Working Paper No. 9

Research Area: Development

NSSO Surveys along India's Periphery: Data Quality and its Implications

Ankush Agrawal and Vikas Kumar

November 2017

Azim Premji University

Available from:
Azim Premji University
Survey No.66 Burugunte Village,
Bikkanahalli Main Road,
Sarjapura, Bengaluru – 562125, India

─────── **About the Working Paper Series** ───────

The working paper series aims at reflecting the range of work done across the Azim Premji Foundation, including the Field Institutes and the University, and disseminating this work to a broader audience. University faculty, members of Field Institutes of the Foundation as well as students graduating from the University may contribute to the series.

The findings, interpretations and conclusions expressed in these papers are entirely those of the authors and do not necessarily represent the views of Azim Premji University.

For all enquiries and feedback, please contact: **wps@apu.edu.in**

NSSO Surveys along India's Periphery: Data Quality and its Implications

Ankush Agrawal and Vikas Kumar*

⸻⸻✦⸻⸻

Abstract: *Sample surveys conducted by the National Sample Survey Office (NSSO) are the most widely used sources of household level information about consumption, employment, and other socio-economic indicators in India. The representativeness of samples, the wide range of topics surveyed, and the availability of a long-time series are some of the reasons for the appeal of NSSO data for research and policy. This paper assesses the quality of the NSSO data for Nagaland and Jammu and Kashmir, which lie in India's politically restive ethno-geographical periphery. It argues that the NSSO data for these states during 1973-2014 lack representativeness and inter-temporal comparability due to faulty sampling frames, frame and sample non-coverage, and biased samples. It quantifies the impact of data quality on statistics of interest to policy-makers. The paper shows that the estimates of monthly per capita consumption expenditure (MPCE) are sensitive to non-coverage and argues that the incidence of poverty is underestimated because NSSO surveys failed to capture the complete distribution of consumption expenditure due to non-coverage. In Nagaland, the degree of non-coverage was so high that in most years between 1993-94 and 2011-12 the state's poverty headcount ratio was the lowest in the country despite the possible overestimation of its poverty line. The paper concludes with a discussion of the implications of the use of non-representative survey data. Put together, the unreliability of government statistics in Jammu and Kashmir and Nagaland highlights systemic problems that have wider implications for our understanding of the relationship between state, statistics, and policy-making.*

Keywords: *Data Quality, Household Surveys, Insurgency, Jammu and Kashmir, Nagaland, Non-coverage, NSSO, Poverty*

NSSO Surveys along India's Periphery: Data Quality and its Implications

Ankush Agrawal and Vikas Kumar

1.0 Introduction

The Indian government attaches high priority to funding development in politically restive and geographically remote ethnic minority states. Until recently, the special category status conferred upon select states that entitled them to preferential federal transfers, was the most visible and institutionalised expression of the Union government's development and financial commitments.[1] However, despite decades of preferential treatment, most of the beneficiary states have remained economically stagnant. In some of these states, insurgency has directly (through physical damage to public infrastructure) and indirectly (by discouraging investment) stunted economic development and employment generation.

There is another aspect of the problem of development deficit that has received insufficient attention. Most national level household surveys in India do not cover the smaller states of the North East. For instance, most states in the region were not covered in various waves of the Rural Economic and Demographic Surveys. Some national surveys do not regularly cover these states. For instance, the District Level Household and Facility Survey did not cover Nagaland in 2007-08. Even the surveys that regularly cover the North East such as those conducted by the National Sample Survey Office (NSSO) – the most comprehensive, reliable, and widely used source of household-level information on India – do not have sufficiently representative samples to generate reliable estimates for all the states of the region.

---

1 The status of Special Category States remains unclear after the Union Government accepted the recommendation of the Fourteenth Finance Commission to raise the tax devolution to states from 32 to 42 per cent. The Union Government has been reassuring northeastern states that their request for continuation of the existing arrangement is under consideration (Nagaland Post, 2015; Economic Times, 2015). More recently, the finance minister indicated that the government was considering Andhra Pradesh's demand for special category status (Swami, 2016) and also seemed to suggest that the states in the North East shall continue to be treated as special category states insofar as the funding for 'core of the core' schemes is considered (Nagaland Post, 2016).

So, while India has some of the most robust government statistical machinery among developing countries,[2] its ethno-geographical periphery suffers from a data deficit — either data are unavailable or the available data are of poor quality. The peripheral states – Jammu and Kashmir and the North Eastern states – are mostly excluded from academic analyses because of the non-availability of data or the poor quality of data. As a result, the states that need more attention owing to persistent development deficit are generally absent from academic discussions. The data deficit is likely to affect development through its effect on policy-making.

However, data deficit could potentially be caused by, among other things, development deficit. The stagnant economies of the insurgency-affected, landlocked, hilly and relatively inaccessible states located along India's northwestern and northeastern peripheries are kept afloat by preferential federal transfers that account for more than half of the state budgets (Bhattacharjee, 2016). In some of these states, communities resort to competitive manipulation of statistics to secure better representation in government bodies and share in public spending and employment (Agrawal and Kumar, 2012b; Kumar and Agrawal, 2016). This obsession with government statistics is a consequence of struggle for scarce public resources in states with stagnant economies, where people depend upon government spending apportioned among communities/regions according to their *reported* numbers and *measured* socio-economic backwardness (Kumar and Agrawal, 2016).[3] In other words, there is a bi-directional relationship between data and development deficits.

Furthermore, it is well known that political unrest is both a cause and consequence of underdevelopment (Alesina et al., 1996, for instance), i.e., democracy[4] and development deficits are interconnected. However, the fact that political unrest can also affect the quality of government data has received insufficient attention. General unrest or boycotts impede data collection partially or fully, due to the inability of surveyors to access sites. Furthermore, the use of flawed data in policy-making affects the relationship between the government and the beneficiaries of the flawed data on the one hand, and aggrieved communities, on the other. The government's neutrality and the fairness of public institutions are compromised in the polarised environment. In short, *development, democracy, and data deficits are inter-related* and as a result *the quality of data is both affected by, and affects political and economic processes.*

This paper makes an attempt to study the aforesaid structural relationship in specific contexts. It examines the problem posed by faulty sampling frames, frame and sample non-coverage, and biased samples in NSSO surveys conducted in insurgency-affected Nagaland and Jammu and Kashmir, and shows that statistics from NSSO surveys are neither representative nor comparable over time for these states. It also tries to quantify the impact of non-coverage on sample statistics such as Monthly Per capita Consumption Expenditure (MPCE) and shows that non-coverage affects the reliability of policy-relevant variables such as the incidence of poverty.

The paper contributes to the literature on the quality of household survey data. The literature is largely focused on sampling errors (Groves, 2004; Groves et al., 2004) and survey non-response (Meyer, Mok and Sullivan, 2015; Hokayem, Bollinger and Ziliak, 2015). Non-coverage has received scant attention, that too mostly from a theoretical perspective (Lessler and Kalsbeek, 1992). More specifically, this paper contributes to the literature on the quality of NSSO data. This literature has mostly focused on the reliability of NSSO data in general (Suryanarayana and Iyengar, 1986), sample and questionnaire design (various contributions to Dandekar and Venkataramaiah, 1975; Deaton, 1997), inconsistencies between the NSSO and other government sources of statistics (Srinivasan, 1994), and the impact of the structure and evolution of the NSSO as an organisation on the quality of data (Vidwans, 2002; Shetty, 2012; Bhalla, 2014). Few, if any, studies assess the quality of NSSO data in specific sites over a long period (Agrawal and Kumar, 2014) and quantify the impact of data quality on survey statistics. Even the *Report of the National Statistical Commission (2001)* did not discuss the deficiencies in NSSO surveys in the context of specific states.[5] The *Report of the Committee on Optimum Sample Sizes for North Eastern States* discussed the problem of small sample size in abstract without engaging with the evolution of NSSO samples or the context within which data are collected in the concerned states [Government of India (GoI), 2011].[6] The Report does not deal with other factors that could affect the representativeness of samples either.[7] For instance, the NSSO samples for some of these states suffer from systematic biases, which are independent of sample size (Agrawal and Kumar, 2014).[8]

---

2   Indian decennial population censuses and household consumption and employment surveys are among the oldest and largest of their kind. In fact, India is one of the very few non-OECD countries that can boast of a continuous series of population census stretching back to the late 19th century. Also, since the 1950s, India has been known for advances in survey methods and has also been an active contributor to the development of international statistical standards (Rao, 1973; Rao, 2010; Shetty, 2012). In his tribute to P. C. Mahalanobis, C. R. Rao refers to Harold Hotelling, who observed that "No technique of random sample has, so far as I can find, been developed in the United States or elsewhere, which can compare in accuracy with that described by Professor Mahalanobis" (Rao, 1973: 472).

3   Under-development also implies that there are fewer resources for activities such as collecting statistics. The resultant data deficit feeds back into development planning.

4   We refer to 'democracy deficit' here because India is a democratic polity and, therefore, political unrest is associated with breakdown of democratic structures/processes, i.e., democracy deficit. A more general nomenclature and framework will be needed to cover problems associated with the impact of political unrest on data quality in non-democratic states.

---

5   The report touches upon several other issues related to the quality of NSSO data, though. These include recall bias and its consistency over time, inter-temporal incomparability of NSS consumption expenditure data due to reduction in cooperativeness of respondents over time, time taken to canvass information and length of the questionnaire, and selection of an appropriate reference period for non-food items such as durables, education, and medical care.

6   The representativeness of NSSO surveys at the sub-state/district level has been explored keeping in mind the necessities of regional planning (Chaudhuri and Gupta, 2009), but these discussions have generally omitted the North Eastern states.

7   There is also a more general problem that affects all Indian states. It has been argued that there has been a reduction in sample size in field surveys because of 'the inadequacy of the organisational strength at the NSSO or at the CSO [Central Statistical Organisation] level' and '[a]s a result, both sample and non-samples errors are feared to be high' (Shetty, 2012: 43). However, in principle, it could be argued that improvements in sampling techniques and accumulated experience in conducting surveys could possibly explain the reduction in the sample size.

8   In statistics, the term 'bias' is generally used in estimation theory, but it also refers to the presence of selective forces in the sampling process (Kruskal and Mosteller, 1979: 247).

The shortcomings of NSSO surveys also affect policy-making. The estimates of consumption expenditure and prices based on NSSO surveys impact poverty estimates and poverty alleviation policies. The related academic literature and public debate on poverty have focused on the level of poverty line being low (Economic Times, 2013), the inability of poverty line to capture the multi-dimensional nature of deprivation (Krishnaji, 2012), the suitability or otherwise of concepts and measures adopted by different expert groups (Suryanarayana, 2011), the use of price indices to update poverty lines (Deaton, 2003, 2008; Minhas et al., 1987), the problem posed by inconsistent reference periods for MPCE across different NSSO surveys (Deaton and Kozel, 2005; Deaton, 1997), and the lack of intertemporal comparability of poverty head counts in the rural sector (Suryanarayana, 2000).[9] Few, if any, studies assess the impact of the quality of NSSO data on poverty estimates.

The paper is also related to the literature on the political economy of statistics insofar as it examines the context dependence of the quality of survey data. This literature has largely focused on deliberate/strategic manipulation as a source of erroneous statistics. For instance, Wade (2012) discusses the refusal of the Chinese government to participate in the International Comparison Programme. Deaton and Kozel (2005: 196) hint at the politically-contested nature of NSSO surveys when they observe that "In evaluating the reforms, the political right had an interest in showing low poverty, and the political left in showing high poverty, and this undoubtedly intensified the debate on survey design and led to the unfortunate compromise design that temporarily undermined the poverty monitoring system." In the run-up to the 2014 Lok Sabha elections, NSS results were not only hotly debated in the media, but it was also argued that quinquennial round of 2009-10 had to be repeated in 2011-12 due to political interference (see Chandrasekhar and Ghosh, 2013, for instance). Michalski and Stoltz (2013) show that certain countries strategically misreport their economic data. Kumar and Agrawal (2016) explore the deliberate misrepresentation of information to secure greater access to public resources. Another strand of this literature examines the link between statistics and legitimacy of the state and the use of statistics/statistical exercises as means of social and political control. For instance, Serra (2014) examines the evolution of household surveys in colonial and post-colonial Ghana from this perspective.

The present paper differs from the existing academic literature and official assessments of NSSO surveys insofar as it provides an in-depth analysis of the quality of NSSO data in specific states over a long duration, situates the problem of data deficit in its political-economic context, quantifies the impact of poor data quality on consumption expenditure, and argues that data quality affects the reliability of policy-relevant variables such as poverty rates. It shows that consumption expenditure and poverty lines are overestimated because of frame and sample non-coverage. In Nagaland, the degree of non-coverage was so high that in most years between 1993-94 and 2011-12 the state's

poverty head count ratio was the lowest in the country despite the possible overestimation of its poverty line. Also, the decline in Nagaland's MPCE and the fluctuation of Jammu and Kashmir's MPCE relative to the rest of the country over the years are partly consequences of the changing operational sampling frames. However, NSS reports do not provide sufficient information for understanding the sampling frame and assessing the nature and extent of non-coverage.

The rest of the paper is organised as follows. Section 2 introduces the problems of frame and sample non-coverage and examines the problem of non-coverage in NSSO surveys in Jammu and Kashmir and Nagaland during 1972-2014. Sections 3 and 4, respectively, discuss faulty sampling frames and biased samples. The paper also briefly discusses similar problems in Manipur and Tripura in the North East and Punjab, but the main analysis is restricted to Nagaland and Jammu and Kashmir, where the problem persisted over a long period (Map 1). Section 5 illustrates the possible consequences of the use of flawed statistics in policy-making by quantifying the impact of non-coverage on sample statistics such as MPCE. The final section draws implications for sample surveys in conflict prone and/or underdeveloped regions.

2.0 Non-coverage in Sample Surveys

The failure to obtain data on parts of the target population leads to 'errors of non-observation' (Kish, 1965: 527-528). In contrast, 'observational errors' refers to the deviations of the respondents' responses from their true values (Groves, 2004: 11). Observational errors affect NSSO data in all states, whereas errors of non-observation are restricted to select states. Depending upon the reason behind non-observation, a three-fold classification of the relevant errors is possible: sampling errors, errors due to non-coverage, and errors due to non-response (Figure 1). Sampling error arises because the statistic is generated from a subset of the population. The 'failure to include some units, or entire sections, of the defined survey population' in the sampling frame is known as non-coverage, whereas non-response is the 'failure to obtain observations on some elements selected and designated for the sample' (Kish, 1965: 527-528). Non-coverage occurs when a perfect sampling frame[10] is not available and has also been referred to as frame coverage error (Biemer and Lyberg, 2003: 63). Such an error could occur even if a census of the frame is attempted (Groves, 2004: 11). On the other hand, non-response is a consequence of the inability of respondents to respond to questions, non-cooperation or refusal on the part of respondents to respond to questions, unavailability of respondents, and non-return or loss of returned survey questionnaires.[11] Non-coverage results from the exclusion of certain units from the sampling frame, whereas non-response results from the inability to collect information on a unit (included in the frame and) selected for the sample.

---

9 Suryanarayana (2009) also points out that the NSSO sampling design permitted the generation of unbiased estimate of the mean consumption expenditure, but not the distribution profiles because of which poverty estimates were unlikely to be reliable.

10 The sampling frame must be accurate, complete, and updated. A frame is perfect if 'every element appears on the list separately, once, only once, and nothing else appears on the list' (Kish, 1965: 53).

11 In their analysis of major household surveys in the United States, Meyer, Mok and Sullivan (2015) observe that the non-response rates have increased markedly during 1984-2013.

The problem of non-observation in case of NSSO samples in Nagaland can be referred to as the *frame non-coverage* because large parts of the state were not included in the sampling frame. In Jammu and Kashmir, too, we observe frame non-coverage due to the exclusion of Leh and Kargil districts from the sampling frame until recently. With regard to the areas in Jammu and Kashmir that could not be surveyed due to unfavourable field conditions, there is no specific term in the literature, and their categorisation into any of the above categories is not obvious. However, we prefer to use the term 'non-coverage' over 'non-response' for two reasons.[12] First, the source of error does not lie with the respondent, who was not approached in the first place and, therefore, it would be inappropriate to classify the error as non-response. Second, (Kish, 1965: 527-528) defines non-coverage as the failure to include units of the target population in the 'operational sampling frame.' In several cases, entire districts of Jammu and Kashmir could not be surveyed because of difficult field conditions and we can refer to the remaining areas where the survey took place as the *operational* sampling frame. Kish (1965: 528) adds that non-coverage 'also refers to "missed units," omissions due to faulty execution of survey procedures' in addition to the problem posed by incomplete frames. We shall refer to the exclusion of units, which were included in frame but could not be surveyed due to unfavorable field conditions, as *sample non-coverage. Frame non-coverage* and *sample non-coverage* together are referred to as *(overall) non-coverage.*

Of the various frame errors that can potentially bias survey estimates, non-coverage is 'perhaps the most serious type of frame error' (Lessler and Kalsbeek, 1992: 48).[13] Non-coverage biases the estimates of attributes because of 'zero selection probabilities for some target population units' (Biemer and Lyberg, 2003: 66). Two characteristics of the bias, referred to as the *coverage bias* (Kish, 1965: 519), are noteworthy. First, a bias would result if the missing and covered populations have distinct values on the survey statistic even if non-coverage rate is small. Therefore, the bias will not disappear with increase in the sample size. Second, in case of a simple linear statistic, such as the sample mean, the bias is a function of the share of target population not included in the frame (viz., non-coverage rate) and the difference between the statistics (viz., the means) of those covered in the survey and those not covered (Groves, 2004: 84-85; Lessler and Kalsbeek, 1992: 58-60). In case of a simple random sample, the bias is the product of these two terms. While there is no straightforward way of calculating the bias in case of multi-stage stratified NSS samples, the population share of the excluded sample First Stage Units (FSUs) can be used to analyse the implications of the bias.

