Building Up and Boosting Productivity: Evidence from the Daylight Spacing Regulation in China

Zeyu Li*

Version: November 4, 2023

Abstract

Estimating the productivity benefits of urban density is challenging due to the endogeneity of density. In this study, I exploit variations arising from a nationwide building spacing policy in China, which imposes restrictions on the distance between buildings. I construct a policy-implied spacing factor and use it as an instrument to estimate the causal effect of density. The results indicate that population density has a positive impact on wages, with an elasticity of approximately 40%. Additionally, the findings suggest that density benefits the economy by enhancing the output of the service sector as well as promoting innovation.

Keywords: Agglomeration Economies, City, Density, Wage, Productivity, Regulation, Developing Country

JEL Codes: R31, R38, R52, R58

1 Introduction

Density, as a measure of agglomeration, is often seen as a positive factor that can increase productivity, foster innovation, and make goods and services more accessible (Duranton and Puga, 2020). In developing economies, rapid urbanization has accompanied income growth. However, agglomeration can also have negative effects, such as traffic congestion, high housing prices, and pollution. Understanding the causal relationship between density and productivity is important not only in location choices for workers but in understanding policies that promote or control agglomeration.

*Department of Economics, University of California, Irvine. Email: zeyu.li@uci.edu. I thank Damon Clark, David Neumark, Jan Brueckner, Matt Freedman, and Miguel Zerecero for their advice and feedback. I also thank all seminar participants at UCI and AMIE for their comments and suggestions. The conclusions in this paper are my own.

Zoning and place-based policies[1], such as enterprise zones, may create losses in efficiency if they incentivize firms and workers to migrate from high-density areas to less dense ones. On the contrary, policies that lead to a reduction in density, such as building height or floor-area ratio restrictions can create significant welfare losses (Bertaud and Brueckner, 2005). In developing countries, removing obstacles that prevent migrant workers from accessing big cities can lead to more efficient resource allocation and drive economic growth. (Duranton, 2008). Density acts as a critical conduit through which diverse policies exert their effects.

Individual workers and firms in big cities worldwide often seem to have higher productivity, as indicated by higher average wages and TFPs. But a problem behind this relationship arises as both density and productivity can be a result of the other, or both are outcomes of something else. This can be further decomposed into two arguments. The first argument concerns the omitted variable bias. Productivity and agglomeration are likely to be associated with unobserved local characteristics. For instance, cities in warmer climates can attract more residents and, at the same time, be more productive compared to cities in colder regions. The second argument relates to the simultaneity bias due to the sorting of workers and firms with high productivity into big cities, which makes agglomeration a result of self-selection. Combes et al. (2008) use French worker data to estimate a model of wage determination across local labor markets and find that individual skills and local employment density are both significant factors in explaining spatial wage disparities.

This paper focuses on addressing the issues mentioned above and examines the causal relationship between urban density and productivity. I use a policy variation related to urban population density from a nationwide daylight building-spacing regulation in China starting in 1994. The regulation is designed to ensure sufficient distance between residential buildings so that each south-facing apartment receives certain hours of direct sunlight during the day, and it constitutes a novel instrument for density. I employ a novel approach to estimate the benefit of population density by instrumenting density with a policy-implied spacing factor (SF), defined as the ratio of building spacing to height. The construction of the spacing factor incorporates the SF guideline provided in the policy text for 43 prominent Chinese cities, with the assumption that all buildings are uniformly sized, oriented east-west, and aligned north-south[2]. I extend this methodology to the remaining cities in my sample, thus providing SF estimates for all cities.

The rationales for exploiting the variation from the above regulation are as follows. First, the distance between buildings is expected to have a significant impact on urban density. As seen

[1]Place-based policies mainly refer to government efforts to improve the economic performance of an area within their jurisdiction. These policies could be focused on underperforming and good-performing areas, and typically take the form of increased job opportunities and higher wages. (Neumark and Simpson, 2015)

[2]This assumption is made to simplify the calculation process.

in Table 1, the spacing factor[3] in Harbin is twice as large as the spacing factor in Guangzhou, which implies that a residential area in Harbin would need to have twice the building spacing of a residential area with similar building heights in Guangzhou. Second, the spacing regulation also varies by the city population and climate zones. Cities that are more populated and in a warmer climate are generally allowed to build denser residential areas. Combining the two facts above, variation of the daylight spacing regulation generally comes from three sources: latitude (affecting sun angles), population, and climate zones. The variation in latitude is considered endogenous because it is correlated with other confounders that may affect productivity, such as temperature. This paper focuses on the latter two variations at the cutoffs of climate zone and population, assuming other variables are continuous across the cutoffs. Unlike a regression discontinuity design exploiting differences at a single cutoff, my instrumental variable approach employs exogenous variations in both the population threshold and zone boundaries.

By combining data recording a rich set of city-level information from 199 prefecture-level cities over 15 years, I find that the urban population density of Chinese cities has a positive impact on the average wage, with an elasticity of 0.4. Density is not estimated to positively affect GDP per capita, but increases the output per worker for the service sector. The benefits of density could be attributed to knowledge spillovers and innovation, as the results also demonstrate a positive elasticity of the number of patents granted per person with respect to density, estimated at 1.6. Another crucial mechanism identified in this study is that density facilitates the transition from an agricultural or manufacturing-oriented economy to a service-oriented economy. This is evidenced by the positive effect of density on both the output share and output per capita of service industries.

This paper contributes to existing research on agglomeration economies by exploiting a nationwide building spacing policy to address the endogeneity problem when estimating the productivity effect of urban density. Some fundamental work on the relationship between population density and productivity does not address this problem properly (Combes et al., 2012). Henderson (2003) uses plant-level data and applies a fixed-effect model to eliminate the effects of the unobservables. Several other papers find instruments for density, such as historical measures of density (Ciccone and Hall, 1996), land fertility (Combes et al., 2010), and land suitability for tall buildings (Combes et al., 2012). Such instruments could still affect productivity through channels other than density. My paper is one of the very first to use a nationwide mandatory regulation in China to build a quasi-experimental design. Compared to similar papers that also use quasi-experimental designs for identification (Greenstone et al., 2010), this paper exploits a nationwide building regulation that alters the level and growth of population density, with the following advantages. First, my

[3] The spacing factor is defined as the minimum ratio of the building spacing to the building height to meet the daylight hour requirement. The policy text in GB50180-93 gives reference standards for each city's spacing factor.

instrument comes from central government regulation and therefore does not depend on local unobserved characteristics. Second, based on data from 199 cities in China, my study provides more representative and generalizable findings than other studies that focus on only one or a few cities. Regarding the investigation of the sources of density's benefits, my paper offers fresh evidence supporting the hypothesis of agglomeration leading to knowledge spillovers and innovation (Carlino et al., 2007; Carlino and Kerr, 2015; Moretti, 2021).

This paper also provides insights into the study of agglomeration economies in China and other developing economies. Although migration and agglomeration are growing rapidly in these regions, the impact of agglomeration is less frequently addressed compared to developed countries. Existing work on China focuses predominantly on the effect of density on economic growth (Zhang and Liu, 2008; Liu, 2014; Zhang, 2018) rather than static productivity measures. There is little consensus on the wage effect. Combes et al. (2015) use a set of historical and other instruments and find that the elasticity of the wage with respect to density in China is 0.11, which is higher than the average of 0.04 in the existing literature (Ahlfeldt and Pietrostefani, 2019). Lu et al. (2015), by contrast, suggest a negative association between density and the average wage in China. A recent meta-analysis by Grover et al. (2023) suggests that there is no statistically significant distinction in agglomeration economies between developing and developed nations. By discovering a novel instrument for density, my paper not only finds a positive wage effect of density which is larger than those previously estimated, but also reports other results such as density fostering the shift from a manufacturing-based economy to a service-based economy, which provides important policy implications for developing countries.

Compared to other research that also exploits variation from China's daylight spacing policy, this paper addresses significant gaps in the empirical design. Since the policy intensity varies between latitudes, Zhang (2018) instruments a city's density with its latitude and finds a positive effect of density on economic growth. Zhang (2019) instead employs a difference-in-difference analysis utilizing data before and after the policy implementation and finds various effects on information diffusion. Both the above approaches depend heavily on variations in density driven by latitude, which is subject to the same limitations as other strategies (e.g. instruments such as land fertility) due to potential correlation with unobserved productivity advantages correlated with latitude. My empirical specification uses a policy-implied spacing factor as an instrument and controls for latitude, and thus uses variation only by climate zones and historical population.

The rest of the paper is organized as follows. Section 2 illustrates the institutional background. Section 3 further introduces the empirical strategy I use. Section 4 introduces data sources and variable definitions. Section 5 presents the main estimation results. Lastly, Section 6 concludes the paper.

