

# Urban Agglomeration Effects in India: Evidence from Town-Level Data

RANA HASAN, YI JIANG, AND RADINE MICHELLE RAFOLS\*

---

Combining multiple data sets for India, we estimate the elasticity of wages with respect to town population and density between 1% and 2%, which is smaller than estimates in the literature based on district-level analysis. We also find that the employment share of firms with 10 or more workers—which typically describes firms that operate in the formal sector—is positively associated with city population and negatively associated with city density. Town characteristics such as infrastructure availability, geographic location, educational services, and industrial structure also play a role in explaining city productivity and the presence of relatively large firms. Overall, we interpret our results to suggest that there is scope to realize more fully urbanization’s potential by addressing issues related to urban planning, infrastructure, and public service delivery, as has been emphasized previously by observers of Indian urbanization.

*Keywords:* agglomeration, formal sector, Indian city

*JEL codes:* O15, R12, R23

---

## I. Introduction

Like other developing economies, India is urbanizing. According to census data, the share of India’s population residing in urban areas increased from 20% in 1971 to 31% in 2011. Expectations are that the process of urbanization in India will continue, if not accelerate. With cities widely believed to be “engines of growth,” urbanization represents an important source of prosperity for India. This is especially true given the scope for further urbanization.

However, the extent to which India’s urbanization will play the positive role that is expected of it is unclear. Ahluwalia, Kanbur, and Mohanty (2014) note that in comparison to other fast-growing economies in Asia, urbanization in India has been relatively slow, largely unplanned, and characterized by underinvestment in urban infrastructure and public service delivery. By way of comparison, the McKinsey

---

\*Rana Hasan: Director, Economic Research and Regional Cooperation Department, Asian Development Bank (ADB). E-mail: rhasan@adb.org; Yi Jiang (corresponding author): Senior Economist, Economic Research and Regional Cooperation Department, ADB. E-mail: yijiang@adb.org; Radine Michelle Rafols: Consultant, ADB. E-mail: radine.rafols@gmail.com. We are grateful to Gilles Duranton for his insightful advice, the participants of the Asian Development Review Conference on Urban and Regional Development in Asia held in Seoul in July 2016, the managing editor, and an anonymous referee for helpful comments and suggestions. In this paper, ADB recognizes “China” as the People’s Republic of China. The usual disclaimer applies.

Global Institute (2010) reports that India's annual capital spending in urban areas is only \$17 per capita, which compares unfavorably with annual capital spending in the People's Republic of China of \$116 per capita. Generally, and common to the situation in many developing economies, there are various aspects of urbanization in India that may act as a brake on economic growth. For example, Duranton (2015) notes that cities in developing economies tend to be much less functionally specialized than cities in advanced economies. Large cities in developing economies tend to be characterized by many ancillary activities that could be undertaken in smaller cities. This adds to urban congestion and detracts from the benefits that urban agglomeration is expected to bring.

In this paper, we use data from various sources, including India's Economic Census of 2005 (EC 2005) and the Labor Force Survey (LFS) of 2004–2005, to shed light on whether Indian cities are functioning as engines of growth. We do this by examining the effects of urban agglomeration on proxies for worker and firm productivity at the town level.<sup>1</sup> The focus on towns as the unit of analysis distinguishes our work from previous literature on India such as Chauvin et al. (2017) and Lall, Shalizi, and Diechmann (2004), who examine urban agglomeration issues at the district level, which is a higher level of geographic aggregation.

We analyze how measures of city-level average (nominal) wages and the share of employment in formal or modern firms (defined as firms with 10 or more employees) are related to measures of urban agglomeration such as population and density. In addition to instrumenting for urban agglomeration, we control for a variety of city characteristics that capture natural amenities, infrastructure availability, geographic location, and educational services in our agglomeration regressions. We also examine how city wages and the prevalence of formal firms relate to the industrial structure of cities; for instance, whether cities with a more diversified industrial structure or a larger share of employment generated through manufacturing activities have higher wages and more formal firms.

While our use of nominal wages is typical in the empirical literature on the effects of urban agglomeration—they are reasonably good indicators of labor productivity as opposed to real wages, which better capture standards of living—the specific measure of city-level wages is less typical. Our city-level wages are derived from information on the sectoral composition of employment across towns and average wages by industry and district. Thus, as a robustness check on our results, we use an approach first employed by Ciccone and Hall (1996) in which district-level wages are related to an index of district-level density that is a nonlinear function of the density of constituent cities.

With regard to our analysis of urban agglomeration based on the share of employment in larger firms, firms with 10 or more employees typically (i) must comply with rules governing industrial labor and workplace safety, (ii) comprise

---

<sup>1</sup> We use the terms city and town interchangeably.

the formal sector, and (iii) pay better wages and are more productive (Asian Development Bank 2009).<sup>2</sup> If economies of agglomeration are important, one would expect to see a greater share of total employment in such firms.

Our analysis indicates that the elasticity of wages with respect to both city population and density is between 1% and 2%, which is smaller than existing estimates as well as estimates from our replication exercise based on analysis at the district level. Our estimates for India are further supported by the results derived from an application of the Ciccone and Hall (1996) approach. We find that a larger employment share in formal firms is positively associated with city population but negatively associated with city density, possibly reflecting the congestion-related effects of higher density. Town characteristics such as natural amenities, infrastructure availability, geographic location, educational services, and industrial structure also play a role in explaining city productivity and the presence of formal firms. Overall, we interpret our results to suggest that urbanization does hold promise for promoting growth and good jobs in India. However, there is scope to further realize urbanization's potential by addressing issues related to urban planning, infrastructure, and public service delivery as emphasized by Ahluwalia, Kanbur, and Mohanty (2014).

## II. Literature Review

Rosenthal and Strange (2004) and Combes and Gobillon (2015) provide comprehensive surveys of studies that estimate agglomeration effects in various economies and regions. Generally, the elasticity of productivity, whether measured by wages or total factor productivity, falls between 1% and 10% with respect to city population or density. According to Combes and Gobillon (2015), studies using city-level productivity measures yield higher elasticity estimates (4%–7%) than those using individual data, which typically reach about 2%.

Melo, Graham, and Noland (2009) undertake a meta-analysis of the empirical estimation of agglomeration effects and conclude that the results are highly context specific depending on factors such as the economy and industries studied, and the controls used for unobserved heterogeneity. Among the large number of empirical studies on the topic, only a small share look at developing economies. Thus, evidence from the developing world is lacking (Duranton 2015).

Lall, Shalizi, and Deichmann (2004) use plant-level data to estimate the effect of district urban density on firm productivity in India. Among the nine manufacturing industries examined, urban population density has a significantly positive effect only on the manufacturing of cotton textiles with a point estimate of

---

<sup>2</sup>The Factories Act, 1948 governs India's industrial sector and the Trade Union Act, 1926 governs the formation and activities of trade unions in the workplace.

about 9%. The effects on other industries are not statistically significant and some are even negative.

More recently, Chauvin et al. (2017) estimate the elasticity of individual income (of prime-age males) to urban population and density at the district level, and instrument population and density with historical population and density in 1980 and 1951, respectively. When nominal income is examined, most estimates of elasticity are around 7%–8%, with some instrumental variable (IV) estimates of the population effect being fairly large. In the case of real income, the estimates are essentially the same at around 6%. Again, urban density demonstrates robustness, while IV estimation generates some unreasonably large effects for population.

Most studies on urban economies in India treat districts—the second administrative level after states—as the urban unit (see, for example, Lall, Shalizi, and Deichmann 2004; Ghani, Kanbur, and O’Connell 2013; Chauvin et al. 2017). This is because most available data from labor and enterprise surveys contain geo-information down to the district level. However, Indian districts often cover large rural areas and many contain multiple geographically independent urban areas. A more proper definition of a city in India would be towns, which is the administrative division below the district level and whose rural counterparts are villages. Ciccone and Hall (1996) show that state-level average density in the United States has no effect on state-level labor productivity once heterogeneity in density within states is accounted for. Therefore, examining agglomeration economies at the town level in India is considered one of the main contributions of this paper.

### III. Data

This paper combines cross-sectional, establishment-level data from India’s 2005 Economic Census; town-level data from the Town Directory of the 2001 Population Census; and individual-level survey data from the 2004–2005 Employment–Unemployment Survey to explore agglomeration effects across cities in India.

