R E S E A R C H A R T I C L E

Discrepancies in Data on Landholdings in Rural India: Aggregate and Distributional Implications

Deepak Kumar*

Abstract: This article examines discrepancies in major sources of data on landholdings in rural India. The Agricultural Census and the Land and Livestock Survey conducted by the National Sample Survey Organisation are two of the most important sources of such data. Since these surveys inform both government policy and academic investigation of many aspects of rural India, it is imperative to examine the sources of these discrepancies. A preliminary view of their aggregate estimates shows significant divergence. While part of the divergence may be explainable on account of differences in definition and non-sampling errors (particularly non-reporting of landholdings), an investigation of their methodology brings out other likely reasons – particularly reasons associated with survey design – for the divergence.

Keywords and phrases: National Sample Survey, Land and Livestock Survey, Agricultural Census, landholding data, inequality, underestimation, land use, size-class distribution.

1. INTRODUCTION

This paper will analyse different sources of data on landholdings in rural India, specifying discrepancies between them and possible reasons for such discrepancies. The paper focuses on the design and methodology of the surveys, particularly as they relate to estimates of the distribution of landholdings and aggregate land use.

An understanding of forms, features, and patterns of distribution of landed property is essential to understand difficulties of sustaining growth in the agricultural sector and in improving the socio-economic conditions of all dependent on agriculture for their livelihoods. Most of academic enquiry and policy prescriptions on the land question in rural India are based on official sources of data on landholdings, primarily the Agricultural Census and the Land and Livestock Survey of the National Sample Survey Organisation (NSSO). The aim of this paper is to critically

* Ph.D. Scholar, Yokohama National University, deepak-k@outlook.com

analyse these sources in light of their stated objectives. I will evaluate the concepts, definitions, and methodology used for the most recent rounds of these surveys, with only passing reference to previous rounds if and when needed. In addition, I will use village-level databases on landholdings from ten villages located in distinct agro-ecological regions in the country to illustrate possible shortcomings in methodology. These data are from the archive of the Project on Agrarian Relations (PARI) of the Foundation for Agrarian Studies (FAS), a significant archive of village data in India today.

In section 2, I discuss the major official sources of data on landholdings in India, focussing on the Agricultural Census and the Land and Livestock Survey (henceforth LHS, for Land Holding Survey) of the National Sample Survey Organisation (NSSO). I critically examine the concepts and definitions employed in these surveys. Section 3 highlights an anomaly in estimation of aggregate land use by the LHS. Section 4 presents an overview of the sampling methodology used by the NSSO for the LHS. Section 5 uses data from census-type surveys of selected villages in order to test the sampling methodology of the LHS for potential shortcomings in generating representative statistics.

2. OFFICIAL SOURCES OF DATA ON LANDHOLDINGS IN INDIA

The two most important official sources of data on landholdings in India are the Agricultural Census and the Land and Livestock Survey (LHS) of the National Sample Survey Organisation (NSSO).¹

Agricultural Census

The Agricultural Census is conducted quinquennially by the Department of Agriculture and Cooperation, Ministry of Agriculture. It is a part of the World Agricultural Census programme of the Food and Agriculture Organisation (FAO). In its present form, the Agricultural Census was first conducted in the year 1970–1; subsequently, a total of nine rounds have been conducted. The most recent round for which data are available was conducted in 2010–1. The stated objective of the survey is "to describe the structure and characteristics of agriculture by providing statistical data on operational holdings."²

The precursor to the present day Agricultural Census were sample surveys conducted by the Directorate of National Sample Survey in 1950 and 1960. Following that, the

¹ In addition to these two sources, data on landholdings are available from the Consumer Expenditure Survey, Employment and Unemployment Survey, Situation Assessment Survey (SAS), and the All India Debt and Investment Survey (AIDIS), all conducted by the National Sample Survey Organisation. However, the stated objectives of these surveys do *not* include analysis of land distribution and are therefore not designed to provide representative estimates thereof. I will, therefore, restrict my enquiry to two of the most important sources. For a critique of the Employment and Unemployment Survey, refer to Rawal (2014). The same critique is also applicable to the CES.

