

# How Long the Shadow of the City? Schooling in Rural Gautam Buddha Nagar \*

Chadha, Nishant <sup>†</sup>      Das, Arijit <sup>‡</sup>

September 20, 2015

[PRELIMINARY DRAFT. PLEASE DO NOT CITE OR CIRCULATE.]

**Abstract** The district of Gautam Buddha Nagar, lying in Uttar Pradesh and bordering Delhi, has seen rapid urbanization in the last decade. This process of urbanization is exogenous in nature and this allows us to convincingly identify the spillovers from urbanization on neighbouring rural areas. Using digitized census maps, census data from 2001 and 2011 and the DISE data on schools we identify the impact of urbanization on schooling in rural GB Nagar. We find that the increase in the number of schools in the last decade is higher in villages closer to the urban agglomeration. While this effect is there for all levels of schooling it is stronger for higher levels of schooling. This change cannot be completely explained by increase in population in villages closer to the city. These results are consistent with the hypothesis that households are making schooling decisions that allow for increased participation in the urban labour market in the future. In line with this hypothesis we find gender differences in educational attainment with outcomes improving more for boys than for girls. We also estimate models with interaction terms to see if different initial conditions in villages shape the future growth in schooling. Our results have important implications for the policy driven urbanization process envisaged in India in the near future.

**Keywords:** Urbanization, spillover, education, schooling, regional science, gender, caste

**JEL Classification Numbers:**

---

\*We would like to thank Shubhashis Gangopadhyay, Dipankar Gupta and the seminar participants at the C-PACT urbanization seminar.

<sup>†</sup>Department of Economics, Shiv Nadar University, Greater Noida, India; e-mail: *nishant.chadha@snu.edu.in*

<sup>‡</sup>Center for Public Affairs and Critical Theory, Shiv Nadar University; e-mail: *airjit.das@snu.edu.in*

# 1 Introduction

There has been a rapid increase in urbanization in India, with the level of urbanization increasing from 27.81% in 2001 Census to 31.16% in the 2011 Census. With initiatives like the smart cities project there is also a stated policy to enhance and quicken this pace of urbanization. A relevant question to then ask is, what is the impact of urbanization on neighbouring rural areas? In particular, what is the change, if any, on the human capital acquisition decisions made by rural residents? This is the question we try and answer in this paper. The basic hypothesis underpinning the empirical work in the paper is that the returns to education are higher in urban areas than they are in rural areas and therefore agents, in anticipation of urbanization, demand more schooling.

To answer this question we study changes in the last decade in the district of Gautam Buddha Nagar (henceforth GB Nagar), specifically focusing on the rural areas of GB Nagar in 2011. GB Nagar is district in Western Uttar Pradesh (UP) and borders the national capital territory of Delhi. A large part of the district is located in what is called the National Capital Region (NCR). The first master plan of Delhi prepared in 1962 suggested that serious consideration should be given for the planned decentralization of large scale economic activities from Delhi and the development of towns around Delhi. Therefore, a need for establishing planned urban centers in the close proximity of Delhi was felt to provide an alternative site for the planned development of small and medium size industrial units. The NCR was thus envisioned to allow de-congestion in Delhi.

GB Nagar was created as a district in 1997, combining parts of two existing districts, Ghaziabad and Bulandshahr. The district was created to allow for rapid expansion of the planned urban centers of NOIDA and Greater NOIDA. The spread of urbanization in the district in the last decade has been impressive (see fig 1). What has been the impact, on schooling in rural areas of the district, of this rapid planned urbanization?

To answer this question we look at the nature of changes in villages of GB Nagar between 2001 - 2011 and see if there are any patterns to these changes in terms of distance from city. That is, is it the case that villages closer to the N-GN agglomeration have changed in a way that is different from those farther away? The key feature of the context that aids our empirical work is the exogenous nature of the urbanization process. The entire process from the acquisition of land, parceling and development, development of industrial areas and residential colonies in the N-GN agglomeration

was planned rather than an endogenous agglomeration process<sup>1</sup>. Thus helps us causally identify the impact of urbanization on schooling in the neighbouring rural areas.