The economic census, which was conducted by the Central Statistics Office of the Ministry of Statistics and Programme Implementation, is a countrywide census of establishments engaged in all economic activities except crop production and plantations. The key purpose of the economic census is to provide a sampling frame for follow-up sample surveys intended to collect more detailed sector-specific information on the nonagricultural economy. In this study, we employ the recently released fifth edition of the Economic Census covering the year 2005. The data allow for the geographic location of establishments to be identified at the town level, which is an administrative level below the state and district levels that is the equivalent to a city in the Indian context.<sup>3</sup>

<sup>3</sup>As mentioned above, a village is the rural counterpart to a town.

Table 1. City Size and Classification

Population	Classification		
	Town	Urban Agglomeration	Total
500,000 and above	13	60	73
100,000 to 499,999	139	173	312
10,000 to 99,999	2,253	141	2,394
Total	2,405	374	2,779

Source: Authors' estimates based on 2005 Economic Census and Town Directory of the 2001 Population Census.

Establishment-level information in the EC 2005 includes number of employees, major activity in terms of a four-digit industrial classification, type of fuel used, registration status with government authorities, and type of ownership. Approximately 17 million establishments surveyed were recorded under the four-digit 2004 National Industrial Classification. The 304 unique four-digit categories can be simplified into 59 two-digit aggregates. The analysis put forth in this paper includes all industries after reclassifying them further into 13 major industrial categories. Table A.1 presents the reclassification of the two-digit National Industrial Classification into 13 categories. Table A.2 reports the distribution of all firms as well as firms with 10 or more employees among these categories.

We also employ the Town Directory of the 2001 Population Census to introduce our town-level agglomeration measures, that is, population and density. This survey contains rich information on geography and climatic amenities, demography, infrastructure provision, government revenues and expenditures, and social and educational services available at the town level. This is the only source we are aware of that has town-level population and land area data that is close to 2005. In addition, we use the town-level characteristics as explanatory variables in our agglomeration regressions to examine how urban productivity varies with city amenities, infrastructure, and access, among other characteristics.<sup>4</sup>

Our final data set allows us to work with about 2,800 Indian cities from around 560 districts with a population of at least 10,000 in 2001. Table 1 presents the distribution of our cities by classification and population. As defined by the Census of India, towns are classified as either census towns or statutory towns, while an urban agglomeration is a construct where contiguous urban areas are included as part of a town for urban planning purposes.<sup>5</sup> Urban agglomerations must consist of

<sup>4</sup>The Appendix describes our efforts to match towns in the Town Directory of the 2001 Population Census with the EC 2005, and defines our town- and district-level variables.

<sup>5</sup>Census towns are administrative units satisfying the following three criteria simultaneously: (i) minimum population of 5,000; (ii) 75% or more of the male working-age population engaged in nonagricultural pursuits; and (iii) population density of at least 400 persons per square kilometer. Statutory towns are defined as urban-like areas with a municipal corporation, municipality, cantonment board, notified town area committee, or town council.

Table 2. City-Level Statistics

Variables	N (1)	Mean (2)	SD (3)	Min (4)	Max (5)
<b>Agglomeration measures</b>					
Population as of 2001	2,779	96,463	546,496	10,018	1.64e07
Density as of 2001 (per km <sup>2</sup> )	2,769	5,024	5,906	87.820	80,305
Population as of 1981	2,520	58,989	315,494	20.100	9.085e+06
Density as of 1981 (per km <sup>2</sup> )	2,510	3,206	3,578	1.702	48,353
Corrected population as of 2001	2,779	261,282	672,523	5,141	1.643e07
Corrected density as of 2001 (per km <sup>2</sup> )	2,776	5,890	3,667	271.300	44,076
Corrected population as of 1981	2,779	138,516	356,875	1,545	9.085e06
Corrected density as of 1981 (per km <sup>2</sup> )	2,776	3,397	2,155	51.050	19,384
<b>Dependent variables</b>					
Derived daily wage (Rs)	2,779	83.330	20.089	32.790	192.800
Share of employment from firms with 10 or more employees	2,779	0.220	0.154	0	0.968
<b>Geographical controls</b>					
Town Area (km <sup>2</sup> )	2,778	22.510	53.690	0	1,135
Maximum temperature (°C)	2,749	37.210	5.689	14.030	52
Minimum temperature (°C)	2,751	14.040	7.140	-4	35
Average rainfall (mm)	2,761	1,043	717.700	33.900	10,270
Total electricity connections	2,779	19,494	188,766	0	8.667e06
Paved road length (km)	2,779	73.590	378.500	0	9,367
Distance to state headquarters (km)	2,759	295	189.400	0	1,094
Number of educational institutions	2,779	2.063	9.166	0	225
Diversity Index	2,779	1.957	0.525	0.601	4.838
Specialization Index	2,779	17.500	48.600	1.648	1,176
Share of manufacturing employment to total employment	2,779	0.211	0.135	0	0.942

°C = degree Celsius, km = kilometer, km<sup>2</sup> = square kilometer, mm = millimeter, sd = standard deviation.

Source: Authors' estimates based on 2005 Economic Census, Town Directory of the 2001 Population Census, and the National Sample Survey Schedule 10. Round 61 (2004–2005).

at least a statutory town and its total population (all constituents combined) should not be less than 20,000. Both towns and urban agglomerations are treated equally in our analysis and referred to as cities.

The majority (86.1%) of cities in India had a population of less than 100,000 in 2001. There were only 73 cities in India, or 2.6% of the total, at the other end of the spectrum with a population of 500,000 or more. Overall, urban agglomerations accounted for 13.5% of all cities in the data and are disproportionately more prevalent in cities with a population of 100,000 or more.

Table 2 reports the summary statistics for our sample. The average city had a population of about 96,000 and India's urban population amounted to 270 million in 2001. The figures suggest India was still at an early stage of urbanization in 2001 given its total population of 1.06 billion at the time. However, Indian cities were larger, denser, and more numerous in 2001 than in 1981 when the average city size was about 59,000 people and the average density was 3,206 people per square

kilometer.<sup>6</sup> These figures would increase by 64% and 57%, respectively, over the next 20 years. We also see considerable variation across cities in terms of both population and density.

Table 2 also presents city-level measures of climate, infrastructure provision, and educational services. The diversity of these characteristics across Indian cities is noteworthy. For instance, the highest annual maximum temperature is three times that of the lowest, and the highest average rainfall is 300 times the lowest. There are towns without any electricity connections, paved roads, or educational institutions, while an average town has 19,494 electricity connections, 74 kilometers of paved roads, and 2 educational institutions.

#### IV. Methodology

As is common in the literature, we estimate agglomeration benefits using data on wages—a proxy for productivity—and testing whether wages increase with a city's population and density. Ideally, we would like to estimate agglomeration regressions of the following type:

$$\log y_{ic} = \alpha + \beta \log S_c + X_i \gamma + Z_c \lambda + \varepsilon_{ic}$$

where  $y_{ic}$  is the wage of person  $i$  living in city  $c$ ;  $S_c$  is the population of the city—so as to capture the concept of scale, or some other measure of agglomeration, such as density (population divided by area of the city);  $X_i$  is a vector of individual characteristics;  $Z_c$  is a vector of city characteristics; and  $\varepsilon_{ic}$  is the error term. The coefficient,  $\beta$ , captures the elasticity of wages with respect to city population (or density), which is our main interest.

The main difficulty with estimating this agglomeration regression in the Indian context is data related. In particular, individual wage information from the LFS only identifies the district and state of the respondent, as well as an urban or rural location. While it is possible to use the urban component of Indian districts as the unit of analysis to estimate agglomeration economies, as done recently by Chauvin et al. (2017)—so that individual-level wages are regressed on district-level measures of agglomeration (district population or density)—this is less than ideal as districts may not be well-defined economic units over which economies of agglomeration may be operating. A district in India can cover between 1 and 22 cities and towns, and these can be spread over an area covering thousands of square kilometers.

To work at a finer geographic unit, several options are available. One approach is to work along the lines of Ciccone and Hall (1996). This would entail working with average wages at the district level and regressing these on an index

<sup>6</sup>Cities in the Town Directory of the 2001 Population Census with missing population data in 1981 are treated as rural areas for that year.

of the density of a district that is defined in terms of the density of its constituent towns. We consider this approach below. A second alternative is to examine if the available data allow one to construct a measure of city-level average wages and then use this measure to see how it varies with city characteristics.