2 Background

2.1 The daylight spacing policy in China

Multi-story buildings are the primary form of housing for urban residents in China, with 93.8% of urban residents living in condominiums in multi-story buildings in 2020 according to the National Bureau of Statistics [4]. However, densely packed buildings can result in lower floors receiving less natural sunlight due to obstruction from adjacent buildings. To address this issue, the Chinese government first introduced regulations in 1980 through the Interim Provisions on Urban Planning Quotas, which stipulated that the spacing between buildings should, in principle, provide at least one hour of daylight on the ground floor of residential buildings on the winter solstice. Further sunshine standards were established in the General Rules for Civil Building Design in 1987, which specified requirements for buildings such as dormitories, childcare centers, elderly and disabled housing, hospitals, and nursing homes.

However, these regulations were found to have certain problems in the late 1980s and early 1990s due to the rapid development of large cities, particularly in northeastern regions where sunlight was heavily blocked to the 3rd and 4th floors in winter. In response, the national mandatory standard Urban Residential Planning and Design District Code (GB50180-93) was implemented in February 1994 (with a minor revision in 2002) to address these challenges. This standard incorporated lessons from foreign standards and adapted them to the specific situation in China. It sets out three levels of daylight requirements based on climate zones and urban population, specifying the minimum number of daylight hours and effective daylight time zone for each level.

The standard requires that large cities (with a population of over 500,000) in climate zones I, II, III, and VII receive daylight during the *Dahan* day [5] for no less than 2 hours, while small and midsize cities (with a population[6] of less than 500,000 [7]) in these zones, as well as large cities in zone IV, must receive effective daylight during *Dahan* for no less than 3 hours. Cities in zones V and VI, as well as small and midsize cities in zone IV, must receive effective daylight time of

[4] Summarized by *mikuang.com*, see https://caifuhao.eastmoney.com/news/20220622070234015231940

[5] The traditional Chinese calendar divides a year into 24 solar terms. *Dahan* refers to the first day of the 24th solar term when the Sun is exactly at the celestial longitude of 300° (Wikipedia).

[6] The policy does not clearly state if the daylight hour requirement adjusts based on population changes. Additionally, there's no evidence to suggest that cities routinely update their guidelines in response to population shifts. Thus, I assume that this population threshold remains at its 1994 value.

[7] The policy text does not make it clear whether this is fixed to the population at the time when the policy takes effect, or changing over time. I do not find any evidence in the local implementation of the national policy that the local standards adjust with population changes.

the winter solstice for not less than 1 hour. Table 1 summarizes the criteria for determining the sunshine hour requirements for different cities.

Table 1: Detailed Daylight Hour Regulations

Climate Zone	I, II, III, VII		IV		V, VI
Urban Population	$\geq 500,000$	<500,000	$\geq 500,000$	<500,000	No restrictions
Reference Date	Dahan (around Jan 21)			Winter solstice (around Dec 23)	
Daylight Hours	≥ 2	≥ 3		≥ 1	

Notes: This table provides a summary of the specific minimum daylight hours and reference dates that each city must adhere to. The first two rows list the determinants of the standards: climate zone and urban population. Following each pair of climate zone and population, the third to the fifth row state the exact number of daylight hours and the reference date of a year. For example, Beijing is a city located in climate zone II and has a population greater than 500,000. According to the second column from the left, the standard for Beijing is to have at least 2+ hours of daylight on the *Dahan* day (around January 21st). Source: GB50180-93, issued by the Ministry of Housing, Urban Rural Development of China.

2.2 Climate zones

China has a system of building climate zoning that divides the country into different zones based on factors such as temperature, humidity, wind speed, and solar radiation. Appendix Figure A1 is a map showing these zones. The climate zones are used to determine the appropriate design and construction methods for buildings, as well as the types of materials that should be used. In this paper, I mainly focus on the difference across the seven major climate zones, as they are an important source of the daylight hour variation (as shown in Table 1). Although the regulation does not explicitly state the reason why the requirement for daylight hours differs by zone, one can infer from Table 1 and Appendix Figure A1 that the number of daylight hours required is larger when a zone experiences colder climates during winter.

2.3 The daylight spacing factor

In addition to setting the minimum daylight hours, the spacing regulations in GB50180-93 also provide examples of the ratio of the minimum building spacing to the building height to meet the daylight hour requirements, named the spacing factor. As shown in Figure 1, the spacing factor is defined as the ratio of building spacing (D) to the height from the ground-floor window to the roof (H_1). It is calculated based on a simplified model where uniform strip buildings extend east-west and have a height of 18.18 meters (59.65 feet). The regulation gives examples of spacing factors for 43 cities across the country.

When determining the spacing factor for the 43 example cities, policymakers implement a two-

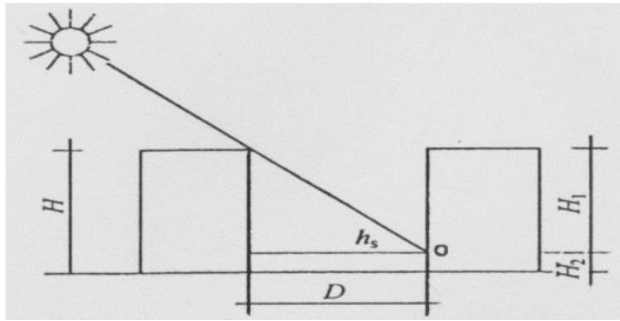

Figure 1: The spacing factor, defined as D/H_1

step procedure. First, they calculate three hypothetical spacing factors for each city based on each of the three daylight hour requirements defined in Table 1 (2 hours on *Dahan*, 3 hours on *Dahan*, and 1 hour on the Winter Solstice) using a sunlight shadow model. Subsequently, with the help of Table 1, which assigns the applicable daylight hour requirement to each city based on climate zone and urban population, they select the appropriate spacing factor from the three calculated ones.

Table 2 illustrates this process for four major cities, listing their corresponding spacing factors. Each row represents a city, with the last three columns showing the hypothetical spacing factors calculated for each daylight hour requirement. The spacing factor in bold aligns with the actual daylight hour requirement for the city, and is thus the true spacing factor.

The within-column variation of spacing factors in Table 2 comes from the variation in latitude. This type of variation is problematic, as latitude may be correlated with other natural conditions that affect productivity. Alternatively, the policy variation used in this paper comes from the remaining two sources of variation in minimum daylight hours: climate zones and the urban population threshold. Constructing this spacing factor measure helps me get a single instrument that incorporates variations in latitude, climate zones, and population. Once controlling for latitude, historical population, and climate zones, the variation in the spacing factor can be attributed to the differential policy intensity at the boundaries of the climate zones, as well as the population threshold.

With the policy guideline of *GB50180-93* offering examples for 43 cities (leading to a total of 129 spacing factors), I am able to predict spacing factors for all other cities in my sample, because holding minimum daylight hours constant, the spacing factor provided in the examples is just a function of latitude. Using these examples to establish my spacing factors has two main advantages. First, it prevents me from having to replicate a sunlight shadow model using complex geographical and construction engineering techniques. Second, because the examples stem from policy guidelines, my extended measure closely mirrors the parameters that local authorities refer to when implementing the policy.

Table 2: The policy-implied spacing factor: examples

		Winter solstice	Dahan (Jan 21)	
City	Latitude	1 hour	2 hours	3 hours
Harbin	45°45' N	2.46	**2.15**	2.24
Tianjin	39°06' N	1.80	**1.61**	1.68
Kunming	25°02' N	**1.06**	0.98	1.03
Guangzhou	23°08' N	0.99	0.92	**0.97**

Notes: This table presents examples of 4 cities and their corresponding spacing factors under each daylight hour requirement (displayed in each column). The specific daylight hour requirement, as shown in Table 1, is determined by the climate zone of the city and its urban population. The **bold** numbers indicate the spacing factors actually applied to the city based on the daylight hour requirement.

2.4 Understanding the definition of a *city* in China

In China, administrative divisions are categorized into several levels, among which *city* is one of them. Nevertheless, the meaning of *city* in the modern Chinese language usually refers to a *prefecture* in English, which includes the central area of the city, suburban area, and surrounding counties. This differs from how the term *city* is used in some other countries such as the United States, where it varies by location and may only refer to the central business district (CBD) but not the surrounding suburban or rural areas. Therefore, when using data on *cities* in China, it is essential to be mindful of this difference, as it may not align with the way the term is used elsewhere.

There are three main levels of cities: directly-administered cities, prefecture-level cities, and county-level cities. Directly-administered cities, also known as municipalities, are the highest level of cities in China, directly under the jurisdiction of the central government. They have administrative power equivalent to provinces and are the only cities in China that are not part of a province. Prefecture-level cities are the second-highest level of city in China, directly under the jurisdiction of a province. They are larger and more populous than county-level cities, which are the lowest level of cities in China and are under the jurisdiction of a prefecture-level city. Each prefecture is further divided into several counties.

