Urban Agglomeration and Firm Innovation

Evidence from Developing Asia

Liming Chen

Rana Hasan

Yi Jiang

WORLD BANK GROUP

World Bank Economic Review *accepted* Paper

November 2021

Abstract

This paper examines the relationship between urban agglomeration and firm innovation using a recently developed dataset that consistently measures city boundaries across Asia together with geo-referenced firm-level data. It finds that the spatial distribution of innovation by firms is highly concentrated within countries. Further, firms in larger cities have substantially higher propensities to introduce product and process innovations and undertake R&D

activities, a result that holds for subgroups of countries and even when the largest cities are excluded from the analysis. Finally, the presence of high quality universities and highly ranked engineering departments in cities is positively associated with firm innovation, lending support to the idea that the accumulation of human capital locally is a key channel through which urban agglomeration affects innovation.

This paper has been accepted for publication in *the World Bank Economic Review*. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at atlimingchen@adb.org, rhasan@adb.org, and yijiang@adb.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Urban Agglomeration and Firm Innovation: Evidence from Asia

Liming Chen, Rana Hasan, and Yi Jiang¹

Keywords: agglomeration economies; firm innovation; R&D; knowledge spillovers

JEL codes: O3, R11, O1

¹ Liming Chen: Young Professional in the South Asia Department, Asian Development Bank (ADB), Mandaluyong City 1550, Metro Manila, Philippines. E-mail: limingchen@adb.org; Rana Hasan: Regional Economic Advisor in the South Asia Department, ADB, Mandaluyong City 1550, Metro Manila, Philippines. E-mail: rhasan@adb.org; Yi Jiang: Principal Economist in the Economic Research and Regional Cooperation Department, ADB, Mandaluyong City 1550, Metro Manila, Philippines. E-mail: yijiang@adb.org. The authors thank Gilles Duranton for detailed comments and suggestions on an earlier draft, and participants at the 2019 SMU Conference on Urban and Regional Economics, ADB Economics Seminar Series, and ADB Economists' Forum for constructive feedback. We are grateful to the Enterprise Analysis Unit of the World Bank for providing geomasked GPS coordinates for the firms in the World Bank Enterprise Survey data used in our analysis. Patricia Thea Basilio, Ma. Adelle Gia Toledo Arbo, Marjorie Villanueva Remolador, and Renz Adrian T. Calub provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of ADB or its Board of Governors or the governments they represent. ADB does not guarantee the accuracy of the data included in this paper, and accepts no responsibility for any consequence of their use.

1. Introduction

Innovation undertaken by firms is regarded as a significant contributor to long-term economic growth in both theory and in practice (see, for example, Romer 1986). One issue that has attracted considerable attention of researchers is on the geographical distribution of firm-level innovation within countries. A large body of evidence from the developed world shows that innovative activities, such as R&D and patenting tend to be more concentrated spatially than production. This pattern has raised various questions, such as whether urban agglomerations play a special role in fostering innovation by firms, and if so, what channels are at work. While evidence on these questions is growing, most of it has been obtained from the developed world. For developing countries, there is a lack of systematic evidence on the spatial distribution of firms' innovation-related activities and whether and how urban agglomeration affects innovation.²

This paper attempts to fill this gap by utilizing geo-referenced firm-level data from World Bank Enterprise Surveys (WBES) conducted in 25 developing Asian economies from 2012 to 2016 and a recently developed dataset that uses nighttime lights (NTL) imagery to consistently measure city boundaries across Asia (Jiang, 2021). In addition to production related data, the WBES collects information on firms' innovation related activities—specifically, whether surveyed firms introduced any product and process innovations and invested in research and development (R&D). As regards our city-level dataset, its urban units are referred to as “natural cities” to distinguish them from administratively defined cities. In comparison with the latter, natural cities are consistently defined and measured across countries. They are also more likely to capture the actual economic “footprint” of a city than its administrative boundaries would suggest. By mapping WBES firms onto the natural city dataset, the relationship between firm innovation and urban agglomeration in developing Asia can be examined. Further, an international ranking of universities and academic disciplines or major fields of study is utilized to investigate whether the presence of high-quality universities has any bearing on firm innovation.

This paper finds that innovative activity by firms is concentrated spatially within several large Asian countries. The share of firms that innovate in just a few cities is substantially greater than the share of urban population these cities account for. Further, the data shows that the population size of a natural city is strongly correlated with the propensity of firms in the city to engage in innovative activities. Specifically, the estimates suggest that larger city size is significantly associated with increases in the propensity of firms to introduce product and process innovations and undertake R&D. A doubling of city size is associated with increases in firms' propensity to engage in product innovation, process innovation, and R&D by 4.3, 3.7, and 2.8 percentage points, respectively, or by 13.5%, 8.5%, and 14.3% against the respective benchmark propensities. Since the biggest natural city is much larger than the smallest one in the data, the effects of agglomeration on firm innovation can be quite large.

² Most research from the developing world focuses on agglomeration effects concerning labor and firm productivity. Duranton (2008), Overman and Venables (2005), Henderson (2005), and Duranton (2014) provided detailed review of this research.

This paper also explores possible heterogeneity in agglomeration effects by the development level of countries and between countries and subregions with distinct historical and institutional backgrounds. The results show that agglomeration effects on firm innovation are prevalent across developing Asia; they are present in both upper-middle income and low and lower-middle income country groups, as well as in China, India and Southeast Asian countries, each considered separately.

Finally, using a widely cited ranking of universities and majors (QS 2019a and QS 2019b), it is possible to identify which cities are home to one or more of the top 500 ranked Asian universities as well as universities with a major in the top 500 global ranking of major fields of study. The presence of a top university is found to be positively and significantly associated with firm innovation, and having a university with a top 500 globally ranked engineering major even more so. These results indicate an important role for locally available human capital, especially that which relates to technical fields, as a channel through which agglomeration affects innovation.³

It must be noted that the analysis here does not estimate the causal effects of city size on innovation. To do so would require instruments that affect current city size but are unrelated to unobserved city characteristics that impact innovation. Given the difficulties in finding plausible instruments, especially in a cross-country setting, this paper focuses on establishing important associations and exploring potential heterogeneity in the relationship between city size and firm innovation while controlling for various city and firm level characteristics.

This paper is closely related to two strands of literature. The first is on the measurement of urban agglomeration economies. Most of the empirical evidence in this strand has focused on whether city size and/or density affect firm and worker productivity.⁴ Studies examining urban agglomeration economies associated with innovation-related activities are limited. Nevertheless, the available evidence—mostly for developed countries—suggests that innovation in a given country is heavily concentrated in a few cities, and that a city’s size or density plays a role in boosting innovation activity within that city. Ó hUallachain (1999) find that large metropolitan areas in the US dominate patenting, with patenting increasing with the size of the metropolitan area. Feldman and Audretsch (1999) also find that innovation appears to be a phenomenon of large cities: less than 4% of innovations occurred outside of the metropolitan areas in which 70% of the country’s population reside. Carlino, Chatterjee, and Hunt (2007) estimate the elasticity of patenting per capita with respect to employment density to be approximately 20% in metropolitan areas in the US. Carlino et al. (2009) further confirm the earlier results by weighting patents with the number of times they are cited. In a more recent study covering 30 OECD countries, Paunov et al. (2019) find that 64% of patent applications from 2010 to 2014 are

³ There are two specific ways universities influence innovation via their effect on the stock of human capital. First, they expand the supply of skilled workers who can undertake innovation in a given location. Second, they may influence innovation through the research of faculty via spillovers of knowledge or directly via technical collaborations with firms. For both, the evidence from developed countries suggests a strong local effect.

⁴ See Rosenthal and Strange (2004), Melo, Graham, and Noland (2009), Puga (2010), and Baum-Snow and Ferreira (2015) for extensive surveys of methodology and evidence.

concentrated in 10% of cities. Evidence on other types of innovation such as firm-level innovation inputs or outputs remains largely missing in the literature.

The second strand of literature this paper relates to attempts to identify the mechanisms underlying agglomeration effects. Duranton and Puga (2004) propose three channels through which urban agglomerations influence firm and labor productivity: sharing, matching, and knowledge spillover. Carlino and Kerr (2015) review the evidence and note that these three mechanisms also apply to innovation, with more empirical support for the knowledge spillover channel.⁵ Knowledge spillovers refer to the intellectual gains made by the exchange of information for which no direct compensation is given to the producer of the knowledge. It is particularly important in explaining the concentration of innovation since innovation arguably depends on the dispersal of knowledge more than other economic activities.

A set of studies closely related to this paper examine the relationship between university-firm colocation and firm innovation and typically find significant effects working through various channels (Jaffe 1989; Audretsch and Feldman 1996; Audretsch and Stephan 1996; Anselin, Varga, and Acs 1997; Andrews 2017; Paunov, Borowiecki and El-Mallakh, 2019). Relatedly, a handful of studies show that academic disciplines or majors belonging to science, technology, engineering and mathematics (STEM) are especially important for such spillover linkages (Bianchi and Giorcelli 2020).

Although both strands of the literature have developed quickly, the evidence mostly draws from the US and other advanced economies. Evidence from developing countries is still extremely thin (Duranton, 2014). One exception is Nieto Galindo (2007), who documents that a significant share of firms in Colombia engage in product and process innovation, with more than 70% of these innovations concentrated in three main cities hosting nearly 40% of the country's population.

To the best of our knowledge, this paper is the first to provide cross-country evidence from the developing world on the relationship between urban agglomeration and firm innovation and how universities may influence it.⁶ While firms in developing countries may not be at the global technological frontier, they face the challenge of surviving local and imported competition. Innovation, though not at the cutting edge globally, may still be critical for firms in developing countries. Moreover, continuous investment in R&D and innovation can enable firms in developing economies to catch up with the frontier (Acemoglu, Aghion, and Zilibotti, 2006). In

⁵ A thick labor market allows for the efficient sharing of access to a pool of specialized and experienced workers, which creates linkages and networks for knowledge to flow rapidly (Helsley and Strange 2002). Gerlach, Ronde, and Stahl (2009) find that firms located in clusters also take greater risks in R&D choices because of this, compared with spatially isolated firms. A thick local labor market also improves the quality of matches between firms and workers (Helsley and Strange 1990; Berliant, Reed, and Wang 2006; Strange, Heijazi, and Tang 2006).

⁶ A recent study that has some similarities is Valero and Van Reenen (2019), which examines the impact of universities on subnational region-specific GDP per capita using data from 78 developed and developing countries and finds university presence to be robustly associated with regional economic growth. The study also considers whether the quality of universities (as proxied by the presence of a PhD program and courses in STEM and “professional services”) drives the university-growth relationship and finds this to be the case only for an “advanced country” subsample. This is taken to suggest that the innovation channel is a more important driver of the university-growth relationship in countries nearer the technology frontier.

this context, understanding what role urban agglomeration plays in firms' undertaking product and process innovations as well as investing in R&D is very relevant.

