentives put in place by the Chinese government to encourage patenting directly (*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:

[1]According to OECD data, the number of full time equivalent R&D workers has nearly tripled between 1998 and 2008 reaching almost 2 million. Over the same time period R&D intensity (R&D-GDP ratio) doubled, from 0.8 to 1.5 percent, while the share of national R&D expenditure spent by business enterprises jumped from 45 to almost 70 percent (MOST, various). Zhou and Leydesdorff (2006) point to China's rapid increase in terms of scientific publications.

[2]In comparison, for example, the average annual growth rate of domestic filings by US residents with the USPTO during the same time period was only 6.5 percent (WIPO Statistics Database, January 2011).

[3]Breznitz 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.

only domestically with SIPO or also with the USPTO. Not only are the direct and indirect costs associated higher in the US, but inventions are required to overcome a higher novelty hurdle in the patent examination during our sample period (see Section 3). These differences suggest that a comparison of patents filed with the USPTO and SIPO reveals information on the underlying invention and the corresponding patentees.

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 — to the best of our knowledge this is the first study to match actual patent data to a large firm-level dataset for China. The period covered 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, a strong increase in exporting, an amendment of the patent law, a swelling tide of highly educated Chinese returnees and an accelerated pace of privatisation and of 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 Chinese companies, concentrated in the ICT equipment industry, accounts for the largest share of the dramatic increase in USPTO patents held by Chinese residents, with underlying technologies mostly related to electronics and semiconductors. This select group of firms also accounts for the overwhelming share of SIPO patents despite there being a relatively larger number of companies across a wider range of industries obtaining domestic patent protection.

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 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 in the country. 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 patenting decision as well as in that of the patent count. We conclude 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 patenting in China. 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

The existing literature on patenting in China is surprisingly sparse. While there are some studies at the aggregate, industry and province level, there is little work at the firm-level. Moreover, the work at the firm-level relies on firms' self-reported patents due to a lack of actual patent information. 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 and Hsu, 2004; 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. 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. The former is interpreted as an indicator of higher patent value while the latter is argued 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 percent per year between 1995-2004. The analysis relies on a concordance table produced by the OECD to allocate patents across industries. Using this concordance table, Hu 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 suggested to be driven by import competition, that is, foreign patenting is positively correlated with the amount imported by China from technologically similar industries in other countries.

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 a substantial impact on Chinese development in general and with regard to innovation activities in particular. It is thus not surprising that output and employment of state-owned and collective 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. 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. Yueh (2009) explores the determinants of aggregate patent counts in 29 Chinese provinces for the period 1991-2003 uncovering substantial differences in patenting activity across provinces, which is 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 NBS dataset 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 to 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.

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.[5]

China's first patent law came into force in 1985 and was since amended three times (in 1992, 2001, and 2009). The second comprehensive amendment of the patent law, adopted on 25th August 2000 and effective from 1st 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, most importantly for our analysis, 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.[6] Yang (2008) points out that different emphasis is put on the 'industrial applicability' criterion during the examination process: whereas the USPTO has a broad interpretation of the potential practical purpose an invention might serve, SIPO

[5]USPTO '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. Sorell (2002), Yang (2008), and Lin and Connor (2008) offer detailed comparisons of the two systems.

[6]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.

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, officially applies a narrower definition more in line with the stance of the European Patent Office (EPO).

The fee structure 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.[7] 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.[8] Additional costs for Chinese firms arise from the need to translate the patent application into English. If a Chinese applicant employs 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.

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).[9]

With regard to patents we are not aware of any existing large-scale dataset for China 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). 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 substantial, potentially non-random mis-reporting.[10]

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

[7] Fees for small firms are half the normal fee. See http://tinyurl.com/66xb774 (USPTO website).

[8] Small entities pay half the standard fee.

[9] 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.

[10] For 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 percent of firms that report to have applied for a patent actually did so.

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 US$600,000). On average, more than 200,000 firms are included each year and they account for around 95 percent of total Chinese industrial output and 98 percent 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 1.5 million observations from about 540,000 firms. 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, output, sales, employment, fixed assets, and industry affiliation.[11]

4.2 Patent Data

The patent data come from the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT), version October 2010. 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).[12] 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.[13]

4.3 Matching/Bridge

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).

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

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

[13]Given the usual 18-month delay between application and publication date at both USPTO and SIPO, 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. Note that the patent data preceding 1999 is used in our descriptive analysis as well as to construct patent stocks.

The Chinese census data contain only firm names using Chinese characters (a), 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 offers a firm-level dataset reporting balance sheet and profit & loss information for individual companies in the Asia-Pacific region. 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 a unique identification number (same id as in ASIE). This allows us to link Oriana to the census through the unique identifier and 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.[14]

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 Table TA-3 in the technical appendix). 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 a 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 Chinese firm names, 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'.[15] In addition, in the case of USPTO patents, we check all matched and unmatched firms manually. Due to the considerably larger number of SIPO patents, we only checked a subsample of matched and unmatched patents.[16] Further details of the matching algorithm can be found in a technical appendix.

5. EMPIRICAL STRATEGY

We investigate two main research questions employing our integrated dataset: (i) 'What drives the recent 'explosion' in patent applications from Chinese firms and are there differences between their patenting in China and the US?' (Q1: patenting versus not patenting), and (ii) 'Are there any differences in the determinants of Chinese firms' patent productivity in China and the US?' (Q2: patenting productivity). We address these questions through descriptive evidence provided in Section 6 as well as a number of alternative empirical models which are discussed in this Section. The corresponding results are shown in Section 7.[17]

[14]Oriana's coverage of firm-level variables is far less comprehensive than the ASIE 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).

[15]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. We verify further whether the equivalent found in this way contains the same assignee name as the matched patent. We do this for both USPTO and SIPO patents.

[16]In 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 percent of the approximately 22,500 unmatched Oriana names among the approximately 145,000 unmatched SIPO assignee names.

