



BACKGROUND PAPER

Urban Agglomeration and Firm Innovation: Evidence from Asia

Liming Chen, Rana Hasan, and Yi Jiang

DISCLAIMER

This background paper was prepared for the report *Asian Development Outlook 2020: What Drives Innovation in Asia?* It is made available here to communicate the results of the underlying research work with the least possible delay. The manuscript of this paper therefore has not been prepared in accordance with the procedures appropriate to formally-edited texts.

The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the Asian Development Bank (ADB), its Board of Governors, or the governments they represent. The ADB does not guarantee the accuracy of the data included in this document and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

Any designation of or reference to a particular territory or geographic area, or use of the term “country” in this document, is not intended to make any judgments as to the legal or other status of any territory or area. Boundaries, colors, denominations, and other information shown on any map in this document do not imply any judgment on the part of the ADB concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

URBAN AGGLOMERATION AND FIRM INNOVATION: EVIDENCE FROM ASIA

Liming Chen¹

Young Professional, Economic Research and Regional Cooperation Department, Asian Development Bank (ADB)
limingchen@adb.org

Rana Hasan

Director, Economic Research and Regional Cooperation Department, ADB
rhasan@adb.org

Yi Jiang

Principal Economist, Economic Research and Regional Cooperation Department, ADB
yijiang@adb.org

January 2020

1. 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. 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.

Abstract

This paper presents evidence on the effects of urban agglomeration on firm innovation. It uses a unique dataset that consistently defines city boundaries and identifies firms' innovation activities across 25 developing countries in Asia. We find substantial gains from increases in city population in firms' propensity to undertake process and product innovation as well as research and development activities. These gains remain even after concerns regarding endogeneity are addressed using historical population data as instruments. In addition, we present evidence that agglomeration effects occur through the knowledge spillover channel, specifically through the presence of a top university in a given city and through firms' efforts in research and development. These findings confirm the existence and significance of urban economies of scale in augmenting the knowledge flows that generate innovation.

Keywords

agglomeration economies, knowledge spillovers, innovation

JEL Codes

O3, R11, O1

I. INTRODUCTION

Urban agglomeration plays an important role in fostering innovation.² Endogenous growth theory states that innovation and knowledge are significant contributors to the economic growth of nations (Romer 1986). As innovation investment has become more globalized, emerging countries in Asia are increasingly basing their economic growth on the capacity to generate new product and process innovations.³ Coinciding with ongoing economic development and increasing innovation, these countries also experience rapid urbanization.⁴

However, most evidence on the urban agglomeration effect on innovation activities is from developed economies. There is a lack of systemic evidence as to whether urban agglomeration affects innovation and, if so, what is the internal channel for it to take place in the developing world, including Asia.

This paper examines the effects of city size on three types of firm innovation activities—product innovation, process innovation, and research and development (R&D). We have constructed a unique dataset that defines agglomeration boundaries, or “natural cities”, using nighttime light (NTL) imagery across 25 developing Asian countries. The advantages of “natural cities” over administratively defined ones for this study are (i) that they are consistently defined and measured across countries and (ii) that they better represent urban agglomerations in which the firms operate as agglomeration units of human activities often extend beyond administrative boundaries. We have also identified the respective population sizes and firm innovation activities within these boundaries, which allows us to assess the correlation between the two. To address the endogeneity issue, we take an instrumental variable (IV) approach, using historical population figures as the source of exogenous variation for current city populations. We find statistically significant effects of city population size on all three firm innovation activities. All else being equal at our sample average level, doubling city size would increase a firm’s propensity for process innovation by 3.1 percentage points, product innovation propensity by 4.5 percentage points, and R&D propensity by 2.0 percentage points. Since, in our data for developing Asia, the biggest natural city is more than 2^{10} times as large of the smallest one, the effects of agglomeration on firm innovation propensity can be substantial.

We also estimate the agglomeration effects by country development levels and for large individual countries. The results show that an agglomeration effect is present in both the upper middle-income and the low- and lower middle-income country groups. Across countries, we find the agglomeration effect most present in the People’s Republic of China (PRC) and India—the two most populous countries in the region—and in countries from regional cooperation organizations such as the Greater Mekong Subregion (GMS) and the Association of Southeast Asian Nations (ASEAN). We do not find agglomeration effect in countries from the Central Asia Regional Economic Cooperation (CAREC) other than the PRC.

² Feldman and Audretsch (1999) and O hUallachain (1999).

³ Lundvall et al. (2009), Mahmood and Singh (2003); and Crescenzi, Rodriguez-Pose, and Storper (2012).

⁴ The urban population in the East Asia and Pacific region grew at an average of 3% per annum from 1960 to 2015, compared with 2.06% globally, and reached 1.2 billion or one third of the world’s urban dwellers by 2015 (Ellis and Roberts 2015 and Baker and Gadgil 2017).

The paper also investigates the channels for the agglomeration effect. We map the top 500 Asian universities into our natural cities, and find a statistically significant and positive effect of university presence on firm innovation. We also find the innovation outcomes resulting from a firm's R&D efforts to vary by city size. Firms investing in R&D are more likely to undertake product innovation in larger cities. Both results suggest knowledge spillover is the main channel for agglomeration effect on innovation to take place.

This paper is closely related to two strands of literature. The first is on the measurement of urban agglomeration economies, i.e., the concept that larger cities enjoy productive advantages over smaller ones (Marshall 1890). Much of the empirical evidence in this strand has focused on externalities on firm and worker productivity,⁵ rather than on innovation itself. Yet the volume of studies that measure the agglomeration effect on innovation activities remains thin. Nevertheless, the empirical evidence so far finds that innovation activity in a given country is heavily concentrated in a few cities, and that a city's size and density play a role in boosting innovation activity within that city. A few studies look at United States (US) patent data as an indicator for innovation. O'Hallachain (1999) find that large metropolitan areas dominate patent activity, which increase with the size of the metropolitan area. Feldman and Audretsch (1999) also find that innovation appear 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) use patents per capita as the dependent variable and find evidence of strong agglomeration effects for innovation. They estimate the elasticity of patenting per capita with respect to employment density to be approximately 20% in metropolitan areas in the US. This number is several times higher than the estimates for the elasticity of wages in the literature. Carlino and Hunt (2009) further confirm the earlier result by weighting patents by their number of citations.

Although these studies provide evidence that agglomeration effects on innovation are quantitatively important and pervasive, they shed little light on the internal mechanisms of how agglomeration augments innovation.

The second strand of related literature focus on identifying potential mechanisms. Duranton and Puga (2004) propose three channels: sharing, matching, and knowledge spillover—for firm and labor productivity to benefit from urban agglomeration.⁶ Carlino and Kerr (2015) argue that these three mechanisms also apply to innovation, with more empirical evidence existing on the knowledge spillover channel (Carlino and Kerr 2015).

The first channel, *sharing*, has to do with thick factor markets that arise when innovative activities cluster locally. 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

⁵ The common technique is to regress the productivity (measured by labor productivity or total factor productivity for firms and by wages for workers) to the agglomeration size (measured by urban population or employment) or density. Rosenthal and Strange (2004); Melo, Graham, and Noland (2009); Puga (2010); and Baum-Snow and Ferreira (2015) provided extensive surveys of methodology and evidence.

⁶ Comprehensive surveys are provided by Rosenthal and Strange (2004), Combes and Gobillon (2015), and Duranton and Kerr (2015).

flow rapidly (Porter 1993 and Helsley and Strange 2002).⁷ 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.

The second channel of agglomeration effect is *matching*. It results from thick local labor markets, in which the quality of matches between firms and workers is improved (Helsley and Strange 1990; Berliant, Reed, and Wang 2006; and Strange, Heijazi, and Tang 2006).

The third channel, *knowledge spillover*, refers to intellectual gains made by exchange of information for which no direct compensation is given to the producer of the knowledge, resulting in higher degrees of innovation in those locations where economic activity is dense. They are particularly important in explaining the concentration of innovation because they depend on new knowledge more than other economic activities.

One approach of studying localized knowledge spillover is the examination of innovation colocation within highly innovative cities. This was led by Jaffe, Trajtenberg, and Henderson (1993), who use patent citations within a given city. They find that patents are 5–10 times more likely to cite other patents from innovators who are spatially closer, compared with similar patents from innovators who resided farther away. Further refinement of this topic can be found in Jaffe et al. (2000), Thompson and Fox-Kean (2005), and Murata et al. (2014).

An alternative approach is to examine the localized knowledge spillover created by the presence of a university or its R&D activities on firms' commercial innovations (Jaffe 1989; Audretsch and Feldman 1996; Audretsch and Stephan 1996; Anselin, Varga, and Acs 1997; Carlino et al. 2007; and Andrews 2017).⁸ There are many ways for knowledge spillover to take place between universities and firms: (i) university-firm collaborations through market-mediated interactions; (ii) unintended knowledge flow from university-based research (D'Este and Iammarino 2010 and D'Este and Patel 2007);⁹ and (iii) university as human capital producer (Toivanen and Vaananen 2016). The quality of academic research also indirectly affects the quality of firm innovation (Baba, Shichijo, and Sedita 2009; Maietta 2015; Mansfield and Lee 1996; and Zucker et al. 1998).

Although both strands of literature have developed quickly, the evidence mostly draws from the US and other advanced economies. Meanwhile, evidence from developing countries is still extremely thin, with considerable knowledge gaps remaining (Duranton 2014). One exception is Nieto Galindo (2007), which finds a significant share of firms in Colombia engage in product and process innovation, but generate minimal patents. More than 70% of these innovations were

⁷ Helsley and Strange (2002) developed a dynamic model of innovation to illustrate the idea that a dense network of input suppliers facilitates innovation by lowering the cost of bringing new ideas to realization.

⁸ Most of the evidence here is from the state or metropolitan level in the US.

⁹ Geographical proximity plays a fundamental role in determining both influence of university-based research (Fischer and Varga 2003) and intensity of collaborations (D'Este, Iammarino, and Guy 2013), with a heterogeneous effect on formal and informal knowledge transmission (Audretsch and Feldman 1996; Morgan 2004; and Leten, Landoni, and Van Looy 2014).

concentrated in three main cities, which hosted less than 40% of the country's population. Available cross-country evidence of the developing world is even scarcer.¹⁰

On the other hand, the innovation activity in the developing countries may differ considerably from the experience with developed countries. Acemoglu, Aghion, and Zilibotti (2006) distinguish between developed economies located at the technological frontier and developing countries within this frontier. They argue that the growth problem for developed economies is to push the frontier, where formal innovations (such as patents) and R&D play a crucial role. For developing countries. However, the problem is one of catching up with the frontier. To some extent, this is about firms' capacity for product and process innovation, i.e., to produce new and better products and to produce them more efficiently. Therefore, it is important to study urban agglomeration with respect to innovation in developing countries.

This paper contributes to the existing literature in three ways. First, it fills the gap with some compelling evidence on the existence and magnitude of agglomeration on innovation activities within developing Asian countries. We do this using a unique dataset that contains consistent measures both of city size and firm innovation across nearly 500 cities across 25 countries, and utilizing historical urban population data to address endogeneity issues. Second, it shows the importance of the knowledge spillover channel in creating an agglomeration effect on innovation. It demonstrates how top universities matter and explores how returns to firm R&D vary by city size.

The remainder of this paper is organized as follows. The next section introduces our data section III describes our empirical strategy, and section IV presents the main results of agglomeration effects on firm innovation activities, followed by robustness test and heterogenous estimation. Section V reports and discusses findings on potential channels. The final section concludes the paper.

II. DATA

We assemble a unique cross-sectional dataset covering nearly 500 cities of 25 developing economies across Asia and the Pacific. The firm data come from World Bank Enterprise Surveys (WBES), which contain firms' innovation activities as well as their basic characteristics. These data are merged with city data constructed using NTL satellite imagery. City characteristics such as population, presence of university, weather, and geography are estimated or obtained from various sources.

A. Data on Firm Innovation

WBES are firm-level surveys conducted in developing economies. Each survey consists of a cross-sectional representative sample of firms from an economy's private sector, excluding the agriculture sector and all state-owned enterprises. Surveyed firms are selected through a

¹⁰ Most research from the developing world is about agglomeration effect on labor and firm productivity. For example, a recent paper by Quintero and Roberts (2018) estimates city productivity premiums for 16 countries in the Latin America and Caribbean regions. Duranton (2008), Overman and Venables (2005), Henderson (2005), and Duranton (2014) provided detailed review of studies from the developing world.

