

On Matching Census Tracts and Electoral Boundaries: The Bottom-up Aggregation Approach

MOHD SANJEER ALAM

Generating socio-economic profiles of electoral constituencies in India has been a long-standing problem. Although efforts have been made in the past by scholars and research agencies, the difficulty is far from being resolved. More often than not, the data on certain socio-economic parameters generated for electoral constituencies have been contested and debunked because of the difficulties and methodological problems facing the exercise of generating such data. A methodology that ensures greater accuracy of estimates of constituency level socio-economic data is attempted here.

A recent debate in the pages of the EPW has brought to the fore an old difficulty that faces students of Indian politics: the problem of generating reliable social profile of political boundaries of parliamentary and assembly constituencies. During the Lok Sabha elections, Indicus Analytics, an economic research firm, derived parliamentary constituency level data on some socio-economic parameters. In his article, "Electoral Politics and Manipulation of Statistics", Himanshu (2009) examined the veracity of constituency level data produced by Indicus. Apart from general problems involved in estimating constituency level data, he points out the extent of inaccuracy of such estimates for certain parliamentary constituencies. Bhandari (2009) responded to Himanshu by defending the Indicus method. Indicus used indirect method of estimation based on the descriptions provided by Delimitation Commission of India. It converted district level information to constituency level by taking weighted average of districts (weighted by population).¹ The data so generated is, in all likelihood, bound to suffer from the problem of under or over estimation as it is an inherent problem of indirect methods employed to project or estimate relevant parameters.

This interesting and occasionally sharp debate remained very limited in its scope. Himanshu views this as a problem of wilful "manipulation" of statistics, while Bhandari assumes this to be a problem that admits of no real solution, thus leaving indirect estimates as the only way out. There is an insufficient realisation in this debate about the nature of the problem, the range of solutions tried in the past and perhaps the best way of resolving this difficulty.

This article² proposes to expand the scope of this debate and suggests a way of resolving this long-standing problem in political analysis that has been evolved at the Centre for the Study of Developing Societies (CSDS) following the latest delimitation. Section 1 spells out the nature of the problem of matching the administrative and political map of India. Section 2 looks at the way this problem has been addressed in the past and the inadequacies of the existing solutions. Section 3 spells out the resolution worked out at the CSDS, following the latest round of delimitation. The approach adopted is spelt out step by step. This section also presents the database that has been developed as a result of the exercise. The concluding section offers a summary and some observations about the research challenges that lie ahead.

1 The Problem of Matching the Boundaries

India is one of the few countries of the world that collects, compiles and divulges wide-ranging information of its population at various administrative levels and on a regular basis. Since 1881,

This paper is based on a project that was partially funded by the Ford Foundation (India Office). Financial support of Ford Foundation to the project "Delimitation of Constituencies and Its Impact on Political Representation in India" is sincerely acknowledged. Special thanks to K C Sivaramakrishnan whose association with the project, though informal, helped the Lokniti team understand the nuggets of delimitation. Sincere thanks to the Lokniti team and the CSDS data unit, Himanshu Bhattacharya in particular. Special thanks to Jeevan Kumar and his team who undertook the painstaking job of downloading electoral rolls and data entry operation.

Mohd Sanjeer Alam (sanjeer.alam@gmail.com) is at Lokniti, Centre for the Study of Developing Societies, Delhi.

the census has been taking place without a break. Besides the census, there are other nodal data collecting agencies, National Sample Survey Organisation (NSSO) and National Family Health Survey (NFHS), in particular, that collect and compile data on many social, economic, demographic and health parameters at various geographical levels. While we have wide-ranging information on such parameters for various geographical and administrative units, we do not have the same at constituency level – the political unit. Lack of availability of constituency level data on various parameters of socio-economic development is largely due to the difficulties involved in translating statistics available at administrative units into constituency level. The source of the problem is that the administrative division of the country in terms of districts and sub-district units bears no direct relationship to the electoral map of the country in terms of Lok Sabha, assembly and panchayat or municipality constituencies. As an initial generalisation, it would not be unfair to say that these are two parallel and largely unconnected divisions of the country that do not map onto each other (Sivaramkrishnan 2008; Yadav 2008). The problem is not that of a lack of a perfect fit or that of incongruity in a small number of cases. As we shall see below, the problem is that of almost no fit at all or that of lack of correspondence in an overwhelming majority of cases.

This lack of fit between the two vital maps of India has led to many administrative and political problems (Sivaramkrishnan 1997, 2000, 2001). But that is not the focus of this article. Here let us focus on some major obstacles that it creates for students of public policy, politics and political economy. First of all, the absence of a reliable socio-economic profile of constituencies has not permitted a meaningful aggregate data analysis of the electoral outcomes.³ It follows that some of the socio-economic variables cannot be related to patterns of political participation and voting behaviour in a rigorous way. This deficit is particularly regrettable, for India provides a unique lab for use of aggregate data for the study of social basis of political behaviour. Since the political boundaries did not change between 1977 and 2004, it provided a unique opportunity for scholars of aggregate data to study political change. This opportunity has largely been neglected because of the absence of socio-economic profile of constituencies for this period.⁴

The second problem is that it has prevented researchers and concerned citizens from gathering any direct evidence of how far political representatives have taken care of those who returned them to legislatures, particularly when political representatives (individual MPs/MLAs) are armed with huge amount of funds to spend on development of their own constituency, apart from scores of universal developmental programmes in place.

