

Chen, Liming; Hasan, Rana; Jiang, Yi

Working Paper

Urban agglomeration and firm innovation: Evidence from Asia

ADB Economics Working Paper Series, No. 616

Provided in Cooperation with:

Asian Development Bank (ADB), Manila

Suggested Citation: Chen, Liming; Hasan, Rana; Jiang, Yi (2020) : Urban agglomeration and firm innovation: Evidence from Asia, ADB Economics Working Paper Series, No. 616, Asian Development Bank (ADB), Manila,
<https://doi.org/10.22617/WPS200195-2>

This Version is available at:

<https://hdl.handle.net/10419/230369>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/igo/>

URBAN AGGLOMERATION AND FIRM INNOVATION

EVIDENCE FROM ASIA

Liming Chen, Rana Hasan, and Yi Jiang

NO. 616

July 2020

ADB ECONOMICS
WORKING PAPER SERIES

ADB Economics Working Paper Series

Urban Agglomeration and Firm Innovation: Evidence from Asia

Liming Chen, Rana Hasan, and Yi Jiang

No. 616 | July 2020

Liming Chen (limingchen@adb.org) is a young professional, Rana Hasan (rhasan@adb.org) is a director, and Yi Jiang (yijiang@adb.org) is a principal economist at the Economic Research and Regional Cooperation Department, Asian Development Bank (ADB).

This paper was prepared as background material for the ADOU 2019 theme chapter on *Fostering Growth and Inclusion in Asia's Cities*. The authors thank Gilles Duranton for detailed comments and suggestions on an earlier draft, and participants at the 2019 SMU Conference on Urban and Regional Economics, ADB Economics Seminar Series, and ADB Economists' Forum for constructive feedback. We are grateful to the Enterprise Analysis Unit of the World Bank for providing geomasked GPS coordinates for the firms in the World Bank Enterprise Survey data used in our analysis. Patricia Thea Basilio, Ma. Adelle Gia Toledo Arbo, Marjorie Villanueva Remolador, and Renz Adrian T. Calub provided excellent research assistance.



Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2020 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

Some rights reserved. Published in 2020.

ISSN 2313-6537 (print), 2313-6545 (electronic)
Publication Stock No. WPS200195-2
DOI: <http://dx.doi.org/10.22617/WPS200195-2>

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication 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.

By making any designation of or reference to a particular territory or geographic area, or by using the term “country” in this document, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Corrigenda to ADB publications may be found at <http://www.adb.org/publications/corrigenda>.

Notes:

In this publication, “\$” refers to United States dollars.
ADB recognizes “China” as the People’s Republic of China.

The ADB Economics Working Paper Series presents data, information, and/or findings from ongoing research and studies to encourage exchange of ideas and to elicit comment and feedback about development issues in Asia and the Pacific. Since papers in this series are intended for quick and easy dissemination, the content may or may not be fully edited and may later be modified for final publication.

CONTENTS

TABLES AND FIGURES	iv
ABSTRACT	v
I. INTRODUCTION	1
II. DATA	3
A. Data on Firm Innovation	3
B. Data on Natural Cities Based on Nighttime Light Satellite Imagery	5
C. Summary Statistics	6
D. Spatial Concentration of Innovation	8
III. EMPIRICAL STRATEGY	10
A. Baseline Probit Model	10
B. Instrumental Variable Estimation	11
IV. MAIN RESULTS	13
A. Baseline and Instrumental Variable Estimates	13
B. Robustness Testing	15
C. Heterogeneity by Development Levels and Countries	17
V. CHANNELS UNDERLYING AGGLOMERATION EFFECTS	22
VI. CONCLUSION	25
APPENDIX	27
REFERENCES	31

TABLES AND FIGURES

TABLES

1	Summary Statistics of Firm and City Variables	7
2	Probit Estimates of Effects of City Population on Firm Innovation	14
3	Instrumental Variable Estimates of Effects of City Population on Firm Innovation	16
4	Robustness Check on Instrumental Variable Estimates of Effects of City Population on Firm Innovation	18
5	Instrumental Variable Estimates of Effects of City Population on Firm Innovation in Low and Lower-Middle-Income Countries and Upper-Middle-Income Countries	20
6	Instrumental Variable Estimates of Effects of City Population on Firm Innovation in the People's Republic of China, India, and ASEAN Countries	21
7	Distribution of Top Universities	22
8	Agglomeration Channels: Presence of Top Universities	23
9	Agglomeration Channels: Differential Effects of Firm Research and Development	24
A1	Matching between Geocoded Firms in the World Bank Enterprise Survey and Natural Cities	30

FIGURES

1	Geographic Distribution of Natural Cities by Population Size	6
2	Cumulative Shares of Innovative Firms versus Urban Population	9
A1	Natural Cities Extracted from the Nighttime Lights of Metro Manila, Philippines, 1992	29

ABSTRACT

This paper presents evidence on the spatial distribution and effects of urban agglomeration on firm innovation. It uses a unique dataset that consistently defines city boundaries and identifies firms' innovation-related activities across 25 developing countries in Asia. We find firm innovation to be highly concentrated at the city level. We also find substantial gains from increases in city population in terms of firms' propensity to introduce process and product innovations and undertake research and development (R&D) activities. These gains remain even after addressing concerns regarding endogeneity through the use of historical population data as instruments. In addition, we present evidence that knowledge spillovers are an important channel through which agglomeration effects occur, specifically through the presence of top-tier universities in a given city and by raising the effectiveness of firms' R&D efforts. These findings confirm the existence and significance of urban economies of scale in augmenting the knowledge flows that generate innovation.

Keywords: agglomeration economies, innovation, knowledge spillovers

JEL codes: O30, R11, O10

I. INTRODUCTION

Innovation undertaken by firms is regarded as a significant contributor to long-term economic growth in both theory and in practice (see, for example, Romer 1986). One issue that has attracted considerable attention from economists is the location of firm-level innovation within countries. A large body of evidence from the developed world shows that innovations tend to be more concentrated spatially than production. Further, there is a strong concentration of research and development (R&D) activities in some large metropolitan areas. These patterns have raised various questions, such as whether urban agglomerations play a special role in fostering innovation by firms, and if so, what channels are at work. While evidence on these questions is growing, much of it has been obtained from the developed world. For developing countries, there is a lack of systematic evidence on the spatial distribution of firms' innovation-related activities and whether and how urban agglomeration affects innovation.¹

This paper attempts to fill this gap by using geocoded firm-level data from World Bank Enterprise Surveys (WBES) conducted in 25 Asian economies from 2012 to 2016. In addition to production-related data, the WBES collects information on firms' innovation-related activities—specifically, whether sample firms introduce process and product innovations and invest in R&D. By mapping WBES firms onto a unique city-level dataset developed using nighttime light (NTL) imagery across the 25 Asian economies, we are able to study the relationship between innovation and agglomeration. We call our cities “natural cities” to distinguish them from administratively defined cities. In comparison with the latter, natural cities are consistently defined and measured across countries. They are also more likely to capture the actual economic “footprint” of a city than its administrative boundaries would suggest.

We find that innovative activity by firms is spatially concentrated within a country. The share of firms that innovate in just a few cities is substantially greater than the share of urban population these cities account for. Further, to identify the causal effect of city size on firm innovation—for example, innovation and city size could be simultaneously determined by unobserved city characteristics—we take an instrumental variable approach, using official urban population half a century ago matched to natural cities as the source of exogenous variation for current city population. All else being equal, our instrumental variable (IV)-probit estimates suggest that doubling city size would increase the average firm's propensity for undertaking process innovation by 3.9 percentage points, product innovation by 4.8 percentage points, and R&D by 2 percentage points. These translate into increases of 8.8%, 14.8%, and 9.4% over baseline propensities, respectively. Since in our data, the biggest natural city is much larger than the smallest one, the effects of agglomeration on the propensity of firms to innovate can be quite large.

We also explore heterogeneity in agglomeration effects by development levels of countries and across countries and subregions with distinct historical and institutional backgrounds. The results show that in general, agglomeration effects are present in both upper-middle income and low and lower-middle-income country groups, as well as in the People's Republic of China (PRC), India, and selected countries of the Association of Southeast Asian Nations (ASEAN), each considered separately.

¹ Most research from the developing world is about agglomeration effects concerning labor and firm productivity. Duranton (2008), Overman and Venables (2005), Henderson (2005), and Duranton (2014) provided detailed review of studies from the developing world.

The paper further investigates possible mechanisms for the estimated agglomeration effects. Mapping the top 500 Asian universities to our natural cities, we find a statistically significant and positive effect of university presence on firm innovation. We also examine the relationship between firms' R&D efforts and their introduction of process or product innovations and find it to vary by city size. Firms investing in R&D are more likely to introduce a product innovation in larger cities as compared to smaller ones. Both results suggest that knowledge spillovers are an important channel through which agglomeration effects take place.

This paper is closely related to two strands of literature. The first is on the measurement of urban agglomeration economies. Most of the empirical evidence in this strand has focused on whether city size and/or density affect firm and worker productivity.² Studies examining urban agglomeration economies associated with innovation-related activities are limited. Nevertheless, the available evidence suggests that innovation in a given country is heavily concentrated in a few cities, and that a city's size or density plays a role in boosting innovation activity within that city. Ó hUallacháin (1999) find that large metropolitan areas in the United States (US) dominate patenting, with patenting increasing with the size of the metropolitan area. Feldman and Audretsch (1999) also find that innovation appears to be a phenomenon of large cities: less than 4% of innovations occurred outside of the metropolitan areas in which 70% of the country's population reside. Carlino, Chatterjee, and Hunt (2007) estimate the elasticity of patenting per capita with respect to employment density to be approximately 20% in metropolitan areas in the US, which is several times higher than the estimated agglomeration effects on productivity in the literature. Carlino et al. (2009) further confirm the earlier results by weighting patents with their number of citations.