² GoI (2010).

National Commission on Agriculture deemed land records – that is, village forms and registers – to be the most appropriate basis for collecting data on agriculture and, from 1970, began a census-type survey of operational holdings. Such an approach, however, required the availability and maintenance of elaborate land records pertaining to every landholding in the country, a reliance that was problematic both because of variations between the quality of record-keeping in different parts of the country and because of deterioration in the quality of record-keeping over time.

Typically, the erstwhile *ryotwari* regions had a better system of land records prior to independence from British rule than other regions, although that quality may not have been always sustained thereafter. Land revenue formed an important source of income for the British Empire and its collection necessitated the maintenance of land records in regions where revenue was supposed to be collected directly from the holder of the title deed (*ryot*). Outside these regions, however, there were other forms of tenure in which the State did not deal directly with the ryots but through intermediaries. These regions typically did not have as well-developed a system of land records as ryotwari regions. Accordingly, different methodologies evolved for conducting the Agricultural Census in "land-record" and "non land-record" States.³ According to the last survey conducted in 2010–1, 91 per cent of all operated area in the country is in regions that have detailed land records, while 9 per cent are in non land-record regions.⁴

The Agricultural Census is conducted in three phases. In the first phase, all operational holdings in villages located in land-record States are enumerated by survey number as per the village land records. In the non land-record States, two-stage stratified sampling is undertaken with villages as the first stage unit. Once the village has been selected, a list of all holdings is made to generate the sampling frame for operational holdings. This list (Schedules L1 and L2) is used to estimate the number and area of operational holdings. The exact proportion of villages drawn differs for each State, and has been specified in the published instruction manuals. In phase II, villages and operational holdings are sampled from land-record States, and, together with the sampled holdings from non land-record states, are surveyed for detailed data on tenancy status, terms of lease, land use, irrigation status, and cropping pattern. For phase III, which is also known as the input survey, a further sample is drawn from all villages selected for phase II of the survey. A total of 4 operational holdings in each size class are surveyed for information on the number of parcels, multiple cropping, agricultural equipment, etc. While the Agricultural Census is the most comprehensive survey of landholdings in terms of coverage, it falls short of its

³ For a note on differences in village-level records in erstwhile ryotwari and zamindari regions, see Bakshi and Ramachandran (2008).

⁴ The non land-record States include States in the North-East Region, West Bengal, Kerala, and Odisha. Punjab, while being a land-record State, employs the methodology of non land-record States.

stated objective of "describing the structure and characteristics of agriculture"⁵ for the following reasons:

- 1. The Agricultural Census is a survey of operational holdings of land put to crop production. Consequently, it does not provide a measure of landlessness in operational holdings as it does not cover rural households that do not cultivate any land. This category of households would include all such households that rely primarily on sale of labour power for their livelihoods and those that engage in allied activities such as fisheries, poultry, livestock raising etc., without supplementing these activities with crop production.⁶ It should be noted that many countries that are part of the FAO World Agricultural Census programme do indeed collect data on households without agricultural land.⁷
- 2. The Agricultural Census does not provide any data on ownership holdings. While it is clearly stated that the objective of the Agricultural Census is the analysis of operational holdings, it can be argued that data on ownership holdings is necessary to "provide benchmark data needed for formulating new agricultural development programmes and for evaluating their progress," and "to lay a basis for developing an integrated programme for current agricultural statistics," both of which are the other stated objectives of the Agricultural Census.
- 3. Land records in India are notorious for *benami* holdings (i.e., "land being held by persons in other, often fictitious names.")⁸ These registrations are done in order to avoid household landholdings from coming under the purview of land reform legislation. Consequently, a single holding may often be recorded as distinct units under different land holders. This may (a) underestimate the real inequality in access to land in rural India, and (b) record the larger holdings as multiple holdings in smaller size-categories.
- 4. The Agricultural Census undercounts tenancy severely. This is so because the most important form of tenancy contracts in rural India today are unregistered oral contracts, which are not recorded by lessors in village records in order to keep their leases outside the ambit of tenancy legislation.⁹
- 5. The instruction manuals state that efforts are to be made to record the *de-facto* position of household operational holdings rather than *de-jure* position. This is necessary to convert individual-based land records to household holdings, to record informal tenancy contracts, correct categorisation of land (dry or