How can the latter be done? While we can observe the scale and density of each town, as well as the industrial composition, thanks to the population and economic censuses, respectively, we do not have a direct measure of wages—the commonly used proxy for productivity in the agglomeration literature. On the other hand, we have data from the LFS carried out around the time of the EC 2005, which contains information on individual earnings and the industry and district of employment. We combine these two sources of information to construct a measure of wages at the town level in the manner described below:

Assume that a district  $d$  consists of multiple towns  $c = 1, 2, \dots, C_d$ . For an individual  $i$  in industry  $j$  and town  $c$ , the wage can be written as

$$y_{ijc} = \delta_c + \varepsilon_i \quad (1)$$

where  $\delta_c$  is a local effect for town  $c$ , which we do not observe directly, and  $\varepsilon_i$  is an independent and identically distributed error term. Let  $N_c \equiv \sum_j n_{jc}$  be the total number of firms in town  $c$ . It sums across all industries  $j$  the number of firms in that industry in town  $c$ ,  $n_{jc}$ . The total number of firms in  $j$  in district  $d$  is  $n_{jd} = \sum_c n_{jc}$ . We can also define the share of industry  $j$  in town  $c$ ,  $s_{jc} = n_{jc}/n_{jd}$ . Given equation (1), the mean wage in  $j$  of district  $d$  is

$$y_{jd} \equiv \frac{\sum_c n_{jc} \delta_c}{n_{jd}} = \sum_c s_{jc} \delta_c \quad (2)$$

We define average wage of town  $c$  as

$$y_c \equiv \frac{\sum_j n_{jc} y_{jd}}{\sum_j n_{jc}} = \frac{\sum_j n_{jc} y_{jd}}{N_c} \quad (3)$$

Plugging equation (2) into equation (3) and altering the notation a bit, we have the relation between  $y_c$  and local effects:

$$y_c = \frac{\sum_j n_{jc} \sum_{c'} s_{jc'} \delta_{c'}}{N_c} = \frac{\sum_j n_{jc} (s_{jc} \delta_c + \sum_{c' \neq c}^{C_d} s_{jc'} \delta_{c'})}{N_c} \quad (4)$$

so that the average wage of town  $c$ , computed using the district average wage, is a linear function of the local effects of both town  $c$  and other towns in the same district.

Suppose the agglomeration regression we wanted to estimate is  $\delta_c = \beta \log S_c + \mu_c$ , where  $S$  is either town population or density. However,  $\delta_c$  is not observed directly. Instead, we can estimate the following regression:

$$y_c = \beta \log \tilde{S}_c + v_c$$

where  $\log \tilde{S}_c = \frac{\sum_j n_{jc}(s_{jc} \log S_c + \sum_{d' \neq c}^{C_d} s_{jc'} \log S_{d'})}{N_c}$ , corresponding to the transformation of  $y_c$  in equation (4).

To implement the procedure described above, we first estimate a wage equation using the LFS data for 2004–2005:

$$\ln y_i = \beta_0 + \beta_1 e_i + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 G_i + \gamma_j + \phi_d + \tau_i \tag{5}$$

where  $y$  is daily wage;  $e$ ,  $X$ , and  $G$  are years of schooling, labor market experience, and gender, respectively; and  $\gamma$  and  $\phi$  are industry and district fixed effects, respectively. The average (log) wage of district  $d$  and industry  $j$  is then computed using the estimated coefficients and fixed effects from the wage equation above and the national average of individual characteristics.<sup>7</sup>

$$\widehat{\ln y_{jd}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{e} + \hat{\beta}_2 \bar{X} + \hat{\beta}_3 \bar{X}^2 + \hat{\beta}_4 \bar{G} + \hat{\gamma}_j + \hat{\phi}_d$$

Plugging  $\widehat{y_{jd}} = \exp(\widehat{\ln y_{jd}})$  into equation (3) gives us the town-level wage,  $\widehat{y}_c$ . Table 2 reports that average  $\widehat{y}_c$  is Rs83 per day with a standard deviation of Rs20 across cities.

The agglomeration model we estimate is

$$\ln \widehat{y}_c = \alpha + \beta \log \tilde{S}_c + v_c \tag{6}$$

where  $\beta$  measures the elasticity of average wages with respect to the population or density of a city.

As  $\widehat{y_{jd}}$  is estimated using the national average characteristics of workers, the estimation of equation (6) is unlikely to be biased by more productive workers selecting into large cities. However, an ordinary least squares (OLS) estimation of  $\beta$  is still subject to endogeneity arising when higher wages attract more workers or in cases where city characteristics are missing (Combes, Duranton, and Gobillon 2011). We address these issues by augmenting the regression model with city characteristics such as infrastructure availability, distance to the state capital, education facilities, climate amenities, and district dummies; and by instrumenting city population and density,  $\log \tilde{S}_c$ , with historical population and density, respectively.<sup>8</sup>

<sup>7</sup>Table A.3 presents the estimated coefficients, industry fixed effects, and sample means. Instead of using district or state averages of individual characteristics, we use national averages to remove the contribution of individual heterogeneity to the average town-level wages. This should help reduce the correlation between city population or density and unobserved individual characteristics at the town level.

<sup>8</sup>The historical population or density is transformed the same way as  $\log S_c$  is transformed into  $\log \tilde{S}_c$ ; that is, by applying equation (4). Summary statistics are provided in Table 3.

In addition to derived town-level wages, we also examine the effects of city scale on an alternative outcome that can be directly measured at the town level. The prevalence of modern or formal firms is one such outcome. Manufacturing firms with 10 or more workers (that use electrical power in the production process) are treated as formal sector firms in India as per the Factories Act, 1948. Generally, it seems safe to assume that India's more dynamic and productive firms tend to be larger. They certainly pay higher wages. (See Hasan et al. 2017 for more information on the relationship between establishment size and wages in the Indian apparel industry). The EC 2005 data show that 75% of Indian firms have two or fewer employees and that 98% of Indian firms have 10 or fewer employees. Thus, a majority of enterprises seem to be self-employment ventures that serve as a means of subsistence for families. In view of this, we consider the share of a city's total employment by firms with 10 or more employees as a proxy for the productivity of a city and investigate how this relates to city size and density.<sup>9</sup> Table 2 shows that on average large firms account for 22% of city employment, with a minimum share of 0% and a maximum share of 97%.

## V. Estimation Results

### A. The Effects of Agglomeration on Wages: District-Level Analysis

In column (1) of Table 3, we report an exercise that replicates Chauvin et al. (2017) to estimate agglomeration effects at the district level. Chauvin et al. (2017) estimate regressions using individual wage data for male workers from districts with an urban population of 100,000 or more, controlling for worker age and level of education. We mimic their sample selection and model specification closely, extending the model to all workers and controlling for additional individual and district characteristics.

As columns (2) and (4) in Table 3 show, when we adopt the same model specification, our agglomeration effects obtained from OLS and IV estimations using 1981 data are qualitatively consistent with those in Chauvin et al. (2017) for both male workers and the sample of all workers. However, when we add additional control variables to the regressions such as individual variables (e.g., labor market experience), industry dummies, and district characteristics (e.g., climate and infrastructure availability) in columns (3) and (5), many of our estimates turn insignificant or even negative.<sup>10</sup>

<sup>9</sup>The correlation between the employment share variable with the derived average town wages is 0.21 and is significant at the 1% level.

<sup>10</sup>We also worked with 1951 data to construct our instrumental variables. However, there are cities and towns with missing population data from 1951 that tend to be smaller (population of between 10,000 and less than 100,000). Since estimates of agglomeration effects could be sensitive to samples with relatively fewer small towns and cities, we dropped 1951 data from the IV construction.

