The Impact of Chinese Innovation Competition on U.S. Firms

Gerard Hoberg, Yuan Li, and Gordon M. Phillips*

November 4, 2018

Abstract

We propose that competitive shocks from China impact U.S. innovation through two margins: the markets for innovation and for existing products. Using Chinese data, we map each industry to province Internet penetration levels using geographic agglomeration data. The resulting industry-year database indicates the ability of Chinese firms to acquire knowledge globally and compete in the market for intellectual property production. Increases in provincial Chinese Internet penetration are followed by sharp reductions in R&D investment and subsequent patents for U.S. firms, and increased patenting by Chinese firms. The new Chinese patents also cite the patents of treated U.S. firms at a high rate, consistent with increased intellectual property competition. In contrast, U.S. firms with fewer growth options and more tangible assets tend to increase R&D and patenting activity. Overall, both competition in intellectual property by Chinese firms and the asset competition of U.S. firms influence U.S. firm innovation.

Keywords: Innovation, competition, China, investment, internet penetration.

JEL Codes: O31, O34, D43, F13

*University of Southern California Marshall School of Business, University of Southern California Marshall School of Business, and Tuck School of Business at Dartmouth College of Business and National Bureau of Economic Research, respectively. Hoberg can be reached at hoberg@marshall.usc.edu, Li can be reached at yuan.li.2019@marshall.usc.edu, and Phillips can be reached at gordon.m.phillips@tuck.dartmouth.edu. All errors are the authors' alone.

> China now has the wealth, commercial sophistication and technical expertise to make its pursuit of technological leadership work. The fundamental issue for the U.S. and other western nations, and the IT sector is how to respond ...
>
> *Office of the United States Trade Representative, March 28, 2018 report*

1 Introduction

A growing body of research focuses on the impact of China's meteoric rise as an economic power and its impact on the innovation spending by established firms in the United States. This growing body of research has been matched by a growing interest in this same issue by policy makers, politicians and the popular press. Issues at stake include job loss, the incentives to innovate, and intellectual property protections. Yet the existing literature disagrees even on the most basic question. Does an increase in foreign competition have a positive or negative impact on the intensity of innovative investment in the U.S?

On the surface, the answer might seem obvious: increased competition is a negative shock and afflicted firms should reduce investment in R&D if this competition is in the form of strategic substitutes, as is true in many markets. Yet this prediction is not a given even if firms compete under strategic substitutes. For example, Aghion, Bloom, Blundell, Griffith, and Howitt (2005) suggest that firms might increase R&D following increased competition as this might facilitate "escaping competition" through increased product differentiation. Bloom, Draca, and Van Reenen (2016) further predict that when firms have "trapped assets" that are difficult to redeploy, or high adjustment costs, that these incentives to increase innovative spending increase further. In particular, these firms will maintain high ex ante production levels despite lower prices as curtailing production is too costly. The increased innovative spending then restores some pricing power through differentiation. It has thus become an empirical question whether increased competition leads to increases or decreases in innovation spending.

The existing empirical evidence is also mixed. Autor, Dorn, Hanson, Pisano, and Shu (2018) find a negative relation between competition shocks measured using trade data and

R&D in the U.S. However, Bloom, Draca, and Van Reenen (2016) finds that competition shocks (measured using trade data) lead to increased R&D spending in a sample of European firms. Hombert and Matray (2018) also examine U.S. firms, and find that firms that are ex ante R&D intensive experience more positive outcomes due to their increased ability to use R&D to escape competition. We consider a new approach to this question that relies on an expanded micro-foundation.

We propose that global competition influences innovation through at least two competitive margins, each having different implications for innovation spending in the U.S. The first is the margin covered by the existing studies: direct import competition in the market for existing products. These existing studies use tariffs and import data, reinforcing their focus on the margin of existing products. The second margin, which has not been studied in the U.S.-China innovation literature, is direct competition in the market for innovation and intellectual property itself. Importantly, shocks to tariffs and imports cannot be used as direct shocks to this margin, as both relate to products that already exist, and hence their impact on intellectual property (IP) competition would only be indirect and observed with delay.

We study the impact of Chinese innovation and its competitive impact on U.S. innovation using direct measures of Chinese ability to access innovation in the U.S. over the internet. Traditional instruments such as tariffs and direct imports apply to existing product competition, and not competition in the race to create new technologies. We propose that industry agglomeration and internet penetration at the province level in China can be used to generate plausible exogenous variation in the capacity of Chinese firms to challenge U.S. firm innovation in particular industries. First and foremost, intellectual property itself is a form of information, and the internet has proven to be an efficient means for accumulating knowledge, especially when the knowledge to be gathered resides overseas and is in electronic form.

In our main analysis, we examine how U.S. firms change their innovative investment in the face of plausibly exogenous changes in intellectual property (IP) competition from China. We find that treated U.S. firms significantly reduce spending in R&D over a lengthy three-year period after treatment. These same firms realize fewer patents over the same horizon, and at the same time, there is a material increase in Chinese patents in these same intellectual property production markets. In particular, there is a strong increase in new patents by Chinese inventors that directly cite the existing technology of the treated U.S. firms. This crowding-out effect is unique to China, as we see no analogous impact of complaints about IP competition from other major international competitors that are related to Chinese Internet penetration.

Competition in the market for intellectual property likely has a strong industry-specific component. We use industry production locations, relying on the agglomeration literature, at the province level in China to identify geographic regions where the most skilled and specialized human capital exists in China for a given industry. We then build industry-specific measures of Chinese internet penetration by mapping province level data on internet penetration to the primary industry locations in each province. Because internet penetration in different geographic regions depends on the ability of unrelated utility companies (internet service providers) to provide digital infrastructure, variation in this Internet penetration is plausibly exogenous. Intuitively, the provision of high quality internet depends in part on the distribution of population in that region, geographic features, and the relative efficiency of ISPs in different regions. Province level penetration thus varies substantially across provinces and over time.[1] This framework allows us to create an industry-year panel of instruments for China's capacity to access innovation information that can plausibly challenge U.S. firms. In turn, this panel data approach allows us adequate power and variation to test our key hypotheses even in the presence of rigid firm and year fixed effects.

Although we are careful to note limitations in our ability to fully establish causal conclusions, we conduct a number of tests aimed at challenging the validity of the exclusion

[1] Roberts and Whited (2013) suggest that variation along geographic dimensions has good properties regarding identification.

requirement. First, we find that our industry-year measures of Chinese internet penetration predict higher ex post incidence of U.S. firms complaining about competition from China, specifically complaints about Chinese competition related directly to technology and intellectual property in their 10-K documents filed with the SEC. Finding increased complaints about Chinese access to their technological and intellectual property indicates successful identification of the second competitive margin of innovation noted above. Moreover, placebo tests indicate no evidence of similar complaints about competition in other regions of the world including Japan, Europe, and neighboring countries such as Canada and Mexico. This test is a powerful placebo as complaints about competition from these other regions are more common unconditionally than are complaints about competition from China.

We thus expect significant increases in competition in the market for intellectual property coming directly from Chinese firms but we should not see increases in competition coming from firms in other parts of the world. We find that our internet penetration measure strongly predicts higher rates of patent citations by Chinese inventors citing the patents of the treated U.S. firms in our sample. We observe no changes in citation rates by inventors from the other regions of the world. Finally, we also find higher rates patents applied for in China itself that cite these same U.S. firm patents.

The strong results specifically for China (and not elsewhere) illustrate the mechanism driving intellectual property competition and indicate that omitted economic state variables, such as worldwide industry supply or demand factors, likely cannot explain our results. Our framework, which includes region, firm and time fixed effects, also ensures that identification is coming from specific Chinese provinces (mapped using industry agglomeration), and not from changes in China that are nation-wide in scope. These findings support the validity of the exclusion requirement as our instrument only measures shocks to innovative potential in China itself, and we observe strong impact on the specific U.S. firms that should be impacted.

Our hypothesis regarding the competitive margin of IP production predicts that our results should be stronger in specific subsamples. We first examine whether our results are stronger in industries with stronger growth options as measured by the market to book ratios of the treated U.S. firms. As predicted, we find that firms with above-median market to book ratios experience more extreme ex post reductions in innovative investment and patents following competitive IP shocks from China.

A second prediction is that trapped assets will moderate these findings, as hypothesized by Bloom, Draca, and Van Reenen (2016). U.S. firms with more trapped assets have incentives to maintain high levels of innovation given high adjustment costs and hence they should reduce innovation less when these competitive shocks materialize. We use the asset tangibility of U.S. firms as our measure of trapped assets and find that firms with more tangible assets and thus more "trapped" assets do increase their relative R&D spending and patents in the face of increased competition to differentiate their existing products. However, consistent with the literature, we view this as driven by competition in existing products rather than competition in intellectual property production.

Our findings regarding growth options and trapped assets provide deeper insights on the importance of an industry's initial conditions, and how they shape the predictions regarding the impact of increased IP competition. These competing forces can help to explain much of the disagreement in the existing empirical literature, where both positive and negative competitive effects on innovation have been found. Key to our conclusion is that at least two margins of competition need to be separately explored. We find that direct competition in the market for intellectual property itself has a sharp negative impact on treated firms due to the intuitive crowding-out effect.

In contrast, if competition only increases in the market for existing products rather than in the market for IP production, it is more plausible that treated U.S. firms might increase innovation in order to escape competition. Such a strategy might be most optimal when, in fact, the Chinese competitors do not have the innovative capacity to compete on this second margin. For example, in such a market, ceding market share in the lowest quality existing products to the entrants, while increasing innovation in order to claim higher quality segments of the market for the incumbents can form the basis for the post-shock equilibrium. This approach can restore some pricing power for incumbents, while accommodating the entering rivals in the market where their competitive advantage of lower cost labor might be most advantageous.

Although our focus is on competitive intensity in the market for innovation, it is natural to ask if our results inform the more controversial issue of intellectual property theft. We believe that our study does not directly address IP theft, although it helps to motivate future research on the topic. A starting point is that IP theft and fair competition should have similar impact on treated U.S. firms. Both will crowd-out innovative spending as the foreign entrants claim a fraction of the rents for themselves. On the surface, the increase in patents we find suggests that IP theft is less likely, as the foreign innovators are securing legally defensible patent protection. However, this alone does not rule out IP theft as the basis for creating the new patents might have roots in stolen trade secrets or other intellectual property as a precursor.

In order to at least partly inform whether our results relate to IP theft, we examine the extent to which U.S. firms complain directly about IP theft in their 10-Ks. First, we note that power is relatively low in such tests given that such direct statements are less prevalent than are comments directly referencing competition. Nevertheless, we find suggestive evidence that our internet penetration instrument predicts a higher incidence of complaints about IP theft by the treated U.S. firms. This evidence suggests that IP theft, or "perceived IP theft", might explain part of the increased competition in these IP markets. Yet we caution readers not to draw strong conclusions from this analysis because power is limited and statements by firms about IP theft do not constitute direct proof that IP theft has in fact occurred. The underlying question of potential IP theft is important for future research to consider, as policy implications differ for IP theft versus fair competition.

2 Literature and Hypotheses

2.1 Existing Literature

We begin by noting that the literatures on innovation, competition and investment are individually massive and we can't discuss all work in these areas. We thus focus our discussion on papers that are more relevant to US-China competition and how it relates to innovation spending and to papers that specifically cover competition in innovative markets.

2.2 Hypotheses

This study examines product market globalization, and the competitive impact of foreign product market competition on innovation outcomes in a domestic market. Empirically, we examine China's impact on the U.S., but we note that the hypotheses we propose here are more general.

The existing global innovation literature typically focuses on the impact of shocks to foreign competition in the market for existing products. We propose that foreign competition plays out on more than one competitive margin and that foreign competitors can challenge domestic firms both on pricing existing products, and also by entering the competitive race for innovation.

The concept of competition in the market for innovation in the domestic U.S. market has been extensively studied by many authors.[2] In an international context, Hombert and Matray (2018), Bloom, Draca, and Van Reenen (2016), and Autor, Dorn, Hanson, Pisano, and Shu (2018) study the impact of competition from international trade on innovation. However, no study to our knowledge has examined the impact of product market globalization on the dual margins of competition in the existing product market and in the market for innovation.

[2]Early work on innovation and competition has been summarized in the survey by Reinganum (1989) with recent contributions by Phillips and Zhdanov (2013) and Bena and Li (2014).

Globalization of product markets results in the opening of borders and the impact on any nation can be modeled using theories of entry in markets with existing incumbents. In classical models of competition with strategic substitutes, such as the Cournot model, the central prediction is that an entrant will cause existing firms to downsize as the new competitor absorbs a fraction of the market share and applies upward pressure on quantities produced and downward pressure on prices. If the value of growth options in such a market is proportional to the scale of the firm, a natural follow-on prediction regarding innovation (our setting) is that such competitive shocks will also lead to reductions in ex-post innovation by incumbents as they reduce scale.