There are a number of sources that collect, collate and divulge information on a range of socio-economic parameters. Notable among them are Population Census, Economic Census (EC), Agricultural Census (AC), NSSO and NFHS. Except for NFHS, other sources are purely official (conceptualised, conducted and data processed by government of India). While, Economic and Agriculture Censuses, as the name implies, are confined to specific items, the decennial Population Census, the NSSO and the NFHS cover a wide range of socio-economic and other aspects of population. The latter two sources are large-scale sample surveys

covering all states of India. The data collected by NSSO can be disaggregated to the district level. The latest 61st round covered 79,298 rural and 45,346 urban households. Using unit level data from this round of NSSO, Chaudhuri and Gupta (2009) have made estimates of levels of living and poverty at district level.⁵ The latest round of NFHS (2005-06) was conducted in all 29 states and collected information of a representative sample of 1,09,041 households. However, unlike NSSO data sets, the estimates drawn from NFHS could be utilised and analysed at state level. The advantage of these sources is while census takes place decennially, NFHS does so quinquennially. NSSO also repeats its round quinquennially. But the variations in geographical coverage and periodicity have proved to be a dampener to researchers/scholars who wish to assess socio-economic performance of a political constituency or understanding the nuances of political processes and outcomes thoroughly. For example NSSO data would be more relevant to assess recent progress made in individual constituency but here the problem, as mentioned earlier, is that one cannot analyse data for units below the district level and the boundaries of a parliamentary constituency rarely coincides with those of districts (Table 1). There are some districts that have two or more parliamentary constituencies, for example, Murshidabad district in West Bengal and East Godavari in Andhra Pradesh. On the other hand, there are many parliamentary constituencies which spread across three or four districts. For example, Aruku parliamentary constituency in Andhra Pradesh is spread across

Table 1: Boundary Fit between Parliamentary Constituency and District in Major States of India as Per the Fourth Delimitation

States	Number of PCs	Number of Districts (2001)	PCs in Perfect Fit with District Boundary	Name of the Parliamentary Constituencies
Andhra Pradesh	42	23	0	-
Bihar	40	37	5	Gopalganj, Begusarai, Nalanda, Bhojpur
Chhattisgarh	11	16	1	Sarguja
Gujarat	26	25	3	Sabarkantha, Anand, Jamnagar
Haryana	10	19	1	Faridabad
Himachal Pradesh	4	12	0	-
Karnataka	28	27	3	Bijapur, Davangere, Dakshin Kannada
Kerala	20	14	0	-
Madhya Pradesh	29	45	4	Satna, Rewa, Jabalpur, Chhindwara
Maharashtra	48	35	1	Beed
Orissa	21	30	2	Sundargarh, Jajpur
Punjab	13	17	2	Jalandhar, Patiala
Rajasthan	25	32	0	-
Tamil Nadu	39	30	3	Krishnagiri, Toothukkudi, Kanyakumari
West Bengal	42	18	0	-
Uttarakhand	5	13	1	Almora
Uttar Pradesh	80	70	9	Mathura, Firozabad, Shahjahanpur, Unnao, Fatehpur, Siddarthnagar, Basti, Maharajganj, Mirzapur
Delhi	7	9	0	-
All	490	472	35	

1. States with one or two parliamentary constituencies such as Goa, Mizoram, Tripura and Sikkim and those states such as Assam, Jammu and Kashmir, Nagaland and Manipur where delimitation was not carried out are not reported in the above table.

2. Delimitation was carried out in Jharkhand but the delimitation plan could not be implemented because of objections to it. The matter is sub-judice in the Supreme Court of India.

Sources: 1. *Changing Face of Electoral India: Delimitation 2008*, Vols I and II, Working Paper VII, Delimitation Commission of India, New Delhi.

2. *Census of India, 2001*, Primary Census Abstracts (State series), Office of Registrar General, government of India.

four districts – Srikakulam, Vizianagram, Vishakhapatnam and East Godavari districts. It is well known that districts, even if lying cheek by jowl, vary a great deal from each other in socio-economic attributes. In such a situation, relying on the physical proximity of districts or calculating averages of their socio-economic characteristics and employing statistical techniques of indirect methods for the purpose of deriving constituency level estimates of given parameters would be misleading. For such methods inevitably run the risk of under or overestimation, as is the case with the data provided by scholars as well as agencies including Indicus Analytics.

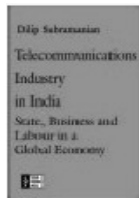
For this and many other reasons, the census remains the best source of generating socio-economic profile of the constituencies. The census is the only source that has universal coverage. It collects information for each and every household of the country in each and every village, hamlet, huts and jhuggis and so on unless some exogenous factors come into play.⁶ The Census of India gives information down to the village and ward levels in rural and urban settlements, respectively, and one can translate census statistics for even the constituency level. Leaving aside the methodological problems, which we discuss in the following section, the only disadvantage of using census data is that it reflects the socio-economic conditions of the past rather than the present. Using 2001 Census data for assessing socio-economic performance of a constituency in 2009 or so could sometimes be

misleading, particularly when things are transforming at rapid pace and a dynamic representative can change much of the socio-economic landscape of his/her constituency in much less time. However, despite such limitations census data is still useful. It can be utilised to track development records of a constituency for almost three decades because the physical make-up of present day electoral constituencies, thanks to the political elite, would not alter for next three decades.⁷ It still remains the most robust source to generate reliable profile of the electorate in terms of the variables that it provides. Especially useful is the census data on the population of the scheduled castes (scs), scheduled tribes (sts), minorities and linguistic communities. In India, social cleavages along the axes of caste, ethnicity, religion, language and so on have been the raw-material of Indian political life and central to political representation. Profiles of such social and linguistic groups have, therefore, been of paramount significance to political analysts. And the census is the only reliable source that provides estimates of such variables.

Translating census statistics into constituency is, nonetheless, an uphill task. This is precisely why authentic constituency level data is not available. There are three kinds of difficulties that arise here. First and foremost, the basic building blocks of the delimitation exercises and census do not match. The census follows the administrative units, the nomenclature of which differs from one group of states to another.⁸ Since 1961, the administrative units



visit us at: www.socialsciencepress.blogspot.com



TELECOMMUNICATIONS INDUSTRY IN INDIA (HB)
State, Business and Labour in a Global Economy
Dilip Subramanian

690 pages Rs 895 978-81-87358-42-8

Dilip Subramanian is Associate Professor at Reims Management School and is affiliated to Ecole des Hautes Etudes en Sciences Sociales, Paris.

'This is an impressive achievement ... that calls on a range of Francophone literature on the sociology of industry and work that is seldom cited in studies of Indian industry.'

Jonathan Parry, FBA
Emeritus Professor of Anthropology, LSE

'[The work]... has the potential to become a benchmark study in government-business-labour relations during the import-substituting industrialization era in India.... The proposed work evidently has a global readership....'

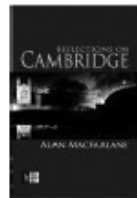
Tirthankar Roy,
Reader in Economic History, LSE



INDIA'S ECONOMIC FUTURE (HB)
Education, Technology, Energy and Environment
Edited by Manmohan Agarwal

294 pages Rs 595 978-81-87358-17-6

Manmohan Agarwal is Senior Visiting Fellow at Centre for International Governance Innovation (CIGI), Canada.