⁵ GoI (2010).

⁶ The Situation Assessment Survey of Farmers (2003) and Situation Assessment Survey of Agricultural Households (2013) attest to the importance of wage employment and allied activities such as raising livestock for the surveyed households. Also see, Chandrasekhar (2016).

⁷ FAO (2010).

 $^{^{\}rm 8}$ Definition taken from the glossary in Basu (2015).

⁹ This stylized fact of Indian agriculture has also been recently noted in the Report of the Expert Committee on Land Leasing, commissioned by the NITI Aayog, Government of India (2016), but with the regrettable suggested remedy of dismantling the old tenancy laws that recognised the right of the tiller to the land.

wetland), etc. However, there is no laid down procedure of how this is to be done except relying on the knowledge of the primary worker.

Land and Livestock Survey of the National Sample Survey Organisation

The Land and Livestock Survey of the National Sample Survey Organisation (LHS) is the most important survey on landholdings in India and is conducted decennially by the NSSO. The most recent survey was conducted as part of the 70th round of the National Sample Survey (NSS) in 2013.¹⁰ The stated objectives of the survey are to "obtain reliable estimates of key characteristics of land and livestock holdings across regions and landholding classes" (GoI 2014). The LHS provides the richest data of any official source on the land economy in the country. It offers us a range of data on the characteristics of household ownership and operational holdings. Nevertheless, there are some aspects of the land economy that are excluded from these surveys, thus limiting their usefulness. These are:

- 1. *Data on sale and purchase of land.* The land market defined in terms of volume of transactions of land is an important aspect of an agricultural economy. While the extent of land owned by a household at the time of survey is implicitly also a result of land transactions in the past, it does not allow us to separate inherited land from land bought or sold, and to that extent limits our understanding of the level of development and the characteristics of the land market.
- 2. Detailed data on irrigation. While the use of irrigation on a particular plot of land for a given season is recorded, the data do not indicate the quality of irrigation. There are, of course, limitations to the detail that a survey like LHS can be expected to record. Nevertheless, the quality of irrigation defined narrowly in terms of availability of the source of irrigation used throughout an agricultural season is an important indicator that, if included, would enrich the data on irrigation substantially. In addition, there are no data collected on the ownership of irrigation sources. The 70th round LHS informs us that the most important source of irrigation in India is groundwater irrigation. This form of irrigation primarily relies on private investment and ownership, and consequently, large private water-markets are possible. This important aspect of the practice of agriculture in India cannot be determined unless NSS records data on the ownership of irrigation sources.
- 3. *Detailed data on terms of lease.* While the LHS provides substantial data on land leases, such as the type of lease, duration of possession, number of lessor/lessee households, type of lessor/lessee, and the type of rent, it does not record any data on the quantum of rent paid by the tenant. This omission has become more important in recent years, since tenancy is reported to have increased from 6 per cent of total operated area to 10 per cent in the past ten years.¹¹ The terms

¹⁰ On the previous surveys, see Ramachandran (1980) and Rawal (2008).

¹¹ In all likelihood, NSS figures are likely to underestimate the actual level of tenancy in the countryside.

of lease differ quite significantly among rural households,¹² and the absence of these data from the LHS prevents us from examining levels of rent and variations in rent across size-classes and social groups.