The remainder of this paper is organized as follows. The next section introduces the data. Section 3 describes the spatial distribution of innovation activities across developing Asian countries. Section 4 describes the empirical approach taken and presents the main results regarding agglomeration effects and firms' innovation-related activities, followed by robustness tests and various extensions. Section 5 investigates the colocation of top university and firms as a channel through which agglomeration effects may be working. The final section concludes.

2. Data

This paper's analysis is based on a unique dataset that maps over 20,000 firms to nearly 500 cities across 25 developing economies in Asia. The firm data comes from the WBES conducted from 2012 to 2016.⁷ The city data was developed using nighttime lights (NTL) satellite imagery, LandScan grid population data and other data sources.

2.1. Data on Firm Innovation

WBES are firm-level surveys conducted in developing economies. Each survey consists of a cross-sectional representative sample of private firms mainly from an economy's formal sector. Those in the agriculture sector and all state-owned enterprises are excluded. Surveyed firms are selected through stratified random sampling, based on the sector of activity and firm size. Since 2002, the data have been collected from face-to-face interviews with high-level managers or company owners with measures taken to ensure high quality of surveys.⁸

The surveys cover a wide range of topics, with some of them country specific. A harmonized version of the WBES data that extracts and standardizes common information from each country is used. Apart from basic information on sample firms—such as number of employees, age, sector, and foreign direct investment (FDI) share—the surveys collect information about firms' innovation-related activities through the following questions:

- (i) During the last 3 years, has this establishment introduced a new or improved product (i.e., a good or service)?
- (ii) During the last 3 years, has [the surveyed] establishment introduced any new or improved process? These include methods of manufacturing products or offering services; logistics, delivery, or distribution methods for inputs, products, or services; or supporting activities for processes.

⁷ The terms firm and establishment are used interchangeably while recognizing that the unit of analysis of the WBES are establishments.

⁸ For instance, the firm identifiers are kept confidential and the survey is conducted by private organizations independent from government. To make sure questions elicit valid answers, various measures including translation and localization checks, different interviewees for the same firm, and post-survey consistency checks are taken before, during and after the survey. See Ayyagari, Demirgüç-Kunt, and Maksimovic (2011) for more discussion.

(iii) During the last fiscal year, did this establishment spend on research and development activities, either in-house or contracted with other companies, excluding market research surveys?

The answers to these questions, coded as binary indicators, measure a firm's engagement in innovation with respect to products, processes, and R&D, respectively. As defined in the questionnaire, these innovations can be more advanced inventions or production processes, or small improvements to existing processes and products.⁹

To address the potential subjectivity of responses to these questions, WBES data are collected by interviewers guided by detailed instructions on what does or does not count as a "new or improved" product or process and as "R&D".¹⁰ Firms responding positively to questions (i) or (ii) were asked to describe the new or improved product or process in detail. (In case of multiple innovations by a firm, the respondent was asked about the product or process considered to be most significant as identified by the value of sales for product innovations and impact on operations for process innovations.) Further, firms responding positively to question (i) were asked whether the new or improved product (good or service) was also new for the establishment's main market.

These questions should prompt respondents to exercise care in providing responses. In view of the above, as well as the general quality control measures adopted in collecting WBES data, the innovation information in WBES is deemed to be reliable. Indeed, it has been used in several peer-reviewed papers that study firm innovation in developing countries (Ayyagari, et al., 2011; Paunov, 2016; and Paunov and Rollo, 2016).¹¹

To examine how urban agglomeration affects firm propensity to undertake innovation-related activities, it is crucial to know the location of firms, ideally measured in a consistent way across countries. Publicly available WBES provide information on the geographic units where each firm is located. However, the scope of the geographic units varies by country and these units do not necessarily correspond to a well-defined urban area. For example, the Kazakhstan data indicate only the region (e.g., north or south) to which surveyed firms belong, while the Indian data reveal only the state where a firm is located. Even if the geographic unit refer to cities, as is the case for some countries, they are not comparable since the official definitions of cities differ greatly across countries (UN 2018). Fortunately, the WBES conducted since 2012 contain the

⁹ Examples of relatively minor product innovations include a shoe manufacturing firm in Bandung, Indonesia introducing shoes made out of nubuck leather with anti-slip and oil resistant soles, or a plastic bag printing firm in Port Moresby, Papua New Guinea printing bags to carry food. Examples of process innovations include a manufacturing firm in Yangon, Myanmar offering a door-to-door sales service or a repair shop running an ad campaign on Facebook. There are more advanced product innovations in the sample as well, including a firm in Ho Chi Minh starting to weld optical fibers for Viettel, a telecom multinational. For process innovation, an example is an auto repair company in Yangon adopting laser-based technology.

¹⁰ See pages 16-20 of "World Bank Group's Enterprise Survey: Understanding the Questionnaire", 2019, World Bank.

¹¹ Robustness tests use as dependent variables an indicator of whether the new or improved product or service was also new for the establishment's main market and an indicator of whether the establishment reported either a product innovation or process innovation. The latter allows for mistaken categorization of innovation as a product or process innovation by respondents. See Section 4.2 for more detail.

geographic coordinates of surveyed firms. These were provided with a random shift of up to 2 kilometers to mask firms' true locations and allow firms to be matched with natural cities as described below.

2.2. Data on Natural Cities Based on Nighttime Lights Satellite Imagery

A novel dataset of more than 1,500 “natural” cities across Asia and the Pacific developed recently (Jiang, 2021) is utilized for capturing cities. The dataset is based on nighttime lights (NTL) satellite imagery data available since 1992 from the National Oceanic and Atmospheric Administration website. Human settlements were delineated with the deblurred NTL data as contiguously illuminated areas. Those meeting certain criteria, such as covering units captured by the Global Rural Urban Mapping Project database with population greater than 100,000 in year 2000 or having a land area greater than 100 km² in 2000 are classified as urban agglomerations. These urban agglomerations are referred to as natural cities to distinguish them from administratively defined cities.

The population of natural cities is estimated using grid population data from LandScan. Other city characteristics including weather (precipitation and annual maximum and minimum temperature), distance to the nearest seaport, and citywide ruggedness are added to the dataset. Jiang (2021) provides additional details on the development of the natural city dataset.

The advantages of using natural cities are many. First, natural cities are uniformly defined and their characteristics are measured consistently across countries and over time. In contrast, official statistics on cities rely heavily on administrative designations, for example, defining cities as areas administered by municipal councils, or combining administrative designations with economic ones, for example, requiring a city to meet some country-specific thresholds on population size and density and even share of agricultural employment (ADB 2019). As may be expected, the boundaries of administratively designated cities differ substantially in their size, institutional status and governance structure across different countries. Second, natural cities offer a better representation of the urban agglomerations in which firms operate. This is because many administratively defined cities contain both dense urban areas and sparsely populated rural areas; additionally, economic activity invariably expands beyond the administrative boundaries of cities and spreads over multiple administrative units (urban and/or rural). These areas are economically integrated and should be considered as part of an urban agglomerations.¹²

To illustrate these points, Figure 1 shows two examples of natural cities and their corresponding administratively defined units. In India, officially designated cities are described as “statutory towns”—i.e., administrative units with a municipal corporation, municipality, cantonment board, notified town area committee, or town council. However, the boundaries of these cities are widely acknowledged as too restrictive, given the complex and time-consuming process of redrawing municipal boundaries as cities expand (Colmer 2016). Thus, the Indian

¹² The literature increasingly recognizes the advantage of defining urban agglomerations through economically integrated areas via remote sensing technologies instead of administrative boundaries. For applications to the developing world, see Dingel, Miscio, and Davis (2019), Galdo, Li, and Rama (2019), and Bosker, Park and Roberts (2020).

census has defined additional urban units, such as “census towns”, administrative units formally classified as villages but which satisfy various criteria common to urban settlements; and “urban agglomerations” (UA), typically an urban area that consists of a statutory town and its adjoining outgrowths. While the concepts of census town and UAs try to capture the urbanization process taking place outside the administrative boundaries of statutory towns, it is widely agreed that these understate the extent of urban growth. Nighttime lights lend strong support to this. Panel (a) of Figure 1 plots three versions of Chennai: the statutory town (dashed red boundary), the UA (dashed black boundary), and the natural city (the shaded area). Further, towns in the vicinity are also plotted. The figure serves to show that urban spread can well exceed the administrative boundaries of officially defined cities and encompass many settlements—towns and villages—which are otherwise treated as wholly separate entities in official statistics.

China presents a different case. In the literature and policy documents, prefectures (nearly 300 in total) and four municipalities directly under the central government are considered administrative cities in China (e.g., Chauvin, Glaeser, and Tobio, 2017). However, there are vast suburban and rural areas in each prefecture. In the official City Statistical Yearbooks, data are reported for both the entire prefecture as well as a core area, called the “city proper”. The latter is often used in studies of Chinese cities (e.g., Anderson and Ge, 2005; Au and Henderson, 2006; Fang, Li and Song, 2017). Panel (b) of Figure 1 shows that Dalian prefecture, known for its seaport and role as an economic center in northern China, contains four natural cities with one located in the city proper and three outside. As the case of Dalian prefecture shows, a focus on the city proper can simultaneously lead to the inclusion of large amounts of rural areas as well as the exclusion of other urban areas inside the prefecture.¹³

The natural city data contains both geocoded locations and boundaries of the natural cities, making them ready for matching with WBES firm-level data. Other publicly available city datasets of developing countries such as the one from World Urbanization Prospects only contain centroids of cities.

WBES firms are mapped to their natural cities using geographic information system (GIS) software. To maximize the number of firms that can be matched to our natural cities, 2016 natural city boundaries are used and firms falling outside a natural city but within a 2km band from the city boundary are assigned to that city.¹⁴ Table A1 shows the results of the matching exercise by country. Overall, a total of 21,857 firms are matched to 489 natural cities, an average of 45 firms per city. The share of WBES firms matched to a natural city is 87%, with the matching rates within individual countries ranging from 40% (Mongolia) to 100% (Myanmar and the PRC).¹⁵ The unmatched firms are likely to be those located in rural areas of very small

¹³ Appendix 1 presents further comparison of the two types of city data by an investigation of Zipf’s law for China and India.

¹⁴ Results do not change if firms falling outside the 2 km band from the analysis are excluded. These are available upon request.

¹⁵ The initial procedure resulted in a low matching rate for the PRC (59%), which is probably due to the country’s regulations on collecting and using geographic information, and the unique coordinate system used for geographic data in the PRC. To address the issue, prefecture information available in the WBES data for the PRC is used to assign each unmatched firm to the largest natural city in its prefecture. This results in a 100% matching rate.

towns far from urban centers. Figure 2 shows the geographic distributions of the natural cities with WBES firms. India alone contributes 8,100 firms, or 37% of the total sample, which were distributed across 207 natural cities.

