

Subaltern Urbanization: The Birth of Census Towns in West Bengal



Saumyabrata Chakrabarti and Vivekananda Mukherjee

1 Introduction

Urban growth in India over the last five decades can broadly be divided into four components, namely (a) growth due to natural increase in population; (b) growth due to net rural-urban migration; (c) growth due to reclassification from rural to urban settlement and (d) growth due to areal expansion. Figure 1 shows the share of these components in Indian urban growth in the last five decades.

From Fig. 1, as Chakrabarti and Mukherjee (2020) pointed out, the largest component of urbanization in India is the urbanization out of natural increase in population that explains more than 50% of urban growth, while the migration-driven urbanization stayed almost stable in the last five decades explaining only 20% of the same. The growth due to areal expansion had never exceeded 10% and remained particularly insignificant in the last decade, 2001–2011. On the other hand, reclassification-driven urbanization whose role remained marginal before 2001–2011, has become more prominent during the same period. As Chakrabarti and Mukherjee (2020) observed in the Indian Census the reclassification of villages in small towns (called Census Towns (CTs)) happens if a village satisfies the following criteria: (1) it has a population of 5000 or more; (2) its population density is at least 400/km²; and (3) 75% of its male main workforce working in non-farm sector. The CTs are administered

This is Chap. 3 of the Ph.D. thesis titled “Growth of Census Towns in India: An Economic Analysis” by Saumyabrata Chakrabarti submitted to Jadavpur University in 2022.

S. Chakrabarti

Department of Economics, Ramsaday College, Howrah, India

V. Mukherjee (✉)

Department of Economics and Finance, BITS Pilani, Hyderabad Campus, Hyderabad, India

e-mail: mukherjeevivek@hotmail.com

Department of Economics, Jadavpur University, Kolkata, India

**COMPONENTS OF URBAN GROWTH
(1961-2011)**

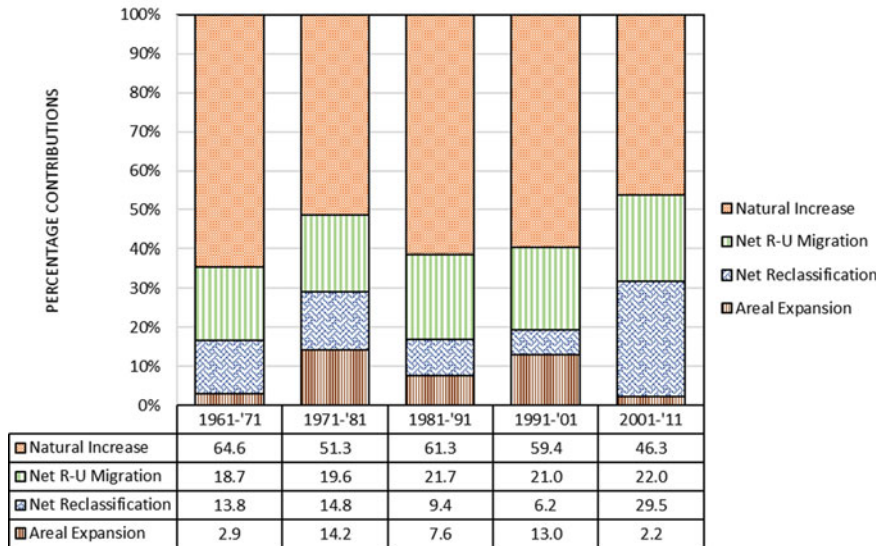


Fig. 1 Components of urban growth in India (1961–2011). *Source* Census of India (1971, 1981, 1991, 2001 and 2011); Chakrabarti and Mukherjee (2022)

by rural local governments.¹ As per Census (2011), the upsurge in number of CTs from 1362 to 3894 during the decade of 2001–2011, had contributed toward almost 30% of urbanization. Denis et al. (2012) call it ‘subaltern urbanization’. This chapter focuses on the determinants of this particular kind of urbanization.

There are three different kinds of theoretical argument that attempt to explain the dynamics of formation of the CTs. The first is related to the impact of developments in the large city in the neighborhood. The factors like fall in formal sector employment, the rise in congestion, house-rent and pollution may incentivize the existing population living in the large city to relocate in the villages bordering the city. It is also possible that the rural population, who looks for opportunities to relocate in the large city for better livelihood, for the same reasons mentioned above, are dissuaded in doing so. Some may resort to commuting rather than living within the city limit. The papers like Kundu (2011) and Sharma (2013) spell out such possibilities. The second is related to the development at the village level. The development of local infrastructure like schools, health facilities, stable electricity connection, development of banking facilities at the villages, generates forces of agglomeration at the village level. They not only facilitate relocation from neighboring villages but also

¹ In India the rural local governments are called ‘Panchayats’. The three tier Panchayat system consists of ‘Gram Panchayat’ at the bottom, ‘PanchayatSamiti’ at the middle and ‘ZillaParisad’ at the top.

from the large city located nearby. The papers like Aggarwal (2018) and Mukhopadhyay et al. (2016) take this view. Third, the development of transport infrastructure connecting the villages to the neighboring large city may play a role in the formation of CTs. The reduced transport cost may either decelerate the formation of the CTs by strengthening the forces of agglomeration toward the existing cities, or it may accelerate the same by strengthening the forces of dispersion away from the existing cities. The effect of transport infrastructure in the formation of CTs is discussed by papers like Ghani et al. (2012) and Chakrabarti and Mukherjee (2022).