More recent research has challenged this classical view. Aghion, Bloom, Blundell, Griffith, and Howitt (2005) suggest that a shock to competition could result in increases in innovation as firms rush to differentiate their products in order to rebuild lost market power. This is the "escape competition" hypothesis. The validity of this alternative depends at least in part on incumbent firms having a technological advantage relative to the new entrants, as only then would they additionally be able to defend their differentiated products from the entrants.

The classical theory and the escape competition theory have opposite predictions. Hence it is not surprising that existing studies find mixed evidence regarding the impact of Chinese competition on the innovation intensity of domestic firms. These studies, however, only examine one competitive margin: competition in the market for existing products. Indeed, on this margin, it is quite plausible that the ideal conditions for the escape competition strategy might hold in some markets.

How do these predictions change if the entrants are also adept at producing innovation? Examining this issue is our main contribution. We propose that the overall effect on a domestic incumbent's innovation spending has two parts: (1) increased competition from the foreign rivals in the market for existing products and (2) increased competition from the foreign rivals in the market for innovation itself. The existing literature illustrates the ambiguous predictions regarding the former, whereas it is largely silent on the competition in intellectual property.

Our first hypothesis relates to the margin of competition for innovation, where we predict that increased competition from entrants on this same margin should crowd-out domestic firm innovation.

Hypothesis H1: Increased foreign competition will reduce the value of growth options and reduce incumbent domestic firm innovation spending in R&D and patenting. We also expect more patenting by the entering foreign firms, especially in technologies strongly related to the incumbent domestic firm's technologies.

Because H1 pertains to an increase in competition on the same margin that we are trying to predict (innovation), H1 intuitively predicts that the classic model's predictions of crowding out should dominate. In contrast, the scenario is more complex for the second margin: competition in the market for existing products with two potential competing forces.

Hypothesis H2a: Increased foreign competition in existing product markets leads domestic incumbents to downsize. We thus predict decreased innovation spending by these incumbent domestic firms.

Hypothesis H2b: Increased foreign competition in existing product markets leads to reduced prices for the existing products. To recapture pricing power, incumbent domestic firms will increase innovation spending in order to escape competition.

Because predictions regarding the impact of innovation in the market for existing products are ambiguous, it is natural to ask if H2a and H2b is more likely under different sets of initial conditions. We follow Bloom, Draca, and Van Reenen (2016) and propose that the existence of trapped assets by the domestic incumbents favors H2b. In particular, if a firm has assets that are not redeployable and adjustment costs are high, it follows that the firm has strong incentives to maintain high production levels. By increasing innovation, such a firm can preserve some pricing power despite its high production rate. This leads to our final hypothesis.

Hypothesis H3: When the domestic incumbent firms have high levels of trapped assets, then increased competition on either margin will favor higher innovation spending, all else equal.

3 Data and Methods

3.1 Sample Selection and Panel Structure

Our sample begins with the universe of Compustat firm-years with available 10-K filings on the EDGAR system. We exclude financial firms and regulated utilities (SIC 6000 - 6999 and 4900 - 4949, respectively) and limit the sample to firm-years with sales and assets of at least $1 million. Since the Chinese internet penetration measures do not exhibit enough industry-province coverage until 2000, our final sample starts from 2001 and ends in 2016, with 61,930 firm-years from 8,474 unique firms. This panel is the base for our analyses.

We construct a set of country-specific competition complaint measures using texts in 10-K filings. For convenience, we utilize the software from meta Heuristica LLC to process our queries. To measure complaints about competition from China, we search for paragraphs that contain at least one word from both the country name list ("China" or "Chinese") and the competition word list ("compete" or "competition" or "competing"). We then use the number of matched paragraphs and normalize it by the total number of paragraphs in the 10-K document as our measure, CNComp. In addition to this generic competition measure, we further construct three additional competition measures by requiring the paragraph to contain a word from a third word list. First, to measure the intensity of competition, we construct the high competition measure, CNCompHi, by requiring the paragraph additionally contains one of the words in (high OR intense OR significant OR face OR faces OR substantial OR significant OR continued OR vigorous OR strong OR aggressive OR fierce OR stiff OR extensive OR severe). Second, we measure the competition in intellectual property, CNIntComp, by requiring the paragraph to additionally contain both ”intellectual” and ”property” in the search. Finally, we measure complaints about intellectual property theft, CNIntTheft, by counting the number of paragraphs that match the country list, contain both ”intellectual” and ”property”, and match one of the words in (protect* OR infringe* OR theft*). In addition to constructing ratio measures of the total number of paragraphs, we also construct dummy variables which equal to one if we hit any matching paragraphs. Similarly, we also construct these measures for three other major economies in the world, namely Europe, North America (Canada and Mexico), and Japan, by changing the words in the country list. Details of these measures can be found in Table 11.

Other firm characteristics variables come from Compustat. We measure firms' R&D intensities by normalizing the R&D expenses (`xrd`) by sales. Following the suggestions from Koh and Reeb (2015), we replace missing R&D intensities by the industry average (2-digit SIC) if the firm has applied for any patents in the past three years, and replace other missing values with 0. Definitions of other variables can be found in Table 11. Finally, we winsorize all ratio variables at the 1% and the 99% level to control for outliers.

3.2 Patent Data

We generate our patent measures from two sources. The first source is Google Patent. Since Oct. 31, 2017, Google, in collaboration with IFI Claims, a global patent research company, has made a set of structured and queryable datasets of patents available to the public[3]. The core part of the datasets contain over 90 million patent publications from the patent offices of 18 countries, including both US and China, among others. The same datasets

[3]More about this announcement at https://cloud.google.com/blog/products/gcp/google-patents-public-datasets-connecting-public-paid-and-private-patent-data. One can access the datasets through Google's BigQuery service

support the searches made through patents.google.com, and to our knowledge represent one of the highest-quality sources for patent research. We also get the patent data from Kogan, Papanikolaou, Seru, and Stoffman (2016) (KPSS hereafter), who kindly shared the data on their website. The key advantage of the KPSS data is that the aforementioned authors have spent huge efforts to link the patents to US public firms. However, the data ends in 2010, thus we will combine the two data sources to generate our patent variables.

We first use patent applications to measure firms' innovation activities. We extend the KPSS data with Google patent data. To link the new Google patent data to public firms, we utilize the links that are already developed by KPSS. First, we take the overlapping part of the Google data and the KPSS data[4] and generate links between permno numbers (from KPSS data) and (first) assignee names (from Google data). Next, we select all the utility patents that are filed in USPTO and granted after Nov. 1, 2010 from Google data. We then merge the permno number to the first assignee of patents using the link file we just generated. In this step we are able to match 77.4% of all the new patents.

Google data also provides the country information of the assignee[5]. Thus we are able to see patents that are assigned to foreign entities but filed in USPTO. We utilize the information by measuring the number of new Chinese patents that cite the existing patents of US firms, providing direct evidence on the intensity of learning from Chinese firms. We also construct similar measures for other major economies, namely Japan, Europe, and Canada and Mexico. These measures would later be used as placebo tests to show that our internet penetration variable is not picking up omitted factors that attract general international competition.

Finally, Google data also includes all the patents filed in China' Patent Office, known as SIPO (State Intellectual Property Office of the Peoples Republic of China). Therefore we are also able to check whether patents filed by SIPO (by Chinese firms) also cite patents

[4]The Google Patent Data covers 99.95% of the patents in the KPSS data matched by the patent number, and covers 99.59% of patents matched by both the patent number and the grant date.

[5]The corresponding variable is `assignee_harmonized.country_code` in the dataset.

from US firms, further enhancing our previous measure using only the patents filed in the US.

3.3 Internet Penetration

The quality and coverage of internet access in China has dramatically changed in the last two decades. While in the early 2000s, only fewer than 1% of the population in China have access to the internet, by 2018, the number of internet users in China has surpassed 800 million, and the internet penetration rate reaches 57.7%. Internet has become the most important medium through which information exchanges. For innovation activities, internet enables inventors to collect information much more efficiently, and is almost a necessary component for any modern day research.

To measure the internet penetration rate in China, we hand collect the number of internet users from the reports issued by China Internet Network Information Center (CNNIC). CNNIC is the official administrator of the internet infrastructure in China, and starting from 1998, it publishes semi-annual reports which describe the recent development of internet infrastructure and the demographic of internet users in China. To our advantage, these reports also provide the number of internet users separately for each province in China[6]. We then collect the number of population for each province from China Data Online[7] and compute the internet penetration ratio for each province in each year.

Note that the internet infrastructure has not grown at similar rates for all the provinces in each year. As one example, Figure 2 plots the year in which each province experienced its largest increase of the internet penetration ratio. The scattering pattern shows that the development of internet is not always in sync for the whole nation. The landscape of telecommunication industry in China has gone through drastic changes in the past two decades. Prior to 1994, China has one government department, the Directorate General

[6]The statistics does not include data for Hongkong or Macau.

[7]Unfortunately, the China Data Center at the University of Michigan has decided to terminate the service from September, 2018. However, one can easily download similar data from alternative sources like http://data.stats.gov.cn/english/

of Telecommunications which is later registered as China Telecom, that provides all the phone and internet services. The monopolistic structure was changed in 1994 when China introduced China Unicom to compete with China Telecom. The deregulation continues in the 1990s as China Telecom was further broken up into two companies, and other new internet service providers like China Net and China Railnet were also established. By the end of 2001, China has seven companies in the telecommunication industry, and these companies tend to focus in different business areas and also different regions. For example, China Net, an internet service provider, mostly operates in the 10 provinces in the northern part of China. The drastic changes continue in the 2000s as the industry went through a round of complicated consolidation, and by the end of 2008, only three companies, each of which now cover all the telecommunication business, are left, namely China Telecom, China Mobile, and China Unicom. These industry changes could generate a direct impact on the internet services. For example, we see from Figure 2 that after China Net was acquired by China Unicom in 2008, three northern provinces, Liaoning, Shandong, and Jilin, experienced their largest increase in the internet penetration rate in 2009.

For each US firm, we want to measure the internet penetration for the potential peer firms in China. To do that, we use a weighted-average measure of the internet penetration of the provinces where the industry of the US firm is important. Indeed, a large literature has documented that industry tend to cluster geographically[8], and China is no exception. Ideally we would want the total assets of all the firms in each industry and province. However, such detailed census data is not publically available, we thus retreat to the second best using data from Chinese public firms. To help address the endogeneity of the industry-province links, we choose to use the industry- in year 2000. This choice is justified by our observation that the number of industries that the Chinese public firms span over becomes sufficiently high and stable in year 2000, as shown in Figure 3. We select all the Chinese public firms that have non-missing headquarter and asset information in 2000. Our final sample includes 864

[8]See Florence (1948); Hoover (1948); Fuchs (1962); Krugman (1993); Ellison and Glaeser (1997); Duranton and Overman (2005, 2008)

firms listed in mainland China (A-share), 74 firms listed in Hongkong, and 5 firms listed in the US[9]. We then assign each firm to the province of its headquarter. To generate the weights, for each 2-digit SIC industry, we first calculate the weights of each province using the total assets of all its public firms in that industry. Then we exclude provinces whose weights are below 10%, and finally recalculate the weights using the remaining provinces. Figure 4 shows the weight loading for all the industry-province pairs.

Using the weights for each industry, we finally calculate the internet penetration measure as the weighted average across all provinces. In the next section, we show that our internet penetration measure significantly predicts the complaints from US firms about competition in intellectual properties. We also find the measure will positively predict the number of Chinese patents that cite the US firms' patents. As placebo tests, we find the internet measure does not predict the complaints about the competition and patent citations from other economies, suggesting our internet measure in not capturing the endogenous factors which affect the overall level of international competition.

4 Summary Statistics and Validation

4.1 Summary Statistics

Table 1 presents summary statistics for our 2001 to 2016 panel of 61,930 firm-year observations having machine-readable 10-K filings. On average, the weighted internet penetration ratio is 36% for each firm-year. We see that about 5% of sample firms explicitly complain about competition from China, and 40% of them specifically mention intellectual property in their complaints. The incidence of U.S. firms complaining about competition, and especially competition in the market for intellectual property also has been rising. Figure 1 plots the time-series of the general Chinese competition complaint measure and the complaint measure about IP competition. Both measures show tremendous increases over the years.

[9]For firms that are dual-listed, we only count it once using its primary exchange

Table 1 not only indicates we have ample power to examine the impact of Chinese innovative capacity on U.S. firms, it also indicates that we have even more power to run placebo tests. For example, sample-wide, U.S. firms complain about Euoropean and North American (Canada and Mexico) competition at even higher rates. As shown in Table 1, the Chinese competition (scaled by document size and x 1000) variable averages 0.15, whereas the analogous variable for Europe is 0.31 and it is 0.24 for North America. Because we use activity in other parts of the world as Placebo tests, this indicates that there is ample power to detect deviations from the exclusion requirement using these other regions of the world as placebos. However, this variable is just 0.04 for Japan, indicating less relative power for Japan in this capacity.