Our results show that the increase in the provision of schooling, both at the extensive and intensive margin, is higher in villages closer to N-GN than in villages farther away. Villages closer to the city are more likely to have a school in 2011, if they did not have a school in 2001, compared to villages farther away. The increase in the number of schools is also higher in villages closer to the city. These results are robust to the inclusion of various controls. In particular, to the inclusion of two controls: the first is whether the village borders the national highway passing through the district and the second is whether the village is close to Dadri, the oldest urban center in this district. This suggests that the change in the last decade is due to the emergence of the new city and not due to an increase in the importance of the traditional means of transport to Delhi.

We also estimate models that include interaction terms of distance from city with the initial conditions in the villages in 2001. The results of these models suggest that there is a substitution between savings through physical capital and savings through human capital. After controlling for distance the growth in schools is lower in villages with initial access to institutional credit.

The results in this paper suggest significant human capital spillovers of urbanization on neighbouring rural areas. This has important policy implications given the push in India towards increasing rates of urbanization and a general spatial re-organization of economic activity.

## 2 Context of study

GB Nagar is in transition from semi-rural to a modern city. Proximity to the national capital Delhi is advantageous for the urbanization of this district. The urbanization process around national capital was stimulated by establishment of The National Capital Planning Board in 1985. Currently this district has 13 census towns and 320 villages. The district GB Nagar was formed in 1997 by carving out portions of Ghaziabad and Bulandshahar. Dadri and Bisrakh blocks carved out of Ghaziabad, while Dankaur and Jewar was taken from Bulandshahar district. Two major urban centers in this district were planned to host industries. The concept of NOIDA was coined in 1976 under UP Industrial Area Development Act., 1976. Greater NOIDA was planned much later, in

---

<sup>1</sup>We provide more details in a later section

1992 and the plan was approved in 1996 by NCR board. But the commonality of these two cities lies in their urban planning. Both the cities are planned and not an urban infringement of Delhi, like Bahadurgarh, which is a classic case of natural urban expansion.

A clear difference can be seen in Dadri. Dadri is the oldest town in GB Nagar that is been in existence for last 100 years. It is a traditional center for trading and exchange. Dadri was classified as town in the 1911 census. Since that time Dadri has grown at a steady rate for and it has 91 thousand inhabitant. Size-wise it is the third largest town in this district. Dadri is a classic example of natural urban agglomeration. It was developed though increase in economic activities and transportation.

This district has witnessed remarkable change in physical and social indicators since its inception. According to the 2011 census this district has a population of 16,48,115 with a density of 1161 inhabitants per square kilometre. Its population growth rate over the decade 2001-2011 was 39.32%. GB Nagar has a sex ratio of 852 female for every 1000 males which is an improvement from 2001. The reason could be rapid increase in female education. Over last 10 years female literacy has increased by 20 percentage points. Due to its close proximity to Delhi, the population is highly literate with 82.2% of population in the district are literate compared to 74.04% national average. The female literacy stands at 72.78% much higher than national average of 65.46%. Out of the total population, 59.12 percent lives in urban areas of this district. Urban GB Nagar has a literacy rate of 83.73% as per 2011 census. Child population figure of Gautam Buddha Nagar district is 13.92% of total urban population.

## **2.1 Progress of urbanization in present day GB Nagar**

On April 17, 1976, the UP government constituted an area about 120 sq. km and a population of approximately 42,000 persons in to the NOIDA. An authority was established under this act and notified 37 villages in 1976 and 14 villages were notified in 1978. The original master plan was prepared by the town and country planning department of UP in 1976 and a population of 10 lakhs was proposed for NOIDA Township (NOIDA Master Plan 2001). The plan document suggests 4500 industrial units in the first phase. A little over 4500 industrial units were planned and more than 300 acres of land was allocated to medium and small scale industries in second phase. The projected working population for NOIDA town was 3 lakhs in 2001 with a total population of 5.5

lakhs, according to the plan document. It was expected that this population of workers would include commuters from Delhi, other towns and villages around NOIDA.