[17]We cluster standard errors at the firm-level in all empirical models. Note that most of the diagnostic tests carried

We begin with the patenting question, `Q1`, where we 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 alternatives (do not patent, in China/US only, in both countries) and to investigate the firm characteristics that influence this choice across the alternatives.[18] 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. A simple test may indicate the jointness of the decision to patent in both countries, which may however be driven by common unobservables (Anand and Khanna, 2000).[19] Perhaps the most interesting element of our analysis is the visual comparison of predicted probabilities, where we focus on firms which *only* patent in China vis-à-vis those which *also* patent in the US. Predicted probabilities for each alternative are plotted against various covariates of interest with an indication of the general tendency.

A final set of results then moves on to analyse patenting productivity, `Q2`, 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.[20] 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. More recent applications have further concerned themselves with the issue of 'excess zeros' in count data:[21] in order to distinguish the 'innovating firms' which chose not to patent in year t from firms which never innovate and therefore never patent (labelled 'certain zeros') 'zero-inflated' versions of the Poisson (ZIP) and Negative Binomial (ZINB) estimators (Lambert, 1992; Winkelmann, 2003) first estimate a nonlinear model for 'certain zeros' and then analyse those observations which are predicted *not* to be 'certain zeros' in a count regression model.[22] In the present data context the large number of zero patents and thus concerns over 'excess zeros' and/or 'overdispersion' are arguably more important than in the standard approaches using US or OECD data: for the USPTO data we have a mere 68 non-zero observations (0.11 percent of the sample), whereas for the SIPO data the figure is 922 (1.43 percent). Given the uncertainty over

out, however, have to be constructed from the standard regression residuals.

[18] Following estimation we evaluate the model by investigating whether coefficients for each covariate are zero across all four alternatives and by analysing LR tests for the combination of alternatives. We also investigate percentage changes in the odds across different alternatives. Here we focus primarily on the 'patent in China' versus the 'patent in China *and* the US' alternatives. This multinomial logit analysis is conducted making extensive use of the routines and examples provided in Long and Freese (2006).

[19] Furthermore we test whether separate probit models fit the data better than the joint model and conduct parameter homogeneity testing across the two equations.

[20] 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, which allow for unobserved time-invariant heterogeneity. In the present case of China, however this 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).

[21] This phenomenon is intimately linked to overdispersion given that both may arise from unobserved heterogeneity (Cameron and Trivedi, 2006).

[22] Note 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.

any differences in the determinants of the patenting decision and the patenting productivity, we follow the standard in the literature by including the same covariates in both equations (Winkelmann, 2003).[23] 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 ('correlated NegBin').[24]

In the following we briefly discuss the choice of covariates employed in the regressions. The firm characteristics considered 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,[25] to avoid confounding the R&D effect with that of the size of the firm (Hall and Ziedonis, 2001), which is measured by employment 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 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 additional indicator for socialist period legacy. Ownership types include 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; our designation here is based on the proportion of paid-in capital in excess of 50 percent, 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. Finally, we add year dummies to all our models which will allow us to chart the changes in patenting over time, accounting for any unobserved common shocks. With the exception of dummy variables all of the above are in logarithms.

6. DESCRIPTIVE EVIDENCE

Tables 1 and 2 list the top 10 companies patenting with the USPTO and SIPO respectively. These tables are constructed using the entire integrated dataset (1985-2006) to provide as complete a picture as possible.

— Tables 1 and 2 about here —

Table 1 illustrates the strong concentration of USPTO patents in the hands of very few companies: the top 10 assignees account for over 85 percent of USPTO patents. Interestingly, three compa-

[23]We employ a number of diagnostic tools as suggested by Trivedi and Munkin (2010). We further consider a number of formal 'goodness-of-fit' tests and statistics (Greene, 1994), as well as various information criteria (see Long and Freese, 2006).

[24]This allows us to conduct parameter homogeneity tests across USPTO and SIPO models akin to the analysis in the bivariate probit model.

[25]We add a dummy variable for firms with no data on R&D expenditure (about 1.7 percent of observations) and a dummy for firms with zero R&D expenditure (about 70 percent of observations).

nies, Hongfujin (1), Fuzhun (3) and Futaihong (6), are subsidiaries of the Taiwanese-owned multinational Foxconn Technology Group, the world's largest contract manufacturer in 3C (Computer, Communication, Consumer electronics) products. These three subsidiaries account for over 40 percent of total USPTO patents in our matched dataset, adding in communications giant Huawei brings the tally to over 60 percent. As shown in the last column of Table 1, with the exception of Sinopec, China International Marine and BYD, all top 10 USPTO patentees are in 3C industries. Table 2 shows SIPO patent holdings, with the top 10 companies accounting for about 75 percent 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 subsidiary, 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 2 are in 3C industries. Note that there is a significant overlap of companies in Tables 1 and 2: six companies appear in both lists, with four of these in 3C industries.

Except for LG Shanghai and Inventec all companies listed in Tables 1 and 2 employ more than 3,000 workers, with Sinopec, Baoshan Iron & Steel and Huawei employing more than 100,000 workers. It is also not surprising that most of the firms turn out to be heavily (though not exclusively) engaged in exporting. According to the Chinese Ministry of Commerce all the companies in Tables 1 and 2 except for BYD, LG Shanghai and Fuzhun are among the top 100 Chinese exporters in 2006.[26]

In summary, 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 ICT equipment industry. In particular Taiwanese-owned contract manufacturer Foxconn emerges as a major player in patenting with the USPTO, whilst two indigenous firms, Huawei and ZTE, play leading roles in domestic patenting.

— Table 3 about here —

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 they protect: product or process innovation or a combination of the two.[27] In the case of USPTO patents, we read all 1,912 patent documents, whereas in the case of SIPO patents we rely on a random subsample of 980 out of over 45,000 patents.[28] Table 3 shows a breakdown of patents filed by Chinese residents according to the innovation type they protect. For USPTO patents nearly half cover product innovations and only about 20 percent cover process innovations. The pattern looks different in the case of SIPO patents: nearly 37 percent protect process innovations and slightly less than 30 percent product innovations. The share of patents protecting both product and process innovations remains approximately the same as for USPTO patents.[29] Thus our analysis suggests

[26]See http://zhs.mofcom.gov.cn/aa/aa.html for "China's top 200 exporters 2006".