In the following sub-sections, we examine the problem of non-coverage in NSSO surveys in Jammu and Kashmir and Nagaland. Both these states are located in the country's ethno-geographical periphery and have been longstanding sites of political unrest.

2.1   Jammu and Kashmir

Jammu and Kashmir was first covered by NSSO in the eighth round (July 1954-March 1955) [Indian Statistical Institute (ISI) 1964, Report No.10]. The sample size was very small in the early rounds, with the number of sample households being comparable to the number of sample villages in the state (ISI 1962a, 1962b, 1964). The coverage has varied markedly over the years due to geographical remoteness, difficult terrain, and, starting in the late 1980s, political unrest. In a few rounds, the results for Jammu and Kashmir are reported separately (for instance NSSO 48th Round), in others it is subsumed under the North West region (NSSO 51st, 52nd, 53rd, and 54th Rounds, Report No. 436, 440, 442, and 448), and while in still others data have not been reported for the state (NSSO 55th Round, Report No. 465 and 470). In a few other rounds (NSSO 59th Round, Report No. 490: 5 and NSSO 60th Round, Report No. 505: 7), the state as a whole had fewer than the 300 households that were required to publish NSS estimates (59th Round, Report No. 490: 4) and yet results were published to facilitate comparison with earlier rounds.[14]

The state is divided into three regions – Jammu, Kashmir, and Ladakh – that differ in terms of ethno-linguistic as well as religious composition (Map 2). These regions, respectively, correspond to the present NSS regions[15] of *Outer Hills* (Doda, Punch, Rajauri, and Udhampur districts of Jammu region) and *Mountainous* (Jammu and Kathua districts of Jammu region), *Jhelam Valley* (the entire Kashmir region), and Ladakh (the entire Ladakh region).[16] The documentation provided with unit-level datasets indicates that Ladakh, a part of Jhelam Valley in earlier rounds, became a separate NSS region in the 66th round (2009-10). A common problem shared by NSSO operations in all NSS regions of the state is related to the non-availability of census data in the 1950s and 1990s as census could not be conducted in the state in 1951 (GoI, 1953: 3) and 1991 (GoI, 2005).

2.1.1   Ladakh[17]

Ladakh was included in NSSO Surveys for the first time in the eighth round, but there is disagreement between reports regarding the field survey. As per one of the reports, Ladakh could not 'be surveyed as it was not possible to contact the survey villages' (Report No 10: 2).[18] Other reports

---

12 The magnitude of the bias in survey statistics is independent of the characterisation of the problem as non-coverage or non-response.

13 Other factors leading to imperfect frame include (1) inclusion of non-population elements, (2) multiplicity problems, (3) failure to account for clustering, (4) incorrect auxiliary information, and (5) incorrect accessing information (Lessler and Kalsbeek, 1992: 48).

14 In some of the earlier rounds, such as the 14th (1958-1959), 'Special weight was given to the State of Jammu and Kashmir [while allotting sample villages to individual states in the rural sector]' to enable generation of separate state level estimates (Murthy, 1967: 530; also see ISI, 1958 for the 8th round).

15 The NSS regions are groups of districts having similar agro-ecological characteristics. At present, there are 88 NSS regions in the country (http://mospi.nic.in/Mospi_New/upload/nsso/nss_regions.pdf). The NSSO survey statistics are believed to be representative at the level of regions (Deolalikar and Dubey, 2008: 409-11). These regions were referred to as 'Natural Divisions' in earlier reports. Initially the whole of Jammu and Kashmir was treated as one Natural Division.

16 Eight new districts, four each in Kashmir and Jammu regions, were created between the 61st and 66th Rounds. All the new districts of Kashmir belong to the Jhelam region. Samba district of Jammu belongs to the Mountainous region, whereas Kishtwar, Ramban, and Reasi belong to the Outer Hills region. The NSS reports covering the period of interest to us continue to work with the older districts.

17 In 1979, the Ladakh district was divided into two districts Leh and Kargil. Ladakh's districts were referred to as Ladakh and Kargil up to the 56th round (2000-01) (see, for instance, GoI, 2004: 27). The reports of the 57th and 58th rounds refer to the two districts as Leh and Kargil, while later reports refer to Leh (Ladakh) and Kargil.

18 In addition to the difficult terrain and climate, international territorial disputes might have affected NSSO operations in Ladakh.

suggest that the survey was delayed. The eighth round was completed in March/April 1955, while Ladakh was surveyed between August and November 1955 (Report 30: 2; ISI, 1964). The report does not provide any reason for the delay (which must have been related to difficult terrain and weather conditions), but mentions that as 'all the villages of the Central sample in Ladakh district of Jammu and Kashmir could not be surveyed, the estimates given in this report on the basis of the Central sample do not cover this district' (ISI, 1964: 110). Furthermore, we are not sure of Ladakh's inclusion between the 9th (1955) and 16th (1960-61) rounds as the reports do not clearly mention if Ladakh was included. However, a comparison of sample sizes of eighth round with ninth and tenth rounds suggests that Ladakh must have been included in the sampling frame of the latter, whether or not it was actually covered by the survey. However, Ladakh was not included in the 14th round (Murthy, 1967: 516). It is likely that Ladakh was not included in the sampling frame between the 17th and 62nd rounds.[19]

In recent times, Ladakh was surveyed for the first time in the 63rd round, but it was included only in the state sample (and not the central sample) in the 63rd and 64th rounds. In the 65th round, Ladakh was included in the central sample, but 'there were no separate sample first-stage units (FSUs) for 'central sample,' and the 'state sample' was also treated as 'central sample' (Report No. 535: B-4). The Directorate of Economics and Statistics of the state provided 'a copy of the filled-in schedules [for the state sample] to Data Processing Division of NSSO for processing' (65th round, Instructions to Field Staff, Vol. 1: A1). This arrangement was followed up to the 67th round.

Separate first-stage units in the central sample were allotted to Ladakh for the first time in the 68th round, which also happened to be the first quinquennial round survey covering the Ladakh region. However, (Report 560: 4) mentions, 'although planned, no sample could actually be surveyed in the districts of Leh (Ladakh), Kargil and Poonch[20] of Jammu & Kashmir by NSSO.' Thus, the 69th round is the first in recent times to have covered Ladakh separately for state and central first-stage units.[21]

2.1.2   Kashmir and Jammu

After the onset of insurgency, all the districts of Kashmir and the adjoining districts of Jammu were irregularly surveyed (Table 1). The degree of coverage varied with the intensity of political unrest (Figure 2). After the 45th round (1989-90), the reach of NSSO surveys dropped sharply in Jammu and Kashmir 'due to unfavourable field conditions' (Report No 407: 2; Report No. 436: 1; Report No. 476: 1; Report No 481: 2), 'due to unavoidable circumstances' (Report No 407: 18), 'owing to

unfavourable field conditions' (Report No. 442: 1; Report No. 448: 2; Report No. 453: 1), 'due to disturbed conditions' (Report No. 505: 7), and because the sample units '[fell] in disturbed areas' (Report No. 490: 4; Report No. 505: 7) or 'belonged to the disturbed areas' (Report No. 508: 3). The reports do not add further details about the nature of field conditions.[22]

Jammu, Kathua, and Udhampur districts of the Jammu region were the only districts that were regularly surveyed between 1973 and 2014. This means that we have an uninterrupted time series for only one of the NSS regions of Jammu and Kashmir, namely, Mountainous region that includes Jammu and Kathua (and Samba, the new district carved out of these two). Other districts of the Jammu region, which constitute the NSS region called Outer Hills, and the districts of the Kashmir region that belong to the NSSO's Jhelam region were not surveyed regularly despite being included in the sampling frame.

During 1991-2012 (47th - 68th rounds), between 2 and 74 per cent of the population in Jammu and Kashmir was not covered by NSSO samples in quinquennial rounds (Table 1).[23] Every round during this period was affected by non-coverage in Jammu and Kashmir and the state alone accounted for more than 51 and 67 per cent of the un-surveyed rural and urban FSUs in the country, respectively (Table 2), although the state's share in the population of the country was about one per cent.

A comparison with other states reveals that the ratio of the villages included in the sample to the villages in the NSSO sampling frame is much higher for Jammu and Kashmir than the national average, but it is lower than that for most northeastern states (Table 3). The NSSO seems to have included a larger number of FSUs to ensure that a sufficient number of sampling units would be surveyed despite possible non-coverage in some areas. However, the ratio of the villages surveyed to the villages allotted is lowest for Jammu and Kashmir for both quinquennial as well as annual rounds.[24] Also, the ratio of villages of Jammu and Kashmir allotted in quinquennial to annual rounds (viz., 1.87) is much higher than the national average (viz., 1.10) and the same holds true for the villages surveyed (1.94 for Jammu and Kashmir compared to the national average of 1.22) (Table 3).

Two other points are noteworthy. First, the NSSO forms several sub-divisions known as hamlet-groups in large villages (with a population of 1200 or more) (GoI, 2001: 15). The criterion for hamlet group formation (the population of the sample FSU) is relaxed in a few hilly districts in

---

19 While we have not been able to access all the relevant reports, those for 17th to 27th and 30th rounds and for the rounds mentioned in Table 1 explicitly mention the exclusion of Ladakh from the sampling frame.

20 NSS reports spell the name of Punch district differently. We use the spelling 'Punch' given in the 2011 District Census Handbook.

21 After Ladakh was included in the sampling frame (viz., 63rd Round), the 2001 Census was used as the sampling frame for Leh and Kargil towns (Report No 535: B1; Report No. 549: B3). Until the 66th Round the urban areas of Ladakh were treated as non-UFS towns (Report No. 547: B2). Ladakh was covered for the first time under the NSSO's Urban Frame Survey (UFS) in the 2007-12 phase (GoI, nd). UFS is the basis of sampling frame in urban sector.

---

22 Only a few reports distinguish between non-coverage in villages in disturbed areas and the rest of Jammu and Kashmir. Report 508 (61st Round), for instance, notes that 'as many as 98 out of 129 allotted villages that could not be surveyed belonged to the disturbed areas of Jammu and Kashmir' (Report 508: 3). Also, interaction with the NSSO staff revealed that the Srinagar office remained completely shut during the peak years of insurgency.

23 We can also use the information on FSUs in Table 2 to arrive at an alternative measure of sample non-coverage (number of allotted FSUs not surveyed divided by the total number of allotted FSUs). As per this alternative measure, the average non-coverage in rural and urban sectors, respectively, was 31 and 37 per cent and the corresponding maxima were about 69 and 71 per cent. Also, note that even in some of the earlier rounds the rural areas of Jammu and Kashmir could not be completely surveyed, e.g., 87.5 per cent of the FSUs could be surveyed in the 38th Round (Table 2).

24 The NSSO conducts two types of surveys: quinquennial rounds, which are conducted once in five years, and annual rounds. The quinquennial rounds, with a much larger sample, mostly focus on consumption expenditure and employment and unemployment.

Jammu (Punch, Rajouri, Udhampur, Doda) and Ladakh (Leh (Ladakh), and Kargil) (GoI, 2014a: B4, for instance). The rationale behind this is not clear. While a similar relaxation is made in case of Himachal Pradesh, Sikkim, and the hill districts of Uttarakhand, it is not clear whether this is done on the basis of criteria such as difficult terrain and low population density – criteria that are satisfied by many other NSS regions. Second, non-coverage in Jammu and Kashmir is different from other states in a crucial respect insofar as the problem affected both rural and urban areas. In all other states, urban areas were almost always part of the sampling frame and, with a very few exceptions, sample non-coverage in urban areas has been very low.

2.2 Nagaland

The whole of Nagaland constitutes one NSS region.[25] The NSSO began operations in Nagaland in 1972 [Government of Nagaland (GoN, 2013: 2, 26)], i.e., nine years after the formation of the state. However, until the 44th round (1988-89), operations were restricted to urban areas (GoI, 2004: 55). Rural Nagaland was included in the 44th round as that annual round focused on the living conditions of tribal population, among other things. The following restrictions were introduced when NSSO surveys were extended to cover the rural areas of Nagaland:

> Due to inaccessible conditions in the Nagaland the (interior) villages located beyond 5 km of bus route (769 [this seems to be a typographical error, the correct figure should be 768] out of a total of 1,119 [the correct figure should be 1,118] villages in the state) were kept outside the coverage of survey. Samples were drawn purposively for the remaining 350 villages of the state, of which 164 were connected by bus-routes and the rest were within 5 km of a bus route. The central sample consisted of 120 villages, of which 82 villages were connected by bus-route and 38 villages were within 5 km of a bus route. (GoI, 1994: 49)

As per the NSSO, there were 1,118 villages in the state at the time of the 44th round, 164 of these were connected by bus, 186 fell within 5 km of bus routes, and 768 were located beyond 5 km of bus routes. All the villages in the last category were not included in the frame. These numbers do not tally with the 1981 Census, which served as the sampling frame. Our compilation from the corresponding District Census Handbooks suggests that there were 1,112 villages in the state of which only 122 were connected by bus, 140 fell within the 5 km and 850 beyond 5 km of bus routes. The difference between our compilation and the NSSO figures can possibly be explained if the NSSO used updated village and bus route lists to account for changes between 1981 and 1988-89. However, this seems unlikely. As per the 1991 Census conducted two years after the 44th round, the total number of villages was 1,225 of which 344 were connected by bus or fell within 5 km of bus routes and 881 were beyond 5 km of bus routes. In any case, over the next two and

a half decades, NSSO surveys continued to exclude 'interior villages of Nagaland situated beyond five km of any bus route' (GoI, 2013a: D1; also see GoI, 1996: 1). Since Nagaland is neither the only insurgency-affected state, nor the only state with difficult terrain, it is not clear why the NSSO imposed distance-based restrictions in this state but not in other comparable states.[26] It bears emphasising that while sample non-coverage was relatively high in the rural sector of other states of the North East, Jammu and Kashmir, and central Indian states (GoI, 1989: 2), similar restrictions did not apply there.

In the 50th round (1993-94), the NSSO changed the sampling frame after the publication of the 1991 Census.[27] While the sampling frame changed, the number of villages excluded continued to be 768 until the 60th round (2004) even though both road coverage and bus routes expanded and the total number of villages in the state increased between 1988-89 and 2004. Yet the number of villages included in the sample did not change. The 2001 Census was adopted as the new sampling frame in the 61st round (2004-05). The number of villages included in the frame increased from 350 (out of 1,118 villages) in the 44th round to 371 (out of 1,317 villages) in the 61st round (GoI, 1994, 2011). As per the District Census Handbooks for 2001, there were 1,322 villages in the state, of which 687 were either connected by bus or were located within 5 km of the bus route and 635 were located beyond 5 km. So, contrary to its claims, the NSSO did not even include all the villages within 5 km of bus routes in the frame. There could be two reasons for these differences. The difference in reference periods could be a reason: while our figures for a year correspond to the preceding census, the NSSO may have used information for 1988, 1993, and 2004, when they changed the frame. The presence of unrecognised or uninhabited villages in the Census figures could be another reason. However, as argued above for the 44th round, the difference between the census and the NSSO is too wide to be explained by these factors.

In the 69th round (2012), the hitherto uncovered villages were included in the sample by forming a State/UT level special stratum comprising these villages (Report No 556: B1). Only four FSUs were allotted to the special stratum compared to 40 FSUs allotted to the non-special stratum. Until the previous round, the non-special strata had 371 villages, while 946 villages were not included in the sampling frame. So, the ratio of villages in the special stratum to that in the non-special stratum is about 2.5, but the corresponding ratio of FSUs allotted to these strata is 0.10.[28]

Three other points are noteworthy. First, in contrast to the persistent and substantial frame non-coverage between 47th and 69th rounds (1991-2012), sample non-coverage was very small in the rural sector and was restricted to just two rounds (56th and 65th) in which one FSU each could not

---

25   The case for dividing the state into two regions – hills and foothills/plains – is strong insofar as they are ecologically, demographically, and economically different. However, Nagaland is a small state both in terms of area and population that makes division into multiple regions infeasible.

---

26   Andaman and Nicobar is the only other state/union territory where until recently certain villages were regularly excluded on grounds of inaccessibility (Report No 538: 5).

27   In some earlier NSS rounds, for instance, 49th (1993), information from the 1991 Census was used only for newly declared towns and for deciding if a district had to be divided into multiple strata (Report No 429: 20).

28   The actual value of the former ratio will be slightly different as the NSSO might have used an updated list of villages.

be surveyed (Table 2). In the urban sector during the same period, five rounds (51st, 53rd, 54th, 56th, and 58th) were affected by sample non-coverage. While the average sample non-coverage (share of allotted sample FSUs that could not be surveyed) for urban areas was less than two per cent during 1991-2012, the average was about eight per cent for the five rounds mentioned above as one-fifth of the FSUs could not be surveyed in the 51st round. It is worth emphasising that until recently a small proportion of the state's population lived in urban areas. The absence of frame non-coverage and a small sample non-coverage in urban areas has limited significance as the towns were not evenly distributed across the state. There was just one town before 1971. Two new towns were notified in 1971, followed by the notification of four new towns, including the first two towns in the districts on the Myanmar border in 1981. Five out of the seven towns were located closer to the Assam border.

Second, the NSSO divides the survey year into four sub-rounds, each of three months' duration, and allots an equal number of sample FSUs to each round. In most rounds this restriction was not strictly enforced in Nagaland because of the arduous field conditions (for instance, GoI, 2013a: D1).[29] This may affect estimates of the characteristics that are influenced by seasonality. While we do not have sufficient information to quantify the impact of the relaxation of survey schedule, unit level data suggest that the distribution of FSUs across sub-rounds was not remarkably different.