Finally, an international ranking of universities and academic disciplines or majors is used to examine the relationship between firms' propensity to undertake innovative activities and the presence of high quality universities and underlying academic disciplines. Specifically, the list of top 500 universities in Asia from the recent QS Asia University Rankings (QS 2019a) publication are mapped to natural cities.¹⁶ Of the 500 top Asian universities, 248 are mapped into 99 natural cities. A dummy indicating the presence of top Asian university is set to 1 for these cities and 0 for the rest.

Unfortunately, the QS rankings of top Asian universities does not provide information on their majors and which ones contributed to their high ranking. To get a sense of the role that different majors might play in any linkages between universities and firms' innovation, information from the QS ranking of universities around the world across five broad areas of study (QS World University Rankings by Subject; QS 2019b) is used.¹⁷ The five areas are arts and humanities, engineering, life sciences, natural sciences, and social sciences and management. This global ranking is used to create dummies for the presence in our cities of universities with a top (globally) ranked major across the five areas. If the city has one or more universities ranked within the top 500 in a major study area, the corresponding dummy is equal to 1.

As may be expected, having a university with a top ranked major globally makes it very likely that a university is also be in the list of top Asian universities. Indeed, with the exception of eight Indian universities located in 6 natural cities, all other Asian universities included in the list of universities with a top major(s) globally are also in the list of top Asian universities. On the flip side, there are 171 top Asian universities without any major in the global top 500 majors.

2.3. Summary Statistics

Table 1 reports the sample summary statistics of firm-level and city-level variables used in the study. Around 36% of the surveyed firms reported that they had introduced new or improved products or services in the past 3 years; 21% introduced products that were also new for the firm's main market. The proportion goes up to 50% for new or improved processes, and 24% for any expenditure on R&D. Overall, 56% of firms carried out at least one of the three innovation related activities.

Following the literature, firms that operated for fewer than 10 years are defined as young firms (Reyes, Robert, and Xu 2017); those that had 50 or fewer permanent employees as small firms (Beck, Demirgüç-Kunt, and Maksimovic 2005; Reyes et al. 2017); and those that had 10% or more foreign ownership as FDI firms.¹⁸ About 21% of firms across the whole sample were

¹⁶ QS stands for Quacquarelli Symonds, a British company that analyzes higher education institutions around the world.

¹⁷ The detailed list of subjects included in each major category is accessible at the QS website.

¹⁸ The literature often follows the International Monetary Fund's standard definition of FDI, i.e., a firm is treated as having FDI if at least 10% of the firm's ownership is foreign. For example, see Kinda (2010).

young, 65% were small, and 6% had FDI present. About two thirds of firms are from the manufacturing sector, with the rest belonging to the services sector. About 41% of the surveyed firms are headquarters and the average share of skilled workers, defined as workers that perform highly skilled or semi-skilled jobs, is around 35%.

To capture time-variant characteristics of natural cities, such as population and weather, 2010 values are used in light of the fact that the innovation activities captured by the surveys were concentrated over 2012–2016. Overall, the sample cities are highly diverse in all the dimensions measured. City population ranges from 8,000 to 35.8 million, with an average of 1.4 million and median of 438,000. The average annual maximum and minimum temperatures are 35 and 9 Celsius degrees, respectively, with minimum temperatures having a larger variation. The average distance of the cities to the nearest seaport is 400 km, with a standard deviation of 500 km. The city with the most rugged terrain has an index 15 times greater than the sample mean for ruggedness.

About 20% of the 489 cities host at least one university that ranks among the top 500 in Asia. Further, based on a dataset that ranks academic disciplines or majors globally, 4% cities also host at least one university whose arts and humanities major ranks among the top 500 globally. The corresponding numbers for other majors are 6% for engineering and technology; 4% for life science and medicine; 6% for natural sciences; and 5% for social sciences and management.

3. Spatial Distribution of Firm Innovation

One stylized fact documented in the literature—which mostly covers developed countries as noted earlier—is that innovation is spatially concentrated to a high degree. Within a given country, there are generally just a few cities that act as “innovation hubs”. These hubs host a large share of a country’s innovative activities, and this share is typically disproportionate to each city’s share of the country’s total population. For instance, Moretti (2019) shows that in the US the top 10 cities accounting for innovations in the fields of computer science, semiconductors, and biology and chemistry host 70%, 79% and 59% of inventors, respectively. Similarly, Nieto Galindo (2007) finds that over 70% of Colombia’s innovations are concentrated in 3 main cities, which together host less than 40% of the country’s population.

Similar patterns with respect to firm innovation are observed in our data. Figure 3 plots the share of innovative firms a natural city accounts for in the country against its share in total urban population for each of the three types of innovation. The natural cities covered are the most innovative cities—the top 10 in the case of China and India, the top three for other countries with more than 10 natural cities, and only the top city for the rest. The left panel displays the plots for China and India while the right panel covers other countries. Cities falling above the 45-degree dashed line have a higher degree of concentration of innovating firms relative to population.

A strong pattern of concentration in spatial innovation is observed in both China and India. For China, with the exception of Beijing for product innovation, all top innovative cities have larger shares of innovative firms than their shares of urban population. Hangzhou, where the headquarter of Alibaba company is located, accounts for 2.6% of total urban population and 4.7%-6.1% in the three firm innovation activities. Suzhou, known for the success of its industrial

park set up in collaboration with the Singaporean government, accounts for 1.6% of urban population, 8.3% of firms undertaking R&D and 10.9% of process innovation. Jointly, the top 10 cities in China account for 71% to 75% of firms engaged in innovation activities with 45% to 55% of urban population.

In India, the top 10 innovative cities are all above the 45-degree line with the exception of Mumbai and Kolkata. Delhi, Chennai, Bengaluru, and Hyderabad host the most innovative firms in India across different type of innovation related activities. Moreover, they display significant gaps between shares of firm innovation and shares of urban population. For instance, 26% of firms undertaking R&D in India are located in Delhi, which accounts for 9.7% of urban inhabitants. Overall, 72% to 79% of innovative firms are located in the top 10 cities in each category while the population of these cities account for 34% to 43% of the urban total nationally.

Similar patterns emerge for the remaining countries. For all three categories of innovation, two-thirds of the top innovative cities in these countries are above the 45-degree line. However, the figure also shows some mega cities with a disproportionately low share of innovative firms. For instance, Almaty, Bangkok, Ho Chi Minh, Jakarta, Kabul, Kathmandu, Manila, and Viangchan all have larger shares of national urban population than their shares of innovative firms in product and process innovation. Interestingly, these cities except Ho Chi Minh are all national capitals.

The spatial concentration of firm innovation at the country level can also be measured as the ratio of the Herfindahl-Hirschman Index (HHI) of firm innovation to the HHI of population at the city level.¹⁹ The results are reported in Table 2. A ratio higher than 1 suggests a higher degree of spatial concentration of firm innovation relative to the urban population. The country-level HHI ratios range between 0.59 and 3.11, 0.53 and 2.68, and 0.67 and 4.12 with simple averages of 1.28, 1.29 and 1.51 for product innovation, process innovation, and R&D, respectively. Across all three innovation categories, two thirds of sample countries have HHI ratios higher than one. Notably, India has the highest degree of concentration for product innovation and R&D and is second to Uzbekistan for process innovation. On the other hand, high population Southeast Asian countries such as Thailand, Viet Nam and Philippines score low in the spatial concentration of firm innovation.

In summary, there is compelling evidence of a high degree of spatial concentration of firm innovation across Asian countries. Nevertheless, some large cities are seen to host disproportionately fewer innovative firms. How is city size associated with firms' innovative activities in developing Asian countries? What are potential channels through which agglomeration may be affecting firms' innovation related activities? These are questions examined in the next two sections.

4. Urban Agglomeration and Firm Innovation

¹⁹ HHI is calculated as $HHI_c = \sum S_i^2$, where S_i is the share of population or firms engaged in innovation of natural city i in country c . We thank a referee for suggesting this measure of relative spatial concentration of firm innovation.

This section aims to investigate the association between urban agglomeration and a firm's propensity to undertake product innovation, process innovation, and R&D activity across Asian cities. Given the nature of the data at our disposal, a Probit model is used to estimate the likelihood of a firm partaking in a particular type of innovation activity as a function of characteristics of the city in which the firm is located as well as those of the firm itself. The underlying latent model can be written as:

$$y_{icnt}^* = \beta_0 + \beta_p \ln P_{cn} + \beta_c CITY_{cn} + \beta_f FIRM_{icnt} + \delta_n + \gamma_t + \epsilon_{icnt}$$

$$Y_{icnt} = \begin{cases} 1 & \text{if } y_{icnt}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where: subscripts i, c, n, t represent firm, city, country, and year, respectively; Y_{icnt} denotes the outcome variable of interest, which equals 1 if the firm responded positively to the question regarding product innovation, process innovation, or R&D expenditure, and 0 otherwise; y_{icnt}^* is the latent variable for outcome variable Y_{icnt} , which can be interpreted as a firm's (unobserved) propensity to innovate; and $\ln P_{cn}$ is the natural log population of city c in 2010, two to six years prior to the actual survey of firms in a country. Of primary interest is its coefficient, β_p . A positive and statistically significant estimate suggests that firms in larger agglomerations are more likely to undertake innovation activities. For better interpretation of the Probit results, the marginal effect on innovation propensity from changes in city population are shown.

$FIRM_{icnt}$ is a vector of firm-level controls, including indicators for firms' age, size, presence of FDI, sector of production, share of skilled workers, and whether the responding establishment is a headquarter or not. These variables are considered to be important influences on a firm's decision to undertake innovation. A number of "first-nature" city characteristics, denoted by $CITY_{cn}$, that can affect both city population and the propensity of firms to innovate in the city are also controlled for (Combes et al. 2010). These are average rainfall, annual maximum and minimum temperature, distance to the nearest port, and average terrain ruggedness. These variables are measured in 2010.

Controls for country (δ_n) and year (γ_t) fixed effects are also included in the model. Country fixed effects can account for cross-country variations in institutions, culture, and nationwide policies that influence firm innovation. They can also capture survey-related differences across countries. For instance, training of interviewers or heterogeneity in respondents' perception of what constitutes innovation may lead to systematic differences in survey responses between countries. The year fixed effects can proxy for time-variant region-wide economic conditions that may affect the dynamics of firm innovation across countries.

Lastly, ϵ_{icnt} denotes a random error term that follows a normal distribution. To be conservative, standard errors are clustered at the country level to account for potential correlation of unobservables within cities as well as across cities of the same country.