[27]An 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.

[28]Claims are not readily available but must be retrieved from the original patent documents which are available only in Chinese.

[29]The last two columns show the distribution of innovation type when we add USPTO equivalents to the SIPO patents. This 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. We use sample weights in computing shares to reflect the small proportion of SIPO patents with USPTO equivalents (2.5 percent) among all SIPO patents. The results change very little.

that inventions that are patented in China but not in the US are more likely to protect process innovations. In contrast the results for USPTO indicate that the share of patents protecting product innovations is substantially higher.[30] These cover mostly mechanical innovations related to the ICT equipment industry embodying relatively modest innovative content. USPTO patents covering processes and the combination of process & product innovations appear to be of more innovative character.[31] When we examine the inventive content of SIPO patents, a similar picture emerges with patents covering processes of more innovative character.[32]

Tables A-1 and A-2 in the appendix contain information on the sample of firms used in our regression analysis. Due to the limited availability of R&D data, the sample collapses to about 64,500 firm-year observations for the years 2001, 2002, 2005 and 2006. Table A-1 provides some basic descriptive statistics regarding firms' patenting activities, indicating 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.

7. RESULTS

The present data poses serious challenges to the empirical estimators typically employed in patent decision and patenting productivity analysis, in particular for the analysis of USPTO patents, where only 0.11 percent of firm-years have non-zero observations. We rely in parts on the predictive power of the models and employ various tests and robustness checks.[33]

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 4 and from a bivariate probit model in Table 5, together with plots for predicted probabilities for the latter in Figure 1.

— Table 4 about here —

Table 4 shows multinomial logit estimates for patenting in China only (column [1]), in the US only ([2]) and in both countries ([3]) in comparison to the omitted alternative not to patent at all. The predicted probabilities for each of the three alternatives are close to the observed

[30]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.

[31]The overwhelming share of these patents protect ICT-related innovations, such as central process units or data processing/transmission methods.

[32]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.

[33]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).

probabilities while the tests for combination of any of the above categories/alternatives rejects in all cases at the 1 percent 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. Formal statistical testing 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.

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 regard to the 'SIPO only' and 'both' alternatives. The relationship seems much stronger for the group of firms that patent in both China and the US, as is also confirmed by the changes in the odds.[34] 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 (mean 32 percent, median 4 percent) implies a 75 percent increase in the odds to patent in both countries relative to in China alone. With regard 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 year dummies are insignificant for the remaining firms which do patent (most of which are contained in the 'both' group in [3]).

A first conclusion from this analysis is that differences in the R&D intensity, firm size, firm age and in particular export-intensity seem to be associated with firms' decisions to patent abroad vis-à-vis exclusively in China. Surprisingly, firm ownership as defined here does not have any discernable relationship with the patenting choice. Recent work on the extension of trade credit in China (Guariglia and Mateut, 2011) finds that differences across traditional ownership categories (e.g. private, state-owned) disappear once the firms' political connections are taken into account. In a similar vein, ownership type per se may not be relevant for firms' patenting behaviour.

— Table 5 about here —

Table 5 reports results for a bivariate probit regression, analysing the patenting decisions of firms in a joint empirical model. We discuss 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 and groups of variables (common symbols indicate groupings). We furthermore compute predicted probabilities and correlate these with some of the key covariates of interest, using fractional polynomial regression to highlight the central tendency (95 percent confidence intervals indicated): this analysis is presented in Figure 1, 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.[35]

— Figure 1 about here —

[34]In the right half of the table we present the percentage changes in the odds for a unit increase in the independent variable — we only present results for differences which are statistically significant at the 10 percent level. 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]).

[35]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 only 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 1 indicates, this is a minor difference in the shape of the innovation effort-patenting relationship between the two. For firm size 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 1 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 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 1: 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 percent export-intensity. The homogeneity test for firm age (in logs) rejects at the 1 percent 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 1 show some noise in the tails of the age distribution of firms, age seems to be positively related with patenting exclusively at SIPO, whereas the relationship appears 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.[36]

In summary, this analysis confirms that patenting with USPTO relative to SIPO in our regression sample is primarily associated with age and export-orientation of firms and 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 ownership type does not seem to impact the patenting choice.

7.2 Patent productivity

We now turn to the empirical analysis of patent productivity, `Q2`, which is analysed using count regression models.[37] For our discussion here we focus on results for the Negative Binomial model which received favourable diagnostics in comparison to rival approaches.[38] Table 6 details results for the USPTO and SIPO equations, which were estimated as seemingly unrelated regression

[36]In Tables TA-4 and TA-5 in the technical appendix we present descriptive statistics and details on sector of operation for each of the four patenting 'groups'. Table TA-5 lists the Top-10 sectors for each of the four groups, based on the share of overall SIPO patents (for 'China only' group) and/or USPTO patents (for 'US only' and 'China and US' groups).

[37]Detailed results for USPTO and SIPO patenting from a number of alternative estimation methods are presented in Tables TA-7 to TA-8 in the technical appendix, with predicted probabilities for all count data models detailed in Table TA-9.

[38]Table TA-10 presents results for the 'correlated' Zero-Inflated Negative Binomial (ZINB) regression model for reference.

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 for 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.

— Table 6 about here —

Parameter homogeneity tests suggest that the R&D effort is not associated differentially with patenting counts at USPTO and SIPO.[39] Firm size is positive significant in both equations, however the parameter homogeneity test suggests that size matters more for USPTO patent count — 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 (parameter homogeneity rejected at the 5 percent level). It is 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.[40] 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 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, this is primarily due to the large standard errors of the USPTO equation estimates, with homogeneity tests for individual years marginally statistically significant for 2002 and 2005.

In summary, our analysis of patent productivity has revealed very similar patterns to our earlier study of the patenting decision. In terms of our comparison of patenting in China and the US we found export-intensity, firm age and firm size of particular importance in distinguishing USPTO patent count from SIPO patent count: firms associated with increased USPTO patenting are larger, younger and more export-oriented than their peers associated with high SIPO patent.[41]

[39]Estimates for the squared log R&D per worker terms are insignificant for USPTO and significant for SIPO, pointing to a linear and convex relationship, respectively; however the formal homogeneity test rejects statistical difference between the two models' parameter estimates.