Third, there are major differences between NSS samples in Nagaland and Jammu and Kashmir. In the latter, whole districts were excluded depending on the intensity of insurgency, whereas in Nagaland the NSSO covered all districts even if some were covered better than others. Moreover, NSS samples were affected by frame non-coverage in Nagaland, whereas in Jammu and Kashmir the samples were affected by sample non-coverage.[30] In most years, frame non-coverage was about two per cent in Jammu and Kashmir, whereas sample non-coverage was less than two per cent in Nagaland.[31]

3.0 Faulty Sampling Frames

The 1991 and 2001 Censuses have been shown to be flawed in the case of Nagaland because of manipulations driven by economic and political considerations. The discrepancy in headcount varied systematically between hills and plains, rural and urban areas, and men and women. The degree of manipulation was higher in rural and inaccessible hilly areas, in areas closer to the Myanmar border, and in areas with higher tribal population share. The discrepancy also bears systematic association with the distance from the nearest town. Between 1991 and 2011, the census figures for the overall population of the state always exceeded the true values. The accumulated error in the overall population exceeded 18 per cent in 2001 and the census populations of certain hill districts and tribes were twice their true populations (Agrawal and Kumar, 2012a, 2013; Kumar and Agrawal, 2016). An incorrect sampling frame affects the representativeness of samples insofar as it does not contain the population characteristics of interest in proper proportions. So, sampling frames based on the 1991 and 2001 Censuses systematically over-represented certain population sub-groups relative to the others. As a result, the Census of Nagaland provided a flawed sampling frame.

The flawed censuses of the past continue to affect the analyses of the NSSO for Nagaland. The NSSO used overestimates of Nagaland's population between the 61st and 66th rounds (2004-10). Afterwards it under-projected the population (Report Nos 532, 536, 537, 538, and 564). The 1991 and 2001 Censuses overestimated the state's population and the NSSO used the inflated headcounts for its projections, even after the state government formally rejected the faulty census data (Agrawal and Kumar, 2012a). The over-count was corrected to an extent in the 2011 Census, when the government tried to check ghost enumeration. The growth rate turned negative because the 2011 population was less than the manipulated head count of 2001. This does not mean that the population contracted during 2001-11. The *correction* was interpreted by the NSSO as a *contraction* in the population of Nagaland and, therefore, it under-projected the population after 2013. This highlights a lack of attention to even essential statistics that decide the sampling frame.

Likewise, the use of the census as a sampling frame in Jammu and Kashmir is problematic. First, census could not be conducted in 1951 and 1991, resulting in the use of outdated sampling frames in the respective decades. Second, the share of Muslims and non-Muslims in the state's population has been hotly contested in almost every census. For decades the Muslim share of the population contracted before picking up in 2001 (Swami, 2014). Only a detailed assessment of the census can

29 Difficult field conditions have also affected geological surveys in Nagaland: 'About 48 % of the area could not be covered by systematic geological mapping on 1:50,000 scale owing to inaccessibility' (GSI, 2011: 2, 70).

30 Frame non-coverage affects two other states of the North East – Manipur and Tripura. Since the adoption of the 2001 Census as the sampling frame, Manipur's sampling frame has excluded three sub-divisions of Senapati district (see Notes to Table 3). Yet this was not mentioned in any of the NSS reports, which is a serious lapse. The three sub-divisions accounted for about 45 per cent (6 per cent) of the population of the district (state) in 2001. Very few NSS reports mention rural Tripura's exclusion from the sampling frame. For instance, a NSSO publication mentions that the 'disturbed villages of Tripura' were outside the sampling frame of the 27th (1972-73) (GoI, 1979b: 289), 28th (1973-1974) (GoI, 2004: 13-14; GoI, 1977: 49), and 30th (1975-76)(GoI, 1984b: 1) rounds. Even in later rounds the number of sampled FSUs was often far less than the allotted number (GoI, 1989: 2). So, it seems the problem of frame non-coverage in rural Tripura transformed into that of sample non-coverage. Between the 47th and 56th rounds, the average sample non-coverage was about 19 per cent (25 per cent excluding the 53rd round). On the other hand, sample non-coverage in the urban sector was less than 2 per cent. Note that in case of Tripura the non-coverage has been estimated using information on FSUs allotted and surveyed. Tripura has not been included in the analysis in Section 5 as the geography of exclusion is not clear from the information provided by the NSSO and there is no straightforward way of translating the sparse information about exclusion of certain FSUs into non-coverage rates in terms of population. Punjab faced a related problem in a few years during the late 1980s and early 1990s, when most rounds suffered from sample non-coverage in the rural areas, while in urban areas non-coverage was restricted to very few rounds, e.g., 38th and 49th.

31 Nagaland also differs from other northeastern states, where frame non-coverage has been minimal and sample non-coverage has been relatively higher. The average sample non-coverage in urban areas was less than 2.5 per cent in these states except Arunachal

Pradesh, where it was about 8 per cent. The persistently high level of sample non-coverage in Arunachal Pradesh is not related to political unrest, though. In all but one of these states, (i.e., Assam), the urban sector suffered from exceptionally high non-coverage in one round – Arunachal Pradesh (81 per cent, 58th), Tripura (17 per cent, 55th), Meghalaya (22 per cent, 58th), Manipur (33 per cent, 54th), Mizoram (56 per cent, 58th), and Sikkim (12 per cent, 58th). In the rural sector, the average sample non-coverage was more than three per cent - Tripura (8.5 per cent), Arunachal Pradesh (6.9 per cent), Assam (3.3 per cent), and Mizoram (3.2 per cent) - except in Sikkim, Manipur, and Meghalaya. Cases of high levels of non-coverage in the rural areas, say more than 10 per cent, were restricted to Arunachal Pradesh and Tripura (except Mizoram in the 58th round). Among these states, only Arunachal Pradesh has not seen a steady decline in non-coverage over the years.

reveal if these changes are driven by manipulation. Field interviews in the state (conducted by the second author) reveal that even government officials who were part of the exercise do not trust the reliability of the census. In fact, it has been suggested that the 2011 Census seems to have been affected by the manipulation of enumeration in the Kashmir Valley (Guilmoto and Rajan, 2013; Bhat, nd), which has seen sustained and high levels of sample non-coverage in the recent past.[32]

4.0 Biased Samples

The distance-based restriction (of including only the villages located within 5 km of bus routes in the sampling frame) applied to surveys in Nagaland skewed the NSSO estimates of socio-economic characteristics such as the monthly per capita consumption expenditure (MPCE). There are a number of reasons for this.

First, a large proportion of villages of Nagaland is located more than 10 km from roads, while NSSO surveys were arbitrarily restricted to villages within 5 km of bus routes. Since bus routes are a subset of roads, the NSSO was not even covering all villages within 5 km of roads. Furthermore, the bus network is not uniformly distributed across the state, with the network being denser in areas closer to the Assam border compared to the areas closer to the Myanmar border. This is likely to have resulted in an overestimation of the average MPCE for the state because households in villages near roads and those in villages in more developed districts closer to Assam enjoy a better standard of living.

Second, the districts closer to the Myanmar border not only have a higher share of rural population, but also a higher share of remote villages. In addition, these districts have lower literacy rates and are economically underdeveloped compared to the districts closer to the Assam border. The share of tribal population also increases with the distance from the nearest road and distance from the Assam border.

Third, there is substantial disparity among tribes in Nagaland and different tribes are concentrated in different parts of the state. This spatial divide is powerfully illustrated by the fact that all tribes residing in districts on the Myanmar border are designated as educationally and economically backward tribes that are eligible for support under affirmative action policies within the state. The NSSO's distance-based sampling restriction implies that the backward tribes, which inhabit districts relatively closer to the Myanmar border, were less likely to be included in the sample compared to the advanced tribes, which inhabit districts closer to the Assam border where road density, bus connectivity, and the level of urbanisation are higher.

Last but not the least, villages further away from road and villages closer to the Myanmar border have relatively difficult terrain and greater forest cover, which increases the possibility of exposure to insurgency as these locations provide easy shelter and escape routes. Exposure to insurgency dampens economic activity. Once again it is clear that the NSSO systematically left out a certain type of villages from its sampling frame.

Data from the 1981 Census, the most reliable of the censuses conducted in the state between 1981 and 2001 (Agrawal and Kumar, 2012a), as well as our field visits to the state support the above claims.[33] In 1981, about 85 per cent villages in the circles bordering Myanmar were located at a distance of more than 10 km from the nearest bus stop compared to 33 per cent along the Assam border. Also, average literacy rate in the circles on the Myanmar border was 16 per cent compared to 44 per cent in the circles on the Assam border. Until recently, there were only two towns in the vast territory adjoining the Myanmar border. Finally, the share of tribes in the total population was 97 per cent in the circles on the Myanmar border compared to 82 per cent in the circles on the Assam border (GoI, 1984a). In fact, as the reach of surveys increased, the share of tribes in the NSSO samples for Nagaland grew from 85 per cent in the 50th Round to 88 in the 55th Round, 94 in the 61st Round, and 96 in the 66th Round.

The underrepresentation of certain parts of Nagaland in the sampling frame seems to have been compensated by an oversampling of areas included in the sampling frame. The NSSO oversamples the villages within 5 km of bus routes in Nagaland perhaps to meet its state level sample size requirement for the central sample. Thus, compared to the North East and the country as a whole, a much higher proportion of Nagaland's villages that are included in the sampling frame are sampled even though the state has a much smaller proportion of villages in the sampling frame (Table 3).

Jammu and Kashmir is the other state with significant and sustained non-coverage. A number of features of non-coverage in the state stand out. First, there is a high correlation between the degree of non-coverage and the intensity of insurgency measured in terms of fatalities (Figure 2).[34] (Nagaland is home to one of the oldest post-colonial insurgencies in India and, therefore, we see a very high degree of frame non-coverage right from the beginning of the NSSO surveys and not just a correlation between non-coverage and the intensity of insurgency in specific years.) Second, in the Jammu region the excluded areas are located either along the international border (Punch and Rajouri districts) or they share a border with Kashmir and Ladakh regions (Doda district), whereas the included areas adjoin the plains of Punjab (Jammu, Kathua, and most of Udhampur districts)

---

32 Nagaland and Jammu and Kashmir are not the only states affected by the possible manipulation of the census head count. In recent times, similar problems have been reported from parts of the northern hill districts of Manipur (GoI, 2014c; Laithangbam, 2004).

33 The 2011 Census was first used as the sampling frame very recently in the 71st round (2014), whereas we are interested in the period before the 69th round (2012).

34 The degree of non-coverage was also highly correlated with insurgency related fatalities even in Tripura (Figure 2). We can infer that the reach of surveys in Tripura was limited due to insurgency. As in Nagaland, Tripura's tribal communities are concentrated in remote rural areas and hilly areas in the eastern part of the state, which were also susceptible to insurgency. The sample systematically excluded underdeveloped, hilly, forested, and largely rural tribal areas. So, there was a systematic demographic variation between the surveyed and un-surveyed areas in Tripura, which must have biased sample statistics.

(Map 3). Third, the excluded areas have relatively inaccessible terrain and greater forest cover (GoI, 2013c, Table III.38). These areas also have a greater exposure to insurgency. (Ladakh that was until very recently excluded is among the most inaccessible parts of the country, but it has not suffered insurgency.) Fourth, the excluded areas have a greater share of Muslim population. Kashmir is almost entirely Muslim, while Doda (58 per cent in 2001), Punch (92 per cent), and Rajouri (60 per cent) have substantial Muslim populations. Fifth, Doda, Punch, and Rajouri have substantial tribal population. In 2001, these three districts together accounted for 35 per cent of the ST population of the state, whereas their share in the state's population was 15 per cent. So, the areas of Jammu and Kashmir affected by frame and sample non-coverage were systematically different from the areas included in the sample.

5.0 Sample Statistics

We have so far argued that the NSSO samples for Jammu and Kashmir and Nagaland are not representative.[35] This affects the estimation of population characteristics based on survey data. Also, since the degree of bias is not uniform across survey years due to, inter alia, the expanding reach of roads and bus routes, or variations in the intensity and the geographical footprint of insurgency, the state-level estimates of socio-economic characteristics are not comparable over time. In this section, we discuss the impact of reliance on faulty sampling frames and biased samples at two levels. We first examine the impact on descriptive statistics and then on second order statistics derived from them that directly affect policy. The analysis in this section is conducted for the quinquennial rounds since estimates from these rounds are mostly believed to be representative at the sub-national level.[36]

5.1 Impact on Descriptive Statistics

Non-coverage biases sample statistics. Consider, for instance, MPCE, which is widely used in the academic literature on India and is also used in the estimation of poverty rates. Following the preceding discussion, we can argue that MPCE estimates obtained from surveys in Jammu and Kashmir and Nagaland are likely to be biased upwards because of the nature of non-coverage, viz, the non-inclusion of relatively inaccessible, economically backward, and/or insurgency prone areas. Table 4 suggests that this might well be the case. The average MPCE of these states is close to the maximum MPCE of all states of India, particularly in the periods associated with high non-coverage,

even though their per capita incomes are far below the maximum for states. In fact, prior to the 66th and 68th rounds, while the MPCE of Jammu and Kashmir and Nagaland approached the maximum MPCE in India, their per capita incomes moved away from the corresponding maximum (Figure 4).[37] In other words, these states supported a much higher level of consumption in comparison to their income levels.

Table 5 examines the relationship between MPCE and non-coverage rate for Jammu and Kashmir and Nagaland. We regress the ratio of state's MPCE to India's MPCE on the non-coverage rate for these states. The results suggest that MPCE is positively correlated with non-coverage, i.e., MPCE is higher in the years with higher non-coverage. We can say that non-coverage led to an overestimation of MPCE as economically backward areas, which are characterised by difficult terrain and climatic conditions and relatively higher exposure to insurgency, were left out. In principle, it can be argued that the expenditure is higher in the two states because of the higher cost of living on account of, among other things, inaccessibility, climatic conditions, dietary practices, and difficult terrain. If that is so, the expenditure should be high relative to other states in all rounds, rather than in select rounds.

Table 5 presents results for three different specifications. The first uses a measure of non-coverage derived from Table 1. The second augments the first with a variable that captures economic growth of the state relative to the median economic growth observed among all the states during the three years preceding the survey, in order to take into account the possibility of change in MPCE driven by economic growth. The coefficient of growth is not statistically significant, however.[38] The last uses an alternative but crude measure of sample non-coverage (insofar as it does not take into account size differences between FSUs) derived from Table 2. The results from the three specifications are comparable, and each of the three models accounts for about two-thirds of the variation in the dependent variable.

The estimated intercept along with the state and sector dummies indicates that the MPCE of rural areas of Jammu and Kashmir (Nagaland) is about 1.16 (1.47) times higher than that of rural India during the period under consideration. The corresponding values for urban sector are lower, 0.95 for Jammu and Kashmir and 1.24 for Nagaland. In Model 1 (Table 5), a one per cent increase in the non-coverage rate increases the ratio of MPCE (state) to MPCE (India) by about 0.31 per cent. Thus, if non-coverage rate is, say, 50 per cent, the ratio of MPCE (state) to MPCE (India) will increase by about 15.5 percentage points. These results indicate that in the period under consideration, the ratio of MPCE (state) to MPCE (India) in the rural areas would have increased by as much as 19.13

---

35 See Kruskal and Mosteller (1979) for different notions of representativeness.

36 We omit the first quinquennial round (viz., 27th round, 1972-73) and instead use the 28th round (1973-74) because the *Task Force on Minimum Needs Projections and Effective Consumption Demands* relied on MPCE from the latter to estimate the incidence of poverty (GoI, 1979a: 10). The report of the Task Force (GoI, 1979a) does not mention the reason for this. There could be two reasons: (i) 1972-73 was a drought year and (ii) Substantial changes were made in the NSSO sampling design in the 28th round, including the use of the 1971 Census as the sampling frame (GoI, 2004). GoI (1986b: S-3) notes that "To meet some ad hoc requirement, a Consumer Expenditure survey was also conducted in 1973-74 (NSS 28th round)." However, the consumption estimates from the 28th round may not have addressed seasonality adequately since the survey was conducted from October 1973 to June 1974 and was not spread over the full agricultural year (Suryanarayana, 2009).

---

37 Since the state with maximum per capita income could be an outlier, we also checked ranking of states by per capita income. Jammu and Kashmir's rank fluctuated between 4 and 6 during 1973-1987 and declined gradually from 10 in 1987-88 to 19 in 2011-12, while that of Nagaland fluctuated between 6 and 18 during 1973-2011.

38 The correlation between both rural and urban MPCE and the state level growth variables is negative and statistically insignificant. The growth variable is not available separately for the rural and urban areas.

per cent in Jammu and Kashmir (for the maximum non-coverage rate, i.e., 62.53 per cent during 1993-2002) and 19.93 per cent in Nagaland (for the maximum non-coverage rate, i.e., 65.13 per cent in 61st, 66th, and 68th rounds). This translates into the overestimation of MPCE by as much as 16 per cent in Jammu and Kashmir and 13 per cent in Nagaland for rural areas, and 20 per cent and 4 per cent in the respective urban areas.[39]

5.2 Impact on Poverty Estimation

The estimation of consumption expenditure-based poverty rates requires information on the poverty line and the distribution of consumption expenditure. As shown in Figure 3, given an expenditure distribution, an over- (under-) estimated poverty line will lead to over- (under-) estimation of poverty rate (head count ratio). Likewise poverty rate will be underestimated if the lower end of the distribution is only partially observed or the left tail is thinner than it ought to be because of the unavailability of the consumption expenditure data on the poorer groups.[40] If the poverty line is overestimated and the distribution is only partly observed, as would be the case with Jammu and Kashmir and Nagaland, the overall effect on poverty rate is not clear *a priori*. We will first examine the impact of the sample survey inadequacies on the poverty lines for the two states estimated by the Expert Groups chaired by Tendulkar (2009) and Rangarajan (2011) (GoI, 2014b).