Despite the comprehensive controls used, estimates of β_p should not be interpreted as capturing the causal effect of city size on firms' innovation. Two types of endogeneity seem

particularly important when estimating the relationship between firms' innovation and city size.²⁰ In one, entrepreneurs do not choose among alternative locations when starting a firm; but once started by a resident entrepreneur, a firm's subsequent decisions on innovation are dependent on city size as well as unobserved city characteristics such as the quality of the local workforce and public resources—characteristics which may also be correlated with city size. Alternatively, firms may be free to choose their locations, with more innovative entrepreneurs likely to locate in larger cities in order to access their (typically) larger pool of skilled workers, better infrastructure, and thicker markets. In both cases, but especially in the second, the common practice of using historical city size as an instrument for current city size is not likely to be helpful in solving the endogeneity problem. Instead, one would need at a minimum, instruments capturing recent immigration, fertility, or mortality shocks occurring in cities—i.e., factors that would affect current city size but be unrelated to city characteristics that impact innovation. More realistically, given that location choice is unlikely to be completely exogenous, valid instruments would need to capture the choice of entrepreneurs across cities but be uncorrelated with firms' innovation.

Given the difficulties in finding plausible instruments, the analysis focuses on the correlation and possible heterogeneity in the relations between city size and firm innovation controlling for various city and firm level characteristics.

4.1. Baseline Estimates

Table 3 presents the baseline results. The coefficients of the main variable of interest, logarithm population in 2010, are positive and significant at 1% for all three outcome variables. In other words, holding everything else constant, firms residing in cities with larger populations are more likely to implement both product and process innovations and undertake R&D. For product innovation, the predicted propensity when the independent variables are at their mean values is 31.4%. The coefficient corresponds to a marginal effect of 4.3 percentage points, or 13.5% higher from the predicted propensity. For process innovation, the estimated marginal effect is 3.7 percentage points, or 8.5% higher than the predicted propensity of 42.8%. For R&D, the estimated marginal effect is 2.8 percentage points, or 14.3% higher than the predicted propensity.

Other city characteristics are also correlated with innovation-related activities. The coefficients on average precipitation are negative and significant; coefficients on maximum temperature are positive and significant in models for process innovation and R&D; coefficients on minimum temperature are positive and significant; lastly, firms based in inland cities and/or cities with more slopes are, on average, more likely to innovate.

In terms of firm characteristics, younger firms are more likely to implement process innovations and conduct R&D. Firm size is positively associated with all three innovation-

²⁰ We are grateful to a referee for this discussion on the requirements of an IV strategy. It may be noted that in an earlier version of the paper (Chen, Hasan and Jiang, 2020), the 1950s population of administratively defined cities corresponding to each natural city (sourced from the World Urbanization Prospects database) was used to instrument for the contemporary population of our natural cities. The IV estimates are largely in line with the results from Probit models.

related activities. Smaller firms are less likely to participate in product or process innovation or invest in R&D. Firms with foreign ownership are more likely to generate product innovations and invest in R&D. Headquarter firms are more likely to implement product innovations. These findings are generally consistent with literature (Bertschek 1995; Huergo and Jaumandreu 2004; Yang 2017).

To summarize, after controlling for a wide range of factors, associations between city population size and each of the three innovation activities undertaken by firms are found to be statistically significant. All else equal, firms in cities twice as large are more likely to engage in product innovation, process innovation, and R&D by 13.5%, 8.5% , and 14.3% respectively. These estimates are either at the high end or greater than estimated elasticities of wage or firm productivity with respect to city employment or urban density in studies of individual countries, and which typically fall between 2% and 10% (Duranton 2014).²¹ On the other hand, the estimates are smaller than the estimated elasticity of patent intensity with respect to employment density, such as the estimate of 20% found in Carlino et al (2007).

4.2. Robustness Tests

A series of robustness tests are performed. First, the results above might be driven by a few megacities that host a large number of innovative firms. To check for this potential “superstar” or outlier effect, the 13 cities with populations of over 10 million are excluded.²² The results are presented in the first three columns of Table 4. The estimation yields larger coefficient estimates and marginal effects than the baseline results. On average, firms in cities twice as large are more likely to engage in innovation by 4.8, 7.3, and 5.1 percentage points, or by 17%, 17%, and 30% against the benchmark propensity of 28%, 42% and 17%, for product innovation, process innovation and R&D investment, respectively. This suggests that while the megacities host many innovative firms, the association between city size and firm innovation activities is more pronounced for firms located in non-mega cities.

Second, it is possible that firms in small cities innovate much less than those in large cities and thus the innovation gradient could turn flat if small cities were excluded from the analysis. Excluding cities with less than 0.5 million population yields a subsample of 225 cities and about 18,000 firms (columns 4 to 6). The predicated probabilities of firm innovation in this subsample are slightly higher than their counterparts for the full sample, implying that the gaps of innovation propensity between firms in small and large cities may not be large enough to be solely responsible for the baseline results. The estimated marginal effects are all positive and statistically significant, with the one on product innovation slightly higher than that of the full

²¹ The higher agglomeration benefit on innovation than on productivity has also been found in developed countries (Duranton 2014).

²² These cities are Bangkok, Beijing, Delhi, Dhaka, Guangzhou, Ho Chi Minh, Jakarta, Karachi, Kolkata, Lahore, Manila, Mumbai, and Shanghai.

sample and the one on R&D two-thirds of the full-sample estimate. Thus, the firm innovation gradients seem ubiquitous along the city size distribution in the region.²³

In addition, to test if results are sensitive to the definition of innovation, two alternative measures of the dependent variable are used. The first indicator concerns whether the product innovation is “also new to the establishment’s main market”. Answers to this question are expected to be more reliable as the question requires the respondents to think more about their answer to the preceding question regarding general product innovation; it also gets at the importance of an innovation. As expected, this follow-up question yields a lower percentage of positive responses reported in the sample—20.5% as opposed to 35.7% for the general question on product innovation. The regression results in column (7) of Table 4 show that doubling city size is associated with a 3.0 percentage point or 22% increase in the propensity of main market product innovation, higher than 13.5% increase in propensity using the general product innovation measure.

The second alternative measure of innovation takes a value of 1 if a firm conducts either product or process innovation. The sample average, equal to 57.5%, is greater than the averages of product innovation and process innovation and smaller than the sum of the two. This implies that respondents were able to differentiate between the two types of innovation and a fair proportion of firms engaged in only one of the two. The estimated marginal effect is 4.5 percentage points, or a 7.8% increase over the predicted propensity, also with 1% statistical significance.

4.3. Heterogeneity by Development Level and Country

The 25 Asian countries included in the analysis are at distinct stages of development. It is possible to examine whether the relationship between agglomeration effects and firm innovation varies with the level of development by dividing the overall sample into upper-middle income countries and low and lower-middle income countries based on the latest World Bank country classifications.²⁴ Probit regressions are estimated for both subsamples and results are reported in Table 5.

For low and lower-middle income countries, encompassing about 75% of firm observations in 345 cities, the agglomeration effects for product innovation and R&D are positive and comparable to the estimates from the full sample with 1% statistical significance. For process innovation, the marginal effect is estimated at 0.019, around half of that for the full sample, with 5% statistical significance.

²³ The sensitivity of estimates to the exclusion of small cities which have only a small number of firms surveyed by the WBES is also considered using thresholds of 10 and 20 firms per city. The former yields 264 cities and around 20,000 firms and results are close to baseline estimates; the city size coefficients for all three innovation categories are positive and significant at 1% level. The more aggressive exclusion results in a subsample of 193 cities and around 19,000 firms. The estimate for product innovation remains almost identical to the full sample estimate. For process innovation and R&D, however, both coefficients halve in size and are marginally significant.

²⁴ The World Bank defines low income economies as those with a gross national income (GNI) per capita of \$1,025 or less, lower-middle income economies between \$1,026 and \$3,995, and upper-middle income economies between \$3,996 and \$12,375.

Firms in upper-middle income countries are more likely to undertake innovation activities. For instance, 36% of firms undertake product innovation and 29% of firms conduct R&D in the upper-middle income countries, as opposed to 25% and 13% in low and lower-middle income countries, respectively. While the upper-middle income subsample accounts for one quarter of the total firms and 30% of cities, strong agglomeration effects on product and process innovations are still obtained. The estimated marginal effect is slightly lower than that for the full sample for product innovation and slightly higher for process innovation. The marginal effect for R&D is half of the estimate of the full sample and statistically insignificant.

Overall, agglomeration effects exist consistently across the samples pertaining to different development levels in the case of product innovation while they vary for process innovation and R&D. At low and lower-middle income levels, firm R&D is more concentrated in large cities while process innovation less so. For countries that have reached higher income levels, process innovation tends to concentrate in large cities whereas R&D is undertaken more universally.

The 25 countries are also diverse in terms of institutional or cultural backgrounds. It is worth examining how the agglomeration effects on firms' innovation varies across these different social settings. Three cases are considered, namely China, India and a group of Southeast Asian countries (which are developing members of the Association of Southeast Asian Nations, or ASEAN, and referred to as so). The results are summarized in Table 6.

The estimated marginal effects are all positive and statistically significant across the three countries and subregions except for those on R&D in China and process innovation in developing ASEAN countries. The predicted propensity of product and process innovation are nearly identical across China and India, whereas China's R&D propensity is higher than India's by 10 percentage points. The estimated marginal effects of city population on product innovation and R&D are much larger in India than in China. These differences appear to echo the stylized facts presented in section 3: At the city level, the most innovative cities in India tend to show a higher concentration of firm innovation relative to population than their Chinese counterparts (Figure 3); at the country level, the Herfindahl-Hirschman Index ratios for India are higher than those for China (Table 2).

Developing ASEAN countries have substantially fewer firms engaged in innovation as compared to China or India. The marginal effects are also smaller than those of these two countries. Owing to their low baseline probabilities, however, the percent changes in product and R&D probabilities in developing ASEAN countries are substantial (27% and 17%, respectively) and comparable to those of China and India.

5. Presence of High Quality Universities

So far, this paper's analysis has established a positive association between urban agglomeration and firm innovation that is pervasive across developing Asia. This section investigates a potentially important channel through which such effects may arise. Specifically, it examines whether the presence of a top-ranked university in cities is associated with firm innovation.

Universities, especially high quality ones, are found to be closely associated with the innovation related activities of local firms in developed countries (for instance, see Jaffe 1989, Anselin, Varga, and Acs 1997, and Andrews 2017). There are several ways for universities and firms in the same city to interact, including university-firm collaborations through market-mediated interactions, unintended knowledge flows from university-based research (D’Este and Iammarino 2010; D’Este and Patel 2007), and universities as suppliers of human capital to local firms (Toivanen and Vaananen 2016). Biasi, Deming, and Moser (2020) provide a comprehensive review of literature on university and innovation.