Power to use other regions of the world as placebo tests is even more impressive when we consider patent citation activity. For example, the average intensity of Chinese firms citing US patents is 2.39, which compares to 26.85, 23.88, and 5.06 for Europe, Japan, and North America respectively. In particular, we find that patenting by Chinese firms that cites the patents of U.S. firms increases strongly with our measure of Internet penetration. Our identifying assumption is that Chinese Internet penetration is first-order driven by the capacity of unrelated IP providers in China and their capacity to expand. If so, our main results should not be driven by underlying state variables such as time-varying industry demand shocks.

Because demand shocks have a global component to them, it follows that if our identifying assumptions are violated, that our Chinese Internet penetration variable should also predict growth in European, Japanese, and North American firms citing the same U.S. firms. Hence we use these regional activities as placebo tests. Because the data is much richer for these regions than it is for China, it follows that these placebo tests should be particularly strong in terms of the power to detect violations of the exclusion requirement. As we document later, we find strong results for Chinese companies and no results for placebo tests using the other regions of the world.

Table 2 displays summary statistics at the firm level rather than at the firm-year panel level (Table 1). In particular, we first calculate the mean value of each variable for each firm, and the table represents the statistics for the resulting firm averages. The primary motive for reporting summary statistics in both dimensions is to examine the the distributions of our key variables, especially the more extreme values. As we will include firm and year fixed effects, for example, major outliers could sway our findings.

As is well-known in the innovation literature, many variables measuring R&D and patenting activity do have distributions that tend to be right-skewed. Consistent with the literature, we therefore winsorize all of our key variables at the 1%/99% level. Overall, we find distributions that are similar to those in other studies. Most noteworthy is that many variables have values of zero even at the 75th percentile, although almost all of our variables have non-zero values at the 95 percentile. Although these distributions are consistent with other studies, we also examine robustness tests examining if our results remain robust in key subsamples including the set of firms with positive R&D activity or in subsamples with above-median patenting activity. Our results remain highly robust.

4.2 Validation: Complaints about Chinese Competition

In this section, we examine if elevated levels of industry-specific Chinese Internet penetration are associated with higher ex-post complaints by US firms that they are facing higher levels of competition specifically from Chinese firms. We use textual analysis of 10-Ks disclosed by U.S. firms during our sample period as explained earlier.

As our hypothesis is that Internet penetration specifically shifts competitive intensity in the market for intellectual property production, we also go one step further. In particular, we also measure the intensity of U.S. firm complaints about competition that appear specifically in paragraphs where the company is discussing innovation. We predict positive results, and such results would serve to validate the economic content of our primary Internet penetration variable.

An analogous framework for other major economies (excluding China) also allows us to rather directly test the validity of the exclusion requirement using powerful placebo tests. We thus examine if our Chinese Internet penetration variable also predicts higher rates of complaints by U.S. firms about competition from Europe, North America (Canada and Mexico) and Japan. If the exclusion requirement holds and if our Internet penetration variable is not related to underlying state variables relating to industry supply or demand shocks, then we predict that these placebo tests should produce insignificant results. As noted earlier, these placebo tests have high power due to the fact that these other economic regions are large in scale and hence U.S. firms frequently summarize the intensity of competition from these regions. They key empirical question is if these complaints are also related to Chinese Internet penetration.

We estimate the following regression

$$Y_{it} = \beta CNInternet_{it-1} + \boldsymbol{\gamma Z_{it-1}} + \alpha_i + \alpha_t + \varepsilon_{it} \tag{1}$$

The dependent variable is a complaint measure as noted above, where all are generated using textual information from U.S. firm 10-K filings. Detailed definitions of these variables can be found in Section 3.1 or Table 11. CNInternet is our key measure of competition and is the weighted-average internet penetration across provinces where Chinese firms agglomerate at the industry level. $\boldsymbol{Z}$ represents the control variables, which include: CNSalesGR, the sales growth of the same industry in China, log(10kSize), log of the total number of paragraphs of each 10-K filing, firm age, size (total asset), and Q. All independent variables are lagged one year relative to the dependent variable and hence are ex-ante measureable. We also include firm and year fixed effects in all regressions, and the standard errors are clustered by firm.

Table 3 shows the results. In the first two columns, we find that the Internet penetration significantly predicts the rate at which treated U.S. firms in the same industry complain about competition specifically from Chinese firms. A one standard deviation increase of the internet penetration ratio leads to a 0.124 standard deviation increase, or a 64% increase from the sample mean of the Chinese competition complaint measure. We obtain similar estimates if the dependent variable is a dummy equal to one if the given U.S. firm has at least one complaint in its 10-K. Columns (3) and (4) of Table 3 show that the high competition measure is also significantly predicted by the internet penetration measure.

Our most direct tests are in the last four columns of Table 3. We find that internet penetration also significantly predicts U.S. firm complaints about competition that are specifically related to intellectual properties (IP) (see Columns (5) and (6)). In Columns (7) and (8), instead of focusing on competition, we consider instances where U.S. firms discuss IP theft. This reflects that fact that IP theft, in an economic sense, is a form of competition and U.S. firm complaints should thus follow similar patterns. We find that indeed they do.

The possibility of IP theft has been a centerpiece of recent public and political debates about recent trade conflicts between US and China. Although we do not draw any strong conclusions with respect to IP theft, our finding that internet penetration significantly predicts IP theft complaints from US firms is suggestive. IP theft, or "perceived IP theft", might thus explain part of the increased competition observed these IP markets. However, we note caution as complaints in 10-Ks do not constitute any proof that any IP theft has, in fact, happened. Moreover, we do document increased patenting by Chinese firms (discussed later), which is not a form of theft given that patents are both transparent and legal. Yet IP theft could be a precursor to such patents as the younger firms in China might use trade secret theft to catch up on overall knowledge capital, which is necessary to create cutting edge patents. Overall, evidence on IP theft is not decisive and this suggestive evidence and the importance of the question indicates that future research examining this issue would be invaluable.

Overall, Table 3 shows that industry-specific internet penetration in China strongly predicts ex-post complaints about competition from US firms, especially competition on the margin of innovation itself. This strong validation indicates that the economic content of our key Internet penetration variable is in line with our predictions.

4.3 Exclusion: Placebo Tests using Other Major Economies

Although the validation documented in the preceding section indicates positive information about content, it alone does not prove the exclusion requirement as other economic forces might also affect this variable. For example industry-specific Internet penetration might be correlated with global supply or demand shocks in the given industry, or it might relate to global competition moreso than Chinese competition alone. In order for our experiment to be ideal, this variable should only identify shifts in the capacity of Chinese firms alone to challenge firms globally on the competitive margin of innovation.

To directly test the exclusion requirement, we construct analogous competition complaint measures for other major economies, namely Japan, Europe, and neighboring countries in North America (Canada and Mexico). If the internet penetration variable contains information about the industry's state thus violating exclusion, we would expect that complaints about competition from these other economies would show similar positive signs. Table 4 shows the results. In Panel (A), we run similar regressions based on Equation 1, but replace the dependent variable by the complaint measures from other countries. For brevity we focus on complaints about competition and intellectual property theft.

Columns (1) - (4) of Table 4 show that Chinese Internet penetration is not significantly related to complaints about competition from Japan or North America. However, Columns (5) and (6) show weakly significant results for the European Union, indicating some potential concerns about exclusion. We examined this issue in-depth and the results indicate that this result is likely spurious. First, the significance of the European Union results are driven fully by the first year of our sample, likely indicating an outlier perhaps relating to the formation of the European Union. If we exclude the first year, the results for the European Union are insigificant whereas our results for China are highly robust.

A second key issue is that our primary measures of industry agglomeration at the province level in China uses geographic headquarter location data from all publicly traded Chinese firms, including those listed in China, Hong Kong, and in the United States. If there is a violation to the exclusion requirement, a most likely source could be that Chinese firms that list in Hong Kong or in the United States have better access to information about innovation in their industries, creating channels for information transmission outside of internet penetration.

We test this issue in Panel (B) of Table 4. In particular, we re-define Internet penetration using industry-specific agglomeration data based *only* on Chinese firms listed in Mainland China (those having A-shares). This more narrow definition of Internet penetration does not load on companies having listings outside of China, and hence limits any alternative channels for information transmission beyond the internet. The results in Panel B lend support to this explanation of the results in Panel A. In particular, all of the placebo tests from all three major economies are insignificantly related to Chinese internet penetration. Although the results in Panel A for the European Union might be spurious and thus less relevant, the results in Panel B indicate a very conservative strategy for our main tests in the paper.

In particular, we run all tests in the paper using our main Internet penetration variable and also separately using our mainland-China-only Internet penetration variable. We note that all of our results are strongly robust to both specifications. Moreover, our results are actually stronger using the mainland-China-only measure. For this reason, we report results using the complete Internet penetration measure in order to be conservative although all resutls are robust to either model.

We briefly note that we later run an additional placebo test later in the paper when we consider patenting activity. We find even stronger support for the exclusion requirement in all of these tests. In particular, our main result is that Chinese firms increase their patenting activity in the markets of the treated US firms after episodes where Internet penetration increases. They also greatly increase cites to the treated U.S. firms in their same industry. The key placebo test we consider later is whether European, North American or Japanese firms do the same. If the exclusion requirement did not hold, we would expect similar results as explained above. As we explain later, we find no significant results for these other economies, and these placebo tests hold regardless of whether we define Internet penetration using all Chinese firms or just those listed in mainland China. As discussed in our summary statistics section, these tests are particularly powerful placebo tests due to the fact that patenting activity overall is more intense for firms from Europe, Japan and North America relative to China.

Collectively, these strong placebo tests suggest that it is unlikely that our internet penetration variable is contaminated by a global factor or by an omitted industry state variable relating to supply or demand shocks. Although we stop short of declaring that our conclusions are fully causal, these findings lend support to the possibility that our results are strongly consistent with internet penetration causing reductions in innovative activities of treated U.S. firms due to a crowding-out effect of increased foreign competition in the market for innovative technologies.

5 Competition and Innovation

In this section, we examine how competition from China, as measured by our industry-specific Chinese internet penetration variable, affects the innovation activities of US firms.

5.1 Impact on U.S. Firms

We first examine how ex ante industry-specific Chinese internet penetration impacts ex post investment in R&D expenses by treated U.S. firms. We do so by estimating a regression model as specified in Equation 1. Our key dependent variables are the R&D/sales and the number of patents/sales of our US firms.

Table 5 shows the results. Column (1), which uses the one-year ahead R&D expense ratio as the dependent variable, shows that Internet penetration significantly negatively predicts ex-post R&D. The coefficient estimate of -0.182 is significant at the 1% level, and indicates that the R&D expense ratio decreases by 0.182 standard deviations when Chinese internet penetration increases by one standard deviation. The coefficient remains significant when we examine the two-year ahead R&D activities in Column 2 and three-year ahead R&D in Column (3).

We find a similar result for the ex post patenting activities of the treated U.S. firms. In Columns (4) - (6) of Table 5, we use the number of patent applications in the next three years divided by sales as the dependent variable. Column (4) shows a highly significant coefficient estimate of -0.108, indicating a decrease of 0.108 standard deviations of patenting activities when Chinese internet penetration increases by one standard deviation. In years two and three, we continue to observe significant and negative coefficients.

To ensure that our results are not driven by the well-known skew-distribution of R&D and patents, we re-estimate the model using Poisson regressions. Table 6 displays the results. To facilitate the Poisson regressions, we drop the firm fixed effects and instead we control for the lagged dependent variable. Overall the negative effects we find for internet penetration on ex post U.S. firm innovation are analogous to those in Table 5.

We thus conclude that plausibly exogenous shocks to the ability of Chinese firms to compete in the market for innovation production are associated with sharp reductions in the ex-post innovation rates of treated U.S. firms. This first main result in our paper is new to the literature, which instead focuses on the margin of competition in the production of existing products.

5.2 Impact on Chinese Firms

We now examine the relationship between ex ante industry-specific internet penetration and ex post increases the number of new Chinese patents that directly cite the existing patents of the impacted US firms. We utilize the country information of the first assignee for each patent to identify patents that are assigned to a Chinese entity. For each firm i in year t, we then count the number of new patents that are (1) applied for through the USPTO, (2) assigned to a Chinese entity, and (3) cite any existing patents of firm i. Following our standard conventions, we then scale this count (PatCiteUS_{CN}) by firm sales.

We use this measure of new Chinese patents (that cite pre-existing same-industry U.S. firm patents) as the dependent variable in our next set of tests. The results are displayed in Table 7. Columns (1) - (3) of Table 7 show that ex ante internet penetration predicts increases in the number of Chinese firms citing patents to these U.S. firms in the next three years. Results are significant at the 1% level in each of the three ex post years. The effects are also large as a one standard deviation increase in internet penetration is followed by a 0.285 standard deviation increase in the number of citing patents by Chinese firms in the following year.