This chapter attempts to find out which of the factors mentioned above explain the birth of CTs in West Bengal, an Eastern state of India, during 2001–2011. We work on West Bengal data, since West Bengal is the state which has experienced maximum increase in terms of number of CTs during 2001 and 2011, as shown in Table 1. It accounts for two-thirds of the urbanization in the state during the period.

The table above shows that all the major states have experienced a jump in the number of CTs between 2001 and 2011 with West Bengal leading the list. In

Table 1 Distribution and growth of census towns in Indian states

States	Number of census towns in 2011	Number of new census towns between 2001 and 2011
West Bengal	780	526
Kerala	461	346
Tamil Nadu	376	269
Uttar Pradesh	267	206
Maharashtra	278	171
Andhra Pradesh	228	137
Jharkhand	188	107
Odisha	116	86
Gujarat	153	83
Karnataka	127	81
Assam	126	80
Rajasthan	112	76
Punjab	76	55
Bihar	60	52
Haryana	74	49
Madhya Pradesh	112	46
Uttarakhand	41	29
Jammu and Kashmir	36	27
Chhattisgarh	14	10
Others	269	162
India	3894	2600

Source Census of India (2011) and author's calculation (Chakrabarti & Mukherjee, 2020, 2022)

West Bengal, out of 780 CTs 526 have born during 2001–2011. In earlier papers Chakrabarti and Mukherjee (2020, 2022) explored the role of neighboring city, development of village-specific factors, and transport infrastructure in the birth of CTs in West Bengal. Chakrabarti and Mukherjee (2020) explain the birth of CTs around existing cities by considering city-specific factors like its formal sector wage, population density, area, and transport cost from the neighborhood village to the city where the jobs are located. They find that high formal sector wage along with a low urban extension is conducive for the birth of Census Towns in the neighborhood of a city. Chakrabarti and Mukherjee (2022), on the other hand, found that the existence of state/national highways in the neighborhood increases the probability of a village, designated as a ‘would be Census Town’, being transformed in a Census Town, whereas the development of rail infrastructure did not stimulate such transformation. Further, Chakrabarti and Mukherjee (2022) noted that density of local road network in a district complements the highways in explaining the formation of the CTs for districts that are not adjacent to Kolkata, the capital city of West Bengal, indicating that in these districts the transport infrastructure creates the forces of dispersion away from the existing cities. On the contrary, the districts that are bordering Kolkata local roads oppose the dispersing influence of the highways in formation of CTs, and decrease the concentration of non-farm activities at the ‘would be CTs’. The paper by Chakrabarti and Mukherjee (2022) did not find significant impact of Golden Quadrangle (GQ) project, implemented during the study period for connecting the four major metropolitan cities of India, in birth of the Census Towns in West Bengal. They also found that the village-specific factors play a positive significant role in conversion of a village to a Census Town.

Since the city-specific, village-specific, and the transport infrastructure-related variables are correlated with each other as one causes the other, we cannot use them as explanatory variables in a single regression framework. However, from the policy perspective it is important to know their relative contribution in formation of Census Towns. The present chapter attempts to solve this problem. It uses Principal Component Analysis (PCA)² to reduce the dimensionality of the explanatory variables. The principal components are a blend of all the control variables used in Chakrabarti and Mukherjee (2020, 2022) and are orthogonal to each other. The varimax rotation of these variables identifies control variables, which have the highest correlation with each of the extracted principal components. We identify the principal components with the variable having the highest correlation with them and use them in the regression analysis to find out how they explain the conversion of villages in Census Towns in West Bengal. The chapter finds, in similar spirit as Chakrabarti and Mukherjee (2022), the explanatory factors play out differently in the districts bordering Kolkata, the capital city of West Bengal, and the other districts. The districts bordering Kolkata are North 24 Parganas, South 24 Parganas, and Haora. While the presence of highways within 5 km. radius of a village plays an important role in creation of CTs for both types of districts, in the districts bordering Kolkata density

² See papers like Savić (2006), Chen and Ma (2015), Nugrahadi et al. (2020), for applications of PCA in different context.

of population in the nearby city/Statutory Town (ST) plays a significant role. In the districts not bordering Kolkata the CTs are created away from the nearby STs and cities. It appears, among the city-specific factors, the importance of density that we find in this chapter is a new finding. It was not significant when Chakrabarti and Mukherjee (2020) analyzed the importance of city-specific factors in formation of CTs in West Bengal. Similarly, when Chakrabarti and Mukherjee (2022) analyzed the transport infrastructure-specific factors, nearness did not appear as a significant variable. Therefore, this is also a new finding of the chapter. We obtain the new results as we take all the three types of variables in a single framework and eliminate correlation between them. The results suggest that the forces of dispersion created from the congestion at the existing cities/STs are important in explaining the birth of CTs in West Bengal both at the districts around Kolkata and the other districts. The highways help the dispersion. In the districts outside Kolkata, commuting to the nearest city/ST is not important for formation of CTs. It highlights the process of the local agglomeration in and around the villages converted into CTs. However, in the districts bordering Kolkata, it seems the density of population in existing cities/STs plays an important role in dispersion process.

The plan of the chapter is as follows. The second section surveys the existing literature on the birth of Census Towns. Section 3 describes the methodology and the data and derives the results. The section following concludes the paper.