The initial planning document envisages development of 3800 hectares of land by the year 2001. The authority kept 49.2 percent land for residential use and 13 percent land for industries. The NCR Planning Board revised the Regional Plan in 2005 for the perspective year of 2021. The Regional Plan for the NCR 2021 has assigned the population of 12 lakhs for NOIDA and accorded it a status of Central NCR town. Another factor, which is likely to have far reaching implications for the growth potential of NOIDA is the development of Greater NOIDA and Yamuna Expressway Industrial area townships on a contiguous territory east of the river Hindon. In 2001 the government of UP notified the entire area between river Hindon and Yamuna under NOIDA, covering approximately 20316 hectare land and 81 villages ( NOIDA master plan 2021).

### 3 Data sources

The data used in the paper are taken from the decadal census of India. We use data from the 2001 and the 2011 census for rural GB Nagar<sup>2</sup>. A crucial step in the analysis is the digitization of the census maps for GB Nagar using Arc GIS. After digitizing these maps (see fig 1) we then calculate, for each village, the distance between this village and the NOIDA-Greater NOIDA (henceforth N-GN) urban agglomeration. This distance is calculated as the distance between the geometrical centroid of the village and that of the N-GN agglomeration. Alternatively we could also use the shortest distance to the N-GN boundary as the distance measure. However, we believe our approach is superior for two reasons, first it is less susceptible to measurement errors because of the digitization of the physical map. Second, we do not know the spatial distribution of economic activity within the urban agglomeration and hence the centroid to centroid distance is a more accurate measure of distance to the city (distance to the administrative boundary would assume a more or less uniform spatial distribution of economic activity).

The census abstracts contain information on the demographics and the availability and number of various public amenities in villages. They also contain information on workforce participation at

---

<sup>2</sup>The variable definitions for some variables do not match exactly in the two census rounds, with the 2011 census containing much more detailed information. In this case we use the 2001 definition and combine the 2011 variables to accord with the 2001 definitions. More details are given in the appropriate sections

the village level. The census definitions of these various categories is given in the data appendix. The summary statistics of the chief variables used in the paper are given in table 1.

## 4 Empirical Methodology

The basic empirical model that we estimate in the paper is the following

$$\Delta Y_v = \beta_0 + \beta_1 \cdot distance_v^{11} + X'_v \beta_2 + \epsilon_v \quad (1)$$

Here  $v$  indexes villages.  $Y_v$  is the outcome variable (no. of schools etc.) and  $\Delta Y_v$  is the change in that outcome between 2001 and 2011.  $distance_v^{11}$  is the distance of the village from the N-GN agglomeration in 2011 and  $X'_v$  are additional controls.

We also estimate models that include interactions with initial village conditions in 2001.

$$\Delta Y_v = \beta_0 + \beta_1 distance_v^{11} + \beta_2 credit_v + \beta_3 (distance_v^{11} X credit_v) + X'_v \beta_4 + \epsilon_v \quad (2)$$

where  $credit_v$  is a dummy of whether the village had any institutional credit facility in 2001.

## 5 Results

The results of the estimation are presented in tables 3 and 4. Table 3 presents the results with the change in the number of schools as the dependent variable, while table 4 presents the results with the change in probability of having a school as the dependent variable. The point estimates on  $distance_v^{11}$  are negative and significant implying that as the distance from the N-GN agglomeration increases the increase in the number of schools between 2001 and 2011 is lower. While this is true for all levels of schooling, the results are stronger for higher levels of schooling.