REFLECTIONS ON CAMBRIDGE (HB)
Alan Macfarlane

243 pages with 20 illustrations Rs 450
978-81-87358-48-0

Alan Macfarlane is Professor of Anthropological Science, University of Cambridge, and Life Fellow, King's College, Cambridge.

'... the Cambridge book - full of wisdom, warmth and great stories, a kind of manifesto for the liberal arts as well as a song of love for a place.'

Michael Chaplin,
film producer and writer



THE EVERYDAY STATE AND SOCIETY IN MODERN INDIA (PB)

Edited by C. J. Fuller and Véronique Béné

231 pages Rs 350 978-81-87358-57-2

C. J. Fuller is Emeritus Professor of Anthropology at LSE. Véronique Béné is Visiting Senior Fellow in Anthropology at LSE.

'... a topical, interesting and significant volume ... the essays make a significant contribution.'

T. N. Madan,
Honorary Professor, Institute of Economic Growth, University of Delhi

'The seven chapters ... along with the extremely lucid introduction by the editors, elevate our understanding of the Indian state and society to a higher conceptual plane.'

Dipankar Gupta,
Former Professor of Sociology, JNU, and Senior Fellow,
Nehru Memorial Museum and Library, New Delhi



Distributed internationally
BERGHAHN BOOKS
www.berghahnbooks.com



Distributed in South Asia by
ORIENT BLACKSWAN
www.orientblackswan.com

Table 2: Matching Delimitation and Census Units

States	Delimitation Units (Building Blocks)		Census Units		Do Delimitation and Census Units Match?
	Rural	Urban	Rural	Urban	
1 Andhra Pradesh	Mandal/village	Ward/EB	Mandal/village	Ward	Yes
2 Bihar,	CD	Ward	CD	Ward	No
3 Orissa,	Block/GP/village		Block/village		
4 West Bengal					
5 Chattisgarh,	Tehsil/RI circle/ Patwari circle	Ward	Tehsil/village	Ward	No
6 Madhya Pradesh	Taluka/Saza	Ward	Taluka/village	Ward	No
7 Goa	Taluka/village	Ward	Taluka/village	Ward	Yes
8 Gujarat	Tehsil/patwari circle	Ward	Tehsil/village	Ward	No
9 Himachal Pradesh	Tehsil/kanungo	Ward	Tehsil/village	ward	No
10 Haryana	circle/patwari circle				
11 Punjab	Taluk/circle/GP	TP/ward	Taluk/village	TP/ward	No
12 Karnataka	Taluk/panchayat	Ward	Taluk/village	Ward	No
13 Kerala	Tehsil/revenue circle/Saza/village	Ward/EB	Tehsil/village	Ward	No
14 Maharashtra	CD block, gram sewak circle	Ward	CD block/village	Ward	No
15 Meghalaya	RD block/village	Ward	RD block/village	Ward	Yes
16 Mizoram	Tehsil/ILRC	Ward	Tehsil/village	Ward	No
17 Rajasthan	Revenue block (RB)	Ward	Subdivision/village	Ward	No
18 Sikkim	Taluk/village	TP/ward	Taluk/village	TP/ward	Yes
19 Tamil Nadu	Tehsil/mouza	Ward	Tehsil/village	Ward	No
20 Tripura	Tehsil/kanungo	Ward	Tehsil/village	Ward	No
21 Uttar Pradesh	Circle/patwari circle				
22 Uttranchal					

1 Delimitation was not carried out in Assam, Jammu and Kashmir, Nagaland and Manipur. The delimitation plan was not implemented in Jharkhand.

2 EB stands for Enumeration Block; TP for Town Panchayat and GP for Gram Panchayat. Sources: 1 *Changing Face of Electoral India: Delimitation 2008*, Vols I and II, Working Paper VI, Delimitation Commission of India, New Delhi.

2 *Census of India, 2001, Primary Census Abstracts (State series)*, Office of Registrar General, government of India.

within the state that census follows are: district and tehsil/taluka/community development block at intermediate level – that is, between district and village/ward.

The Delimitation Commission (DC) however does not follow this approach. They do use district information but do not follow census conventions when it comes to sub-district units. Table 2 lists the units used by the Fourth DC for different states. Except Andhra Pradesh, Gujarat, Tamil Nadu and Mizoram, the DC has not followed the same units as the census. Usually, the authority that determines the boundary of constituencies has limited objectives. The basic concern of such authorities is to enclose a territory and a population as fixed by rules and laws to elect political representatives. In doing so, the unity of administrative units in many countries is often neglected and India is no exception, although the guidelines for delimitation provided by the DC categorically states that all constituencies shall be demarcated as far as practicable, be geographically compact areas and in delimiting them regard shall be given to the physical features, existing boundaries of administrative units, facilities of communication and public convenience (see guidelines of Fourth DC of India). However, the prime objective of the DC has been to partition a territory into constituencies whose populations are more or less equal. Article 81 (b) of the Constitution states:

....each State shall be so divided into territorial constituencies in such manner that the ratio between the population of each constituency and the number of seats allotted to it is, so far as practicable, the same throughout the State.

The second problem is also obvious from Table 2. Not only does the DC not follow the census units, it does not follow the same pattern in every state. To serve its purpose and for the sake of convenience, the DC chose different building blocks (that make up constituency) in different states, often at odds with the census units within the state. As a result, political boundaries hardly coincide with those of administrative units at which socio-economic information is normally available, particularly in the population census. It is a matter of debate as to why the DC did not choose the building blocks in a manner that had a better fit with the existing administrative units or at least the boundaries of the newly created local bodies (panchayat and nagarpalika) of panchayati raj.⁹ One of the practical problems is that administrative units (e.g., districts, tehsil/taluka/CD block) in India are not uniform in size of population and the DC is bound by the rule that inter-constituency disparity in population should be within the range fixed as “norm of tolerance”. The primacy of equal size population has resulted in parliamentary constituencies criss-crossing districts and assembly constituencies criss-crossing tehsils/talukas/CD blocks and so on. The net outcome of this difficulty is that it is not possible to use the same solution for every state if we use any intermediate level of aggregation.