State-specific poverty lines are computed from the national poverty line using state-wise relative (to all-India) prices for the corresponding sector (viz., rural/urban).[41] The Expert Group (Tendulkar) followed a different approach and used the urban poverty ratio for 2004-05, estimated using the Expert Group (Lakdawala) methodology, to recalibrate the urban poverty line for 2004-05.[42] It then derived the state-specific poverty lines for urban areas using state-wise relative (to all-India) urban prices. Once the urban poverty line for a state was arrived at, the Expert Group applied a within-state

rural-relative-to-urban price index to estimate the poverty line for the rural areas. The state-level poverty lines for subsequent quinquennial rounds were estimated by applying a state-level price index on the corresponding figures for 2004-05. The 2011 Expert Group (Rangarajan) too followed similar criteria, though it reverted to the practice of estimating separate all-India poverty lines for rural and urban areas. Therefore, only state relative to all-India price indices were required and within-state rural-relative-to-urban prices were not needed. In both the sectors (urban and rural), the price relatives were arrived at based on NSSO surveys and Consumer Price Index/CPI (Agricultural Labourers/AL and Industrial Workers/IW). Thus, assuming the all-India consumption distribution of the NSSO is reliable, the price relatives would be the main source of distortion of the poverty lines in Nagaland and Jammu and Kashmir since the poverty lines for states are based on an adjustment of all-India consumption distribution using state-specific price relatives. Both the expert groups, as noted earlier, relied on the NSSO data for price relatives. The Expert Group (Tendulkar) used implicit prices from NSSO surveys for most eatable items, many durable goods, and education and health services and CPI (AL) and CPI (IW) for entertainment, personal care, and other miscellaneous goods and services and durables. The Expert Group (Rangarajan) followed a broadly similar approach (GoI, 2014b: 16, 63). Insofar as the NSSO data suffer from frame and sample non-coverage, the price data used for rural and urban state poverty lines are biased as well.

Table 6 (Panel A) presents the poverty lines estimated by expert groups for Nagaland and Jammu and Kashmir, along with the highest poverty line for Indian states. In the previous section, we have shown that NSSO Surveys overestimated the MPCE for Jammu and Kashmir and Nagaland by systematically excluding remote and insurgency affected areas. Non-coverage would have affected the estimation of the poverty lines for the two states via its effect on prices. Insofar as NSSO samples in Nagaland and Jammu and Kashmir were biased in favour of relatively well-to-do villages and urban areas within the state, prices are likely to be overestimated.[43]

Indeed this appears to be the case with Nagaland, which had the highest or fairly close to the highest poverty line in all of the country in 2004-05, 2009-10, and 2011-12. In 2009-10, the poverty lines (Tendulkar) for Nagaland were Rs 1017 (rural) and Rs 1148 (urban) compared to the corresponding all-India figures of Rs 673 and Rs 860, respectively. For the sake of comparison, note that the next highest poverty lines in the rural sector were Rs 850 (Mizoram, a state which is among the most urbanised states of the country and the only north-eastern state to have settled insurgency through negotiation), and Rs. 830 (Punjab, the leader in the field of agriculture and among the most prosperous states). The corresponding figures for the urban sector were Rs. 1040 (Delhi, the national capital), Rs. 1035 (Sikkim), and Rs. 1025 (Goa in the urban sector). The last two are major tourist attractions in their respective regions. In most of the years, the poverty lines of Nagaland are close to the maximum among states because of, among other things, the high degree

---

39 The estimates of overestimation have been arrived at after assuming that the non-coverage in the two states does not affect the average MPCE of the country (viz., the denominator in the dependent variable). This is not an unreasonable assumption as the combined share of the two states in the country's population and GDP is about one per cent. Also, recall that bias has two components, namely, the degree of non-coverage and the magnitude of the difference between the true MPCE of the villages included in and excluded from the sampling frame. We have been able to control only the former in regressions.

40 Truncation shown in the first panel of Figure 3 is an oversimplification. In case of the distributions such as the one shown in the second panel of the figure, all observations to the left of the point of truncation are unobserved. In our case, the observations are not available for those situated beyond the five km of the bus route in Nagaland or in remote areas of insurgency affected districts of Jammu and Kashmir. While most of these points are likely to be located at the bottom of the distribution it may not be possible to define a truncation point because the lower end of the distribution is not entirely unobserved. So, a combination of the second and third panels in Figure 3 better captures our situation as the lower end of distribution is under-observed rather than entirely un-observed.

41 The Expert Group (Lakdawala) was the first to provide separate poverty estimates for states, but it had used Assam's poverty ratio for all the states of the North East, including Nagaland. Therefore, analysis in this section is confined to the poverty lines and ratios estimated by the post-Lakdawala Expert Groups (viz., Tendulkar and Rangarajan), which provide separate poverty lines and poverty ratios for each state.

42 While the poverty ratios for urban India (2004-05) estimated using Lakdawala and Tendulkar methods are identical, the corresponding poverty lines are not because they differ with respect to the reference period for some items of poverty line basket used in the NSSO surveys. The Expert Group (Lakdawala) used URP (Uniform Reference Period) and the Expert Group (Tendulkar) used MRP (Mixed Reference Period). In the former case, expenditure is recorded for a "uniform" reference period of 30 days preceding to the date of survey and in the latter the expenditure for some items is recorded with a reference period of 365 days and for the rest, with 30 days.

43 There is substantial rural-urban difference in price levels in India (see, for instance, Bhattacharya and Chatterjee, 1971; Majumder, Ray, and Sinha, 2012).

of non-coverage, inadequate sample size, and sampling bias in NSSO surveys in favour of better off areas. The poverty lines of Jammu and Kashmir are not close to the maximum among states because during 2004-12 the degree of non-coverage was much lower than Nagaland. Also, there is a difference in the nature of non-coverage of the two states. In rural Nagaland, all the remote areas were excluded. However, the non-coverage in Jammu and Kashmir was confined to a few districts and the remaining districts were not affected. Thus, the remote areas of the remaining districts were included in the sample for Jammu and Kashmir. This difference in nature of non-coverage is likely to translate into more biased estimates for Nagaland. Since it is the price level that matters for the poverty line estimation (and not the expenditure distribution), the difference in the nature of non-coverage partly explains the fact that Nagaland's poverty line is almost always the highest among the Indian states, whereas that of Jammu and Kashmir is not.

While only Nagaland's poverty lines were strongly affected, the poverty ratio seems to have been underestimated in both the states. Both the states have reported the lowest incidence of poverty among all Indian states at least once in the rural sector (Table 6: Panel B). Barring a few instances, Jammu and Kashmir has ranked in the topmost quantile, when the states are arranged in the ascending order of the incidence of poverty. Nagaland, on the other hand, has shown substantial volatility.[44] Nagaland, which was among the two states with the least incidence of rural poverty in 1993-94 and 2004-05, ranks 10th in 2009-10 in the rural sector. Likewise, the ranking in terms of urban poverty were 7th, 1st, and 21st in 1993-94, 2004-05, and 2009-10, respectively. Very low incidences of poverty were reported for Nagaland despite the fact that the poverty lines in Nagaland are on the higher side. With high poverty lines, one expected a higher incidence of poverty in Nagaland, and certainly not the lowest. Are Nagaland and Jammu and Kashmir really better off than most of the Indian states in terms of rural poverty? Or, are a high poverty line and a low poverty rate in Nagaland and a low poverty rate in Jammu and Kashmir artefacts of the inadequacies of NSSO surveys that only partially observe the lower end of the consumption distribution (recall Figure 3)?[45] A high poverty line, a low poverty rate *compared* to other states (which unlike Nagaland are not economically stagnant), and a low per capita state domestic product *compared* to other states cannot all be simultaneously true. An incomplete MPCE distribution due to the non-coverage of the population at the bottom end seems to have led to an underestimation of the incidence of poverty.[46] In Nagaland, the degree of exclusion is so much that despite the overestimation of the poverty line the state had the lowest poverty head count in the entire country in several years between 1993-94 and 2011-12 (Table 6). The decline in Nagaland's MPCE relative to the rest of the country over the

years is partly a consequence of the changing sampling frame. While further research is needed in case of Jammu and Kashmir, non-coverage of remote and disturbed districts seems to be one reason for underestimation of poverty, especially, in rural areas.

6.0 Concluding remarks

We began the discussion by suggesting that development, democracy, and data deficits are interrelated and then examined at length the problem of non-coverage in NSSO surveys in insurgency affected Jammu and Kashmir and Nagaland to illustrate the relationship between political unrest and data quality that in turn affects policy-making. We also briefly discussed other North Eastern states, particularly, Tripura. Non-coverage and incidence of insurgency (as reflected in overall annual fatalities) were shown to be correlated. We also showed that non-coverage affects estimates of consumption expenditure and prices, which in turn affect poverty estimates. More specifically, the decline in Nagaland's MPCE and the fluctuation of Jammu and Kashmir's MPCE relative to the rest of the country over the years are shown to be partly a consequence of the changing reach of surveys.

Academic analyses of development problems such as poverty mostly omit Jammu and Kashmir and North Eastern states due to lack of data. However, those concerned with poverty profiles at the sub-state level occasionally uncritically use the poverty estimates for these two states. Studies in the first category are likely to suffer from sample selection bias because of the omission of certain types of states, where data and development deficits are structurally interconnected. The importance of these nine states (Jammu and Kashmir and eight North Eastern states) should be evident from the fact that together they constitute about a third of India's states, even though their combined population share is small (about five per cent of India's population).

The lack of data is mostly explained by irregular surveys and unrepresentative samples. Small sample size has often been blamed for unrepresentative samples in the North Eastern states (for instance GoI, 2011). However, there are other factors, which affect the representativeness of NSSO samples, to which this paper has drawn attention. We argued that NSSO samples are unlikely to be representative for Nagaland because of the arbitrary restriction of its surveys to villages within 5 km of bus routes, whereas the bulk of the rural and tribal population is located farther from roads. The spatial distribution of population vis-à-vis roads varies across both districts and NSS rounds that makes inter-temporal comparison difficult. Even the recent change that extends the coverage of NSSO samples to hitherto untouched rural areas fails to address this problem as it allocates very few FSUs to these areas compared to their population share. However, it is likely that the NSSO will gradually expand the size of the special stratum in Nagaland. We also highlighted the sharp variations in non-coverage in Jammu and Kashmir (and Tripura). Just as in Nagaland, there were systematic differences between the FSUs surveyed and those left out, which affected the representative character of the samples. Since insurgency was the reason for this inadequacy in both the states, the rural samples were often restricted to relatively accessible

---

44   For instance, poverty in rural Nagaland halved between 1993-94 and 2004-05 but bounced back to the 1993-94 level by 2009-10. In urban Nagaland, the swings are even more unrealistic.

45   The estimates from surveys conducted by the respective governments provide a very different poverty profile. In rural Nagaland, the incidence of poverty was 62.24 per cent in 2002 (GoN, nd) and in Jammu and Kashmir the same was 26.14 per cent in rural and 7.96 in urban in 2007-08 (J&K Government, 2008). The Expert Group's (Tendulkar) estimate of the incidence of poverty for 2004-05 was 10 per cent for Nagaland (Rural), 14 per cent for Jammu and Kashmir (Rural), and 10 per cent for Jammu and Kashmir (Urban).

46   The degree of underestimation depends on inter alia the density of the MPCE distribution in the neighbourhood of the poverty line.

villages. Accessible villages are likely to be better connected to the urban economy and better covered by government's developmental programs. This means that the rural sample was skewed toward relatively well-off villages. Therefore, the sample is unrepresentative and will yield biased estimates of indicators. We also argued that census is not a reliable sampling frame in case of Nagaland and Jammu and Kashmir.

In all fairness, it has to be noted that in some cases the NSS reports either do not provide estimates or highlight inter-temporal incomparability on account of non-coverage. For instance, NSS Report 401 (50th Round) neither includes Jammu and Kashmir among major states, nor provides MPCE estimates for the state. Likewise, Report 521 (61st Round) compares estimates from 50th and 61st Rounds without including Jammu and Kashmir. Similarly, Report 555 (68th round) cautions against direct comparison of the estimates of the 66th and 68th Rounds. However, the NSSO's cautionary notes are buried in the reports and seem to fail to prevent unintentional misuse of its data.

Despite major changes in the extent of non-coverage that limit inter-temporal comparability of NSS results particularly at the level of NSS regions, social scientists (Atkin, 2013; Palmer-Jones and Sen, 2003; Jha and Sharma, 2003, for instance) have compared NSS region level estimates of characteristics for different rounds, at times even for those regions that were not covered in a few rounds. A related, but different aspect of the above problem is found in the Sachar Committee's report (GoI 2006) on the social, economic and educational conditions of the Muslim community in India. The NSSO data were crucial for the Sachar Committee because among other things, unlike the Census, NSSO data allow the identification of social categories such as Other Backward Classes (OBCs) among Muslims (GoI, 2006: 5, 189-190, 203). The Sachar Committee Report relies inter alia upon 43rd to 61st Rounds of NSSO (GoI, 2006), which are not comparable in the case of Jammu and Kashmir. Being the only Muslim majority state in the country, Jammu and Kashmir was particularly important for the Committee. The discussion on the impact of non-coverage on estimation of population characteristics in this paper directly applies to the use of NSSO data in the Report (see, for instance, Table 10.3, Appendix Tables 1.1-13, 4.3).

The NSS reports were found lacking in information that is necessary for understanding the sampling frame. This paper highlighted a few instances of lack of attention to essential information. First, the NSSO used the inflated headcounts of Nagaland for its projections even after the state government had formally rejected the faulty census data. It could have at least alerted the readers to this discrepancy in census and discussed the possible impact of the discrepancy on its survey statistics. Second, Nagaland's population was over-counted in 2001 and the actual population in 2011 was less than the manipulated head count of 2001. The correction was interpreted by the NSSO as a contraction in the population of Nagaland and it under-projected the population after 2013. Third, since the publication of the results of the 2001 Census, the NSSO sampling frame for Manipur has excluded three sub-divisions of Senapati district. Yet this was not mentioned in any of the NSS reports. Also, there are discrepancies between the reports and unit level data, e.g.,

as per NSS reports all urban FSUs allocated to Nagaland were surveyed in the 55th round (Table 2), even though the unit level data indicate the exclusion of three districts (Table 1). Fourth, non-coverage and differences across rounds in this regard are not explicitly discussed in reports. For example, non-coverage in case of Tripura and Manipur are not recorded in reports (see Footnote 30). Fifth, there are discrepancies between different government sources. For example, GoN (2013: 2) suggests that the NSSO began operations in Nagaland in 1972, whereas GoI (1976: 1) mentions that NSSO began surveys in urban Nagaland in the 23rd round (July 1968 – June 1969). Sixth, the NSSO did not explain how it identified the villages within five kms of bus routes in Nagaland. As a result it is not clear why the number of such villages did not change between 1988 and 2004.

If we step back from the immediate problem of non-coverage in NSSO surveys, we can explain the poor quality of data from a political-economic perspective by arguing that census and household survey statistics that feed into government's development planning process can be treated as public goods. Historically, governments have been weak in low population density areas (Herbst, 2000; also Jerven, 2013: 2). We can add that a weak state with a poor public good provision system comes in contact with its subjects infrequently and, to that extent, is less likely to be capable of collecting reliable information about them. Also, a state's ability to collect information about the lives of its citizens is a reflection of its power and/or legitimacy (Jerven, 2013: 112). If a state lacks in power and legitimacy, then it may not have the means to generate reliable information. Jammu and Kashmir is a low population density state (124 persons per sq km[47] compared to the national average of 382 in 2011, see GoI, 2013d), where the power and legitimacy of the government are circumscribed by a three decade old insurgency in the Kashmir Valley. Such a government cannot provide good quality public goods including demographic and development statistics. Nagaland's case is comparable insofar as it is a low density (119 persons per sq km, GoI 2013d)[48] area where tribes that are exempt from taxation account for the bulk of the population. The tax exemption was initially introduced keeping in mind poverty, and also to avoid open rebellion in a politically restive state.

There is another way of looking at statistics in ethnically diverse politically restive regions. Ethnic competition often negatively affects the provision of public goods (Alesina and La Ferrera, 2005). So, the quality of government statistics is likely to be poor in ethnically divided societies. Jammu and Kashmir's ethno-linguistic and religious diversities are geographically distributed across three regions. Competition among these regions and communities has been the defining feature of the state's politics. While it might not affect NSSO surveys directly, it has an indirect effect through census that is used to draw NSSO samples. Guilmoto and Rajan (2013) draw attention to how religious polarisation has affected the quality of census data for Jammu and

---

47   There is considerable variation in population density within the state ranging from 3 persons per sq km in Leh to 1148 in Ganderbal.

48   The actual density is about 3.6 per cent lesser than the official estimate (Agrawal and Kumar, 2017).

Kashmir. Likewise Nagaland, which accounts for about 0.2 per cent of India's population and 0.5 per cent of its area, is home to about 20 per cent of India's non-scheduled languages reported in the census (authors' calculations based on GoI, 2008) and about 2.5 % of all mother tongues of India (authors' calculations based on The Hindu, 2013). Intense competition among tribes ahead of the 2002 delimitation resulted in large-scale manipulation of the 2001 Census (Agrawal and Kumar, 2012b; Kumar and Agrawal, 2016).