Across various academic disciplines or majors, those belonging to science, technology, engineering and mathematics (STEM) appear to be particularly important for spillover linkages between universities and firms’ innovation. Bianchi and Giorcelli (2020) find that an exogenous increase in STEM majors in Italy led to more innovation in general, with effects concentrated in chemistry, medicine, and information technology. Similarly, Toivanen and Vaananen (2016) find engineering education to have a positive and significant effect on patenting in Finland.

In summary, based on analysis of developed country data, firms’ innovation could benefit from geographical proximity to a university, with certain majors playing an especially important role.

An examination of the data on top ranked universities and majors suggests there is a high degree of spatial concentration of top Asian universities and top global majors in developing Asia. First, the top Asian universities are unevenly distributed across countries (Table 7). Only 9 of the 25 countries in the sample have a top university, and the majority of them (211 or 85% of total) are located in five countries—PRC (78), India (64), Malaysia (25), Pakistan (22), and Indonesia (22). Second, within these countries, the quality universities are also unevenly distributed across cities. For example, all 6 top universities in Bangladesh are located in Dhaka; 14 of the 25 top universities in Malaysia in Kuala Lumpur; and 12 of 18 in Thailand in Bangkok. Third, the majority of top Asian universities and universities with top global majors are located in larger cities, with the latter slightly more so (Table 8). Among the 248 universities in our sample cities, less than 1 in 7 are located in cities with population less than 1 million; for the universities with top majors, the share is between 3% (arts and humanities) to 10% (life sciences). The large city bias is especially apparent in China, where all the top universities and top majors are in cities with a population above 1 million.

A formal investigation of the association between firm innovation and presence of a top Asian university and academic disciplines is carried out as follows. First, a dummy for co-location of a top Asian university is included in the Probit model. The results are reported in columns 1-3 of Table 9 and the presence of a top university is found to be significantly associated with a higher propensity for firms to introduce product and process innovations as well as conduct R&D. The association is more pronounced for process innovation. In addition, once the presence of a top university is controlled for, the coefficients on city size diminish in magnitude as well as significance for process innovation and R&D. This implies that the colocation of high-quality universities and firms serves as an important channel through which urban scale promotes firm innovation through these activities. In contrast, the coefficient on city

size remains positive and significant in the regression for product innovation (though diminished in magnitude) suggesting that agglomeration influences innovation through several channels and not just the one working through universities.

Second, there may be sector heterogeneity in the relationship between universities and firms' innovation activities. Firms are thus distinguished by sector (manufacturing or services) and the university dummy is interacted with the firm sector dummy. The results presented in columns 4-6 of Table 9 suggest that the presence of a top university mainly promotes product innovation in manufacturing firms with little effect seen for services firms. On the other hand, it plays a bigger role for services firms than manufacturing firms in terms of R&D investment. While more research is needed to pin down the drivers of these patterns, the findings imply that the universities-firms linkages vary by sector.

The role a university plays in promoting firm innovation may also depend on its strength in different fields, which may vary across universities. To assess this possibility, dummies for the presence of a top major (across five majors) are included in the model, in addition to the dummy for a top university. The results are shown in Table 10. Most notably, the presence of a top engineering major is found to be strongly associated with increased propensity of all three firm innovation activities. In particular, the coefficients are higher for product innovation and R&D than for process innovation. Institutions with top life science majors also catalyze firm innovation, particularly in process innovation and R&D, in the same city. On the contrary, the estimates suggest that the presence of a top arts and humanities major or natural sciences is associated with decreased propensity for firm innovation. As for having a top social sciences and management major, the estimates are positive for product innovation and negative for R&D. Lastly, the top university dummy remains positive across the three regressions and statistically significant for process innovation and R&D.

Though by no means definitive, these results are interesting for the region studied. Local governments such as Shenzhen in China seem to be on the right track with their efforts to establish universities or attract new campuses of universities with strong technological and engineering majors as they aspire to move up the global value chain of production. Nevertheless, it is unclear to what extent our estimates represent causal relations. Universities with strong arts and humanities departments may not discourage firm innovation at all. Instead, it may be that they are located in cities with rich cultural or historical backgrounds that are not particularly attractive to entrepreneurs.

6. Conclusion

To the best of our knowledge, this paper is among the first to systematically examine the spatial distribution of developing country firms' engagement in innovation-related activities at the city level, and whether and how urban agglomeration affects the propensity of firms to innovate. The analysis shows that firm innovation is highly concentrated in larger cities, although not all megacities host disproportionately more innovative firms. Overall, a doubling of city size is associated with increases in firms' propensity to introduce a product innovation, process innovation, and conduct R&D by 13.5%, 8.5%, and 14.3% against their respective benchmark

propensities, all else being equal. The implied percent increases in the propensities of firms to undertake innovation related activities are prominent given their moderate starting points in the countries studied. Furthermore, the relationship between city size and innovation propensities is found to be robust and pervasive across developing Asia. Urban agglomeration effects are heterogeneous across different subsamples defined in terms of development level or subregion, but the differences seem to be secondary.

This paper's findings also suggest that firms in cities with top universities are more likely to carry out innovation-related activities. In addition, across academic disciplines, the presence of top ranked engineering majors is especially closely linked with firm innovation.

These results have some clear policy implications. First, in light of the importance of innovation to economic growth, policymakers in developing economies should be more welcoming of the growth of larger urban agglomerations than they often are. In particular, calls for regionally balanced development and a belief that large cities have become too large can lead to policies that inadvertently undermine the potential benefits from larger urban agglomerations. Second, policies to improve the quality of universities and promote interactions between universities and local firms are likely to have high payoffs. Further, the results on the importance of high-quality engineering majors in enhancing the propensity to undertake innovation related activities is consistent with the findings of developed country studies emphasizing the role of STEM education in bolstering innovation. Overall, setting up new universities or improving the quality of existing ones may be prioritized for those larger cities without a university (or a high-quality one).

The results also suggest several issues for future research. First, an effort to establish causality would be clearly worthwhile, though it would most probably require adopting a country-specific approach given the need to find natural experiments that yield plausible instrumental variables. Second, it would be useful to explore how other city characteristics, such as the quality of transport connectivity within and between cities, might affect the spatial distribution of process and product innovations. Third, there are some differences across the results for product and process innovations. Not only do they have unequal baseline probabilities, they also respond to agglomeration differently across low and lower-middle income countries and upper-middle income countries. Similarly, the analysis on the quality of different academic disciplines suggests clear differences regarding their effect on propensities to innovate. These differences could be explored. Finally, gaining in-depth understanding of the channels through which the co-location of universities and firms influences innovation in a developing country setting would be very worthwhile.

References

- Acemoglu, D., P. Aghion, and F. Zilibotti. (2006) Distance to Frontier, Selection, and Economic Growth. *Journal of the European Economic Association*. 4 (1). pp. 37–74.
- Anderson, G. and Ge, Y., 2005. The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35(6), pp.756-776.
- Andrews, M. (2017) The Role of Universities in Local Invention: Evidence from the Establishment of US Colleges. *Job Market Paper*.
- Anselin, L., A. Varga, and Z. Acs. (1997) Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics*. 42 (3). pp. 422–448.
- Asian Development Bank (2019). *Asian Development Outlook Update: Fostering Growth and Inclusion in Asia's Cities*. Manila.
- Au, C.C. and Henderson, J.V., 2006. Are Chinese cities too small? *The Review of Economic Studies*, 73(3), pp.549-576.
- Audretsch, D.B., and M.P. Feldman. (1996) R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*. 86 (3). pp. 630–640.
- Audretsch, D.B., and P.E. Stephan. (1996) Company-Scientist Locational Links: The Case of Biotechnology. *The American Economic Review*. 86 (3). pp. 641–652.
- Ayyagari, M., Demirgüç-Kunt, A. and Maksimovic, V., (2011) Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial and Quantitative Analysis*, 46(6), pp.1545-1580.
- Baum-Snow, N., and F. Ferreira. (2015) Causal Inference in Urban and Regional Economics. In *Handbook of Regional and Urban Economics*. Vol. 5. pp. 3–68. Elsevier.
- Beck, T., A.S.L.I. Demirgüç-Kunt, and V. Maksimovic. (2005) Financial and Legal Constraints to Growth: Does Firm Size Matter? *The Journal of Finance*. 60 (1). pp. 137–177.
- Berliant, M., R.R. Reed III, and P. Wang. (2006) Knowledge Exchange, Matching, and Agglomeration. *Journal of Urban Economics*. 60 (1). pp. 69–95.
- Bertschek, I. (1995) Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investment. *The Journal of Industrial Economics*. pp. 341–357.
- Bianchi, N. and Giorcelli, M., 2020. Scientific education and innovation: from technical diplomas to university STEM degrees. *Journal of the European Economic Association*, 18(5), pp.2608-2646.

- Biasi, B., Deming, D.J. and Moser, P., 2020. Education and innovation. In *The Role of Innovation and Entrepreneurship in Economic Growth*. University of Chicago Press.
- Bosker, M., Park, J. and Roberts, M. (2020) Definition matters. Metropolitan areas and agglomeration economies in a large-developing country, *Journal of Urban Economics*, p.103275
- Carlino, G.A., S. Chatterjee, and R.M. Hunt. (2007) Urban Density and the Rate of Invention. *Journal of Urban Economics*. 61 (3). pp. 389–419.
- Carlino, G.A., R.M. Hunt, G. Duranton, and B.A. Weinberg. (2009) What Explains the Quantity and Quality of Local Inventive Activity? *Brookings-Wharton Papers on Urban Affairs*. pp. 65–123.
- Carlino, G.A., and W.R. Kerr. (2015) Agglomeration and Innovation. In *Handbook of Regional and Urban Economics*. Vol. 5. pp. 349–404. Elsevier.
- Chauvin, J.P., Glaeser, E., Ma, Y. and Tobio, K., (2017) What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98, pp.17-49.
- Chen, L., Hasan, R. and Jiang, Y., 2020. Urban Agglomeration and Firm Innovation: Evidence from Asia. *Asian Development Bank Economics Working Paper Series*, No. 616.
- Colmer, Jonathan, 2016. Urbanisation, Growth, and Development: Evidence from India. *Unpublished Review Paper*.
- Combes, P.P., G. Duranton, L. Gobillon, and S. Roux. (2010) Estimating Agglomeration Economies with History, Geology, and Worker Effects. In *Agglomeration Economics*. pp. 15–66. University of Chicago Press.
- Dingel, J.I., Miscio, A. and Davis, D.R. (2019) Cities, lights, and skills in developing economies. *Journal of Urban Economics*, p.103174.
- Duranton, G. (2008) From Cities to Productivity and Growth in Developing Countries. *Canadian Journal of Economics/Revue Canadienne d'économie*. 41 (3). pp. 689–736.
- Duranton, G. (2014) *Growing through Cities in Developing Countries*. Washington, DC: World Bank.
- Duranton, G., and D. Puga. (2004) Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics*. Vol. 4. pp. 2063–2117. Elsevier.
- D'Este, P., and S. Iammarino. (2010) The Spatial Profile of University-Business Research Partnerships. *Papers in Regional Science*. 89 (2). pp.335–350.