Finally, there is a chronological misalignment between the census and the delimitation process. The census took place in 2001 and reflected the administrative boundaries as they existed on 1 March 2001. The delimitation exercise reflects the boundaries as they existed on 15 February 2004. As the changes in administrative boundaries are an ongoing process (Kumar and Somanathan 2009), these two do not match even if they were to follow the same principles. For example, there were 593 districts in 2001, which have increased to 618 in 2007 (see <http://india.gov.in/knowindia/districts.php>). It can be reasonably assumed that many of them were created between 2001 and 2004. In fact, creation of sub-district units has been even more frequent than the districts.

2 Existing Solutions and Limitations

This section presents a review of the efforts made, largely in the 1960s and 1970s, to overcome the problem of mismatch between census tracts and boundaries of electoral constituencies. Basically, there are three levels chosen by the scholars: state, district or assembly constituency. And there are two methods for profiling: estimation techniques or the more precise disaggregation and reaggregation technique.

State Level Aggregations

Historically, Indian states do not compare with one another in terms of socio-economic and cultural attributes. Politically too, they have treaded varying courses. Hence, state is a critically important political unit in India. Also, carrying out ecological analysis of electoral behaviour at state level poses no problem because electoral constituencies are in perfect fit with state boundaries and it is easier to obtain current socio-economic data at the state level. Given this ease at state level, ecological analysis of electoral behaviour and patterns has been carried out by a handful of scholars including Heginbotham (1971) and Weiner (1968). While such

analysis can and does illustrate innate differences and clarify the systematic properties that distinguish the political system of one state from another (Brass 1978), it inevitably misses interesting patterns within the state. Second, analysis at the state level may actually obscure the most relevant socio-economic and cultural boundaries.

Region or District Level Aggregation

As matching census tracts with constituency boundary has been well-nigh impossible, scholars have tried to overcome the problem by creating an artificial construct such as regions within the state. Chandidas (1967) and Frankel (1977) used this strategy to relate electoral behaviour with socio-economic variables drawn from census. The advantage of regional level analysis is that it helps relate strength of smaller parties, which tend to contest few seats and whose support base is localised, with ecological variables. Second, regions so created are a group of assembly constituencies as well as districts. Assembly constituencies often do criss-cross sub-districts but fit perfectly within the district boundary. However, region as a unit in ecological analysis of electoral variables has many limitations. The main problems with the region as a unit of analysis is that it is too large a unit that tends to overlook as well as obscures differences among sub-divisions within the region. To obviate the problem arising at regional level, some scholars have used district as a basic unit of analysis in order to relate electoral variables with the ecological ones. Notable among them are Baxter (1969), Dasgupta and Morris-Jones (1975) and Hardgrave (1975). District as a unit of ecological analysis of electoral variables is relatively easy, for assembly constituencies are contained within the boundary of districts and socio-economic data are much richer for the districts than for subordinate units.¹⁰ But it is still problematic to choose district as a unit for parliamentary constituencies.

While such techniques overcome the problem of incongruity between electoral and census units and allow ecological analysis of voting behaviour using aggregate data to a great extent, they are insufficient to make a nuanced account of the problem. The reason is that: regions or districts are too large a unit to be treated as internally homogeneous and cohesive. Reducing a number of assembly seats into one district and then treating the district as a unit with the assumption that there is a common underlying behaviour in the district as a whole is bound to suffer from ecological fallacy. Each constituency, as Blair (1973) points out, is at least in some sense *sui generis*. Pointing out methodological problems involved in district as a unit of analysis, Frankel remarks:

Frequently, the district represents too large an area to meet minimum standards of internal cohesion. District averages, which smooth out variations in the distribution of electoral data, often cancel out – and conceal – sharply divergent patterns among constituencies. It is, therefore, methodologically unsound to make an a priori assumption that district is a politically relevant unit (Frankel 1977: 150).

The other difficulty is that a district is an artificial administrative construct, often truncating natural subdivisions with the state at regional level. Most often it conceals the sharp divergence between the urban core and the rural periphery within a district. In India, geographic divisions coincide with the pattern of

development. There is a tendency to concentrate investment in industries, and social, economic and administrative infrastructure in areas with favourable factor endowment, which perpetuates pre-existing disparities between the so-called “forward” and “backward” states. For this and other reasons, the district is not an ideal unit to relate electoral variables with socio-economic ones. Frankel further clarifies,

when the district is used as the basic unit of comparison in correlation carried out at the state or national level, relationships which do exist between electoral and economic variables within some regional divisions are often submerged or weakened by the absence of such relationship in others (Frankel 1977: 150).

In brief, choosing the district as a basic unit to relate electoral variables with socio-economic ones is not immune from methodological fallacies. Finally, choosing districts or regions as units for ecological analysis of electoral behaviour involves massive aggregation of electoral and socio-economic data disaggregated down to the units below region and districts.

Constituency Level Aggregation

Given the limitations inherent in aggregation of larger units such as district, some scholars have tried to build the data below district level and to match the administrative units with the boundaries of constituencies. In the past such attempts have been made but for few constituencies, using ethnographic techniques or selecting a constituency that fitted census tracts such as tehsil/taluka/CD block. Works done by Kothari and Shah (1965) and others¹¹ in 1960 are most prominent ones that employed such techniques. However, Brass (1975) attempted to match assembly constituencies with “tehsils” for whole of Punjab. He aggregated the subdivisions/tehsils up to a point where constituency and administrative boundaries did coincide. The constituencies which did not fit administrative (tehsil) boundary were thus excluded from analysis. In this way, some groupings comprised only one tehsil, even one constituency; others comprised as many as four tehsils and 11 constituencies (the whole district). Weiner and Field (1977) also employed similar a “exclusionary” technique in their study of electoral behaviour in urban constituencies. They approached the urban constituencies by size of city population (big cities) and identifying constituencies lying in the core of urban settlements.

Besides the above, some scholars have employed sophisticated, though complex, techniques to match constituency with census tracts. Most notable among them is Blair’s matching of assembly constituencies with census tracts. Blair (1972) selected a sample of assembly constituencies that did fit the census tracts exactly. In other words, he identified the assembly constituencies in Bihar that was made up of whole blocks. The problem was that there were only few constituencies that actually did fit the census tracts (e.g. blocks). To overcome this problem, Blair (1973) used a cartographic method in his later study. He prepared isopleth maps that connected areas of equal concentration of population and estimated Muslim population for each assembly constituency. Such efforts, useful though they are, do not exactly fit the census tracts with constituencies. They are at best approximate fits. Second, disaggregation at units below constituency makes the scale of operation too large to apply this to all India level.