2 The Survey of Literature

Kundu (2011) discussed the issue of low rates of in-migration in metropolis and large cities like Delhi, Chandigarh, Kolkata, Hyderabad, Chennai, and Mumbai. According to this study, in urban metropolis various drives like removal of encroachments, shantytowns, petty commercial establishments, squatter settlements, and judicial interventions to curb the unplanned extension of urban growth caused the in-migration less attractive for poorer section of the population. Sharma (2013) points out that due to improved transport facilities workers travel daily to cities for jobs, which in turn led to the developments of CTs near the existing urban centers.

In a theoretical paper Marjit and Kar (2014) argue that in an open economy reform in labor market may lead to development of urban formal sector and shrinkage of urban informal sector coupled with a decline in wage in urban informal sector and an increase in the same in rural informal sector. These lead to a reverse migration in the economy with a fall in the average wage until there is a substantial growth of urban organized sector. Krugman and Elizondo (1996) with the example of Mexico point out that due to a shift of the economy from an inward looking strategy to liberalization of international trade the manufacturing sector scatters with the mitigation of backward and forward linkages in Mexico City to northern states adjacent to US border causing the dispersion of non-farm labor force in the economy. Zhu (2017) recognizes that the phenomenon of in situ urbanization in a wide range of areas in south eastern coastal

provinces of China after 1970s was an outcome of China's urbanization process after liberalization in line with Krugman and Elizondo (1996).

Among village centric views, according to Aggarwal (2018) development of roads and transport facilities in rural areas leads to a greater market integration, reduced price of non-local goods, and wider variety of consumption basket in rural areas and more participation of local youths in the expanded labor market. Mukhopadhyay et al. (2016) in the context of certain CTs in northern India observed that, improved transport and communication facilities along with growing rural income stimulate the growth of small scale non-tradable services which are the main sources of non-farm employment in these settlements.

There are some notable studies that discuss the role of transport infrastructure and urbanization. The study by Krugman (1991) on economic geography, argues that lowering of transport cost encourages agglomeration by attracting both labor and capital to the city. However, Helpman (1998) predicts that rising housing price due to rise in population in the larger region acts against agglomeration in the city, in favor of dispersion of economic activities and consequently the labor force. However, the technological progress in manufacturing sector may mitigate the dispersing forces (Zhu, 2017). Proost and Thisse (2019) argue that owing to the improvement in transport infrastructure and falling transport cost there is first agglomeration and then dispersion of production and consequent regional disparity and in the later stage regional integration occurs. Chandra and Thompson (2000), with the result of a study on the effect of highways show that the highways improve the income of the rural countries they go through and reduce the same in the adjacent countries in USA between 1969 and 1993. Baum-Snow et al. (2020) show with the example of China that investing in local transport infrastructure to promote the growth of hinterland often has an opposite impact of losing economic activities and specialization in agriculture. In Indian context, Ghani et al. (2012) found that district-level infrastructure is partly dispersing organized manufacturing to rural locations while the unorganized manufacturing is relocating to urban locations. Such movement seems to be partially explained by the development of national-level highways especially the construction of Golden Quadrangle, a highway project undertaken by the Government of India that was implemented in the study period. However, there is a very limited impact of Golden Quadrangle on unorganized manufacturing outside the nodal districts where more than one highways meet. Balakrishnan (2013) cited instances of Bangalore-Mysore highway and Pune-Nasik highway to establish that the urbanization along highways has been the emerging pattern of urbanization in developing countries. According to the study of Mahajan and Nagraj (2017) using NSSO 55th round (1999–2000), 66th round (2009–2010), and 68th round (2011–2012) datasets, the construction of highways and rural road networks during 2000–2012 boosted up construction demand and employment in rural areas. Chakrabarti and Mukherjee (2020, 2022) in search of the reasons behind the birth of CTs show the relative significance of certain city-specific, transport infrastructure-specific, and village infrastructure-specific variables in formation of CTs in the state of West Bengal in India. They find that factors like high formal sector wage rate at the nearest city of a village, the larger area of the city, development of highways and local roads

in the village-neighborhood, provision of banking facility, and availability of stable electricity connection at the village help conversion of the village in a CT.

3 The Principal Component Analysis

The Principal Component Analysis (PCA) is a technique that reduces the dimension of correlated variables into new variables, called principal components (PC) which are uncorrelated with each other and describe most of the information in the full dataset to explain its common variation.

The control variables used in the literature for explaining the birth of CTs are mainly of three different types: (1) variables related to transport infrastructure; (2) variables related to village-specific infrastructure; (3) variables related to the nearest city/statutory towns.

An improvement of transport infrastructure in a village can occur in three different ways: (i) development of a highway that connects cities may pass through the village neighborhood; (ii) improvement of rail-connectivity of the village to the cities; (iii) improvement of local road network near the village (Chakrabarti & Mukherjee, 2022). The improvement of transport infrastructure reduces the cost of traveling to the city, which may have both positive and negative influence on a village in terms of its conversion to a CT. On the one hand, it facilitates migration from village to city and reduces its chance of being converted to a CT; on the other, it shelters the firms and workers who relocate from the city for avoiding the congestion and increases its chance of being converted to a CT. The increased commuting of non-farm workers, due to ease of commuting, to the nearest city/ST also helps the conversion of a village to a CT. The set of variables related to transport infrastructure are:

HIGHWAY 5: A dummy variable that takes value of 1, if a highway passes through a neighborhood of 5 km radius around the center of village i in district j ;

RAIL 5: A dummy variable that takes value of 1, if there is at least one railway station within 5 km radius neighborhood around the center of village i in district j ;

NEARNESS: Reciprocal of the nearest city/ST distance from the center of village i in district j .