Put together, the unreliability of government statistics in Jammu and Kashmir and Nagaland highlights systemic problems that have wider implications for our understanding of the relationship between state, statistics, and policy-making. First, it raises doubts about the reliability of statistics generated from sample surveys in other conflict-ridden and inaccessible areas. For instance, it is not clear to what extent NSSO is able to reach villages in Naxalism-affected Central India, where the reach of the government is limited (see Hindustan Times, 2017, which suggests that parts of Chhattisgarh are being subject to field survey for the first time). Second, while we have examined only NSSO samples, other demographic and socio-economic surveys conducted in Nagaland and in Jammu and Kashmir are also likely to suffer from similar problems. Third, insofar as federal redistribution is linked to the level of poverty measured using NSSO's MPCE data (as in GoI, 2013b, for instance), faulty data distort redistribution.[49] Fourth, economically underdeveloped, ethnically divided, and politically disturbed states that need better policies are also the areas where government statistics are relatively less reliable. Among other things, poorly designed policies contribute to underdevelopment and underdevelopment contributes to political instability, which in turn affects the availability and the quality of statistics by hindering government surveys. In short, democracy, development, and data deficits are closely related in states such as Nagaland and Jammu and Kashmir. Piecemeal reforms to address the problems in specific surveys might not succeed unless the other links of the structural relationship are reworked.

---

49 "The funding of the IAY [Indira Awaas Yojna] is shared between the Centre and State in the ratio of 75:25. From the year 2005-06, the allocation of funds under the Indira Awaas Yojana to the States/UTs is being made on the basis of rural housing shortage estimated by the Office of the Registrar General of India based on 2001 Census and the poverty ratio, as approved by the Planning Commission, giving 75% weightage to housing shortage and 25% weightage to poverty ratio" (Annual Report 2006-07: 49, Ministry of Rural Development).

References:

Alesina, A., Ozler S., Roubini, N., Swagel, P. (1996). Political Instability and Economic Growth. *Journal of Economic Growth*, 1, 189-211.

Alesina, A., La Ferrara, E. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3), 762-800.

Agrawal, A., Kumar, V. (2012a). An Investigation into Changes in Nagaland's Population between 1971 and 2011. *IEG Working Paper No 316*, Institute of Economic Growth, Delhi.

Agrawal, A., Kumar, V. (2012b, July 12). Number games in Nagaland. *The Hindu*, p. 11 (Corrigendum July 14, p. 13)

Agrawal, A., Kumar, V. (2013). Nagaland's Demographic Somersault. *Economic and Political Weekly*, 48(39), 69-74.

Agrawal, A., Kumar, V. (2014). Infirmities in NSSO Data for Nagaland. *Economic and Political Weekly*, 49(12), 20-22.

Agrawal, A., Kumar, V. (2017). Notes on Nagaland's Area. *Working Paper No 4*, Azim Premji University, Bengaluru.

Atkin, D. (2013). Trade, Tastes, and Nutrition in India. *The American Economic Review*, 103(5), 1629-1663.

Bhalla, S. (2014). Behind the Post-1991 'Challenge' to the Functional Efficiency of Established Statistical Institutions. *Economic and Political Weekly*, 49(7), 43-50.

Bhat, B. A. (nd). *Miss Missed or Miss Missing: Reality behind declining child sex ratio in Kashmir-India*. Retrieved from http://iussp.org/sites/default/files/event_call_for_papers/Extended%20Abstract%20for%20iussp%20Conference.pdf

Bhattacharjee, G. (2016). *Special Category States of India.* Oxford University Press: New Delhi.

Bhattacharya, N., Chatterjee, G. S. (1971). On Rural-Urban Differentials in Consumer Prices and Per Capita Household Consumption in India by Levels of Living. *Sankhya: The Indian Journal of Statistics, Series B*, 33 (3/4), 355-370.

Biemer, P. P., Lyberg, L. E. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc: New Jersey.

Chandrasekhar, C. P., Ghosh, J. (2013, July 8). The employment bottleneck. *Business Line*, July 8.

Chaudhuri, S., Gupta, N. (2009). Levels of Living and Poverty Patterns: A District-Wise Analysis for India. *Economic and Political Weekly*, 44(9), 94-110.

Dandekar, V. M., Venkataramaiah, P. (1975). *Database of Indian Economy: Role of Sample Surveys (Vol II)*. Calcutta: Statistical Publishing Society and Hyderabad: The Indian Econometric Society.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Washington, DC: The World Bank.

Deaton, A. (2003). Prices and Poverty in India, 1987-2000. *Economic and Political Weekly*, 38(4), 362-68.

Deaton, A. (2008). Price Trends in India and Their Implications for Measuring Poverty. *Economic and Political Weekly*, 43(6), 43-49.

Deaton, A., Dreze, J. (2002). Poverty and Inequality in India: A Re-examination. *Economic and Political Weekly*, 37(36), 3729-48.

Deaton, A., Kozel, V. (2005). Data and Dogma: The Great Indian Poverty Debate. *The World Bank Research Observer*, 20(2), 177-99.

Deolalikar, A., Dubey, A. (2008). *National Sample Survey (India)*. International Encyclopedia of the Social Sciences. NY: Thomson Gale.

Economic Times. (2013). "Poverty line low, need to revisit methodology, says Montek Singh Ahluwalia." July 30. Retrieved from http://economictimes.indiatimes.com/news/economy/policy/poverty-line-low-need-to-revisit-methodology-says-montek-singh-ahluwalia/articleshow/21460171.cms

Economic Times. (2015). "NITI Aayog's flexible fund may benefit special category states," May 20. Retrieved from http://articles.economictimes.indiatimes.com/20150520/news/62413535_1_nitiaayogcategorystatusspecialcategorystates

Government of India/GoI. (nd). Urban Frame Survey. *National Sample Survey Office.* Retrieved from http://mospi.nic.in/Mospi_New/site/inner.aspx?status=3&menu_id=55

Government of India/GoI. (1953). *Census of India 1951, Volume I India, Part I - A – Report.* Registrar General, India: New Delhi.

Government of India/GoI. (1977). Part II: Survey Results, Consumer Expenditure, NSS 28th Round (1973-74). *Sarvekshana - Journal of the National Sample Survey Organisation,* 1(1), 49-51, S1-S131.

Government of India/GoI. (1979a). *Report of the Task Force on Projections of Minimum Needs and Effective Consumption Demand.* Perspective Planning Division, Planning Commission: New Delhi.

Government of India/GoI. (1979b). Introduction: Consumer Expenditure, NSS 27th Round (October 1972 – September 1973). *Sarvekshana - Journal of the National Sample Survey Organisation*, 3(1), S289-290.

Government of India/GoI. (1984a). *Census of India 1981, Series 15 Nagaland, District Census Handbook" (for different districts).* DCO Nagaland: Kohima and RGI: New Delhi.

Government of India/GoI. (1984b). Introduction. *Sarvekshana - Journal of the National Sample Survey Organisation*, 8(1), 1-2.

Government of India/GoI. (1986a). Some Results on the Second Quinquennial Survey on Consumer Expenditure. *Sarvekshana - Journal of the National Sample Survey Organisation*, 9(3), S51-S184.

Government of India/GoI. (1986b). A Report on the Third Quinquennial Survey on Consumer Expenditure. *Sarvekshana - Journal of the National Sample Survey Organisation*, 9(4), S154-S184.

Government of India/GoI. (1989). A Note on Pattern of Consumer Expenditure of Scheduled Caste and Scheduled Tribe Households. *Sarvekshana - Journal of the National Sample Survey Organisation*, 12(1), 1-23.

Government of India/GoI. (1993). *Report of the Expert Group on Estimation of Proportion and Number of Poor.* Planning Commission.

Government of India/GoI. (1994). A Note on an Exploratory Survey of Living Conditions of Tribals of Nagaland, NSS 44th Round (July 88 - June 89). *Sarvekshana - Journal of the National Sample Survey Organisation*, 17(1), 49-84.

Government of India/GoI. (1996). *Level and Pattern of Consumer Expenditure, 5th Quinquennial Survey, NSS Report No 402.* National Sample Survey Organisation: New Delhi.

Government of India/GoI. (2001). *Concepts and Definitions Used in NSS.* Golden Jubilee Publication, National Sample Survey Organisation: New Delhi.

Government of India/GoI. (2004). *Evolution of the Sample Design in the Indian National Sample Survey from 1st to 55th Round.* Golden Jubilee Publication, National Sample Survey Organisation: New Delhi.

Government of India/GoI. (2005). *General Population Tables: India, States and Union Territories, Part-1.* Registrar General and Census Commissioner: New Delhi.

Government of India/GoI. (2006). *Social, Economic and Educational Status of the Muslim Community of India: A Report.* Prime Minister's High Level Committee, Cabinet Secretariat, New Delhi.

Government of India/GoI. (2008). *Census of India 2001, Series 1-India, Language: India, States and Union Territories, Table C-18.* Office of the Registrar General, India: New Delhi.

Government of India/GoI. (2009). *Report of the Expert Group to Review the Methodology for Estimation of Poverty.* Planning Commission: New Delhi.

Government of India/GoI. (2011). *Report of the Committee on Optimum Sample Sizes for North Eastern States.* National Sample Survey Office, MOS&PI: New Delhi.

Government of India/GoI. (2013a). *Key Indicators of Household Consumer Expenditure in India', NSS 68th Round, July 2011 – June 2012.* National Sample Survey Office, MOS&PI: New Delhi.

Government of India/GoI. (2013b). *Report of the Committee for Evolving a Composite Development Index of States.* Ministry of Finance: New Delhi.

Government of India/GoI. (2013c). *State Development Report, Jammu and Kashmir.* Planning Commission: New Delhi.

Government of India/GoI. (2013d). *Census of India. 2011, Primary Census Abstract Data Tables (India & states/UTs - District Level)* (Excel format)." Retrieved from http://www.censusindia.gov.in/2011census/population_enumeration.aspx

Government of India/GoI. (2014a). *Level and Pattern of Consumer Expenditure 2011-12, NSS Report No 555.* National Sample Survey Office: New Delhi.

Government of India/GoI. (2014b). *Report of the Expert Group to Review the Methodology for Measurement of Poverty.* Planning Commission: New Delhi.

Government of India/GoI. (2014c). *Order (dated 07-01-2014).* Office of the Registrar General, India. Retrieved from http://www.censusindia.gov.in/2011-Documents/ordermanipur_201401081351.pdf

Government of Nagaland/GoN. (nd). *Census, Dimapur district.* Retrieved from https://www.nagaland.gov.in/portal/portal/StatePortal/UsefulLinks/Census

Government of Nagaland/GoN. (2013). *Housing Conditions and Amenities in Nagaland, National Sample Survey 65th Round.* Directorate of Economics and Statistics: Kohima.

Ghosh, B., De, P. (2005). *India Infrastructure Database 2005.* Bookwell: New Delhi.

GSI. 2011. *Geology and Mineral Resources of Manipur, Mizoram, Nagaland and Tripura.* Miscellaneous Publication, No. 30, Part IV, Vol 1 (Part-2). Geological Survey of India: Kolkata.

Groves, R. M. (2004). *Survey Errors and Survey Costs.* John Wiley & Sons.

Groves, R. M., Floyd J. F. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R. (2004). *Survey Methodology.* John Wiley & Sons.

Guilmoto, C. Z., Irudaya Rajan, S. (2013). Fertility at the District Level in India: Lessons from the 2011 Census. *Economic and Political Weekly*, 48(23), 59-70.

Herbst, J. (2000). *States and Power in Africa: Comparative Lessons in Authority and Control.* New Jersey, Princeton University Press.

Hokayem, C., Bollinger, C., Ziliak, J. P. (2015). The Role of CPS Nonresponse in the Measurement of Poverty. *Journal of the American Statistical Association*, 110(511), 935-945.

Indian Statistical Institute/ISI. (1953). The National Sample Survey. General Report No 1. First round: October 1950-March 1951. *Sankhyā: The Indian Journal of Statistics, (1933-1960)*, 13(1/2), 47-214.

Indian Statistical Institute/ISI. (1958). The National Sample Survey. Eighth Round: July 1954-March 1955. *Sankhyā: The Indian Journal of Statistics, (1933-1960)*, 19(1/2), 29-180.

Indian Statistical Institute/ISI. (1962a). The National Sample Survey. Eleventh and Twelfth Rounds: August 1956-August 1957. Number 33. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 24(1/2), 43-200.

Indian Statistical Institute/ISI. (1962b). "The National Sample Survey. Ninth Round: May-November 1955. Number 40." *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 24:3/4, 277-400.

Indian Statistical Institute/ISI. (1964). The National Sample Survey. Eighth Round: July 1954-April 1955. Number 30. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 26(1/2), 103-99.

Hindustan Times. (2017, May 01). *"This may be the most dangerous mission an Indian official has undertaken since 1947."* Retrieved from www.hindustantimes.com/india-news/chhattisgarh-maoist-affected-area-undertakes-first-govt-survey-since-independence/story-ZcIN9DG9QK5r1ucx1hUQGP.html

J&K Government. (2008). *Below Poverty Line (BPL) Survey-2008: J&K State.* Directorate of Economics and Statistics. Retrieved from http://ecostatjk.nic.in/publications/BPL200809.pdf

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it.* Ithaca, Cornell University Press.

Jha, R., Sharma, A. (2003). Spatial Distribution of Rural Poverty: Last Three Quinquennial Rounds of NSS. *Economic and Political Weekly*, 38(47), 4985-4993.

Kish, L. (1965). *Survey Sampling.* John Wiley and Sons.

Krishnaji, N. (2012). Abolish the Poverty Line. *Economic and Political Weekly*, 47(15), 10-11.

Kruskal, W., Mosteller, F. (1979). Representative Sampling, III: The Current Statistical Literature. *International Statistical Review*, 47(3), 245-65.

Kumar, V., Agrawal, A. (2016). Winning Censuses. *Annual Meeting of the European Public Choice*, Freiburg, March 30-April 02, 2016

Laithangbam, I. (2004, Nov 13). Naga tribal council plans blockade. *The Hindu.* Retrieved from http://www.hindu.com/2004/11/13/stories/2004111301591300.htm

Lessler, J. T., William D. K. (1992). *Nonsampling Error in Surveys.* John Wiley and Sons.

Majumder, A., Ray, R., Sinha, K. (2012). Calculating Rural-Urban Food Price Differentials from Unit Values in Household Expenditure Surveys: A Comparison with Existing Methods and a New Procedure. *American Journal of Agricultural Economics,* 94(5), 1218-1235.

Meyer, B. D., Mok, W. K. C., Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives,* 29(4), 199-226.

Michalski, T., Stoltz, G. (2013). Do Countries Falsify Economic Data Strategically? Some Evidence That They Do. *Review of Economics and Statistics*, 95(2), 591-616.

Minhas, B. S., Jain, L. R., Kansal, S. M., Saluja, M. R. (1987). On the choice of Appropriate Consumer Price Indices and Data Sets for Estimating the Incidence of Poverty in India. *Indian Economic Review*, 22:1, 19-49.

Murthy, M. N. (1967). *Sampling Theory and Methods.* Calcutta: Statistical Publishing Society.

Nagaland Post. (2015, October 10). *Centre examining NE's spl. category status: Minister.* Retrieved from http://www.nagalandpost.com/ChannelNews/Regional/RegionalNews.aspx?news=TkVXUzEwMDA4Nzk1Mg%3d%3d

Nagaland Post. (2016, November 27). *Jaitley clarifies changes in funding pattern for NE.* Retrieved from http://nagalandpost.com/ChannelNews/National/NationalNews.aspx?news=TkVXUzEwMDEwNjA5Mw%3d%3d

Palmer-Jones, R. and Sen, K. (2003). What has luck got to do with it? A regional analysis of poverty and agricultural growth in rural India. *Journal of Development Studies,* 40(1), 1-31.

Panandiker, V. A. P., Umashankar, P. K. (1994). Fertility Control and Politics in India. *Population and Development Review.* 20, 89-104.

Rao, C. R. (1973). Prasantha Chandra Mahalanobis 1893-1972. *Biographical Memoirs of Fellows of the Royal Society,* 19(454). doi:10.1098/rsbm.1973.0017

Rao, T. J. (2010). Official Statistics in India: The Past and the Present. *Journal of Official Statistics,* 26(2), 215-31.

Serra, G. (2014). An Uneven Statistical Topography: The Political Economy of Household Budget Surveys in Late Colonial Ghana, 1951–1957. *Canadian Journal of Development Studies,* 35(1), 9-27.

Shetty, S. L. (2012). Dealing with a Deteriorating Statistical Base. *Economic and Political Weekly,* 44(18), 41-44.

Srinivasan, T.N. (1994). Data Base for Development Analysis: An overview. *Journal of Development Economics,* 44, 3-27.

Suryanarayana, M. H. (2000). How Real is the Secular Decline in Rural Poverty? *Economic and Political Weekly,* 35(25), 2129-40.

Suryanarayana, M. H. (2009). Pursuing Inclusion in India: A Story of Specification Errors. *Indian Growth and Development Review,* 2(2), 155-72.

Suryanarayana, M. H. (2011). Expert Group on Poverty: Confusion Worse Confounded. *Economic and Political Weekly,* 46(46), 36-39.

Suryanarayana, M. H. and Iyengar, N. S. (1986). On the Reliability of NSS Data. *Economic and Political Weekly,* 21(6), 261-64.

Swami, P. (2014). Demography and Discontent: Crisis of Modernity and Displacement in Undivided Jammu and Kashmir. In K. Warikoo (Ed), *The Other Kashmir: Society, Culture and Politics in the Karakoram Himalayas,* IDSA/Pentagon Press: New Delhi, 201-221.

Swami, P. (2016, Aug 09). Talks with Andhra Pradesh on special status at fairly advanced stage - Centre. *The Indian Express.* Retrieved from http://indianexpress.com/article/india/india-news-india/andhra-pradesh-special-status-bill-protest-ysr-congress-arun-jaitley-monsoon-session-2962570/

Vidwans, S. M. (2002). Indian Statistical System at the Crossroads (Part II: Expansion of National Sample Survey). *Economic and Political Weekly,* 37(38), 3943-3955.