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.

The survey questions cover a wide range of topics, with some of them country specific. The World Bank offers a harmonized dataset that extracts and standardizes common information from each survey. This is the dataset we use for this study. In addition to a firm's basic information—such as number of employees, age, sector, and foreign direct investment (FDI) share—the dataset records information about a firm's innovation activity through three questions:

- (i) During the last 3 years, has this 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.
- (ii) During the last 3 years, has this establishment introduced new or improved product or service?
- (iii) During 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 process, product, and inputs, respectively.¹¹

The subjectivity of surveys may raise concerns to the validity of the self-reported innovation variables or WBES in general as they may not measure innovation as objectively as peer-reviewed sources, e.g., patents. But these innovation measures and WBES can be considered reliable in a few ways. First, the survey organization (World Bank) has taken several steps to ensure reliability (Ayyagari, Demirgüç-Kunt, and Maksimovic 2011) detailed the steps taken by y. The firm identifiers are kept confidential and the survey is conducted by private organizations independent from government. Various logistic measures are taken before, during, and after the survey to ensure that questions elicit valid answers, including translation and localization checks, two interviewees for the same firm, and post-survey consistency checks. Second, to the extent that there may exist systematic measurement error for innovation variables because innovation may be interpreted differently depending on country's development level or technology advancement, we include country year fixed effects in all our specifications to mitigate this concern. Third, detailed descriptions of product and process innovation reveals that firms have valid reasons to consider their reported products or practices as "innovative", such as increased profit, reduced cost, or reduced production time as a direct result. Fourth, in robustness test we use two additional innovation measures based on the raw measures and the results are not sensitive to them. The first measure is "main market product innovation" from a follow-up

¹¹ WBES also record descriptions of process and product innovations when available. As defined in the questionnaire, these innovations need not necessarily be cutting-edge inventions or revolutionary production processes. They could be small improvements on existing processes and products. From our sample, process innovations can be activities such as a firm in Yangon, Myanmar including a door-to-door sales service or running an ad campaign on Facebook. Product innovation can be a firm in Cebu, Philippines introducing Panda pao, or a sawmilling enterprise in Bago, Myanmar now producing shelves for Buddha images.

question to question (ii) on product innovation. The question states: Were any of the new or improved products or services also new for the establishment's main market? The answer to this question is also yes or no. We consider this question a stronger version of product innovation.¹² The second measure is "firm innovation" constructed from questions (i) and (ii). The variable takes value of 1 if firm has reported product or process innovation and 0 otherwise. Last, there has been a large number of studies using data from the survey (Djankov et al. 2003, Barth et al. 2009, and Besley 2018), including using innovation variables (Ayyagari et al. 2011, Paunov 2016, and Paunov and Rollo 2016).¹³

To study how urban agglomeration affects firms' innovation activities, we need to access firms' urban locations, ideally measured in a consistent way across countries. WBES indicate 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 show only the region (e.g., north or south) of the surveyed firms, and the Indian data show only states. Even if the geographic units represent cities in some countries, they are not comparable since the official definitions of cities differ greatly across countries (United Nations 2018).

Alternatively, the WBES conducted since 2012 contain the geographic coordinates of the firms surveyed. In the publicly available data, firms are randomly shifted up to 2 kilometers (km) to mask their true locations. Based on these coordinates, we can match the firms with the geocoded natural city data to associate firms with the cities in which they are located.

B. Data on Natural Cities Based on Nighttime Satellite Imagery

We construct a dataset of more than 1,500 cities across Asia and the Pacific, using NTL satellite data available since 1992. The imagery obtained tracks the footprints of these cities from 1992 to 2016 as contiguous illuminated areas. Individual city characteristics, including population, presence of university, weather, geography, and historical population are ascertained using various data sources. For example, to estimate population, we fill the delineated urban area with grid population data from LandScan, which is available at approximately 1 km spatial resolution. We then tally all cells falling within, or interacting with, the city contour. We refer to these urban agglomerations as natural cities to distinguish them from administratively defined cities.¹⁴

The advantages of using natural cities, at least for this study, are threefold. Firstly, natural cities are uniformly defined and their characteristics are measured consistently across countries and time. Second, they offer a better representation of the urban agglomerations in which the firms operate, compared with the administratively defined cities (many of which contain both dense urban areas and sparsely populated rural areas). The urban areas of natural cities often expand beyond the administrative boundaries and spread over multiple administrative units. Third, the natural cities are geocoded and this information can readily be merged with WBES firm data.

¹² There is no corresponding follow-up question on process innovation.

¹³ Literature has also found that some subjective firm responses from the survey have a high degree of correlation with those objective outcomes from external data sources (Hallward-Driemeier and Aterido 2009).

¹⁴ For a more detailed account of the process of developing the dataset, please go to Appendix 1 or see Jiang (2019).

We map the WBES firms into natural cities using geographic information system software. To optimize the matching rate, we use 2016 natural city boundaries and allow those firms falling 2 km or less outside a natural city boundary to be assigned to that city.¹⁵ Table A1 in Appendix 1 shows the matching results by country. The overall success rate is 87%, with matching within individual countries ranging from 40% (Mongolia) to 100% (Myanmar and the PRC)¹¹. The process matches a total of 21,857 firms to 489 natural cities, an average of 45 firms per city. India alone contributes 8,100 firms, or 37% of the total sample, which were distributed in 207 natural cities. Figure 1 shows the geographic distributions of our natural cities with WBES firms.

C. Summary Statistics

Table 1 reports the summary statistics of firm-level and city-level variables used in the study (upper panel for the full sample). Around 50% of the surveyed firms reported that they had introduced new or improved processes in the past 3 years. The proportion goes down to 36% for new or improved products or services (21% for products that are also new for the establishments' main market), and 24% for any expenditure on R&D.

Following the existing literature, we define those firms that had operated for fewer than 10 years as young firms (Reyes, Robert, and Xu 2017); those that had 50 or fewer permanent employees as small firms (Beck, Demirgüç-Kunt, and Maksimovic 2008 and Reyes et al. 2017); and those that had 10% or more coming from FDI as FDI firms¹⁶. About 21% of firms across the whole sample were young, 65% were small, and only 6% were FDI firms. About two thirds of firms are from the manufacturing sector, with the rest from the services sector. About 41% of the firms are headquarters and the average share of skilled workers is around 35%.

We use 2010 values for time-variant city characteristics, such as population and weather, taking into account that the innovation activities captured by the surveys were concentrated in 2012–2016. Overall, the sample cities are highly diverse in terms of the dimensions measured. City population ranges from 8,000 to 35.8 million, with the average at 1.4 million and the median at 438,000. About 20% of the 489 cities host at least one university that rank among the top 500 in Asia, while there is no top 500 university in the remaining 80% of cities. The average distance of the cities to the nearest seaport is 400 km, with standard deviation of 500 km. Minimum temperature has a larger variance than does maximum temperature, and the city with the most rugged terrain has an index 15 times greater than the sample mean for ruggedness.

D. Spatial Concentration of Innovation

One stylized fact that is documented in the existing literature is that innovation is highly spatially concentrated. Within a given country, there are generally a few cities acting as “innovation hubs”. These hubs host a large share of the country’s innovative activities, and this share is typically disproportionate to each city’s share of the country’s total population. For instance,

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, we use the prefecture information available in the WBES data for the PRC, and assign each unmatched firm to the largest natural city in its prefecture. By doing so, we achieve a 100% matching rate.

¹⁶ Literature often follows the International Monetary Fund standard definition of FDI, i.e., the dependent variable (FDI) takes the value of 1 if at least 10% of the firm’s ownership is foreign, and the value of 0 otherwise. For example, see Kinda (2010).

Nieto Galindo (2007) found that over 70% of Colombia's innovations are concentrated in three main cities, which together host less than 40% of the country's population. In the US, Moretti (2019) shows that the top 10 innovative cities in computer science, semiconductors, and biology and chemistry account for 70%, 79%, and 59% of inventors, respectively.

Similar patterns are clearly observed in our data. We calculate each city's share of firms with process innovation, product innovation, and R&D expenditure, then compare these figures with each city's share of population as a percentage of the national total. A city with a higher share of firms engaged in innovation in comparison with its share of total population reflects a concentration of innovation activities. In the PRC, 10 cities host 64% of the country's firms involved in process innovation, 72% of firms involved in product innovation, and 64% of firms investing in R&D, while together accounting for only 55% of the national population. The contrast is even starker for India, where the top 10 cities host 70% of firms involved in process innovation, 76% involved in product innovation, and 76% undertaking R&D, while together accounting for only 43% of the country's population.

Figure 2 plots the cumulative share of firms against cumulative population share for India, Indonesia, Kazakhstan, Malaysia, and the PRC—the five countries with the most natural cities in our sample. Each data point in the figure represents one natural city, and the cities are ranked by their share of firms involved in each of the three innovation activities. All the curves across the three panels in the figure demonstrate a Lorenz-type convex feature, implying consistent spatial concentration of innovation across different countries and innovation activities. In general, the curves of India and Malaysia are more convex than those of the PRC, suggesting a higher degree of innovation concentration in those two countries.

Figure 2 also shows that many large cities have a high share of innovative firms (segments with a relatively long projection on the horizontal axis and a slope greater than 45 degrees). However, we also see some mega cities with a disproportionately low share of innovative firms (segments with a long projection on the horizontal axis and a slope less than 45 degrees). For instance, Kuala Lumpur in Malaysia, Jakarta in Indonesia, and Almaty in Kazakhstan all have a share of national population larger than the corresponding share of innovative firms.

How does city size affect firms' innovative activities? To what extent is this relationship causal? What are the channels used in order for an agglomeration effect to take place? These are the questions the rest of this paper is designed to answer.

III. EMPIRICAL STRATEGY

A. Baseline Probit Model

Our main aim is to investigate the impact of urban agglomeration on a firm's propensity to undertake product innovation, process innovation, and R&D activity. Given this research proposition and the nature of the data we have at our disposal, the econometric analysis is most appropriately undertaken by employing a probit model. This allows us to estimate the likelihood of a firm partaking in a particular type of innovation activity as a function of characteristics of the firm itself as well as the city in which it is domiciled. In the probit model, we control country

fixed effects. This allowed us to account for unobservable differences in the innovation environment across the 25 developing Asian countries, including variations in institutions, culture, and nationwide policies that promote innovation. We controlled year fixed effects, which captures difference across year such as macroeconomic trends that affect firm dynamics across country and sectors. We also control sector fixed effects. Therefore, our results derive from linking firm innovation across different cities to city-level characteristics, while holding country, year, and sector factors constant.

The key explanatory variable of interest is city population in log form. To partially mitigate potential bias induced by endogeneity or reverse causation, we chose population as at 2010, or 2–6 years prior to the administration of firm surveys. A number of first-nature geographic characteristics could affect innovation and city population (Combes et al. 2010). These include average rainfall, maximum highest temperature, minimum lowest temperature, average terrain ruggedness, and city's distance to the nearest port. We incorporated these as city-level controls. Firm-level controls included firm size (small firm if total employment is fewer than 50 people), firm age (young firm if operation duration is fewer than 10 years), and whether a firm has a significant FDI component (over 10% foreign ownership).

We estimate three baseline models, relating to each of the three types of firm innovation activity: product innovation, process innovation, and R&D activity. The underlying latent model for our probit model can be written as:

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

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

where:

Subscripts i, c, n, s, t represent firm, city, country, sector, and year, respectively.

Y_{icnst} 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_{icnst}^* is the latent variable for outcome variable Y_{icnst} , which can be interpreted as a firm's propensity to innovate. Its realization is not observed.

$\ln P_{cn}$ is the natural log population of 2010 in city c . Of primary interest is its coefficient, β_p . A positive estimate suggests there exists agglomeration effects on firms' innovation activities.

$CITY_{cn}$ is a vector of city controls. These include average rainfall, maximum and minimum temperature, average terrain ruggedness, and city's distance to the nearest port. These variables are measured at 2010 and are time invariant.

$FIRM_{icnst}$ is a vector of firm controls. These include firm age, size, and FDI ownership.

$\delta_n, \sigma_s, \gamma_t$ are country, sector, and year fixed effects.

ϵ_{icnst} is the random error term. We clustered the standard errors at the country level in the estimation to account for factors driving correlations in the unobservables, such as national technology policies.