Apart from unavailability of certain data, there are some shortcomings with respect to the definitions employed by the NSSO for the LHS.

1. The LHS defines a household operational holding as "land that was used wholly or partly for agricultural production and was operated (directed/managed) by one household member alone or with assistance of others, without regard to title, size or location." Agricultural production includes allied activities such as kitchen garden, floriculture, pisciculture, beekeeping, raising livestock, poultry farming etc. If the household undertakes any of these activities, it must have an operational holding, the extent of which is not only the area used for these activities, but *all land* (including house-site, paths, buildings etc. – that is, non-agricultural land) possessed by the household for most of the agricultural year.¹³

This is done in order to account for livestock holdings, which are an important source of livelihood for households in arid and semi-arid regions. However, by clubbing together holdings used for cultivation and those used for allied activities (that may be, and often are, undertaken on non-agricultural land), both the *number* of operational holdings and the *area* of operational holdings extend beyond crop production.

The implication of this definition is that it will underestimate landlessness with respect to crop production, as small holdings (as long as they are larger than 0.002 ha) possessed by the household are included in the extent of household operational holdings, even if they only possess livestock or poultry and do not undertake crop production. Secondly, it underestimates the average size of operational holdings (which are usually assumed to be for crop production) as the smaller holdings used for allied activities are aggregated along with holdings employed for crop production.

A partial correction for this was made in the 70th round of the LHS. While for the 59th round, the "area operated" was the area of the household operational holding, which included all non-agricultural land put to allied activities, in the 70th round, "area operated" was distinguished from the area of the operational holding. In the 70th round, area operated refers to the area of plots used for

¹² For an illustration, see Ramachandran, Rawal, and Swaminathan (2010).

¹³ "...once a household is identified to operate some land, all the plots possessed by the household during the major part of the reference period is taken into account in determining the size or area of the operational holding, irrespective of whether all the plots included in the holding are put to agricultural production or not. In case a household is found to possess more than one holding, plots possessed by the household during the major part of the reference period and put to uses other than agricultural production, such as house-sites, paths, buildings, etc., are also included in the operated area and all such plots are considered as part of operational holding number '1'"GoI (2006b), p. 6. The same definition and terms are employed for the 70th round as well (see GoI 2014, p. B-1).

agricultural production, i.e., it excludes the area of plots that were not put to any agricultural or allied production. However, given that the definition of agricultural production includes farming of animals/fishery, other agricultural uses, etc., plots such as homestead, which are non-agricultural in their primary land use, may still be counted in operational holdings, as long as they were put partially to agricultural or allied production.

2. In terms of ownership holdings, "a plot of land was considered to be owned by the household if permanent heritable possession, with or without the right to transfer the title, was vested in a member or members of the household. Land held in owner-like possession under long-term lease or assignment was also considered as land owned."

As per this definition, even households that do not have full legal ownership of their land are considered to be households with ownership holdings. This makes it impossible to study the proportion of households that have possession of a land holding but not full ownership over it.¹⁴

In addition, the measure of landlessness in the reports follows this definition of ownership holding. Even though the definition is explicitly stated as encompassing *all land*, it may be misleading to an observer unacquainted with the NSSO methodology. The measure of landlessness with respect to ownership of arable land should also be included in the published reports with due mention of its definition.

3. Estimation of Aggregate Area in the Land Holding Survey

A serious shortcoming of the LHS is with respect to the estimation of aggregate operated area in the country. *This survey has consistently estimated the aggregate extent of operational holdings to be less than the corresponding figure in the Agricultural Census and Land Use Statistics.* Table 1 compares the estimate of total operated area in the country from the LHS and the Agricultural Census.