To avoid confusions, the data used in this paper is constructed from 199 prefecture-level cities in China, and the measurement is restricted to the urban area. Therefore I use the term *city* as it is commonly used in English, while the corresponding administrative region that includes the city is referred to as a *prefecture*.

3 Empirical Strategies

3.1 Urban population density and productivity

My empirical analysis starts from a naïve regression of various productivity measures on a city's density. Following the work by Glaeser et al. (1995), I use the following regression equation:

$$y_{ipt} = \alpha + \beta_1 density_{ipt} + X_{ipt}\eta + \mu_p + \phi_t + \epsilon_{ipt} \tag{3.1}$$

where y_{ipt} is one of the productivity measures of city i, located in province p, and observed in year t. *density* is the logarithm of the *naive* population density of city i, defined as the urban population divided by urban built area (in km^2). X_{ipt} includes a set of control variables that may affect both density and productivity. μ_p and ϕ_t denote province and time fixed effects accounting for heterogeneity in both dimensions. ϵ_{ipt} is the error term.

An apparent challenge when estimating Equation (3.1) is the effects of unobserved city characteristics that correlate with both density and productivity, in which case the estimate ($\hat{\beta_1}$) would be biased. A classic example is some form of productivity advantage in ϵ that positively affects both density and productivity, in which case the estimate will be biased upwards. Conversely, if a proactive city government artificially lowers density by investing in urban expansion projects, we might see a $\hat{\beta_1}$ that is underestimated. Additionally, simultaneity bias arises when high-productivity workers are more likely to gravitate towards larger cities, causing a concurrent rise in both density and productivity. To mitigate these biases, researchers have developed various instrumental variables, such as land fertility, land suitability for tall buildings, and historical measures of density. However, these instruments may not be entirely convincing, as they are all associated with local natural conditions. This raises concerns about the validity of these instruments, as they might be correlated with other unobserved factors that can influence productivity, leading to biased estimates. Alternatively, this paper addresses this issue by exploiting exogenous policy variations in density from the daylight building spacing regulation in China.

3.2 The IV approach

To accurately estimate Equation (3.1) and address both the omitted variable bias and sorting, it is imperative to leverage external variation in population density via use of an instrument. This entails finding a variable that exogenously impacts urban density without affecting other factors tied to the city's economic productivity.

The regulation on the spacing between buildings in China leads to variations in the distance between buildings in different cities and hence, the population density. Specifically, the regulation states that the residential building spacing should be sufficient for a minimum daylight hours requirement (as listed in Table 1). As illustrated in the table, the intensity of the spacing policy, as measured by the specific number of daylight hours required, depends on two local conditions, the climate zone in which the city is located and an urban population threshold of 500,000. The regulated daylight hours vary significantly around the boundaries of climate zones, as well as at the population thresholds.

This paper instruments the endogenous population density with the policy-implied building spacing factor, defined as the minimum ratio of building spacing to building height to guarantee the number of daylight hours. The spacing factor I use is drawn and extended from the examples for 43 cities in the policy document (GB50180-93) and is therefore assumed to be uncorrelated with local characteristics other than geographic location and population.

The exogenous variation I utilize comes from the discontinuity in the spacing factor at the boundaries (cutoffs) of climate zones and historical population[8]. Supposing that cities were subject to the same minimum daylight hour requirement, the corresponding spacing factors are solely determined by the trajectory of the sun, which varies continuously by the city's latitude. However, as the minimum daylight hour requirement (shown by different columns of Table 1) is a non-linear function of the climate zone and population, the implied spacing factor is then a non-continuous function of the latitude, the climate zone and population.

Compared to other measures of building spacing, there are several advantages of using the implied spacing factor as an instrument. First, it is a simple approach to quantify the intensity of daylight spacing regulation, given the fact that the required spacing is determined by multiple factors (latitude, climate zone, and urban population). Second, the implied spacing factor is uncorrelated with local unobserved characteristics, such as the actual implementation of the regulation, which is likely to be correlated with local characteristics such as the government's capacity.

Similar to a regression discontinuity design[9], I implicitly exploit variations only from the discontinuous variation from climate zone boundaries and the population cutoff, with the latitude, historical population, and region fixed effects controlled on the right-hand side. The goal of controlling these is to rule out continuous variation and leave only the discontinuity part to identify the causal effect of density. The first stage of my IV-2SLS estimation of population density on

[8] I do not find any evidence suggesting that the daylight hour requirement changes as the local population grows. When calculating the spacing factor, I use the urban population in 1994 to determine the population threshold.

[9] A regression discontinuity design is not preferable in this context for the following reasons: (1) There are two dimensions of discontinuity, which include both the population cutoff and the climate zone boundary. (2) Some borders do not imply a policy difference, such as the borders between zones I, II, III, and IV. On the other hand, borders that have a policy difference might not have many cities close to them, as zones V and VI are the least populous areas of the country.

productivity is specified as:

$$density_{ipt} = \delta_0 + \delta_1 SpacingFactor_{ip} + \delta_2 latitude_{ip} + X_{ipt}\xi + \lambda_p + \tau_t + e_{ipt} \tag{3.2}$$

As in Equation 3.1, $density_{ipt}$ denotes the log of the urban population density in city i, located in province p, and observed in year t. The variable $SpacingFactor_{ip}$ is the spacing factor introduced by the daylight spacing policy, defined as the minimum spacing-to-height ratio to satisfy the daylight hour requirement. The city's *latitude* is included to ensure the conditional exogeneity of the spacing factor, given that it geographically affects the SF but may be correlated with unobserved productivity advantages. X_{ipt} is a set of control variables. λ_p, the province fixed effect, is specified to eliminate the time-invariant differences across provinces. τ_t is the year fixed effect.

4 Data and Variables

4.1 Data sources

The data used in this paper are drawn from the following sources: (1) Measures of city area and population come from the *China Urban Construction Statistical Yearbook*. (2) City-level average wages, GDP, and other outcomes of interest are from the *China City Statistical Yearbook*. (3) Patent information at the prefecture level is from the published information on the website of the China National Intellectual Property Administration (CNIPA).

4.1.1 China Urban Construction Statistical Yearbook

The China Urban Construction Statistical Yearbook (CUCSY) has been published annually since 2002 by the Ministry of Housing, Urban, and Rural Development. The data record detailed information about the urban population and built-up area, from which I can calculate the naive density simply by dividing the population by land area (Duranton and Puga, 2020). An advantage of using CUCSY is its clear distinction between city-level and prefecture-level measurements, which avoids confusion commonly found in other datasets labeled as city-level in China. As noted by Combes et al. (2015), many of these datasets are actually at the prefecture level, which results in information being only available for the overall prefecture rather than the urban (metropolitan) area. The CUCSY, designed to track the development of urban constructions, provides a clearer distinction between the two levels. I use the yearbooks from 2002 to 2019 to construct a panel dataset, which serves as the baseline time frame for my analysis.

4.1.2 China City Statistical Yearbook

The productivity measures and other outcome variables used in this paper come from the *China City Statistical Yearbook* (CCSY) published by the National Bureau of Statistics. It provides detailed information on city-level GDP, employment, average wages, and output by sector, which serve as the key outcomes of interest in this paper. One advantageous aspect of the yearbook is the inclusion of two distinct versions of prefecture-level data: the entirety of the prefecture and exclusively the urban (metropolitan) region. The latter precisely aligns with the scope of this paper and correlates with our data extracted from the *China Urban Construction Statistical Yearbook*.

4.2 Variables

Based on data on urban population and different measures of city area, this paper calculates a naive density (Duranton and Puga, 2020) for each city by dividing the population (in 10,000) size by the area of built land (in km^2) in the city, and taking the logarithm. The density differs from the official statistic published in the CCSY because I calculate the density based on the built land exclusively, rather than the inclusion of undeveloped land. Such an approach provides a more precise gauge of the concentration of economic pursuits within a given urban region.

This paper uses the city-level average wage as the core productivity measure, where the wage is the most commonly used measure for productivity. In contrast to per-capita GDP, the wage data in China is a more meticulous indicator of economic activity, as it derives from social security records. The reliability of GDP measurements in China has been a topic of scrutiny, owing to its inability to capture the informal sector output and data manipulation due to inter-regional competition among local officials in achieving GDP growth targets. To delve deeper into the underlying mechanisms behind the benefits of density, this paper estimates the effect of density on the output shares and labor shares of both the secondary (production) and tertiary (service) sectors, as well as the per-worker output by sectors. The objective is to determine whether density has a significant impact on a city's ability to upgrade its industrial structure, particularly in terms of transitioning from a manufacturing-based economy to one that is more service-oriented. In addition, this paper examines the impact of density on innovation, represented by the number of patents granted per 10,000 population[10].