Recently, Indicus Analytics generated socio-economic profiles of parliamentary constituencies, which were reported in the media. With the help of descriptions provided by DCs with regard to the building blocks of constituencies, Indicus used weighted average of districts (weighted by the population) and converted district level information to constituency level (Bhandari 2009). However, a scrutiny of such information found them quite problematic (Himanshu 2009).

In sum, the efforts made so far to estimate socio-economic indicators of electoral constituencies and using them for ecological analysis of electoral behaviour suffer from serious methodological and analytical problems. One, socio-economic data aggregated at larger unit such as region and district and correlating them with electoral variables fails to capture the political nuances at micro level (constituency) and is thus prone to ecological fallacy. Second, the attempts to disaggregate ecological variables at constituency level are not very useful as they fail to make a perfect fit between census tracts and constituency boundary. They offer, if any, only an ad hoc solution to the problem.

A review of methods (both direct and indirect method of disaggregation and aggregation) employed to generate constituency level data thus suggests that the methods suffer from serious limitations and have actually failed to yield reliable estimates of socio-economic parameters. But are there ways of making direct computation of socio-economic parameters for individual constituencies?

The readjustment of boundaries of electoral constituencies that began under the Fourth DC constituted in 2002 has thrown a challenge to resolve seemingly an non-solvable riddle of generating socio-economic profile of electoral constituencies. The problem once again seemed to defy solution because the current DC, like its predecessors, chose the building blocks to re-delimit the constituencies that were not useful for researchers to profile electoral constituencies. Unlike in the past, the only advantage this time was that one could have access to, thanks to the RTI Act, the details (contained in various Working Papers of the Commission) of how delimitation was carried out immediately after the delimitation plan was approved by the Parliament.

It was, nonetheless, a remarkable intellectual opportunity because if the problem could be resolved, it would be a great contribution to Indian social sciences as a whole, leave alone ecological analysis of electoral behaviour. Moreover, present day researchers are technologically more advanced and better equipped to take on this problem than those who worked in this field in the 1960s and the 1970s.

The Lokniti team decided to seize this opportunity as soon as the final reports of DC came out even without attending to the question of "how". After much deliberation, discussions and research, the Lokniti team realised that the catch lies in identifying the unit common to delimitation and census. But the most striking question arose was: what would help identify common units?

3 Bottom-up Aggregation Approach

Obviously, the challenge that the Lokniti team at the CSDS faced was of the nature that seemingly defied solutions. To summarise the premises with which the team began its work:

(a) It was quite clear that ecological analysis of the mass of data

on political participation and voting behaviour required a reliable socio-economic profile of electoral constituencies.

(b) Such a profile was needed at the level of assembly constituencies which could be aggregated at the level of parliamentary constituencies.

(c) Such a profile needed to be based on census data rather than any other surveys. This would require reconciling the census data with assembly constituencies.

(d) It would be ideal if this could be done by disaggregating and re-aggregating the census data instead of using any estimation technique.

All this could be achieved if and only if the lowest unit of the census and the lower unit of the constituency could be matched with one another. This was not simple for there was no document that did this matching. The DC provided working papers, especially Working Paper VI that gave details of building blocks for assembly constituency (see Final Orders of Fourth DC).¹² But, unfortunately, the description contained in the paper was not comprehensive for our purpose. This was also a massive job, for there was no way it could be done without going through every single one of the 4,078 assembly constituencies in the country.

The puzzle of getting to common unit was, however, resolved through electoral rolls (ERS). It was found that the cover page of ERS of each polling booth contained some information that could have served our purpose. The cover page, amongst other things, gives the details of the location and extent of polling stations. The information includes the tehsil/taluka/circle/panchayat and so on in which the polling station is located and finally the name of village(s), ward number (in urban areas) falling in the polling station. As a village and ward in rural and urban areas respectively are the lowest units against which census gives out most of information, aggregation of information of all the villages as well as the wards in the given constituency would lead us to have accurate constituency level data, first at assembly constituency level and finally at parliamentary constituency level, for a parliamentary constituency is the aggregation of assembly constituencies.

Thus, a matching of the lowest units of census tracts and electoral constituencies and then a bottom-up aggregation provided a possible way-out of the methodological impasse. This method was time and labour consuming, but the effort seemed worth it when matched against the long-term rewards of generating a robust constituency profile that may be helpful for the next three decades. This is the approach Lokniti team at the CSDS decided to follow.¹³ For the sake of clarity, it would be worthwhile to detail out the exercise here. The whole exercise could be divided into the following steps.

Step I: The very first stage involved getting in possession of the cover page of ERS for each polling booth in each assembly constituency of the country. The problem appeared slightly less daunting as ERS were available electronically in the web sites of Chief Electoral Officer (CEO) of each state. Using some sophisticated fast downloaders, the ERS could be mechanically downloaded. However, we only took out the information relevant to the current exercise, that is, detailed information about the villages/wards (the lowest unit) falling under the polling booths

and their location in terms of taluka/tehsil/block/circle/panchayat, etc. The other difficulty was the language, for the ER in most of the states is available only in the state language. This required engaging a team of translators who would read the relevant part of the ER in different languages and enter the data in English in the relevant field in the database.

Step II: This stage marked the compilation of all the villages and wards in the given assembly constituency by taluka/tehsil/CD block/circle/panchayats (whatever relevant in the given state). Proper matching of the villages/wards with those recorded in census followed. Each village so recorded was assigned census codes so as to eliminate repetitions of villages. This was likely because a village does not itself form a polling booth. Where villages are small, a polling booth is made up of many villages. By contrast, where villages are large enough is divided into many polling booths. However, there are no established rules about how small or big a polling booth may be. This stage was critical for constituency level estimates. In urban areas, except for Mumbai and Delhi, ward numbers were the building blocks of assembly constituency and ERs bore information about the ward falling in the given polling booth. But in the case of Delhi and Mumbai, localities instead of wards were mentioned. It was possibly because the building blocks in these two cities were enumeration blocks rather than wards. Given the paucity of sufficient information, we preferred to leave out these two cities for the time being.