Source	Agricultural	Agricultural	Agricultural
	Census (1990–1)	Census (2000–1)	Census (2010–1)
	LHS (1992)	LHS (2003)	LHS (2013)
Agricultural Census (000 Ha)	163922	157629	158050
LHS (000 Ha)	125100	107650	98614
Difference (000 Ha)	38822	49979	59436
% Difference	24	32	38

Table 1 Estimate of aggregate area of operational holdings in India according to theAgricultural Census and the Land and Livestock Survey of the NSS, 1992–2013 in thousand ha

Source: Agricultural Census and NSSO LHS Reports

¹⁴ Households that possess institutional land over which they have no ownership rights are recorded as having land "otherwise possessed." This category does not include private land that is encroached upon, which is recorded as leased land (GoI 2006a, p. 4). The two data sources show considerable discrepancy in the total area under operational holdings in the country. This discrepancy has increased from 24 per cent in 1991–92 to 38 per cent in 2011–13. It should be noted that the definition of operational holding for the Agricultural Census refers only to agricultural land, whereas for the LHS, the extent of operational holding also includes allied activities. However, this should only be reason for the aggregate extent of land recorded by the NSS to be greater, not less, than the aggregate extent recorded by the Agricultural Census.

In the preceding section, I had stated the shortcomings of the Agricultural Census in accounting for the pattern of *distribution* of landholdings in the country. However, being a *census* of landholdings (at least for 91 per cent of the country), it is likely to provide a reasonably useful indication of *aggregate area*. A factor that may contribute to an overestimation of total extent under operational holdings in the Agricultural Census is the lack of updating of records when land use changes from agricultural to non-agricultural purposes. However, this explanation may not explain the scale of disparity between the LHS and the Agricultural Census, which differed by 38 per cent in 2011–12.

A similar discrepancy is found when results from LHS are compared to the *Land Use Statistics* published by the Directorate of Economics and Statistics of the Ministry of Agriculture.¹⁵ Table 2 shows a comparison of aggregate extent of agricultural land from the LUS and the aggregate area of operational holding from the LHS.¹⁶ The former are consistently and increasingly higher than the latter over the previous three surveys. This discrepancy, which was 32 per cent of total extent in 1992, rose to 46 per cent in 2012.

Source	1992-3	2002-3	2012-13
LUS: Total agricultural land (in 000 ha)	184875	183450	181983
LHS: Aggregate operational holdings (in 000 ha)	125100	107650	98614
Difference (in 000 ha)	59775	75800	83369
% Difference	32	41	46

Table 2 Estimate of aggregate area as per LUS and LHS in thousand ha

Notes: LUS=Land Use Statistics.

LHS=Land and Livestock Holdings Survey, National Sample Survey Organisation.

Sources: DES (2014), GoI (2014).

¹⁵ The *Land Use Statistics* are "latest figures of geographical area of the State/Union Territories as provided by the Office of the Surveyor General of India." These are aggregated from land records. Where no land records exist, they are based on estimates of classification from the Agricultural Census. In some cases, where there are no land records, for instance, forest areas, "the magnitude of such area is known" and is included, without clearly indicating how it is known (Directorate of Economics and Statistics (2010–11), "Land Use Statistics: Concepts and Definitions," Department of Agriculture, Cooperation, and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India, New Delhi, available at http://eands.dacnet.nic.in/LUS-2010-11/ Concept.pdf, viewed on June 18, 2016.).

¹⁶ The aggregate extent of agricultural/cultivable land in the *Land Use Statistics* includes "net area sown, current fallows, fallow lands other than current fallows, culturable waste land, and land under miscellaneous tree crops." (*ibid.*) In other words, it includes all land that can be put to cultivation. The NSS definition includes these categories of land use plus land put to non-cultivation agricultural use, usually referred to as allied activities.

With reference to this divergence, the National Sample Survey states in its report (Report 571, pp. 4–5) that any comparison between these two sources needs to account for: (1) exclusion of land held by urban residents as LHS only covers households resident in rural areas, while LUS includes the whole geographical region covered by the Office of the Surveyor General of India; (2) exclusion of institutional land from the NSS on account of it being a household survey; (3) the distinction between total/gross cropped area or total area sown in the LUS and operational holding in the NSS; and (4) the distinction between total/gross irrigated area in the LUS and irrigation on operational holding of the NSS (GoI 2015).