The control variables in both X's in equations (3.1) and (3.2) can be categorized into two types. First, I control for a set of city's geographic conditions (slope, elevation, distance to the coastline,

[10] The data on prefecture-level patents granted are published by the *China National Intellectual Property Administration (CNIPA)*. While the scope of prefecture-level patent data may differ from that of city-level data, it can serve as a useful proxy for measuring innovation at the city level. This is because a significant portion of scientific and educational activities occur within cities.

and latitude) to capture the productivity advantage that may affect both density and productivity. The latitude, in particular, also helps guarantee the conditional exogeneity of my instrument in the first stage (as in equation 3.2). Most of these variables are calculated using GIS. Second, to account for accounts for local government's ability to expand urban sprawl, I include the total land area of the city (including undeveloped land) and the area of the prefecture as control variables, which come from the yearbooks.

Table 3 summarizes the descriptive statistics of the core variables used in this paper.

Table 3: Descriptive statistics

	Mean	SD	Min.	Max.	Observations
Explanarory Variable					
Urban density (log)	8.997	0.315	8.026	10.133	2985
Instrument					
Spacing factor	1.349	0.34	0.785	2.279	2985
Productivity measures					
Average wage (log)	10.578	0.534	7.586	12.062	2985
GDP per capita (log)	10.656	0.697	8.327	15.675	2984
Sector composition					
Employment share: primary	1.271	4.391	0	54.64	2985
Employment share: secondary	46.129	14.162	7.4	85.7	2985
Employment share: tertiary	52.696	13.914	14.2	100	2985
Output share: primary	6.304	6.411	0.03	60.47	2985
Output share: secondary	48.553	12.293	8.1	90.4	2985
Output share: tertiary	45.132	11.611	8.6	83.5	2985
Output per capita: primary (log)	10.303	1.621	4.733	14.892	2801
Output per capita: secondary (log)	7.958	0.651	5.65	10.181	2984
Output per capita: tertiary (log)	7.738	0.609	5.928	9.451	2984
Innovation					
Total patents granted per 10,000 population	6.546	16.937	0.011	307.119	2786
Invention patents granted per 10,000 population	0.94	3.368	0.002	48.13	2769
Design patents granted per 10,000 population	2.091	6.465	0.002	100.066	2767
Utility patents granted per 10,000 population	3.535	8.576	0.005	165.468	2786
Geographic controls					
Latitude	32.546	6.723	18.253	47.728	2985
Distance to coastline	0.437	0.337	0.002	1.435	2985
Average slope	10.984	5.691	1.592	27.139	2985
Average elevation (km^2)	0.486	0.557	0.008	3.128	2985
Total area including undeveloped land (10000 km^2)	0.24	0.306	0.008	4.326	2985
Total area of prefecture (10000 km$\hat{2}$)	1.521	1.239	0.166	9.066	2985

5 Results

5.1 First-stage results

Table 4 shows the first-stage estimates from regressing the urban density on the spacing factor. I start with column (1) with only the latitude controlled. The estimate is negative and statistically significant, which is expected as urban density should decrease when building spacing increases.

Table 4: First stage estimates of population density on the spacing factor

	(1)	(2)	(3)	(4)	(5)	(6)
Spacing factor	-0.774***	-2.549***	-0.746**	-2.494***	-0.757**	-2.476***
	(0.287)	(0.516)	(0.294)	(0.529)	(0.302)	(0.522)
Latitude	0.044***	0.108***	0.041***	0.105***	0.041***	0.102***
	(0.014)	(0.024)	(0.015)	(0.025)	(0.015)	(0.025)
Distance to coastline (km)			0.032	-0.054	0.038	-0.003
			(0.072)	(0.144)	(0.074)	(0.145)
Average slope			-0.002	-0.002	-0.002	-0.005
			(0.005)	(0.005)	(0.005)	(0.006)
Average elevation (km)			0.045	-0.036	0.042	-0.026
			(0.050)	(0.074)	(0.050)	(0.072)
Total area including undeveloped land ($10000km^2$)					-0.057	-0.202***
					(0.080)	(0.061)
Area of prefecture ($10000km^2$)					0.009	0.029*
					(0.017)	(0.016)
Province FE	-	Yes	-	Yes	-	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	8.997	8.997	8.997	8.997	8.997	8.997
First-stage F-stat	7.265	24.409	6.424	22.192	6.287	22.498
R-squared	0.214	0.447	0.221	0.449	0.224	0.464
Observations	2985	2985	2985	2985	2985	2985

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). The dependent variable is the logarithm of the urban population density (in $10{,}000/km^2$). The *Spacing factor* is defined as the ratio of building distance to building height and is calculated based on same method used in the daylight regulation policy . Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.

Column (2) adds province fixed effects to eliminate any potential similarities within a given province. This leads to a substantial increase in the magnitude of the estimate, which triples in size. Additionally, the R^2 value experiences a sharp increase, indicating that a large share of the variations in density is captured by the fixed effects. Columns (3) and (4) repeat the same exercises as in the previous columns but add more geographical controls, and yield similar point estimates and standard errors. In columns (5) and (6), I further control the local government's ability to expand the urban area by converting undeveloped land. My most saturated specification in column (6) shows that the point estimate for the spacing factor is -2.48. Specifically, we find that a 0.01 increase in the spacing factor leads to a reduction in density of approximately 2.5%. An F-statistic value of 22.50 not only indicates significance at the 1% level but also effectively dispels concerns of a weak instrument problem.

While the estimate in column (6) of Table 4 may seem surprisingly large, it is important to note that the variation in spacing factor is relatively small for locations at the same latitude. For instance, based on the data in Table 2, the average difference in spacing factor between the 2-hour and 3-hour requirements on *Dahan* is 0.06. Therefore, relaxing the minimum daylight hours on *Dahan* from 3 hours to 2 hours is estimated to lead to an increase in urban density of approximately 15%.

Despite the promising results from the first stage regressions, there are still some potential concerns that should be addressed. One such concern is the possibility of unobserved location characteristics that could influence the spacing factor, even conditional on province fixed effects and other geographic controls. For example, the fixed effects may not fully capture the endogenous variation of density across different locations. To address this concern, I residualize the spacing factor by climate zones, latitude, and population threshold, using regressions that incorporate four different functional forms of latitude: linear, quadratic, cubic, and exponential. Appendix Figure A5 displays maps of the residualized spacing factor using each of these functional forms, and no noticeable patterns of location correlations are apparent in any of these maps. These results suggest that the spacing factor is not likely to be driven by any unobserved location patterns. Additionally, this study re-runs the first-stage regressions while controlling for alternative fixed effects at both the administrative-zone and climate-zone levels, as shown in Table A2. I start with no region fixed effects, as in column (1), then add the climate zone fixed effects (as in column (2)), administrative zone fixed effects, and province fixed effects. The adjusted R-squared value increases consistently from column (1) to (4), indicating that my preferred specification, with provincial fixed effects, captures the largest unobserved variations in density across all four specifications.

A second potential concern is that the estimated correlation between the spacing factor and density may be coincidental (i.e. confounded with some unobservable). I conduct a falsification test and check the F-statistics of the estimate on spacing factor computed based on randomly assigned climate zones and population thresholds. The rationale behind this approach is that if the correlation between the spacing factor and density is purely coincidental, then randomizing climate zones and the population threshold should not significantly weaken the significance of the first-stage estimate. I start with creating 9 placebo population cutoffs from 100,000 to 900,000 at intervals of 100,000. Then I create 100 randomly assign climate zones for each city. Interacting with the placebos I get from the two steps above provides me with 900 placebo spacing factor sets. I then re-run the first stage regression using these placebo spacing factors and plot the cumulative distribution of the F-statistic for the estimates for the placebo spacing factor. Appendix Figure A5 presents the resulting graphs. Each of the 4 panels shows a different specification (linear and quadratic latitude, with and without covariates). The results of all four model specifications indicate that the F-statistic of the true instrument is greater than that of 99% of the placebo instruments. This suggests that the observed variation in population density can be confidently attributed to the differences in policy across climate zones and historical population factors as of 1994.

Additionally, to check whether the 1994 population cutoff introduces endogenous variation into my estimation[11], and whether the estimate is sensitive to the functional form of latitude, I first re-estimated the first-stage regression using three different functional forms of latitude, as shown in Panel A of Appendix Table A1. Secondly, I repeat the same regressions but add a dummy for whether the urban population in 1994 exceeds the threshold (500,000). The goal is to fully eliminate the variation from population. The results in Panel A indicate that the estimated coefficient for the spacing factor is robust to all three alternative functional forms of latitude, suggesting that my first-stage specification effectively captures the exogenous variation of the spacing factor while leaving out the endogenous variation driven by latitude. Comparing the estimates in Panels A and B, removing the historical population variation does not substantially alter the size of the estimates, though it does reduce their precision by increasing the standard errors. Generally, using variation from the historical population does not affect my first-stage estimate but does help to strengthen the power of my instrument.