## 6 Conclusion

## 7 Figures and Tables

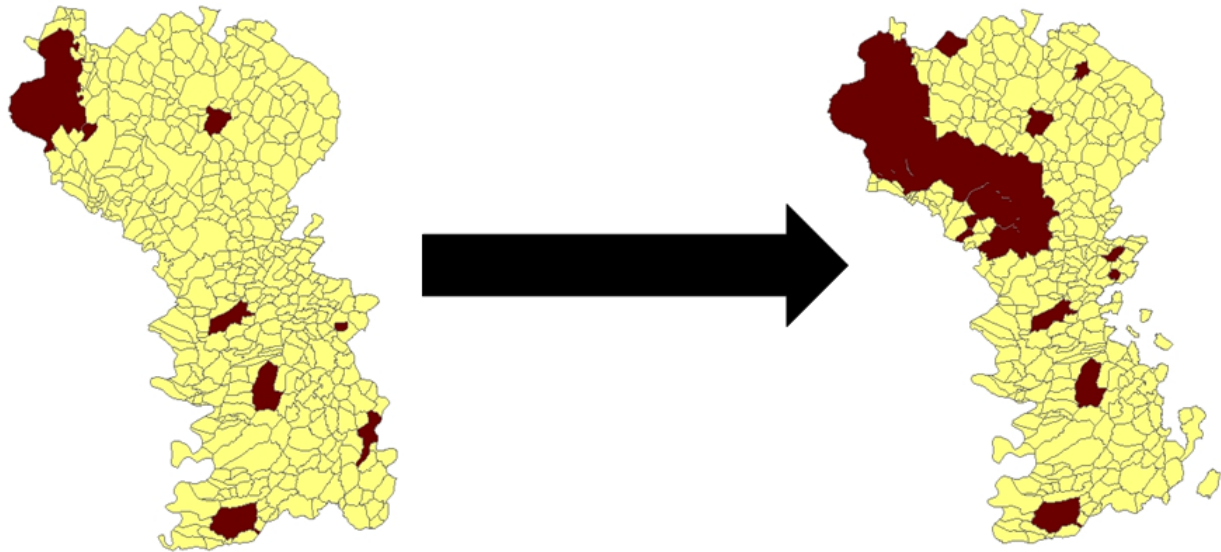


Figure 1: Urbanization in GB Nagar 2001-2011

Table 1: Summary statistics

Variable	2001			2011		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Distance from city				303	21.428	12.075
<i>Demographics</i>						
Total Population ('000)	303	1.827	1.538	303	2.231	1.880
Average HH size	289	6.831	.913	303	5.908	.852
Sex ratio	289	.854	.084	303	.871	.116
Proportion SC	289	.192	.154	303	.195	.181
Literacy rate	289	.51	.106	303	.618	.098
Male literacy	291	.642	.104	303	.715	.094
Female literacy	290	.357	.105	302	.509	.117
<i>Workforce</i>						
Prop. workers	289	.297	.087	303	.339	.11
Prop. main workers	291	.813	.157	303	.737	.209
Prop. main AL	291	.085	.115	303	.104	.144
Prop. main CL	291	.419	.182	303	.286	.2
Prop. main HH	291	.031	.041	303	.04	.056
Prop. main OT	291	.279	.189	303	.307	.217
Prop. marginal workers	291	.187	.157	303	.263	.209
Prop. marg AL	291	.072	.09	303	.091	.147
Prop. marg CL	291	.043	.074	303	.044	.069
Prop. marg HH	291	.01	.024	303	.024	.048
Prop. marg OT	291	.062	.096	303	.103	.116
<i>Number of Schools</i>						
Primary	303	1.142	.961	303	1.815	1.082
Middle	303	.267	.55	303	.931	1.022
Secondary	303	.079	.282	303	.426	.899
Senior Secondary	303	.063	.243	303	.29	.751
Colleges	303	.01	.099			
<i>Facilities</i>						
Communication	289	.208	.406			
Bus service	303	.168	.375	303	.297	.458
Banking	289	.045	.208			
Commercial bank	303	.033	.179			
Agr. credit society	303	.056	.231			
Paved road	289	.747	.435			

*Notes:* The data is from the decadal census of India for 2001 and 2011. The data for all variables was not comparable across the two years. In this case data is reported for 2001. *Prop. main AL* is the proportion of the total work force engaged in agriculture labour as their main activity. *CL* is for cultivators. *HH* is for household industries. *OT* is for others.