As NSS notes in its report, LHS data are collected through a sample survey of rural households whereas LUS data are based on the latest figures of geographical area provided by the Surveyor General of India. There are two possible explanations for the disparity between LHS and LUS based on the source of data. First, the LUS estimate of total extent of agricultural area will be more than the aggregate extent of operational holdings from LHS if there is substantial cultivation in urban areas, which is completely excluded by the LHS.¹⁷ There is no evidence of practice of urban agriculture in India on such a significant scale. Secondly, the LHS estimated extent of aggregate operational holding will be lesser than agricultural area of LUS if there is substantial land held by urban households in rural areas that will, then, be excluded from the LHS sampling frame as it only covers households resident in rural areas. If we assume substantial ownership of agricultural land in rural areas by urban households, there are two possibilities. Either the urban household will keep the land uncultivated or lease it out to rural households. In the first case, there should be significant tracts left fallow. However, the extent of fallow land other than current fallow land in 2012 was, according to Land Use Statistics, 5.86 per cent of total agricultural land. In the second case, there should be significant *unexplained* tenancy. According to the LHS from the 70th round of the NSS, an estimate of the total extent of such land under tenancy that is not included can be generated from the survey itself, by taking the extent of land leased in minus the extent leased out, which is 6,00,000 ha. This is less than 1 per cent of the estimated aggregates from both the NSS and LUS. In addition, this would also include non-reporting of leases from rural landowning households, which is thought to be significant by itself. Therefore, while the ownership of rural land by urban households may well be a real phenomenon, circumstantial evidence does not support the explanation that it accounts for a large part of the disparities in estimation.

The total extent of institutional land in the country as per the Agricultural Census of 2010–11 was 1,541,914 ha, or 1.1 per cent of the total LUS estimate. The extent of institutional holdings from the Agricultural Census may be more reliable than data from the same source on the distribution of household operational holdings. This is possible as these institutional holdings are wholly legal and there is no reason for land records to be

¹⁷ The sampling frame for first stage unit selection in the LHS, in fact, includes all first stage units designated census villages, even when located in urban centres such as the Delhi National Capital Region (NCR).

incorrect in this regard. In fact, it is their registration as institutional land that excludes them from the purview of land reform legislation such as land ceilings. Further, only a part of these institutional holdings are engaged directly in agricultural production. It seems unlikely, then, that the inclusion of institutional land in the LUS contributes significantly to the divergence in estimates of extent between the LHS and LUS.

It should be noted that point 3 and 4 are irrelevant to our comparison, or for that matter any other researcher's, as the report misleadingly compares two different categories of data and judges them incomparable. The gross/total cropped area is the extent of *crops* in terms of land under such cultivation and operational holding is the extent of *land* as such, as the NSS report clarifies. LUS, however, also releases figures on the *extent of agricultural land*, which is directly comparable to the operational holding of LHS.¹⁸ It is, then, not immediately clear why the NSS report refers to the non-comparability of the extent of operational holdings from the LHS to gross cropped area from the LUS.

The concepts and definitions employed by the LHS, though having significant shortcomings on their own, do not convincingly explain the underestimation of aggregate extent when compared to other sources of data on land use. We now consider the sampling process employed in the LHS for a possible explanation.

4. SAMPLING DESIGN OF THE NSS LAND HOLDING SURVEY

In considering the sampling design of LHS, we shall focus on the methodology adopted for the 70th round of the NSS. This survey is based on a two-stage stratified sampling design. At the first stage of sampling, the First Stage Unit (FSU) is the village for rural areas and blocks for urban areas. The number of FSUs is decided in advance based on the availability of resources with the NSSO. FSUs are selected in multiples of two for each district, with a minimum of two villages from each district.