Wade, R. (2012). The Politics behind World Bank Statistics: The Case of China's Income." *Economic and Political Weekly*, 47(25), 17-18.

Map 1: Geographies of non-coverage

Note: The states with sustained and high levels of frame non-coverage (Nagaland), sustained and high levels of sample non-coverage and low levels of frame non-coverage (Jammu and Kashmir), high levels of sample non-coverage (Tripura), and low levels of frame non-coverage (Manipur) are highlighted in the above map.

Source: Generated using blank template available at www.mapsofindia.com

Map 2: NSS Regions of Jammu and Kashmir
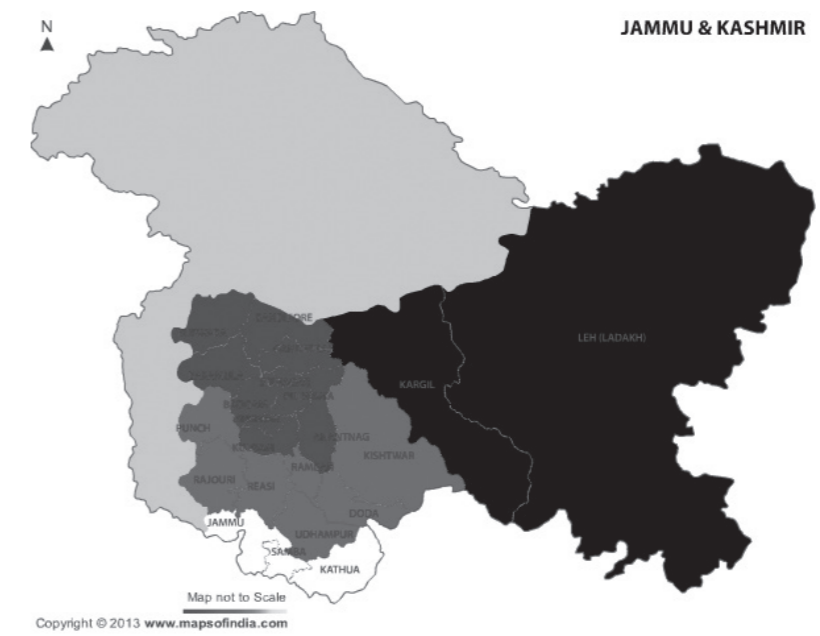
Note: The districts shaded in black fall in NSSO's Ladakh region; dark grey, Jhelam Valley region; and light grey, Outer Hills region. The districts that are not shaded correspond to the Mountainous region (see Section 2.1). Areas shaded in very light grey are under the occupation of Pakistan.

Source: Same as Map 1.

Map 3: Districts of Jammu and Kashmir inadequately covered by NSSO Surveys
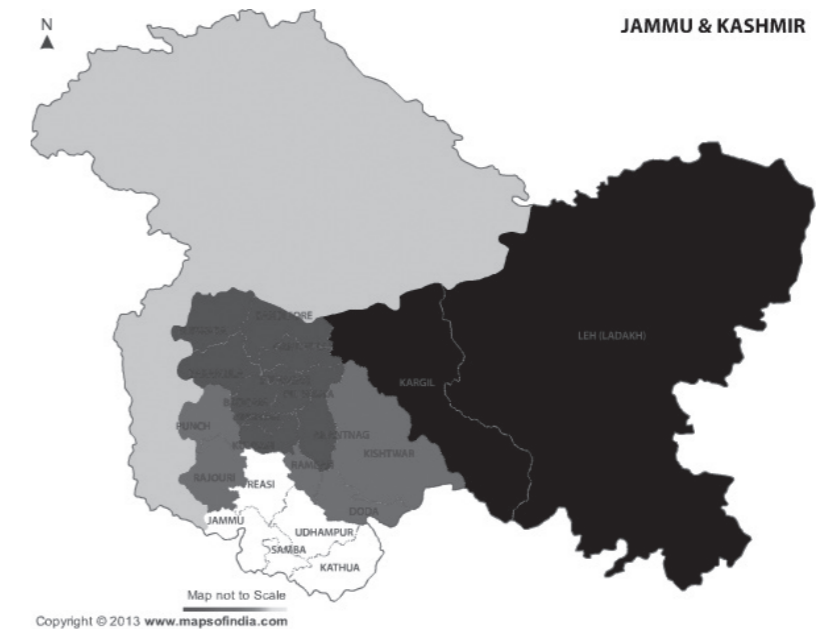
Note: The Districts shaded in black were affected by sustained frame non-coverage, while those shaded in relatively darker shades of grey were affected by sustained sample non-coverage. Areas shaded in very light grey are under the occupation of Pakistan.

Source: Same as Map 1.

Figure 1: A Taxonomy of Errors in Sample Surveys
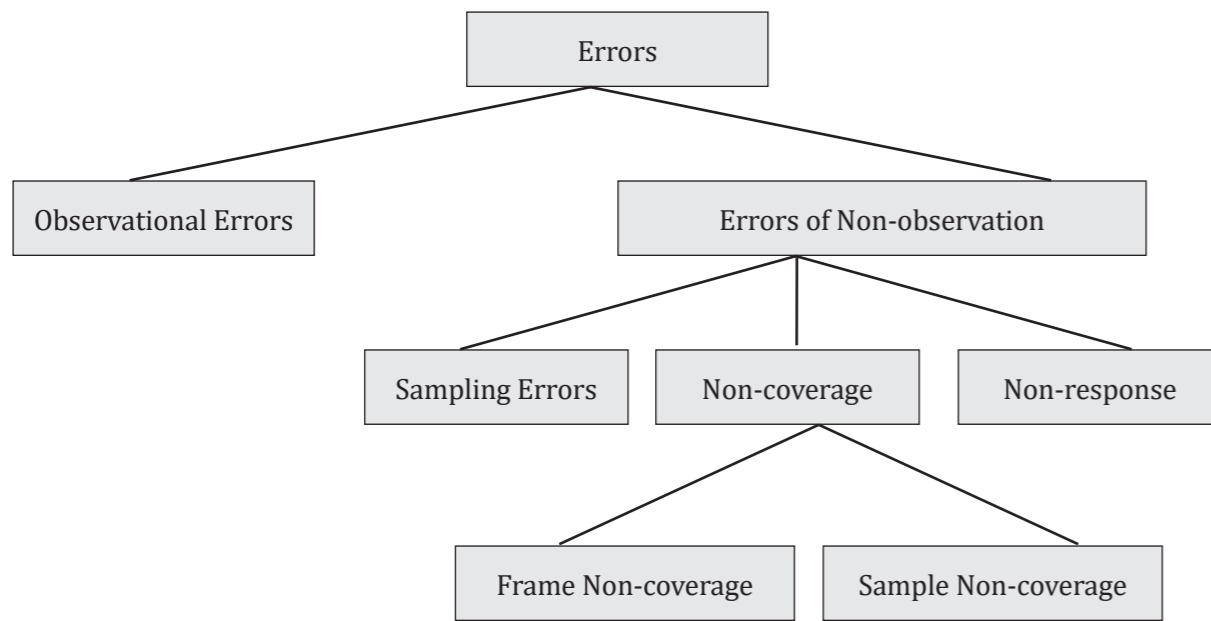
Source: Section 2 and Groves (2004: 10)

Figure 2: Total annual fatalities and non-coverage (per cent)

Jammu and Kashmir (1988-2014)
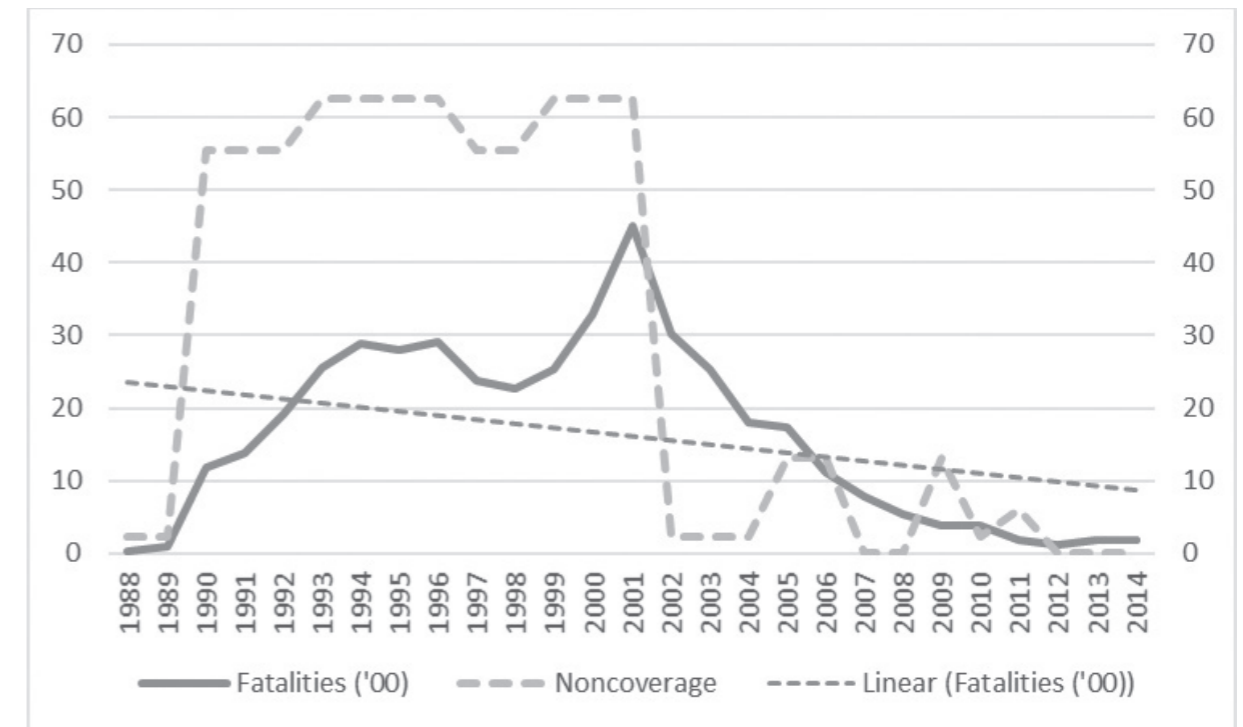
Nagaland (1992-2014)
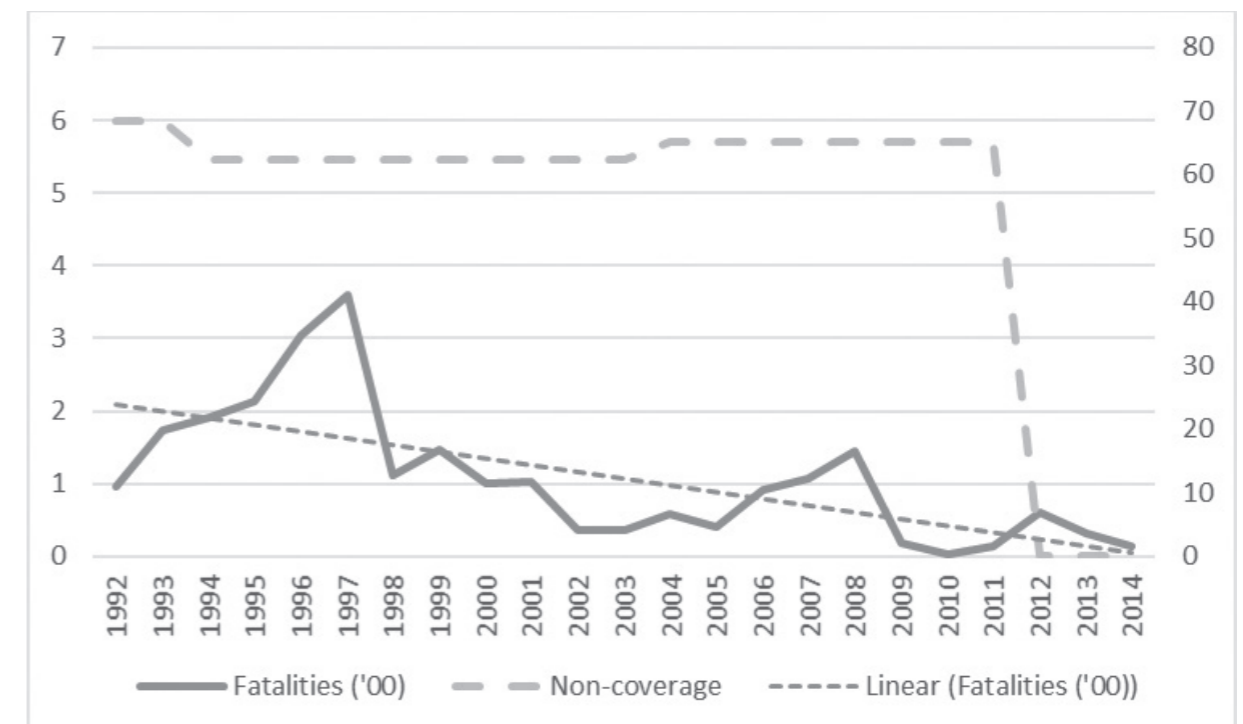
Tripura (1992-2014)
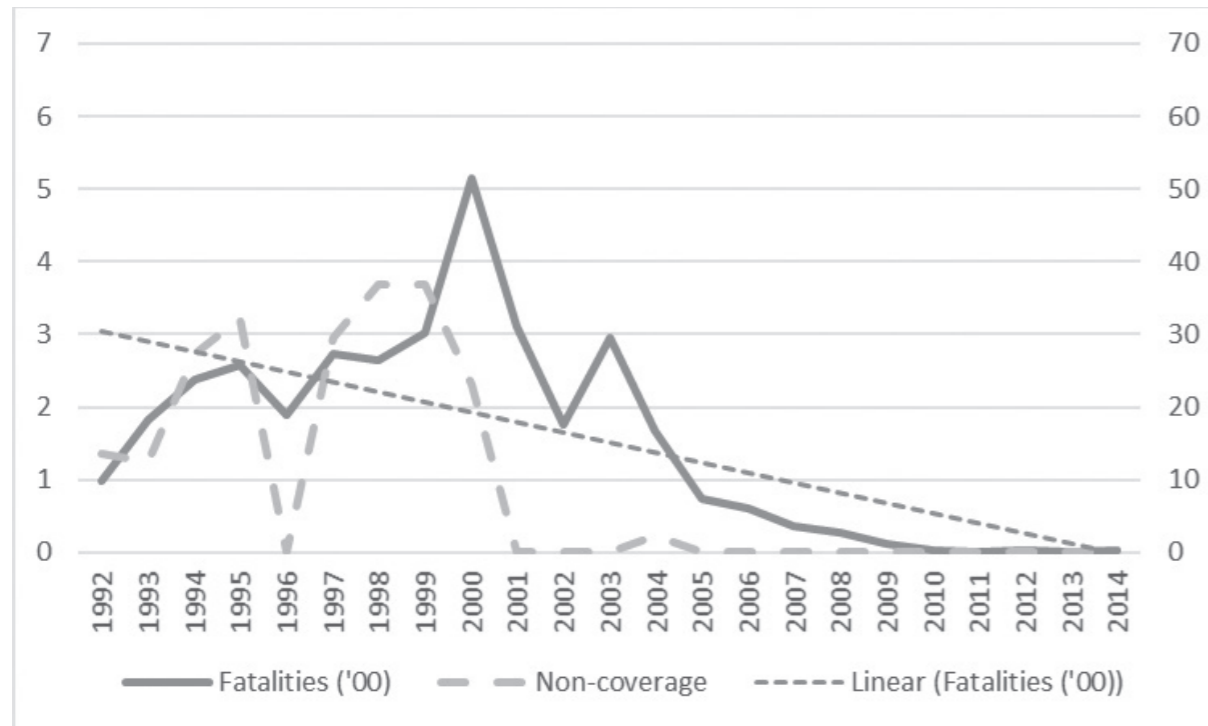
Note: (i) The annual fatalities (in 100s) indicated on the left axis include civilian, security force, and terrorist/insurgent fatalities. The degree of non-coverage (in per cent) is indicated on the right axis. (ii) The terminal points of graphs vary depending on the onset of insurgency and the availability of data on fatalities.

Source: Table 1 (non-coverage in Jammu and Kashmir and Nagaland), Table 2 (non-coverage in Tripura), South Asia Terrorism Portal http://www.satp.org (fatalities)

Figure 3: Poverty line and consumption distribution

Panel A: Consumption distrbution is fully observed

True consumption distribution and poverty line

True consumption distribution and over-estimated poverty line

Panel B: Consumption distribution is truncated at the bottom

Truncated consumption distribution and true poverty line

Truncated consumption distribution and over-estimated poverty line

Panel C: Consumption distribution is partially observed at the bottom end

Skewed distribution (thinner left tail) and true poverty line

Skewed distribution (thinner left tail) and over-estimated poverty line

Note: The horizontal axis in each Panel represents log MPCE, while the vertical axis represents density. The hashed portion is a measure of the incidence of poverty.