Step III: By this stage, villages/wards (with their unique codes and other locational details) were arranged by assembly constituency. Many rounds of checks, cross-checks and verification were made to ensure that no villages were left out nor were they repeated or misrepresented. After this, what followed in was the process of extracting information for each village/ward from Primary Census Abstract 2001 (State Series) through computer programming.

Step IV: Most of the difficulties and complexities were resolved up until the third stage. In the fourth stage, the major task was to collate and aggregate the data for the variables listed below gathered for villages/wards to obtain the same for the assembly constituency.

Step V: As the parliamentary constituencies are the aggregation of assembly constituencies, the final stage was to group them the way they constitute a parliamentary constituency and then to aggregate the information obtained for the assembly constituencies. Thus the same information as for assembly constituencies is available for parliamentary constituencies.

With this, the following information for each assembly and parliamentary constituency has been collated. Table 3 provides an example of the dataset for Valmiki Nagar parliamentary constituency of Bihar.

A caveat is in order. For the estimates of religious groups, same technique as for other parameters was not applicable because religion data is not available at village level. The lowest unit at which religion data is available is tehsil/taluka/CD block, etc (wherever relevant) and in many cases these administrative units split into many assembly constituencies. For estimating religious group, an indirect method was the only way out. Given this, the principle of proportionality was employed to estimate religious groups at constituency level. To illustrate this, let us take the specific example of Bihar. In Bihar, except for few instances, the lowest unit of delimitation is gram panchayat (GP). Consider that an assembly constituency (x) comprises 'i' and 'j' blocks and 'n' number of GPs of third block (k). The population of relevant religious group of 'i' and 'j' blocks can be added straightaway as they do not split. To estimate the religious population of GPs of 'k' block that fall in constituency 'x', the first step is to determine the mean population of the relevant religious group for GPs therein. This is achieved by dividing the population of given religious groups by total number of GPs in the given block. The second step is to estimate the population of religious groups for the GPs in 'k' falling in constituency 'x'. This is achieved by aggregating the mean population of those GPs falling in constituency 'x' (or simply by multiplying the mean population of GPs by number GPs that fall in the constituency 'x'). The population so achieved when added to the sum of the population of given religious group in blocks 'i' and 'j', it gives the final estimate of the religious group in the given constituency.

However, it must be noted that this is a crude method as it assumes that the given religious group is uniformly distributed across the GPs within the block. Most often, it is actually not the case. However, when averages are aggregated, there is also a probability that the problem of under or overestimation of each entity finally cancels out each other, which eventually results in minimising, if not eliminating, the errors of estimation.

It should also be noted that above exercise could not be carried out for all the constituencies in all the states for two reasons. One, delimitation was not carried out in a number of states. Therefore, only those states were taken up only those states

where delimitation was carried out. In states like Jammu and Kashmir, Arunachal Pradesh, Assam, Jharkhand, Nagaland and Manipur delimitation of constituencies were not carried out. In other words, profiling these states via electoral constituencies will have to wait until delimitation is carried out. Second, all the assembly constituencies of Delhi and many assembly constituencies in two districts of Maharashtra – Mumbai suburban and Mumbai city districts were also left out because the ERs in such cases were not as useful as in others. In the first place, the ERs contained localities by name instead of wards. Second, the building blocks of delimitation were enumeration blocks, for which detailed census information was

Table 3: Select Socio-economic Parameters for Valmiki Nagar Parliamentary Constituencies of Bihar

Parameters	
Total number of households	4,59,771
Total population	26,01,988
% Females	47.8
% Literates	43.9
% Urban population	7.9
% SC population	15.0
% ST population	2.1
% Muslim population	16.5
% Total workers	41.8
% Main workers	68.7
% Marginal workers	31.3
% Non-workers	58.2
% Cultivators	34.2
% Agricultural labourers	49.0
% Workers engaged in household industries	4.2
% Workers engaged in other activities	24.3

Source: CSDS Data Unit; computed and compiled from Census 2001.

not available. The net outcome of these difficulties was that we decided to shelve the idea of profiling all the constituencies of Delhi and about 35 assembly constituencies in Maharashtra (falling under Mumbai suburb and Mumbai city). In these two instances an accurate profiling has to wait for an ethnographic or GIS matching of the administrative and political maps of these metros. Thus, we have created socio-economic profile of 490 out of 543 parliamentary constituencies and 3,540 out of 4,078 assembly constituencies. We propose to add to this list as soon as the remaining states are brought under the new delimitation and when the required matching is done in the case of the two metros.

How accurate are our computations? This needed to be validated by measuring its deviations from actual data in such cases as was available. The best way to validate our profiles was to juxtapose these with the actual figures of SCs and STs for each assembly and parliamentary constituency in Working Papers VI and VII of the DC. The DC got population figures of SCs and STs for the smallest delimitation units (which in some instances generally not available in census, for example panchayat wise data) from local authorities. After extracting census information for assembly constituencies, it was double checked with the total population and proportion of SC/ST in assembly constituency. The systematic method that we followed for estimating certain parameters for constituencies bore fruits. For most of the states, the estimates of SCs/STs were perfect. In other words, the estimates were not in variance with the official figures at all. In rest of the states, the deviation was within the range of $\pm 1\%$.

4 Concluding Remarks

As noted above, the Lokniti team at the CSDS has generated information on a range of social and economic parameters at constituency level which avoid most of the difficulties that have characterised past efforts in this field. Even as these profiles are immensely useful for researchers, they do not transcend the whole spectrum of socio-economic indicators. Creating a socio-economic profile of political constituencies is an ongoing process and the Lokniti team at the CSDS would continue to expand the columns in the existing database with other information available at village level. It is now clear that whatever information is available at village or ward levels can be translated into constituency level. Therefore, the first and obvious step is to extract data from the District Census Handbook which gives out detailed information

including schools, healthcare facilities and other amenities at village level. It would be of great value if we could estimate proportion of households below poverty line from BPL census. The next challenge would be to incorporate the information available at sub-districts level, for example landholding ownership and other relevant information from agricultural census, by discovering methods that keep the estimation error to minimum. The third step would be look for information that is relevant to political analysis but for which there is no official data. The most important of these is of course the caste composition of the constituency. This step would require some rough approximation. Lokniti team is in constant search for discovering such methods and expanding the number of relevant variables at the constituency level.