Let X_{icnst} be the vector of covariates and β be the vector of coefficients listed above, so $y_{icnst}^* = X'_{icnst}\beta + \epsilon_{icnst}$. Then denote $\Phi(\cdot)$ and $\phi(\cdot)$ the CDF and PDF of standard normal distribution. The probability of firm innovation activity conditional on X_{icnst} can thus be written as

$$P(Y_{icnst} = 1|X_{icnst}) = \Phi(X^T\beta).$$

The marginal effect to innovation probability from percentage change in city population can be derived as

$$\frac{\Delta P(Y_{icnst} = 1|X_{icnst})}{\Delta P_{cn}/P_{cn}} = \frac{\Delta P(Y_{icnst} = 1|X_{icnst})}{\Delta \ln P_{cn}} = \phi(X^T\beta) \cdot \beta_p,$$

Results in the next section will present probit model coefficients as well as marginal effects from city population.

B. Instrument Variable

Endogeneity issues could arise in our baseline specification, resulting in bias in the estimated coefficients. One concern is that a city's size and a firm's innovation activities may be determined simultaneously (Moomaw 1981). This could happen if the innovative outputs of firms expand production scale, thereby attracting more employment to a city and resulting in reverse causality (Duranton 2007 and 2014). Another explanation for simultaneity is missing local variables that are correlated with both city size and innovation. For example, highly skilled workers at innovative firms could be attracted to large cities by amenities not adequately controlled in our regressions.

Following Ciccone and Hall (1996) and Combes et al. (2010), we perform instrumental variables (IV-Probit) regression to address these endogeneity possibilities, using historical populations from World Urbanization Prospects (WUP) to instrument for the 2010 population used in the baseline specification. WUP collects urban populations for over 200 countries and covers cities with populations over 300,000 population from 1950.¹⁷ A detailed discussion on the pros and cons of WUP data quality can be found in Buettner (2015). Despite limitations, the data have been found useful in studying urban issues, especially in the developing world where available datasets are scarce (Henderson 2000, Montgomery et al. 2004, and World Bank 2008).

Cities within the WUP data are defined by administrative boundaries. We reconciled this with our natural cities by mapping the WUP cities into our natural city boundaries using their latitude and longitude coordinates. Of 789 WUP cities available across 23 of the 25 countries included in our study,¹⁸ 767 WUP cities could be mapped into 287 of our 489 natural cities. Of the 202

¹⁷ WUP only includes cities with population of 300,000 and above at the most recent year (2018). Data and documentation of this dataset can be accessed at <https://population.un.org/wup/>.

¹⁸ WUP data do not cover Bhutan and Timor-Leste. This brings us down by three cities and 292 firm observations in IV analysis.

natural cities that did not have corresponding WUP cities, the majority were relatively small, with an average population of 277,829 in 2010. This compares with the average population of 2,227,414 in 2010 for the other 287 natural cities (Table 1). As a result, although we are down to 289 natural cities in our IV analysis, we retain 88% of our WBES firm sample because most of the surveyed firms are not located in the smaller cities. We consider this slight loss of estimation power to be a small sacrifice for dealing with endogeneity. For our instrument, we use the annual average of 1950–1959 population, the earliest available period from WUP. Using the earliest period helps further guarantee exogeneity, and a 10-year average smoothing helps avoid data inaccuracies or outliers in WUP historical data.

The validity of this instrument relies on two conditions. First, there is some persistence in the spatial distribution of population. Second, the local drivers of the dependent variable, in our case innovation activity, at the time of sampling period (2012–2016) differ from those of the past (Combes et al. 2010). The historical population data we chose should satisfy these two conditions.

For the first condition to hold, the instrument must have a strong correlation to our 2010 instrumented variable. The correlation of log population (average of 1950–1959) and log population (2010) was 0.87 and was significant at 1%, indicating the persistence in the population spatial distribution. The first stage IV regression, shown later, further confirmed its relevance.

For the second condition to hold, our instrument must be orthogonal to the error term. In other words, we require the instrument that affects innovation only through the spatial distribution of population. We argue that our instrument satisfies the exogeneity condition for following reasons.

First, long-lagged values of the same variable will remove any simultaneity or reverse causality bias. It is highly implausible that city population distribution in the 25 developing Asian countries would be determined by firm innovation activities in the 2010s, or by any other contemporaneous local shocks. That being said, some permanent city geographic characteristics, such as climate or proximity to coastal areas, may indeed explain both past population spatial distribution and current innovation outcomes. We solved this problem by directly controlling for these first-nature city characteristics in $CITY_{cn}$, e.g., temperature, rainfall, and distance to port.

Second, after controlling time-invariant geographic characteristics, the drivers of urban agglomeration in the 1950s are not related to determinants of local innovation activities in the 2010s. This is because the current economies of the 25 developing Asian countries are very different to what they were 60 years before. Take the two biggest countries in our sample, India and the PRC, for example. In the early 1950s, both India and the PRC had just experienced drastic regime changes. India had regained independence from British colonialism, while the PRC was recovering from decade-long wars and the main economic activity was reconstruction. In addition, during the latter half of the 20th Century, both countries experienced significant reforms. The PRC initiated its reform program and opened its economy in 1978. These reforms are directly responsible for the country's structural transformation and 500 million increase in

urban population from 1978 to 2012 (World Bank 2014). What is more relevant in the case of the PRC is that this period saw the rise of some of the most innovative cities, such as Shenzhen, a small aquaculture port city in the 1970s. This illustrates significant spatial change in both urban population concentration and innovation activities. India also has undergone a series of economic reforms, including the liberalization program initiated in 1991 (Bhagwati and Panagariya 2013). The sectoral composition of India's economy has also changed significantly. The agriculture share of gross domestic product was 55% in the 1950s and fell to 22% in the 2000s. Meanwhile, the services sector increased from 30% in the 1950s to 54% in the 2000s (Mukherjee 2013).

Other countries in our sample also experienced significant economic changes. For instance, most countries in Central Asia and Wwest Asia were a part of the former Soviet Union until its dissolution. With much change across these developing Asian countries, it is highly plausible that past determinants of population spatial distribution are not major drivers of current innovation activities.

IV. MAIN RESULT

A. Baseline and IV Estimates

Table 3 presents our baseline results. The first three columns report three baseline probit models for full sample estimates. The last three columns report results using the same baseline models, but restricting observations to those with historical urban population data (thereafter IV sample). The IV sample probit estimates provide direct comparison to the IV estimates and robustness check results.

The coefficients of our main variable of interest, logarithm population in 2010, are positively significant at 1% in all three models. In other words, holding everything else constant, firms residing in cities with larger populations are more likely to partake in R&D and implement both product and process innovations. For process innovation, the predicted propensity when the independent variables are at mean values is 42.7%. The coefficient corresponds to a marginal effect of 0.0336. This implies that, if the population size in a city doubles while keeping the rest of the independent variables constant at their mean values, the propensity of a firm to implement process innovation increases on average by 3.36 percentage points, or an 7.9% increase from the predicted propensity of 42.7%. For product innovation, the marginal effect is 4.21 percentage points, or a 13.4% increase from the predicted propensity of 31.4%. For R&D, the marginal effect is 2.68 percentage points, or a 13.7% increase from the predicted propensity of 19.5%. In column (4), (5), and (6) of Table 3, we perform the same regressions as in columns (1), (2), and (3) to the IV sample and obtained comparable results. The marginal effects and predicted propensities are slightly different: a 7.0% increase for process innovation, a 16.4% increase for product innovation, and an 11.4% increase for R&D, if the population size doubles.

Other city characteristics are also correlated with innovation 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; and, perhaps surprisingly, the coefficients for both terrain ruggedness and distance to the nearest port are found to be positive in three models. In

other words, 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. The coefficients on product innovation are found to be positive although not significant. Firm size matters for all three innovation activities. Smaller firms are less likely to invest in R&D or participate in either product or process innovation. Firms with foreign ownership are more likely to invest in R&D and generate product innovations. Headquarters are more likely to implement product innovation. These findings on firm characteristics are consistent with existing literature (Bertschek 1995, Huergo and Jaumandreu 2004, and Yang 2017).

In Table 4, we instrument city population size with the average of 1950–1959 population size. First, the instrument is quite strong, with first-stage F-statistics above 5,000. This further validates the first condition for the instrument.

Second, a larger city size is still associated with a higher propensity for firm innovation, even after addressing endogeneity. Compared with their corresponding probit coefficients in column (4), (5), and (6) of Table 2, the second-stage coefficients remain significant at the 1% level, but the marginal effects are now slightly different. For process innovation, if population size doubles, the predicted propensity for process innovation will increase by 3.14 percentage points from an average propensity of 44.0%, or a 7.1% increase. For product innovation, the effect becomes smaller: an increase of 4.54 percentage points from a predicted average propensity of 32.6%, or a 13.9% increase. For firm R&D, the coefficient in the IV estimates is now significant at the 5% level; the marginal effect is an increase of 1.84 percentage points from predicted propensity of 20.9%, or a 8.8% increase.

To summarize, after accounting for endogeneity, we still find economically and statistically significant effects of city size on firms' innovation inputs and outputs. The average propensity for firms to engage in process innovation, product innovation, and R&D increased from 1.8 percentage points to 4.5 percentage points, or 7% to 14% when city size doubles. These estimates are larger than the elasticity of wage or firm productivity with respect to city employment or urban density, which is found to be between 2% and 10% (Duranton 2014). On the other hand, the estimates are smaller than the 20% elasticity of patent intensity with respect to employment density (Carlino et al. 2007).

B. Robustness Testing and Country Heterogeneity

To examine a few factors that might potentially swing our results, we perform robustness checks by restricting our sample in a few specifications. We first consider the possibility that the megacities might dominate our results. In addition, we consider alternative definitions of innovation. Most of the results are not affected for the tested specifications.

First, we test whether our results might be driven by a few megacities (such as Guangzhou in the PRC and Delhi in India) that host a large number of innovative firms. In other words, the agglomeration effect only significantly contributes to firm innovation activities in a few very

large cities, while not so much in others. We check this potential outlier effect by excluding 13 cities with populations of over 10 million.¹⁹ This leave us with 274 natural cities and 13,567 observations. The results are presented in columns (1), (2), and (3) of Table 4. The first stage F-statistics drops to over 3,600, although they still indicate very strong IV. In the second stage, none of the significance levels for population size is affected by this test. The marginal effects for process innovation and R&D are much larger, indicating a possible stronger agglomeration effect for non-mega cities.

Second, we use two alternative innovation measures to test if our results are sensitive to the definition of innovations. The first measure we use is from a follow-up question to product innovation that gauges whether the product innovation is “also innovative to the establishment’s main market”. In the sample, 21% of the firms reported main market product innovation compared with 36% firms that reported product innovation. The difference indicates that some firms consider their new product innovative for themselves but not for the market (for example, they could be adopting existing product modifications to their lineup) while the remaining consider their new products are truly innovative, at least in their main market. This follow-up question could provide a “stricter” measure for product innovation. Result shows that coefficient on population is still significant, and the percentage increase from predicted probability is 14.2% if city size doubles (Table 4, column 4), only slightly different from 13.9% using product innovation measure. The second measure is an “aggregated innovation” dummy which takes value of 1 if firm conducts either product or process innovation. The coefficient is significant at 1% level and the marginal effect is 5.0 percentage points or 8.5% increase from predicted probability.

There exists considerable heterogeneity of development progress among our 25 selected countries. To test the extent that the agglomeration effect on firm innovation varies at different stages of a country’s development, we categorize our sample into upper-middle income countries and low- or lower middle-income countries, using the latest World Bank country classifications.²⁰ We then perform endogenous probit analysis on each category (Table 5). For low- and lower middle-income countries, the agglomeration effects on product innovation and R&D are statistically significant, with effects slightly smaller than for the pooled sample, but the coefficient for process innovation is small and insignificant. For upper middle-income countries, the agglomeration effects on process and product innovation are relatively consistent with baseline IV results, while the effect on R&D is positive but not significant. Overall, we find some differences in the agglomeration effects with respect to R&D investment and process innovation when assessing various stages of country development. Meanwhile, agglomeration effects exist consistently for product innovation, regardless of a country’s development level.

¹⁹ Natural cities in our sample with populations over 10 million include: Bangkok, Beijing, Delhi, Dhaka, Guangzhou, Ho Chi Minh, Jakarta, Karachi, Kolkata, Lahore, Mumbai, Quezon City, and Shanghai.