ROAD: The length of rural roads per 1000 km². area in district j . We assume uniform distribution of rural roads across all villages of the district.

The development of non-tradable services like availability of electricity services, rural banks, other financial services, etc., offers a better quality of life to village residents, creates more non-farm jobs, and therefore, attracts more population in it (Chakrabarti & Mukherjee, 2022). The set of variables related to local development are:

RBANK: A dummy variable that takes value of 1, if at least one branch of a commercial bank is present in village i at district j ;

POWER: Represents percentage of electrified villages in district j , where village i is located. This indicates the probability of availability of power in village i in district j .

The presence of an urban center in the neighborhood may impact CT dynamics meaningfully (Chakrabarti & Mukherjee, 2020). Let us describe the variables related to the developments at the nearest city/statutory towns. First, in the short run a rise in the formal wage although reduces the employment in urban formal sector, has an uncertain effect on the expected wage in the urban sector, and consequently on the CT dynamics. As the wage expected in the urban area rises, labor starts migrating to the city; the city with its defined boundary and density of population cannot accommodate the migrants from the rural area, causing them to revert back to the village and be absorbed in rural non-farm sector. However, the wage rate in the rural area rises and the farm sector is forced to shed-off labor. They find jobs in the rural non-farm sector causing non-farm employment to expand and helping a village to transform to a CT. On the contrary, if the expected wage in the city falls, reverse migration starts leading to a fall in rural wage. Then the farm sector with limited absorption capacity accommodates more labor compared to rural non-farm sector and the village ceases to transform to CT. Since the wage in the formal sector is credited through banks, the number of branches of commercial banks has been used as proxy variable for formal wage in the urban sector. Second, a fall in the population density in the city causes a fall in the participation in the urban informal sector. These labors find jobs in rural non-farm sector since the rural farm sector with limited absorption capacity does not employ them. Therefore, the villages near the city have greater chance to be transformed into CTs. Last, an expansion in the boundary of the city reduces the chance of a neighboring village being transformed into a CT. This happens as given the size of its formal sector, expansion of the city leads to an expansion of urban informal sector. The labor relocates from the rural non-farm sector to urban informal sector, which reduces the chance of the neighboring village transforming into a CT. The city-specific variables used in analysis are:

DENS: Population density in the nearest city/ ST; a proxy variable for ‘formalization and sanitization’ in the city/ST; a lower value of DENS implies more ‘formalization and sanitization’;

AREA: Area of the nearest city/ST;

UBANK: Number of commercial bank branches in the nearest city/ST, a proxy variable for formal wage in the nearest city.

The three types of variables described above are likely to be correlated with each other. For example, the improvement of transport infrastructure connecting a village to a city or in the neighborhood of a village may improve the village-specific infrastructure as agglomeration forces are created around the village. The number of bank branches can increase. The power supply can improve. It may also increase population density in the city as both the migration and the daily commuting to the city becomes less costly. The improvement of local infrastructure also leads to the development of transport infrastructure. The development of the city, which leads to congestion, as argued by Helpman (1998) obviates development of infrastructure away from the city. Therefore, we cannot use all these variables together in a regression analysis for finding their influence on the growth of Census Towns.

Hence the empirical exercise consists of two parts. The first part uses PCA to reduce the number of variables described above into few principal components (PCs)

which are the linear combination of original variables containing most of the information of original variables capturing majority of the variation in the dataset. The analysis helps us to extract those PCs which have the highest Eigen values. The varimax rotation of them tells us which variables have the highest association with the chosen PCs. Then, for finding out relative significance of the variables which have the highest association with chosen PCs (dominant variables) in explaining the formation of CTs, we run the following regression specification:

$$y_{ij} = \alpha + \mu\beta + D_j + \varepsilon_{ij} \quad (1)$$

In Eq. (1) the dependent variable y_{ij} represents the status of a village in Census 2001 which was identified as ‘would be CT’. It takes a value of 1 on the successful conversion of the village in Census 2011. Otherwise, it takes a value of 0. X_{ij} represents the set of dominant variables that indicate the dominant variable chosen from significant PCs at the i th village in the j th district. The district-specific fixed factors of district j , which are shared by all the villages located in district j , are captured through the dummy variable D_j . The unobserved village-specific factors are captured through ε_{ij} , which we assume to be independently identically distributed across the villages.

3.1 The Data

We use the dataset pertaining to CTs/Villages Directory of West Bengal for 2001 and 2011 extracted from the Census of India (2001 and 2011)) and also from the State Statistical Handbook of West Bengal for various years. Table 2 describes the data.

The descriptive statistics presented in Table 2 depicts that majority of the villages (80%) designated as ‘would be’ census town in 2001 got transformed into CT in 2011. Among the set of explanatory variables related to connectivity to city/local connectivity of village i , the dummy variable HIGHWAY5 has a mean of 0.78 implying 78% of such villages were located in the 5 km radius neighborhood of either nearest national highway or state highway (Chakrabarti & Mukherjee, 2022). Similarly, the dummy variable RAIL5 having the mean of 0.38 implies that 38% of the ‘would be CTs’ villages in the Census 2001 were located in the 5 km radius neighborhood of the nearest rail station. For the NEARNESS variable, for the sampled villages the mean is 0.11, with maximum of 2 and minimum of 0.01. There exists, an inverse relation between distance of a village from the nearest city/ST and the value of NEARNESS variable. The minimum distance that we have found in our data was 0.5 km (the reciprocal of 2) and the maximum distance was 100 km (the reciprocal of 0.01). The average distance was 9.09 km (the reciprocal of 0.11) (Chakrabarti & Mukherjee, 2022). For the variable ROAD, data was unavailable at the village level. Therefore, we used the data on the roads maintained by the local administrative bodies like Gram Panchayats, Panchayat Samitis, and Zilla Parishads aggregated at the district level as a proxy for the ROAD variable as in Chakrabarti and Mukherjee (2022). The