The second strand of literature our paper relates to attempts to identify mechanisms underlying agglomeration effects. Duranton and Puga (2004) propose three channels through which urban agglomerations influence firm and labor productivity: sharing, matching, and knowledge spillover. Carlino and Kerr (2015) review evidence and note that these three mechanisms also apply to innovation, with more empirical support for the knowledge spillover channel.³ Knowledge spillover refers to the intellectual gains made by the exchange of information for which no direct compensation is given to the producer of the knowledge. It is particularly important in explaining the concentration of innovation since innovation arguably depends on dispersal of knowledge more than other economic activities. One approach closely related to this paper examines the effect of university presence on local innovations and suggests significant effect through various channels (Jaffe 1989; Audretsch and Feldman 1996; Audretsch and Stephan 1996; Anselin, Varga, and Acs 1997; Carlino, Chatterjee, and Hunt 2007; Andrews 2017).

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

² See Rosenthal and Strange (2004), Melo, Graham, and Noland (2009), Puga (2010), and Baum-Snow and Ferreira (2015) for extensive surveys of methodology and evidence.

³ A thick labor market allows for the efficient sharing of access to a pool of specialized and experienced workers, which creates linkages and networks for knowledge to flow rapidly (Helsley and Strange 2002). Gerlach, Rønde, and Stahl (2009) find that firms located in clusters also take greater risks in R&D choices because of this, compared with spatially isolated firms. A thick local labor market also improves the quality of matches between firms and workers (Helsley and Strange 1990; Berliant, Reed, and Wang 2006; Strange, Hejazi, and Tang 2006).

To the best of our knowledge, there is no large-scale, cross-country evidence from the developing world on this topic. Yet, it is important to study the effects of urban agglomeration with respect to innovation in developing countries. While firms in developing countries may not be at the global technological frontier, they face the challenge of catching up with the frontier (Acemoglu, Aghion, and Zilibotti 2006). In this context, understanding the conditions under which firms undertake product and process innovations to produce new and better products and to produce them more efficiently, respectively, as well as conduct R&D is very relevant.

The remainder of this paper is organized as follows. The next section introduces the data, while Section III describes our empirical strategy. Section IV presents the main results regarding agglomeration effects and firms' innovation-related activities, followed by robustness tests and various extensions. Section V investigates potential channels through which agglomeration effects may be working. The final section concludes.

II. DATA

We assemble a unique cross-sectional dataset that assigns firms covered by the WBES from 2012 to 2016 to nearly 500 cities across 25 Asian economies. The WBES provide information on production-related data on firms, as well as their engagement in innovation-related activities. The data on cities utilizes NTL satellite imagery to define “natural cities.” 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 mainly formal firms from an economy's private sector, excluding the agriculture sector and all state-owned enterprises.⁴ Surveyed firms are selected through stratified random sampling, based on the sector of activity and firm size. Since 2002, the data have been collected from face-to-face interviews with high-level managers or company owners.⁵

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

- (i) During the last 3 years, has [the surveyed] establishment introduced any new or improved process? These include methods of manufacturing products or offering services; logistics,

⁴ While the sample firms comprising the WBES are unlikely to be representative of the industrial structure of a city, this should not be much of a disadvantage for our purpose. Moreover, there seems to be no alternative to the WBES, especially when studying a cross-section of developing countries.

⁵ The World Bank has taken several steps to ensure quality of the survey data as detailed by Ayyagari, Demirgüç-Kunt, and Maksimovic (2011). For instance, the firm identifiers are kept confidential and the survey is conducted by private organizations independent from government. To make sure questions elicit valid answers, various measures including translation and localization checks, different interviewees for the same firm, and post-survey consistency checks are taken before, during, and after the survey.

delivery, or distribution methods for inputs, products, or services; or supporting activities for processes.

- (ii) During the last 3 years, has this establishment introduced a new or improved product or service?
- (iii) During the last fiscal year, did this establishment spend on R&D 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 R&D, respectively. As defined in the questionnaire, these innovations need not necessarily be cutting-edge inventions or revolutionary production processes. They could be small improvements to existing processes and products.⁶

The subjectivity of responses may raise concerns that the self-reported innovation variables may not measure innovation as objectively as peer-reviewed sources, for example, patents. The WBES questionnaire therefore also asks for detailed descriptions of process and product innovations. When provided, these descriptions reveal that firms have valid reasons to consider their reported improvements in products or practices as "innovative," pointing to increased profit, reduced cost, or reduced production time as a direct result. It is not surprising, therefore, to find that the innovation measures from the WBES have met the test of being used and reported in peer-reviewed academic journals (see, in particular, Ayyagari, Demirgüç-Kunt, and Maksimovic 2011; Paunov 2016; and Paunov and Rollo 2016).

Nevertheless, there may exist systematic differences across countries regarding what constitutes innovation depending, for example, on a country's development level or technological capabilities. To address this issue, we include country and year fixed effects in all our regressions.

Furthermore, we use two additional innovation measures derived from the WBES as robustness tests. The first measure is labeled "main market product innovation" and is based on a follow-up to question (ii) above and asks whether any of the new or improved products or services were also new for the establishment's main market. The second measure is labeled "firm innovation" and is constructed from question (i) and (ii). The variable takes a value of 1 if a firm has reported undertaking either a product or process innovation and 0 otherwise.

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

⁶ In our sample, examples of process innovations include a manufacturing firm in Yangon, Myanmar starting to offer a door-to-door sales service or a repair shop now running an ad campaign on Facebook. Examples of product innovation include a restaurant in Cebu, Philippines introducing Panda pao, or a plastic bags printing firm in Port Moresby, Papua New Guinea now also printing bags to carry food.

Fortunately, the WBES conducted since 2012 contain the geographic coordinates of surveyed firms.⁷ However, 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 (described below), so as to associate firms with the cities in which they are located.

B. Data on Natural Cities Based on Nighttime Light Satellite Imagery

We use a novel dataset of more than 1,500 “natural” cities across Asia and the Pacific. The dataset is constructed using NTL satellite imagery data available since 1992 obtained from the National Oceanic and Atmospheric Administration website. In a nutshell, NTL data is used to track the footprint of urban agglomerations from 1992 to 2016 as captured by contiguously illuminated areas, which are thus referred to as natural cities to distinguish them from administratively defined cities.

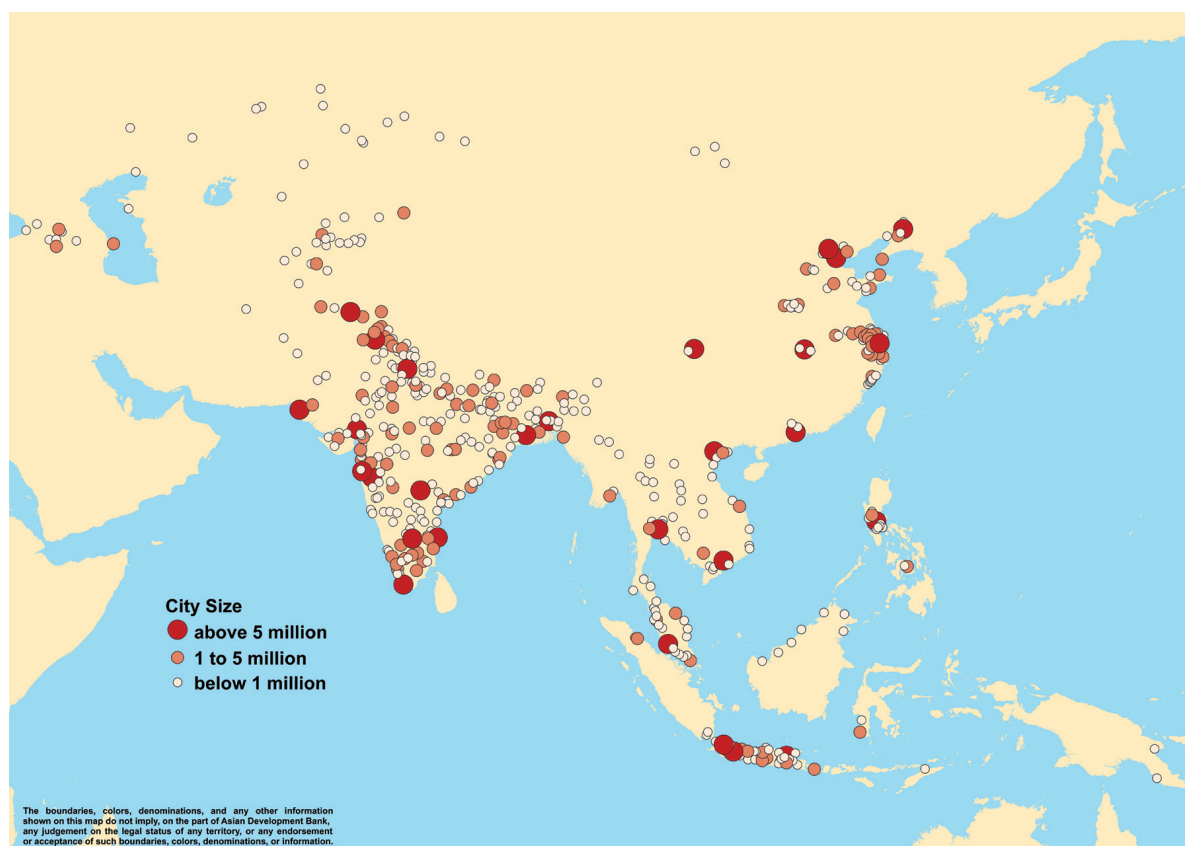
Individual city characteristics, including population, presence of university, weather, geography, and historical population are ascertained using various data sources. For example, to estimate population, we fill the delineated urban areas corresponding to our natural cities with grid population data from LandScan, which is available at approximately 1-km spatial resolution. We then tally all cells falling within, or intersecting with, the contour of each natural city. Details on the sources of data and how information from them is used to develop the database on natural cities is provided in Appendix 1 and Jiang (2020).