- D'Este, P., and P. Patel. (2007) University–Industry Linkages in the UK: What Are the Factors Underlying the Variety of Interactions with Industry? *Research policy*. 36 (9). pp. 1295–1313.
- Fang, L., Li, P. and Song, S., 2017. China's development policies and city size distribution: An analysis based on Zipf's law. *Urban Studies*, 54(12), pp.2818-2834.
- Feldman, M.P., and D.B. Audretsch. (1999) Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition. *European Economic Review*. 43 (2). pp. 409–429.
- Galdo, V., Li, Y. and Rama, M. (2019) Identifying urban areas by combining human judgment and machine learning: An application to India, *Journal of Urban Economics*, p103229
- Gerlach, H., T. Rønde, and K. Stahl. (2009) Labor Pooling in R&D Intensive Industries. *Journal of Urban Economics*. 65 (1). pp.99–111.
- Helsley, R.W., and W.C. Strange. (1990) Matching and Agglomeration Economies in a System of Cities. *Regional Science and Urban Economics*. 20 (2). pp. 189–212.
- Helsley, R.W., and W.C. Strange. (2002) Innovation and Input Sharing. *Journal of Urban Economics*. 51 (1). pp. 25–45.
- Henderson, J.V. (2005) Urbanization and Growth. In *Handbook of Economic Growth*. Vol. 1. pp. 1543–1591). Elsevier.
- Huergo, E., and J. Jaumandreu. (2004) How Does Probability of Innovation Change with Firm Age? *Small Business Economics*. 22 (3-4). pp. 193–207.
- Jaffe, A.B. (1989) Real Effects of Academic Research. *American Economic Review*. 79 (5). pp. 957–970.
- Jiang, Y., 2021. Asian cities: spatial dynamics and driving forces. *The Annals of Regional Science*, pp.1-46.
- Kinda, T. (2010) Investment Climate and FDI in Developing Countries: Firm-Level Evidence. *World Development*. 38 (4). pp. 498–513.
- Melo, P.C., D.J. Graham, and R.B. Noland. (2009) A Meta-Analysis of Estimates of Urban Agglomeration Economies. *Regional Science and Urban Economics*. 39 (3). pp.332–342.
- Moretti, E. (2019) *The Effect of High-Tech Clusters on the Productivity of Top Innovators*. No. w26270. National Bureau of Economic Research.
- Nieto Galindo, V.M. (2007) La Aglomeración Como Una Causa de la Innovación en Colombia. *Planeación & Desarrollo*. 38 (1). pp. 97–139.

- Ó hUallicháin, B. (1999) Patent Places: Size Matters. *Journal of Regional Science*. 39 (4). pp. 613–636.
- Overman, H.G., and A.J. Venables. (2005) *Cities in the Developing World*. No. 695. Centre for Economic Performance, London School of Economics and Political Science.
- Paunov, C., (2016) Corruption's asymmetric impacts on firm innovation. *Journal of Development Economics*, 118, pp.216–231.
- Paunov, C., M. Borowiecki and N. El-Mallakh (2019), "Cross-country evidence on the contributions of research institutions to innovation", *OECD Science, Technology and Industry Policy Papers*, No. 77, OECD Publishing, Paris
- Paunov, C., D. Guelleci, N. El-Mallakhii, S. Planes-Satorrai and L. Nüse (2019), On the concentration of innovation in top cities in the digital age, *OECD Science, Technology and Industry Policy Papers*, No. 85, OECD Publishing, Paris
- Paunov, C. and Rollo, V., (2016) Has the internet fostered inclusive innovation in the developing world?. *World Development*, 78, pp.587–609.
- Puga, D. (2010) The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science*. 50 (1). pp. 203–219.
- QS. (2019a) Quacquarelli Symonds (QS) Asia University Rankings. <<https://www.topuniversities.com/university-rankings/asian-university-rankings/2019>> accessed 15 May 2019.
- QS. (2019b) Quacquarelli Symonds (QS) World University Rankings by Subject. <<https://www.topuniversities.com/subject-rankings/2019>> accessed 10 January 2021.
- Reyes, J.D., M. Roberts, and L.C. Xu. (2017) *The Heterogeneous Growth Effects of the Business Environment: Firm–Level Evidence for a Global Sample of Cities*. Washington, DC: World Bank.
- Romer, P.M. (1986) Increasing Returns and Long-Run Growth. *Journal of Political Economy*. 94 (5). pp. 1002–1037.
- Rosenthal, S.S., and W.C Strange. (2004) Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics*. Vol. 4. pp. 2119–2171. Elsevier.
- Strange, W., Hejazi, W. and Tang, J., (2006) The uncertain city: competitive instability, skills, innovation and the strategy of agglomeration. *Journal of Urban Economics*, 59(3), pp.331–351.
- Toivanen, O., and Väänänen, L. (2016) Education and invention. *Review of Economics and Statistics*, 98(2), pp.382–396.

UN (United Nations). (2018) World Urbanization Prospects, the 2018 Revision: Methodology. New York: United Nations.

Valero, A. and J. Van Reenen (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*. Volume 68, pages 53-67.

Yang, J.S. (2017). The Governance Environment and Innovative SMEs. *Small Business Economics*. 48 (3). pp. 525–541.

Figure 1. Natural Cities and Administratively Defined Cities

(a) Chennai urban agglomeration and nearby towns (2011) vs. Chennai natural city (2010)

(b) Dalian prefecture and its city proper vs. natural cities within Dalian prefecture (2010)

Source: Authors' analysis based on nighttime lights images from the National Oceanic and Atmospheric Administration, 2011 town boundary database for India from World Bank and database of Global Administrative Areas, version 3.6 from University of California, Berkeley.

Figure 2: Geographic Distribution of Natural Cities by Population Size

Source: Authors' analysis based on nighttime lights images from the National Oceanic and Atmospheric Administration and World Bank Enterprise Surveys

Note: Each dot represents a natural city in our sample.

Figure 3: Shares of Innovative Firms versus Urban Population

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Note: This figure plots the share of natural city population in a country against the share of innovative firms in a country, alongside a dashed 45-degree line. Population share is measured as the population in each natural city over the total population in all sample natural cities. The same method applies for the share of innovative firms. In each innovation category, the cities with the most innovative firms in each country are shown. For China and India (left), the top 10 innovative cities are plotted. For other countries (right hand side), three cities are plotted if the number of cities is greater than 10, and one city is plotted otherwise.

Table 1: Summary Statistics of Firm and City Variables

VARIABLES	N	Mean	Standard Deviation	Min	Median	Max
Firm performance						
Product innovation	21,741	0.357	0.479	0	0	1
Main market product innovation	21,741	0.205	0.403	0	0	1
Process innovation	20,578	0.493	0.500	0	0	1
Firm R&D	20,675	0.244	0.430	0	0	1
Product or process innovation	21,119	0.563	0.496	0	1	1
Firm characteristics						
Young firm	21,655	0.209	0.407	0	0	1
Small firm	21,795	0.645	0.478	0	1	1
FDI	21,857	0.061	0.239	0	0	1
Manufacturing	21,857	0.657	0.475	0	1	1
Headquarter dummy	21,857	0.413	0.492	0	0	1
Share of skilled workers	21,857	0.348	0.327	0	0.313	1
City characteristics						
Population (2010)	489	1.43e+06	3.33e+06	8,353	437,813	3.58e+07
Log population (2010)	489	13.19	1.278	9.030	12.99	17.39
Average precipitation (mm)	489	3.664	2.521	0.0973	2.887	12.38
Max temperature	489	34.78	4.091	16.13	33.92	43.84
Min temperature	489	9.381	11.21	-30.26	10.91	23.49
Distance to port (km)	489	400.9	499.1	0.462	170.1	2,232
Log distance to port	489	4.937	1.769	-0.773	5.137	7.711
Terrain ruggedness index (m)	489	33.20	43.31	0	20.34	503.9
With top Asian university	489	0.202	0.402	0	0	1
With top arts & humanities major	489	0.039	0.193	0	0	1
With top engineering major	489	0.063	0.244	0	0	1
With top life sciences major	489	0.041	0.198	0	0	1
With top natural sciences major	489	0.055	0.229	0	0	1
With top social sciences & management major	489	0.047	0.212	0	0	1

Source: Authors' analysis based on firm data from World Bank Enterprise Survey, nighttime lights images from the National Oceanic and Atmospheric Administration, and university data from QS World University Rankings and Asia University Rankings.

Note: FDI = foreign direct investment dummy, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, mm = millimeter, R&D = research and development, SD = standard deviation.

Table 2. Herfindahl-Hirschman Index Raito

Country	Product innovation HHI ratio	Process innovation HHI ratio	Firm R&D HHI ratio
Afghanistan	0.59	0.53	0.67
Armenia	1.14	1.27	1.27
Azerbaijan	0.97	1.13	-
Bangladesh	1.36	1.47	1.66
Bhutan	1.14	1.02	1.12
Cambodia	1.2	1.2	1.19
China	1.45	1.33	1.21
Georgia	1.03	0.88	1.66
India	3.11	2.68	4.12
Indonesia	1.76	1.36	3.26
Kazakhstan	1.59	1.21	3.09
Kyrgyz Republic	2.12	2.15	1.89
Lao PDR	0.94	0.84	0.87
Malaysia	2.23	1.52	1.16
Mongolia	1.26	1.22	0.96
Myanmar	0.98	1.45	0.85
Nepal	0.95	0.82	1.62
Pakistan	1.17	1.17	1.32
Papua New Guinea	0.85	1.02	1.14
Philippines	0.87	0.84	0.75
Tajikistan	1.03	0.99	0.89
Thailand	0.59	0.39	0.78
Uzbekistan	1.58	3.62	3.74
Viet Nam	0.79	0.82	1.08

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Notes: The HHI ratio is the firm innovation HHI over the population HHI. It measures the degree of concentration of innovation activities relative to population. An HHI ratio higher than one indicates relative spatial concentration of firm innovation to population. City population HHI in country c is defined as $HHI_c = \sum S_i^2$, where S_i is the share of population in natural city i in country c . The same method applies to city innovation HHI. For Azerbaijan, none of the surveyed firms reported R&D activity.