Table 3. Estimates of Elasticities of Wages with Respect to District Population and Density

	Chauvin et al. (2017)	Male Only		All Workers	
	(1)	(2)	(3)	(4)	(5)
<b>OLS regressions</b>					
Log of urban population	0.0770*** (0.0264) R <sup>2</sup> = 0.251	0.0689*** (0.0202) R <sup>2</sup> = 0.472	0.0199 (0.0248) R <sup>2</sup> = 0.555	0.0863*** (0.0207) R <sup>2</sup> = 0.521	0.0416** (0.0171) R <sup>2</sup> = 0.600
Log of density	0.0760*** (0.0195) R <sup>2</sup> = 0.257	0.115** (0.0444) R <sup>2</sup> = 0.468	0.0150 (0.0314) R <sup>2</sup> = 0.555	0.144*** (0.0449) R <sup>2</sup> = 0.517	0.0208 (0.0345) R <sup>2</sup> = 0.600
Observations	9,778	18,119	17,998	29,355	29,186
<b>IV-1981 regressions</b>					
Log of urban population	0.160 (0.0998) R <sup>2</sup> = 0.237	0.0536*** (0.0165) R <sup>2</sup> = 0.471	-0.0998*** (0.0311) R <sup>2</sup> = 0.552	0.0710*** (0.0178) R <sup>2</sup> = 0.520	-0.0820*** (0.0257) R <sup>2</sup> = 0.598
Log of density	0.0828*** (0.0218) R <sup>2</sup> = 0.253	0.0412 (0.0374) R <sup>2</sup> = 0.466	-0.0636*** (0.0379) R <sup>2</sup> = 0.553	0.0706* (0.0393) R <sup>2</sup> = 0.514	-0.0583 (0.0387) R <sup>2</sup> = 0.598
Observations	7,627	17,930	17,809	29,074	28,905
Age	Yes	Yes	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes	Yes	Yes
Gender	No	No	No	Yes	Yes
Industry dummies	No	No	Yes	No	Yes
Labor market experience	No	No	Yes	No	Yes
District characteristics	No	No	Yes	No	Yes

IV = instrumental variables, OLS = ordinary least squares.

Notes: Robust standard errors in parentheses. Clustered at state level. Weighted to national level with National Sample Survey Organization sample weights. District characteristics include climate, infrastructure availability, weighted distance to state headquarters of constituent cities or towns and number of educational institutions. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

(1) is copied from the last column of Table 8 of Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India, and the United States." *Journal of Urban Economics* 98: 17–49. Regressions are at the individual level and restricted to urban prime-age males in districts with urban population of 100,000 or more.

(2) and (3) are samples restricted to male workers aged 25–55 from districts with urban population above 100,000. (4) and (5) are samples restricted to districts with an urban population above 100,000 and include both genders aged 15–65.

(2) and (4) control for state dummies, individual age, and educational attainment by categories.

(3) and (5) are additional controls for potential years in the labor market, squared potential years in the labor market, industry dummies, and district characteristics.

Source: Authors' estimates based on 2005 Economic Census, Town Directory of the 2001 Population Census, and National Sample Survey Schedule 10. Round 61 (2004–2005).

We also estimate individual wage regressions at the district level separately for manufacturing and service sectors. This exercise allows us to check whether agglomeration effects are the same across economic sectors. The results presented in Table A.4 show that agglomeration effects are stronger for services than for manufacturing. However, similar to the aggregate estimates, both sector-specific estimates are sensitive to additional controls at the district level.

Table 4. Ordinary Least Squares Estimates of Elasticities of Derived Average City Wages—City Population and Density, 2005

Variables	Log Average City Wages			
	(1)	(2)	(3)	(4)
Corrected log population	0.0329*** (0.00613)	0.0258*** (0.00775)		
Corrected log density			0.0348*** (0.0119)	0.0273** (0.0124)
<i>City characteristics</i>				
Minimum temperature in centigrade		0.000362 (0.000242)		0.000346 (0.000242)
Maximum temperature in centigrade		-0.000218 (0.000289)		-0.000177 (0.000284)
Log average rainfall in millimeters		-0.00398 (0.00689)		-0.00401 (0.00684)
Log number of electricity connections		-1.77e-05 (0.000383)		4.44e-05 (0.000385)
Log paved road length in kilometers		0.00176*** (0.000602)		0.00190*** (0.000601)
Log distance in kilometers to state headquarters		-0.00239* (0.00140)		-0.00238* (0.00141)
Log number of educational institutions		0.00361*** (0.000929)		0.00383*** (0.000936)
Observations	2,779	2,723	2,776	2,723
R <sup>2</sup>	0.992	0.992	0.992	0.992

Notes: All regressions include district dummies. Boot-strapping district-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census, Town Directory of the 2001 Population Census, and National Sample Survey Schedule 10. Round 61 (2004–2005).

In summary, we are able to replicate the main results found in the recent literature by using the same model specification, which suggests that there are significant positive agglomeration effects at the district level in India. Meanwhile, the estimated effects seem quite sensitive to the model specification, which may be partly due to the fact that districts are not well-defined economic units in India.

## B. The Effects of Agglomeration on Wages: City-Level Analysis

Table 4 presents OLS estimates of the elasticities of wages with respect to city population and population density, applying the methodology described in section IV. Columns (1) and (3) show that, when town characteristics are not controlled for, a 10% increase in town population or urban density will lead to a 3.3%–3.5% increase in a town's average wages. Both estimates are statistically significant at the 1% level. While the estimates fall in the broad range of agglomeration effects found in the literature, they are half the size estimated by Chauvin et al. (2017) for India. One possible driver for the difference could be the

different geographic scales examined—a town in our case and a district in that of Chauvin et al. (2017).

Columns (2) and (4) present models controlling for town characteristics. We use data on the annual minimum and maximum temperatures and average rainfall to measure climatic amenities, electricity connections, and paved road length to measure infrastructure availability, distance to state headquarters as a proxy for market access, and the number of educational institutions as a proxy for opportunities to accumulate human capital. The estimated coefficients of log population and log density drop to 2.6%–2.7% when these town characteristics are included in the models, suggesting that there are city features that are correlated with productivity and attract more workers to a city.<sup>11</sup>

In addition to addressing the endogeneity concern about omitted city characteristics, the controlled variables themselves are of interest in understanding urban agglomeration. The results show that road length and number of educational institutions have positive and statistically significant effects on urban wages, and the farther away a town is located from the state headquarters, the lower the urban wage. To the extent that the availability of roads and distance to the state center proxy for a town's market access, the results are consistent with the idea that better market access increases firm productivity through higher demand for outputs and cheaper intermediate inputs. The results support the view that local educational resources play a positive role in raising citywide productivity. The results also show that city wages are positively associated with more desirable climates (e.g., higher annual minimum temperature, lower annual maximum temperature, and lower average rainfall). However, none of the correlations are statistically significant.

To further mitigate the endogeneity of city population and density, we instrument these variables with their corresponding values in 1981. Table 5 reports the estimated elasticities of wages with respect to city population and density using instruments based on 1981 data.<sup>12</sup> The upper panel presents the first stage estimates and the lower panel presents the two-stage least squares estimates. Again, we show models with and without town characteristics.

First, the instrumental variables are powerful and robust in explaining contemporaneous city population or density. The IV coefficients are quite stable at around 0.72 regardless of whether it is population or density being instrumented, and whether town characteristics are included or not.

The two-stage least squares estimates of the agglomeration effects are not statistically significant as the estimated standard errors are double those

<sup>11</sup>We also added (log) town area to the models and found (i) area has a positive but smaller effect than that of density by one order of magnitude, and (ii) adding area has little impact on the coefficient estimates of density.

<sup>12</sup>Using 1951 population or density as instrumental variables yields zero or negative elasticity of wages to city scale. (These results are available upon request.) This may be due to the fact that one-third of towns in the Town Directory of the 2001 Population Census have 1951 population missing. This is less an issue for 1981 since the number of missing values is much smaller. Hence, we focus on results with instruments using data from 1981.