²⁰ When our analysis was performed, the World Bank defines low-income economies as those with a gross national income (GNI) per capita (calculated using the World Bank Atlas method) of \$1,025 or less in 2018. Lower middle-income economies are those with a GNI per capita between \$1,026 and \$3,995. Upper middle-income economies are those with a GNI per capita between \$3,996 and \$12,375. High-income economies are those with a GNI per capita of \$12,376 or more.

The 25 countries in our sample are diverse. It includes two of the most populous economies and countries that are similar socioeconomically. Even though we included country fixed effect in our previous analysis, it is worth examining how the agglomeration effects on firms' innovation differ in India, the PRC, and countries from several regional cooperation groups such as GMS, ASEAN, and CAREC. The results are summarized in Table 6. For India, all three coefficient estimates are sizable and significant at the 1% level, suggesting substantial impacts of city size on Indian firms' innovation activities. For the PRC, we find agglomeration effects in both process and product innovation, but not for R&D investment, which is largely in line with the results for upper middle-income countries. Due to the limited number of cities in the rest of the countries, we group them together through regional cooperation organizations. We find agglomeration effects in GMS/ASEAN countries pooled together for all innovation activities, except for R&D in ASEAN. The same cannot be said for CAREC countries other than the PRC, where we detect agglomeration effect in none of the innovation activities.

V. CHANNELS FOR AGGLOMERATION

So far, we have established the agglomeration effect of city size on firm innovation. In this section, we investigate the potential channels for such an agglomeration effect. We test two hypotheses that could take place at the city level: (i) that the agglomeration effect matters differently for R&D firms and non-R&D firms and (ii) that the presence of a top-ranked university is associated with a higher propensity for firms to implement product and process innovations. Both hypotheses are related to the knowledge spillover channel studied in existing literature.

To benchmark the analysis of the knowledge spillover channel, we first include a firm R&D dummy as an explanatory variable for the process innovation and product innovation models. In other words, we here focus on innovation outputs. The results are reported in columns (1) and (2) of Table 7. Not surprisingly, firm R&D exhibits significant impacts on innovation outputs, while the agglomeration effects remain positive and significant.²¹

Local characteristics are found to have heterogeneous effects on the output of innovative firms and those non-innovative (Yang 2017). In the context of this paper, the indication that city size augments the return of firms' R&D efforts may arise for two reasons. First, as shown in previous analysis, firms in larger cities are more likely to invest in R&D. These firms could have higher efficiency of R&D investments through formal collaborations or informal exchanges of information with other R&D firms in the same locality. Second, a larger city often implies a larger pool of experts because of the higher population and more R&D job opportunities. Therefore, firms in larger cities would be more efficiently matched with the most suitable R&D specialists.

To test this hypothesis, we add the interaction term between city size and the R&D dummy to the models, which is instrumented by the interaction between firm R&D and historical population. A

²¹ The agglomeration effects were also nearly identical to the baseline case when using the same sample. When not controlling firm R&D but using only the same sample in Table 7, the coefficient of the agglomeration effect was 0.105 for process innovation and 0.151 for product innovation. Both are significant at the 1% level.

positive coefficient on this term implies that R&D firms reap more agglomeration benefits than do non-R&D firms. Results for process and product innovations, respectively, are reported in columns (2) and (4) of Table 7. The first-stage estimates for both endogenous variables are strong. In the second stage, two results are of interest. First, the coefficient of the interaction term is small and insignificant, suggesting little heterogeneous agglomeration effect on process innovation by R&D and non-R&D firms. Further, including the interaction term has only little effect on the coefficient of population. However, for product innovation, the coefficient of the interaction term is positive and significant at the 1% level, indicating that the effect of R&D on innovative product outputs increases with city size. Meanwhile, the estimated effect of population size decreases by half compared to the no interaction case.

Our second hypothesis is that the identified agglomeration benefits may result from the presence in the local area of one or more top tier higher education institutions. Universities, especially the prestigious ones, are the pioneers of pushing knowledge frontier and harbingers of exploring the uncharted. Countries allocate abundance of resources to the top ranked universities to boost innovation capacity. Geographical proximity to a university plays a fundamental role in determining firm innovation. There are several forms of university-firm interaction through which the spillover could take place: university-firm collaborations through market-mediated interactions, unintended knowledge flow from university-based research (D'Este and Iammarino 2010 and D'Este and Patel 2007), and university as human capital producer (Toivanen and Vaananen 2016).

To measure the availability of quality tertiary education, the study utilizes the recent QS World University Rankings, which identify the top 500 universities in Asia. Of the top 500 universities in Asia, 248 are mapped into 99 cities among nine developing Asian countries. Of the top 500 universities in Asia, 248 are mapped into 99 cities among nine developing Asian countries (Table 8). The spatial pattern shows that there is a high degree of concentration of top universities. First, the top universities are unevenly distributed across countries. Only nine of the 25 developing Asia countries have top university presence, and 211 (85% of total) of them are located among five countries: the PRC (78), India (64), Malaysia (25), Pakistan (22), and Indonesia (22). Second, within a country the universities are also unevenly distributed across cities, with some countries more concentrated than others. For example, 14 of 25 top universities in Malaysia are located in Kuala Lumpur and 12 of 18 in Thailand are located in Bangkok, making them the second and third cities with the most top universities in developing Asia.²² On the other hand, the degree of concentration is less in India: 64 top universities are distributed in 43 cities, with Delhi being the city with the most (7) universities.

The aforementioned high degree concentrations of university and firm innovation activities suggest that there could exist correlation. To investigate whether firms that are in cities with top university presence are more likely to innovate, we adopt a logit model with firm innovation activity as dependent variable and university dummy as main explanatory variable. The university dummy indicates whether the city has one or more top universities. Results show that the presence of a top university is strongly associated with a higher propensity of product and

²² The city with most top universities is Beijing, hosting 20 of 78 top universities in the PRC.

process innovation (Table 10). The coefficients are positive and significant in all specifications. The odds that firms innovate in new production processes are 72% higher in cities with at least one top university than for firms in cities without any top universities (Table 10, column 1). After controlling for firm characteristics, sector, year, and county fixed effects, the odds are slightly higher at 77% (Table 10, column 3). The positive effect of top universities on process innovation is even more pronounced after controlling for city characteristics (Table 10, column 4), indicating that firms in cities with top university presence have twice the odds to undertake process innovation than the firms in cities without any top university presence. For product innovation, the correlation is smaller (30% more odds for local firms to innovate) but remain significantly positive (Table 10, column 9).

To investigate whether more university further promotes local firm innovation, the same model is applied with the number of universities as main explanatory variable. For an additional top university in the city where a firm is located, there is a 5%–8% increase in the odds that firms innovate in new production processes (Table 5, columns 4); for product innovation, there is a 4%–7% increase in the odds that firms innovate for an additional top-ranked university nearby firms depending on the model specification (Table 5, columns 6–8). However, the coefficient becomes insignificant when city level controls are included (Table 2, column 9).

To summarize, through examining the knowledge spillover channel for the identified agglomeration effects, we identify that the agglomeration effects would benefit R&D firms more in terms of product innovation. Firms that locate in cities with top university presence are more likely to innovate in new product and process. Holding everything else constant, firms in cities with local university presence are 111% more likely to undertake process innovation, and 32% more likely to undertake product innovation. In addition, more top universities are related to higher chance of process innovation by local firms.

VI. CONCLUSION AND DISCUSSION

The main objective of this paper is to identify the urban agglomeration effect on firm innovation activities in developing Asian countries. The results obtained show that doubling city size would, on average, increase a firm's propensity for process innovation by 3.14 percentage points, for product innovation by 4.54 percentage points, and for R&D investment by 1.84 percentage points, with all else being equal at average levels. Because of large variances in Asian city sizes, the implied agglomeration gains of firm innovation propensity are substantial.

The second objective of this paper is to identify the channels for this agglomeration effect to take place. We present two such channels at city level: through firm R&D and through the local presence of a quality university. For the first channel, firms investing in R&D would reap a higher agglomeration benefit in the form of higher product innovation propensity. For the second channel, the presence of a top university has a significant correlation with innovation activities by firms in the same city.

We found heterogeneity in the agglomeration effect across different countries, with cities in India and the PRC enjoying agglomeration benefits that were not evident for the rest of the countries. However, the reason for such heterogeneity remains largely unidentified. To precisely

investigate this question, one might need to examine other variations, such as industrial composition or institutional differences between countries. Another direction worth exploring is the mechanism of the knowledge spillover by university presence. Detailed data, such as innovation characteristics and precise proximity to university, are needed to further identify and fine-tune the spillover mechanism. These are topics for future research on urban agglomeration in the developing world.

We find discrepancies between IV estimates and probit estimates that move at different directions for process and product innovations. For process innovation, the marginal effect becomes 0.56 percentage points larger after accounting for endogeneity, while for product innovation the marginal effect becomes 0.46 percentage points smaller. This implies that the endogeneity issues at play, such as simultaneity, lead to underestimation of an agglomeration effect on process innovation, but to overestimation for product innovation. One possible explanation for this divergence is the impact of innovative activities on urban population, i.e., the direction of reverse causality between urban population and innovation. Duranton (2014) proposed a theoretical argument that product innovation is expected to add to a city's employment, while process innovation within a city can cut both ways. Employment increases with process innovation only if greater productive efficiency and lower prices lead to a more than proportional increase in demand; otherwise, process innovation implies a contraction of local employment. Drawing from this, the reverse causality between population and product innovation would explain the overestimate of the coefficient in the probit model. Conversely, if process innovation drives urban decline, it explains the underestimation before accounting for endogeneity. However, whether this hypothesis holds statistically needs further testing with stronger evidence, e.g., finer data at within city level.

REFERENCES

- Abrahams, A., C. Oram and N. Lozano-Gracia. 2018. Deblurring DMSP Nighttime Lights: A New Method Using Gaussian Filters and Frequencies of Illumination. *Remote Sensing of Environment*. 210. pp. 242–258.
- 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.
- 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.
- 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., A. Demirgüç-Kunt, and V. Maksimovic. 2011. Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial and Quantitative Analysis*, 46(6), pp. 1545–1580.
- Baba, Y., N. Shichijo, and S.R. Sedita. 2009. How Do Collaborations with Universities Affect Firms' Innovative Performance? The Role of “Pasteur Scientists” in the Advanced Materials Field. *Research Policy*. 38 (5). pp. 756–764.
- Baker, J.L. and G.U. Gadgil, eds. 2017. *East Asia and Pacific Cities: Expanding Opportunities for the Urban Poor*. Washington, DC: World Bank.
- Barth, J.R., C. Lin, P. Lin, and F.M. Song. 2009. Corruption in bank lending to firms: Cross-country micro evidence on the beneficial role of competition and information sharing. *Journal of Financial Economics*, 91(3), pp. 361–388.
- 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.
- Besley, T. 2015. Law, regulation, and the business climate: The nature and influence of the World Bank Doing Business project. *Journal of Economic Perspectives*, 29(3), pp. 99–120.
- 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.
- Bhagwati, J. and A. Panagariya. 2013. *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries*. United Kingdom: Hachette.
- Buettner, T. 2015. Urban Estimates and Projections at the United Nations: The Strengths, Weaknesses, and Underpinnings of the World Urbanization Prospects. *Spatial Demography*. 3 (2). pp. 91–108.
- 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.
- Cohen, W.M. and S. Klepper. 1996. Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. *Review of Economics and Statistics*. 78 (2). pp. 232–243.
- 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.
- Combes, P.P. and L. Gobillon. 2015. The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*. Vol. 5. pp. 247–348. Elsevier.
- Crescenzi, R., A. Rodríguez-Pose, and M. Storper. 2012. The Territorial Dynamics of Innovation in China and India. *Journal of Economic Geography*. 12 (5). pp. 1055–1085.
- Djankov, S., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2003. Courts. *The Quarterly Journal of Economics*, 118(2), pp. 453–517.
- Donaldson, D. and A. Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*. 30 (4). pp. 171–98.
- Duranton, G. 2007. Urban Evolutions: The Fast, the Slow, and the Still. *American Economic Review*. 97 (1). pp. 197–221.
- Duranton, G. 2008. From Cities to Productivity and Growth in Developing Countries. *Canadian Journal of Economics/Revue Canadienne d'économique*. 41 (3). pp. 689–736.