Lastly, considering my exploration of a policy impacting residential building spacing, questions might arise regarding its influence on the production side. A significant number of Chinese cities carry the legacy of Soviet-style urban planning, characterized by an intermingling of commercial, residential, and industrial structures. As a result, any alteration in residential building spacing, driven by the daylight spacing policy, indirectly affects commercial building density. This, in turn, has ripple effects on various non-residential metrics, notably employment and innovation densities.

5.2 The effect on the average wage

My main results are provided in Table 5, which presents both OLS and 2SLS regressions of the average wage on urban density. I use both the log of the average wage (shown in Panel A) and its level (shown in Panel B). In columns (1)-(4) of Table 5, the OLS estimates of the effect of density on wages are negative and statistically insignificant. However, as I add more control variables, the point estimate goes toward zero and the standard errors decrease. The main results are displayed in columns (5) and (6) of Panel A, where the 2SLS estimates show a strong and positive impact of density on wages, with an estimated elasticity of approximately 0.4. Combined with the first-

[11]The IV estimation omits the urban population in 1994 as a control due to concerns that including population may influence the interpretation of the regression as in Equation 3.1.

stage estimates, and taking into account the average spacing factors under different daylight hour requirements, relaxing the minimum daylight hours by one is expected to result in an increase in the average by approximately 5.9%. The estimate is significantly larger in magnitude than many existing findings, particularly those based on developed economies. It is important to note that urbanization rates in China are still relatively low compared to other countries. Therefore, it is expected that the marginal benefits of agglomeration would be higher in this context.

A comparison of the 2SLS estimates with the OLS estimates suggests that the latter is biased towards zero as a result of omitted variables or simultaneity issues. This could occur as larger cities in China typically impose stricter restrictions on migrant workers from less urbanized areas, leading to productivity losses (Au and Henderson, 2006a,b). Alternatively, local governments may invest in ineffective place-based policies, leading to the development of low-density urban areas that artificially inflate wages in the short term due to an increase in investment.

To obtain a more straightforward idea of the impact of density on wages, I repeat the same exercise as in Panel A, but use the average wage level instead. The estimates, as presented in Panel B of Table 5, are similar in direction and precision to the previous findings. According to the most saturated specification presented in column (6), doubling the urban density is associated with an increase of 15,096.96 yuan (1,849.62 USD[12]) in the average annual wage. The annual wage increases by 4,769.13 yuan in response to a one-standard-deviation rise in density.

In addition to the results I get on wages, it is important to look at the effect of density on the GDP per capita, which measures productivity from the output side. As displayed in Appendix Table A3, I re-run the same set of estimations on city-level GDP per capita. The results suggest that all four specifications using OLS yield negative estimates of density on productivity. IV estimates are positive but not statistically significant. Although there are reasons to believe that GDP data are of low quality in China[13], the results do suggest a pattern that OLS regressions underestimate the true effect of density, which is consistent with my findings in Table 5.

5.3 Sources of the productivity benefit

While most existing literature suggests positive impacts of density on productivity, the specific sources of these benefits and the mechanisms through which they influence productivity have been comparatively less explored. This paper focuses on two potential sources of productivity advantages derived from density: innovation and transitions in sector composition

[12]Based on the historical exchange rate between US Dollar and Chinese Yuan on December 31, 2005.

[13]Rawski (2001) identifies an overstatement in China's GDP data and explores the possible reasons behind this phenomenon. Among the contributing factors are the government's incentive to exaggerate economic performance and the double-counting of output at multiple levels. Similar studies also raise concerns about the exclusion of output from informal sectors in GDP calculations.

Table 5: 2SLS estimates on the average wage

Panel A: Dependent variable: log average wage						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	2SLS	2SLS
Log urban density	-0.032	-0.013	-0.010	-0.012	0.403**	0.393**
	(0.068)	(0.047)	(0.044)	(0.039)	(0.176)	(0.177)
Latitude			-0.003	-0.000	0.003	0.007
			(0.008)	(0.008)	(0.009)	(0.009)
Distance to coastline (km)			-0.100	-0.119	-0.125	-0.161
			(0.088)	(0.088)	(0.118)	(0.121)
Average slope			-0.007**	-0.006	-0.006	-0.004
			(0.004)	(0.004)	(0.005)	(0.005)
Average elevation (km)			0.129***	0.125***	0.173***	0.164***
			(0.047)	(0.046)	(0.063)	(0.061)
Total area including undeveloped land (10000 km^2)				-0.033		0.055
				(0.115)		(0.108)
Area of prefecture (10000 km^2)				-0.012		-0.022
				(0.011)		(0.013)
Province FE	-	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	10.578	10.578	10.578	10.578	10.578	10.578
R-squared	0.826	0.901	0.903	0.904		
First-stage F-stat					22.192	22.498
Observations	2985	2985	2985	2985	2985	2985
Panel B: Dependent variable: average wage						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	2SLS	2SLS
Log urban density	-401.64	-465.02	-442.01	38.35	14738.63**	15096.96**
	(3721.70)	(1845.26)	(1755.85)	(1714.66)	(6488.78)	(6799.01)
Latitude			-244.12	-192.72	-0.67	65.16
			(291.43)	(302.94)	(352.92)	(343.82)
Distance to coastline (km)			-4068.17	-5060.70	-4986.15	-6613.76
			(3384.91)	(3519.99)	(4474.07)	(4733.00)
Average slope			-289.56**	-235.81	-256.23	-169.72
			(141.28)	(147.91)	(187.97)	(204.59)
Average elevation (km)			4344.01**	4185.93**	5960.86**	5648.94**
			(1749.07)	(1679.16)	(2365.73)	(2264.18)
Total area including undeveloped land (10000 km^2)				4017.63		7291.31
				(4784.36)		(4502.76)
Area of prefecture (10000 km^2)				-546.87		-905.96*
				(446.58)		(514.02)
Province FE	-	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	44787.17	44787.17	44787.17	44787.17	44787.17	44787.17
R-squared	0.780	0.883	0.885	0.886		
First-stage F-stat					22.192	22.498
Observations	2985	2985	2985	2985	2985	2985

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). The dependent variable is the log of the average wage (in Panel A) and the level of the average (in Panel B). In both panels, OLS regression results are shown in columns (1)-(4), and 2SLS results are shown in columns (5) and (6). Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.

5.3.1 Innovation

In agglomeration economics, theoretical models consider the concentration of innovation as a critical mechanism contributing to enhanced productivity. To examine the first part of this argument—whether urban density stimulates higher innovative activities—using my instrumental variable design, I estimate the effect of density on the prefecture-level average number of patents granted. I define the outcome as the number of patents divided by the population (in units of 10,000). There are three types of patents in China: invention patents, utility patents, and design patents. In general, invention patents are granted for the design of novel inventions that have not been created before. In contrast, the latter two categories concentrate on enhancing existing products or services, either by improving their usability or appearance.

As displayed in Panel A of Table 6, urban density exhibits positive effects on innovative activities, as indicated by the log number of patents granted. Column (1) presents the impact on all three types of patents, yielding an elasticity of patents with respect to density at 1.63. Columns (2)-(4) provide separate estimates for each patent type, emphasizing that invention patents demonstrate the highest elasticity with respect to density. This result indicates that invention patents are particularly sensitive to changes in urban density, exhibiting an elasticity twice as large as the overall elasticity and those of the other two patent types. Nonetheless, as pointed out by Carlino and Kerr (2015), invention patents do not always correlate with production, as a significant portion of them never reach commercialization. Consequently, it is reasonable to consider the effects on utility patents and design patents as more closely tied to productivity. These two patent types, which focus on enhancing existing products or services, are likely to have a more immediate and tangible impact on productivity levels within the economy.

5.3.2 Sector composition

The economic growth path of many economies can be characterized by a decline in the output share of the primary sector (agriculture, forestry, fishing, and mining), a gradual increase and slowdown in the secondary sector (manufacturing, construction, and utilities) growth, and an expansion of the tertiary sector (services and information-based industries). This transition reflects the nation's efforts to shift towards a more sustainable, service-oriented, and consumption-driven growth model. It is important especially in a developing country's context, to examine whether agglomeration plays a role in shaping this transition.

In the first three columns of Table 7, I estimate the effect of density on each sector's share in total employment. The results indicate that a 10% increase in density corresponds to a decrease in the primary sector's employment share by 1.1 percentage points, which is approximately 90% of

Table 6: The effect of urban density on patents granted

Dependent variable: log number of patents per 10,000 population				
	(1) All Patents	(2) Invention patents	(3) Utility patents	(4) Design patents
Log urban density	1.631**	3.474***	1.588*	1.785**
	(0.808)	(1.301)	(0.827)	(0.773)
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Dependent Mean	0.592	-1.825	0.051	-0.860
First-stage F-stat	20.139	20.139	20.139	20.139
Observations	2752	2752	2752	2752

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). 233 observations (7.81%) observations are dropped because of missing patent data. The dependent variable is the log number of patents granted. Each column includes the same specification as in column (6) of Table 5. Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.

the primary sector's average share. The same increase in density results in a 0.21-percentage-point growth in the employment share of the secondary sector with marginal significance. The same exercises are repeated in columns (4)-(6) of Table 7 for the output shares by sector. While the estimates for primary and secondary shares are negative but not significant, the result in column (6) suggests that increasing density by 10% is estimated to increase the output share of the tertiary (service) sector by 1.84% percentage points. In the last set of columns, (7)-(9), I estimate the influence of density on per-worker outputs across all three sectors. The findings reveal that density enhances productivity in the tertiary (service) sector, whereas it leads to a decline in productivity in the secondary (manufacturing) sector[14]. The estimated elasticities are respectively 0.72 and -0.78.