[40]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.

[41]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 rapidly 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 chart the developments from 1985-2006 and to investigate the factors associated with the Chinese patent explosion during the period 1999-2006.

Our answer to what lies behind the Chinese patent explosion is unambiguous: a handful of companies in the ICT sector account for the overwhelming share of patents, with concentration more pronounced in USPTO than SIPO 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 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 degree) 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 most likely not: patenting is concentrated in very few industries and even within these is 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,' some Chinese companies appear to be truly innovative, potentially even pushing the global technology frontier in certain niches. 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,[42] there are at present very few such companies driving this development.

[42]Mahmood and Singh (2003), for example, point to a similarly strong concentration of USPTO patents (1970-1999) among assignees in South Korea and Singapore as the top 50 assignees hold 85 percent and 70 percent of each country's USPTO patents, respectively.

ACKNOWLEDGEMENTS

We thank seminar/session participants at the 'UC Berkeley – Peking University Workshop on the Evolution of Chinese Patenting and its Implications for Research and Innovation' in Beijing, a workshop at Oxford, the 6th annual EPIP conference in Brussels and the RES IO Autumn School in Birmingham for their useful comments and suggestions. Lei Hou and Jiarui Zhang provided excellent assistance and we thank Max Ernicke for his advice on the identification of innovation types in patent claims. The usual disclaimers apply. 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.

APPENDIX

A. DATA CONSTRUCTION AND DESCRIPTIVE STATISTICS

Following the merging of ASIE and Oriana data we match the integrated dataset with PATSTAT data for SIPO innovation and USPTO utility patents. We drop firms that are only contained in Oriana. This results in 1,942 USPTO and 45,765 SIPO patents matched to firms for the 1985-2006 period. For the regression analysis ('R&D sample', constrained to the years 2001, 2002, 2005, 2006) we furthermore exclude those operating outside the manufacturing sector, which yields 64,652 firm-year observations from 19,956 firms. A more detailed discussion can be found in a technical appendix.

REFERENCES

[1] 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.

[2] Breznitz, Dan and Michael Murphree. 2011. *Run of the Red Queen*, Yale University Press.

[3] Cameron, Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press.

[4] 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.

[5] 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.

[6] 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.

[7] Gallini, Nancy. 2002. "The economics of patents: Lessons from recent U.S. patent reform." *Journal of Economic Perspectives*, Vol. 16(2), pp. 131-154.

[8] 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

[9] 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.

[10] Guariglia, Alessandra, Xiaoxuan Liu and Lina Song. 2011. "Internal Finance and Growth: Microeconometric Evidence on Chinese Firms." *Journal of Development Economics*, Vol. 96, No. 1, pp. 79-94.

[11] Guariglia, Alessandra and Simona Mateut. 2011. "Political affiliation and trade credit extension by Chinese firms." University of Nottingham, GEP Research Paper Series 2011/12.

[12] 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.

[13] 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.

[14] Hall, Bronwyn H. and Rose Marie Ziedonis 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.

[15] 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.

[16] 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.

[17] Hilbe, Joseph M. 2011. *Negative binomial regression*. 2nd edition, Cambridge University Press.

[18] Hu, Albert Guangzhou. 2010. "Propensity to patent, competition and China's foreign patenting surge." *Research Policy*, Vol. 39, pp. 985-993.

[19] Hu, Albert Guangzhou 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.

[20] Hu, Albert Guangzhou and Gary Jefferson. 2008. "Science and Technology in China." In: Brandt, Loren and Thomas Rawski (eds.) *China's Great Economic Transformation*. New York City: Cambridge University Press.

[21] Hu, Albert Guangzhou 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.

[22] Hu, Albert Guangzhou, Gary Jefferson, and Qian Jinchang. 2005. "R&D and Technology Transfer: Firm-Level Evidence From Chinese Industry." *The Review of Economics and Statistics*, Vol. 87, No. 4, pp. 780-786.

[23] Hu, Mei-Chih and John A. Mathews. 2008. "China's national innovative capacity." *Research Policy*, Vol. 37, No. 9, pp. 1465-1479.

[24] 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.

[25] King, Gary, and Langche Zeng. 2001a. "Logistic Regression in Rare Events Data." *Political Analysis*, Vol. 9, No. 2, pp. 137-163.

[26] King, Gary, and Langche Zeng. 2001b. "Explaining Rare Events in International Relations." *International Organization*, Vol. 55, No. 3, pp. 693-715.

[27] Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing." *Technometrics*, Vol. 34, No. 1, pp. 1-14.

[28] 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.

[29] 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.

[30] 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, No. 3, pp. 163-173.

[31] Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd edition, College Station, TX: Stata Press.

[32] Mahmood, Ishtiaq P. and Jasjit Singh. 2003. "Technological Dynamism in Asia." *Research Policy*, Vol. 32, No. 6, pp. 1031-1054.

[33] 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.

[34] Naughton, Barry. 2007. *The Chinese economy: transitions and growth*, Boston/MA: MIT Press.

[35] Park, Walter. 2008. "International patent protection: 1960-2005." *Research Policy*, Vol. 37, No. 4, pp. 761-766.

[36] 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, No. 1, pp. 64-76.

[37] 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.

[38] Subramanian, Arvind. 2011. "The Inevitable Superpower: Why China's Dominance Is a Sure Thing." *Foreign Affairs*, Vol. 90, No. 5, September/October.

[39] Sun, Yifei. 2000. "Spatial Distribution of Patents in China." *Regional Studies*, Vol. 34, No. 5, pp. 441-454.

[40] Trivedi, Pravin and Murat Munkin. 2010. "Recent Developments in Cross Section and Panel Count Models." In: A. Ullah and D. Giles (eds.) *Handbook of Empirical Economics and Finance*, Chapter 4, pp. 87-131.

[41] van Pottelsberghe de la Potterie, Bruno. 2010. "The Quality Factor in Patent Systems." ECARES Working Paper 2010-27.