Once an FSU is selected, Schedule 0.0 is used for house listing. This provides the sampling frame for Schedule 18.1 (Land and Livestock), Schedule 18.2 (Debt and Investment), and Schedule 33 (Situation Assessment Survey). For sampling for the LHS, questions are asked on the *land possessed* by the household, that is, the total land owned plus the extent of land leased in minus the extent of land leased out. All households in the village are then classified into four Second Stage Strata (SSS) based on the extent of land possessed by the households are surveyed from each of these SSS, adding up to 8 households surveyed per FSU. The demarcation of distinct second stage strata from which further sampling of households is undertaken has changed over the previous three rounds of the survey. They are stated in Table 3.

For the 48th round, all households in a village that did not possess land or land less than 0.005 ha were classified in SSS 1. For the rest, all households in the village were arranged

¹⁸ In fact, the LUS definition takes into account only cultivable area while the LHS definition also includes land that is put to agricultural activity other than cultivation.

	48th round	59th round	70th round
SSS 1	Landless	Landless	Landless
SSS 2	А	0.005 to X	0.005 to 1 ha
SSS 3	В	X to Y	1 to 2 ha
SSS 4	С	> Y	> 2 ha

Table 3 Classification for sub-stratification for LHS for the 48th, 59th, and 70th rounds of theNational Sample Survey

Notes:

A: All households in the village are arranged in ascending order of extent of land possessed. "A" includes the first set of households that collectively possess a third of all land in the village.

B: All households that collectively possess the next third of all land in the village

C: The top households that collectively possess the next third of all land in the village

X: Extent such that 40 per cent of all households in the State possessed less than X in 1991

Y: Extent such that 80 per cent of all households in the State possessed less than Y in 1991

Source: NSSO reports for the 48th, 59th, and 70th rounds of LHS

in ascending order of possession of land and divided in three groups (A, B, and C) such that each of the groups roughly possessed an equal proportion of all land. These were SSS 2, SSS 3 and SSS 4 respectively. For the 59th round, SSS 1 remained unchanged. The other substrata were based on two cutoffs (X and Y) determined from the State level results of the 48th round survey. These cutoffs were such that 40 per cent of landpossessing households in the State possessed land less than extent X, and 80 per cent of all land-possessing households in the State had landholdings whose extent was less than Y. For the 70th round, these village and State-relative measures were done away with and a uniform scheme was used for all regions. Under this classification, all households in the village were classified as landless (SSS 1), possessing 0.005 to 1 ha (SSS 2), 1 to 2 ha (SSS 3), and more than 2 ha (SSS 4).

This uniform cut-off for the sub-stratification may not be the best practice, as patterns of distribution of landholdings across size-classes differ significantly across regions and States on account of various factors, including agro-ecological conditions, the history of land tenure, and land reform legislation. For example, according to the 1991 LHS, 40 per cent of all landowning households in Rajasthan held land whose extent was below 1.085 ha and, and 80 per cent of all land-possessing households held landholdings whose extent was less than about 4.125 ha each. The corresponding figures for West Bengal were 0.138 ha and 0.850 ha respectively.¹⁹ Uniform stratification could thus defeat the very purpose of the stratification, which is to define strata that straddle the distribution in order to reduce the variance within the strata and increase the variance between them, so as to form more accurate characterisations of and comparisons between the different constituent strata of the distribution of land.

Once all households in the village have been classified according to the criteria described above, the Second-Stage Units (SSU), i.e. households, are sampled from within these strata. The method used is Simple Random Sampling Without Replacement

19 GoI (2006a), p. B-4.

(SRSWOR). Once households are surveyed, aggregates based on multipliers derived from the proportion of surveyed population to the total are computed.