The advantages of using natural cities, at least for this study, are threefold. First, 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 firms operate, compared to 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.

We map the WBES firms into natural cities using geographic information system software. To maximize the number of firms that can be matched to our natural cities, 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. Table A1 shows the results of the matching exercise by country. The share of WBES firms matched to a natural city is 87%, with matching within individual countries ranging from 40% (Mongolia) to 100% (Myanmar and the PRC).⁸ Overall, we are able to match a total of 21,857 firms to 489 natural cities, an average of 45 firms per city. Figure 1 shows the geographic distributions of our natural cities with WBES firms. India alone contributes 8,100 firms, or 37% of the total sample, which were distributed across 207 natural cities.

⁷ The authors are grateful to the World Bank for sharing geocoded enterprise survey data.

⁸ 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.

Figure 1: Geographic Distribution of Natural Cities by Population Size

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

Source: Authors' estimates.

C. Summary Statistics

Table 1 (upper panel) reports the summary statistics of firm-level and city-level variables used in the study. 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 literature, we define firms that had operated for fewer than 10 years as young firms (Reyes, Roberts, and Xu 2017); those that had 50 or fewer permanent employees as small firms (Beck, Demirgüç-Kunt, and Maksimovic 2005; Reyes, Roberts, and Xu 2017); and those that had 10% or more coming from FDI as FDI firms.⁹ 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 service sector. About 41% of the surveyed firms are headquarters and the average share of skilled workers, defined as workers that perform highly skilled or semi-skilled jobs, is around 35%.

⁹ The literature often follows the International Monetary Fund's standard definition of FDI, that is, 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).

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	1
Firm R&D	20,675	0.244	0.430	0	0	1
Firm innovation	21,119	0.563	0.496	0	1	1
Firm characteristics						
Young firm	21,655	0.209	0.407	0	0	1
Small firm	21,795	0.645	0.478	0	1	1
FDI	21,857	0.061	0.239	0	0	1
Manufacturing	21,857	0.657	0.475	0	1	1
Headquarter dummy	21,857	0.413	0.492	0	0	1
Share of skilled workers	21,857	0.348	0.327	0	0.313	1
City characteristics						
Population (2010)	489	1.43e+06	3.33e+06	8,353	437,813	3.58e+07
Log population (2010)	489	13.19	1.278	9.030	12.99	17.39
With top 500 university	489	0.202	0.402	0	0	1
No. of top universities	489	0.507	1.727	0	0	20
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
Instrumental Variable 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 market product innovation	19,120	0.210	0.407	0	0	1
Firm R&D	18,102	0.253	0.435	0	0	1
Firm innovation	18,539	0.571	0.495	0	1	1
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
Headquarter dummy	19,220	0.434	0.496	0	0	1
Share of skilled workers	19,220	0.354	0.327	0	0.333	1
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
With top 500 university	287	0.328	0.470	0	0	1
No. of top universities	287	0.843	2.189	0	0	20
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, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, mm = millimeter, N = number of observations, R&D = research and development, SD = standard deviation.

Source: Authors' estimates.

We use 2010 values for time-variant city characteristics, such as population and weather, to account for the fact that the innovation activities captured by the surveys were concentrated during 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 ranks 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 a standard deviation of 500 km. Minimum temperature has a larger variance than 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

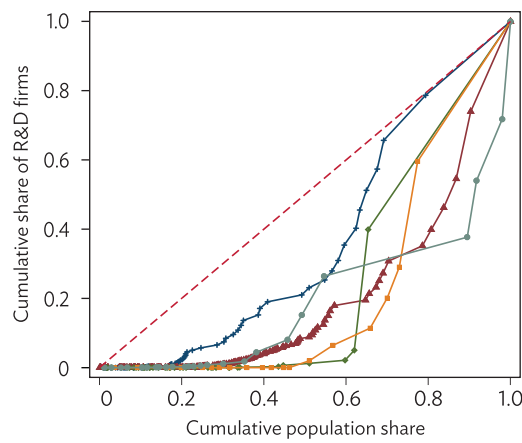
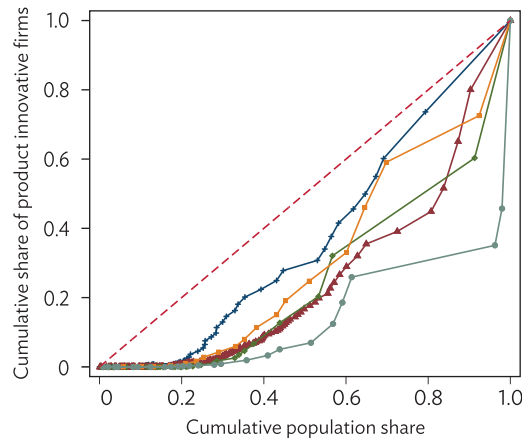
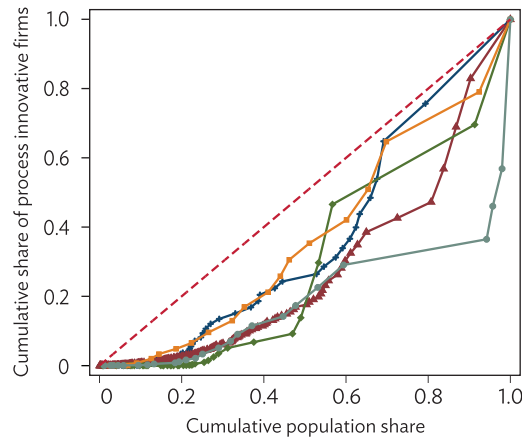
One stylized fact documented in the literature—which mostly covers developed countries as noted earlier—is that innovation is highly spatially concentrated. Within a given country, there are generally just a few cities that act as “innovation hubs.” These hubs host a large share of a country’s innovative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

Similar patterns are observed in our data. We calculate for each city the share of firms that have undertaken process innovations, product innovations, and R&D expenditures, and then compare these figures to each city’s share of population in the national total. A city with a higher share of firms engaged in innovation in comparison to 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. Together, these cities account 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, but together account 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-related activities. In general, the curves of Malaysia and India are more convex than those of the PRC, suggesting a higher degree of innovation concentration in those two countries.

The figure 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.

Figure 2: Cumulative Shares of Innovative Firms versus Urban Population



— China, People's Rep. of — India — Indonesia — Kazakhstan — Malaysia

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: Authors' estimates.

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 process innovation, product innovation, and R&D activity. Given the nature of the data at our disposal, we employ a probit model to estimate the likelihood of a firm partaking in a particular type of innovation activity as a function of the characteristics of the firm itself as well as the city in which it is located. We also control for country fixed effects in our analysis, allowing us to account for unobservable time-invariant differences in the innovation ecosystem across the 25 countries we study. These can include cross-country variations in institutions, culture, and nationwide policies that influence innovation. We also control for year fixed effects, thereby capturing differences across years, such as global and regionwide economic conditions that may affect firm dynamics across country and sectors. We also control for sector fixed effects. Therefore, our results are driven by factors that link firms' decisions regarding innovation to city-specific characteristics, 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 in 2010, that is, 2–6 years prior to the actual survey of firms in a country. We also control for a number of “first-nature” geographic characteristics that could affect innovation and city population (Combes et al. 2010). These are average rainfall, maximum highest temperature, minimum lowest temperature, average terrain ruggedness, and city's distance to the nearest port, and they are our city-level controls. Firm-level controls include firm size (defining a firm as small if its total employment is fewer than 50), firm age (defining a firm as young if it has been in operation for 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: process innovation, product 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 based on 2010 data and are time invariant.

$FIRM_{icnst}$ is a vector of firm controls, including indicators for firm age, size, FDI ownership, and headquarter, and share of skilled workers.

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

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 independent variables and β be the vector of coefficients listed above, so $y_{icnst}^* = X_{icnst}^T \beta + \epsilon_{icnst}$. Then denote $\Phi(\cdot)$ and $\phi(\cdot)$ as the cumulative distribution function and probability density function of standard normal distribution, respectively. The probability of firm innovation activity conditional on X_{icnst} can thus be written as

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

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

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

Results in the next section present our estimated probit model coefficients as well as the marginal effects for city population.