[42] Winkelmann, Rainer. 2003. *Econometric Analysis of Count Data.* Heidelberg, 4th edition.

[43] WIPO [World Intellectual Property Organisation of the United Nations]. 2011. "International Patent Classification (IPC)" WIPO website accessed 31st May 2011 at http://www.wipo.int/classifications/ipc/en/.

[44] Yang, Deli. 2008. "Pendency and grant rations of invention patents: A comparative study of the US and China." *Research Policy*, Vol. 37, No. 6-7, pp. 1035-1046.

[45] 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.

[46] Yueh, Linda. 2009. "Patent Laws and Innovation in China." *International Review of Law and Economics*, Vol. 29, No. 4, pp. 304-313.

[47] Zhao, Minyuan. 2006. "Conducting R&D in Countries with Weak Intellectual Property Rights Protection." *Management Science*, Vol. 52, No. 8, pp. 1185-1199.

[48] 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.

TABLES AND FIGURES

Table 1: Top 10 Chinese companies filing with USPTO (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	Petroleum, Natural Gas Exploration (079)
5	Semiconductor Manufacturing Int.	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	100.0	

Notes: ♮ Chinese GB/T 3-digit industry code in brackets.

Table 2: Top 10 Chinese companies filing with SIPO (1985-2006)

Rank	Company	Patents	Share	Main Industry affiliation♮
1	Huawei Technology	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	Petroleum, 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	CCO (409)
9	Baoshan Iron & Steel	756	1.65	Ferrous metal smeltering and rolling (320)
10	Inventec Shanghai	711	1.55	CCO (409)
	Other	11,423	24.96	
	Total	45,765	100.00	

Notes: ♮ Chinese GB/T 3-digit industry code in brackets. CCO – Communications, computers & other electronic equip.

Table 3: Product vs. Process Innovation (1985-2006)

Innovation Type	USPTO		SIPO			
			excl. US Equivalents		incl. US Equivalents[♮]	
	Share	Patents	Share	Patents	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	100.00	1,912	100.00	980	100.00	2,130

Notes: Patents are classified manually using patent claims. ♮ Equivalents with the USPTO and SIPO patents are weighted with the respective sample share.

Table 4: R&D Sample – Multinomial Logit regression (2001/2, 2005/6)

	SIPO only	**USPTO only**	**both**	Ind‡ (p)	**Changes in the Odds†** percentage change for unit increase in x		
	[1]	[2]	[3]		[2]vs.[1]	[3]vs.[1]	[3]vs.[2]
log R&D pw	0.326	0.589	0.639	(0.00)		36.80	
	[0.027]***	[0.175]***	[0.098]***				
(log R&D pw)^2	0.019	-0.047	0.098	(0.00)		8.30	15.60
	[0.008]*	[0.062]	[0.016]***				
log Workers	0.588	0.627	1.567	(0.00)		166.30	156.00
	[0.048]***	[0.223]**	[0.161]***				
log Exp/Sales	-0.118	1.696	0.444	(0.00)	513.70	75.30	-71.40
	[0.025]***	[0.494]***	[0.157]**				
log Firm age	-0.061	-0.711	-0.593	(0.01)	-47.80	-41.30	
	[0.055]	[0.271]**	[0.234]*				
FIE (other)	0.041	-0.072	1.213	(0.22)		222.60	
	[0.166]	[1.147]	[0.450]**				
FIE (HMT)	-0.016	0.162	1.244	(0.28)		252.60	
	[0.182]	[0.983]	[0.572]*				
Private	0.073	-0.066	0.338	(0.83)			
	[0.122]	[0.889]	[0.454]				
Collective	-0.472	0.473	-0.785	(0.20)			
	[0.286]	[1.294]	[0.661]				
Other	0.143	0.818	1.099	(0.49)			
	[0.246]	[1.294]	[0.781]				
Zero R&D	-1.290	-1.686	-0.164	(0.00)		208.40	385.40
	[0.101]***	[0.671]*	[0.420]				
Zero Exports	0.171	-1.676	0.178	(0.00)	-84.20		538.40
	[0.110]	[0.949]	[0.525]				
2002	0.437	0.272	0.455	(0.18)			
	[0.144]**	[0.730]	[0.522]				
2005	1.234	0.449	0.704	(0.00)			
	[0.148]***	[0.714]	[0.522]				
2006	1.382	-0.153	0.428	(0.02)	-78.50	-61.50	
	[0.147]***	[0.745]	[0.545]				
Combine χ^2 ♮					35.7	51.6	320.7
(p-value)					(0.00)	(0.00)	(0.00)

Notes: $n = 64,652$ observations for $N = 19,956$ firms. The omitted category is no patents (98.53 percent predicted probability, 98.53 percent true probability). † We only report differences statistically significant at the 10 percent level. We report percentage changes in the odds of moving between alternatives for a unit increase (for dummy variables: discrete change) in the independent variable. Column heads indicate alternatives tested. ‡ Wald test for the effect of independent variables (p-values reported), which has the H_0 that all coefficients associated with a given variable are zero. ♮ 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). We omit the results for 'Missing R&D' and the intercept to save space.

Table 5: R&D Sample – Bivariate Probit regression (2001/2, 2005/6)