To ensure that our tests are not driven by changes in the overall intensity of patents to a given U.S. firm's existing patents, we consider an alternative scaling that accounts for the cites to these same patents by other U.S. firms. In particular, we define PatCiteUS_{US} as the number of cites to the focal firm's patents by U.S. firms. Columns (4) - (6) of Table 7 show the results of regressions where the dependent variable is PatCiteUS_{CN} / (PatCiteUS_{CN} + PatCiteUS_{US} + 1). The added one in the denominator avoids division by zero and this construction ensures this variable is bounded in [0,1] and thus does not have outliers. We find that the results in Columns (4) to (6) are very similar to our baseline results in Columns (1) to (3). Hence our results are not driven by broad increases in patent cites, but rather are unique to the Chinese firms citing these patents.

The Google patent database also includes all patents filed with SIPO, the Chinese Patent Office. Hence we can also construct a similar measure of Chinese patents that cite the U.S. firm patents, but that are filed in China. The dependent variable for Columns (7) - (9) of Table 7 is PatCiteCN, which is the number of new patents that are applied with SIPO that cite the existing patents of the US firm, and we scale this quantity by the focal firm's sales. We find that the coefficient estimates for internet penetration once again are highly significant and economically large. A one standard deviation increase in internet penetration is associated with an increase of 0.180 to 0.250 standard deviations of these SIPO patents over the three ex post years.

Columns (10) - (12) of Table 7 repeat this exercise using the same scaling convention discussed above for Columns (4) to (6), where the goal is to ensure our results are not explained by broad-based increases in cites to the focal U.S. firm's patents. Our results remain significantly positive in all three years.

Overall, we find consistent evidence that the internet penetration predicts strong ex post patenting activity by Chinese firms, and that these new patents are directly in the technological areas previously covered by the treated U.S. firms. These results suggest that high quality internet service facilitates increased learning by Chinese firms about the existing technologies used by U.S. firms in their industry. Put together with our finding that U.S. firms decrease patenting in these same technological markets, our results suggest that internet penetration is followed by a strong crowding-out effect. As Chinese firms enter these markets for innovation, they absorb a fraction of the associated rents, and thus crowd-out the treated U.S. firms that previously had greater control of these marekts.

5.3 Impact on Firms in Placebo Economies

Analogous to our earlier placebo tests in Table 4 that examined complaints by U.S. firms about competition from rivals in various economic centers, we perform a similar set of placebo tests regarding the ex post patenting results we found for Chinese firms in the previous section.

If the exclusion requirement is strongly violated, we would expect to see significant increases in patenting activity that cites these same U.S. firms by other firms in other major economies including Europe, North America and Japan. As noted earlier in our summary statistics section, these placebo tests are strong due to the fact that patenting activity by firms in these other regions is more active in our sample overall than is patenting activity by Chinese firms. Hence even if relatively modest industry supply and demand effects were driving our results, these placebo tests should produce significant links to our Chinese internet penetration variable for firms in these economies.

We therefore consider regressions analogous to those in Table 7, except that we replace the dependent variable with patenting activity associated with firms in each of these alternative economies. Table 8 displays the results. In Columns (1) to (3), the dependent variable is based on patents filed with USPTO by assigned entities in Japan. Columns (4) to (6) are based on North American entities and (7) to (12) are based on European Union entities.

The results in Table 7 show that, across all columns and thus all economic regions, we find no evidence that our Chinese Internet penetration variable predicts ex post patenting activity by firms in these regions. The absence of results also holds uniformly over the first, second and third years following the increases in internet penetration.

Furthermore, the economic size of the coefficients are much smaller than those for Chinese patents documented earlier. In fact, six of the nine regressions show a negative sign, whereas the results for China are positive and highly significant. Especially when combined with our results for Table 4, these placebo tests indicate that our internet penetration measure rather cleanly measures the ability of Chinese firms uniquely to compete in the market for innovation at a global level. We find no impact for firms in other nations, suggesting that the exclusion requirement likely holds in a first order way.

5.4 Competition and Two Fronts

As we noted in our discussion of hypotheses, the impact of foreign competition on the innovation activities of U.S. firms can vary based on the specific threats posed by the foreign entrants, and also based on the asset composition of the affected U.S. firms. For example, theory suggests that competition in the market for existing products can either increase or decrease innovation activities by affected U.S. firms. Moreover, U.S. firms having trapped assets might have particularly strong incentives to increase innovation spending on the margin. In particular, innovation can help them to "escape competition" and serve higher quality market segments while conceding low quality segments to the entrants.

5.4.1 High versus Low Growth Options

Because our primary focus is on competition in the market for innovation, it also follows that our predictions should be particularly strong for U.S. firms that have stronger growth options, as innovation is a large fraction of firm value for these firms. Analogously, firms with few growth options are likely more impacted by competition on the other margin: competition in the market for existing products.

In this section, we thus consider the two subsample tests motivated by these hypotheses: high versus low growth option value and high vs low asset tangibility (indicating trapped assets with higher likelihood).

We first examine whether our results are stronger for U.S. firms with high versus low growth options as measured by each firm's market to book ratio. To do so, we start with the models we ran in prior sections of this study, but add an interaction between the internet dummy and an additional dummy HighQ, which equals to one if the firm has an above-median market-to-book ratio in the prior year. We also include the HighQ dummy itself in the model. The dependent variables include the complaint measures from Table 3, and the innovation measures from Table 5. Table 9 shows the results.

Columns (1) to (3) show that higher market-to-book firms complain more about competition from China, especially such complaints that occur in the context of paragraphs discussing innovation. As documented in the existing literature, these high valuation firms tend to have more growth options and are more innovative. As a result, their overall valuations load highly on their ability to control markets for innovation in their sectors, and direct competition from Chinese peers on the margin of innovation production should be particularly strong. The coefficient of the interaction term is generally one-third as large as the coefficient of the internet penetration level alone, suggesting an economically large difference between the high Q and low Q firms.

We also find that these high value firms have innovation activities that are also more sensitive to Chinese internet penetration. As shown in Columns (4) to (7), these high market-to-book ratio firms more severely scale back on their R&D expenses and patenting activities when internet penetration is high. The coefficient of the interaction term for R&D in Column (4) is -0.063, almost half the size of the coefficient of the internet penetration variable itself, high is -0.147. The effect is also economically large for patenting activities.

We conclude that our results for competition in the market for innovation are stronger for U.S. firms that that have more valuable growth options and thus more potential exposure to competitive threats that are uniquely in the market for innovation production.

5.4.2 Trapped Assets

The theory of Bloom, Draca, and Van Reenen (2016) suggests that firms with more trapped assets will have stronger incentives to increase innovation following shocks to competition. This is due to the possibility that innovation can facilitate an escape from competition into higher quality market segments (Aghion, Bloom, Blundell, Griffith, and Howitt (2005)). We note that this prediction is also squarely about irreversibility in production and adjustment costs. Hence this prediction pertains specifically to shocks to competition in the market for existing products. When such competition increases, the affected firms become more innovative even if they were not highly innovative before the shock's arrival. Hence the prediction would be that U.S. firms increase innovation following such competitive shocks.

A material related fact obtains from the results of the previous section. Just as our results are stronger when U.S. firms have more growth options, it follows that they are weaker in the diametric opposite subsample: when firms have few growth options. We find in the previous section that such firms have less negative reactions to the competitive shocks than do firms with high growth options. The results in the previous section thus suggest that results for competition in the market for existing products should be different, and might be more positive on U.S. firm innovation activities.

We now test whether the likely existence of trapped assets also favors higher innovation levels for the affected U.S. firms as the aforementioned theories predict. We measure the likely existence of trapped assets using the level of asset tangibility of the U.S. firms. We then consider regressions similar to those in the previous section, but we interact internet penetration with a dummy indicating above-median asset tangibility in the prior year (instead of a high market-to-book dummy).

Table 10 displays the results. Columns (1) to (3)show that firms with higher asset tangibility complain more about the Chinese competition. This supports the notion that these firms face fewer options to adapt to the increased competition because they cannot easily downsize as some theories would predict as optimal. Hence these results are consistent with the existence of trapped assets. Moreover, despite these additional complaints, we find that high asset tangibility firms favor increases in innovation relative to firms with less asset tangibility as the cross terms in Columns (4) to (7) are all positive and highly significant at the 1% level or the 5% level. These findings are consistent with the possibility of increased innovation to plausibly escape competition.

Although these results support the theories of Bloom, Draca, and Van Reenen (2016) and Aghion, Bloom, Blundell, Griffith, and Howitt (2005) for this well-motivated subsample, we note that our broader results show that this outcome is not observed in all situations. In particular, the sample-wide results strongly favor down-sizing of innovative activities when the competitive shock is in the market for innovation production. Yet, these results echo the differences in outcomes when the competitive shock is primarily in the market for existing products, as here the predictions (and the empirical results) are more ambiguous.

As the existing empirical literature, which focuses on the margin of competition in the market for existing products also finds ambiguous results, we believe our more refined analysis of two competitive margins, plus accounting for the role of asset composition of the treated U.S. firms, helps to explain much of the disagreement in the literature regarding the impact of foreign competition on U.S. firm innovative activities. Collectively, our results stress the importance of analyzing competition on multiple margins when the competitive threats are more sophisticated. They also stress the importance of initial conditions such as asset composition, as these characteristics strongly moderate the incentives to increase or decrease innovation.

6 Conclusion

We examine the impact of increased Chinese competition in intellectual property on U.S. firms' R&D and patents. We use Chinese province-level data on internet penetration and geographic industry-specific agglomeration data to generate plausibly exogenous variation in the capacity of Chinese firms to challenge U.S. firm innovation. Our tests show that this instrument is strong. We find higher rates of U.S. firms ex post complaining about high competition from Chinese firms, especially when discussing their innovation. Moreover, we find direct evidence of realized competition as Chinese firms apply for more patents that cite the patents of the U.S. firms that are exposed to the internet penetration. In placebo tests, we find limited evidence that the Chinese internet penetration impacts R&D and patenting for firms from other major economies.

Our main conclusion is that increased intellectual property competition has a strong and robust negative impact on U.S. firm R&D spending and realized patents. This indicates a crowding-out effect as the foreign rivals capture some of the rents of innovation. This is in contrast to the existing empirical literature, which focuses on a competition for existing products, where the impact of foreign competition is more ambiguous.

Our results show vacation by firms with high growth options versus high assets-in-place. The impact of foreign competition in IP production on U.S. firm innovation is particularly negative for firms that have higher valued growth options as measured by their market-to-book ratios. In contrast, the impact is less severe when U.S. firms have assets-in-place and high adjustment costs. As predicted by existing theories, firms with high assets-in-place are likely attempting to differentiate their existing products and thus invest more in R&D and patents. Overall our results help to reconcile disagreement in the literature on whether competition leads to increases or decreases in domestic firm innovation. Given the importance of these issues in political and regulatory circles, we believe more work examining multiple competitive margins and potential IP theft would be invaluable.

References

Aghion, Philippe, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, 2005, Competition and innovation: an inverted u relationship, *Quarterly Journal of Economics* 120, 701–28.

Autor, David, David Dorn, Gordon H Hanson, Gary Pisano, and Pian Shu, 2018, Foreign competition and domestic innovation: Evidence from U.S. patents, MIT working paper.

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, *Journal of Finance* 69, 1923–1960.

Bloom, Nicholas, Mirko Draca, and John Van Reenen, 2016, Trade induced technical change? the impact of chinese imports on innovation, IT and productivity, *Review of Economics Studies* 83, 87–117.

Duranton, Gilles, and Henry G Overman, 2005, Testing for localization using micro-geographic data, *The Review of Economic Studies* 72, 1077–1106.

———, 2008, Exploring the detailed location patterns of uk manufacturing industries using microgeographic data, *Journal of Regional Science* 48, 213–243.

Ellison, Glenn, and Edward L Glaeser, 1997, Geographic concentration in us manufacturing industries: a dartboard approach, *Journal of political economy* 105, 889–927.

Florence, Philip Sargant, 1948, *Investment, location, and size of plant* . vol. 7 (CUP Archive).

Fuchs, Victor R, 1962, The determinants of the redistribution of manufacturing in the united states since 1929, *Review of Economics and Statistics* 44, 167–177.

Hombert, Johan, and Adrien Matray, 2018, Can innovation help us manufacturing firms escape import competition from china?, *The Journal of Finance* 80, 2003–2039.

Hoover, Edgar M, 1948, *Location of economic activity* (McGraw-Hill Book Company, Inc., New York).

Kogan, Leonid, Dimtris Papanikolaou, Amit Seru, and Noah Stoffman, 2016, Technological innovation, resource allocation and growth, *Quarterly Journal of Economics* forthcoming.

Koh, Ping-Sheng, and David Reeb, 2015, Missing r&d, *Journal of Accounting and Economics* pp. 73–94.

Krugman, Paul R, 1993, *Geography and trade* (MIT press).

Phillips, Gordon M., and Alexei Zhdanov, 2013, R&d and the incentives from merger and acquisition activity, *Review of Financial Studies* 34-78, 189–238.

Reinganum, Jennifer F, 1989, The timing of innovation: Research, development, and diffusion, *Handbook of industrial organization* 1, 849–908.

Roberts, Michael R, and Toni M Whited, 2013, Endogeneity in empirical corporate finance1, in *Handbook of the Economics of Finance*vol. 2 . pp. 493–572 (Elsevier).