- Duranton, G. 2014. *Growing through Cities in Developing Countries*. Washington, DC: World Bank.
- Duranton, G. and W.R. Kerr. 2015. *The Logic of Agglomeration*. No. w21452. National Bureau of Economic Research.
- 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., F. Guy, and S. Iammarino. 2012. Shaping the Formation of University–Industry Research Collaborations: What Type of Proximity Does Really Matter? *Journal of Economic Geography*. 13 (4). pp. 537–558.
- 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.
- Ellis, P. and M. Roberts. 2015. *Leveraging Urbanization in South Asia: Managing Spatial Transformation for Prosperity and Livability*. Washington, DC: World Bank.
- 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.
- Fischer, M.M. and A. Varga. 2003. Spatial Knowledge Spillovers and University Research: Evidence from Austria. *The Annals of Regional Science*. 37 (2). pp. 303–322.
- Gerlach, H., T. Rønde, and K. Stahl. 2009. Labor Pooling in R&D Intensive Industries. *Journal of Urban Economics*. 65 (1). pp.99–111.
- Hallward-Driemeier, M. and Aterido, R., 2009. *Comparing apples with.... apples: how to make (more) sense of subjective rankings of constraints to business*. Washington, DC: World Bank.
- 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. 2000. *The Effects of Urban Concentration on Economic Growth*. No. w7503. Cambridge, MA: National Bureau of Economic Research.
- 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.
- Jaffe, A.B., M. Trajtenberg, and R. Henderson. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*. 108 (3). pp. 577–598.
- Jaffe, A.B., M. Trajtenberg, and M.S. Fogarty. 2000. Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review*. 90 (2). pp. 215–218.
- Jiang, Y. 2019. Asian Cities: Spatial Dynamics and Driving Forces. *ADB Economics Working Paper* (forthcoming).
- Kinda, T. 2010. Investment Climate and FDI in Developing Countries: Firm-Level Evidence. *World Development*. 38 (4). pp. 498–513.
- Leten, B., P. Landoni, and B. Van Looy. 2014. Science or Graduates: How Do Firms Benefit from the Proximity of Universities? *Research Policy*. 43 (8). pp. 1398–1412.
- Liu, F. and Y. Sun. 2009. A Comparison of the Spatial Distribution of Innovative Activities in China and the US. *Technological Forecasting and Social Change*. 76 (6). pp. 797–805.
- Lundvall, B.Å., K.J. Joseph, C. Chaminade, and J. Vang, eds. 2011. *Handbook of Innovation Systems and Developing Countries: Building Domestic Capabilities in a Global Setting*. Edward Elgar Publishing.
- Mahmood, I.P. and J. Singh. 2003. Technological Dynamism in Asia. *Research Policy*. 32 (6). pp. 1031–1054.
- Maietta, O.W. 2015. Determinants of University–Firm R&D Collaboration and Its Impact on Innovation: A Perspective from a Low-Tech Industry. *Research Policy*. 44 (7) pp. 1341–1359.
- Mansfield, E. and J.Y Lee. 1996. The Modern University: Contributor to Industrial Innovation and Recipient of Industrial R&D Support. *Research policy*. 25 (7). pp. 1047–1058.
- Marshall, A. *Principles of Economics*. pp.1–627. London: Mac-Millan, 1895.
- 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.
- Montgomery, M.R., R. Stren, B. Cohen, and H.E. Reed. 2013. *Cities Transformed: Demographic Change and Its Implications in the Developing World*. Routledge.

- Moomaw, R.L. 1981. Productivity and City Size: A Critique of the Evidence. *The Quarterly Journal of Economics*. 96 (4). pp. 675–688.
- Moretti, E. 2019. *The Effect of High-Tech Clusters on the Productivity of Top Innovators*. No. w26270. National Bureau of Economic Research.
- Morgan, K. 2004. The Exaggerated Death of Geography: Learning, Proximity and Territorial Innovation Systems. *Journal of Economic Geography*. 4 (1). pp. 3–21.
- Mukherjee, A. 2013. The Service Sector in India. *Asian Development Bank Economics Working Paper Series*. 352.
- Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura. 2014. Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *Review of Economics and Statistics*. 96 (5). pp. 967–985.
- Nieto Galindo, V.M. 2007. La Aglomeración Como Una Causa de la Innovación en Colombia. *Planeación & Desarrollo*. 38 (1). pp. 97–139.
- Nunn, N. and D. Puga. 2012. Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics*. 94 (1). pp. 20–36.
- Ó hUallcháin, 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. London: 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. and Valentina Rollo. Has the internet fostered inclusive innovation in the developing world? *World Development* 78 (2016): pp. 587–609.
- Porter, M.E. 1993. *The Competitive Advantage of Nations*. pp. 73–93). Cambridge: Harvard Business School Management Programs.
- Puga, D. 2010. The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science*. 50 (1). pp. 203–219.
- Quintero, L.E. and M. Roberts. 2018. *Explaining Spatial Variations in Productivity: Evidence from Latin America and the Caribbean*. Washington, DC: World Bank.
- 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.

Riley, S.J., S.D. DeGloria, and R. Elliot. 1999. Index that Quantifies Topographic Heterogeneity. *Intermountain Journal of Sciences*. 5 (1–4). pp. 23–27.

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.

Rotemberg, J.J. and G. Saloner. 2000. Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade. *Regional Science and Urban Economics*. 30 (4). pp. 373–404.

Small, C., F. Pozzi, and C.D. Elvidge. 2005. Spatial Analysis of Global Urban Extent from DMSP–OLS Night Lights. *Remote Sensing of Environment*. 96 (3–4). pp. 277–291.

Thompson, P. and M. Fox–Kean. 2005. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *American Economic Review*. 95 (1). pp. 450–460.

Toivanen, O. and Väänänen, L. 2016. Education and invention. *Review of Economics and Statistics*, 98(2), pp. 382–396.

United Nations. 2018. *World Urbanization Prospects, the 2018 Revision: Methodology*. New York.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

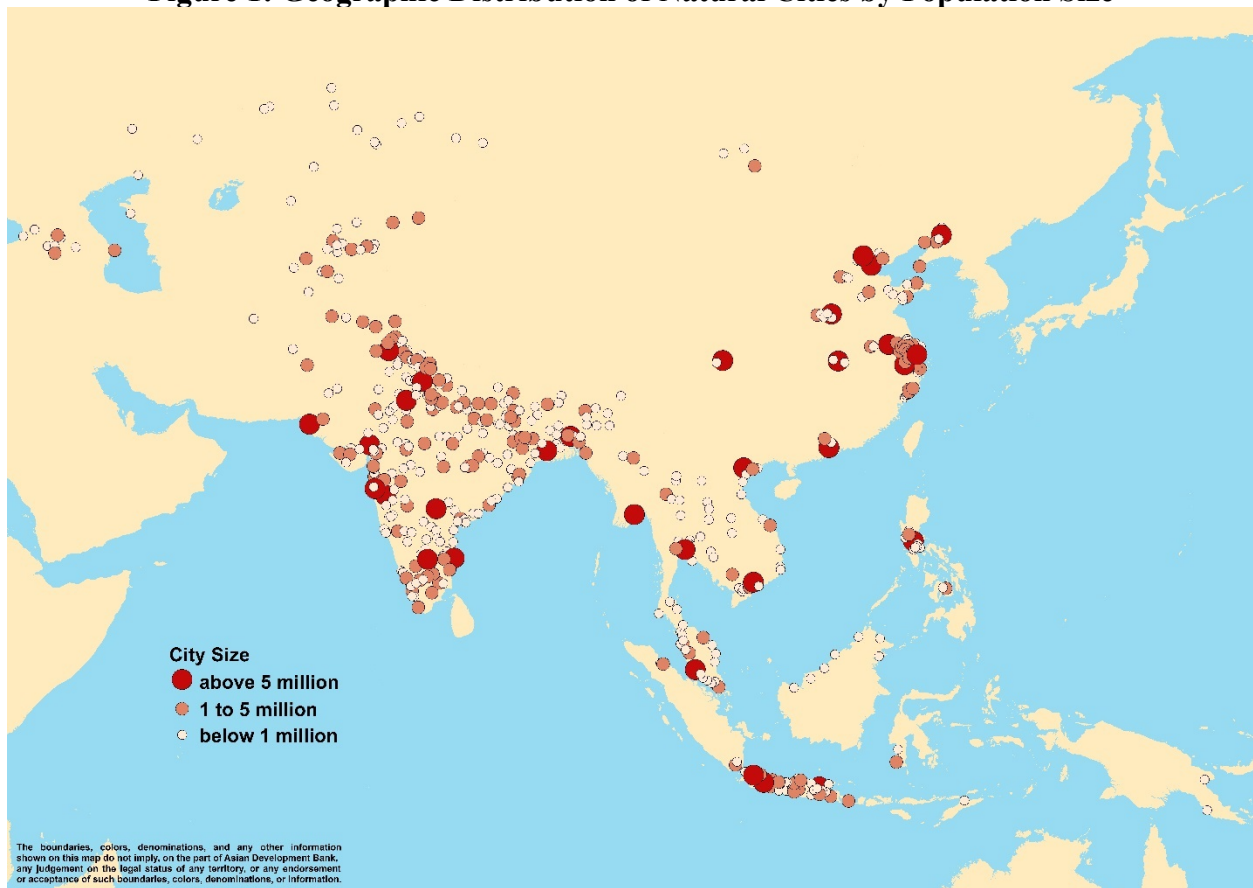
World Bank. 2009. *World Development Report 2009: Reshaping Economic Geography*. Washington, DC.

World Bank. 2014. *Urban China: Toward Efficient, Inclusive, and Sustainable Urbanization*. Development Research Center of the State Council, the People’s Republic of China. Washington, DC.

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

Zucker, L.G., M.R. Darby, and J. Armstrong. 1998. Geographically Localized Knowledge: Spillovers or Markets? *Economic Inquiry*. 36 (1). pp. 65–86.

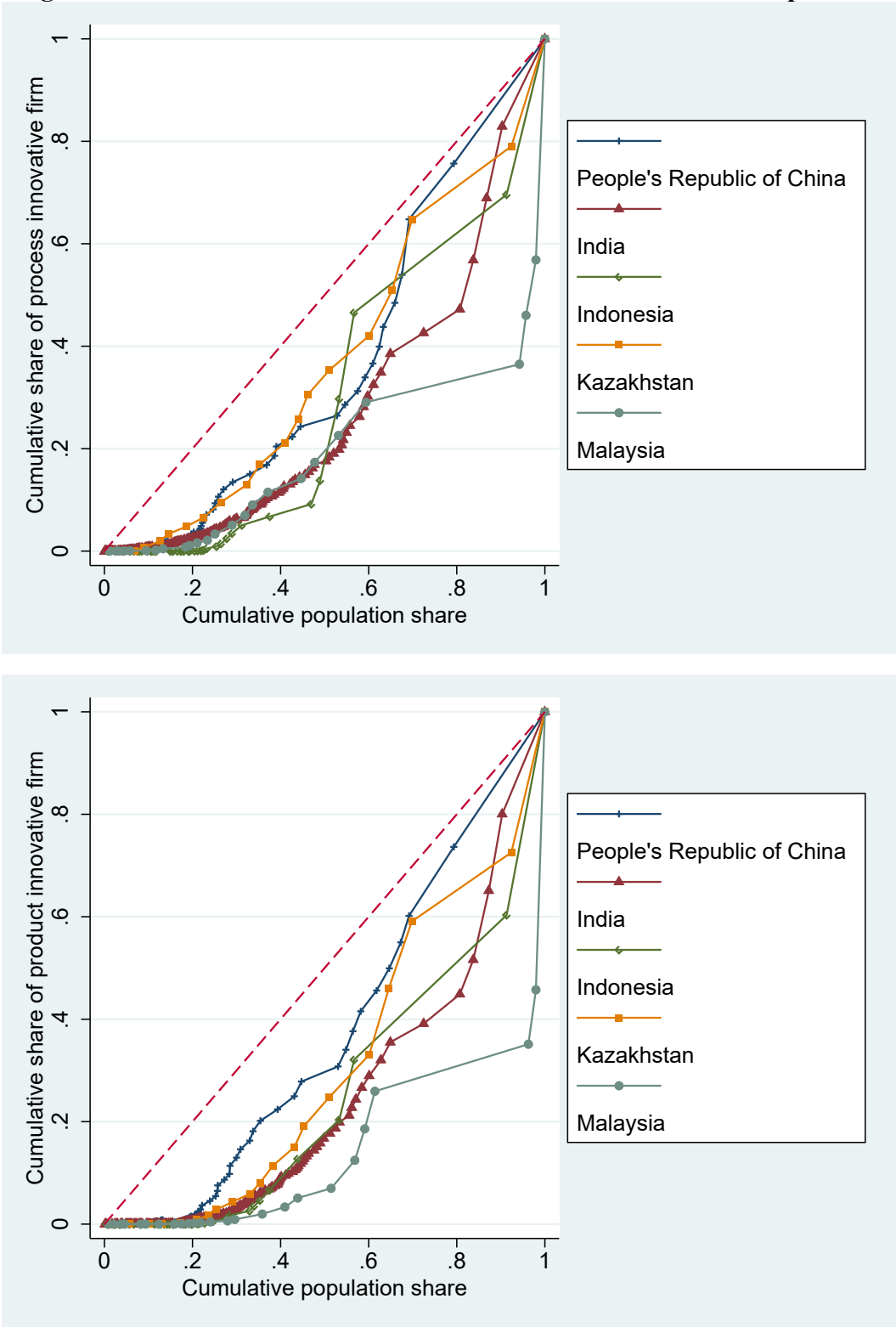
Figure 1: Geographic Distribution of Natural Cities by Population Size