Table 5. Instrumental Variable Estimates of Elasticities of Derived Average City Wages—City Population and Density, 2005

Variables	Instrumental Variables 1981			
	(1)	(2)	(3)	(4)
<b>First-stage estimates</b>				
Corrected log population	0.724*** (0.0895)	0.726*** (0.0895)		
Corrected log density			0.714*** (0.0924)	0.720*** (0.0909)
City characteristics controls	No	Yes	No	Yes
F-statistic	65.35	65.75	59.79	62.67
<b>Two-stage least squares estimates</b>				
Corrected log population	0.0229 (0.0167)	0.0135 (0.0157)		
Corrected log density			0.0212 (0.0266)	0.0102 (0.0248)
<i>City characteristics</i>				
Minimum temperature in centigrade		0.000346 (0.000254)		0.000335 (0.000252)
Maximum temperature in centigrade		−0.000170 (0.000246)		−0.000139 (0.000245)
Log average rainfall in millimeters		−0.00374 (0.00669)		−0.00368 (0.00654)
Log number of electricity connections		6.43e-05 (0.000421)		0.000114 (0.000423)
Log paved road length in kilometers		0.00185*** (0.000615)		0.00193*** (0.000631)
Log distance in kilometers to state headquarters		−0.00240 (0.00155)		−0.00240 (0.00155)
Log number of educational institutions		0.00386*** (0.00100)		0.00402*** (0.000993)
Observations	2,779	2,723	2,776	2,723
R <sup>2</sup>	0.992	0.992	0.992	0.992

Notes: All regressions include district dummies and a constant. District-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census, Town Directory of the 2001 Population Census, and National Sample Survey Schedule 10. Round 61 (2004–2005).

of the OLS estimation. Considering coefficient estimates only, the elasticity of wages to population or density drops by about 1 percentage point to around 2% without controlling for town characteristics. When city characteristics that may be correlated with both city scale and productivity are taken into account, the point estimates of the wage elasticity decrease further to 1.4% for population and about 1% for density. In contrast to the results in Chauvin et al. (2017), our IV estimates suggest that the agglomeration benefits in India seem to be quite small and not statistically distinguishable from zero.

As far as the estimates of city characteristics are concerned, the coefficients of paved road length, distance to the state headquarters, and number of educational

institutions remain qualitatively the same in the IV estimated models as in the OLS estimation. More paved roads and educational institutions lead to higher wages, and being located farther from state headquarters lowers wages.

We also include town-level measures of industrial diversity and specialization, share of manufacturing employment, and share of employment from firms with 10 or more employees as control variables and present the results in Table 6.<sup>13</sup> Using OLS, the elasticities of wages to city population and density are estimated at 2.4% and 2.8%, respectively. The IV estimate of wage elasticity to city population is slightly lower at 1.8% and statistically significant at the 5% level; the wage elasticity to city density is around 2%, but is not statistically significant.

All four city-level industry-related variables added to the models are positively correlated with average town wages and are statistically significant, except the share of manufacturing. In other words, town-level wages are higher when towns are both more diversified and specialized, have larger manufacturing sectors, and host relatively larger firms. The estimated coefficients of city characteristics do not change qualitatively although they are often smaller in magnitude and weaker in terms of statistical significance. This may be because the added town-level industrial variables pick up the effects of these city characteristics.

### C. Alternative Approach

We have argued that urban agglomeration effects should be assessed at the city level because districts are not the proper unit for defining cities in India. Since we do not observe wages at the firm or city level in the EC 2005 data, we construct the city-level wage variable based on the average district-industry wages and the distribution of industries across towns and sectors. Our estimation shows that Indian cities have generated limited agglomeration effects that are smaller than the estimates obtained at the district level.

Ciccone and Hall (1996) offer another approach to address the problem that a measure of productivity such as wages may be missing at the city level. They are interested in examining the extent to which agglomeration economies at the county level in the United States may be driving state-level gross domestic product per worker. Essentially, they construct a density index at the state level as a nonlinear function of density of constituent counties. Recognizing that the state–county relationship is similar to our district–city relationship, their approach entails estimating the following regression:

$$\log y_d = \alpha + \eta \log e_d + \log D_d(\theta) + u_d \quad (7)$$

---

<sup>13</sup>We define diversity and specialization indexes based on examples from the literature. See the Appendix for details.

Table 6. Estimates of Elasticities of Derived Average City Wages—City Population and Density with City Industrial Variables, 2005

Variables	Ordinary Least Squares		Instrumental Variable 1981	
	(1)	(2)	(3)	(4)
<b>First-stage results</b>				
Corrected log population			0.746*** (0.0872)	
Corrected log density				0.743*** (0.0868)
City characteristics controls			Yes	Yes
F-statistic			73.28	73.21
<b>Two-stage least squares results</b>				
Corrected log population	0.0239*** (0.00553)		0.0178** (0.00883)	
Corrected log density		0.0279*** (0.00851)		0.0194 (0.0139)
Diversity Index	0.00925*** (0.00203)	0.00963*** (0.00206)	0.00954*** (0.00203)	0.00986*** (0.00206)
Specialization Index	0.000123** (5.53e-05)	0.000123** (5.55e-05)	0.000122** (4.90e-05)	0.000121** (4.89e-05)
Share of manufacturing employment to total employment	0.00846 (0.00691)	0.00739 (0.00675)	0.00861 (0.00834)	0.00790 (0.00822)
Large firms employment share	0.0552*** (0.00604)	0.0564*** (0.00610)	0.0558*** (0.00702)	0.0569*** (0.00702)
<i>City characteristics</i>				
Minimum temperature in centigrade	0.000327 (0.000235)	0.000314 (0.000235)	0.000321 (0.000253)	0.000310 (0.000250)
Maximum temperature in centigrade	-7.96e-05 (0.000314)	-4.83e-05 (0.000311)	-5.62e-05 (0.000273)	-2.95e-05 (0.000272)
Log average rainfall in millimeters	-0.00250 (0.00429)	-0.00261 (0.00434)	-0.00243 (0.00415)	-0.00249 (0.00414)
Log number of electricity connections	-0.000155 (0.000362)	-0.000123 (0.000358)	-0.000125 (0.000386)	-9.59e-05 (0.000385)
Log paved road length in kilometers	0.000921 (0.000603)	0.00101* (0.000606)	0.000946 (0.000580)	0.00101* (0.000586)
Log distance in kilometers to state headquarters	-0.00127 (0.00145)	-0.00123 (0.00146)	-0.00126 (0.00151)	-0.00124 (0.00151)
Log number of educational institutions	0.00182** (0.000833)	0.00192** (0.000840)	0.00191** (0.000883)	0.00199** (0.000884)
Observations	2,723	2,723	2,723	2,723
R <sup>2</sup>	0.993	0.993	0.993	0.993

Notes: All regressions include district dummies and a constant. District-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census, Town Directory of the 2001 Population Census, and National Sample Survey Schedule 10. Round 61 (2004–2005).

where  $y_d$  is an aggregate productivity measure (average district wage in our case),  $e_d$  is aggregate education of labor force, and  $D_d(\theta)$  is the density index function defined as

$$D_d(\theta) = \frac{1}{N_d} \sum_c a_c D_c^\theta$$

where  $N_d$  is employment in district  $d$ ,  $a_c$  is the area of constituent city  $c$ , and  $D_c$  is the density of city  $c$ .  $\theta$  is the key parameter to be estimated, with  $\theta > 1$  implying a positive agglomeration effect. Note that function  $D_d(\theta)$  is distinct from the multiplication of log of linear average of city density and  $\theta$ , which is commonly used in district-level agglomeration models.

We apply the Ciccone and Hall (1996) approach and obtain  $\hat{\theta} = 0.94$  with a standard error equal to 0.031. This suggests that there is unlikely to be appreciable agglomerations effects among Indian cities. We consider this result as being supportive of our results reported in section V.

#### D. Employment Shares of Formal Firms

As noted in the introduction and in section IV, we also estimate how employment shares of firms above a certain scale vary in response to city population and density. The main rationale is that these firms are more productive than microenterprises and hence the share of employment accounted for by these firms is a proxy for the average productivity of the city. As noted earlier, we choose 10 employees as the cutoff point to define scale—in line with various regulations governing the industrial workplace and trade union activity—and calculate the share of total employment that is hired by firms with 10 or more employees for each town.

Table 7 presents OLS estimates of the coefficients of log city population and density on the employment share of firms with 10 or more employees. First, city population and density show distinct effects on the dependent variable. When town variables are not controlled, urban population has a significantly positive impact on the employment share, while urban density has a small positive effect that is not distinguishable from zero. When city size doubles, the share of firms with 10 or more employees increases by 2.9 percentage points, or by about 13% as the average share is 22%. The impact of city size falls to about 1% and is statistically insignificant when town characteristics are controlled for in the model. On the other hand, density has a negative and statistically significant effect on the presence of the relatively large firms when town characteristics are accounted for. When urban density doubles, the employment share of the relatively large firms decreases by nearly 1 percentage point.