The results in Table 7 are consistent with the general beliefs about the sector transitions in developing economies. Agglomeration is expected to benefit service industries by both increasing their share in GDP and enhancing their productivity. While this study is among the first to investigate the influence of density on sector composition, the findings corroborate several hypotheses regarding mechanisms like accessibility and knowledge spillovers, both of which are more prevalent in the service sector.

[14] The estimation for the primary sector, presented in column (7), is positive and statistically significant. Nevertheless, considering that farmers are classified as employees in the dataset, it is plausible to assume that per-worker output may be overestimated. The elevated mean value of the dependent variable in column (7) further supports this hypothesis.

Table 7: 2SLS estimates on labor/output shares, and per-worker output, by sector

	Sector labor share			*Sector output share*			*Sector output per worker*		
	(1) Primary	(2) Secondary	(3) Tertiary	(4) Primary	(5) Secondary	(6) Tertiary	(7) Primary	(8) Secondary	(9) Tertiary
Log urban density	-10.901**	21.338*	-10.560	-5.516	-12.747	18.441*	2.059*	-0.782**	0.724**
	(4.776)	(12.661)	(11.495)	(4.535)	(11.760)	(10.788)	(1.109)	(0.382)	(0.335)
Latitude	0.131	1.014	-1.138	0.010	0.213	-0.225	-0.013	-0.030	0.005
	(0.185)	(0.707)	(0.727)	(0.272)	(0.547)	(0.514)	(0.068)	(0.024)	(0.019)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	1.271	46.129	52.696	6.304	48.553	45.132	10.303	7.958	7.738
First-stage F-stat	22.498	22.498	22.498	22.498	22.498	22.498	19.870	19.870	19.870
Observations	2985	2985	2985	2985	2985	2985	2801	2801	2801

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). 184 observations (6.16%) observations are dropped in the last three columns because of missing employment count in the primary sector. In this table, the dependent variables are as follows: A. For columns (1)-(3): The sector employment as a share of total employment. B. For columns (4)-(6): The sector output as a share of total GDP. C. For columns (7)-(9): The sector output divided by the sector employment. Each column includes the same specification as in column (6) of Table 5. Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

6 Conclusion

This study investigates the causal relationship between urban population density and a city's productivity by leveraging a policy variation arising from the daylight building spacing policy initiated in 1994. This policy affects building density in all cities by imposing restrictions on the minimum daylight hours for ground-floor rooms, with specific requirements differing based on a city's climate zone and urban population. To capitalize on this policy, a unique instrumental variable approach is employed, utilizing the policy-implied spacing factor that incorporates variations from both aforementioned dimensions. The spacing factor, defined as the ratio of the distance between buildings to building height, serves as an instrument for density. By analyzing the spacing factors for 43 major cities stipulated in the regulation, I derive each city's specific spacing factor and employ it as an instrument for density. Although the daylight spacing policy is designed to regulate the construction of residential buildings, it impacts the city's total density significantly. One plausible explanation is that Chinese cities inherit the Soviet style, characterized by a mix of residential and commercial buildings within the urban landscape. Increasing the distances between residential buildings can effectively reduce both residential and non-residential densities.

My findings suggest a strong positive effect of urban density on the average wage, with an estimated elasticity of approximately 0.4. Comparing 2SLS and OLS estimates on the average wage suggests that OLS estimates are biased downwards potentially due to omitted productivity advantages or sorting of productive workers. Doubling the urban density is estimated to yield an increase of 15,096.96 yuan in the average annual wage of all employees. I also explore two potential sources of productivity advantages derived from density: innovation and transitions in sector composition. Results indicate that urban density has a positive effect on innovative activities, with invention patents demonstrating the highest elasticity with respect to density. In terms of sector composition, agglomeration is found to benefit service industries by increasing their share in GDP and enhancing their per-worker output. The findings support hypotheses about the mechanisms, such as accessibility and knowledge spillover, which are more prevalent in the service sector.

This paper contributes to the research analyzing the productivity effect of agglomeration in the following ways. First, I explore a novel policy variation in urban population density. Compared to other works using quasi-experimental approaches to address the issue of endogenous density, my instrument is introduced by a central government policy based on differences across climate zone boundaries and a historical population cutoff, and thus not determined by local unobserved characteristics. Second, this paper runs an analysis based on 199 prefecture-level cities in China, which makes my findings more generalizable. Third, this paper stands out as one of the few studies examining the effect of agglomeration in a developing economy context.

My empirical design and findings also provide insights for future research. For example, if more data at the employer or employee level become available, one can use them to examine the effects of agglomeration on a more detailed measure of productivity, and potentially focus more on its dynamics, i.e. growth of worker productivity over time in large cities. Furthermore, this line of research enables the definition of agglomeration by industry and the implementation of regressions with more controls at the firm or industry level. This approach helps provide empirical evidence on the driving forces behind agglomeration benefits, such as innovation, and their impact on productivity. Lastly, the unique method of quantifying a policy and employing it as an instrument in this study illuminates the potential for using similar policy-based instruments to explore a wide range of research topics.

The findings of this paper hold implications for policymakers when formulating policies that influence urban density. Firstly, the strong impact of high density on productivity in the service sector, as evidenced by the per-worker output estimates, suggests that policies could focus on increasing density in areas with a high concentration of service-sector firms. In contrast, manufacturing firms may not necessarily benefit from being located in densely populated areas, indicating that density-related policies for these firms should be approached differently. This nuanced understanding of the relationship between density and productivity across various sectors can help policymakers make more informed decisions on urban planning and development.

References

Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.

Au, C.-C. and Henderson, J. V. (2006a). Are chinese cities too small? *The Review of Economic Studies*, 73(3):549–576.

Au, C.-C. and Henderson, J. V. (2006b). How migration restrictions limit agglomeration and productivity in china. *Journal of development economics*, 80(2):350–388.

Bertaud, A. and Brueckner, J. K. (2005). Analyzing building-height restrictions: predicted impacts and welfare costs. *Regional Science and Urban Economics*, 35(2):109–125.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and innovation. *Handbook of regional and urban economics*, 5:349–404.

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3):389–419.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review*, 86(1):54–70.

Combes, P.-P., Démurger, S., and Li, S. (2015). Migration externalities in chinese cities. *European Economic Review*, 76:152–167.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of urban economics*, 63(2):723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration economics*, pages 15–66. University of Chicago Press.

Duranton, G. (2008). From cities to productivity and growth in developing countries. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(3):689–736.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.

Glaeser, E. L., Scheinkman, J., and Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of monetary economics*, 36(1):117–143.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Grover, A., Lall, S., and Timmis, J. (2023). Agglomeration economies in developing countries: A meta-analysis. *Regional Science and Urban Economics*, page 103901.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of urban economics*, 53(1):1–28.

Liu, X. (2014). Spatial efficiency and regional balance: An examination of agglomeration effects at the provincial level in china (in chinese). *The Journal of World Economy*, (1):55–80.

Lu, M., Zhang, H., and Liang, W. (2015). How a biased midwestern land supply has pushed up wages in the east (in chinese). *Social Sciences in China*, (5):59–83.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–3375.

Neumark, D. and Simpson, H. (2015). Place-based policies. In *Handbook of Regional and Urban Economics*, volume 5, pages 1197–1287. Elsevier.

Rawski, T. G. (2001). What is happening to china's gdp statistics? *China Economic Review*, 12(4):347–354.

Zhang, H. (2018). Sunlight space constraint, population density, and urban growth in china (in chinese). *China Economic Quarterly*, (1):333–354.

Zhang, Q. (2019). Sunlight, urban density and information diffusion.

Zhang, Y. and Liu, X. (2008). Agglomeration economies and economic growth: empirical evidence from china (in chinese). *The Journal of World Economy*, (3):60–70.