	[1]	[2]	Homogeneity Testing‡		
dep. var.	**USPTO**	**SIPO**	Individual		Joint
log R&D pw	0.190 [0.027]**	0.159 [0.011]**	(0.25)	×	(0.08)
(log R&D pw)^2	0.025 [0.007]**	0.014 [0.003]**	(0.07)	×	
log Workers	0.397 [0.041]**	0.271 [0.022]**	(0.00)		
log Exp/Sales	0.286 [0.063]**	-0.052 [0.011]**	(0.00)		
log Firm age	-0.213 [0.060]**	-0.040 [0.023]	(0.00)		
FIE (other)	0.204 [0.182]	0.055 [0.067]	(0.41)	⋄	(0.81)
FIE (HMT)	0.244 [0.168]	0.041 [0.074]	(0.21)	⋄	
Private	0.123 [0.136]	0.043 [0.051]	(0.56)	⋄	
Collective	0.071 [0.247]	-0.171 [0.111]	(0.33)	⋄	
Other	0.280 [0.217]	0.111 [0.104]	(0.44)	⋄	
Zero R&D	-0.303 [0.117]**	-0.479 [0.039]**	(0.13)		
Zero Exports	-0.231 [0.162]	0.092 [0.045]*	(0.04)		
2002	0.082 [0.128]	0.150 [0.053]**	(0.61)	▹	(0.01)
2005	0.211 [0.131]	0.463 [0.053]**	(0.06)	▹	
2006	0.092 [0.140]	0.525 [0.054]**	(0.00)	▹	
Obs	64,652	64,652			
Non-zero obs (percent)	0.11	1.43			
ρ (st.error)	0.733 (0.034)				
Wald ($\rho = 0$)	165.2 (0.00)				
LL	-4257				

Notes: $n = 64,652$ observations for $N = 19,956$ firms. *, ** indicate statistical significance at the 1 and 5 percent level respectively. ρ indicates the interrelatedness of the two probit models estimated jointly. ‡We test the null of parameter homogeneity across the two models with results presented in the column marked 'Individual'. Joint tests for groups of variables are indicated by the same symbol, results reported in the final column. Note that we do not report the results for 'Missing R&D' or the intercept to save space. 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).

Table 6: R&D Sample – Correlated NegBin estimation

	Negative Binomial Model				
	[1]	[2]	Homogeneity Testing		
	USPTO	**SIPO**	Individual		Joint
log R&D pw	0.730 [0.160]***	0.605 [0.046]***	0.60 (0.44)	×	0.60 (0.74)
(log R&D pw)^2	0.034 [0.038]	0.046 [0.018]***	0.08 (0.77)	×	
log Workers	1.906 [0.219]***	1.064 [0.160]***	15.04 (0.00)		
log Exp/Sales	1.019 [0.250]***	-0.144 [0.071]**	21.55 (0.00)		
log Firm age	-0.971 [0.235]***	-0.414 [0.137]***	4.79 (0.03)		
FIE (other)	1.988 [0.915]**	1.445 [0.427]***	0.32 (0.57)	⋄	34.32 (0.00)
FIE (HMT)	2.101 [0.753]***	1.160 [0.537]**	1.81 (0.18)	⋄	
Private	0.352 [0.671]	0.083 [0.263]	0.14 (0.70)	⋄	
Collective	-0.626 [1.011]	-0.174 [0.593]	0.15 (0.70)	⋄	
Other	3.671 [0.898]***	1.475 [0.610]**	9.56 (0.00)	⋄	
Zero R&D	-1.220 [0.430]***	-1.118 [0.261]***	0.04 (0.84)		
Zero Exports	0.331 [0.566]	0.672 [0.300]**	0.53 (0.47)		
2002	-0.341 [0.585]	0.616 [0.174]***	2.62 (0.11)	▹	3.74 (0.29)
2005	0.503 [0.645]	1.541 [0.246]***	2.46 (0.12)	▹	
2006	0.926 [0.672]	1.621 [0.229]***	1.05 (0.31)	▹	
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			
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 a 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 report p-values for the hypothesis test of parameter homogeneity across models. We also tested groups of variables (grouping indicated by the symbols) jointly. See Table TA-7 in the appendix for further details on estimation and diagnostics.