B. Instrumental Variable Estimation

Endogeneity issues could arise in our baseline specification, resulting in bias in the estimated coefficients. One reason is that a city's size and a firm's innovation-related activities may be determined simultaneously (Moomaw 1981). This could happen if the innovative outputs of firms expand their scale of production, thereby attracting more employment to a city and resulting in reverse causality (Duranton 2007, 2014). Another explanation for simultaneity is missing local variables that are correlated with both city size and innovation. For example, highly skilled workers in 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 population data from the World Urbanization Prospects (WUP) to instrument for the 2010 population used in the

baseline specification. The WUP collects urban populations for over 200 countries and covers cities with populations over 300,000 from 1950.¹⁰ 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. 2013, World Bank 2009).

Cities within the WUP data are defined by administrative boundaries. We reconciled this with our use of natural cities by mapping the WUP cities into our natural city boundaries using their latitude and longitude coordinates. Of the 789 WUP cities available across 23 of the 25 countries included in our study, 767 WUP cities could be mapped into 287 of our 489 natural cities.¹¹ Of the 202 natural cities that did not have corresponding WUP cities, the majority were relatively small, with an average population of 276,780 in 2010. This compares with the average population of 2,241,738 in 2010 for the other 287 natural cities (Table 1 lower panel). 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 the 1950–1959 population, the earliest available period from the WUP. Using the earliest period available helps further guarantee exogeneity, and using 10-year averages helps avoid data inaccuracies or outliers in historical WUP data.

The validity of our instrument relies on two conditions. First, there is some persistence in the spatial distribution of population. The satisfaction of this condition will govern the degree of correlation between the instrument and instrumented. Second, the local drivers of the dependent variable, in our case innovation-related activities undertaken from 2012 to 2016 differ from drivers of innovation in the past (Combes et al. 2010). The satisfaction of this condition will govern the degree of orthogonality between the instrument and error term. The historical population data we chose should satisfy these two conditions.

For the first condition to hold, the instrument must have a strong correlation with 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 confirm its relevance.

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

First, long-lagged values of the same variable will remove any simultaneity or reverse causality bias. It is highly implausible that the distribution of city population in the 25 developing Asian economies we examine would be determined by firms' innovation-related 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 $CITY_{cn}$, for example, temperature, rainfall, and distance to port.

¹⁰ The WUP only includes cities with population of 300,000 and above in the most recent year (2018). Data and documentation of this dataset can be accessed at: <https://population.un.org/wup/>.

¹¹ WUP data do not cover Bhutan and Timor-Leste. This brings us down by three cities and 292 firm observations in the IV analysis.

Second, after controlling for time-invariant geographic characteristics, the drivers of urban agglomeration in the 1950s are unlikely to be related to determinants of local innovation activities in the 2010s. This is because the current structure of the economies of the 25 developing Asian countries are very different compared to what they were 60 years before. Take the two biggest countries, India and the PRC, for example. In the early 1950s, both India and the PRC had just experienced drastic regime changes. India had gained independence from British colonialism, while the PRC was recovering from decade-long wars and the main economic activity undertaken 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 the 500 million increase in urban population experienced from 1978 to 2012 (World Bank; Development Research Center of the State Council, the People's Republic of China 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 currently, such as Shenzhen, a small aquaculture port city in the 1970s. This illustrates significant spatial change in both the concentration of urban population and innovation-related activities. India also has undergone a series of economic reforms, including the liberalization of trade and industrial policies 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 service 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 and West Asia were a part of the 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 RESULTS

A. Baseline and Instrumental Variable Estimates

Table 2 presents our baseline results. The first three columns report baseline probit model estimates using the full sample. The coefficients of our main variable of interest, logarithm population in 2010, are positive and 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 process and product innovations. For process innovation, the predicted propensity when the independent variables are at their mean values is 42.8%. The coefficient corresponds to a marginal effect of 0.037. 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 3.65 percentage points, or an 8.5% increase from the predicted propensity of 42.8%. For product innovation, the marginal effect is 4.25 percentage points, or a 13.5% increase from the predicted propensity of 31.4%. For R&D, the marginal effect is 2.79 percentage points, or a 14.3% increase from the predicted propensity of 19.5%.

In columns (4), (5), and (6) of Table 2, we perform the same regressions restricting observations to those with historical urban population data (thereafter referred to as the IV sample). The results are similar to those from the full-sample estimation and all estimated coefficients effects are statistically significant at the 1% level: a 7.6% increase in the propensity for process innovation, a

16.5% increase for product innovation, and an 11.5% increase for R&D, if the city population doubles. The probit estimates based on the IV sample provide direct comparison to the IV estimates and robustness check results reported in Tables 3 and 4.

Table 2: Probit Estimates of Effects of City Population on Firm Innovation

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)	.0929*** (.0215)	.12*** (.0254)	.101*** (.0246)	.0846*** (.0269)	.149*** (.0292)	.0837*** (.029)
Marginal effect	.0365*** (.0084)	.0425*** (.0091)	.0279*** (.0066)	.0334*** (.0106)	.0538*** (.0107)	.024*** (.0081)
Predicted probability	.428*** (.0114)	.314*** (.0114)	.195*** (.0106)	.44*** (.0115)	.326*** (.0121)	.209*** (.011)
Log distance to port (km)	.159*** (.0233)	.108*** (.0257)	.0248 (.0248)	.174*** (.024)	.101*** (.0266)	.0174 (.024)
Average precipitation (mm)	-.117*** (.0286)	-.124*** (.0295)	-.0827** (.0349)	-.126*** (.0311)	-.138*** (.0312)	-.0843** (.0374)
Max temp	.0389*** (.0146)	-.0047 (.0162)	.0909*** (.0164)	.0401*** (.0154)	-.0097 (.0169)	.0963*** (.0165)
Min temp	.03*** (.0073)	.0318*** (.009)	.0154* (.0084)	.0331*** (.0077)	.032*** (.0095)	.0158* (.0086)
Terrain ruggedness index (m)	.0028** (.0011)	.0015 (.0011)	.0024** (.0011)	.0029*** (.0011)	6.9e-04 (.0012)	.0022** (.0011)
Young firm	.196*** (.0654)	.13 (.0893)	.122* (.072)	.215*** (.0681)	.116 (.0893)	.155** (.0749)
Small firm	-.524*** (.0596)	-.214*** (.0633)	-.408*** (.0603)	-.541*** (.0601)	-.229*** (.0643)	-.422*** (.0633)
FDI	.116 (.133)	.329** (.139)	.222* (.124)	.109 (.139)	.315** (.144)	.237* (.128)
Headquarter dummy	-.0961 (.0946)	.322* (.176)	.0855 (.131)	-.0839 (.0964)	.347* (.18)	.0966 (.133)
Share of skilled workers	.142 (.145)	-.0914 (.136)	.0442 (.139)	.248 (.154)	-.122 (.138)	.101 (.145)
Sector/Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,374	21,496	20,218	17,838	18,903	17,673
No. of cities	483	489	480	285	287	283
F-statistic	33.42	21.82	25.48	31.73	20.16	24.46

FDI = foreign direct investment, FE = fixed effect, IV = instrumental variable, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, 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' estimates.

Other city characteristics are also correlated with innovation-related activities. The coefficients on average precipitation are negative and significant; coefficients on maximum temperature are positive and significant in models for process innovation and R&D; coefficients on minimum temperature are positive and significant; and, perhaps surprisingly, the coefficients for both terrain ruggedness and distance to the nearest port are found to be positive in two out of three innovation activities. 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-related activities. Smaller firms are less likely to participate in either product or process innovation or invest in R&D. Firms with foreign ownership are more likely to generate product innovations and invest in R&D. Headquarters are more likely to implement product innovations. These findings on firm characteristics are generally consistent with existing literature (Bertschek 1995, Huergo and Jaumandreu 2004, Yang 2017).

In Table 3, we instrument city population with the average of the 1950–1959 population size. Firstly, the instrument is strongly positively correlated with the contemporary city population, with coefficient estimates around 0.76 and first-stage F-statistics above 4,000. This further validates the first condition for the instrumental variable.

The IV estimation indicates that with endogeneity being addressed, a larger city size is still associated with a higher propensity for firm’s innovation activities. For process innovation, if population size doubles, the predicted propensity for process innovation will increase by 3.87 percentage points from an average propensity of 44%, or an 8.8% increase. For product innovation, the marginal effect suggests an increase of 4.81 percentage points from a predicted average propensity of 32.6%, or a 14.8% increase. For firm R&D, the marginal effect is an increase of 1.97 percentage points over the predicted propensity of 20.9 percentage points, or a 9.4% increase. All the estimates are statistically significant at the 5% or 1% level. Compared to their corresponding probit estimates in columns (4), (5), and (6) of Table 2, the IV estimates imply a slightly smaller impact of city size on product innovation and R&D investment.