Appendix

Figure A1: Climate zones in China

Figure A2: Geographic regions in China

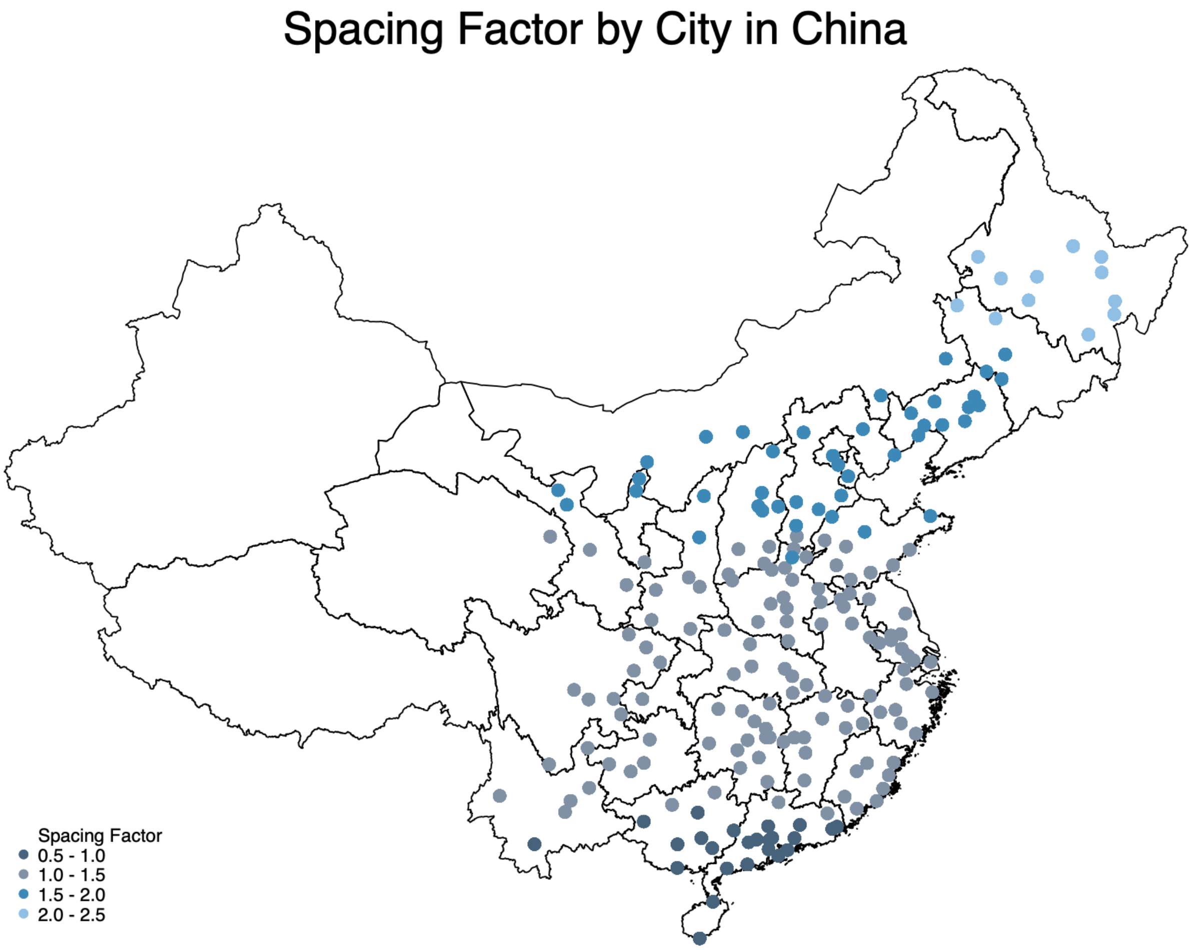

Figure A3: Maps of Chinese cities by spacing factors

Notes: Dots on the map represent the 199 cities in China, with colors representing the size of the policy-implied spacing factors.

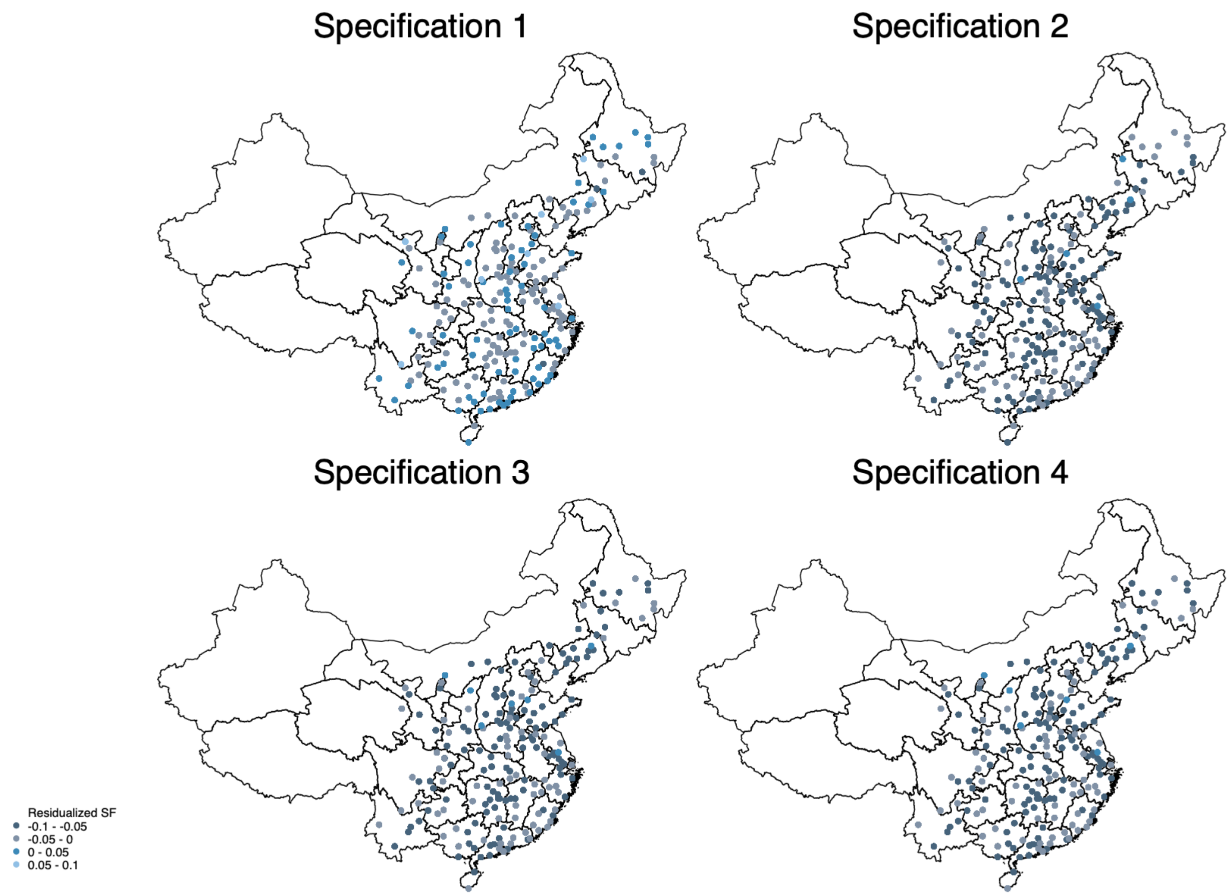

Figure A4: Maps of residualized spacing factors

Notes: Dots on the map represent the 199 cities in China, with colors representing the size of the residualized policy-implied spacing factors. The residuals come from regressions of the spacing factor on a function of latitude, population threshold, other covariates and FE's. Specifications 1-4 correspond to different functional forms of latitude: linear, quadratic, cubic, and exponential.