Table 2 Descriptive statistics

Variable name	Mean	Standard deviation	Minimum value	Maximum value
Dependent variable (=1, for villages transformed to CT) (=0, for villages not transformed to CT)	0.80	0.40	0	1
<i>Transport related variables</i>				
HIGHWAY5 (=1, neighborhood of highways within 5 km radius from the village center; 0, otherwise)	0.78	0.41	0	1
RAIL5 (=1, neighborhood of railheads within 5 km radius from the village center; 0, otherwise)	0.38	0.48	0	1
NEARNESS (reciprocal of the distance from nearest city/ST from the village)	0.11	0.18	0.01	2
ROAD (the length of rural road in the district 2007–2008 per 1000 km ²)	11.46	9.50	3.09	30.77
<i>Variables related to local non-traded services</i>				
RBANK (=1, if 1 commercial banks in the village 2007–2008; 0 otherwise)	0.51	0.50	0	1
POWER (percentage of electrified village in the district 2001)	93.7	11.86	53.7	100
<i>City/ST specific variables</i>				
DENS (population/ km ²) Population density of the nearest city/ST	12,618.62	6901.34	1884	24,841
AREA (km ²) Area of the nearest city/ST	56.11	62.92	5.85	185

(continued)

Table 2 (continued)

Variable name	Mean	Standard deviation	Minimum value	Maximum value
UBANK Number of bank branches of the nearest city/ST	189	369	15	1007

Source Census of India (2001 and 2011) and State Statistical Handbook of West Bengal (various years)

data has been normalized per 1000 km² area of the district. The average length of local roads per 1000 km² in the districts of West Bengal by 2007–08 was 11.46 km with a minimum of 3.09 km in the district of Darjeeling and maximum of 30.77 km in the district of South 24 Parganas. We assume that the road allocation is uniformly distributed among all villages in a district (Chakrabarti & Mukherjee, 2022).

Among the non-tradable services RBANK is taken at village level but in the absence of proper village-level data, the data on POWER is taken at the district level. RBANK is a binary variable with mean of 0.51 which implies that 50% of ‘would be CTs’ villages have had at least one commercial branch in 2007–2008 before the Census operation of 2011. On an average, 93.70% of villages were electrified across all districts in West Bengal in 2001 with a minimum of 53% and a maximum of 100%.

It is not that all the villages in West Bengal, which were identified as ‘would be CTs’ in Census 2001, had a nearby city with population exceeding 1 lac. In the absence of such cities we have used those urban bodies, which have the status of Statutory Towns (STs) and we have taken their population density and area as explanatory variable for our analysis. But, all STs are not large enough to trigger migration/commuting from nearby villages. For solving this problem, we looked at the distribution of bank branches at the city and STs of West Bengal and considered only those STs in which the number of bank branches is above the median (turns out as 15).

Among the city-specific attributes, population density of the nearest city/ST from a ‘would be CTs in 2001’ vary from 1884/km² to 24,841/km² with the mean at 12,618.6/km². The maximum is corresponding to the city of Kolkata. The area of such city/ST has a mean of 56.11 km² with a standard deviation 62.92. The number of branches of the banks in the nearest urban bodies from the ‘would be CTs in 2001’, which has been considered as a proxy for wage at the nearest urban locality, vary from 15 to 1007 with a mean of 189. The maximum number of bank branches belongs to the city of Kolkata. We take log transformation of the variables POWER, DENS, AREA, and UBANK while doing the empirical exercise.

We also run separate regressions for the districts bordering Kolkata and the other districts of West Bengal. The districts bordering Kolkata are North 24 Parganas, South 24 Parganas, and Haora. Of the 551 villages, which were ‘would be CTs in 2001’, 171 belonged to these districts.

3.2 *The Principal Components*

We perform the PCA with the variables described above. First, we calculate Eigen values and Eigen vector of each Principal Components. While the Eigen values represent the variances of the dataset, the Eigen vectors are the coefficients of the original variables in each principal component representing the correlation between a variable and the principal component. A PC qualifies for incorporation in the analysis if it has its Eigen value greater than one. Then we perform the orthogonal varimax rotation of the original Eigen vector matrix corresponding to Eigen values. The objective is to determine the association between the variables and corresponding principal components clearly which the original Eigen vector matrix fails to show in some cases. The rotated component vectors after varimax rotation represent clearly the correlation between the original variables and the chosen Principal Component. This method helps us to identify those variables with the highest load (dominant variables) in each PC.

3.3 *The Regression*

The regression specification (1) has binary dependent variable. Hence, the Probit regression method is applied. We allow the corresponding probability distribution to be associated with a cumulative normal distribution,

$$Z = \Phi(x\mu + \varepsilon) \in (0, 1) \text{ so that } x\mu + \varepsilon = \Phi^{-1} \quad (2)$$

Notice, $Z' = x\mu + \varepsilon$

Since the dependent variable takes the value 0 and 1, we assume, Z' takes the value 0 and 1. The above function $Z' = x\beta + \varepsilon$ is called the Probit function whose parameters are estimated.