To summarize, after accounting for endogeneity, we still find economically and statistically significant agglomeration effects on firms’ innovation inputs and outputs. The average increase in firms’ propensity to engage in process innovation, product innovation, and R&D ranges between 8% and 15% when city size doubles. These estimates are at the high end or larger than the estimated elasticities of wage or firm productivity with respect to city employment or urban density, which typically falls between 2% and 10% (Duranton 2014). On the other hand, the estimates are smaller than the estimated elasticity of patent intensity with respect to employment density of 20% reported in Carlino, Chatterjee, and Hunt 2007,

B. Robustness Testing

We perform robustness tests to examine a few factors that could potentially dampen our results. First, we test whether our results might be driven by a few megacities that host a large number of innovative firms. We check this potential “superstar” or outlier effect by excluding 13 cities with populations of over 10 million. This leaves us with about 70% of firm observations in 274 cities. The results are presented in columns (1), (2), and (3) of Table 4. The first stage F-statistics indicate that the IV remains

Table 3: Instrumental Variable Estimates of Effects of City Population on Firm Innovation

Variables	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D
First-stage estimates			
	log Population (2010)		
Log population (1950–1959)	.768*** (.0131)	.757*** (.0117)	.77*** (.0132)
Other controls	Yes	Yes	Yes
F-statistic	4,642.2	4,334.6	4,690.0
Second-stage estimates			
Log population (2010)	.0982*** (.0288)	.133*** (.0311)	.0685** (.0286)
Marginal effect	.0387*** (.0113)	.0481*** (.0113)	.0197** (.0081)
Predicted probability	.44*** (.0115)	.326*** (.0122)	.209*** (.011)
Log distance to port (km)	.175*** (.0241)	.101*** (.0265)	.0164 (.0238)
Average precipitation (mm)	-.128*** (.031)	-.135*** (.0313)	-.0832** (.0374)
Max temp	.0388** (.0155)	-.0085 (.0171)	.0962*** (.0165)
Min temp	.0324*** (.0078)	.0326*** (.0096)	.0168* (.0086)
Terrain ruggedness index (m)	.003*** (.0011)	5.5e-04 (.0012)	.002* (.0011)
Young firm	.215*** (.068)	.117 (.0894)	.155** (.0748)
Small firm	-.54*** (.0601)	-.229*** (.0642)	-.422*** (.0633)
FDI	.103 (.139)	.323** (.143)	.243* (.128)
Headquarter dummy	-.084 (.0955)	.348* (.181)	.1 (.133)
Share of skilled workers	.245 (.154)	-.119 (.139)	.104 (.145)
Sector/Year/Country FE	Yes	Yes	Yes
Observations	17,838	18,903	17,673
No. of cities	285	287	283
F-statistic	31.53	20.38	24.77

FDI = foreign direct investment, FE = fixed effect, km = kilometer, log = logarithm, m = meter, max = maximum, min = minimum, 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' estimates.

relevant. The IV estimation yields considerably larger coefficient estimates and marginal effects. As compared with the baseline probabilities, which are lower than those of the full sample, the propensity to engage in innovation would increase by 27%, 23%, and 27% for process innovation, product innovation, and R&D investment, respectively, if city size doubled. This result suggests that although firms in these megacities were more likely to innovate, city size can have more significant impacts on firms located in nonmega cities.

Second, we test how sensitive the estimates are to including cities in which only a small number of firms were surveyed in the WBES. We reestimate the models restricting our sample to cities hosting 20 or more surveyed firms. This leaves us with about 95% firm observations in just 151 cities. The results are shown in columns (4), (5), and (6) of Table 4. For process innovation, the coefficient estimate drops by a third to 0.065 but remains statistically significant at the 5% level. The estimate for product innovation remains similar to the full-sample IV estimate. The effect of city size on R&D outcome, however, diminishes and becomes insignificant. To the extent that the excluded cities are smaller ones, the results imply that above a certain population level, firm R&D, and process innovation to a lesser degree are not much affected by city size. However, the propensity gradient with respect to product innovation appears consistent over the whole city size distribution.

In addition, we use two alternative dependent measures to test if our results are sensitive to the definition of innovation. The first indicator concerns whether the product innovation is “also new to the establishment’s main market.” Answers to this question are expected to be more reliable as the question requires the respondents to think more about their answer to the preceding question regarding general product innovation and also gets at the importance of an innovation. As expected, this follow-up question yields a lower percentage of positive responses reported in the sample—21% as opposed to 36% for the general question on product innovation. The regression results in column (7) show that doubling city size could lead to 2.8 percentage points or 19% increase in the propensity of main market product innovation, slightly higher than 15% using the general product innovation measure.

The second alternative measure of innovation takes a value of 1 if a firm conducts either product or process innovation. The sample average equal to 56% is greater than the averages of both process innovation and product innovation and smaller than the sum of the two. This implies that respondents were able to differentiate between the two types of innovation and a fair proportion of firms engaged in only one of the two. The estimated marginal effect is 5.6 percentage points, or a 9.4% increase over the predicted probability, with 1% statistical significance.

C. Heterogeneity by Development Levels and Countries

The 25 Asian countries we study are at various stages of development. It is interesting to examine whether or how the agglomeration effects on firm innovation vary with development levels. We thus divide our sample into upper-middle income-countries, and low and lower-middle-income countries, according to the latest World Bank country classifications. We then estimate the IV-probit regressions with each subsample. The results are reported in Table 5.

For low and lower-middle-income countries that encompass about 80% of firm observations in 195 cities, the agglomeration effects are positive and statistically significant for product innovation and R&D, but small and insignificant for process innovation. The estimated marginal effect for R&D is comparable to that of the full sample. However, the increase of 13% exceeds the 9% estimated for the full sample.

Table 4: Robustness Check on Instrumental Variable Estimates of Effects of City Population on Firm Innovation

Variables	Excluding Cities with >10M Population			Excluding Cities with <20 Firms			Using Other Measures of Firm Innovation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D	Main Product Innovation	Firm Innovation
First-stage estimates								
	Log population (2010)							
Log population (1950–1959)	.613*** (.0137)	.592*** (.0115)	.614*** (.0138)	.785*** (.0099)	.783*** (.0092)	.787*** (.0099)	.847*** (.0099)	.757*** (.0124)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	2,723.4	3,047.9	2,742.9	6,606.2	6,832.6	6,629.1	4,334.6	4,421.9
Second-stage estimates								
	Log population (2010)							
Log population (2010)	.299*** (.0459)	.193*** (.0489)	.185*** (.0439)	.0646** (.0319)	.124*** (.0381)	.0266 (.0336)	.122*** (.0337)	.143*** (.0304)
Marginal effect	.118*** (.0181)	.0665*** (.0169)	.0498*** (.012)	.0255** (.0126)	.0449*** (.0139)	.0077 (.0097)	.0284*** (.0078)	.0555*** (.0119)
Predicted probability	.439*** (.0137)	.293*** (.0129)	.187*** (.0159)	.445*** (.0118)	.331*** (.0127)	.213*** (.0113)	.15*** (.0101)	.588*** (.0109)
Log distance to port (km)	.152*** (.0272)	.0967*** (.0294)	.0034 (.027)	.184*** (.0249)	.103*** (.0273)	.0196 (.0239)	.152*** (.0384)	.143*** (.0214)
Average precipitation (mm)	-.0137 (.0368)	-.0809** (.0371)	.0958* (.0384)	-.132*** (.033)	-.132*** (.0335)	-.0877** (.0364)	-.193*** (.0308)	-.107*** (.0309)
Max temp	.031** (.0149)	-.0141 (.0163)	.0867*** (.0152)	.0468*** (.0161)	-.0034 (.0179)	.103*** (.0169)	-.0378** (.0163)	.0666*** (.0147)
Min temp	.019** (.0088)	.0149 (.0111)	.0156* (.0089)	.0373*** (.0081)	.0326*** (.0101)	.0208** (.0084)	.0594*** (.0114)	.022*** (.0077)
Terrain ruggedness index (m)	.0041*** (.0011)	.0024** (.0012)	.0029** (.0013)	.0029*** (.0011)	5.6e-04 (.0012)	.0018 (.0012)	.0043** (.002)	.002* (.0011)
Young firm	.0735 (.0802)	-.141 (.102)	-.0333 (.0886)	.2*** (.071)	.103 (.0924)	.151** (.0757)	.169* (.0876)	.152** (.0681)

continued on next page

Table 4 continued

Variables	Excluding Cities with >10M Population			Excluding Cities with <20 Firms			Using Other Measures of Firm Innovation	
	(1) Process Innovation	(2) Product Innovation	(3) Firm R&D	(4) Process Innovation	(5) Product Innovation	(6) Firm R&D	(7) Main Product Innovation	(8) Firm Innovation
Second-stage estimates								
Small firm	-.478*** (.0725)	-.298*** (.0656)	-.483*** (.0762)	-.525*** (.062)	-.207*** (.0667)	-.384*** (.0658)	-.306*** (.0671)	-.395*** (.0608)
FDI	.326* (.181)	.143 (.156)	.261* (.153)	.103 (.14)	.313** (.146)	.249* (.13)	-.251 (.184)	.271** (.131)
Headquarter dummy	.0743 (.113)	.315 (.199)	.249** (.113)	-.0785 (.0939)	.361* (.191)	.104 (.139)	-.23** (.105)	.264** (.107)
Share of skilled workers	-.0159 (.185)	-.129 (.16)	.0409 (.196)	.131 (.16)	-.207 (.146)	-.0138 (.15)	-.0632 (.154)	.429** (.185)
Sector/Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,556	13,348	12,368	16,986	18,013	16,818	16,427	18,345
No. of cities	272	274	270	151	151	149	232	286
F-statistic	26.49	17.17	19.63	33.27	21.93	26.67	18.41	36.45

FDI = foreign direct investment, FE = fixed effect, km = kilometer, log = logarithm, m = meter, M = million, max = maximum, min = minimum, mm = millimeter, R&D = research and development.