Table 2: Outcomes by distance of village from the N-GN agglomeration

Dependent Variable	Distance (1)	Distance (2)
<i>Demographics</i>		
Population density	-0.1664 (0.0522)***	-0.1809 (0.0648)***
Literacy rate	-0.0024 (0.0004)***	-0.0020 (0.0005)***
Male female literacy gap	0.0016 (0.0002)***	0.0016 (0.0002)***
Sex ratio (of children < 6)	0.0001 (0.0005)	-0.0005 (0.0005)
SC proportion	0.0027 (0.0006)***	0.0031 (0.0006)***
Avg. HH size	-0.0094 (0.0029)***	-0.0090 (0.0032)***
<i>No. of Schools</i>		
Govt. primary	0.0041 (0.0018)**	0.0052 (0.0020)***
Pvt. primary	-0.0153 (0.0038)***	-0.0134 (0.0041)***
Govt. middle	-0.0023 (0.0024)	-0.0029 (0.0026)
Pvt. Middle	-0.0096 (0.0028)***	-0.0103 (0.0029)***
Govt. secondary	-0.0057 (0.0020)***	-0.0044 (0.0019)**
Pvt. Secondary	-0.0078 (0.0025)***	-0.0075 (0.0025)***
Govt. Sr. secondary	-0.0051 (0.0019)***	-0.0043 (0.0018)**
Pvt. Sr. Secondary	-0.0071 (0.0019)***	-0.0065 (0.0018)***
<i>Workforce categories (Main workers)</i>		
Agricultural labour	0.0010 (0.0007)	0.0018 (0.0009)
Cultivators	0.0038 (0.0009)***	0.0038 (0.0010)***
HH industries	0.0001 (0.0003)	0.0000 (0.0003)
Others	-0.0072 (0.0009)***	-0.0072 (0.0010)***

*Notes:* The standard errors are reported in the parenthesis. Column (1) does not contain any controls in the regression while column (2) contains additional controls.

Table 3: Regression results of changes in number of schools on distance from the N-GN agglomeration

	$\Delta$ Primary	$\Delta$ Primary	$\Delta$ Middle	$\Delta$ Middle	$\Delta$ Secondary	$\Delta$ Secondary	$\Delta$ Sr. secondary	$\Delta$ Sr. secondary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-.011 (.005)**	-.009 (.005)*	-.009 (.005)**	-.011 (.005)**	-.012 (.004)***	-.011 (.004)***	-.011 (.003)***	-.010 (.003)***
Does village border NH 91 (Y=1, N=0)		.069 (.292)		.028 (.285)		-.216 (.235)		-.155 (.283)
Is Dadri closest town (Y=1, N=0)		.111 (.222)		-.172 (.245)		.267 (.207)		.212 (.200)
Constant	.900 (.121)***	.859 (.132)***	.861 (.126)***	.914 (.136)***	.614 (.109)***	.547 (.110)***	.470 (.098)***	.415 (.092)***
N	303	303	303	303	303	303	303	303
$R^2$	.014	.015	.011	.014	.031	.039	.038	.046

*Notes:* Robust standard errors are reported in the paranthesis. \*, \*\*, \*\*\* denote significance at the 90%, 95% and 99% levels respectively. The dependent variable in each column is the change in that level of schooling in that village between 2001 and 2011. So, for example, in columns (1) and (2),  $\Delta$  Primary=No. of primary schools in 2011-No. of primary schools in 2001. NH91 is the national highway passing through GB Nagar. Dadri is the oldest and second most populous town.