Figure 4: Per Capita Income (PCI) and MPCE: Jammu and Kashmir and Nagaland

Legend: — PCI J & K   — MPCE (Rural) J & K   - - MPCE (Urban) J & K

Legend: — PCI Nagaland   — MPCE (Rural) Nagaland   - - MPCE (Urban) Nagaland
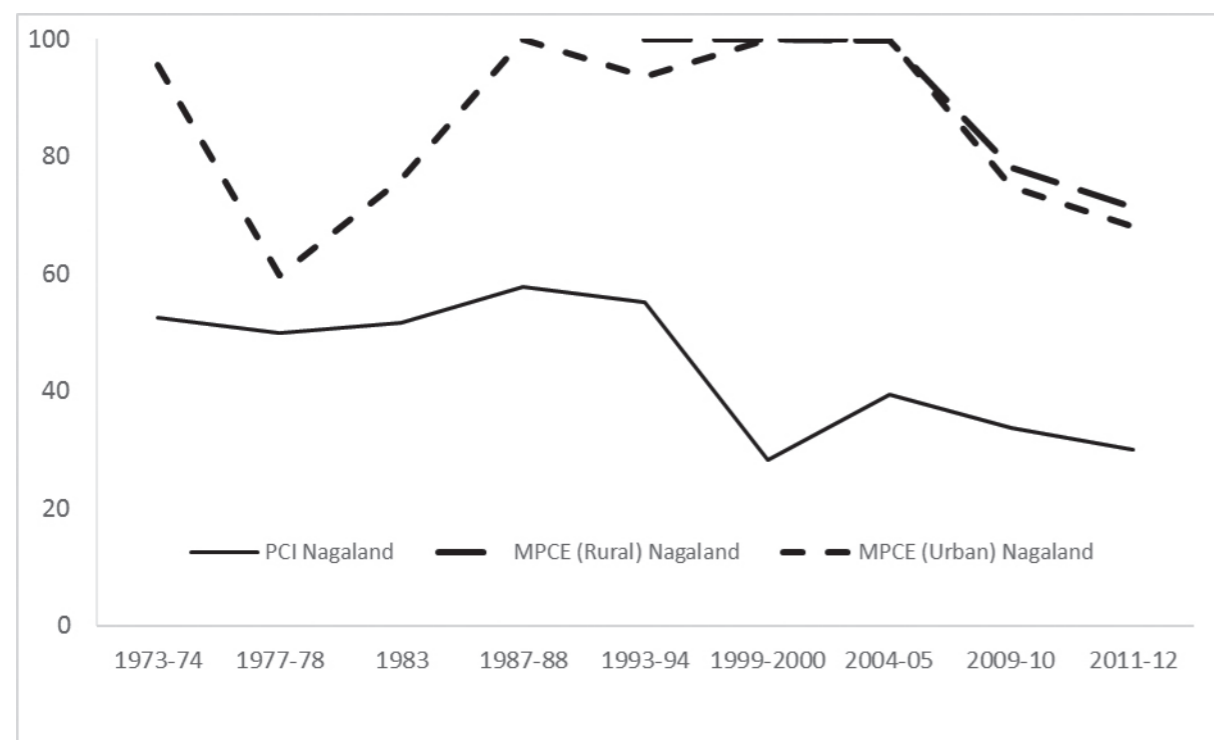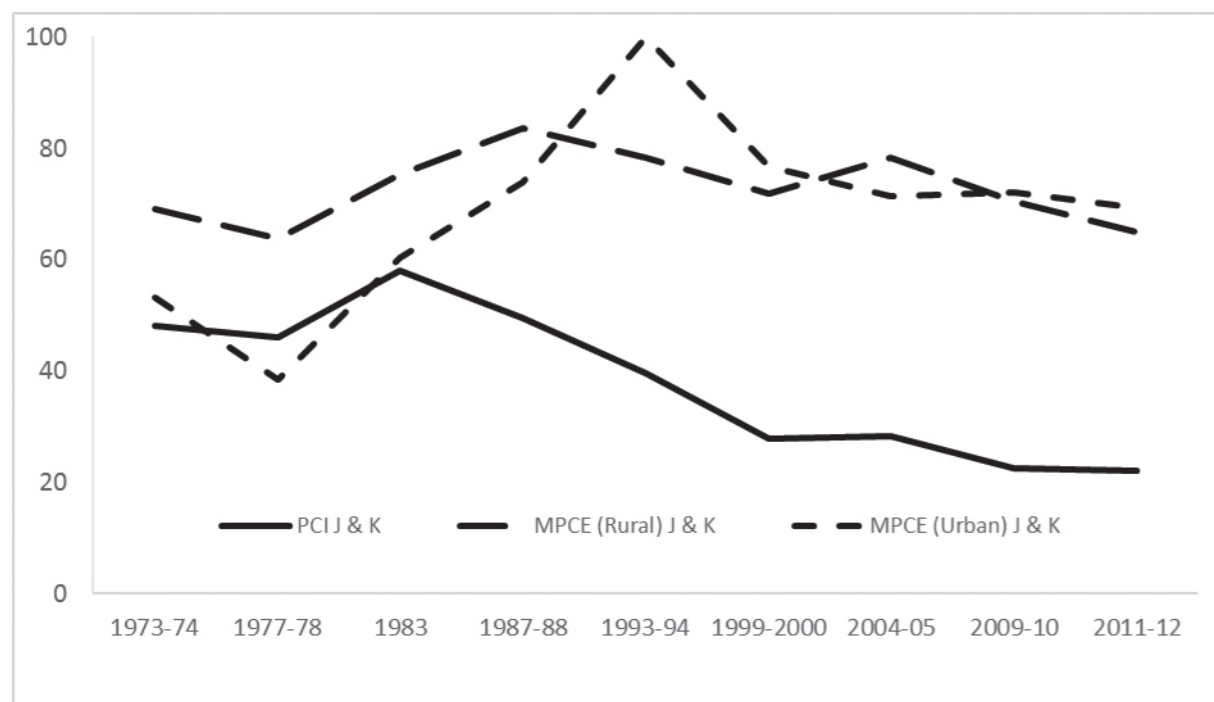
Source: Each curve shows the ratio of variable (PCI or MPCE) of the corresponding state to that of the maximum among all the states. In case of Nagaland, MPCE (Rural) curve starts in 1993-94 as the rural areas of the states were not covered in earlier rounds of NSSO surveys. The pronounced dip in the PCI curve for Nagaland is partly an artefact of population over count in 2001.

Source: Table 4.

Table 1: Areas excluded from NSSO Samples and Non-coverage Rate

Round/ Period	Jammu and Kashmir					Nagaland				
	Census†	Districts excluded (by NSS Region)			Overall (frame + sample) non-coverage rate	Rural			Urban	
		Ladakh (Ladakh)	Kashmir (Jhelam Valley)	Jammu†† (Outer Hills)		Census†	Areas excluded	Frame non-coverage rate#	Districts excluded	Sample Non-coverage rate
28 (Oct 73-Jun 74)	1971	Ladakh	None	None	2.28	-	All	100	NA	NA
32 (Jul 77-Jun 78)	1971	Ladakh	None	None	2.28	-	All	100	NA	NA
38 (Jan 83-Dec 83)	1981	Ladakh, Kargil	None	None	2.24	-	All	100	NA	NA
43 (Jul 87-Jun 88)	1981	Ladakh, Kargil	None	None	2.24	-	All	100	None	00.00
44 (Jul 88-Jun 89)	1981	Ladakh, Kargil	None	None	2.24	1981	768 villages	68.52		
45 (Jul 89-Jun 90)	Economic Census 1980	Ladakh, Kargil	None	None	2.24	1981	768 villages	68.52		
46 (Jul 90-Jun 91)	Economic Census 1980	Ladakh, Kargil	All six districts	None	2.24+53.19 = 55.43	1981	768 villages	68.52		
47 (Jul 91-Dec 91)	1981	Ladakh, Kargil	All six districts	None	55.43	1981	768 villages	68.52		
48 (Jan 92-Dec 92)	1981	Ladakh, Kargil	All six districts	None	55.43	1981	768 villages	68.52		
49 (Jan 93-Jun 93)	1981	Ladakh, Kargil	All six districts	Doda	55.43+ 7.10 = 62.53	1981	768 villages	68.52		
50 (Jul 93-Jun 94)	1981	Ladakh, Kargil	All six districts	Doda (C, E) / Punch, Rajauri (E)	62.53 (C) / 62.53+11.41 = 73.94 (E)	1991	768 villages	62.27	NA	NA

Round / Period	Jammu and Kashmir					Nagaland				
	Census†	Districts excluded (by NSS Region)			Overall (frame + sample) non-coverage rate	Census†	Rural		Urban	
		Ladakh (Ladakh)	Kashmir (Jhelam Valley)	Jammu†† (Outer Hills)			Areas excluded	Frame non-coverage rate#	Districts excluded	Sample Non-coverage rate
51 (Jul 94-Jun 95)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27		
52 (Jul 95-Jun 96)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27		
53 (Jan-Dec 1997)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27		
54 (Jan 98-Jun 98)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27		
55 (Jul 99-Jun 00)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27	Mon, Wokha, Zunheboto	17.60$
56 (Jul 00-Jun 01)	1981	Ladakh, Kargil	All six districts	Doda	62.53	1991	768 villages	62.27		
57 (Jul 01-Jun 02)	1981	Kargil, Leh	All six districts	Doda	62.53	1991	768 villages§	62.27		
58 (Jul 02-Dec 02)	1981	Kargil, Leh	None	None	2.24	1991	768 villages	62.27		
59 (Jan 03-Dec 03)	1981	Leh (Ladakh), Kargil	None	None	2.24	1991	768 villages	62.27		
60 (Jan 04-Jun 04)	1981	Leh (Ladakh), Kargil	None	None	2.24	1991	768 villages	62.27		
61 (Jul 04-Jun 05)	2001	Leh (Ladakh), Kargil	None	Punch, Rajauri, Doda (Rural)	2.33 + 10.92 + 6.35 = 19.60	2001	946 villages§§	65.13	None	00.00

Round / Period	Jammu and Kashmir					Nagaland				
	Census†	Districts excluded (by NSS Region)			Overall (frame + sample) non-coverage rate	Census†	Rural		Urban	
		Ladakh (Ladakh)	Kashmir (Jhelam Valley)	Jammu†† (Outer Hills)			Areas excluded	Frame non-coverage rate#	Districts excluded	Sample Non-coverage rate
62 (Jul 05-Jun 06)	2001	Leh (Ladakh), Kargil	None	Punch, Rajauri	2.33 + 10.92 = 13.25	2001	946 villages	65.13		
63 (Jul 06-Jun 07)	2001	Leh (Ladakh), Kargil (central sample)	None	Punch, Rajouri	13.25	2001	946 villages	65.13		
64 (Jul 07-Jun 08)	2001	Leh (Ladakh), Kargil (central sample)	None	None	2.33	2001	946 villages	65.13		
65 (Jul 08-Jun 09)	2001	Leh (Ladakh), Kargil (central sample)	None	None	2.33	2001	946 villages	65.13		
66 (Jul 09-Jun 10)	2001	Kargil, Leh	None	Punch, Rajouri (Sub-rounds 1,2,4), Doda (Rural, Sub-round 3)	13.25	2001	946 villages	65.13	None	00.00
67 (Jul 10- Jun 11)	2001	Leh (Ladakh), Kargil (central sample)	None	None	2.33	2001	946 villages	65.13		
68 (Jul 11- Jun 12)	2001	Leh (Ladakh), Kargil	None	Punch	2.33 + 3.68 = 6.01	2001	946 villages	65.13	None	00.00
69 (Jul 12 - Dec 12)	2001*	None	None	None	00.00	2001*	None	00.00		

Round/ Period	Jammu and Kashmir					Nagaland				
	Districts excluded (by NSS Region)				Overall (frame + sample) non-coverage rate	Rural			Urban	
	Census†	Ladakh (Ladakh)	Kashmir (Jhelam Valley)	Jammu†† (Outer Hills)		Census†	Areas excluded	Frame non-coverage rate#	Districts excluded	Sample Non-coverage rate
70 (Jan 13-Dec 13)	2001*	None	None	None	00.00	2001*	None	00.00		
71 (Jan 14-Jun 14)	2011	None	None	None	00.00	2011	None	00.00		

Notes: (i) † indicates the census used as the sampling frame in the survey. * indicates the use of the 2001 Census updated for changes as the sampling frame. †† Jammu region also includes the NSS region called *Mountainous*, which was fully covered in all rounds and has not been listed here. §§ The figure is based on GoI (2011), which analyses the 61st to 67th rounds, and mentions total villages in the state to be 1317 and those in the frame to be 371. 'C' ('E') indicates the *Consumer Expenditure (Employment and Unemployment) Survey*. (ii) Ladakh, a part of *Jhelam Valley* in earlier rounds, became a separate NSS Region in the 66th round. (iii) Nagaland contains only one NSS region. (iv) Non-coverage rate has been calculated using data from 1981, 2001, and 2011 censuses in case Jammu and Kashmir; and from 1981, 1991, 2001, and 2011 Censuses in case Nagaland. (v) Ladakh suffers from frame non-coverage, whereas Kashmir and parts of Jammu suffer from sample non-coverage. Since both bias sample statistics, these have been summed up to get overall non-coverage. (vi) In some rounds the extent of coverage, as indicated in the NSS reports, varies across surveys conducted in the same year: The Table mentions the names of the districts which were not covered as per at least one of the reports. (vii) § The reports for Rounds 57-59 do not specify the number of villages and merely mention that "interior villages of Nagaland situated beyond 5 kilometres of the bus route" were excluded from the sampling frame. However, a report for the 60th round (No 506), which was the last round with the 1991 Census as sampling frame, mentions the number of excluded villages to be 768.

Notes for Nagaland (Rural): (a) See Appendix 1 for the calculation of non-coverage. (b) # Sample non-coverage has been omitted in case of rural Nagaland because the only one FSU each was not covered two rounds (see Table 2).

Notes for Nagaland (Urban): (a) We have relied on unit level data to retrieve information on districts excluded. Non-coverage is defined as the urban population share of districts excluded from the sample. (b) We could not get the unit level data for 28th and 32nd rounds. In the 28th round, 29 households from 24 blocks were sampled (GoI, 1977). In the 32nd round, 432 households from 36 blocks were sampled (GoI, 1986a). The estimates from the 28th round appear unreliable, but note that NSSO began its operation in Nagaland just around that period (Section 2.2). During the 1970s, only two districts (Kohima and Mokokchung) had urban population. (c) 38th round: we could not retrieve the district variable from the unit level data. (d) In the 50th round, the dataset does not contain any district variable. (e) $ As per NSS reports for the 55th round all urban FSUs were surveyed. However, unit level data indicate the exclusion of the districts mentioned in the table. (f) In the 66th round, there were no observations in the sample from the three new districts created between 2001 and 2004. In 2001, the constituent parts of these three districts did not have any urban area.

Sources: Compiled from relevant NSS and Census reports, *Sarvekshana* issues, and unit level NSSO data.

Table 2: Sample FSUs allotted/sampled in different rounds of NSSO Surveys

Round/Report	Nagaland	Tripura	Jammu and Kashmir	India	Un-surveyed	
					J&K/India	NE/India
38/332	0/0 32/32	208/85 48/28	328/287 152/145	8598/8039 4572/4379	41/559 7/193	274/559 86/193
43/Sarvekshana Apr-Jun 1992	0/0 32/30	208/153 48/36	328/322 152/150	8518/8375 4648/4599	6/143 2/49	96/143 16/49
47/393	32/32 16/16	108/103 28/28	164/80 80/27	4468/4373 2564/2503	84/114^ 53/61	17/114^ 3/61
48/407 (437, only surveyed figures) [419, only surveyed figures]	32/32 16/16	104/97 (96) 24/24	160/80 (77) 84/28	4328/4231 (4208) 2484/2420 [2411] [2419]	80/114^ 56/64	16/114^ 6/64
49/429	36/36 20/20	124/64 32/32	184/64 94/27	5112/4822 2928/2791	120/290 67/137	57/290 3/137
50/409 (401)	48/48 (46) 24/24	176/154 (153) 56/56	264/84 (82) 168/53	7248/6983 (6951) 4792/4670 (4650)	180/265 115/122	47/265 1/122
51/435	112/112 48/39	208/151 80/79	304/95 192/55	8536/8214 5536/5258	209/322 137/278	93/322 17/278
52/439	96/96 48/48	192/130 72/72	312/203 200/81	7888/7663 5112/4991	109/225 119/121	110/225 1/121

Round/Report	Nagaland	Tripura	Jammu and Kashmir	India	Un-surveyed	
					J&K/India	NE/India
53/444	32/32 40/39	32/32 24/24	48/36 64/50	6055/5988 7169/7138	12/67 6/31	10/67 10/31
54/452	56/56 12/11	108/76 20/20	192/116 60/27	5242/5115 1788/1745	76/127 33/43	43/127 10/43
55/454 (468)	40/40 24/24	136/86 48/40 (48)	208/131 128/84	6208/6046 (6048) 4176/4116 (4125)	77/162 44/60	62/162 10/60
56/476 (477)	56/55 (56) 56/52 (56)	152/117 136/136	180/127 (128) 364/214 (216)	5696/5433 (5586) 9092/8918 (8942)	53 (52)/263 (110) 150 (148)/174 (150)	55 (45)/263 (110) 7 (2)/174 (150)
57/481	68/68 51/51	120/120 128/128	180/140 246/204	6513/6381 9356/9297	40/132 42/59	6/132 2/59
58/484 [489] [487] {485}	44/44 20/19 (20)	104/104 [102] {103} 40/40	152/96 [95] 100/72 {66}	4786 {4828}/4675 (4769) [4646] {4637} 3552 {3628}/3498 (3538) {3354}	56/111 {191} 28 {34}/54 {274}	12 {78}/111 {191} 6 {129}/54 {274}
59/490 (491; 497 for rounds) [500] {504}	48/48 16/16	128/128 40/40	196/115 100/52	6784 (6760)/6553 (6638 rnd 1; 6634 rnd 2) [6552] {6551} 3824/3757 (3764) {3746}	81/231 48/67	46/231 8/67
60/505	24/24 12/12	84/82 24/24	136/83 68/38	4908/4755 2704 (2708)/2669 (2668)	53/153 30/35	12/153 0/35