Tables

Table 1: Summary Statistics

This table shows the summary statistics of the variables used in our analyses. Detailed variable definitions can be found in Table 11

Variable	N	Mean	Std. Dev.	Median	75th	95th	99th
CNInternet	61930	0.36	0.23	0.31	0.56	0.75	0.77
CNComp % x 1000	61930	0.15	0.77	0.00	0.00	0.00	5.64
CNComp Dummy	61930	0.05	0.21	0.00	0.00	0.00	1.00
CNCompHi % x 1000	61930	0.09	0.51	0.00	0.00	0.00	3.86
CNCompHi Dummy	61930	0.03	0.18	0.00	0.00	0.00	1.00
CNIntTheft % x 1000	61930	0.04	0.27	0.00	0.00	0.00	2.2
CNIntTheft Dummy	61930	0.02	0.14	0.00	0.00	0.00	1.00
CNIntComp % x 1000	61930	0.05	0.32	0.00	0.00	0.00	2.52
CNIntComp Dummy	61930	0.02	0.15	0.00	0.00	0.00	1.00
EUComp % x 1000	61930	0.31	1.08	0.00	0.00	2.46	6.58
EUCompHi % x 1000	61930	0.16	0.69	0.00	0.00	1.37	4.62
EUIntComp % x 1000	61930	0.11	0.55	0.00	0.00	0.00	3.88
JPComp % x 1000	61930	0.04	0.30	0.00	0.00	0.00	2.39
JPCompHi % x 1000	61930	0.01	0.13	0.00	0.00	0.00	1.28
JPIntComp % x 1000	61930	0.03	0.35	0.00	0.00	0.00	0.00
NAComp % x 1000	61930	0.24	0.93	0.00	0.00	1.96	6.15
NACompHi % x 1000	61930	0.10	0.53	0.00	0.00	0.00	3.85
NAIntComp % x 1000	61930	0.05	0.32	0.00	0.00	0.00	2.53
XRD/Sales	61930	0.14	0.54	0.00	0.06	0.48	4.24
NPatent/Sales	61930	0.03	0.12	0.00	0.00	0.10	0.99
PatCiteCN	61930	3.33	35.78	0.00	0.00	6.00	67.00
PatCiteUS$_{CN}$	61930	2.39	32.09	0.00	0.00	4.00	41.00
PatCiteUS$_{EU}$	61930	26.85	237.32	0.00	1.00	57.00	549.00
PatCiteUS$_{JP}$	61930	23.88	286.82	0.00	0.00	34.00	357.71
PatCiteUS$_{NA}$	61930	5.06	53.76	0.00	0.00	11.00	93.00
PatCiteUS$_{US}$	61930	226.84	2118.64	0.00	14.00	499.00	4558.55
Age	61884	18.06	13.48	14.00	24.00	47.00	53.00
CNSalesGR	61930	0.08	0.29	0.09	0.25	0.57	0.86
Q	61831	1.95	1.78	1.36	2.09	5.03	11.19
Sales	59849	2702.18	12607.52	283.5	1236.07	10519.42	43890.60
log(TA)	61790	6.20	2.18	6.24	7.70	9.88	11.42
AssetTangibility	59483	0.16	0.20	0.07	0.22	0.62	0.92
CNInternet_Macro	61930	0.27	0.20	0.23	0.46	0.62	0.70
CNInternet_Top1	61930	0.35	0.24	0.29	0.55	0.75	0.78

Table 2: Summary Statistics at the firm level

We first calculate the mean value of each variables for each firm, and the table shows the summary statistics of the firm-averages. Detailed variable definitions can be found in Table 11

Variable	N	Mean	Std. Dev.	Median	75th	95th	99th
CNInternet	8474	0.33	0.20	0.31	0.48	0.70	0.76
CNComp % x 1000	8474	0.16	0.69	0.00	0.00	0.98	4.47
CNComp Dummy	8474	0.05	0.18	0.00	0.00	0.36	1.00
CNCompHi % x 1000	8474	0.09	0.44	0.00	0.00	0.48	2.70
CNCompHi Dummy	8474	0.03	0.15	0.00	0.00	0.20	1.00
CNIntTheft % x 1000	8474	0.04	0.25	0.00	0.00	0.00	1.73
CNIntTheft Dummy	8474	0.02	0.13	0.00	0.00	0.00	1.00
CNIntComp % x 1000	8474	0.05	0.26	0.00	0.00	0.17	1.56
CNIntComp Dummy	8474	0.02	0.12	0.00	0.00	0.08	0.83
EUComp % x 1000	8474	0.30	0.85	0.00	0.00	2.06	4.36
EUCompHi % x 1000	8474	0.15	0.54	0.00	0.00	1.13	2.82
EUIntComp % x 1000	8474	0.11	0.43	0.00	0.00	0.79	2.31
JPComp % x 1000	8474	0.04	0.23	0.00	0.00	0.12	1.35
JPCompHi % x 1000	8474	0.01	0.10	0.00	0.00	0.00	0.49
JPIntComp % x 1000	8474	0.03	0.26	0.00	0.00	0.00	0.91
NAComp % x 1000	8474	0.23	0.76	0.00	0.00	1.56	4.18
NACompHi % x 1000	8474	0.10	0.42	0.00	0.00	0.65	2.32
NAIntComp % x 1000	8474	0.05	0.24	0.00	0.00	0.20	1.27
XRD	8474	35.92	279.66	0.00	9.42	89.83	564.53
NPatent	8474	15.21	220.09	0.00	0.27	16.11	211.06
PatCiteCN	8474	1.79	21.34	0.00	0.00	2.21	33.74
PatCiteUS$_{CN}$	8474	1.30	18.51	0.00	0.00	1.58	21.12
PatCiteUS$_{EU}$	8474	15.53	164.67	0.00	0.5	26.19	308.52
PatCiteUS$_{JP}$	8474	13.44	193.68	0.00	0.14	15.72	191.35
PatCiteUS$_{NA}$	8474	2.91	34.42	0.00	0.00	5.10	50.60
PatCiteUS$_{US}$	8474	132.00	1485.49	0.00	5.9	228.06	2467.89
Age	8465	13.97	12.10	9.50	17.50	44.00	48.00
CNSalesGR	8474	0.06	0.18	0.06	0.13	0.32	0.47
Q	8459	2.00	1.70	1.42	2.18	5.17	9.84
Sales	8158	1811.31	9251.19	177.30	764.93	6705.79	30456.79
log(TA)	8443	5.73	2.15	5.75	7.22	9.38	10.94
AssetTangibility	8215	0.16	0.20	0.07	0.22	0.61	0.81
CNInternet_Macro	8474	0.25	0.16	0.26	0.37	0.54	0.66
CNInternet_Top1	8474	0.32	0.21	0.30	0.48	0.72	0.77

Table 3: Competition complaints and Chinese internet penetration

The table displays OLS regressions in which the dependent variables are textual measures of competition complaints in 10K filings. We search for four types of complaints in the 10K filings. CNComp measures competition in general; CNCompHi measures competition with high intensity; CNIntComp measures intellectual property competition; CNIntTheft measures intellectual property theft. All these competition measures are China-specific, meaning the words "China" or "Chinese" appear in the the same paragraph as the competition complaint phrases. We exclude instances if other countries are in the same paragraph to ensure the compeitition disucssion is truely about China. More detailed variable construction procedures can be found in Table 11 in the Appendix. In Columns (1), (3), (5), and (7), the dependent variables are the number of paragraphs containing the above search instances divided by the total number of paragraphs of the 10K filing. In Columns (2), (4), (6), and (8), the dependent variables are dummies that equal to 1 if we found any of the phrases in the search. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables, except for *log(10kSize)*, are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2001 to 2015 with 10K filings. We exclude all observations where the total asset or sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	CNComp		CNCompHi		CNIntComp		CNIntTheft	
	%	dummy	%	dummy	%	dummy	%	dummy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CNInternet	0.124***	0.158***	0.119***	0.137***	0.127***	0.148***	0.126***	0.137***
	(0.041)	(0.044)	(0.039)	(0.042)	(0.040)	(0.041)	(0.039)	(0.041)
CNSalesGR	0.001	0.006*	−0.001	0.002	−0.001	0.001	−0.004	−0.004*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
log(10kSize)	−0.108***	−0.033***	−0.111***	−0.052***	−0.097***	−0.063***	−0.042***	−0.023***
	(0.011)	(0.008)	(0.011)	(0.009)	(0.011)	(0.010)	(0.008)	(0.007)
log(Age + 1)	−0.059**	−0.059**	−0.061**	−0.061**	−0.032	−0.030	−0.008	−0.004
	(0.024)	(0.024)	(0.025)	(0.025)	(0.027)	(0.027)	(0.024)	(0.024)
log(TA)	0.050*	0.030	0.067**	0.045*	0.037	0.029	0.046	0.044
	(0.029)	(0.028)	(0.028)	(0.027)	(0.031)	(0.030)	(0.031)	(0.030)
Q	−0.015**	−0.015**	−0.013**	−0.013**	−0.015**	−0.015**	−0.004	−0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	61,930	61,930	61,930	61,930	61,930	61,930	61,930	61,930
Adjusted R^2	0.586	0.519	0.523	0.489	0.472	0.473	0.605	0.604

Table 4: Placebo tests - Competition from other countries and Chinese internet penetration

The table displays OLS regressions in which the dependent variables are textual measures of competition complaints from 10K filings. The dependent variables are constructed in a similar way as in Table 3. However, instead of measuring China-related competition complaints, we now search for competition complaints about other regions of the world. More specifically, Columns (1) - (2) report searches using European Union countries, Column (3) - (4) using Japan, and Columns (5)-(6) using Canada and Mexico. All the dependent variables are the count of matched paragraphs divided by the total number of paragraphs in the 10K filings. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables, except for *log(10kSize)*, are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2001 to 2015 with 10K filings. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

Panel A: Weights from A-share, HK-, and US-listed firms

	JP		NA		EU	
	IntComp	IntTheft	IntComp	IntTheft	IntComp	IntTheft
	(1)	(2)	(3)	(4)	(5)	(6)
CNInternet	0.021	0.009	0.059	0.005	0.085*	0.052**
	(0.038)	(0.008)	(0.044)	(0.005)	(0.048)	(0.025)
CNSalesGR	−0.002	−0.0001	0.001	−0.00001	−0.0002	−0.001
	(0.003)	(0.001)	(0.005)	(0.0003)	(0.004)	(0.002)
log(10kSize)	−0.081***	−0.016***	−0.156***	−0.006***	−0.251***	−0.081***
	(0.014)	(0.004)	(0.014)	(0.001)	(0.020)	(0.009)
log(Age + 1)	0.061***	−0.004	−0.035	−0.001	−0.059**	−0.013
	(0.023)	(0.006)	(0.026)	(0.003)	(0.027)	(0.014)
log(TA)	0.024	0.006	0.126***	0.008**	0.227***	0.085***
	(0.046)	(0.008)	(0.033)	(0.004)	(0.041)	(0.018)
Q	0.007	−0.003	0.014*	0.001	−0.008	−0.0004
	(0.008)	(0.003)	(0.008)	(0.001)	(0.011)	(0.006)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	61,930	61,930	61,930	61,930	61,930	61,930
Adjusted R^2	0.379	0.298	0.345	0.437	0.326	0.399

Panel B: Weights from A-share listed firms only

	JP		NA		EU	
	IntComp	IntTheft	IntComp	IntTheft	IntComp	IntTheft
	(1)	(2)	(3)	(4)	(5)	(6)
CNInternet	0.032	0.009	0.053	0.005	0.020	0.027
	(0.036)	(0.008)	(0.038)	(0.005)	(0.049)	(0.025)
CNSalesGR	−0.002	−0.0002	0.001	−0.00002	−0.0003	−0.001
	(0.003)	(0.001)	(0.005)	(0.0003)	(0.004)	(0.002)
log(10kSize)	−0.081***	−0.016***	−0.156***	−0.006***	−0.251***	−0.081***
	(0.014)	(0.004)	(0.014)	(0.001)	(0.020)	(0.009)
log(Age + 1)	0.061***	−0.004	−0.035	−0.001	−0.059**	−0.013
	(0.023)	(0.006)	(0.026)	(0.003)	(0.027)	(0.014)
log(TA)	0.024	0.006	0.126***	0.008**	0.226***	0.085***
	(0.046)	(0.008)	(0.033)	(0.004)	(0.041)	(0.018)
Q	0.007	−0.003	0.014*	0.001	−0.008	−0.0005
	(0.008)	(0.003)	(0.008)	(0.001)	(0.011)	(0.006)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	61,930	61,930	61,930	61,930	61,930	61,930
Adjusted R^2	0.379	0.298	0.345	0.437	0.326	0.399