3.4 *The Results*

First we report the results as we use data from all the districts of West Bengal. The Eigen values of the principal components and corresponding rotated varimax table of Eigen vector matrix are shown in the following tables.

Table 3 shows the number of principal components to be extracted based on Eigen values. From the table we choose first four PCs whose Eigen values are greater than one explaining 70% common variation in data. The second column of this table shows each PC's individual contribution in explaining the data. The principal components are ordered according to their ability to explain the variation in the data in decreasing order. We observe that the first principal component individually explains 25% of the

Table 3 Principal components and their Eigen values: all districts

Component	Eigen value	Proportion	Cumulative
PC1	2.30	0.25	0.25
PC2	1.50	0.21	0.46
PC3	1.17	0.13	0.59
PC4	1.05	0.11	0.70
PC5	0.89	0.10	0.80
PC6	0.85	0.09	0.89
PC7	0.77	0.08	0.97
PC8	0.33	0.02	0.99
PC9	0.11	0.01	1.00

variation in the data, the second, third, and fourth individually explain 21%, 13%, and 11% respectively.

In order to identify the mostly associated control variable with a particular PC, the varimax orthogonal rotation for the first four PCs is carried out. The results are shown in Table 4. The varimax orthogonal rotation represents the correlation between a variable and a principal component. We have chosen that variable from a principal component, which has the highest association with the principal component. The dominant variables in the four principal components that we have chosen are as follows: ROAD for PC1, DENS for PC2, RAIL5 for PC3, and POWER for PC4. The associations indicate that transport-specific rural road network and railheads within 5 km from a village has been the dominant variables in PC1 and PC3. City-specific population density and village non-tradable rural electrification also find their place in terms of dominance in PC2 and PC4 respectively. Keeping their dominant role in explaining the PCs in mind, henceforth, in our analysis we will identify PC1 by ROAD, PC2 by DENS, PC3 by RAIL5, and PC4 by POWER. These variables are regressed on the dependent variable as in (1). The results obtained are reported in Table 5.

We run two different specifications of (1). The first is without the inclusion of the district-specific fixed effect and the second is with the inclusion of it. In both the specifications, the variables DENS and POWER turn out to be significant at 1% level. The result is interesting on the one hand because the partial approach taken by Chakrabarti and Mukherjee (2020) did not find a significant impact of DENS on the birth of Census Towns in West Bengal. On the other, Chakrabarti and Mukherjee (2022) found ROAD as a significant variable affecting the birth of CTs, which is no longer true in the present analysis. However, since from Chakrabarti and Mukherjee (2022) we know that because of uneven development the districts neighboring Kolkata show a different trend in the case of the transport infrastructure and the village-specific variables, we bifurcate the entire dataset in two parts and repeat the same exercise with each of them as above. First, we report the results of the dataset that contains data only on the villages in the districts bordering Kolkata. The districts bordering Kolkata are North 24 Parganas, South 24 Parganas, and Haora.

Table 4 Principal component table: all districts (rotated component matrix with varimax rotation)

Variable	PC1	PC2	PC3	PC4
HIGHWAY5	-0.19	0.01	0.51	0.33
RAIL5	0.04	-0.04	0.71	-0.03
NEARNESS	0.03	-0.06	0.35	-0.58
ROAD	0.60	-0.06	0.05	-0.09
BANK	0.15	0.42	0.28	-0.13
POWER (LOG TRANSFORMED)	0.06	-0.02	0.16	0.70
DENS (LOG TRANSFORMED)	-0.08	0.74	-0.05	0.02
AREA (LOG TRANSFORMED)	0.54	-0.30	-0.02	0.12
U.BANK (LOG TRANSFORMED)	0.51	0.40	-0.06	0.10
N	551	551	551	551

Table 5 The regression results: all districts

Explanatory variables	Probit1	Marginal effect	Probit2	Marginal effect
ROAD	0.01 (0.01)	0.01	0.01 (0.01)	0.01
DENS	0.36*** (0.07)	0.11	0.49*** (0.13)	0.16
RAIL5	0.10 (0.12)	0.02	0.11 (0.12)	0.04
POWER	0.92*** (0.41)	0.41	1.01*** (0.52)	0.45
District fixed effect	No		Yes	
N	551		551	
Pseudo R ²	0.02		0.03	

Dependent Variable Transformation of a village which was ‘would be Census Town’ of Census 2001 to Census Towns in Census 2011

Note Robust standard errors in the parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Of the 653 villages present in the entire dataset, which were ‘would be CTs in 2001’ in West Bengal, 171 belonged to these districts.

The Eigen values of the principal components of the above exercise and the corresponding varimax table of Eigen vectors are reported in Tables 6 and 7.