Round/Report	Nagaland	Tripura	Jammu and Kashmir	India	Un-surveyed	
					J&K/India	NE/India
61/508	96/96 32/32	176/176 56/56	288/190 144/89	8128/7999 4660/4602	98/129 55/58	5/129 2/58
62/523 (522)	47/47 32/32	103/103 80/80	152/115 200/178 (180)	4847/4750 (4798) 5150/5120 (5125)	37/97 22/30	0/97 2/30
63/527	32/32 65/65	120/120 85/85	96/78 181/165	5601/5573 7721/7698	18/28 16/23	1/28 1/23
64/530 (533)	128/128 48/48	216/216 72/72	144/120 (118) 112/95	7984/7953 (7921) 4704/4682 (4668)	24/31 17/22	1/31 0/72
65/535 [534] [536]	80/79 32/32	216/216 72/72	176/127 88/64 [63]	8188/8130 [8109] 4764/4735 (4738) [4719]	49/58 24/29	5/58 1/29
66/537 (538)	88/88 40/40	164/164 68/68	276/181 (183) 180/160	7524/7402 (7428) 5284/5252 (5263)	95/122 20/32	10/122 2/32
67/546	104/104 32/32	136/136 64/64	168/135 112/111	8380/8296 7620/7602	33/84 1/19	42/84 8/19
68/554	84/84 44/44	164/164 68/68	260/254 172/171	7508/7469 5276/5268	6/39 1/8	7/39 0/8

Round/Report	Nagaland	Tripura	Jammu and Kashmir	India	Un-surveyed	
					J&K/India	NE/India
69/556 (561)	44/44 28/28	104/104 72(73)/72(73)	92/92 70(74)/70(74)	4500/4475 3524(3835)/3522(3832)	0/25 0/2	19/25 2/2
70 (Unit level data; report on KIs)	44/44 28/28	104/104 72/72	92/92 68/68	4534/4529 3508/3507	0/5 0/1	2/5 0/1
71 (Unit level data; report on KIs)	44/44 28/28	104/104 72/72	92/92 58/68	4580/4577 3720/3720	0/3 0/0	0/3 0/0

Notes: (i) The figures in the first row in each cell correspond to rural areas and those in the second, to urban. In each row the first entry is the number of FSUs allotted and the second, FSUs surveyed. (ii) Multiple thematic reports are issued after each round of survey. As far as possible, the table captures all variations among the reports. For example, 69/556 (561) corresponds to the 69th round and Report Nos. 556 and 561. The relevant entry under Tripura $\frac{104/104}{72(73)/72(73)}$ indicates that both reports agree that all 104 FSUs allotted were surveyed in the rural sector, whereas they differ with regard to urban sector. Report No. 556 (561) indicates that 72 (73) FSUs allotted were surveyed. (iii) ^ Excluding Arunachal Pradesh, where the number of surveyed villages exceeds the number of allotted villages due to a different survey design used in earlier rounds. (iv) In four rounds (51-54), some reports provide results for Jammu and Kashmir as part of the North West region, whereas results for the North Eastern states, except Assam, are clubbed together. (v) We could not retrieve information for the 28th and 32nd rounds.

Source: Compiled from relevant NSS reports and Notes on Sample Design & Estimation Procedure (Rounds 61-64).

Table 3: Some characteristics of NSSO central samples (2004-05 – 2011-12)

Description	Nagaland	Tripura	Assam	Arunachal Pradesh	Meghalaya	Manipur	Mizoram	Sikkim	North East	Jammu and Kashmir	India
Panel A											
Total number of villages (as per 2001 Census)#	1,317	870	26,312	4,065	6,034	2,391	817	452	42,258	6,652	6,38,596
Villages in NSSO frame	371	870	26,312	4,065	6,034	2,228	817	452	41,149	6,404*	6,36,127
Average number of villages allotted in quinquennial (annual) rounds**	89 (78)	168 (158)	332 (278)	145 (94)	111 (92)	188 (128)	80 (58)	81 (60)	1,195 (947)	275 (147)	7,720 (7,000)
Average number of villages surveyed in quinquennial (annual) rounds**	89 (78)	168 (158)	329 (278)	141 (86)	110 (92)	188 (128)	80 (58)	81 (60)	1,187 (938)	208 (107)	7,623 (6,267)
Panel B											
Proportion of villages in the frame	28.17	100	100	100	100	93.18†	100	100	97.38	96.27	99.61
Ratio of villages allotted to that in the frame in quinquennial (annual) rounds	24.08 (21.08)	19.31 (18.18)	1.26 (1.06)	3.58 (2.32)	1.83 (1.53)	8.44 (5.75)	9.79 (7.12)	17.99 (13.36)	2.9 (2.3)	4.13 (2.21)	1.21 (1.1)
Ratio of villages surveyed to those allotted in quinquennial (annual) rounds	100 (100)	100 (100)	99.2 (100)	97.25 (91.3)	99.7 (100)	100 (100)	99.58 (100)	100 (100)	99.39 (98.97)	75.85 (72.83)	98.75 (89.53)
Ratio of villages allotted in quinquennial to annual rounds	1.14	1.06	1.19	1.54	1.21	1.47	1.38	1.35	1.26	1.87	1.10
Ratio of villages surveyed in quinquennial to annual rounds	1.14	1.06	1.18	1.64	1.20	1.47	1.38	1.35	1.27	1.94	1.22

Notes: (i) # Source: GoI (2011); http://censusindia.gov.in/Census_Data_2001/Census_data_finder/A_Series/Number_of_Village.htm (for Jammu and Kashmir). The number of villages include uninhabited villages (GoI, 2011). (ii) * This figure is arrived at by subtracting the 2001 Census villages of Leh (viz., 112) and Kargil (viz., 127) adjusted for uninhabited villages. Since we do not have district level information on uninhabited villages, we have distributed the total number of uninhabited villages of the state (viz., 235) across districts in proportion to their population. (iii) ** These figures are the averages of the corresponding rounds during the period, 2004-05 to 2011-12. (iv) † *Note on Sample Design and Estimation Procedure* (63rd round: 27) indicates that three sub-divisions of Senapati district are not included in the sampling frame.

Sources: Panel A (GoI, 2011, NSS Report Nos 508, 523, 527, 530, 535, 537, 546, and 554) and Panel B (Authors' calculations)

Table 4: Monthly per Capita Consumption Expenditure (MPCE) and State Domestic Product (SDP) per capita (₹)

Panel A: MPCE

State	1973-74 (Oct-June) 28th round		1977-78 32nd round		1983 (Jan-Dec) 38th round		1987-88 43rd round		1993-1994 50th Round		1999-2000 55th Round		2004-05 61st Round		2009-10 66th Round		2011-12 68th Round	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Jammu and Kashmir	52.24	55.71	72.86	88.37	128	155	204	271	360	556	677	953	793	1070	1244	1667	1601	2320
Nagaland	NA	100.22	NA	137.30	NA	196	NA	367	460	521	941	1242	1011	1498	1379	1732	1757	2279
All India (average)	53.01	70.77	68.89	96.15	112	166	158	250	281	458	486	855	559	1052	953	1856	1287	2477
India (Maximum)	75.51	104.9	114.39	229.58	170	257	244	367	460	556	941	1243	1013	1498	1763	2315	2461	3346
J and K/Max (%)	69.18	53.11	63.69	38.49	75.29	60.31	83.61	73.84	78.26	100.00	71.94	76.67	78.28	71.43	70.56	72.01	65.05	69.34
Nagaland / Max (%)		95.54#		59.80		76.26		100.00	100.00	93.71	100.00	99.92	99.80	100.00	78.22	74.82	71.39	68.11

Panel B: SDP per capita*

State	1973-74 (Oct-June) 28th round	1977-78 32nd round	1983 (Jan-Dec) 38th round	1987-88 43rd round	1993-1994 50th Round	1999-2000 55th Round	2004-05 61st Round	2009-10 66th Round	2011-12 68th Round
Jammu and Kashmir	291.06	327.35	2428	2954	6543	12373	21734	33650	46734
Nagaland	317.08	354.84	2159	3445	9129	12594 (15004)**	30441	50263	63781
India (Maximum)	604.31	709.90	4181	5966	16558	44349	76968	149164	211570
J and K/Max (%)	48.16	46.11	58.07	49.51	39.52	27.90	28.24	22.56	22.09
Nagaland/Max (%)	52.47	49.98	51.64	57.74	55.13	28.40 (33.82)**	39.55	33.70	30.15

Notes: (i) While NSSO began operating in Nagaland in 1972, prior to the 50th Round (1993-94) NSSO samples did not cover the rural areas. (ii) Due to the changes in recall period in consumption questions, the estimates from 1999-2000 are not comparable with the rest of the period (Deaton and Dreze, 2002). (iii) The figures for 2009-10 and 2011-12 correspond to the 'Mixed Reference Period.' (iv) Estimates for the 50th round for Nagaland and Jammu and Kashmir have been computed using the distribution of MPCE by fractile groups. (v) India (Max) is the maximum of the average MPCE or SDP per capita excluding UT's and Delhi. Including Delhi does not affect our conclusions. Report No 502 (50th round) provides information on average MPCE of only 15 states, and India (Max) values for MPCE for this round are based on 17 states (15 states in the Report; Jammu and Kashmir and Nagaland computed by the authors). (vi) * SDP per capita figures for 1973-74 and 1977-78 correspond to 1960-61 prices; 1983 and 1987-88, to 1980-81 prices; 1993-94 and 1999-2000, to 1993-94 prices; and the rest, to 2004-05 prices. We have divided the SDP per capita figures by 12 to make it comparable with MPCE. (vii) Note that per capita figures – which use Census population in the denominator – are affected by the anomalies in Nagaland's census (Agrawal and Kumar, 2013). (viii) # This entry appears to be an outlier as this was one of the first NSSO surveys in Nagaland. It is likely that actual non-coverage in this round was much higher than our estimate based on NSSO's reports. Unit level data are not available for the 28th Round. Also, see note (b) to Table 1. (ix) **The figures in parentheses were obtained by correcting the population of Nagaland for over count detected in 2001 (see Kumar and Agrawal, 2016 for the correction procedure). The correction has not been applied to other intercensal years in the 1990s and 2000s as we do not have information to estimate correction factors, which will lie between 1 and 2, for those years. (x) Unless otherwise specified the survey year is the agricultural year (July-June).

Sources: Part A: GoI (1977) for 28th round; GoI (1986a) for 32nd round; NSS reports of the corresponding rounds; Part B: Ghosh and De (2005) and *Handbook of Statistics on Indian Economy* for the years 2012-13 and 2014-15, published by the Reserve Bank of India

Table 5: Non-coverage rate and MPCE - Regression Results

Dependent variable: MPCE (state)*100/ MPCE (India)	Model 1	Model 2	Model 3
Non-coverage rate (Table 1)	0.306**	0.308**	
Non-coverage rate (Table 2)			0.236*
State Dummy (Nagaland=1)	30.095***	30.662***	
Sector Dummy (Urban =1)	-22.1328***	-21.675***	-26.012***
SDP Growth (difference of state to median)		-0.728	
Intercept	116.426***	114.279***	123.013***
R-square	0.65	0.66	0.76
Observations	32	32	14

Note: (i) No of observations = 32 (18 for Jammu and Kashmir - 28th, 32nd, 38th, 43rd, 50th, 55th, 61st, 66th, and 68th rounds - and 14 for Nagaland, where only urban areas were covered in the 28th, 32nd, 38th, 43rd). (ii) 'SDP Growth' is constructed as follows. First, growth rate of per capita SDP is calculated for states for three years preceding each round. Second, the three year averages of the growth rates are computed for all states. In the third step, we find the median growth rate across states. Finally, we subtract the median obtained in step 3 from the actual growth rates of Nagaland and Jammu and Kashmir. The variable thus arrived at captures the relative position of the state vis-à-vis the state with median growth. (iv) Note that Model 3, where non-coverage is based on Table 2 that is restricted to sample non-coverage, could only be run for Jammu and Kashmir as sample non-coverage is close to zero for Nagaland across various NSS rounds. The data on frame non-coverage (Table 1) and sample non-coverage (Table 2) have been calculated using two different methods and cannot be combined directly. In Table 2, we could not use the data from 28th and 32nd rounds even for Jammu and Kashmir because of unavailability of sample non-coverage data. (v) ***, **, and * indicate statistically significant coefficient at 1, 5, and 10 per cent level, respectively.

Sources: Estimated based on information from Tables 1, 2 and 4. We assume zero non-coverage for 'NA' values for Nagaland (Urban) in Table 1.

Table 6: Poverty Line and Incidence of Poverty

State / India / Ratio / Rank	2009 Expert Group (Tendulkar)								2011 Expert Group (Rangarajan)			
	Rural				Urban				Rural		Urban	
	1993-94	2004-05	2009-10	2011-12	1993-94	2004-05	2009-10	2011-12	2009-10	2011-12	2009-10	2011-12
Panel A: Poverty Line (₹, Current prices)												
Jammu and Kashmir	289	522	723	891	281	603	845	988	848	1045	1200	1403
Nagaland	382	687	1017	1270	410	783	1148	1302	985	1230	1424	1616
India (Maximum)	382	687	1017	1270	410	783	1148	1302	1026	1231	1424	1704
All-India	NA**	447	673	816	NA**	579	860	1000	801	972	1198	1407
Jammu and Kashmir / Max (in %)	76	76	71	70	69	77	74	76	83	85	84	82
Nagaland /Max (in %)	100	100	100	100	100	100	100	100	96	100	100	95
Panel B: Poverty Incidence (%)												
Jammu and Kashmir	32.5	14.1	8.1	11.5	6.9	10.4	12.8	7.2	14.4	12.6	32.4	21.6
Nagaland	20.1	10	19.3	19.9	21.8	4.3	25	16.5	11.1	6.1	37.2	32.1
All-India	50.1	41.8	33.8	25.7	31.8	25.7	20.9	13.7	39.6	30.9	35.1	26.4
*Jammu and Kashmir's Rank**	5	3	1	7	2	4	7	8	5	7	13	12
*Nagaland's Rank**	2	1	10	14	7	1	21	19	3	2	20	21
No. of states	28	28	28	28	28	28	28	28	28	28	28	28

Notes: Maximum and minimum are based on 28 States (Telangana included within Andhra Pradesh); Union territories are not included in the above table. * Rank indicates the position of the State when the 28 states are arranged in ascending order of the incidence of poverty. ** For 1993-94 (All-India), GoI (2009) provides the estimates of the incidence of poverty and not the poverty line.

Sources: GoI (2009; 2014b)

Appendix 1

Calculation of non-coverage rate for Nagaland (for Table 1)

The NSSO does not provide information on the population of the excluded villages in Nagaland. We can use the share of number of excluded villages in the state as a crude estimate of the non-coverage, but we will have to assume that the population of villages is identical across the state. A relatively better estimate can be obtained using the information given in the District Census Handbooks on the distribution of villages by the availability of bus service. Note that we still have to assume that population of all the villages within the same district is identical. Had NSSO provided information about the spatial distribution of non-coverage we could have exploited the data on distance of habitations provided in the census to arrive at a more precise estimate of non-coverage.

Since the Census and the NSSO data on the distribution of villages by the availability of bus service do not match (see Section 2.2), we use the census data on population to estimate non-coverage, while maintaining that the NSSO figures for the total number of villages of Nagaland not included in the sampling frame are correct.

Let n_{ij} be the number of villages in the jth circle of the ith district, d_{ij}, the number of villages (out of n_{ij}) beyond 5 km of the bus routes, P_i, the population of the district, and r_i, the share of rural population of the district. The rural population of the ith district is given by $r_i P_i$. The rural population excluded (viz., population of distant villages) in ith district is given by:

$$R_i^e = r_i * P_i * \frac{\sum_j d_{ij}}{\sum_j n_{ij}} \qquad (1)$$

The sum of rural population excluded for all districts gives us the excluded rural population of the state. We can now find the average population per excluded village in the state as follows:

$$\overline{P^e} = \frac{\sum_i R_i^e}{\sum_{i,j} d_{ij}} \qquad (2)$$

The non-coverage (the share of rural population outside the coverage of survey) is given by the following expression, where d^{NSSO} is the number of villages excluded from NSSO sample in the state (note that $d^{NSSO} \neq \sum_{i,j} d_{i,j}$):

$$Non - coverage = \left(\frac{\overline{P^e} * d^{NSSO}}{\sum_i (r_i * P_i)} \right) * 100 \qquad (3)$$

The above procedure is repeated for three census years, 1981, 1991, and 2001. Note that the above procedure over-estimates non-coverage insofar as it has to assume uniform population density within districts, whereas in reality the excluded villages are relatively sparsely populated. However, this does not affect our key results in Table 5 because to the extent that non-coverage is overestimated we are underestimating the sensitivity of MPCE to non-coverage, i.e., the coefficients will be smaller than they should be.

About the Authors

Ankush Agrawal

Ankush Agrawal teaches economics at Indian Institute of Technology, Delhi.
Email: ankush@hss.iitd.ac.in

Vikas Kumar

Vikas Kumar teaches economics at Azim Premji University, Bengaluru.
Email: vikas.kumar@apu.edu.in

About Azim Premji University

Azim Premji University was established in Karnataka by the Azim Premji University Act 2010 as a not-for-profit University and is recognized by The University Grants Commission (UGC) under Section 22F. The University has a clearly stated social purpose. As an institution, it exists to make significant contributions through education towards the building of a just, equitable, humane and sustainable society. This is an explicit commitment to the idea that education contributes to social change. The beginnings of the University are in the learning and experience of a decade of work in school education by the Azim Premji Foundation. The University is a part of the Foundation and integral to its vision. The University currently offers Postgraduate Programmes in Education, Development and Public Policy and Governance, Undergraduate Programmes in Sciences, Social Sciences and Humanities, and a range of Continuing Education Programmes.