Table 3: Probit Estimation Results of Firm Innovation

Variables	(1) Product Innovation	(2) Process Innovation	(3) Firm R&D
log City Population (2010)	0.12*** (0.0254)	0.0929*** (0.0215)	0.101*** (0.0246)
Marginal effect	0.0425*** (0.0091)	0.0365*** (0.0084)	0.0279*** (0.0066)
Predicted probability	0.314*** (0.0114)	0.428*** (0.0114)	0.195*** (0.0106)
log Distance to Port (km)	0.108*** (0.0257)	0.159*** (0.0233)	0.0248 (0.0248)
Average precipitation (mm)	-0.124*** (0.0295)	-0.117*** (0.0286)	-0.0827** (0.0349)
Max temperature	-0.0047 (0.0162)	0.0389*** (0.0146)	0.0909*** (0.0164)
Min temperature	0.0318*** (0.009)	0.03*** (0.0073)	0.0154* (0.0084)
Terrain Ruggedness Index (m)	0.0015 (0.0011)	0.0028** (0.0011)	0.0024** (0.0011)
Young firm	0.13 (0.0893)	0.196*** (0.0654)	0.122* (0.072)
Small firm	-0.214*** (0.0633)	-0.524*** (0.0596)	-0.408*** (0.0603)
FDI	0.329** (0.139)	0.116 (0.133)	0.222* (0.124)
HQ dummy	0.322* (0.176)	-0.0961 (0.0946)	0.0855 (0.131)
Share of skilled workers	-0.0914 (0.136)	0.142 (.145)	0.0442 (0.139)
Sector/Year/Country FE	Yes	Yes	Yes
Observations	21496	20374	20218
No. of cities	489	483	480
F statistic	21.82	33.42	25.48

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Notes: FDI = foreign direct investment, FE = fixed effect, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, mm = millimeter, R&D = research and development. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Robust standard errors clustered by country in parentheses.

Table 4: Robustness Checks

Variables	Excluding cities with >10M population			Excluding cities with <0.5M population			Using other measures of firm innovation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Product Innovation	Process Innovation	Firm R&D	Product Innovation	Process Innovation	Firm R&D	Main Product Innovation	Firm Innovation
log City Population	0.142*** (0.0409)	0.187*** (0.0294)	0.203*** (0.0327)	0.16*** (0.0329)	0.093*** (0.0294)	0.061* (0.0336)	0.138*** (0.0273)	0.115*** (0.0216)
Marginal effect	0.0477*** (0.0137)	0.0732*** (0.0115)	0.0513*** (0.0082)	0.058*** (0.0120)	0.0367*** (0.0116)	0.0179* (0.0097)	0.0301*** (0.0058)	0.0451*** (0.0085)
Predicted probability	0.28*** (0.012)	0.421*** (0.0134)	0.17*** (0.0152)	0.33*** (0.0125)	0.438*** (0.0123)	0.215*** (0.0116)	0.136*** (0.0093)	0.575*** (0.0107)
log Distance to Port (km)	0.106*** (0.028)	0.125*** (0.0262)	0.0092 (0.0264)	0.101*** (0.0271)	0.17*** (0.0245)	0.007 (0.026)	0.144*** (0.0363)	0.133*** (0.0207)
Average precipitation (mm)	-0.0773** (0.0348)	-0.0204 (0.033)	0.0993*** (0.0341)	-0.145*** (0.0324)	-0.116*** (0.0327)	-0.0761** (0.0378)	-0.189*** (0.0303)	-0.0928*** (0.0285)
Max Temperature	-0.0071 (0.0158)	0.0286** (0.0141)	0.0849*** (0.0147)	-0.0111 (0.0182)	0.0515*** (0.0173)	0.116*** (0.0182)	-0.04** (0.0156)	0.0687*** (0.014)
Min Temperature	0.0173* (0.0102)	0.0171** (0.0082)	0.0124 (0.0084)	0.0334*** (0.0097)	0.0319*** (0.008)	0.0165* (0.0086)	0.0541*** (0.0107)	0.0213*** (0.0073)
Terrain Ruggedness Index (m)	0.0032*** (0.0012)	0.0033** (0.0013)	0.0029** (0.0011)	4.2e-04 (0.0013)	0.0023* (0.0013)	0.0025* (0.0014)	0.0044*** (0.0015)	0.002* (0.0011)
Young firm	-0.0876 (0.107)	0.0626 (0.0757)	-0.0618 (0.0838)	0.109 (0.0916)	0.196*** (0.0695)	0.159** (0.0762)	0.161* (0.0833)	0.14** (0.0658)
Small firm	-0.27*** (0.0663)	-0.455*** (0.0731)	-0.454*** (0.0709)	-0.215*** (0.0668)	-0.542*** (0.0631)	-0.396*** (0.0635)	-0.307*** (0.0639)	-0.373*** (0.0607)
FDI	0.157 (0.148)	0.327* (0.171)	0.209 (0.145)	0.296** (0.148)	0.118 (0.142)	0.211 (0.131)	-0.228 (0.175)	0.292** (0.126)
HQ dummy	0.283 (0.195)	0.0514 (0.112)	0.23** (0.108)	0.363** (0.182)	-0.0816 (0.0974)	0.0891 (0.135)	-0.232** (0.102)	0.254** (0.107)
Share of skilled workers	-0.096 (0.16)	-0.151 (0.174)	-0.0374 (0.182)	-0.132 (0.146)	0.16 (0.159)	0.0819 (0.149)	-0.0286 (0.144)	0.362** (0.175)
Sector/Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15941	15092	14913	17845	16792	16638	18877	20897
No. of cities	476	470	467	225	223	222	415	486
F statistic	19.259	29.095	21.461	25.608	30.861	25.201	18.520	38.555

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Notes: FDI = foreign direct investment, FE = fixed effect, km = kilometer, log = logarithm, m = meter, M = million, max = maximum, min = minimum, mm = millimeter, R&D = research and development. *** = p<0.01, ** = p<0.05, * = p<0.1. Robust standard errors clustered by country in parentheses. Columns (1) to (3) are samples restricted to cities with <=10M population, and columns (4) to (6) are samples restricted to cities with >=0.5M population. Columns (7) and (8) use the full sample.

Table 5: Estimation Results by Country Income level

Variables	Low and Lower-middle Income			Upper-middle Income		
	(1) Product Innovation	(2) Process Innovation	(3) Firm R&D	(4) Product Innovation	(5) Process Innovation	(6) Firm R&D
log City Population	0.119*** (0.0225)	0.0496** (0.0232)	0.133*** (0.0293)	0.109*** (0.0397)	0.116*** (0.0362)	0.0438 (0.0397)
Marginal effect	0.0378*** (0.0071)	0.0185** (0.0086)	0.028*** (0.0055)	0.0405*** (0.0149)	0.0463*** (0.0144)	0.0149 (0.0135)
Predicted probability	0.251*** (0.0122)	0.36*** (0.014)	0.129*** (0.013)	0.358*** (0.0154)	0.512*** (0.0162)	0.288*** (0.0156)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector/Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16332	16210	16290	5164	4164	3928
No. of cities	345	344	344	144	139	136
F statistic	17.472	25.993	20.350	21.434	24.954	19.333

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Notes: FE = fixed effect, log = logarithm, R&D = research and development. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Robust standard errors clustered by country in parentheses. Columns (1), (2), and (3) are samples restricted to low income and lower-middle income countries, and columns (4), (5), and (6) are samples restricted upper-middle income countries.

Table 6. Estimation Results of Firm Innovation in China, India, and ASEAN countries

Variables	China			India			ASEAN		
	(1) Product Innovation	(2) Process Innovation	(3) Firm R&D	(4) Product Innovation	(5) Process Innovation	(6) Firm R&D	(7) Product Innovation	(8) Process Innovation	(9) Firm R&D
log City Population	0.114*** (0.0396)	0.151*** (0.0368)	0.0376 (0.0415)	0.189*** (0.0232)	0.148*** (0.0206)	0.177*** (0.0216)	0.148*** (0.0467)	0.05 (0.0445)	0.0855* (0.047)
Marginal effect	0.0447*** (0.0156)	0.0591*** (0.0146)	0.0144 (0.0158)	0.0739*** (0.0091)	0.0578*** (0.008)	0.0598*** (0.0075)	0.0229*** (0.0073)	0.0141 (0.0126)	0.01* (0.0052)
Predicted probability	0.43*** (0.0173)	0.581*** (0.0172)	0.386*** (0.0161)	0.423*** (0.0122)	0.582*** (0.0123)	0.282*** (0.012)	0.0843*** (0.0104)	0.203*** (0.0177)	0.0586*** (0.0116)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year/Country FE	No	No	No	No	No	No	Yes	Yes	Yes
Observations	2619	1640	1629	8067	8054	8050	5697	5618	5693
No. of cities	74	69	68	207	207	207	124	123	123
F statistic	4.744	19.380	12.293	18.973	37.288	34.549	12.519	10.588	7.763

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Notes: FE = fixed effect, log = logarithm, R&D = research and development, ASEAN = Association of Southeast Asian Nations. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Robust standard errors clustered by sector and/or country in parentheses. Columns (1), (2), and (3), and (4), (5), and (6) are samples restricted to cities in China and India, respectively. Columns (7), (8), (9) are samples restricted to ASEAN countries (i.e., Cambodia, Indonesia, Lao PDR, Myanmar, Malaysia, Philippines, Thailand, and Viet Nam).

Table 7. Distribution of Top Asian Universities by Country

Country	Number of top Asian universities	Number of cities with top Asian universities	City with the most top Asian universities
Bangladesh	6	1	Dhaka (6)
PRC	78	17	Beijing (20)
India	64	43	Delhi (7)
Indonesia	22	9	Bandung (5)
Malaysia	25	10	Kuala Lumpur (14)
Pakistan	22	5	Lahore (9)
Philippines	6	2	Metro Manila (5)
Thailand	18	7	Bangkok (12)
Viet Nam	7	5	Ha Noi (2)
Total	248	99	

Source: Authors' analysis based on firm data from World Bank Enterprise Survey, nighttime lights images from the National Oceanic and Atmospheric Administration, and university data from QS World University Rankings.

Table 8. Distribution of Top Asian Universities and Top Global Majors

Number of universities	China			India			Other countries			Total		
	<=1M	>1M	Total	<=1M	>1M	Total	<=1M	>1M	Total	<=1M	>1M	Total
Top Asian university	0	78	78	15	49	64	18	88	106	33	215	248
University with top arts & humanities major	0	17	17	0	5	5	1	12	13	1	34	35
University with top engineering major	0	30	30	2	8	10	3	16	19	5	54	59
University with top life sciences major	0	18	18	0	2	2	3	8	11	3	28	31
University with top natural sciences major	0	28	28	2	9	11	1	7	8	3	44	47
University with top social sciences & management major	0	17	17	1	9	10	3	16	19	4	42	46

Source: Authors' analysis based on firm data from World Bank Enterprise Survey, nighttime lights images from the National Oceanic and Atmospheric Administration, and university data from QS World University Rankings and Asia University Rankings.