Table 5: Innovation activities and Chinese internet penetration

The table displays OLS regressions in which the dependent variables are firms' innovation activities. The dependent variable in Columns (1) - (3) is the R&D expenses over sales. For missing R&D, we follow the Koh and Reeb (2015) and replace the missing with industry average if the firm files for any patent patents applications in the past three years (including the current year), and 0 otherwise. The dependent variables are measures from 1, 2, or 3 years in the future. The dependent variable in Columns (4) - (6) is the total number of patent applications each year (by filing date) dividend by sales. The patent data comes from Google Patents, and we match the patents to Compustat firms using the links from Kogan, Papanikolaou, Seru, and Stoffman (2016). The dependent variables are measures from 1, 2, or 3 years in the future. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2003 to 2015. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	XRD/Sales			NPatent / Sales		
	t+1	t+2	t+3	t+1	t+2	t+3
	(1)	(2)	(3)	(4)	(5)	(6)
CNInternet	-0.182^{***}	-0.181^{***}	-0.175^{***}	-0.108^{***}	-0.086^{**}	-0.065^{*}
	(0.042)	(0.042)	(0.040)	(0.041)	(0.041)	(0.038)
CNSalesGR	0.001	0.003	0.002	-0.002	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log(Age + 1)	-0.101^{***}	-0.096^{***}	-0.084^{***}	-0.106^{***}	-0.087^{***}	-0.067^{***}
	(0.017)	(0.019)	(0.021)	(0.021)	(0.020)	(0.021)
log(TA)	0.034	0.048^{*}	0.007	-0.058^{**}	-0.033	-0.031
	(0.027)	(0.028)	(0.029)	(0.029)	(0.028)	(0.026)
Q	-0.002	0.011	0.001	-0.016	-0.007	-0.003
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	61,768	53,799	46,726	61,768	53,799	46,726
Adjusted R^2	0.736	0.736	0.740	0.724	0.752	0.785

Table 6: Innovation activities and Chinese internet penetration - Poisson Regression

The table displays poisson regressions in which the dependent variables are firms' innovation activities. The dependent variable in Columns (1) - (3) is the R&D expenses over sales. For missing R&D, we follow the Koh and Reeb (2015) and replace the missing with industry average if the firm files for any patent patents applications in the past three years (including the current year), and 0 otherwise. The dependent variables are measures from 1, 2, or 3 years in the future. The dependent variable in Columns (4) - (6) is the total number of patent applications each year (by filing date) dividend by sales. The patent data comes from Google Patents, and we match the patents to Compustat firms using the links from Kogan, Papanikolaou, Seru, and Stoffman (2016). The dependent variables are measures from 1, 2, or 3 years in the future. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2003 to 2015. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	XRD/Sales			NPatent / Sales		
	t+1	t+2	t+3	t+1	t+2	t+3
	(1)	(2)	(3)	(4)	(5)	(6)
CNInternet	-0.348***	-0.474***	-0.542***	-0.427***	-0.512***	-0.518***
	(0.0574)	(0.0551)	(0.0606)	(0.101)	(0.105)	(0.114)
CNSalesGR	-0.0555***	-0.0189	-0.0420**	-0.0601***	-0.0244	-0.0358*
	(0.0174)	(0.0144)	(0.0190)	(0.0203)	(0.0168)	(0.0187)
log(Age + 1)	-0.189***	-0.187***	-0.154***	0.0433	0.00330	-0.0136
	(0.0207)	(0.0214)	(0.0215)	(0.0326)	(0.0305)	(0.0328)
log(TA)	-0.568***	-0.537***	-0.555***	-0.385***	-0.325***	-0.337***
	(0.0277)	(0.0283)	(0.0277)	(0.0377)	(0.0388)	(0.0372)
Q	0.0287	0.0575***	0.0525***	0.0402*	0.0671***	0.0616**
	(0.0184)	(0.0177)	(0.0188)	(0.0215)	(0.0223)	(0.0241)
Lagged XRD/Sales	0.242***	0.242***	0.246***			
	(0.0261)	(0.0265)	(0.0285)			
Lagged NPatent/Sales						
				0.206***	0.216***	0.219***
				(0.0223)	(0.0230)	(0.0268)
Year FE	Y	Y	Y	Y	Y	Y
Observations	56,032	53,400	46,404	56,032	53,400	46,404

Table 7: Patent citations and Chinese internet penetration

The table displays OLS regressions in which the dependent variables are the annual number of citations by Chinese firms on the firm's existing patents. In Columns (1) - (3), for each firm we count the number of new patents that have cited the firm's existing patents in each year. We further require the first assignee of the citing patent is a Chinese company, and the patent is filed in the US with USPTO. The dependent variables in Columns (1) - (3) are the total count number, PatCiteUS$_{CN}$, divided by sales in the next three years, respectively. In Columns (4) - (6), we further compare PatCiteUS$_{CN}$ to the number of citations from news patents which are filed with USPTO and assigned to US firms. The dependent variables in Columns (4) - (6) are PatCiteUS$_{CN}$/(PatCiteUS$_{CN}$ + PatCiteUS$_{US}$ + 1) in the next three years, respectively. In Columns (7) - (9), PatCiteCN counts the number of new patents filed with Chinese Patent Office (SIPO) that have cited the firm's existing patents. We exclude patents that are filed in SIPO but are assigned to US companies. In Columns (10) - (12), we use PatCiteCN / (PatCiteCN + PatCiteUS + 1) as the dependent variables, where the PatCiteUS is the total counts of new citing patents filed in the US. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2003 to 2015. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	$\frac{PatCiteUS_{CN}}{Sales}$			$\frac{PatCiteUS_{CN}}{PatCiteUS_{CN}+PatCiteUS_{US}+1}$			$\frac{PatCiteCN}{Sales}$			$\frac{PatCiteCN}{PatCiteCN+PatCiteUS+1}$		
	t+1	t+2	t+3	t+1	t+2	t+3	t+1	t+2	t+3	t+1	t+2	t+3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CNInternet	0.285***	0.256***	0.227***	0.262***	0.264***	0.224***	0.250***	0.181***	0.180***	0.336***	0.294***	0.298***
	(0.053)	(0.053)	(0.054)	(0.045)	(0.046)	(0.047)	(0.046)	(0.046)	(0.047)	(0.048)	(0.052)	(0.056)
CNSalesGR	−0.0003	0.002	0.004	0.001	0.006	0.015***	0.002	0.005*	0.008**	−0.008*	0.002	0.012***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
log(Age + 1)	0.050**	0.051**	0.026	−0.073***	−0.056**	−0.030	−0.006	−0.015	−0.030	−0.467***	−0.447***	−0.426***
	(0.022)	(0.023)	(0.024)	(0.023)	(0.023)	(0.024)	(0.020)	(0.021)	(0.021)	(0.031)	(0.031)	(0.031)
log(TA)	−0.298***	−0.251***	−0.166***	−0.116***	−0.104***	−0.090***	−0.345***	−0.293***	−0.208***	−0.027	−0.050*	−0.051
	(0.035)	(0.036)	(0.037)	(0.027)	(0.029)	(0.031)	(0.034)	(0.036)	(0.036)	(0.029)	(0.030)	(0.032)
Q	−0.059***	−0.052***	−0.051***	−0.050***	−0.048***	−0.039***	−0.037***	−0.046***	−0.039***	−0.008	−0.009	−0.005
	(0.013)	(0.014)	(0.015)	(0.009)	(0.010)	(0.012)	(0.013)	(0.013)	(0.015)	(0.008)	(0.008)	(0.008)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	61,768	53,799	46,726	61,768	53,799	46,726	61,768	53,799	46,726	61,768	53,799	46,726
Adjusted R^2	0.401	0.435	0.470	0.245	0.268	0.283	0.456	0.493	0.521	0.322	0.341	0.364

Table 8: Placebo tests - patent citations from other counties and Chinese internet penetration

The table displays OLS regressions in which the dependent variables are the annual number of citations by firms in other economies on the firm's existing patents. We define the dependent variables as in the Columns (1)-(3) of Table 7. PatCiteUS$_{it}^{JP}$ are the number of patents, which are filed by Japanese firms with USPTO in year t, that cite firm i's existing patents. Similarly, PatCiteUS$_{it}^{NA}$ are the patent counts filed by firms from Canada or Mexica, and PatCiteUS$_{it}^{EU}$, the firms from European Union. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2003 to 2015. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	$\frac{PatCiteUS_{JP}}{Sales}$			$\frac{PatCiteUS_{NA}}{Sales}$			$\frac{PatCiteUS_{EU}}{Sales}$		
	t+1	t+2	t+3	t+1	t+2	t+3	t+1	t+2	t+3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CNInternet	−0.041	−0.062	−0.068	0.024	0.030	0.039	−0.014	−0.029	−0.043
	(0.044)	(0.046)	(0.049)	(0.044)	(0.046)	(0.046)	(0.045)	(0.046)	(0.048)
CNSalesGR	−0.001	0.002	−0.002	−0.004*	0.001	−0.002	−0.0001	0.002	−0.0003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
log(Age + 1)	0.066***	0.035*	0.030	0.075***	0.065***	0.065***	0.062***	0.036*	0.026
	(0.017)	(0.020)	(0.022)	(0.019)	(0.021)	(0.022)	(0.017)	(0.019)	(0.020)
log(TA)	−0.198***	−0.093***	−0.029	−0.222***	−0.157***	−0.107***	−0.207***	−0.123***	−0.062*
	(0.030)	(0.032)	(0.036)	(0.030)	(0.032)	(0.032)	(0.030)	(0.032)	(0.033)
Q	−0.037***	−0.024*	−0.002	−0.032**	−0.014	−0.002	−0.040***	−0.028**	−0.006
	(0.012)	(0.013)	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	61,768	53,799	46,726	61,768	53,799	46,726	61,768	53,799	46,726
Adjusted R^2	0.672	0.682	0.687	0.524	0.542	0.563	0.697	0.706	0.719

Table 9: Subsample analysis - by Q

This table re-estimates regressions in Table 3 and 5 with an additional variable, HighQ, which equals to 1 if a firm's Q is higher than the median Q in each year, and 0 otherwise. We interact the HighQ dummy with the Chinese internet penetration variable and test whether high- and low-Q firms have different reponses in their innovation activities to Chinese competition. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample construction follows the same procedure as in previous tables. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	CNComp	CNCompHi	CNIntComp	XRD/Sales		NPatent/Sales	
	t+1	t+1	t+1	t+1	t+3	t+1	t+3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CNInternet x HighQ	0.036***	0.030**	0.034***	−0.063***	−0.056***	−0.046***	−0.046***
	(0.012)	(0.012)	(0.013)	(0.010)	(0.012)	(0.011)	(0.012)
CNInternet	0.100**	0.098**	0.095**	−0.147***	−0.147***	−0.083**	−0.043
	(0.043)	(0.041)	(0.040)	(0.038)	(0.036)	(0.038)	(0.035)
CNSalesGR x HighQ	−0.005	−0.005	0.002	−0.001	−0.004	−0.001	0.008**
	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
CNSalesGR	0.004	0.002	−0.002	0.001	0.004*	−0.001	−0.003
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
HighQ	−0.050***	−0.047**	−0.068***	0.085***	0.077***	0.043**	0.049**
	(0.018)	(0.019)	(0.020)	(0.019)	(0.020)	(0.021)	(0.020)
log(10kSize)	−0.051***	−0.057***	−0.046***				
	(0.008)	(0.009)	(0.009)				
log(Age + 1)	−0.057**	−0.060**	−0.034	−0.104***	−0.086***	−0.108***	−0.068***
	(0.024)	(0.025)	(0.026)	(0.017)	(0.021)	(0.021)	(0.021)
log(TA)	0.037	0.055**	0.022	0.041	0.014	−0.054*	−0.026
	(0.029)	(0.028)	(0.032)	(0.026)	(0.029)	(0.029)	(0.026)
Q	−0.014**	−0.012*	−0.011	−0.006	−0.004	−0.015	−0.006
	(0.006)	(0.006)	(0.007)	(0.015)	(0.015)	(0.015)	(0.015)
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
N	60,730	60,730	60,730	61,768	46,726	61,768	46,726
Adjusted R^2	0.587	0.526	0.478	0.737	0.740	0.725	0.785