One plausible interpretation of the results is that firms benefit from better input–output linkages, thicker local labor markets, and improved learning in larger cities. However, instead of using the family house or backyard, formal firms need relatively large tracts of land in appropriately zoned areas to operate on. When city density increases, land and property prices may be expected to rise and outweigh the agglomeration benefits of city size, and thus bid some firms out of the market. Indeed, the relatively rapid growth of manufacturing in peri-urban and rural areas documented by Ghani, Goswami, and Kerr (2012) lends support to this possibility.

Table 7. Ordinary Least Squares Estimates of Effects of City Population and Density on Employment Shares of Firms with 10 or More Employees, 2005

Variables	Share of Employment among Firms with 10 or More Employees			
	(1)	(2)	(3)	(4)
Log population	0.0290*** (0.00270)	0.00941 (0.00919)		
Log density			0.00258 (0.00355)	-0.00928** (0.00374)
<i>City characteristics</i>				
Minimum temperature in centigrade		-0.000672 (0.00135)		-0.000668 (0.00134)
Maximum temperature in centigrade		-0.00285 (0.00191)		-0.00266 (0.00188)
Log average rainfall in millimeters		0.00468 (0.0173)		0.00501 (0.0169)
Log number of electricity connections		-0.00153 (0.00321)		0.00138 (0.00274)
Log paved road length in kilometers		0.0139** (0.00552)		0.0179*** (0.00397)
Log distance in kilometers to state headquarters		-0.0143 (0.00887)		-0.0145 (0.00886)
Log number of educational institutions		0.0116** (0.00560)		0.0168*** (0.00461)
Observations	2,520	2,468	2,510	2,461
R <sup>2</sup>	0.572	0.563	0.540	0.566

Notes: All regressions include district dummies and a constant. District-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census and Town Directory of the 2001 Population Census.

While this “spread” of manufacturing is in some ways equalizing and natural, it may well lead to suboptimal economic outcomes if it undercuts economies of agglomeration and leads to locational decisions that discount proximity to input suppliers and markets—features that are widely believed to be key drivers of sustained growth of manufacturing and associated service industries, and thus the creation of modern jobs.<sup>14</sup> Similar effects may also occur in the service sector more generally.

The coefficients on the city characteristics show that there is more employment in formal firms if the city has better transport infrastructure, as measured by paved road length, and more educational institutions. These estimates are qualitatively consistent with the models with derived average city wages as the

<sup>14</sup>As Henderson (2014, 9) notes, “We have painted decentralization here as a positive development. However, there may be premature decentralization in India and a loss of agglomeration benefits. Firms are driven from cities because of poor environments: poorly allocated infrastructure investments, a lack of public utilities, and inappropriate land market regulations. Bertaud and Brueckner (2005) discuss how land market regulations (limiting floor-area ratios) in Mumbai have led to sprawl and inefficiently low densities near the city center. This may result in a costly lessening of agglomeration benefits.”

Table 8. Instrumental Variable Estimates of Effects of City Population and Density on Employment Shares of Firms with 10 or More Employees, 2005

Variables	Instrumental Variable 1981			
	(1)	(2)	(3)	(4)
<b>First-stage estimates</b>				
Log population	0.992*** (0.0285)	0.761*** (0.0887)		
Log density			0.925*** (0.0325)	0.902*** (0.0346)
City characteristics controls	No	Yes	No	Yes
F-statistic	1,215	73.51	811.90	679.80
<b>Two-stage least squares estimates</b>				
Log population	0.0300*** (0.00227)	0.0144* (0.00738)		
Log density			0.00147 (0.00326)	-0.00818** (0.00338)
<i>City characteristics</i>				
Minimum temperature in centigrade		-0.000647 (0.00121)		-0.000676 (0.00119)
Maximum temperature in centigrade		-0.00293* (0.00169)		-0.00267 (0.00167)
Log average rainfall in millimeters		0.00474 (0.0154)		0.00506 (0.0150)
Log number of electricity connections		-0.00258 (0.00271)		0.00126 (0.00243)
Log paved road length in kilometers		0.0121** (0.00484)		0.0178*** (0.00353)
Log distance in kilometers to state headquarters		-0.0139* (0.00802)		-0.0146* (0.00787)
Log number of educational institutions		0.00948** (0.00474)		0.0167*** (0.00410)
Observations	2,520	2,468	2,510	2,461
R <sup>2</sup>	0.572	0.563	0.540	0.566

Notes: All regressions include district dummies and a constant. District-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census and Town Directory of the 2001 Population Census.

dependent variable, implying that two dependent variables do measure something in common.

Table 8 reports IV estimates for the models of employment shares of relatively large firms. The results are essentially the same as the OLS estimates. City population has a significant positive effect—with a coefficient estimated at 0.03—on the shares of these firms when 1981 population is used as the instrument variable. When city characteristics are included, the estimated population effect of 1.4% is slightly higher than the OLS estimates and is significant at the 10% level. When city density replaces population and town characteristics are taken into account, the IV estimates are similar to the OLS estimates, confirming that higher city density leads to fewer large firms. This is consistent with the possibility that higher land and

property prices resulting from increased urban density can outweigh the benefits of urban agglomeration and drive some productive enterprises out of the local market. This does not necessarily contradict the finding that higher density is associated with higher local wages in that those driven out of the dense cities are likely to be less productive among the formal firms.

The estimated coefficients of city variables are largely consistent with their OLS counterparts. In particular, the paved road length and number of educational institutions show robust positive effects on the presence of formal firms. Notably, the results suggest that larger firms are more likely to be located in cities closer to the state headquarters.

From Table 9, we see that the above results (e.g., the positive effects of city population and negative effects of city density on the employment shares of larger and formal firms) hold when a city's diversity, specialization, and employment share of manufacturing are added to the models. Not surprisingly, the specialization index and diversity measure are positively and negatively correlated with the dependent variables, respectively. However, our estimates show that the greater the share of manufacturing employment in a city, the smaller the share of employment from formal firms, though the relationship is not statistically significant.

## VI. Conclusions

To the best of our knowledge, this paper provides the first evidence of agglomeration effects in India at the city or town level. We examine two outcome variables: (i) a measure of average town wages and (ii) employment shares of formal firms (firms with 10 or more employees). The former is derived from individual earnings information provided in the LFS and the sectoral distribution of employment at the city level, while the latter is computed with micro data from the EC 2005. We address potential endogeneity issues by including relevant city characteristics and instrumenting the contemporaneous city population and density with historical values.

We find positive though small elasticity of wages with respect to town population and density. The estimates we find most credible suggest that the average wage will increase by 1%–2% if the city size or density doubles. This agglomeration effect is smaller than those from the recent literature (e.g., Chauvin et al. 2017) and from our replication exercise, which estimates the effect at 7%–8%. We think the different unit of analysis—cities in our case and districts in the literature—drives some of the difference in the estimates. We also find that city size has a positive effect and density has a negative effect on the presence of formal, or relatively large, firms. This suggests that agglomeration benefits for some firms may be offset by rising land and property prices as cities become more dense.

Our results on city characteristics generally conform to the literature. Higher city productivity is associated with better infrastructure and geographic location,

Table 9. Estimates of Effects of City Population and Density on Employment Shares of Firms with 10 or More Employees with City Industrial Variables, 2005

Variables	Ordinary Least Squares		Instrumental Variable 1981	
	(1)	(2)	(3)	(4)
<b>First-stage results</b>				
Log population			0.756*** (0.089)	
Log density				0.902*** (0.0347)
City characteristics controls			Yes	Yes
F-statistic			72.12	675.20
<b>Two-stage least squares results</b>				
Log population	0.0163* (0.00846)		0.0215*** (0.00719)	
Log density		-0.00592* (0.00330)		-0.00493* (0.00298)
Diversity Index	-0.0700*** (0.0104)	-0.0676*** (0.0105)	-0.0705*** (0.00929)	-0.0676*** (0.00937)
Specialization Index	0.000603*** (0.000173)	0.000598*** (0.000172)	0.000603*** (0.000154)	0.000599*** (0.000153)
Share of manufacturing employment to total employment	-0.0797 (0.0567)	-0.0732 (0.0548)	-0.0825 (0.0504)	-0.0734 (0.0488)
<i>City characteristics</i>				
Minimum temperature in centigrade	-0.00127 (0.00139)	-0.00130 (0.00138)	-0.00124 (0.00124)	-0.00131 (0.00123)
Maximum temperature in centigrade	-0.00263* (0.00149)	-0.00235 (0.00147)	-0.00271** (0.00132)	-0.00236* (0.00131)
Log average rainfall in millimeters	0.0193 (0.0144)	0.0193 (0.0142)	0.0194 (0.0129)	0.0193 (0.0126)
Log number of electricity connections	-0.000584 (0.00298)	0.00332 (0.00257)	-0.00168 (0.00250)	0.00321 (0.00226)
Log paved road length in kilometers	0.0129*** (0.00492)	0.0191*** (0.00361)	0.0111** (0.00441)	0.0191*** (0.00322)
Log distance in kilometers to state headquarters	-0.0144* (0.00765)	-0.0154** (0.00757)	-0.0139** (0.00689)	-0.0155** (0.00672)
Log number of educational institutions	0.0165*** (0.00554)	0.0240*** (0.00446)	0.0143*** (0.00478)	0.0239*** (0.00396)
Observations	2,468	2,461	2,468	2,461
R <sup>2</sup>	0.620	0.620	0.620	0.620

Notes: All regressions include district dummies and a constant. District-cluster standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on 2005 Economic Census and Town Directory of the 2001 Population Census.

as well as opportunities for accumulating human capital through schools. We also find that industrial diversity and specialization and the employment share of manufacturing in a city are all positively associated with city wages.