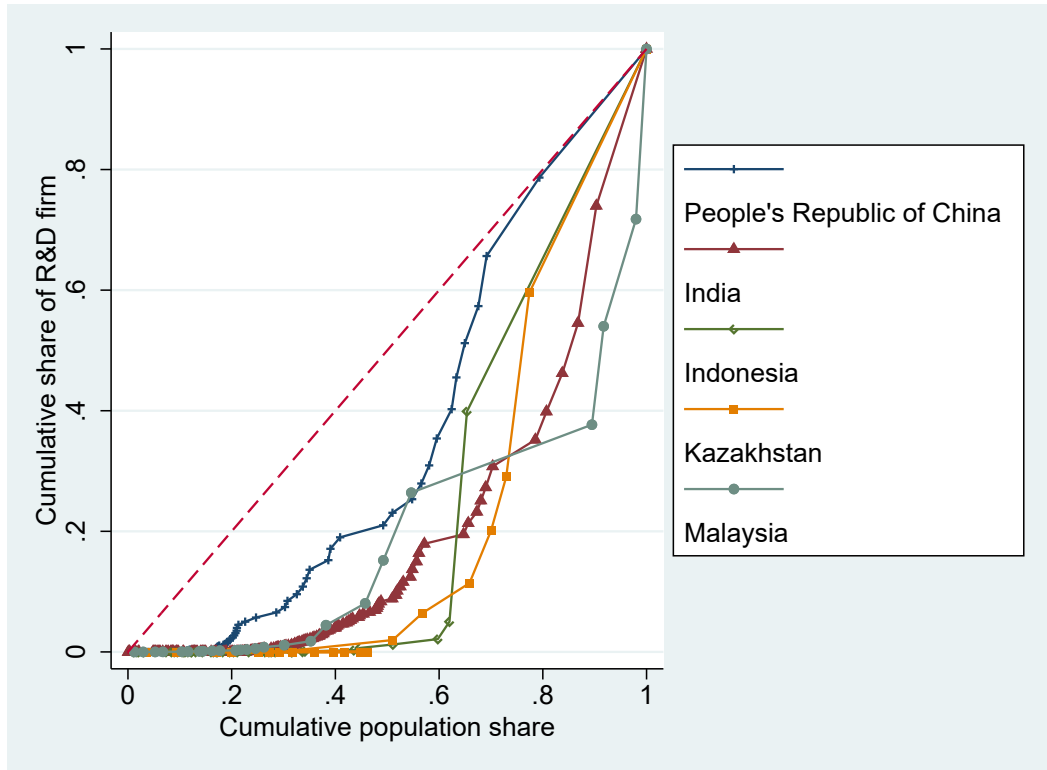


Note: Each dot represents a natural city in our sample.

Source: ADB estimates using nighttime lights images from the National Oceanic and Atmospheric Administration (accessed 1 April 2017 and 10 August 2019) and grid population data from LandScan Datasets of the Oak Ridge National Laboratory (accessed 31 August 2017 and 31 August 2018).

Figure 2: Cumulative Shares of Innovative Firms versus Urban Population





R&D = research and development.

Note: Each data point represents a natural city. Natural cities are sorted by innovative firm share in ascending order. Population share is measured by the population in each natural city over the total population in all natural cities. The same method applies for innovative firm share.

Source: ADB estimates using data from the World Bank's Enterprise Surveys (accessed 11 August 2018) and nighttime lights images from the National Oceanic and Atmospheric Administration (accessed 1 April 2017 and 10 August 2018).

Table 1: Summary Statistics of Firm and City Variables

Variables	N	Unweighted Mean	SD	Min	Median	Max
Full Sample						
Firm performance						
Process innovation	20,578	0.493	0.500	0	0	1
Product innovation	21,741	0.357	0.479	0	0	1
Main market product innovation	21,741	0.205	0.403	0	0	0
Firm R&D	20,675	0.244	0.430	0	0	1
Firm innovation	21,119	0.563	0.496	0	0	0
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
HQ dummy	21,857	0.413	0.492	0	0	0
Share of skilled workers	21,857	0.348	0.327	0	0	0
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
Ave. annual population growth (2000–2010)	489	0.0447	0.0519	-0.107	-0.00859	0.0126
With top-500 university	489	0.202	0.402	0	0	1
No. of top universities	489	0.507	1.727	0	0	0
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
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
Terrain ruggedness index (m)	489	33.20	43.31	0	20.34	503.9
IV Sample						
Firm performance						
Process innovation	18,015	0.500	0.500	0	0	1
Product innovation	19,120	0.366	0.482	0	0	1
Main product innovation	19,120	0.210	0.407	0	0	0
Firm R&D	18,102	0.253	0.435	0	0	1
Firm innovation	18,539	0.571	0.495	0	0	0
Firm characteristics						
Young firm	19,040	0.202	0.401	0	0	1
Small firm	19,166	0.632	0.482	0	1	1
FDI	19,220	0.061	0.239	0	0	1
Manufacturing	19,220	0.664	0.472	0	1	1
HQ dummy	19,220	0.434	0.496	0	0	0
Share of skilled workers	19,220	0.354	0.327	0	0	0
City characteristics						
Population (2010)	287	2.24e+06	4.16e+06	98,863	931,236	3.58e+07
Log population (2010)	287	13.85	1.142	11.50	13.74	17.39
Population (ave. 1950–1959)	287	310,319	646,080	3,531	104,409	5.57e+06
Log population (ave. 1950–1959)	287	11.60	1.475	8.169	11.56	15.53
Ave. annual population growth (2000–2010)	287	0.0407	0.0394	-0.0529	-0.00122	0.0148
With top-500 university	287	0.328	0.470	0	0	1
No. of top universities	287	0.843	2.189	0	0	0
Distance to port (km)	287	376.3	471.8	0.462	167.1	2,232
Log distance to port	287	4.888	1.748	-0.773	5.118	7.711
Average precipitation (mm)	287	3.464	2.350	0.0973	2.786	11.97
Max temperature	287	35.03	3.993	23.30	34.46	43.84
Min temperature	287	9.159	10.70	-30.23	10.71	23.41
Terrain ruggedness index (m)	287	29.88	29.41	0	20.36	187.2

ave = average, FDI = foreign direct investment, HQ = headquarters, IV = instrumental variable, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, mm = millimeter, R&D = research and development, SD = standard deviation.

Source: Authors' calculations.

Table 2: Baseline Model

Variables	Full Sample			IV Sample		
	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D	(4) Process Innovation	(5) Product Innovation	(6) Firm R&D
Log population (2010)	.0856*** (.0221)	.119*** (.0264)	.0972*** (.0258)	.0783*** (.0272)	.148*** (.0297)	.0831*** (.0301)
Marginal effect	.0336*** (.0086)	.0421*** (.0095)	.0268*** (.0069)	.0309*** (.0107)	.0534*** (.0108)	.0239*** (.0084)
Predicted probability	.427*** (.0113)	.314*** (.0114)	.195*** (.0106)	.44*** (.0115)	.326*** (.0121)	.209*** (.011)
Average annual population growth (2000–2010)	-1.77 (1.1)	-.279 (1.06)	-.916 (1.24)	-2.82* (1.67)	-.702 (1.66)	-.248 (1.84)
Log distance to port (km)	.167*** (.0243)	.109*** (.0256)	.0295 (.0257)	.189*** (.0261)	.105*** (.0267)	.0187 (.0264)
Average precipitation (mm)	-.106*** (.0292)	-.123*** (.0303)	-.0761** (.0347)	-.108*** (.0318)	-.135*** (.0327)	-.0826** (.0367)
Maximum temperature	.0422*** (.0147)	-.0043 (.0163)	.0936*** (.0165)	.0464*** (.0157)	-.0083 (.017)	.097*** (.0167)
Minimum temperature	.0311*** (.0074)	.0321*** (.0088)	.0159* (.0085)	.0347*** (.008)	.0328*** (.0092)	.016* (.0089)
Terrain ruggedness index (m)	.0028** (.0011)	.0015 (.0011)	.0025** (.0012)	.0029*** (.0011)	7.5e-04 (.0012)	.0022** (.0011)
Young firm	.199*** (.0654)	.13 (.0893)	.123* (.0719)	.214*** (.0678)	.116 (.0893)	.155** (.0749)
Small firm	-.528*** (.0597)	-.215*** (.0632)	-.411*** (.0598)	-.55*** (.0603)	-.23*** (.0643)	-.423*** (.0624)
Foreign direct investment	.119 (.133)	.329** (.139)	.224* (.124)	.118 (.139)	.316** (.144)	.238* (.128)
HQ dummy	-.1 (.0961)	.321* (.176)	.0813 (.132)	-.0918 (.0987)	.345* (.181)	.0955 (.132)
Share of skilled workers	.133 (.145)	-.0922 (.136)	.0405 (.139)	.235 (.154)	-.124 (.139)	.101 (.145)
Sector/year/country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20374	21496	20218	17838	18903	17673
Number of cities	483	489	480	285	287	283
F statistic	32.790	22.078	25.705	31.063	20.392	24.688

FE = fixed effects, HQ = headquarters, IV = instrumental variable, km = kilometer, log = logarithm, m = meter, mm = millimeter, R&D = research and development.

Notes: *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$

Robust standard errors clustered by country in parentheses. Columns (1)–(3) use the full sample, and (4)–(6) are samples restricted to cities with IV.

Source: Authors' calculations.

Table 3: Instrumental Variable Estimates of Elasticities of Firm Innovation–City Population

Variables	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D
First-stage estimates	Log Population (2010)		
Log population (1950–1959)	.809*** (.0118)	.809*** (.0116)	.812*** (.0118)
Other controls	Yes	Yes	Yes
F statistic	5713.3	5707.6	5714.8
Second-stage estimates			
Log population (2010)	.0796*** (.0295)	.126*** (.0302)	.0642** (.0306)
Marginal effect	.0314*** (.0116)	.0454*** (.011)	.0184** (.0087)
Predicted probability	.44*** (.0114)	.326*** (.0121)	.209*** (.011)
Average annual population growth (2000–2010)	-2.81* (1.68)	-1.03 (1.6)	-.46 (1.89)
Log distance to port (km)	.189*** (.0261)	.106*** (.0266)	.0183 (.0264)
Average precipitation (mm)	-.108*** (.0316)	-.13*** (.0326)	-.0789*** (.0364)
Maximum temperature	.0463*** (.0156)	-.0056 (.0172)	.0986*** (.0166)
Minimum temperature	.0346*** (.0081)	.0341*** (.0092)	.017* (.0091)
Terrain ruggedness index (m)	.0029*** (.0011)	6.7e-04 (.0012)	.0021* (.0011)
Young firm	.214*** (.0678)	.117 (.0894)	.154** (.0747)
Small firm	-.55*** (.0602)	-.231*** (.0642)	-.424*** (.0624)
Foreign direct investment	.118 (.139)	.329** (.144)	.247* (.128)
HQ dummy	-.0918 (.0986)	.345* (.182)	.0974 (.133)
Share of skilled workers	.235 (.154)	-.123 (.139)	.104 (.145)
Sector/year/country FE	Yes	Yes	Yes
Observations	17,838	18,903	17,673
Number of cities	285	287	283
F statistic	31.00	20.29	24.77

FE = fixed effects, HQ = headquarters, IV = instrumental variable, km = kilometer, log = logarithm, m = meter, mm = millimeter, R&D = research and development.