Table 10: Subsample analysis - by Asset Tangibility

This table re-estimates regressions in Table 3 and 5 with an additional variable, HighT, which equals to 1 if a firm's asset tangibility is higher than the median asset tangibility in each year, and 0 otherwise. We interact the HighT dummy with the Chinese internet penetration variable and test whether high- and low-asset tangibility firms have different reponses in their innovation activities to Chinese competition. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample construction follows the same procedure as in previous tables. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	CNComp	CNCompHi	CNIntComp	XRD/Sales		NPatent/Sales	
	t+1	t+1	t+1	t+1	t+3	t+1	t+3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CNInternet x HighT	0.035**	0.028*	0.039**	0.061***	0.040***	0.045***	0.032**
	(0.014)	(0.015)	(0.016)	(0.012)	(0.014)	(0.013)	(0.013)
CNInternet	0.090**	0.092**	0.083**	−0.223***	−0.204***	−0.136***	−0.083*
	(0.044)	(0.044)	(0.042)	(0.049)	(0.048)	(0.047)	(0.044)
CNSalesGR x HighT	0.006	0.006	0.006	0.003	0.008*	0.007*	0.004
	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)	(0.004)
CNSalesGR	−0.002	−0.004	−0.004	−0.001	−0.003	−0.006*	−0.001
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
HighT	−0.004	−0.0003	0.021	−0.092***	−0.040*	−0.048*	−0.014
	(0.023)	(0.024)	(0.026)	(0.022)	(0.023)	(0.027)	(0.025)
log(10kSize)	−0.053***	−0.059***	−0.048***				
	(0.009)	(0.009)	(0.010)				
log(Age + 1)	−0.047*	−0.052*	−0.022	−0.084***	−0.075***	−0.095***	−0.060***
	(0.026)	(0.027)	(0.028)	(0.017)	(0.021)	(0.021)	(0.022)
log(TA)	0.045	0.063**	0.034	0.032	0.006	−0.062**	−0.032
	(0.030)	(0.029)	(0.033)	(0.028)	(0.030)	(0.030)	(0.027)
Q	−0.012*	−0.011*	−0.009	−0.006	0.0004	−0.017	−0.002
	(0.006)	(0.006)	(0.007)	(0.015)	(0.014)	(0.015)	(0.014)
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
N	58,415	58,415	58,415	59,396	44,907	59,396	44,907
Adjusted R^2	0.587	0.526	0.478	0.736	0.739	0.724	0.785

Figures

Figure 1: Complaints about Chinese competition

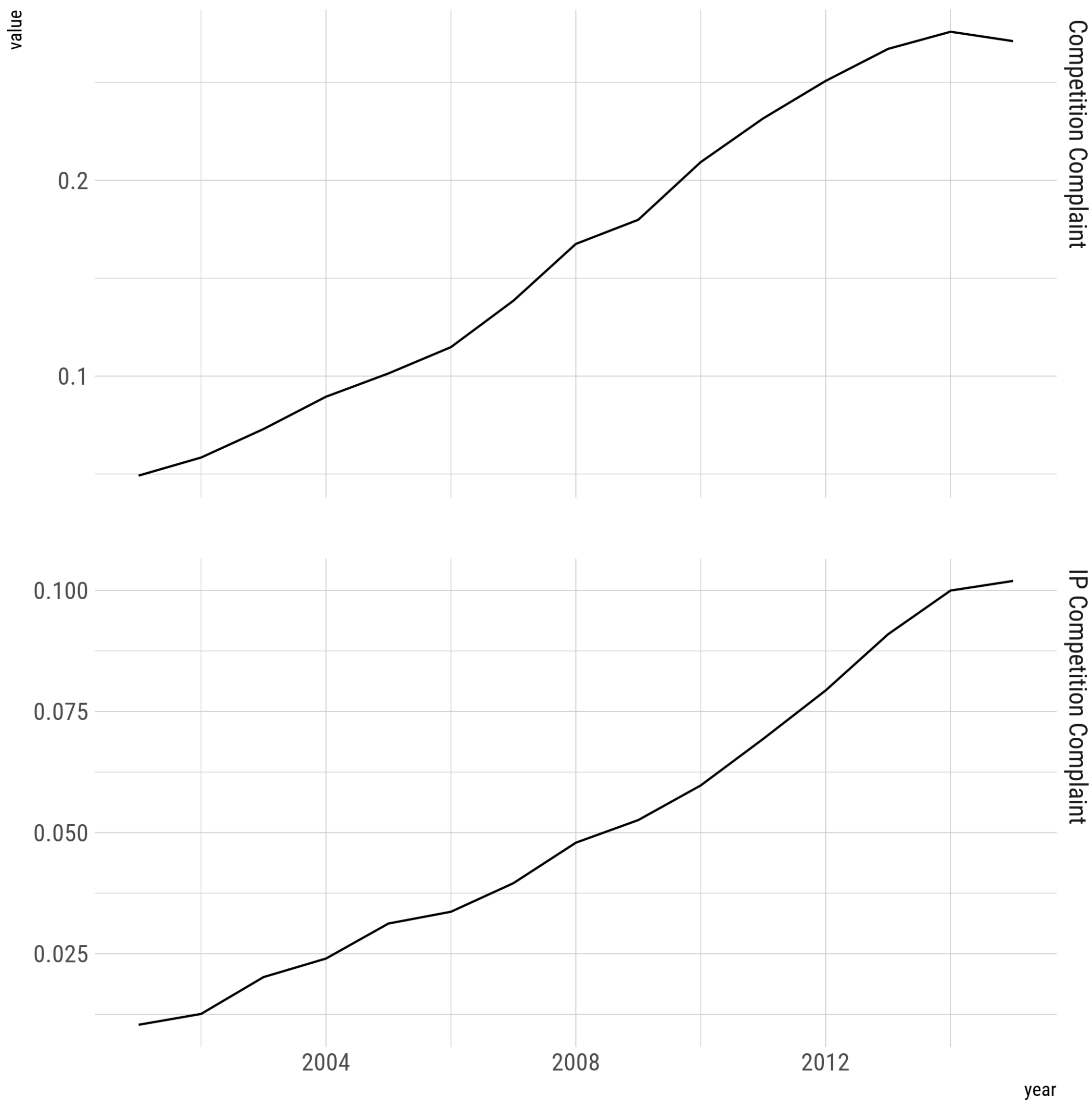


Figure 2: Internet penetration growth variation

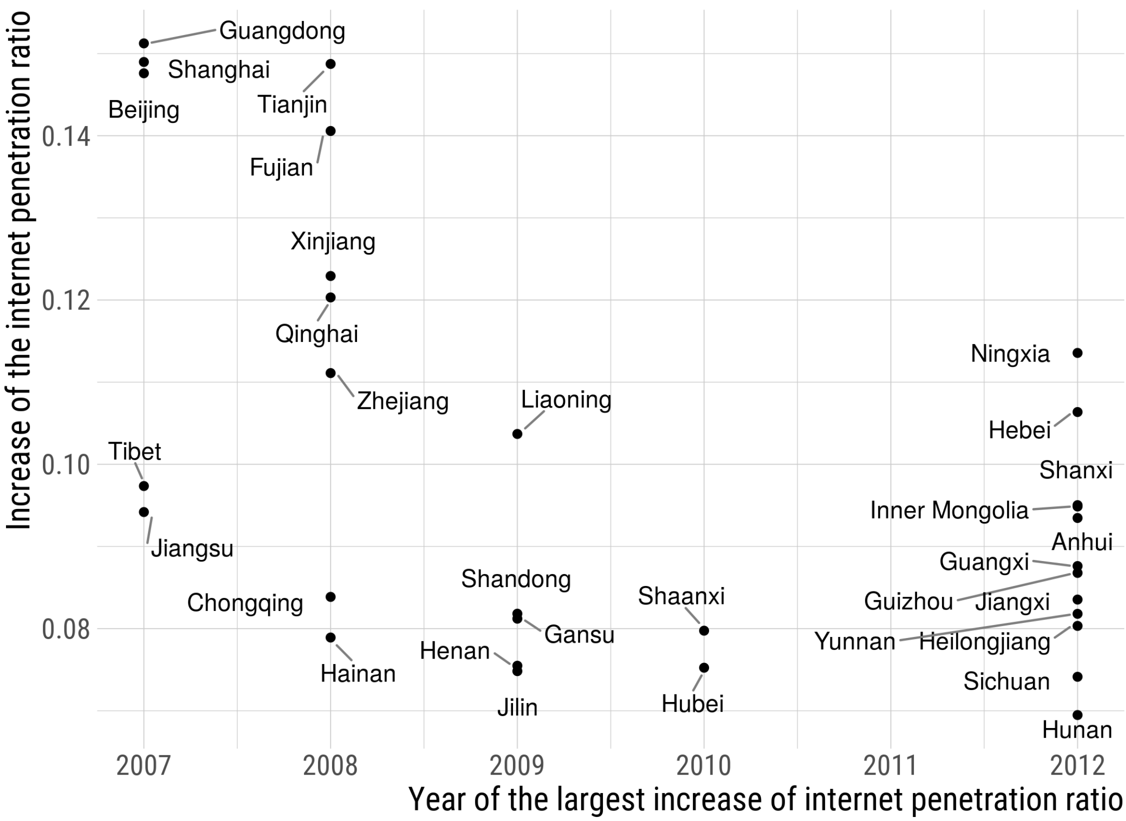


Figure 3: Number of industries (SIC2) covered by Chinese public firms

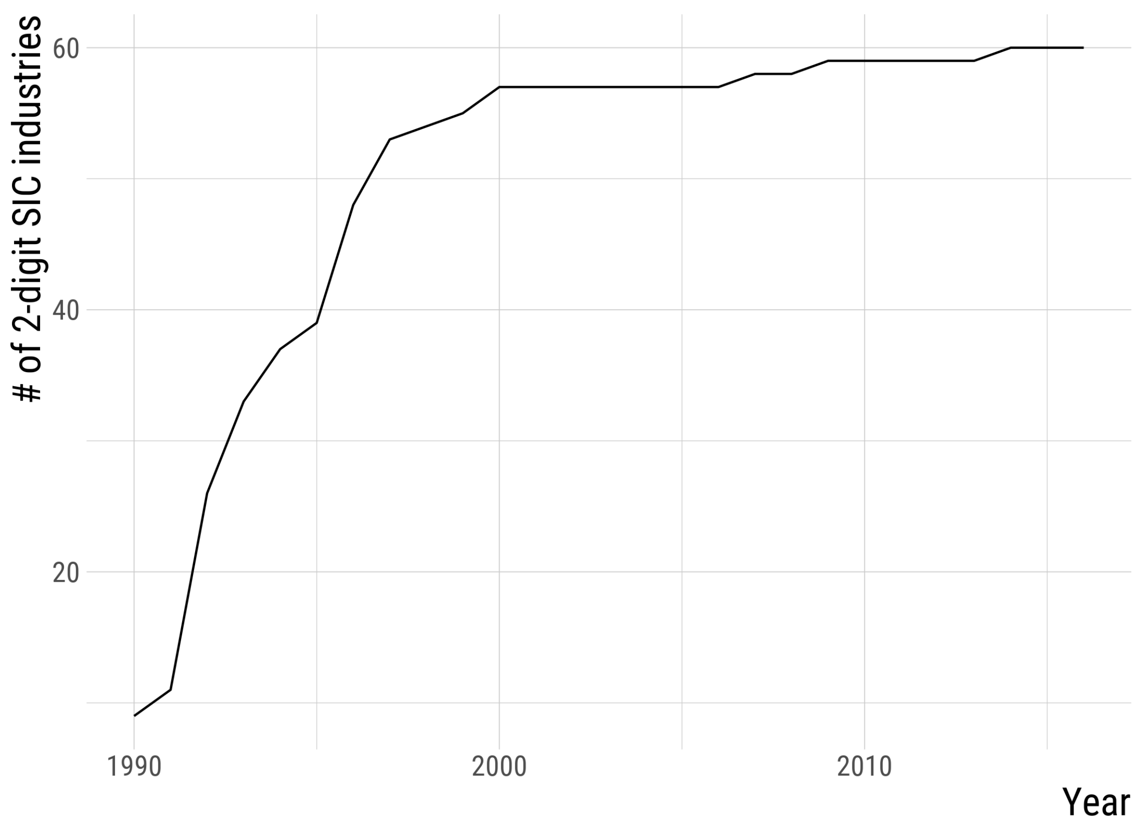
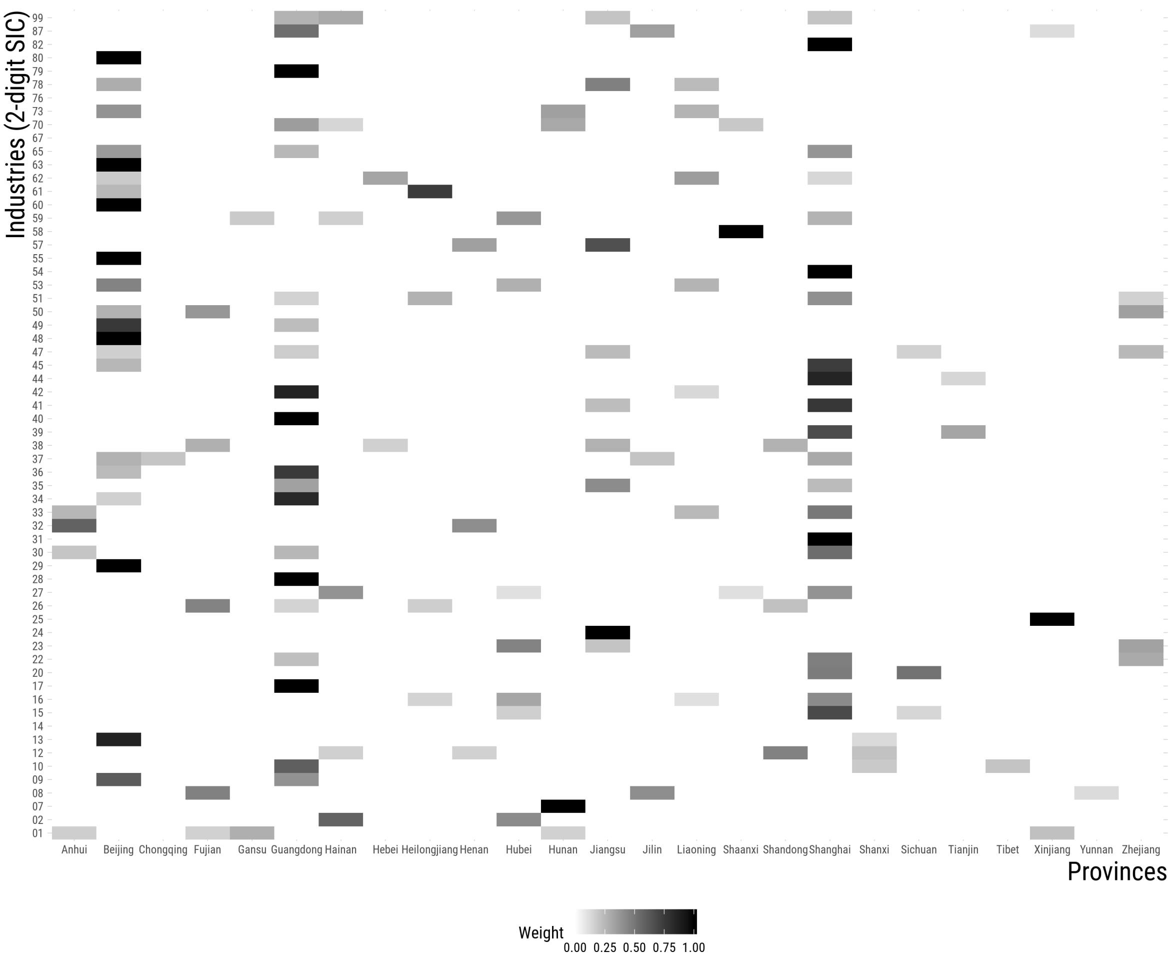