To conclude, constituency level data pertaining to socio-economic parameters is immensely useful. Politics in India is undergoing significant transformation. Both the language and content of politics all over the country have made significant departure from those of yesteryears. The politics of parochialism seems to be rapidly giving way to the politics of development and good governance. Implementation of schemes and delivery of public goods and services are what people talk about and return incumbent governments to power or oust them from it. Moreover, individual MPs and MLAs, unlike in the past, get huge funds meant for spending on creation of durable assets and infrastructure as well as all round development of their own constituency. While it is impossible to collate every bit of information about a constituency, some broad and conventional developmental parameters that could be easily compiled might give some insights into how individual constituency has performed over decades. Also, availability of such statistics would, to a great extent, help the students of politics, political economy and public policy and those having interest in the field understand the political dynamics therein.

Indirect estimates of constituency level data are problematic and in most of instances could be misleading. Its usage, in all probability, is likely to be contested and debunked. On the other hand, accurate or near accurate data at constituency level is difficult to generate. Matching of the lowest unit of census and constituencies and a bottom-up aggregation thereafter seems to offer a way past this difficulty. This method is time-consuming, cumbersome and expensive. But given the sheer academic value of such a database, we hope that this may have been worth it.

NOTES

- 1 In his response, Bhandari however does not detail out the method used to estimate constituency level data. He would have made immense contribution to existing body of knowledge and scholarship in this particular field if he had given a little more detail about the method and procedures involved in generating constituency level data on the chosen parameters.
- 2 This article is an initial product of the project "Delimitation of Constituencies and Its Impact on Political Representation in India". The project had begun in April 2008 and is still in progress. In a sense it aims at integrating as much information in the data set as possible. As the new delimitation got under way, numerous issues emerged. The most prominent among them was how this new political geography would shape the political outcome. Estimates of changes varied from wild

- guesses to efforts to systematically gauge some changes. The Lokniti programme of the CSDS decided to undertake a painstaking route of matching the boundaries of the new constituencies and the census data for 2001. The author was the coordinator of this work while the work itself is the outcome of the collective efforts of colleagues at Lokniti. Yogendra Yadav, Sanjay Kumar and Suhas Palshikar helped in the conceptualisation and designing of the project.
- 3 Despite problems and difficulties, some scholars have tried to match census tracts with constituency boundary. For example, Blair (1973) has tried, though through complex technique, to map demographic and electoral data that matches all constituencies with census units in approximate fit. Weiner and Field (1977) examine voting behaviour and their correlates in urban and tribal constituencies using census data. Also see Brass

(1975), Dasgupta and Morris-Jones (1975) and Frankel (1977). Aggregation of socio-economic data available at lower level administrative units to constituency level in these studies suffers from severe limitations, which are discussed in the following section in greater detail.

- 4 The CSDS data unit had made some rough estimates of some social variables such as the proportion of SC/ST and urban in assembly and parliamentary constituencies as these existed during 1977-2004. The estimates of SC/ST population were derived from the data offered by the Third Delimitation Commission. The estimates of urban and in some cases, Muslim, population were based upon rough geographical matching. As it could not be done very systematically, the profiles generated by the CSDS data unit naturally suffered from the malaise of serious under/overestimations, especially in the case of urban population.

- 5 The NSSO surveys prior to the 61st round did not use districts as strata in the urban sector (in the two-stage stratified sampling design). While this allowed generation of reliable estimates of population parameters at NSS regional level, it did not help researchers to do the same at more disaggregated level. However, a significant departure was made in the 61st round in which the sampling design defined rural and urban parts of districts as strata for selection of sample villages and urban blocks, respectively. As Choudhuri and Gupta put it, "This has paved the way for generating unbiased estimates of important socio-economic parameters at the district level adequately supported by the sample design" Choudhuri and Gupta (2009: 94).
- 6 For example, census was not held in Assam in 1981 and Jammu and Kashmir in 1991 due to the problem of insurgency in the two states.
- 7 Originally, as per the provisions of the Constitutions, readjustment and delimitation of constituencies was to take place after every decennial census (Article 82). Accordingly, Delimitation Commissions were constituted in 1952, 1963 and 1972. However, the Parliament by passing the 42nd Constitution Amendment Act 1976 (which amended Articles 82 and 170) imposed a moratorium on allocation of seats and readjustment of territorial boundaries of constituencies as determined by Delimitation of Parliamentary and Assembly Constituencies Order 1976 until the publication of population figures of 2001 Census. The Parliament through the Constitution (84th Amendment) Act 2001, the Constitution (87th Amendment) Act 2003, inter alia, provided that (i) the total number of existing seats as allocated to various states in the House of People (Lok Sabha) on the basis of 1971 Census shall remain unaltered till the first census to be taken after the year 2026; (ii) the total number of existing seats in the Legislative Assemblies of all States as fixed on the basis of 1971 Census shall also remain unaltered till the first census to be taken after the year 2026; (iii) the number of seats to be reserved for scheduled castes (SCs) and scheduled tribes (STs) in the House of People and State Legislative Assembly shall be re-worked out on the basis of 2001, and (iv) each State shall be re-delimited into territorial parliamentary and assembly constituencies on the basis of 2001 Census and the extent of such constituencies as delimited now shall remain frozen till the first census to be taken after the year 2026. Also see *Changing Face of Electoral India: Delimitation 2008*, Delimitation Commission of India. The first census after 2026 will take place in 2031. The net effect is that the next (fifth) delimitation will be held only after 2031.
- 8 There is no uniformity among the administrative unit among the states themselves. In order to avoid confusion the census authorities in 1961 decided to disregard every unit other than tehsil/taluka and CD block at intermediate level, that is, between village/ward and district.
- 9 For a critique see Sivaramkrishnan (2008).
- 10 Matching the assembly constituency with the district is possible but not always very easy. However, it must be noted that the perfect matching obtains only at the starting point of delimitation but the neatness disappears very soon. By the end of last delimitation there were too many cases of the assembly constituency which fell in two or even three districts.
- 11 For a collection of such studies, see Weiner and Kothari (1965).
- 12 Fourth Delimitation Commission prepared seven Working Papers. Paper I contains district-wise 2001 population data and entitlement of assembly seats for each district. Paper II provides entitlement of seats for scheduled castes in the assembly and distribution of SC seats among the districts. Paper III gives population in the proposed assembly constituencies and seats proposed to be reserved for SCs. Paper IV gives the same information for STs as for SCs in Paper III. Paper V provides abstract statement of proposed assembly constituencies and total, SC and ST population in each constituency. Paper VI contains district-wise

statement showing details of extent, total/SC/ST population in each of the proposed assembly constituencies and, finally, Paper VII provides information on proposed Lok Sabha constituencies and their extent in terms of proposed assembly constituencies. (See *Changing Face of Electoral India: Delimitation 2008*, Delimitation Commission of India.)