From a policy perspective, our results serve to provide a note of caution in line with the arguments of Ahluwalia, Kanbur, and Mohanty (2014) and Henderson (2010). While cities are widely believed to be engines of growth and good jobs, the

mere fact of an expansion in urban agglomerations does not necessarily mean that agglomeration benefits will flow automatically. Perhaps cities require a certain level and type of planning and infrastructure investments to play the beneficial role many economists and policy makers have come to expect of them. Without such planning and infrastructure investments, Indian cities may well fail to capture agglomeration benefits.

In closing, the use of Economic Census data for research on policy issues is relatively new. With its universal coverage and the location of firms at the town level, the data allow us to study urban economic topics in a developing country context in considerable depth.

## References

- Ahluwalia, Isher Judge, Ravi Kanbur, and Prasanna Kumar Mohanty. 2014. "Challenges of Urbanisation in India: An Overview." In *Urbanisation in India: Challenges, Opportunities, and the Way Forward*, edited by Isher Judge Ahluwalia, Ravi Kanbur, and Prasanna Kumar Mohanty, 1–28. New Delhi: SAGE Publications.
- Asian Development Bank. 2009. *Key Indicators 2009: Enterprises in Asia—Fostering Dynamism in SMEs*. Manila.
- Bertaud, Alain, and Jan K. Brueckner. 2005. "Analyzing Building-Height Restrictions: Predicted Impacts and Welfare Costs." *Regional Science and Urban Economics* 35 (2): 109–25.
- Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India, and the United States." *Journal of Urban Economics* 98: 17–49.
- Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1): 54–70.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011. "The Identification of Agglomeration Economies." *Journal of Economic Geography* 11 (2): 253–66.
- Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics, Volume 5*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 247–348. Amsterdam: Elsevier.
- Duranton, Gilles. 2015. "Growing through Cities in Developing Countries." *The World Bank Research Observer* 30 (1): 39–73.
- Ghani, Ejaz, Arti Goswami, and William Kerr. 2012. "Is India's Manufacturing Sector Moving Away from Cities?" NBER Working Paper No. 17992.
- Ghani, Ejaz, Ravi Kanbur, and Stephen D. O'Connell. 2013. "Urbanization and Agglomeration Benefits: Gender Differentiated Impacts on Enterprise Creation in India's Informal Sector." World Bank Policy Research Working Paper No. 6553.
- Hasan, Rana, Nidhi Kapoor, Aashish Mehta, and Asha Sundaram. 2017. "Labor Regulations, Employment, and Wages: Evidence from India's Apparel Sector." *Asian Economic Policy Review* 12 (1): 70–90.
- Henderson, J. Vernon. 2010. "Cities and Development." *Journal of Regional Science* 50 (1): 515–40.
- \_\_\_\_\_. 2014. "Urbanization and the Geography of Development." World Bank Policy Research Working Paper No. 6877.

- Lall, Somik V., Zmarak Shalizi, and Uwe Deichmann. 2004. "Agglomeration Economies and Productivity in Indian Industry." *Journal of Development Economics* 73 (2): 643–73.
- McKinsey Global Institute. 2010. *India's Urban Awakening: Building Inclusive Cities Sustaining Economic Growth*. New York: McKinsey & Company.
- Melo, Patricia, Daniel Graham, and Robert Noland. 2009. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics* 39 (3): 332–42.
- National Sample Survey Organization. 2005. "National Sample Survey 2004–2005 (61st round) Schedule 10—Employment and Unemployment." Government of India, Ministry of Statistics and Programme Implementation. [http://mospi.nic.in/?status=3&menu\\_id=74](http://mospi.nic.in/?status=3&menu_id=74) (accessed February 1, 2016).
- Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." *Handbook of Regional and Urban Economics* 4: 2119–71.

## Appendix

### Matching Towns in the Town Directory of the 2001 Population Census and the 2005 Economic Census

The spatial boundaries used in the Town Directory of the 2001 Population Census and the 2005 Economic Census do not match one-to-one at the town level due to changes in local administrative boundaries. We used secondary sources such as the Administrative Atlas of India, pin code database, digital maps, and other online resources to track geographical variabilities between 2001 and 2005. This effort resulted in a one-to-one mapping of location codes in which all observations in the Economic Census were linked to their corresponding equivalent in the Town Directory of the 2001 Population Census.

### Definitions of Town-Level Variables

- (i) Total electricity connections refer to a city's electricity supply, which covers the total number of connections for domestic, industrial, and commercial purposes.
- (ii) *Pacca* roads (in the Town Directory) refer to paved roads. The latter term is used in the paper.
- (iii) Number of educational institutions covers colleges, universities, and polytechnics located in a city.
- (iv) Diversity and specialization indexes are defined as  $Div_c = 1 / \sum_j |s_{jc} - s_j|$  and  $Spec_c = \max_j (s_{jc} / s_j)$ , respectively, wherein  $s_{jc}$  refers to the share of

industry  $j$  in city  $c$ 's total employment, and  $s_j$  is industry  $j$ 's share in national total employment. Calculations for these indexes are based on a 59-industry breakdown of the two-digit National Industrial Classification 2004.

### Definitions of District-Level Variables

- (i) Minimum temperature: lowest value of minimum temperature among all towns within the district
- (ii) Maximum temperature: highest value of maximum temperature amongst all towns within the district
- (iii) Average rainfall: average of average rainfall of all towns within the district
- (iv) Number of electricity connections: sum of electricity connections across towns within the district
- (v) Paved road length: sum of paved road length across towns within the district
- (vi) Number of educational institutions: sum educational institutions across towns within the district
- (vii) Distance to state headquarters: population-weighted average distance to state headquarters across towns within the district

Table A.1. 2004 Industrial Categorization

Broad Categorization	Classification at the Two-Digit Level
Primary	1 Agriculture, hunting and related service activities
	2 Forestry, logging and related service activities
	5 Fishing, aquaculture, and service activities incidental to fishing
Mining	10 Mining of coal and lignite; extraction of peat
	11 Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying
	12 Mining of uranium and thorium ores
	13 Mining of metal ores
	14 Other mining and quarrying
Labor-intensive manufacturing	15 Manufacture of food products and beverages
	16 Manufacture of tobacco products
	17 Manufacture of textiles
	18 Manufacture of wearing apparel; dressing and dyeing of fur