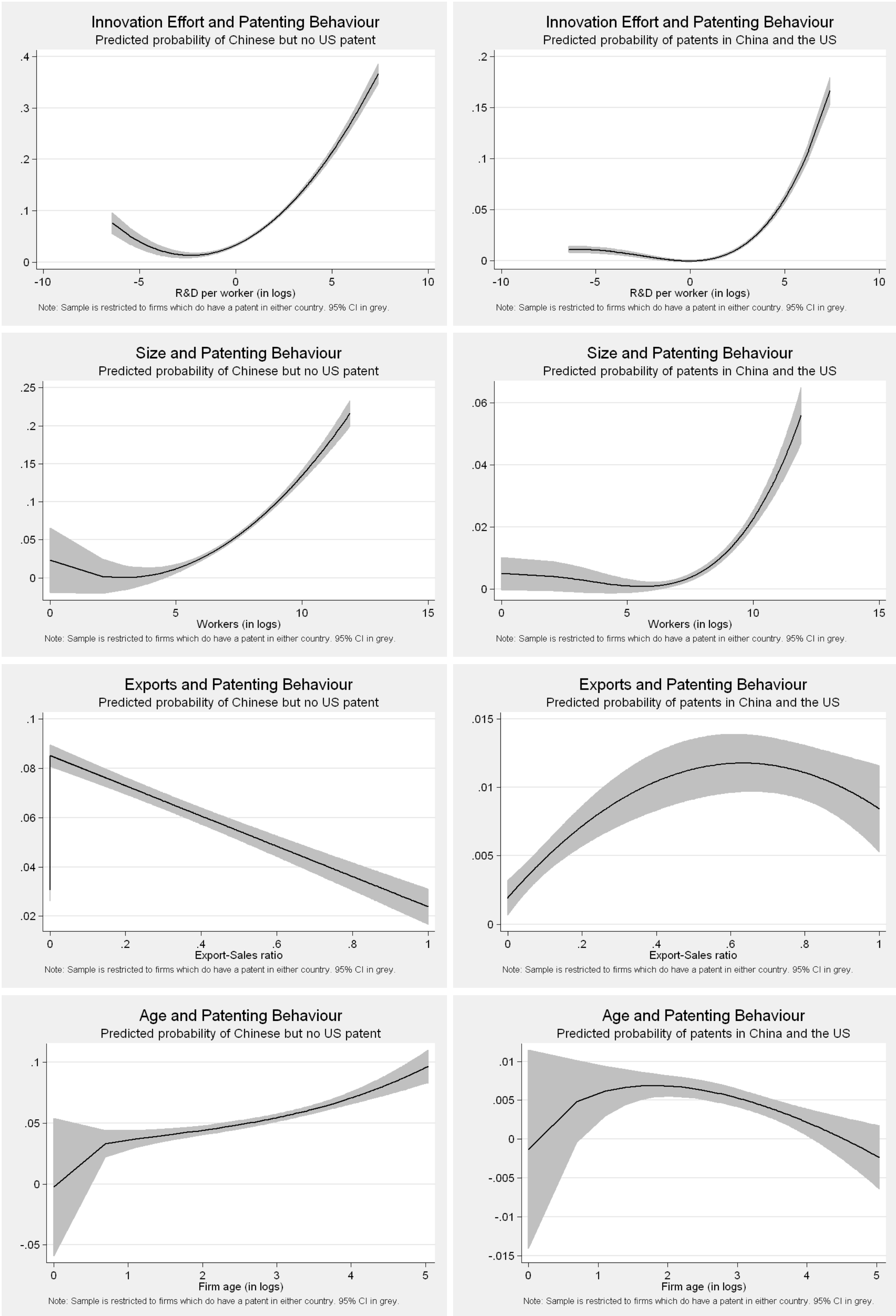

Figure 1: Predicted Patenting Probabilities

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 percent 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) — these amount to 1,937 observations from 849 firms (around 3 percent of all firms in our sample).

APPENDIX TABLES

Table A-1: R&D Sample – Chinese and US Patents

USPTO									
	All sectors		ICT equipment		All sectors				
year	**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									
	All sectors		ICT equipment		All sectors				
year	**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: The columns marked 'ICT equipment' refer to the subsample of 2,831 firms ($n = 9,552$ observations) in sectors 40 and 41 of the Chinese GB/T 2-digit industry code (Communications Equipment & Computers; Instruments & Office Machinery) which dominate the USPTO patent count.

Table A-2: R&D Sample – Descriptive Statistics

variable	type	obs	mean	median	sd	min	max
Patents							
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
Innovation effort							
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
Firm characteristics							
Export/Sales	continuous	64,576	0.319	0.040	0.406	0	1
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
Ownership type							
FIE (Other)	dummy	64,652	0.166	0		0	1
FIE (HMT)	dummy	64,652	0.153	0		0	1
Private	dummy	64,652	0.464	0		0	1
SOE	dummy	64,652	0.125	0		0	1
Collective	dummy	64,652	0.076	0		0	1
Other	dummy	64,652	0.024	0		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 percent 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).

TECHNICAL APPENDIX — Not Intended For Publication

A. DATA MATCHING

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 (with Chinese residency) respectively. The assignees contain a large range of different assignee types, including private individuals, police, military, universities, and public research institutes (e.g., the Chinese Academy of Sciences and other not-for-profit organisations). 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 TA-1 shows the resulting matching rates. In a third step, we define equivalent groups.[43] 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 step ensures consistency between the USPTO and SIPO matches and adds a number of patents to our matched sample (see Table TA-1). We successfully match 42 percent and 11 percent of all USPTO and SIPO patents filed between 1985 and 2006, respectively.

Table TA-1: Benchmarking the matching outcome (1985-2006)

	Assignee names					**Patents**			
	Raw Data♮	Cleaned Data§	Matched	Equivalents Corrected	Match Success (percent)	Patents	Matched	Equivalents Corrected	Match Success (percent)
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.

[43] We apply a definition that assigns patents into the same equivalent group if patents share the same priority documents.

B. DATA CLEANING

The merged ASIE-Oriana sample contains 143,458 firm-year observations from 23,915 individual firms spanning the period of 1999-2006 (see Table TA-2). There are 1,467 firms that are contained in Oriana but not in ASIE with most of these firms in non-manufacturing industries. 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 GB/T 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 (final two columns of Table TA-2). All variables employed in the regression analysis are discussed in detail in Section 5 in the main text. Note that for our descriptive analysis of patenting in Section 6, we make use of the entire data span for which we have patent data which covers the period 1985 to 2006.

Table TA-2: Oriana-ASIE dataset

	Oriana		**Oriana-ASIE**		**Share**	**R&D Sample**	
Year	observations	percent	observations	percent	percent	observations	percent
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

Notes: 'Share' indicates the number of observations in the integrated ASIE-Oriana data as a proportion of all observations in Oriana.

Finally, in order to verify potential sample selection issues due to the use of only those 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-3.

Table TA-3: Testing Sample Representativeness

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

Notes: Results presented in columns [1]-[3] represent the regression coefficients from pooled regressions of log sales on all covariates listed. In each case 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 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.

C. MORE DESCRIPTIVES

Table TA-4: R&D Sample – Firm Characteristics by patenting behaviour

	Firm Always Patents in China						Firm Only Patents in the US					
	obs	mean	median	sd	min	max	obs	mean	median	sd	min	max
Patents												
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
Innovation Effort												
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
Firm characteristics												
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
Ownership type												
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
Patents												
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
Innovation Effort												
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
Firm characteristics												
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
Ownership type												
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. 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 percent 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 percent are for the 19,238 firms that did not patent.

Table TA-5: R&D Sample – Sector distribution by patenting behaviour

Firm Always Patents in China

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		

Firm Always Patents in United States

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

Firm Never Patents

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					

Firm Patents in China and the United States

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 in percent with reference to these overall counts.

D. ADDITIONAL REGRESSION RESULTS

Table TA-6: R&D Sample – Rare Events Logit

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

Notes: $n = 64,652$ observations for $19,956$ firms (2001/2, 2005/6). *, ** indicate statistical significance at the 1 percent and 5 percent level respectively. We reprint the Bivariate (SUR) probit estimates from Table 5 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.