- 13 The project of this nature would involve huge cost. As financial assistance was not forthcoming from any source especially the government institutions that fund research projects in social sciences, CSDS submitted the proposal for financial assistance to Ford Foundation (India Office), which kindly agreed to finance the project.

REFERENCES

- Baxter, Craig (1969): *District Voting Trends in India: A Research Tool* (Columbia: Columbia University Press).
- Bhandari, L (2009): "Socio-economic Performance of Constituencies: A Response", *Economic & Political Weekly*, October 3, Vol XLIV (40): 61-63.
- Blair, H W (1972): "Election and Aggregate Data Analysis at Block Level in Bihar" Mimeo.
- (1973): "Minority Electoral Politics in a North Indian State: Aggregate Data Analysis and Muslim Community in Bihar, 1952-1972", *The American Political Science Review*, Vol 67 (4): 1275-87.
- Brass, Paul R (1975): "Ethnic Cleavages in the Punjab Party System, 1952-1972" in M Weiner and John Osgood Field (ed.), *Studies in Electoral Politics in the Indian States: The Impact of Modernisation*, Vol IV, pp 3-62 (Delhi: Manohar Book Service).
- (1978): "Indian Election Studies", *South Asia*, Vol 1 (2): 91-108.
- Chandidas, R (1967): "The Fourth General Elections in Madhya Pradesh: A Case Study" in Rajni Kothari (ed.), *Context of Electoral Change in India*, pp 113-29 (New Delhi: Academic Books).
- Chaudhuri, S and Nivedita Gupta (2009): "Levels of Living and Poverty Patterns: A District-Wise Analysis for India" *Economic & Political Weekly*, February 28, Vol XLIV (9): 94-110.
- Dasgupta, B and W H Morris-Jones (1975): *Patterns and Trends in Indian Politics* (New Delhi: Allied Publishers).
- Frankel, F (1977): "Problem of Correlating Electoral and Economic Variables: An Analysis of Voting Behaviour and Agrarian Modernisation in Uttar Pradesh" in M Weiner and John Osgood Field (ed.), *Studies in Electoral Politics in the Indian States: The Impact of Modernisation*, Vol III, pp 149-82 (Delhi: Manohar Book Service).
- Hardgrave, Robert L Jr (1975): "The Communist Parties of Kerala: An Electoral Profile" in M Weiner and John Osgood Field (ed.), *Studies in Electoral Politics in the Indian States: The Impact of Modernisation*, Vol IV, pp 163-224 (Delhi: Manohar Book Service).
- Himanshu (2009): "Electoral Politics and the Manipulation of Statistics", *Economic & Political Weekly*, May 9, Vol XLIV (9): 31-35.
- Heginbotham, Stanley J (1971): "The 1971 Revolution in Indian Voting Behaviour", *Asian Survey*, Vol 11, (December): 1133-52.
- Kothari, Rajni and Ghanshyam Shah (1965): "Caste Orientation of Political Factions: Modasa Constituency" in M Weiner and Rajni Kothari (ed.), 1965.
- Kumar, Hemanshu and Rohini Somanathan (2009): "Mapping Indian Districts Across Census Years, 1971-2001", *Economic & Political Weekly*, 10 October, Vol XLIV (41): 69-73.
- Sivaramkrishnan, K C (1997): "Under-Franchise in Urban Areas: Freeze on Delimitation of Constituencies and Resultant Disparities", *Economic & Political Weekly*, 20-26 December, Vol 32 (51): 3275-3281.
- (2000): "North-South Divide and Delimitation Blues", *Economic & Political Weekly*, 26 August, Vol 35 (35-36): 3093-97.
- (2001): "Constituencies Delimitation: Deep Freeze Again?", *Economic & Political Weekly*, Vol 36 (51), 22-28 December: 4694-4696.
- (2008): "A Fear of Change" *Seminar*, 586, June, pp 14-19.
- Weiner, M and J Osgood Field (1977): "India's Urban Constituencies" in M Weiner and John Osgood Field (ed.), 1974, *Studies in Electoral Politics in the Indian States: The Impact of Modernisation*, Vol III, pp 1-47 (Delhi: Manohar Book Service).
- Weiner, M (1968): "Political Development in Indian States" in Myron Weiner (ed.), *State Politics in India*, pp 3-58 (Princeton: Princeton University Press).
- Weiner, M and Rajni Kothari ed, (1965): *Indian Voting Behaviour: Studies of the 1962 General Elections*, (Calcutta: Firma K L Mukhopadhyay).
- Yadav, Yogendra (2008): "The Paradox of Political Representation", *Seminar*, 586, June, pp 52-59.

Economic & Political WEEKLY

REVIEW OF AGRICULTURE

June 26, 2010

- Labels for GM Foods: What Can They Do? – Sangeeta Bansal, Bharat Ramaswami
- Agricultural Price Policy, Farm Profitability and Food Security – S Mahendra Dev, N Chandrasekhara Rao
- Climate Change and Water Supplies: Options for Sustaining Tank Irrigation Potential in India – K Palanisami, Ruth Meinzen-Dick, Mark Giordano
- Changes in Land Relations: The Political Economy of Land Reforms in a Kerala Village – Suma Scaria
- Pesticides in Agriculture – A Boon or a Curse? A Case Study of Kerala – Indira Devi P
- Social Organisation of Shared Well Irrigation in Punjab – Rakesh Tiwari

For copies write to:

Circulation Manager,

Economic and Political Weekly,

320-321, A to Z Industrial Estate, Ganpatrao Kadam Marg, Lower Parel, Mumbai 400 013.

email: circulation@epw.in