Table 4: Regression results of changes in probability of presence of schools on distance from the N-GN agglomeration

	$\Delta$ Primary	$\Delta$ Primary	$\Delta$ Middle	$\Delta$ Middle	$\Delta$ Secondary	$\Delta$ Secondary	$\Delta$ Sr. Secondary	$\Delta$ Sr. Secondary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	.002 (.002)	.002 (.002)	.0005 (.003)	-.0005 (.003)	-.005 (.002)***	-.004 (.002)**	-.004 (.001)***	-.003 (.001)**
Does village border NH 91 (Y=1, N=0)		-.027 (.047)		.051 (.166)		.019 (.162)		.014 (.149)
Is Dadri closest town (Y=1, N=0)		-.067 (.073)		-.124 (.129)		.146 (.091)		.099 (.081)
Constant	.032 (.038)	.055 (.045)	.306 (.069)***	.341 (.076)***	.270 (.046)***	.222 (.048)***	.186 (.040)***	.154 (.042)***
N	303	303	303	303	303	303	303	303
R <sup>2</sup>	.006	.009	.0001	.004	.026	.039	.026	.035

*Notes:* Robust standard errors are reported in the paranthesis. \*, \*\*, \*\*\* denote significance at the 90%, 95% and 99% levels respectively. The dependent variable in each column is the change in the probability of having at least one school of that level in that village between 2001 and 2011. So, for example, in columns (1) and (2),  $\Delta$  Primary is 0 if the village had a primary school in both 2001 and 2011 or if the village did not have a school in neither 2001 or 2011. It is given a value 1 if the village did not have any primary school in 2001 but had at least one school in 2011. NH91 is the national highway passing through GB Nagar. Dadri is the oldest and second most populous town.

Table 5: Demographic changes on distance from the city

Variable	Distance	
	(1)	(2)
$\Delta$ SC proportion	.0001 (.0003)	.0001 (.0003)
$\Delta$ population	-8.129 (4.283)*	-8.380 (4.885)*
$\Delta$ literacy rate	.0005 (.0003)	.0006 (.0003)*
$\Delta$ male-female literacy gap	.0007 (.0003)**	.0007 (.0003)**
$\Delta$ sex ratio (among children < 6)	.00007 (.0007)	-.0005 (.0008)
$\Delta$ prop. cultivators	-.0004 (.0009)	.0004 (.001)
$\Delta$ prop. other workers	.0006 (.0009)	-.0007 (.001)

*Notes:* Robust standard errors are reported in the parenthesis. \*, \*\*, \*\*\* denote significance at the 90%, 95% and 99% levels respectively. Each entry in the table is the coefficient from a regression of that dependent variable on distance from the N-GN urban agglomeration. Column (1) does has no controls in the regressions while column (2) contains whether village borders GT road and whether Dadri is closest town as control variables.

Table 6: Interaction with the presence of credit facility in 2001

	$\Delta$ primary	$\Delta$ primary	$\Delta$ middle	$\Delta$ middle	$\Delta$ secondary	$\Delta$ secondary	$\Delta$ Sr. secondary	$\Delta$ Sr. secondary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-.009 (.004)**	-.008 (.004)*	-.007 (.005)	-.009 (.005)*	-.012 (.004)***	-.009 (.004)**	-.010 (.003)***	-.008 (.003)***
Credit facility	.353 (.646)	.369 (.653)	.964 (.427)**	.947 (.437)**	.790 (.389)**	.901 (.396)**	.493 (.299)*	.572 (.292)*
Distance X credit facility	-.024 (.031)	-.024 (.031)	-.030 (.015)**	-.030 (.015)*	-.017 (.011)	-.020 (.012)*	-.014 (.008)*	-.016 (.008)**
Constant	.876 (.121)***	.832 (.129)***	.797 (.130)***	.842 (.139)***	.562 (.115)***	.478 (.113)***	.437 (.103)***	.371 (.095)***
N	303	303	303	303	303	303	303	303
R <sup>2</sup>	.021	.022	.025	.026	.051	.064	.047	.057

*Notes:* Robust standard errors are reported in the parenthesis. \*, \*\*, \*\*\* denote significance at the 90%, 95% and 99% levels respectively. Each entry in the table is the coefficient from a regression of that dependent variable on distance from the N-GN urban agglomeration. Odd columns have no additional controls in the regressions while even columns contain whether village borders GT road and whether Dadri is closest town as control variables.