Notes: Robust standard errors clustered by country in parentheses. *** = p<0.01, ** = p<0.05, * = p<0.1. Columns (1)–(3) are samples restricted to cities with <=10M population, and columns (4)–(6) are samples restricted to cities with >=20 firms. Columns (7) and (8) use the full sample.

Source: Authors' estimates.

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

Variables	Low and 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)	.847*** (.0097)	.847*** (.0097)	.848*** (.0097)	.699*** (.0125)	.699*** (.0099)	.7*** (.0126)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	11,576.2	11,865.7	11,780.8	7,944.9	13,511.2	8,266.7
Second-stage estimates						
Log population (2010)	.042 (.0287)	.102*** (.0281)	.0839** (.0338)	.149*** (.0394)	.163*** (.0434)	.0518 (.0396)
Marginal effect	.016 (.0109)	.0335*** (.0092)	.0191*** (.0073)	.0596*** (.0157)	.0612*** (.0167)	.0179 (.0137)
Predicted probability	.38*** (.0143)	.267*** (.0128)	.145*** (.0137)	.513*** (.0155)	.366*** (.016)	.297*** (.0153)
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
No. of cities	195	195	195	90	92	88
F-statistic	23.76	15.82	19.42	23.01	19.24	18.96

FE = fixed effect, 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' estimates.

The upper-middle-income subsample is only one-third of the size of the lower-middle and low-income subsample. But we still obtain strong agglomeration effects on process and product innovations. The estimated marginal effect and percent increase for process innovation are substantially larger than those for the full sample. On the other hand, the effect on R&D is positive but statistically insignificant.

Overall, the analysis suggests that at earlier stages of development, firm R&D is more concentrated in large cities, while process innovation spreads out relatively equally across city sizes. Once countries are at higher income levels, process innovation tend to concentrate in large cities whereas R&D is undertaken more universally across cities. Meanwhile, agglomeration effects exist consistently for product innovation regardless of a country's development level.

The 25 countries are also diverse in terms of institutional or cultural backgrounds. It is worth examining how the agglomeration effects on firms' innovation change in these different social settings. We consider three cases, namely the PRC, India, and selected ASEAN countries. The results are summarized in Table 6.

The estimated marginal effects are all positive and statistically significant across the three countries and subregions except those on R&D in the PRC and ASEAN countries. The insignificance for

Table 6: Instrumental Variable Estimates of Effects of City Population on Firm Innovation in the People's Republic of China, India, and ASEAN countries

Variables	China, People's Rep. of			India			ASEAN		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D	Process Innovation	Product Innovation	Firm R&D
	Log population (2010)			Log population (2010)			Log population (2010)		
Log population (1950–1959)	.674*** (.0105)	.682*** (.0079)	.675*** (.0107)	.807*** (.0083)	.807*** (.0082)	.807*** (.0083)	.872*** (.0121)	.873*** (.0121)	.875*** (.012)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F–statistic	7,937.0	9,318.1	7,781.3	4,197.8	4,216.0	4,205.0	4,870.8	5,040.4	4,973.1
Second-stage estimates									
Log population (2010)	.148*** (.0418)	.167*** (.0459)	.0537 (.0417)	.194*** (.0262)	.202*** (.0327)	.137*** (.028)	.131** (.0578)	.137** (.0611)	.0668 (.0661)
Marginal effect	.0576*** (.0164)	.0655*** (.0182)	.0207 (.016)	.0755*** (.0102)	.0791*** (.0128)	.0465*** (.0097)	.0363** (.0163)	.0227** (.0105)	.008 (.008)
Predicted probability	.589*** (.0165)	.432*** (.0176)	.395*** (.0166)	.584*** (.0128)	.426*** (.0127)	.287*** (.013)	.198*** (.0185)	.0922*** (.0124)	.0608*** (.0118)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year/Country FE	No	No	No	No	No	No	Yes	Yes	Yes
Observations	1,544	2,476	1,533	7,450	7,463	7,445	4,894	4,971	4,966
No. of cities	53	55	53	130	130	130	59	59	59
F–statistic	18.83	5.94	13.46	37.86	14.26	23.68	9.64	10.65	7.05

ASEAN = Association of Southeast Asian Nations, FE = fixed effect, 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), and (4), (5), and (6) are samples restricted to cities in the People's Republic of China and India, respectively. Columns (7), (8), (9) are samples restricted to ASEAN countries which are Cambodia, Indonesia, the Lao People's Democratic Republic, Myanmar, Malaysia, the Philippines, Thailand, and Viet Nam.

Source: Authors' estimates.

R&D in Chinese cities may be due to the small sample size since the point estimate (0.021) is of comparable magnitude. India had larger marginal effects on all three innovation measures than the PRC did, and hence greater agglomeration effects. ASEAN countries' marginal effects are smaller than those of these two countries. Owing to their low baseline probabilities, however, the percent changes in product and process innovations in ASEAN countries are substantial (18% and 25%, respectively, with city size doubling), outweighing those of the PRC and India.

V. CHANNELS UNDERLYING AGGLOMERATION EFFECTS

So far, we have established the effect of city size on firm innovation, and which appears pervasive across developing Asia. In this section, we investigate two potential channels through which such agglomeration effects may arise. Specifically, we examine whether the presence of a top-ranked university in large cities encourages innovation-related activities of firms; we also examine whether city size influences R&D-related outcomes.

Universities, especially prestigious ones, are often pioneers in pushing the knowledge frontier outward and exploring the uncharted. There are several ways for universities and firms in the same city to interact including university-firm collaborations through market-mediated interactions; unintended knowledge flows from university-based research (D'Este and Iammarino 2010, D'Este and Patel 2007); and universities as suppliers of human capital to local firms (Toivanen and Väänänen 2016). In sum, firm innovation could benefit from geographical proximity to a university.

We measure the availability and number of quality tertiary institutions using the list of top 500 universities in Asia identified in the recent QS Asia University Rankings (QS 2019). Of the 500 universities, 248 are mapped into 99 cities in our sample. Table 7 shows that there is a high degree of spatial concentration of these top universities. First, the top universities are unevenly distributed across countries. Only nine of the 25 developing Asian countries have a top university, and the majority of them (211 or 85% of total) are located in five countries—the PRC (78), India (64), Malaysia (25), Pakistan (22),

Table 7: Distribution of Top Universities

Country	Number of Universities	Number of Cities with Top Universities	City with the Most Top Universities
Bangladesh	6	1	Dhaka (6)
China, People's Rep. of	78	17	Beijing (20)
India	64	43	Delhi (7)
Indonesia	22	9	Bandung (5)
Malaysia	25	10	Kuala Lumpur (14)
Pakistan	22	5	Lahore (9)
Philippines	6	2	Metro Manila (5)
Thailand	18	7	Bangkok (12)
Viet Nam	7	5	Ha Noi (2)
Total	248	99	

Source: Authors' estimates.

and Indonesia (22). Second, the universities are also unevenly distributed within these countries. For example, all six top universities in Bangladesh are located in Dhaka; 14 of the 25 top universities in Malaysia are in Kuala Lumpur; and 12 of the 18 top universities are in Thailand in Bangkok.

We include university dummy and number of universities in the IV-Probit models, respectively, and present the results in Table 8. First, the presence of a top university is strongly associated with a higher propensity of firms introducing product and process innovations and investing in R&D. The three coefficient estimates are positive and significant, with the effects more pronounced for process innovation. The association is twice as strong for process innovation (20.6 percentage points higher when a top university is present) compared to product innovation and R&D (9.2 and 10.6 percentage points, respectively). The number of top universities is also positively correlated with the probability of process innovation and R&D investment by firms. The estimates suggest that if a city has one additional top university, its firms on average are 1 percentage point more likely to introduce a process innovation or invest in R&D. However, the association between product innovation and number of top universities is much weaker.

Table 8: Agglomeration Channels: Presence of Top Universities

Variables	(1) Process Innovation	(2)	(3) Product Innovation	(4)	(5) Firm R&D	(6)
First-stage estimates	Log population (2010)		Log population (2010)		Log population (2010)	
Log population (1950–1959)	.738*** (.0164)	.636*** (.0137)	.744*** (.0146)	.638*** (.0117)	.74*** (.0165)	.638*** (.0138)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	4,885.1	5,982.1	4,342.4	8,138.2	4,936.6	6,293.2
Second-stage estimates						
Log population (2010)	-.0374 (.0394)	.0253 (.0429)	.066 (.0424)	.111** (.0445)	-.0282 (.0402)	-.0072 (.0436)
University dummy	.521*** (.0885)		.257*** (.0962)		.368*** (.0951)	
No. of universities		.0331** (.0142)		.0085 (.0131)		.0324** (.0146)
Marginal effect	.206*** (.0351)	.0131** (.0056)	.0925*** (.0347)	.003 (.0047)	.106*** (.0282)	.0093** (.0042)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector/Year/Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,838	17,838	18,903	18,903	17,673	17,673
No. of cities	285	285	287	287	283	283
F-statistic	32.08	30.60	19.92	19.86	24.29	24.12

FE = fixed effect, 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' estimates.

After controlling for the university dummy or number of universities, the effect of city size diminishes in magnitude and statistical significance, except for product innovation with number of universities as regressor. This implies that colocation of high-quality universities and firms serves as a key channel for urban agglomeration to promote firm innovation. Nevertheless, it is noted that the university variables could be endogenous. For instance, more innovative firms would probably choose to move to cities with good universities. Lacking information on an exogenous determinant of the location of top universities in the sample, we consider our results as suggestive.