Table TA-7: R&D Sample – US Patent Production Functions

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

Notes: $n = 64,652$ observations for $19,956$ firms (2001, 2002, 2005, 2006). Dependent variable in all models is USPTO patent count. For ZIP and ZINB we also report the 'inflation' equation which predicts being in the 'always zero' group via logit regression. Standard errors (in brackets) clustered at the firm-level. 'Non-Zero Obs' indicates the percentage 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 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 (see Table TA-9. 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-8: R&D Sample – Chinese Patent Production Functions

	Poisson [1]	NegBin [2]	ZIP [3]		ZINB [4]	
			logit	Poisson	logit	Neg Bin
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]
Zero R&D	-0.610 [0.624]	-1.118 [0.261]***	1.235 [0.283]***	0.122 [0.632]	1.784 [0.231]***	0.785 [0.284]***
Zero Exports	-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 (p)				7.14 (.00)		9.89 (.00)
LR-Test $\theta = 0$ (p)		59,000 (.00)				26,754 (.00)
$\Sigma\lvert\hat{y}_i - y_i\rvert$	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 (2001/2, 2005/6). Dependent variable in all models is SIPO patent count. See Table TA-7 for further details on estimation and diagnostics.

Table TA-9: 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**	$\Sigma\lvert\hat{y}_i - y_i\rvert$
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	
$\lvert\hat{y}_i - y_i\rvert$	0.289	0.226	0.037	0.020	0.011	0.006	0.003	0.002	0.001	0.001	0.018	0.61
NegBin	99.895	0.062	0.015	0.007	0.004	0.003	0.002	0.001	0.001	0.001	0.010	
$\lvert\hat{y}_i - y_i\rvert$	0.005	0.002	0.005	0.007	0.004	0.003	0.002	0.001	0.001	0.001	0.030	0.06
ZIP	99.896	0.033	0.016	0.010	0.007	0.005	0.004	0.003	0.002	0.002	0.022	
$\lvert\hat{y}_i - y_i\rvert$	0.006	0.027	0.006	0.010	0.007	0.005	0.004	0.003	0.002	0.002	0.018	0.09
ZINB	99.898	0.045	0.015	0.008	0.005	0.004	0.003	0.002	0.002	0.001	0.018	
$\lvert\hat{y}_i - y_i\rvert$	0.008	0.015	0.005	0.008	0.005	0.004	0.003	0.002	0.002	0.001	0.022	0.07

Panel B: SIPO models

Patents	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**>9**	$\Sigma\lvert\hat{y}_i - y_i\rvert$
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	
$\lvert\hat{y}_i - y_i\rvert$	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	
$\lvert\hat{y}_i - y_i\rvert$	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	
$\lvert\hat{y}_i - y_i\rvert$	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	
$\lvert\hat{y}_i - y_i\rvert$	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 percent). The deviations from the observed frequencies (at the top of each panel) are reported in the lines marked $\lvert\hat{y}_i - y_i\rvert$, 'Sum' simply adds up these absolute deviations.

Table TA-10: R&D Sample – Correlated ZINB estimation

	Zero-Inflated Negative Binomial Model									
	logit		Homogeneity Testing			**NegBin**		Homogeneity Testing		
	USPTO	**SIPO**	Indiv.		Joint	**USPTO**	**SIPO**	Indiv.		Joint
log R&D pw	-0.634	-0.313	0.14	×	0.01	-0.085	0.329	0.36	×	0.00
	[0.216]***	[0.067]***				[0.451]	[0.062]***			
log R&D pw^2	-0.071	-0.003	0.20	×		0.019	0.079	0.64	×	
	[0.052]	[0.016]				[0.127]	[0.026]***			
log Workers	-0.510	-0.438	0.88			1.682	0.756	0.17		
	[0.499]	[0.067]***				[0.654]**	[0.089]***			
log Exp/Sales	-0.327	0.308	0.32			0.837	0.188	0.34		
	[0.642]	[0.073]***				[0.691]	[0.074]**			
log Firm age	-0.172	-0.097	0.92			-1.765	-0.386	0.26		
	[0.763]	[0.136]				[1.195]	[0.195]**			
FIE other	0.631	0.678	0.98	$\diamond$	0.20	2.659	1.998	0.84	$\diamond$	0.03
	[2.068]	[0.384]*				[3.320]	[0.705]***			
FIE HMT	1.228	0.585	0.78	$\diamond$		3.892	1.723	0.52	$\diamond$	
	[2.329]	[0.361]				[3.388]	[0.664]***			
Private	1.404	-0.100	0.40	$\diamond$		2.930	0.047	0.28	$\diamond$	
	[1.777]	[0.248]				[2.696]	[0.299]			
Collective	1.708	1.064	0.83	$\diamond$		2.916	0.965	0.71	$\diamond$	
	[3.021]	[0.418]**				[5.173]	[0.447]**			
Other	1.466	-0.286	0.51	$\diamond$		5.057	0.386	0.33	$\diamond$	
	[2.625]	[0.701]				[4.871]	[0.749]			
Zero R&D	-0.557	1.784	0.06			-2.465	0.785	0.12		
	[1.233]	[0.231]***				[2.079]	[0.284]***			
Zero Exports	1.579	-0.499	0.19			2.361	-0.307	0.27		
	[1.594]	[0.224]**				[2.397]	[0.297]			
2002	-0.908	-0.401	0.60	$\triangleright$	0.07	-0.991	0.062	0.47	$\triangleright$	0.27
	[0.926]	[0.264]				[1.418]	[0.300]			
2005	-0.814	-1.138	0.75	$\triangleright$		-0.310	0.489	0.57	$\triangleright$	
	[0.967]	[0.253]***				[1.353]	[0.321]			
2006	0.752	-1.223	0.17	$\triangleright$		2.154	0.714	0.55	$\triangleright$	
	[1.383]	[0.264]***				[2.401]	[0.340]**			
Constant	7.537	5.666				-11.884	-6.249			
	[4.449]*	[0.899]***				[4.837]**	[1.294]***			
lnθ						1.705	2.376			
						[1.287]	[0.384]***			
Non-Zero Obs	0.11	1.43				0.11	1.43			
LL						-511	-5855			
Vuong (p)						4.36 (0.00)	9.89 (0.00)			
LR-Test $\theta = 0$ (p)						377.31 (0.00)	26,754 (0.00)			
$\Sigma\lvert\hat{y}_i - y_i\rvert$						0.074	0.380			

Notes: $n = 64,652$ observations for $19,956$ firms (2001/2, 2005/6). 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. Note that we do not report the results for the 'R&D missing' dummy.