Notes: *** = p<0.01, ** = p<0.05, * = p<0.1. Robust standard errors clustered by country in parentheses.

Source: Authors' calculations.

**Table 4: Robustness Check on Instrumental Variable Estimates of Elasticities of Firm Innovation–
City Population**

Variables	Excluding Cities with >10M Population			Using Other Measures of Firm Innovation	
	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D	(4) Main Product Innovation	(5) Firm Innovation
First-stage estimates	Log Population (2010)			Log Population (2010)	
Log population (1950–1959)	.644*** (.0139)	.623*** (.0129)	.645*** (.014)	.86*** (.0115)	.803*** (.0116)
Other controls	Yes	Yes	Yes	Yes	Yes
F statistic	3,668.7	4,283.8	3,686.4	5,707.6	5,939.1
Second-stage estimates					
Log population (2010)	.247*** (.0504)	.168*** (.0469)	.208*** (.0529)	.0918** (.0359)	.128*** (.0306)
Marginal effect	.0972*** (.0199)	.0578*** (.0161)	.0559*** (.0138)	.0214*** (.0083)	.0498*** (.0119)
Predicted probability	.439*** (.0136)	.293*** (.0128)	.187*** (.0162)	.15*** (.0097)	.589*** (.0109)
Average annual population growth (2000–2010)	-5.01*** (1.79)	-1.96 (1.6)	1.77 (1.86)	-4.13** (1.74)	-2.19 (1.72)
Log distance to port (km)	.159*** (.028)	.0997*** (.0294)	.0017 (.027)	.155*** (.0374)	.153*** (.0239)
Average precipitation (mm)	.0187 (.0378)	-.0755* (.039)	.0856** (.0361)	-.177*** (.0322)	-.093*** (.031)
Maximum temperature	.0407*** (.0149)	-.0112 (.0167)	.0834*** (.0154)	-.0345** (.0161)	.0722*** (.0147)
Minimum temperature	.0146 (.009)	.0147 (.011)	.0172** (.0087)	.0526*** (.0124)	.0241*** (.0081)
Terrain ruggedness index (m)	.0037*** (.0011)	.0023** (.0012)	.0031** (.0013)	.0035 (.0022)	.0019* (.0011)
Young firm	.0693 (.0784)	-.139 (.102)	-.0314 (.0892)	.16* (.0858)	.152** (.0678)
Small firm	-.489*** (.0732)	-.298*** (.0656)	-.479*** (.0756)	-.301*** (.0671)	-.401*** (.0608)
Foreign direct investment	.346* (.186)	.154 (.155)	.256* (.154)	-.237 (.186)	.281** (.132)
HQ dummy	.063 (.116)	.309 (.198)	.255** (.113)	-.238** (.105)	.258** (.109)
Share of skilled workers	-.0503 (.188)	-.149 (.16)	.0491 (.197)	-.0912 (.155)	.424** (.186)
Sector/year/country FE	Yes	Yes	Yes	Yes	Yes
Observations	12,556	13,348	12,368	16,427	18,345
Number of cities	272	274	270	232	286
F statistic	26.79	17.15	19.42	19.48	35.19

FE = fixed effects, HQ = headquarters, IV = instrumental variable, km = kilometer, log = logarithm, m = meter, mm = millimeter, R&D = research and development.

Notes: *** = p<0.01, ** = p<0.05, * = p<0.1. Robust standard errors clustered by country in parentheses.

Source: Authors' calculations

Table 5: Instrumental Variable Estimates of Elasticities of Firm Innovation—City Population in Low-, Lower Middle-, and Upper Middle-Income Countries

Variables	Low/Lower Middle-Income			Upper Middle-Income		
	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D	(4) Process Innovation	(5) Product Innovation	(6) Firm R&D
First-stage estimates	Log Population (2010)			Log Population (2010)		
Log population (1950–1959)	.851*** (.0118)	.852*** (.0117)	.853*** (.0118)	.755*** (.0139)	.77*** (.0117)	.757*** (.0139)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
F statistic	11,671.7	11,967.4	11,854.5	14,609.9	26,653.5	14,879.9
Second-stage estimates						
Log population (2010)	-9.3e-04 (.0332)	.0735** (.0315)	.065 (.0425)	.134*** (.0406)	.158*** (.0396)	.0423 (.0428)
Marginal effect	-3.5e-04 (.0127)	.0242** (.0104)	.0148 (.0093)	.0535*** (.0162)	.0596*** (.0152)	.0146 (.0148)
Predicted probability	.381*** (.0139)	.268*** (.0126)	.145*** (.0135)	.513*** (.0155)	.366*** (.0158)	.297*** (.0152)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector/year/country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,120	14,235	14,192	3,718	4,668	3,481
Number of cities	195	195	195	90	92	88
F statistic	25.26	16.29	19.43	21.71	18.18	19.62

FE = fixed effects, IV = instrumental variable, log = logarithm, R&D = research and development.

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

Source: Authors' calculations.

Table 6: Instrumental Variable Estimates of Elasticities of Firm Innovation–City Population in India, the People’s Republic of China, GMS, ASEAN, and CAREC

Variables	People’s Republic of China			India		
	(1)	(2)	(3)	(4)	(5)	(6)
	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D
First-stage estimates	Log Population (2010)			Log Population (2010)		
Log population (1950–1959)	.752*** (.0126)	.774*** (.0106)	.753*** (.0126)	.87*** (.0085)	.87*** (.0085)	.87*** (.0085)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
F statistic	10,565.3	13,015.0	11,211.8	3,775.9	3,791.0	3,765.2
Second-stage estimates						
Log population (2010)	.162*** (.0442)	.173*** (.0417)	.0359 (.0463)	.114*** (.0317)	.116*** (.0363)	.165*** (.0368)
Marginal effect	.063*** (.0174)	.0679*** (.0166)	.0138 (.0178)	.0443*** (.0123)	.0454*** (.0142)	.0562*** (.0126)
Predicted probability	.59*** (.0165)	.431*** (.0176)	.395*** (.0166)	.586*** (.0128)	.426*** (.0127)	.287*** (.013)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,544	2,476	1,533	7,450	7,463	7,445
Number of cities	53	55	53	130	130	130
F statistic	18.09	5.75	13.03	38.08	15.59	21.67

FE = fixed effects, IV = instrumental variable, log = logarithm, R&D = research and development.

Notes: Robust standard errors clustered by sector and/or country in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Columns (1), (2), and (3) are samples restricted to cities in the People’s Republic of China and columns (4), (5), and (6) to cities in India,. Source: Authors’ calculations.

Table 7: Instrumental Variable Estimates of Elasticities of Firm Innovation—City Population in the People’s Republic of China, India, GMS, ASEAN, and CAREC

Variables	GMS			ASEAN			CAREC		
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D
First-stage estimates	Log Population (2010)			Log Population (2010)			Log Population (2010)		
Log population (1950–1959)	.987***	.988***	.994***	.874***	.875***	.877***	.666***	.668***	.662***
	(.0152)	(.015)	(.0124)	(.0122)	(.0121)	(.0121)	(.0218)	(.0217)	(.022)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F statistic	74,634.3	95,719.8	72,873.2	4,545.0	4,706.1	4,664.0	12,181.89	14,374.15	15,051.54
Second-stage estimates									
Log population (2010)	.323***	.307**	.652***	.124**	.139**	.073	-.0257	-.165	-.14
	(.116)	(.128)	(.2)	(.0618)	(.0609)	(.0778)	(.195)	(.234)	(.24)
Marginal effect	.106***	.0762**	.0903***	.0343**	.0232**	.0088	-.0065	-.0421	-.0097
	(.0381)	(.0309)	(.0265)	(.0174)	(.0107)	(.0091)	(.0494)	(.0599)	(.0167)
Predicted probability	.267***	.165***	.0728***	.197***	.0934***	.0608***	.171***	.172***	.0309***
	(.0252)	(.019)	(.0159)	(.018)	(.0123)	(.0119)	(.0174)	(.0156)	(.0052)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector/year/country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,078	2,113	2,113	4,894	4,971	4,966	2,263	2,298	2,039
Number of cities	27	27	27	59	59	59	32	32	32
F statistic	9.13	7.26	7.37	11.48	11.58	7.87	14.27	10.28	9.14

ASEAN = Association of Southeast Asian Nations, CAREC = Central Asia Regional Economic Cooperation, FE = fixed effects, GMS = Greater Mekong Subregion, log = logarithm, R&D = research and development.

Notes: Robust standard errors clustered by sector/country in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Columns (7), (8), and (9) are samples restricted to GMS countries which are Cambodia, the People’s Republic of China (PRC) (specifically Yunnan Province and Guangxi Zhuang Autonomous Region), the Lao People’s Democratic Republic, Myanmar, Thailand, and Viet Nam. Columns (10), (11), and (12) are samples restricted to ASEAN countries which are Cambodia, Indonesia, the Lao People’s Democratic Republic, Myanmar, Malaysia, the Philippines, Thailand, and Viet Nam. Columns (13), (14), and (15) are samples restricted to CAREC countries (excluding the PRC) which are Afghanistan, Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyz Republic, Mongolia, Pakistan, Tajikistan, and Uzbekistan. We exclude the PRC from CAREC analysis because the large number of firms and cities of PRC will dominate the results, and thus provide little insight to the rest of CAREC countries.

Table 8: Knowledge Spillover and Firm Innovation

	(1)	(2)	(3)	(4)
First-Stage Estimates	Process Innovation		Product Innovation	
	Log Population (2010)	Log Population (2010)	Log Population (2010)	Log Population (2010)
Log population (1950–1959)	.809*** (.0119)	.805*** (.0133)	.81*** (.0119)	.806*** (.0133)
Firm R&D * log population (1950–1959)		-.228 (.271)	.0507*** (.0195)	-.221 (.271)
Other controls	Yes	Yes	Yes	Yes
F statistic	5,713.8	5,483.9	5,748.0	5,525.8
		Firm R&D * log Population (2010)		Firm R&D * log Population (2010)
Log population (1950–1959)		.0119*** (.0036)		.0121*** (.0036)
Firm R&D * log population (1950–1959)		5.04*** (.274)		5.05*** (.273)
Other controls		Yes		Yes
F statistic		42,884.5		43,437.5
Second-stage estimates				
Log population (2010)	.0676** (.0301)	.0731** (.0327)	.138*** (.0264)	.0636** (.0275)
Firm R&D	1.32*** (.072)	1.8** (.913)	1.33*** (.066)	-2.53*** (.775)
Firm R&D * log population (2010)		-.0311 (.058)		.248*** (.0498)
Other controls	Yes	Yes	Yes	Yes
Sector/year/country FE	Yes	Yes	Yes	Yes
Observations	17,741	17,741	17,861	17,861
Number of cities	285	285	285	285
F statistic	35.73	35.23	32.51	32.16

FE = fixed effects, log = logarithm, R&D = research and development

Notes: Robust standard errors clustered by country in parentheses. *** = p<0.01, ** = p<0.05, * = p<0.1.

Source: Authors' calculations.

Table 9: Distribution of Top Universities by Country

Country	Number of Universities	Number of Cities with Top Universities	Number of Cities in Sample
Bangladesh	6	1	11
China, People's Republic of	78	17	74
India	64	43	207
Indonesia	22	9	40
Malaysia	25	10	25
Pakistan	22	5	11
Philippines	6	2	11
Thailand	18	7	20
Viet Nam	7	5	17
Total	248	99	416

Source: Authors' calculations.

Table 10: Top 10 cities with the most top universities

Country	City	No. of Top Universities
China, People's Republic of	Beijing	20
Malaysia	Kuala Lumpur	14
Thailand	Bangkok	12
China, People's Republic of	Shanghai	11
China, People's Republic of	Nanjing	10
Pakistan	Lahore	9
China, People's Republic of	Guangzhou	8
India	Delhi	7
China, People's Republic of	Wuhan	7
Bangladesh	Dhaka	6

Source: Authors' calculations.