Another channel through which agglomeration may influence firm innovation positively is by raising the effectiveness of investments in innovative activity, such as R&D. This may happen 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 implies a larger pool of skilled labor and experts. Firms in larger cities would therefore be more efficiently matched with the most suitable R&D employees.

To benchmark the analysis, we first include the firm R&D dummy as an explanatory variable for the process innovation and product innovation models. It is expected that R&D investment is positively correlated with the introduction of process and/or product innovations at the firm level. The regression results are reported in columns (1) and (3) of Table 9. Not surprisingly, the firm R&D indicator is strongly correlated with both types of innovation. Meanwhile, the coefficient estimates for city size remain positive and unchanged in magnitude.

Table 9: Agglomeration Channels: Differential Effects of Firm Research and Development

Variables	(1)	(2)	(3)	(4)
	Process Innovation		Product Innovation	
First-stage estimates	Log population (2010)		Log population (2010)	
Log population (1950–1959)	.768*** (.0133)	.769*** (.0149)	.769*** (.0133)	.77*** (.0149)
Firm R&D x log Population (1950–1959)		–.0026 (.023)		–.003 (.023)
Other controls	Yes	Yes	Yes	Yes
F-statistic	4,582.1	4,459.8	4,650.4	4,525.0
	Firm R&D x Log population (2010)		Firm R&D x Log population (2010)	
Log population (1950–1959)		–4.4e-05 (.0031)		2.7e-04 (.0031)
Firm R&D x Log population (1950–1959)		.778*** (.0211)		.778*** (.021)
Other controls		Yes		Yes
F-statistic		42,288.9		43,662.8

continued on next page

Table 9 *continued*

Variables	(1)	(2)	(3)	(4)
	Process Innovation		Product Innovation	
Second-stage estimates				
Log population (2010)	.0913*** (.0291)	.0955*** (.0327)	.152*** (.0251)	.074*** (.0267)
Firm R&D	1.31*** (.0721)	1.66* (.884)	1.33*** (.0658)	-2.59*** (.768)
Firm R&D x Log population (2010)		-.0223 (.0563)		.252*** (.0495)
Other controls	Yes	Yes	Yes	Yes
Sector/Year/Country FE	Yes	Yes	Yes	Yes
Observations	17,741	17,741	17,861	17,861
No. of cities	285	285	285	285
F-statistic	36.70	35.95	33.67	33.17

FE = fixed effect, 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' estimates.

We then 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 positive coefficient on this term implies that firms that conduct R&D reap more benefits from agglomeration than firms that do not conduct R&D. Results for process and product innovations, respectively, are reported in columns (2) and (4) of Table 9. The first-stage estimates for both endogenous variables are strong. In the second stage, two results are of interest. First, for process innovation, the coefficient of the interaction term is small and insignificant, suggesting little difference in how agglomeration affects process innovation between firms that conduct R&D and those that do not. Furthermore, including the interaction term has no effect on the coefficient of city population. However, for product innovation, the estimated coefficient of the interaction term is positive and significant at the 1% level. Meanwhile, the estimated coefficient on population size decreases by half compared to the no interaction case. Thus, we obtain some evidence that firms' investment in R&D is more effective in generating product innovation in larger cities, shedding light on a channel through which agglomeration effects on firm innovation may work.

VI. CONCLUSION

To the best of our knowledge, this paper is among the first to examine the spatial distribution of developing country firms' engagement in innovation-related activities, and whether and how urban agglomeration affects the propensity of firms to innovate. Our analysis shows that firm innovation is highly concentrated in larger cities, although not all megacities host disproportionately more innovative firms. Doubling city size would, on average, increase a firm's propensity to introduce a process innovation by 3.9 percentage points and product innovation by 4.8 percentage points, and for R&D investment by 2 percentage points, all else being equal. The implied percent increases in the probabilities of firms undertaking innovation activities are prominent given the moderate starting points in the countries studied. Furthermore, we find that the relationship between city size and

innovation propensities is robust and pervasive in developing Asia. Urban agglomeration effects are heterogeneous across different subsamples defined in terms of development level or country background, but the differences seem to be secondary.

We also explore two potential channels through which agglomeration benefits may accrue: through the presence of a quality university locally, and by enhancing the effectiveness of R&D undertaken by firms. We find suggestive evidence that firms in cities with top universities are more likely to carry out innovation-related activities, especially investing in R&D and introducing process innovations. Further, firms investing in R&D are more likely to introduce a product innovation in larger urban agglomerations.

Our results have some clear policy implications. First, in light of the importance of innovation to economic growth, policy makers in developing economies should be more welcoming to the growth of larger urban agglomerations than they often are. In particular, calls for regionally balanced development and a belief that large cities have become too large can lead to policies that inadvertently undermine the potential benefits from larger urban agglomeration. Second, policies to improve university quality and promote interactions between universities and local firms should have high payoffs. Further, setting up new universities or new campuses may be prioritized for those larger cities without one.

There are plenty of areas identified for future research. For instance, we detect certain differences between process innovation and product innovation. Not only do they have unequal baseline probabilities, they also respond to agglomeration differently across low and lower-middle-income countries and upper-middle-income countries. The two knowledge spillover channels also seem to have distinct impacts on the two types of innovations. It is worth gaining an in-depth understanding of the differences and their policy implications. It would also be interesting to study how other city characteristics such as the quality of transport connectivity within cities affects the spatial distribution of process and product innovations.

It is also important to understand the channels through which knowledge spillovers arise between colocated universities and firms in developing economies. An increasing amount of studies look at interactions between the two, but largely focus on developed economies. When firm innovation does not involve patented cutting-edge technologies—a case very pertinent for developing country firms—how do universities help firms become more efficient and competitive? Does human capital or resource outputs from universities play a more significant role?

APPENDIX: DEVELOPMENT OF NATURAL CITY DATASET OF ASIA AND THE PACIFIC

Development of the natural city dataset involved three main steps. Firstly, 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 (km²) at the equator up to 2013, with the resolution of the pixel improved to about 0.22 km² after 2013. For a comprehensive introduction of the NTL data (DMSP-OLS part) and its applications in economics, see 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, Pozzi, and Elvidge 2005). The post-2013 VIIRS images have much less blooming due to their higher resolution. We applied the latest methodology developed in Abrahams, Oram, and Lozano-Gracia (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, 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.¹² The exercise yielded between 88,000 and 187,000 geocoded polygons for Asia across 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).¹³ This database contains geocoded centers, names, populations, and the upper administrative divisions they belong to, for over

¹² A pixel with positive luminosity value is illuminated. It is noted that 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 lifecycle. 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.

¹³ 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>. It is worth noting that the units in the GRUMP database are not from the same administrative level within, as well as across, countries.

70,000 human settlements across the world. We focused on over 1,900 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.¹⁴

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.¹⁵ In terms of our example of Metro Manila, also known as the National Capital Region, there were nine natural cities as shown in the right panel of Figure A1: 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, Quezon City, etc.

Some natural cities have 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, so as 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, historical population, etc. We filled the delineated areas of the natural cities with grid population data from LandScan. LandScan provides global population counts at approximately 1-km 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.¹⁶ 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 Asia 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 Asia, 248 were found in 99 cities of nine developing Asian countries. A binary indicator was created equal to 1 if a natural city hosted at least one top 500 university.

¹⁴ 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—as well as in developed economies such as the metropolitan areas surrounding Tokyo in Japan or the capital city of Taipei, China.

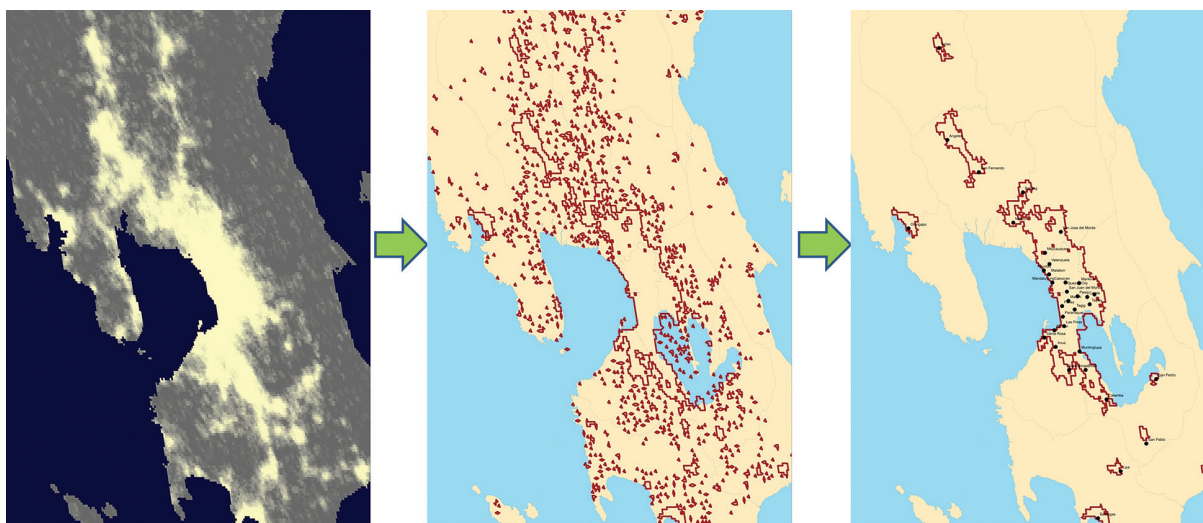
¹⁵ Please note the population threshold adopted refers to the population of the GRUMP units. This helped us to capture the most sizable urban agglomerations in the region. However, by no means does it imply that our natural cities had population over 100,000 prior to 2000, or even after 2000, because natural cities could cover very different geographic scopes from those of GRUMP units, and their populations were estimated based on grid population data.