Figure 4: Weight loadings by Province-Industry



Appendix A. Variable definitions

Table 11: Variable definitions

Table 11

Variable	Definition	Source
CNInternet	The weighted average internet penetration ratio across provinces in China. We first collect the number of internet users from annual reports. We then get the number of population for each province-year from China Data Online and calculate the internet penetration ratio. Next, for each industry, we calculate the weights across provinces using the total assets of all the Chinese public firms (mainland A-share, Hongkong, and US) in 2000, and the same weights are used in all later years. We assign each public firm to the province of its headquarter. In calculating the weights for each industry, we keep only provinces whose weights are above 10%, and then calculate CNInternet as the weighted-average of the internet penetration ratio, where the weights are the total asset of the public firms of the industry from the province.	CNNIC Reports; CSMAR; Capital IQ; China Data Online
CNComp %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [China, Chinese]; List 2: [compete, competition, competing]	10-K Filing
CNComp Dummy	A dummy variable that equals to one if CNComp % is larger than 0, and 0 otherwise.	10-K Filing
CNCompHi %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [China, Chinese]; List 2: [compete, competition, competing]; List 3: [high, intense, significant, face, faces, substantial, significant, continued, vigorous, strong, aggressive, fierce, stiff, extensive, severe]	10-K Filing
CNCompHi Dummy	A dummy variable that equals to one if CNCompHi % is larger than 0, and 0 otherwise.	10-K Filing
CNIntComp %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [China, Chinese]; List 2: [compete, competition, competing]; List 3: [intellectual]; List 4: [property]	10-K Filing
CNIntComp Dummy	A dummy variable that equals to one if CNIntComp % is larger than 0, and 0 otherwise.	10-K Filing
CNIntTheft %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [China, Chinese]; List 2: [protect, infringe, theft]; List 3: [intellectual]; List 4: [property]	10-K Filing
CNIntTheft Dummy	A dummy variable that equals to one if CNIntTheft % is larger than 0, and 0 otherwise.	10-K Filing
EUIntComp %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Europe, European]; List 2: [compete, competition, competing]; List 3: [intellectual]; List 4: [property]	10-K Filing
EUIntTheft %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Europe, European]; List 2: [protect, infringe, theft]; List 3: [intellectual]; List 4: [property]	10-K Filing
JPIntComp %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Japan, Japanese]; List 2: [compete, competition, competing]; List 3: [intellectual]; List 4: [property]	10-K Filing

Continued on next page

Table 11 – *Continued from previous page*

Variable	Definition	Source
JPIntTheft %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Japan, Japanese]; List 2: [protect, infringe, theft]; List 3: [intellectual]; List 4: [property]	10-K Filing
NAIntComp %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Mexico, Mexican, Canada, Canadian]; List 2: [compete, competition, competing]; List 3: [intellectual]; List 4: [property]	10-K Filing
NAIntTheft %	# of paragraphs that contain at least one words from the following word lists divided by the total number of paragraphs of the 10-K filing. List 1: [Mexico, Mexican, Canada, Canadian]; List 2: [protect, infringe, theft]; List 3: [intellectual]; List 4: [property]	10-K Filing
XRD	R&D expenses from Compustat. We replace the missing R&D expense ratio (over sales) by the industry average if the firms has applied for any patents in the past three years. We replace the other missing variables with 0.	Compustat
NPatent	The number of patents that the firm applies in a year. For patents granted prior to Nov. 1, 2010, we use the KPSS data; For patents granted after Nov. 1, 2010, we use the patent data from Google patents.	Google Patent; Kogan, Papanikolaou, Seru, and Stoffman (2016)
PatCiteCN	The total number of new patents that (1) are applied in SIPO (China Patent Office), (2) assigned to a Chinese firm, and (3) cite any existing patents of the firm	Google Patent
PatCiteUS$_{CN}$	The total number of new patents that (1) are applied in USPTO, (2) assigned to a Chinese firm, and (3) cite any existing patents of the firm	Google Patent
PatCiteUS$_{EU}$	The total number of new patents that (1) are applied in USPTO, (2) assigned to an European firm, and (3) cite any existing patents of the firm	Google Patent
PatCiteUS$_{JP}$	The total number of new patents that (1) are applied in USPTO, (2) assigned to a Japanese firm, and (3) cite any existing patents of the firm	Google Patent
PatCiteUS$_{NA}$	The total number of new patents that (1) are applied in USPTO, (2) assigned to a Mexican or Canadian firm, and (3) cite any existing patents of the firm	Google Patent
PatCiteUS$_{US}$	The total number of new patents that (1) are applied in USPTO, (2) assigned to an American firm, and (3) cite any existing patents of the firm	Google Patent
Age	Number of years that the firm has been public	Compustat
CNSalesGR	The average sales growth of the Chinese public company of the same 2-digit SIC industry	CSMAR; Capital IQ
Q	Market to book ratio	Compustat
Sales	Sales of the firm	Compustat
TA	Total asset of the firm	Compustat
AssetTangibility	property, plant and equipment over total assets	Compustat
CNInternet_Macro	The variable is constructed similarly to CNInternet. Instead of using the weights from public firms, we use the industry weights from the total assets information from China Data Online. We hand-matched each industry to 2-digit SIC industries.	CNNIC Reports; China Data Online
CNInternet_Top1	The variable is constructed similarly to CNInternet. Instead of using the value-weighted measure using all the provinces whose weights are above 10%, we put 100% weight on the province with the highest total assets of the industry	CNNIC Reports; Capital IQ; China Data Online

Appendix B. Robustness Tests

Table 12: Robustness - Weights from Macro Data

This table estimates the robustness of our main results using

	CNComp	CNCompHi	CNIntComp	$\frac{XRD}{Sales}$	$\frac{NPatent}{Sales}$	$\frac{PatCiteUS_{CN}}{Sales}$	$\frac{PatCiteCN}{Sales}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CNInternet_Macro	0.198***	0.159***	0.162***	−0.069***	−0.092***	0.343***	0.351***
	(0.044)	(0.044)	(0.043)	(0.021)	(0.032)	(0.048)	(0.042)
log(10kSize)	−0.107***	−0.110***	−0.097***				
	(0.011)	(0.011)	(0.011)				
log(Age + 1)	−0.044*	−0.049**	−0.020	−0.106***	−0.113***	0.075***	0.021
	(0.024)	(0.025)	(0.027)	(0.018)	(0.021)	(0.023)	(0.020)
log(TA)	0.053*	0.069**	0.039	0.034	−0.059**	−0.292***	−0.339***
	(0.029)	(0.027)	(0.031)	(0.027)	(0.029)	(0.035)	(0.034)
Q	−0.013**	−0.012**	−0.014**	−0.003	−0.017	−0.057***	−0.034***
	(0.006)	(0.006)	(0.007)	(0.014)	(0.014)	(0.013)	(0.013)
CNSalesGR	0.001	−0.002	−0.001	0.001	−0.002	−0.002	0.0005
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
N	61,930	61,930	61,930	61,768	61,768	61,768	61,768
Adjusted R^2	0.586	0.523	0.473	0.735	0.724	0.401	0.457

Table 13: Robustness - Top 1 provinces

	CNComp	CNCompHi	CNIntComp	$\frac{XRD}{Sales}$	$\frac{NPatent}{Sales}$	$\frac{PatCiteUS_{CN}}{Sales}$	$\frac{PatCiteCN}{Sales}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CNInternet_Top1	0.116***	0.105***	0.112***	−0.107***	−0.111***	0.260***	0.245***
	(0.034)	(0.033)	(0.032)	(0.027)	(0.030)	(0.040)	(0.036)
CNSalesGR	0.001	−0.002	−0.001	0.001	−0.001	−0.002	0.0001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
log(10kSize)	−0.107***	−0.110***	−0.097***				
	(0.011)	(0.011)	(0.011)				
log(Age + 1)	−0.061**	−0.063**	−0.034	−0.100***	−0.104***	0.046**	−0.009
	(0.024)	(0.025)	(0.027)	(0.017)	(0.020)	(0.022)	(0.020)
log(TA)	0.050*	0.066**	0.037	0.035	−0.058**	−0.298***	−0.345***
	(0.029)	(0.028)	(0.031)	(0.027)	(0.029)	(0.035)	(0.034)
Q	−0.014**	−0.012**	−0.015**	−0.003	−0.017	−0.058***	−0.035***
	(0.006)	(0.006)	(0.007)	(0.014)	(0.014)	(0.013)	(0.013)
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
N	61,930	61,930	61,930	61,768	61,768	61,768	61,768
Adjusted R^2	0.586	0.523	0.472	0.736	0.725	0.401	0.456

Table 14: Robustness of Table 5 Excluding Zero R&D Firms

This table tests the robustness of Table 5 by using subsample excluding observations where XRD/Sales equals 0. The dependent variable for Columns (1) - (3) is the R&D expenses over sales. For missing R&D, we follow the Koh and Reeb (2015) and replace the missing with industry average if the firm files for any patent patents applications in the past three years (including the current year), and 0 otherwise. The dependent variables are measures from 1, 2, or 3 years in the future. The dependent variable in Columns (4) - (6) is the total number of patent applications each year (by filing date) dividend by sales. The patent data comes from Google Patents, and we match the patents to Compustat firms using the links from Kogan, Papanikolaou, Seru, and Stoffman (2016). The dependent variables are measures from 1, 2, or 3 years in the future. The key independent variable CNInternet is the Chinese internet penetration ratio. All independent variables are one-year lagged relative to the dependent variables. All the variables are normalized by their standard deviations for easier interpretation. The sample covers all Compustat firms from 2003 to 2015. We exclude all observations where the total asset or the sales are smaller than one million dollars. Robust standard errors clustered by firms are reported in the parentheses. Detailed definitions of the variables can be found in Table 11 in the Appendix.

	XRD/Sales			NPatent / Sales		
	t+1	t+2	t+3	t+1	t+2	t+3
	(1)	(2)	(3)	(4)	(5)	(6)
CNInternet	-0.354^{***}	-0.325^{***}	-0.268^{***}	-0.156^{*}	-0.133^{*}	-0.092
	(0.082)	(0.076)	(0.071)	(0.084)	(0.080)	(0.078)
CNSalesGR	0.0004	0.008	0.004	-0.007	-0.0003	0.0002
	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
log(Age + 1)	-0.271^{***}	-0.213^{***}	-0.163^{***}	-0.310^{***}	-0.258^{***}	-0.210^{***}
	(0.046)	(0.046)	(0.049)	(0.055)	(0.055)	(0.057)
log(TA)	0.014	0.052	-0.001	-0.158^{***}	-0.093^{*}	-0.095^{*}
	(0.054)	(0.052)	(0.053)	(0.058)	(0.055)	(0.051)
Q	-0.015	0.004	0.007	-0.037^{**}	-0.018	-0.017
	(0.020)	(0.017)	(0.015)	(0.019)	(0.018)	(0.018)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	27,950	24,131	20,913	27,950	24,131	20,913
Adjusted R^2	0.738	0.731	0.733	0.709	0.751	0.785