Note: <=1M are natural cities with less than or equal to one million population. >1M are natural cities with greater than one million population.

Table 9. Estimation Results of Agglomeration Channels

Variables	(1) Product Innovation	(2) Process Innovation	(3) Firm R&D	(4) Product Innovation	(5) Process Innovation	(6) Firm R&D
log City Population	0.0801** (0.0332)	-0.0147 (0.0305)	0.0405 (0.0347)	0.0782** (0.0331)	-0.0139 (0.0305)	0.0445 (0.0347)
Top Asian university	0.181** (0.0874)	0.465*** (0.0841)	0.263*** (0.0907)	0.0431 (0.11)	0.5*** (0.111)	0.476*** (0.116)
Manufacturing	0.0472 (0.0983)	0.093 (0.108)	0.0254 (0.0971)	-0.157 (0.116)	0.129 (0.12)	0.238** (0.118)
Manufacturing x Top Asian university				0.276** (0.119)	-0.0566 (0.109)	-0.31*** (0.113)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21496	20374	20218	21496	20374	20218
No. of cities	489	483	480	489	483	480
F statistic	21.486	33.823	24.905	21.447	32.979	23.759

Source: Authors' analysis based on firm data from World Bank Enterprise Survey, nighttime lights images from the National Oceanic and Atmospheric Administration, and university data from QS World University Rankings.

Notes: FE = fixed effect, log = logarithm, R&D = research and development. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Robust standard errors clustered by country in parentheses.

Table 10. Estimation Results of Agglomeration Channels

Variables	(1) Product Innovation	(2) Process Innovation	(3) Firm R&D
log City Population	0.0643 (0.0404)	0.0081 (0.0315)	0.0332 (0.0378)
Top Asian university	0.147 (0.104)	0.302*** (0.0903)	0.188* (0.104)
Top arts & humanities major	-0.68*** (0.139)	-0.68*** (0.137)	-0.548*** (0.153)
Top engineering major	0.652*** (0.159)	0.338** (0.167)	0.959*** (0.166)
Top life sciences major	0.149 (0.139)	0.408*** (0.148)	0.635*** (0.168)
Top natural sciences major	-0.298** (0.132)	0.0685 (0.128)	-0.472*** (0.141)
Top social sciences & management major	0.224* (0.131)	0.0374 (0.131)	-0.424*** (0.11)
Other controls	Yes	Yes	Yes
Sector/Year/Country FE	Yes	Yes	Yes
Observations	21496	20374	20218
No. of Cities	489	483	480
F statistic	20.246	31.173	23.079

Source: Authors' analysis based on firm data from World Bank Enterprise Survey, nighttime lights images from the National Oceanic and Atmospheric Administration, and university data from QS World University Rankings and Asia University Rankings.

Notes: FE = fixed effect, log = logarithm, R&D = research and development.

*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Robust standard errors clustered at country in parentheses.

Table A1: Matching between Geocoded Firms in the World Bank Enterprise Survey and Natural Cities

Country	Survey Year	Total Number of Firms	Number of Natural Cities Matched	Number of Matched Firms	Number of Unmatched Firms	Matching Rate
Afghanistan	2014	410	5	392	18	95.61%
Armenia	2013	360	3	313	47	86.94%
Azerbaijan	2013	390	2	249	141	63.85%
Bangladesh	2013	1,442	11	1,345	97	93.27%
Bhutan	2015	253	2	168	85	66.40%
Cambodia	2016	373	3	163	210	43.70%
China	2012	2,700	74	2,700	0	100.00%
Georgia	2013	360	4	258	102	71.67%
India	2014	9,281	207	8,100	1,181	87.28%
Indonesia	2015	1,320	40	1,185	135	89.77%
Kazakhstan	2013	600	20	547	53	91.17%
Kyrgyzstan	2013	270	3	189	81	70.00%
Laos	2016	368	3	232	136	63.04%
Malaysia	2015	1,000	25	888	112	88.80%
Mongolia	2013	360	3	144	216	40.00%
Myanmar	2016	607	5	607	0	100.00%
Nepal	2013	482	4	389	93	80.71%
Pakistan	2013	341	11	334	7	97.95%
Papua New Guinea	2015	65	2	64	1	98.46%
Philippines	2015	1,335	11	1,234	101	92.43%
Tajikistan	2013	359	4	280	79	77.99%
Thailand	2016	1,000	20	660	340	66.00%
Timor-Leste	2015	126	1	124	2	98.41%
Uzbekistan	2013	390	9	364	26	93.33%
Viet Nam	2015	996	17	928	68	93.17%
Total		25,188	489	21,857	3,331	86.78%

Source: Authors' analysis based on firm data from World Bank Enterprise Survey and nighttime lights images from the National Oceanic and Atmospheric Administration.

Appendix 1. Zipf's Law analysis using NTL-based natural cities and traditional city data in China and India

Zipf's law on city size states that the population of the N^{th} largest city in a given country is $1/N$ times the population of the largest city. Testing data fit for Zipf's law is considered a useful exercise to describe the distribution of urban population in a country. Recently, Rozenfeld et al. (2011) find cities built with a bottom-up algorithm using granular population distribution data fit Zipf's law better than those defined by administrative or legal borders in the USA and UK.

Studies examining the city size distribution in developing countries typically use official statistics, which are based on administratively-defined cities or extended areas encompassing large rural areas (e.g. Chauvin et al., 2017; Schaffar and Dimou, 2012; Soo, 2014; and Colmer, 2016). It is instructive to estimate Zips' law with our natural city data and compare with estimates using conventional data. The analysis is performed using data on China and India, the two countries that contain a sufficiently large number of cities.¹

Specifically, the following regression equation is estimated:

$$\log\left(r_i - \frac{1}{2}\right) = \alpha + \beta \log Pop_i + \varepsilon_i$$

where r_i is the rank and Pop_i is the size of population of city i . The coefficient of city size, β , is equal to -1 if Zipf's law holds. If the coefficient is less than -1 (absolute value greater than 1), it implies that small cities are too big and/or large cities are too small as compared to what Zipf's law predicts, and vice versa.

Figure A1 plots the raw data (black dots) and fitted Zipf's law (red lines) with different data (left panels for administratively defined cities and right panels natural cities). Except the prefecture city proper in China, all samples contain cities both greater and smaller than 100,000 in population. Thus, Zipf's law is estimated with both full sample (solid line) and the subsample with population above 100,000. The estimation regressions are reported in the footnote of the figure.

In the case of China, the two datasets produce distinct estimates. Using official statistics, the β is estimated at -1.26, significantly different from -1.0, with R^2 equal to 0.92. The estimate suggests that the small cities are bigger and large cities are smaller than what Zipf's law predicts, which probably arise, respectively, when the data leaves out other small cities and towns outside city proper (as in the case of Ningbo in Figure 1) and adjacent urbanized areas are not counted as part of the city proper. The results from natural cities depict a different picture. When all natural cities are considered, the β is estimated at -0.79 implying that as of 2010 the urban system of China was characterized by a number of small cities and/or some mega cities. This deviation from the Zipf's law turns out to be driven by the presence of small cities. When the focus is on

¹ Following the literature, administratively-defined cities are the prefecture city proper in the case of China (with their 2010 population drawn from the City Statistical Yearbook of China), and town and urban agglomeration in the case of India (with their population drawn from Census 2011). An urban agglomeration consists of one or more contiguous towns and their adjoining outgrowths in India. The term town also includes cities in this context.

natural cities with 100,000 or more population, the coefficient is virtually -1 and the goodness-of-fit measured as R^2 reaches 0.96.

India's administrative city data fits Zipf's law quite well. The estimated β is moderately above and below -1.0 for all cities and cities with population above 100,000, respectively. These patterns are consistent with what Colmer (2016) obtains, which uses similar city definitions based on census data up to 2001. The estimates using natural city data suggests slightly larger and one-way deviation from the Zipf's law for both samples. With natural cities below 100,000 dropped, the estimated coefficient of population is -0.92, implying that urban population is more concentrated in a few mega cities in India than what Zipf's law would predict.

In summary, results on how well Zipf's law holds are mixed when comparing natural cities and administratively defined cities in China and India. However, as Duranton (2021) indicates, the degree of fit for Zipf's law may not be used to validate a city definition even though it is an informative way to describe an urban system. Defining cities is a key challenge in studying cities. The natural city approach used here attempts to provide an alternative, spatially-consistent measurement and perspective on cities in Asia, which allows for meaningful comparison across countries.

References

Chauvin, J.P., Glaeser, E., Ma, Y. and Tobio, K., 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98, pp.17-49.

Colmer, Jonathan, 2016. Urbanisation, Growth, and Development: Evidence from India. *Unpublished Review Paper*.

Duranton, Gilles, 2021. Classifying locations and delineating space: An introduction. mimeo.

Rozenfeld, H.D., Rybski, D., Gabaix, X. and Makse, H.A., 2011. The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, 101(5), pp.2205-25.

Schaffar, A. and Dimou, M., 2012. Rank-size city dynamics in China and India, 1981–2004. *Regional Studies*, 46(6), pp.707-721.

Soo, Kwok Tong, 2014. "Zipf, Gibrat and geography: evidence from China, India and Brazil." *Papers in Regional Science* 93, no. 1: 159-181.

Figure A1. City size distribution and fitted Zipf's law in China and India

Source: Authors' analysis based on nighttime lights images from the National Oceanic and Atmospheric Administration.

Note: (a) Solid line: $\log(\text{ranking}-1/2) = 22.0 - 1.259 \cdot \log(\text{population})$, $N = 286$, $R^2 = 0.919$
 (b) Solid line: $\log(\text{ranking}-1/2) = 15.33 - 0.786 \cdot \log(\text{population})$, $N = 678$, $R^2 = 0.881$
 Dash line: $\log(\text{ranking}-1/2) = 18.13 - 0.997 \cdot \log(\text{population})$, $N = 542$, $R^2 = 0.961$
 (c) Solid line: $\log(\text{ranking}-1/2) = 16.69 - 0.914 \cdot \log(\text{population})$, $N = 5,749$, $R^2 = 0.988$
 Dash line: $\log(\text{ranking}-1/2) = 18.23 - 1.046 \cdot \log(\text{population})$, $N = 468$, $R^2 = 0.99$
 (d) Solid line: $\log(\text{ranking}-1/2) = 15.39 - 0.827 \cdot \log(\text{population})$, $N = 320$, $R^2 = 0.921$
 Dash line: $\log(\text{ranking}-1/2) = 16.63 - 0.92 \cdot \log(\text{population})$, $N = 295$, $R^2 = 0.965$