¹⁶ 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.

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: Authors’ 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 PDR	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%

Lao PDR = Lao People's Democratic Republic.

Notes: 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' estimates.

REFERENCES

- Abrahams, Alexei, Christopher Oram, and Nancy Lozano-Gracia. 2018. "Deblurring DMSP nighttime lights: A New Method Using Gaussian Filters and Frequencies of Illumination." *Remote Sensing of Environment* 210: 242–258.
- Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. 2006. "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association* 4 (1): 37–74.
- Andrews, Michael. 2017. "The Role of Universities in Local Invention: Evidence from the Establishment of US Colleges." *Job Market Paper*.
- Anselin, Luc, Attila Varga, and Zoltan Acs. 1997. "Local Geographic Spillovers between University Research and High Technology Innovations." *Journal of Urban Economics* 42 (3): 422–448.
- Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *The American Economic Review* 86 (3): 630–640.
- Audretsch, David B., and Paula E. Stephan. 1996. "Company-Scientist Locational Links: The Case of Biotechnology." *The American Economic Review* 86 (3): 641–652.
- Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2011. "Firm Innovation in Emerging Markets: The Role of Finance, Governance, and Competition." *Journal of Financial and Quantitative Analysis* 46 (6): 1545–1580.
- Baum-Snow, Nathaniel, and Fernando Ferreira. 2015. "Causal Inference in Urban and Regional Economics." In *Handbook of Regional and Urban Economics*. Vol. 5: 3–68. Elsevier.
- Beck, Thorsten, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2005. "Financial and Legal Constraints to Growth: Does Firm Size Matter?." *The Journal of Finance* 60 (1): 137–177.
- Berliant, Marcus, Robert R. Reed III, and Ping Wang. 2006. "Knowledge Exchange, Matching, and Agglomeration." *Journal of Urban Economics* 60 (1): 69–95.
- Bertschek, Irene. 1995. "Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investment." *The Journal of Industrial Economics*, 341–357.
- Bhagwati, Jagdish, and Arvind Panagariya. 2013. *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries*. United Kingdom: Hachette.
- Buettner, Thomas. 1991. "Urban Estimates and Projections at the United Nations: The Strengths, Weaknesses, and Underpinnings of the World Urbanization Prospects." *Spatial Demography* 3 (2): 91–108.
- Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. 2007. "Urban Density and the Rate of Invention." *Journal of Urban Economics* 61 (3): 389–419.
- Carlino, Gerald, Robert Hunt, Gilles Duranton, and Bruce A. Weinberg. 2009. "What Explains the Quantity and Quality of Local Inventive Activity?." *Brookings-Wharton Papers on Urban Affairs*, 65–123.

- Carlino, Gerald A., and William R. Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Regional and Urban Economics*. Vol. 5: 349–404. Elsevier.
- Ciccone, Antonio, and Robert E. Hall. 1996. Productivity and the Density of Economic Activity. *The American Economic Review*, Vol. 86: 54–70.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Effects." In *Agglomeration Economics*, 15–66. University of Chicago Press.
- Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98.
- Duranton, Gilles. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review* 97 (1): 197–221.
- Duranton, Gilles. 2008. "From Cities to Productivity and Growth in Developing Countries." *Canadian Journal of Economics/Revue Canadienne d'économie* 41 (3): 689–736.
- Duranton, Gilles. 2014. *Growing through Cities in Developing Countries*. Washington, DC: World Bank.
- Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 4: 2063–2117. Elsevier.
- D'Este, Pablo, and Simona Iammarino. 2010. "The Spatial Profile of University - Business research Partnerships." *Papers in Regional Science* 89 (2): 335–350.
- D'Este, Pablo, and Pari Patel. 2007. "University-Industry Linkages in the UK: What are the Factors Underlying the Variety of Interactions with Industry?." *Research Policy* 36 (9): 1295–1313.
- Feldman, Maryann P., and David B. Audretsch. 1999. "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition." *European Economic Review* 43 (2): 409–429.
- Gerlach, Heiko, Thomas Rønde, and Konrad Stahl. 2009. "Labor Pooling in R&D Intensive Industries." *Journal of Urban Economics* 65 (1): 99–111.
- Helsley, Robert W., and William C. Strange. 1990. "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics* 20 (2): 189–212.
- Helsley, Robert W., and William C. Strange. 2002. "Innovation and Input Sharing." *Journal of Urban Economics* 51 (1): 25–45.
- Henderson, J. Vernon. 2000. *The Effects of Urban Concentration on Economic Growth*. No. w7503. National Bureau of Economic Research.
- Henderson, J. Vernon. 2005. Urbanization and Growth. In *Handbook of Economic Growth*. Vol. 1: 1543–1591. Elsevier.

- Huergo, Elena, and Jordi Jaumandreu. 2004. "How Does Probability of Innovation Change with Firm Age?" *Small Business Economics* 22 (3-4): 193–207.
- Jaffe, Adam B. 1989. "Real Effects of Academic Research." *The American Economic Review*, 957–970.
- Jiang, Yi. 2020. "Asian Cities: Spatial Dynamics and Driving Forces." ADB Economics Working Paper Series. (forthcoming).
- Kinda, Tidiane. 2010. "Investment Climate and FDI in Developing Countries: Firm-level Evidence." *World Development* 38 (4): 498–513.
- Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics* 39 (3): 332–342.
- Montgomery, Mark R., Richard Stren, Barney Cohen, and Holly E. Reed. 2013. *Cities Transformed: Demographic Change and Its Implications in the Developing World*. Routledge.
- Moomaw, Ronald L. 1981. "Productivity and City Size: A Critique of the Evidence." *The Quarterly Journal of Economics* 96 (4): 675–688.
- Moretti, Enrico. 2019. *The Effect of High-Tech Clusters on the Productivity of Top Innovators*. No. w26270. National Bureau of Economic Research.
- Mukherjee, Arpita. 2013. The Service Sector in India. *Asian Development Bank Economics Working Paper Series*. 352.
- Nieto Galindo, Victor Manuel. 2007. "La Aglomeración Como Una Causa de la Innovación en Colombia." *Planeación & Desarrollo*. 38 (1): 97–139.
- Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics*. 94 (1): 20–36.
- Ó hUallcháin, Breandán. 1999. "Patent Places: Size Matters." *Journal of Regional Science* 39 (4): 613–636.
- Overman, Henry G., and Anthony J. Venables. 2005. *Cities in the Developing World*. No. 695. Centre for Economic Performance, London School of Economics and Political Science.
- Paunov, Caroline. 2016. "Corruption's Asymmetric Impacts on Firm Innovation." *Journal of Development Economics* 118: 216–231.
- Paunov, Caroline, and Valentina Rollo. 2016. "Has the Internet Fostered Inclusive Innovation in the Developing World?" *World Development* 78: 587–609.
- Puga, Diego. 2010. "The Magnitude and Causes of Agglomeration Economies." *Journal of Regional Science* 50 (1): 203–219.
- Quacquarelli Symonds (QS). 2019. QS Asia University Rankings 2019. <http://www.topuniversities.com/university-rankings/asian-university-rankings/2019> (accessed May 15, 2019).

Reyes, José-Daniel, Mark Roberts, and Lixin Colin Xu. 2017. *The Heterogeneous Growth Effects of the Business Environment: Firm-level Evidence for a Global Sample of Cities*. The World Bank.

Riley, Shawn J., Stephen D. DeGloria, and Robert Elliot. 1999. "Index that Quantifies Topographic Heterogeneity." *Intermountain Journal of Sciences* 5 (1-4): 23–27.

Romer, Paul M. 1986. "Increasing Returns and Long-run Growth." *Journal of Political Economy* 94 (5): 1002–1037.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 4: 2119–2171. Elsevier.

Small, Christopher, Francesca Pozzi, and Christopher D. Elvidge. 2005. "Spatial Analysis of Global Urban Extent from DMSP-OLS Night Lights." *Remote Sensing of Environment* 96 (3-4): 277–291.

Strange, William, Walid Hejazi, and Jianmin Tang. 2006. "The Uncertain City: Competitive Instability, Skills, Innovation and the Strategy of Agglomeration." *Journal of Urban Economics* 59 (3): 331–351.

Toivanen, Otto, and Lotta Väänänen. 2016. "Education and Invention." *Review of Economics and Statistics* 98 (2): 382–396.

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

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

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

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

Urban Agglomeration and Firm Innovation

Evidence from Asia

This paper presents evidence on the spatial distribution and effects of urban agglomeration on firm innovation. It uses a unique dataset that consistently defines city boundaries and identifies firms' innovation-related activities across 25 developing countries in Asia. The analysis finds firm innovation to be highly concentrated at the city level. It finds substantial gains from increases in city population in terms of firms' propensity to introduce process and product innovations and undertake research and development activities. The paper also presents evidence that knowledge spillovers are an important channel through which agglomeration effects occur.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68 members—49 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.



ASIAN DEVELOPMENT BANK

6 ADB Avenue, Mandaluyong City

1550 Metro Manila, Philippines

www.adb.org