Here as before four principal components have their Eigen values greater than 1, which are taken up for the analysis. These PCs explain 73% of the common variation in the data. The dominant variables in the four principal components that we have chosen are as follows: AREA for PC1, DENS for PC2, RAIL5 for PC3, and HIGHWAY5 for PC4. Keeping their dominant role in explaining the PCs in mind, henceforth, in our analysis we will identify PC1 by AREA, PC2 by DENS, PC3 by

Table 6 Principal components and their Eigen values: the districts bordering Kolkata

Component	Eigen value	Proportion	Cumulative
PC1	2.86	0.31	0.31
PC2	1.61	0.18	0.49
PC3	1.04	0.13	0.62
PC4	1.00	0.11	0.73
PC5	0.87	0.09	0.82
PC6	0.70	0.07	0.89
PC7	0.62	0.06	0.95
PC8	0.19	0.03	0.98
PC9	0.04	0.02	1.00

Table 7 Principal component table: the districts bordering Kolkata (rotated component matrix with varimax rotation)

Variable	PC1	PC2	PC3	PC4
HIGHWAY5	-0.07	-0.01	0.11	0.83
RAIL5	0.04	-0.02	0.81	0.10
NEARNESS	-0.19	-0.09	0.36	-0.23
ROAD	0.56	0.07	0.15	-0.17
BANK	0.09	0.50	0.32	0.05
POWER (LOG TRANSFORMED)	0.29	0.03	-0.19	0.44
DENS (LOG TRANSFORMED)	-0.11	0.71	-0.11	-0.01
AREA (LOG TRANSFORMED)	0.57	-0.28	-0.01	0.02
U.BANK (LOG TRANSFORMED)	0.44	0.37	-0.04	-0.07
N	171	171	171	171

RAIL5, and PC4 by HIGHWAY5. As these variables are regressed on the dependent variable, the results obtained are reported in Table 8.

The first specification of the regression is without the inclusion of the district-specific fixed effect and the second is with inclusion of it. In both the specifications, the variables DENS and HIGHWAY5 turn out to be significant at 1% level. Notice that the role of density of the nearest city/ST, which we derived in the regression consisting of all the districts is preserved in this regression as well. However, in the districts bordering Kolkata it seems the existence of a highway within 5 km radius of a village plays a significantly positive role in its transformation to a CT. While the result confirms the finding of Chakrabarti and Mukherjee (2020), it is interesting to note that POWER, the variable which turned out to be significant in overall regression, no longer plays a role here.

Table 8 The regression results: the districts bordering Kolkata

Explanatory variables	Probit1	Marginal effect	Probit2	Marginal effect
AREA	0.05 (0.11)	0.01	0.24 (0.12)	0.05
DENS	0.60*** (0.18)	0.10	0.65*** (0.18)	0.15
HIGHWAY5	0.62*** (0.19)	0.15	0.47** (0.20)	0.10
RAIL5	0.17 (0.19)	0.04	0.24 (0.19)	0.05
DSFE	NO		YES	
N	171		171	
Pseudo R^2	0.18		0.21	

Dependent Variable Transformation of a village which was ‘would be Census Town’ of Census 2001 to Census Towns in Census 2011

Note Robust standard errors in the parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we report the results of the dataset that contains data on the villages in the districts not bordering Kolkata. The Eigen values of the principal components and the corresponding varimax table of Eigen vectors are reported in Tables 9 and 10.

The four principal components having their Eigen values greater than 1 here explain 64% of common variation in the data. The dominant variables in these PCs are: AREA for PC1, ROAD for PC2, HIGHWAY5 for PC3, and NEARNESS for PC4. Keeping their dominant role in explaining the PCs in mind, our analysis identifies PC1 by AREA, PC2 by ROAD, PC3 by HIGHWAY5, and PC4 by NEARNESS. As these variables are regressed on the dependent variable, the results are reported in Table 11.

Table 9 Principal components and their Eigen values: the districts not bordering Kolkata

Component	Eigen value	Proportion	Cumulative
PC1	1.81	0.22	0.22
PC2	1.44	0.17	0.39
PC3	1.25	0.13	0.52
PC4	1.08	0.12	0.64
PC5	0.98	0.10	0.74
PC6	0.78	0.09	0.83
PC7	0.77	0.08	0.91
PC8	0.50	0.06	0.97
PC9	0.35	0.03	1.00

Table 10 Principal component table: the districts not bordering Kolkata (rotated component matrix with varimax rotation)

Variable	PC1	PC2	PC3	PC4
HIGHWAY5	0.05	0.03	0.66	0.13
RAIL5	-0.18	0.35	0.47	0.01
NEARNESS	-0.05	-0.15	0.13	0.73
ROAD	0.01	0.75	0.13	0.09
BANK	0.12	-0.17	0.50	0.22
POWER (LOG TRANSFORMED)	0.05	-0.35	0.10	-0.59
DENS (LOG TRANSFORMED)	-0.35	-0.29	0.16	-0.09
AREA (LOG TRANSFORMED)	0.65	0.11	-0.04	-0.04
U.BANK (LOG TRANSFORMED)	0.62	-0.14	0.12	-0.08
N	380	380	380	380

Table 11 The regression results: the districts not bordering Kolkata

Explanatory variables	Probit1	Marginal effect	Probit2	Marginal effect
AREA	0.05 (0.09)	0.02	-0.01 (0.12)	-0.01
ROAD	0.01* (0.03)	0.01	0.01* (0.04)	0.01
HIGHWAY5	1.15*** (0.20)	0.18	1.10*** (0.27)	0.11
NEARNESS	- 0.90*** (0.35)	-0.10	-1.01*** (0.40)	-0.15
DSFE	No		Yes	
N	380		380	
Pseudo R^2	0.21		0.25	

Dependent Variable Transformation of a village which was 'would be Census Town' of Census 2001 to Census Towns in Census 2011

Note Robust standard errors in the parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notice that the variables HIGHWAY5 and NEARNESS turn out to be significant at 1% level. While the HIGHWAY5 has a positive effect on the formation of CTs as expected, NEARNESS has a negative effect. The negative sign of NEARNESS implies that in the districts, which are not bordering Kolkata, commuting is not an important factor for the birth of the CTs. In the districts not bordering Kolkata, it appears that only the transport infrastructure-related variables play a significant role in explaining the emergence of CTs: there is no significant role played by either the village infrastructure-specific factors or the city-specific factors.