*Continued.*

Table A.1. *Continued.*

Broad Categorization	Classification at the Two-Digit Level
Capital-intensive manufacturing	19 Tanning and dressing of leather harness and footwear; manufacture of luggage, handbags, saddlery,
	20 Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
	26 Manufacture of other nonmetallic mineral products
	28 Manufacture of fabricated metal products, except machinery and equipment
	36 Manufacture of furniture; manufacturing n.e.c.
	37 Recycling
	21 Manufacture of paper and paper products
	22 Publishing, printing, and reproduction of recorded media
	23 Manufacture of coke, refined petroleum products and nuclear fuel
	24 Manufacture of chemicals and chemical products
	25 Manufacture of rubber and plastics products
	27 Manufacture of basic metals
	29 Manufacture of machinery and equipment n.e.c.
	30 Manufacture of office, accounting, and computing machinery
	31 Manufacture of electrical machinery and apparatus n.e.c.
	32 Manufacture of radio, television and communication equipment and apparatus
	33 Manufacture of medical, precision and optical instruments, watches and clocks
	34 Manufacture of motor vehicles, trailers, and semi-trailers
	35 Manufacture of other transport equipment
Utility	40 Electricity, gas, steam, and hot water supply
	41 Collection, purification, and distribution of water
Construction	45 Construction
Sales and trade	50 Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of automotive fuel
	51 Wholesale trade and commission trade, except of motor vehicles and motorcycles
	52 Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
Hotels and restaurant	55 Hotels and restaurants
Transport and telecommunication	60 Land transport; transport via pipelines
	61 Water transport
	62 Air transport
	63 Supporting and auxiliary transport activities; activities of travel agencies
	64 Post and telecommunications
Finance, insurance, real estate, and rental	65 Financial intermediation, except insurance and pension funding

*Continued.*

Table A.1. *Continued.*

Broad Categorization	Classification at the Two-Digit Level	
	66	Insurance and pension funding, except compulsory social security
	67	Activities auxiliary to financial intermediation
	70	Real estate activities
	71	Renting of machinery and equipment without operator and of personal and household goods
Business services and research and development	72	Computer and related activities
	73	Research and development
	74	Other business activities
Public administration, health, and education	75	Public administration and defence; compulsory social security
	80	Education
	85	Health and social work
Other services	90	Sewage and refuse disposal, sanitation, and similar activities
	91	Activities of membership organizations n.e.c.
	92	Recreational, cultural, and sporting activities
	93	Other service activities
	99	Extraterritorial organizations and bodies

n.e.c. = not elsewhere classified.

Source: National Industrial Classification 2004.

Table A.2. **Number of Establishments by Industry in the 2005 Economic Census**

Industries	All Firms		Firms with 10 or More Employees	
	Number	%	Number	%
Capital-intensive manufacturing	483,153	2.91	24,959	6.94
Labor-intensive manufacturing	2,718,507	16.38	56,618	15.73
Business services and research and development	431,978	2.60	10,928	3.04
Construction	151,249	0.91	2,574	0.72
Finance, insurance, real estate, and rental	347,857	2.10	27,236	7.57
Hotels and restaurant	705,394	4.25	14,574	4.05
Mining	24,928	0.15	1,222	0.34
Other services	1,031,312	6.21	12,217	3.39
Primary	480,813	2.90	2,991	0.83
Public administration, health, and education	919,749	5.54	139,286	38.70
Sales and trade	8,217,526	49.51	47,880	13.30
Transport and telecommunication	1,061,985	6.40	15,487	4.30
Utility	23,004	0.14	3,915	1.09
Total	16,597,455	100.00	359,887	100.00

Source: Authors' compilation from the 2005 Economic Census (urban data full sample).

Table A.3. Individual Wage Regression Using Data from the National Sample Survey (Labor Force Survey), 2004–2005

Variables	Log Wages $\beta$	National Average Characteristics
Years of schooling	0.0873*** (2.03e-05)	8.5874
Gender (male = 1, female = 0)	0.361*** (0.000247)	0.8338
Potential years in the labor market	0.0579*** (2.44e-05)	20.2891
Squared potential years in the labor market	-0.000714*** (4.91e-07)	563.8307
Capital-intensive manufacturing	2.729*** (0.00899)	0.0907
Labor-intensive manufacturing	2.514*** (0.00898)	0.1798
Business services and research and development	2.948*** (0.00900)	0.0284
Construction	2.686*** (0.00898)	0.1144
Finance, insurance, real estate, and rental	3.032*** (0.00899)	0.0366
Hotels and restaurant	2.553*** (0.00900)	0.0274
Mining	3.218*** (0.00902)	0.0148
Other services	2.380*** (0.00900)	0.0201
Primary	2.373*** (0.00899)	0.0476
Public administration, health, and education	2.936*** (0.00898)	0.2190
Sales and trade	2.403*** (0.00898)	0.1120
Transport and telecommunication	2.808*** (0.00899)	0.0960
Utility	3.146*** (0.00902)	0.0132
District dummies	Yes	
Observations (nonweighted)	33,172	33,172
Observations with frequency weights	43,585,749	43,585,749
R <sup>2</sup>	0.986	

Note: Standard errors in parentheses. \*\*\* =  $p < 0.01$ , \*\* =  $p < 0.05$ , and \* =  $p < 0.1$ .

Source: Authors' estimates based on the National Sample Survey Schedule 10. Round 61 (2004–2005) (weighted to national level with National Sample Survey Organization sample weights).

Table A.4. Agglomeration Effects at the District Level by Industry

	Male Only				All Workers			
	Manufacturing		Services		Manufacturing		Services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>OLS regressions</b>								
Log of urban population	0.503* (0.0267) R <sup>2</sup> = 0.421	-0.0119 (0.0380) R <sup>2</sup> = 0.469	0.0829*** (0.0206) R <sup>2</sup> = 0.475	0.0376 (0.0368) R <sup>2</sup> = 0.536	0.0727*** (0.0230) R <sup>2</sup> = 0.502	0.0288 (0.0281) R <sup>2</sup> = 0.545	0.100*** (0.0195) R <sup>2</sup> = 0.517	0.0433 (0.0278) R <sup>2</sup> = 0.578
Log of density	0.0966* (0.0510) R <sup>2</sup> = 0.419	-0.0176 (0.0523) R <sup>2</sup> = 0.469	0.162*** (0.0445) R <sup>2</sup> = 0.473	0.0497 (0.0308) R <sup>2</sup> = 0.537	0.115** (0.0529) R <sup>2</sup> = 0.498	-0.0112 (0.0642) R <sup>2</sup> = 0.545	0.201*** (0.0378) R <sup>2</sup> = 0.514	0.0685** (0.0313) R <sup>2</sup> = 0.578
Observations	4,013	4,010	10,397	10,305	6,623	6,618	16,457	16,333
<b>IV-1981 regressions</b>								
Log of urban population	0.0395 (0.0246) R <sup>2</sup> = 0.421	-0.0713 (0.0700) R <sup>2</sup> = 0.468	0.0680*** (0.0170) R <sup>2</sup> = 0.474	-0.102** (0.0485) R <sup>2</sup> = 0.534	0.0572*** (0.0210) R <sup>2</sup> = 0.502	-0.0636 (0.0607) R <sup>2</sup> = 0.545	0.0858*** (0.0161) R <sup>2</sup> = 0.516	-0.0815*** (0.0262) R <sup>2</sup> = 0.576
Log of density	0.0461 (0.0490) R <sup>2</sup> = 0.418	-0.0601 (0.0611) R <sup>2</sup> = 0.468	0.0881** (0.0398) R <sup>2</sup> = 0.470	-0.0434 (0.0549) R <sup>2</sup> = 0.535	0.0461 (0.0538) R <sup>2</sup> = 0.496	-0.0741 (0.0615) R <sup>2</sup> = 0.545	0.128*** (0.0322) R <sup>2</sup> = 0.512	-0.0172 (0.0505) R <sup>2</sup> = 0.577
Observations	3,992	3,989	10,259	10,167	6,584	6,579	16,251	16,127
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	No	No	No	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	No	Yes	No	Yes	No	Yes
Labor market experience	No	Yes	No	Yes	No	Yes	No	Yes
District characteristics	No	Yes	No	Yes	No	Yes	No	Yes

IV = instrumental variables, OLS = ordinary least squares.  
 Notes: Robust standard errors in parentheses. Clustered at the state level. \*\*\* = p < 0.01, \*\* = p < 0.05, and \* = p < 0.1. Weighted to the national level with National Sample Survey Organization sample weights. District characteristics include climate, infrastructure availability, weighted distance to state headquarters of constituent cities or towns, and number of educational institutions. (1) to (4) are samples restricted to male workers aged 25–55 years old from districts with an urban population above 100,000. (5) to (8) are samples restricted to districts with an urban population above 100,000 and include both genders aged 15–65 years old. Odd-numbered regressions control for state dummies, individual age, and educational attainment by categories. Even-numbered regressions have additional controls for potential years in the labor market, squared potential years in the labor market, industry dummies, and district characteristics.  
 Source: Authors' estimates based on 2005 India Economic Census, 2001 Census of India, and National Sample Survey Schedule 10, Round 61 (2004–2005).