Table 11: Firm performance near an educational institution

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Process Innovation					Product Innovation				
University dummy	1.72*** (.14)	1.6*** (.134)	1.77*** (.164)	2.11*** (.302)	2.29*** (.339)	1.99*** (.187)	1.96*** (.181)	1.71*** (.182)	1.32* (.193)	1.33* (.201)
Firm characteristics	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
City characteristics	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Sector/year/country FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Observations	20,578	20,374	20,374	20,374	17,838	21,741	21,496	21,496	21,496	18,903
Number of cities	483	483	483	483	285	489	489	489	489	287

F statistic	44.08	45.64	29.54	27.54	26.33	54.55	24.54	20.15	17.80	16.33
-------------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

FE = fixed effects.

Notes: Robust standard errors clustered by country in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Logit model used in all columns. All columns use the full sample except for columns (5) and (10) which are restricted to samples with instrumental variable. Coefficients on top university are expressed as odds ratio. Firm characteristics include age, size, foreign direct investment dummy, headquarters dummy, and share of skilled workers. City characteristics include log of population 2010, log of distance to port, average precipitation, maximum temperature, minimum temperature, and terrain ruggedness.

Source: Authors' calculations.

Table 12: Firm performance and number of top universities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables	Process Innovation					Product Innovation				
No. of top universities	1.08*** (.0126)	1.07*** (.0125)	1.06*** (.0142)	1.05** (.0213)	1.07*** (.0231)	1.06*** (.0143)	1.07*** (.0143)	1.04*** (.0149)	1.01 (.0201)	1 (.0209)
Firm characteristics	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
City characteristics	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Sector/year/country FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Observations	20,578	20,374	20,374	20,374	17,838	21,741	21,496	21,496	21,496	18,903
No. of cities	483	483	483	483	285	489	489	489	489	287
F statistic	41.59	44.24	29.43	27.00	25.67	20.75	19.50	19.93	17.62	16.15

FE = fixed effects.

Notes: Robust standard errors clustered by country in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. Logit model used in all columns. All columns use the full sample except for columns (5) and (10) which are restricted to samples with instrumental variable. Coefficients on number of top universities are expressed as odds ratio. Firm characteristics include age, size, foreign direct investment dummy, headquarters dummy, and share of skilled workers. City characteristics include log of population 2010, log of distance to port, average precipitation, maximum temperature, minimum temperature, and terrain ruggedness.

Source: Authors' calculations.

Appendix 1: Development of Natural City Dataset of Asia and the Pacific

Development of the natural city dataset involved three main steps. First, we delineated human settlements with nighttime light (NTL) satellite imagery. Satellites from the United States (US) Air Force Defense Meteorological Satellite Program (DMSP) with Operational Linescan System (OLS) sensors recorded the intensity of earth-based lights and stored them in a digital archive from 1992 to 2013. Since 2013, the DMSP-OLS data were succeeded by those recorded by the Visible and Infrared Imager/Radiometer Suite (VIIRS) flown on spacecraft launched under the Joint Polar Satellite System of the US. Scientists at the National Oceanic and Atmospheric Administration processed the raw data and distributed a yearly version of NTL data to the public, which we used for this study. The data contains luminosity measures for every latitude-longitude grid equivalent to about 0.86 square kilometers (km^2) at the equator up to 2013, with the resolution of the pixel improved to about 0.22 km^2 after 2013. For a comprehensive introduction of the NTL data (DMSP-OLS part) and its applications in economics (Donaldson and Storeygard 2016).

The left panel of Figure A1 shows the area encompassing Metro Manila, Philippines in NTL. Due to the problem known as “blooming” or “overglowing”, the boundaries of the illuminated areas appear quite blurry. It is caused by the relatively coarse spatial resolution of the OLS sensors, the large overlap in the footprints of adjacent OLS pixels, and the accumulation of geolocation errors in the compositing process (Small et al. 2005). The post-2013 VIIRS images have much less blooming due to their higher resolution. We applied the latest methodology developed in Abrahams et al. (2018) to deblur the imagery.

With the deblurred NTL data, we delineated polygons consisting of pixels with positive luminosity values (i.e., threshold equal to 0) as human settlements,¹ then aggregated those with 1 pixel gaps between them into one polygon to allow for measurement errors as well as unlit areas (such as roads) within an integrated human settlement. The exercise yielded between 88,000 and 187,000 geocoded polygons for Asia in various years. The middle panel of Figure A1 shows the identified human settlements in the Metro Manila area.

The majority of the polygons obtained are very small and discrete, likely representing rural settlements. The second step is to identify urban areas from all human settlements, for which we referred to the database of the Global Rural Urban Mapping Project (GRUMP)². This database contains geocoded centers, names, populations, and the upper administrative divisions they belong to, for over 70,000 human settlements across the world. We focused on over 1,900

¹ A pixel with positive luminosity value is illuminated. Different values are adopted as thresholds to draw urban boundaries in the existing literature. Examples include 5 in Zhang and Seto (2011); 13 in Ellis and Roberts (2015) and in Zhou, Hubacek, and Roberts (2015); 33 in Tewari et al. (2017); and 35 in Harari (2016). A positive threshold is needed if one uses the pre-deblurring data to delineate urban extent. However, a uniform threshold across years may not yield consistent definition of urban scope given that different sensors (OLS versus VIIRS) were used over the period, and the same sensor performed differently over its life cycle. Moreover, a proper threshold value to define urban extent, if it exists, should probably vary across regions. In view of these issues, we considered it less arbitrary to define human settlements as all illuminated pixels.

² GRUMP data were generated by the Columbia University Center for International Earth Science Information Network in collaboration with the International Food Policy Research Institute, the World Bank, and Centro Internacional de Agricultura Tropical, through combining census and geospatial datasets. It can be accessed at <http://sedac.ciesin.columbia.edu/data/collection/grump-v1>. The units in the GRUMP database are not from the same administrative level within, as well as across, countries.

GRUMP units in Asia and the Pacific, all of which had a population greater than 100,000 in 2000. We identified more than 1,400 independent NTL polygons in 1992 that either cover these units or turn out to be the most relevant upon visual checking. These were treated as our natural cities and were named after the corresponding GRUMP units or the unit with the largest population if a natural city contained multiple units.³

To maximize country coverage and include large cities that are missing in the GRUMP database, we added to the data 115 polygons that either were related to major cities from small countries (mostly in the Pacific) or had an area greater than 100 km² in 2000, despite the associated GRUMP units having populations below 100,000 in 2000. With this approach, we reached a final set of 1,527 natural cities in Asia and the Pacific.⁴ As shown in the right panel of Figure A1, there were nine natural cities in the area: Angeles, Batangas, Lipa, Lucena, Metro Manila, Olongapo, San Pablo, San Pedro, and Tarlac. The largest one, Metro Manila, contained 29 GRUMP units, including Makati, Manila, and Quezon City.

Some natural cities had expanded and became connected with each other over time. To retain the individual natural cities as our primary units for analysis, we separated the connected ones where the luminosity was the lowest to obtain the footprint of each. Thus, we obtained a balanced panel of natural cities from 1992 to 2016.

The third step to develop the dataset was to assign city characteristics to the natural cities. These characteristics included population, presence of a university, weather, distance to seaports, and historical population. We filled the delineated areas of the natural cities with grid population data from LandScan. LandScan provides global population counts at approximately 1-kilometer spatial resolution, which are generated through spatial modeling and image analysis, with inputs from census data, high-resolution imagery, land cover, and other spatial data such as various boundaries, coastlines, elevations, and slopes.⁵ We overlaid the natural city polygons with the grid population data. The population of a natural city is the sum of all cells falling within or interacting with the city contour.

To measure the availability of quality tertiary education in the city, we utilized the recent QS World University Rankings, which identify the top 500 universities in Asia, based on six metrics: academic reputation, employer reputation, faculty-and-student ratio, citations per faculty, international faculty ratio, and international student ratio. We mapped these universities into natural cities using their geolocations obtained from Google Maps. Of the top 500 universities in

³ The number of the natural cities was lower than the number of units from GRUMP because some units in GRUMP were located close to each other, and thus were covered under the same polygons. Such cases arose in the relatively advanced areas of developing countries, such as the Pearl River Delta area centered around Guangzhou city in the People's Republic of China (PRC), as well as in developed economies such as the metropolitan areas surrounding Tokyo in Japan or Taipei in Taipei, China.

⁴ Please note that the population threshold adopted refers to the population of the GRUMP units. This helped us to capture most sizable urban agglomerations in the region. However, it by no means implies that our natural cities had population over 100,000 prior to 2000, or even after 2000, because natural cities could cover very different geographic scopes from those of GRUMP units, and their populations were estimated based on grid population data.

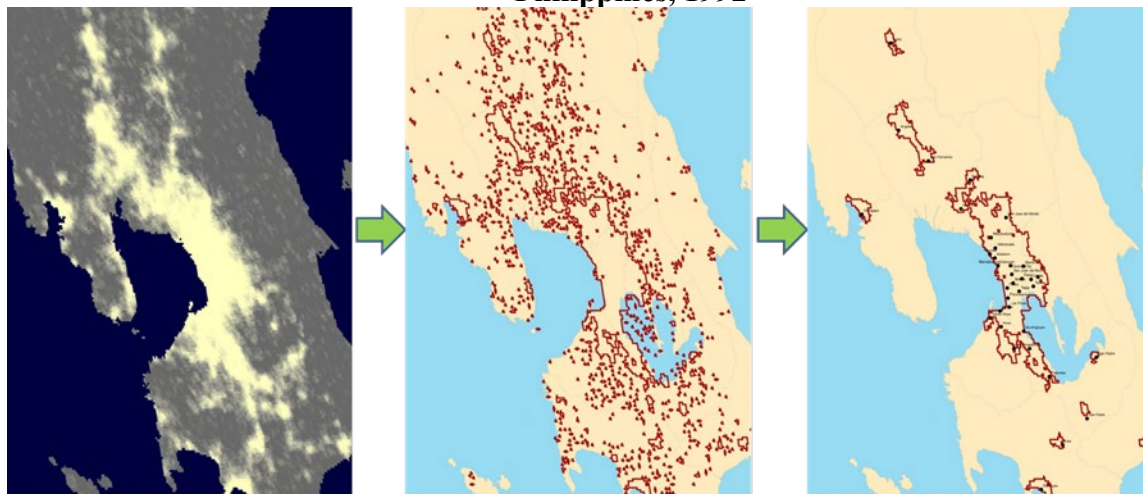
⁵ Essentially, the census population counts were disaggregated to each cell with a multivariate dasymetric modeling approach. Data precision was improved through manual verification and modification as well as refinements to the input datasets. LandScan data have been widely used in fields such as demographics, urban planning, and remote sensing.

Asia, 248 were found in 99 cities of nine developing Asian countries. A binary indicator was created equal to one if a natural city hosted at least one top-500 university.

For weather indicators, we referenced the United Kingdom's Climatic Research Unit, which publishes global monthly gridded weather data with a 0.5-degree variance, from 1901 onward. We used this monthly data to obtain the annual average daily precipitation and annual maximum and minimum temperatures for each grid. The averages across the grids surrounding the centroid of the natural city were taken as the weather measures for the natural city.

Two geographic factors were considered. One was the distance to the nearest seaport. A list of seaports was obtained from the World Port Index developed by the National Geospatial-Intelligence Agency in the US. The distance was measured as the Euclidian distance from the centroid of the natural city to the nearest domestic seaport. For cities of landlocked countries, it was the distance to the nearest seaport abroad. The other geographic factor is citywide ruggedness. Nunn and Puga (2012) calculated the Terrain Ruggedness Index—originally devised by Riley, DeGloria, and Elliot (1999)—to quantify topographic heterogeneity for every 30 by 30 arcsecond cell across the world. We obtained the grid Terrain Ruggedness Index at <https://diegopuga.org/data/rugged/> and averaged all grid cells in each natural city to obtain citywide ruggedness.

Figure A1: Natural Cities Extracted from the Nighttime Lights of Metro Manila, Philippines, 1992



Source: ADB estimates using nighttime lights images from the National Oceanic and Atmospheric Administration (accessed 1 April 2017 and 10 August 2018).

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, People's Republic of	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
Kyrgyz Republic	2013	270	3	189	81	70.00
Lao People's Democratic Republic	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

Note: Firm data came from the World Bank Enterprise Surveys conducted from 2012 to 2016. Natural city data were developed by economists at the Asian Development Bank. The two datasets were matched using geographic information system software. Source: Authors' calculations.