4 Conclusions

The burgeon of CTs in India during 2001–2011 has been astounding and allures economists to analyze in deep the reasons behind it. The earlier literature pointed out the importance of city-specific factors like its formal-sector wage and area (Chakrabarti & Mukherjee, 2020); transport infrastructure-related factors like availability of highways near the villages and the local road network, the village infrastructure-related factors like availability of electricity and banks (Chakrabarti & Mukherjee, 2022) in conversion of villages in CTs. However, since these variables have mutual dependence on each other, no study before takes up all these factors together in a single framework to study their relative importance in formation of the CTs. The present chapter takes it up in a dataset of West Bengal, which have seen the birth of maximum number of CTs among the states in India during 2001–2011. The analysis has been carried out using the Principal Component Analysis, which consolidates the set of nine explanatory variables in four uncorrelated principal components. We identify the principal components with the dominant variable correlated with them and subsequently use them in the regression analysis to find their role in explaining the formation of CTs. The results show that the factors play their roles differently in the districts bordering Kolkata, the capital city of West Bengal, and the other districts of the state. While in both types of districts the availability of a highway within 5 km radius of a village, which was a ‘would be CT’ in 2001, helped the village to get converted into a CT in 2011, the density of nearby city/ST played a significant role in the districts bordering Kolkata. Commuting was not an important factor in the districts not bordering Kolkata. The results suggest that the forces of dispersion away from the existing cities/STs are important to understand subaltern urbanization in West Bengal. The improvement of the highway infrastructure is an important instrument in the process. The partial analysis that is present in the existing literature often emphasizes the role of local infrastructure in the villages and city-specific developments like rise in urban wages and areal expansion of the cities, in explaining the emergence of CTs. But a more general framework adopted in the present study do not support these views. The chapter shows that the subaltern urbanization in all over West Bengal is a fallout of policy of improving highway infrastructure. Whether it has improved welfare of the state, remains a future research agenda.

References

- Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, 133, 375–395.
- Balakrishnan, S. (2013). Highway urbanization and land conflicts: The challenges to decentralization in India. *Pacific Affairs*, 86(4), 785–811.

- Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103–124.
- Census of India. Various Years. Town and Village Directory, Office of the Registrar General and Census Commissioner, Government of India, New Delhi.
- Chakrabarti, S., & Mukherjee, V. (2020). Birth of Census Towns in India; An Economic Analysis. *South Asian Journal of Macroeconomics and Public Finance*, 9(2), 139–166.
- Chakrabarti, S., & Mukherjee, V. (2022). Role of transport infrastructure in birth of census towns in West Bengal. *Asia-Pacific Journal of Regional Science*, 1–24.
- Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4), 457–490.
- Chen, M. M., & Ma, J. L. (2015). Application of principal component regression analysis in economic analysis. In *3rd International conference on management science, education technology, arts, social science and economics* (pp. 1205–1208). Atlantis Press.
- Ghani, E., Goswami, A. G., & Kerr, W. R. (2012). *Is India's manufacturing sector moving away from cities?* (No. w17992). National Bureau of Economic Research.
- Helpman, E. (1998). The size of regions in topics in public economics: Theoretical and applied analysis. In D. Pines, E. Sadka & I. Zilcha (Ed.) (pp. 33–54). Cambridge University Press.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483–499.
- Krugman, P., & Elizondo, R. L. (1996). Trade policy and the third world metropolis. *Journal of Development Economics*, 49(1), 137–150.
- Kundu, A. (2011). Politics and economics of urban growth. *Economic and Political Weekly*, 46(20), 10–12.
- Mahajan, K., & Nagaraj, R. (2017). Rural construction employment boom during 2000–12. *Economic & Political Weekly*, 52(52), 54–63.
- Marjit, S., & Kar, S. (2014). Broader implications of labor market reforms in India: A general equilibrium perspective. *Indian Economic Review*, 49(1), 27–35.
- Mukhopadhyay, P., Zerah, M. H., & Denis, E. (2012). Subaltern urbanisation in India. *Economic and Political Weekly*, 47(30), 52–62.
- Mukhopadhyay, P., Zerah, M. H., Samanta, G., & Maria, A. (2016). Understanding India's urban frontier: What is behind the emergence of census towns in India?. *World Bank Policy Research Working Paper*, (7923).
- Nugrahadi, E. W., Maipita, I., & Situmeang, C. (2020). Dominant socio-economic indicators on the growth of small scale industrial sector: Empirical evidence with principal component analysis. *RevistaEspacios*, 41(01), 2020.
- Proost, S., & Thisse, J.-F. (2019). What can be learnt from spatial economics? *Journal of Economic Literature*, 57(3), 575–643.
- Savić, M. (2006). Principal components analysis of employment in Eastern Europe. *Panoeconomicus*, 53(4), 427–437.
- Sharma, A. (2013). Exclusionary urbanization and changing migration pattern in India: Is commuting by workers a feasible alternative. In *International union for the scientific study of population conference, Busan, Korea* (vol. 26).
- Zhu, Y. (2017). The urban transition and beyond: Facing new challenges of the mobility and settlement transitions in Asia. In *United Nations Expert Group Meeting on Sustainable Cities, Human Mobility and International Migration, New York*.