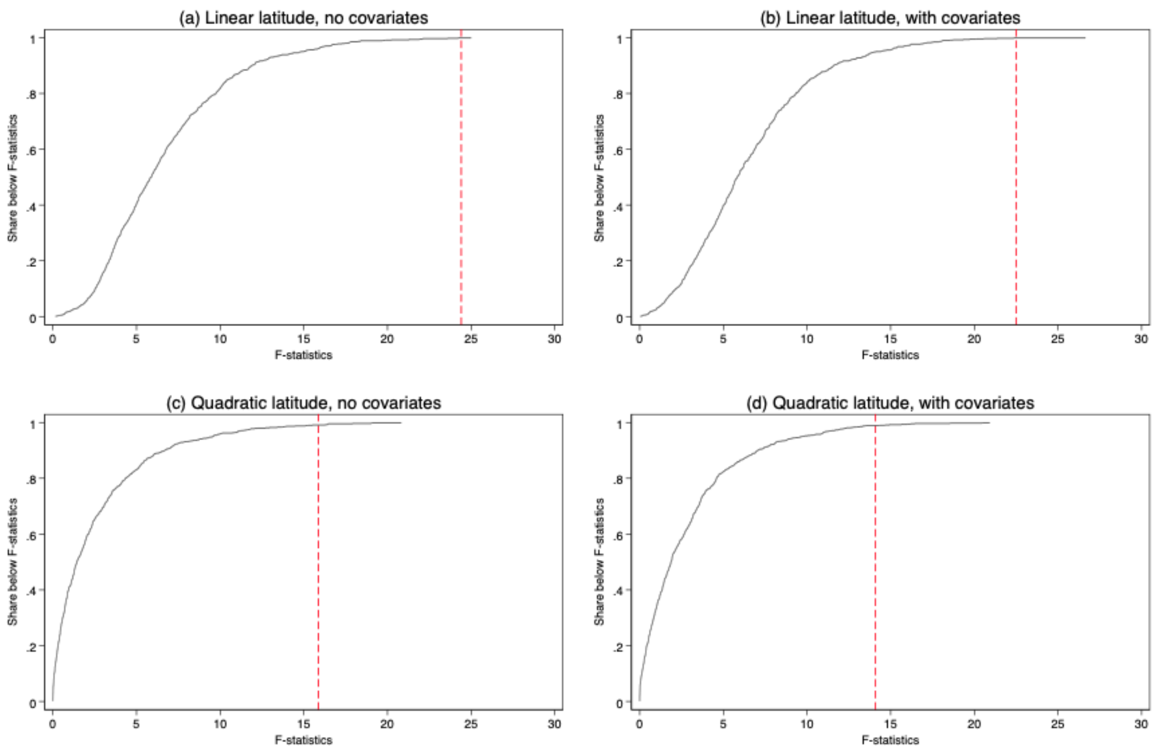

Figure A5: Falsification Tests

Notes: The F-statistics come from running regressions on 900 placebo spacing factors, constructed by randomizing the climate zone of each city and the population threshold. Panels (a) and (c) display the distribution of F-stats from regressions without other covariates (as those included in the first two column of Table 4), while panels (b) and (d) plot the same distribution from regressions with covariates as those in column (6) of Table 4. Panels (a) and (b) include the original latitude control, while panels (c) and (d) include a quadratic function of latitude, in order to check the robustness of first-stage estimates to alternative functional forms.

Table A1: First stage estimates of population density on the spacing factor: alternative functional forms

Panel A: No Population Control				
	(1)	(2)	(3)	(4)
	Linear	Quadratic	Cubic	Exponential
Spacing factor	-2.339***	-2.247***	-2.407***	-2.571***
	(0.512)	(0.596)	(0.615)	(0.525)
Latitude	0.086***	0.108	0.634**	0.094***
	(0.024)	(0.079)	(0.299)	(0.024)
Latitude^2		-0.000	-0.016*	
		(0.001)	(0.009)	
Latitude^3			0.000*	
			(0.000)	
Exp(latitude)				0.000
				(0.000)
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Dependent Mean	8.997	8.997	8.997	8.997
F-stat	20.882	14.196	15.338	23.944
R-squared	0.476	0.476	0.481	0.480
Observations	2880	2880	2880	2880
Panel B: Controlled for 1994 Population				
	(1)	(2)	(3)	(4)
	Linear	Quadratic	Cubic	Exponential
Spacing factor	-2.649***	-2.841**	-3.054**	-3.201***
	(0.779)	(1.176)	(1.182)	(0.786)
Latitude	0.101***	0.087	0.616**	0.124***
	(0.037)	(0.082)	(0.296)	(0.037)
Latitude^2		0.000	-0.016*	
		(0.002)	(0.009)	
Latitude^3			0.000*	
			(0.000)	
Exp(latitude)				0.000
				(0.000)
Province FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Dependent Mean	8.997	8.997	8.997	8.997
F-stat	11.555	5.832	6.671	16.576
R-squared	0.477	0.477	0.481	0.481
Observations	2880	2880	2880	2880

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). 105 records are missing due to missing values in 1994 population. The dependent variable is the logarithm of the urban population density (in 10,000/km^2). The *Spacing factor* is defined as the ratio of building distance to building height and is calculated based on the same method used in the daylight regulation policy. Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: First stage estimates of population density on the spacing factor: alternative region fixed effects

Panel A: No Population Control				
	(1)	(2)	(3)	(4)
	No FE	Climate Zones	Admin Zones	Provinces
Spacing factor	-0.771**	-1.495***	-2.147***	-2.339***
	(0.301)	(0.498)	(0.443)	(0.512)
Latitude	0.042***	0.074***	0.089***	0.086***
	(0.015)	(0.025)	(0.019)	(0.024)
Region FE	-	Climate	Geo	Province
Year FE	Yes	Yes	Yes	Yes
Dependent Mean	8.997	8.997	8.997	8.997
First-stage F-stat	6.551	9.017	23.498	20.882
Adj. R-squared	0.220	0.237	0.333	0.467
Observations	2880	2880	2880	2880
Panel B: Controlled for 1994 Population				
	(1)	(2)	(3)	(4)
	No FE	Climate Zones	Admin Zones	Provinces
Spacing factor	-0.531*	-0.930	-2.133***	-2.649***
	(0.310)	(0.636)	(0.515)	(0.779)
Latitude	0.029*	0.045	0.088***	0.101***
	(0.015)	(0.031)	(0.022)	(0.037)
Population(1994) $\geq$ 500K	0.100**	0.073	0.003	-0.028
	(0.045)	(0.055)	(0.043)	(0.052)
Region FE	-	Climate	Geo	Province
Year FE	Yes	Yes	Yes	Yes
Dependent Mean	8.997	8.997	8.997	8.997
First-stage F-stat	2.934	2.140	17.169	11.555
Adj. R-squared	0.237	0.242	0.333	0.467
Observations	2880	2880	2880	2880

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). 105 records are missing due to missing values in 1994 population. The dependent variable is the logarithm of the urban population density (in 10,000/km^2). The *Spacing factor* is defined as the ratio of building distance to building height and is calculated based on same method used in the daylight regulation policy. In both panels, column (1) does not include any region FE, and column (2) includes FE at the climate zone level. column (3) instead include FE at the administrative zone level. column (4) narrows this down to province level. All columns in Panel B add a dummy for whether the population in 1994 is greater than 500K (threshold). Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: 2SLS estimates on GDP per capita

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dependent variable: the log of GDP per capita						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	2SLS	2SLS
Log urban density	-0.310**	-0.263**	-0.264**	-0.256**	0.288	0.213
	(0.131)	(0.120)	(0.111)	(0.109)	(0.390)	(0.387)
Latitude			-0.026	-0.006	-0.017	0.002
			(0.023)	(0.023)	(0.025)	(0.024)
Distance to coastline (km)			-0.463	-0.608**	-0.496	-0.657**
			(0.301)	(0.299)	(0.327)	(0.323)
Average slope			-0.041***	-0.031**	-0.040***	-0.029**
			(0.012)	(0.012)	(0.013)	(0.013)
Average elevation (km)			0.546***	0.517***	0.605***	0.562***
			(0.154)	(0.145)	(0.176)	(0.162)
Total area including undeveloped land ($10000 km^2$)				-0.037		0.065
				(0.167)		(0.175)
Area of prefecture ($10000 km^2$)				-0.089***		-0.100***
				(0.030)		(0.031)
Province FE	-	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	10.656	10.656	10.656	10.656	10.656	10.656
R-squared	0.358	0.539	0.572	0.583		
First-stage F-stat					22.188	22.492
Observations	2984	2984	2984	2984	2984	2984
Panel B: Dependent variable: the level of GDP per capita						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	2SLS	2SLS
Log urban density	-13642.84*	-15027.11*	-18196.28**	-17214.97**	25534.18	23114.27
	(7892.77)	(8060.19)	(8314.97)	(8596.60)	(21338.11)	(21079.91)
Latitude			-493.88	385.21	207.42	1075.84
			(1709.89)	(1611.27)	(1875.18)	(1782.15)
Distance to coastline (km)			644.79	-6497.62	-1999.63	-10656.94
			(27995.24)	(30049.11)	(29209.87)	(30879.26)
Average slope			-2415.29***	-1935.75***	-2319.27***	-1758.77**
			(703.56)	(730.89)	(787.64)	(802.90)
Average elevation ()			-4344.02	-5721.58	313.58	-1803.42
			(25673.39)	(25213.90)	(25723.82)	(25019.08)
Total area including undeveloped land ($10000 km^2$)				4254.66		13022.07
				(10949.94)		(10690.84)
Area of prefecture ($10000 km^2$)				-4257.07*		-5218.75**
				(2260.19)		(2324.07)
Province FE	-	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	55146.80	55146.80	55146.80	55146.80	55146.80	55146.80
R-squared	0.032	0.048	0.053	0.054		
First-stage F-stat					22.192	22.498
Observations	2985	2985	2985	2985	2985	2985

Notes: The sample includes 199 prefecture-level cities in China observed over 15 years (2005-2019). The dependent variable is the log of GDP per capita (in Panel A) and the level of GDP per capita (in Panel B). In both panels, OLS regression results are shown in columns (1)-(4), and 2SLS results are shown in columns (5) and (6). Robust standard errors are clustered at the city level. All financial variables in this research are adjusted by inflation and are in the real value of the Chinese Yuan in 2005. *** p<0.01, ** p<0.05, * p<0.1.