Once the aggregates are computed, estimates of the distribution of households by size-class are tabulated. The tabulation follows a six-fold size classification: landless, less than 1 ha, 1 to 2 ha, 2 to 4 ha, 4 to 10 ha, and more than 10 ha. So, in the 70th round, SSS 4, which included all households that had extent greater than 2 ha, is now further subdivided into 3 size-classes: 2 to 4 ha, 4 to 10 ha, and more than 10 ha.

5. VILLAGE-LEVEL CALCULATION

I have undertaken an illustrative exercise to examine the sampling methodology employed by the LHS (NSS 70th round). I use the LHS sampling design for selection of households (Second-Stage Unit in the LHS) on a database of rural households drawn from census surveys of villages by PARI. Based on this exercise, I examine the chance of picking households representative of different substrata in the FSU. I used data from census-type surveys of 10 villages located in distinct agro-ecological regions.²⁰ They were surveyed between 2007 and 2011. The study villages have been listed in Table 4.

These villages differ significantly in terms of their characteristics. Table 5 provides some basic parameters of distribution of landholdings (land possessed) in the villages. We see that there is considerable variation in terms of the size of holdings and in the pattern of distribution of land. The household population varied from 79 in Siresandra in Kolar, Karnataka, to 648 in Hakamwala in Punjab. The average holding ranged from 0.27 hectares per household in Amarsinghi to 5.31 hectares per household in 25F Gulabewala.

It should be noted that these villages were not sampled based on the criteria employed for the selection of FSUs by the NSSO, and, as such, may not be representative.

Village	District	State	Year of survey
25 F Gulabewala	Sri Ganganagar	Rajasthan	2007
Rewasi	Sikar	Rajasthan	2010
Alabujanahalli	Mandya	Karnataka	2009
Siresandra	Kolar	Karnataka	2009
Zhapur	Gulbarga	Karnataka	2009
Amarsinghi	Malda	West Bengal	2010
Kalmandasguri	Koch Bihar	West Bengal	2010
Panahar	Bankura	West Bengal	2010
Tehang	Jalandhar	Punjab	2011
Hakamwala	Budhlada	Punjab	2011

 Table 4 List of study villages with district, State and year of survey

²⁰ For further details, see www.fas.org.in.

Village	Number of holdings	Mean extent of holdings (ha)	Coefficient of variation	Skewness of distribution
25F Gulabewala	204	5.31	241.2	5.24
Alabujanahalli	243	1.09	119.1	2.89
Amarsinghi	127	0.27	113.4	1.66
Hakamwala	497	2.04	241.0	5.84
Kalmandasguri	147	0.40	114.7	2.14
Panahar	248	0.34	204.2	5.39
Rewasi	219	2.43	91.3	1.73
Siresandra	79	1.89	142.3	4.56
Tehang	648	1.00	356.6	6.49
Zhapur	109	2.16	211.0	3.38

 Table 5 Basic information on land distribution in the study villages

Source: Survey data

However, this is somewhat irrelevant to our analysis as our focus is on the effects of picking households within a village and not on picking the village itself. For this purpose, I picked villages that covered a range of characteristics and for which reliable survey data were available. This exercise, then, is *illustrative* of a phenomenon rather than *representative* of it.

Stratified sampling is employed by the NSSO in order to decrease intra-strata variance in the extent of possession of land in order to obtain better estimates. If we stratify households in our study villages based on the 70th round SSS classification, we obtain the following distribution.

Table 6 shows that the intra-strata variance is fairly low for SSS 1, SSS 2, and SSS 3 for all our study villages. However, SSS 4 is an exception. The variance within SSS 4 is

State	Village	SS	SS1	SS	SS2	S	SS3	S	SS4
		Ν	Var	N	Var	N	Var	N	Var
Karnataka	Alabujanahalli	16	0.00	128	0.08	57	0.06	42	2.60
	Siresandra	3	0.00	30	0.08	21	0.05	25	14.70
	Zhapur	21	0.00	51	0.07	9	0.06	28	38.28
Punjab	Hakamwala	211	0.00	104	0.06	54	0.05	128	58.79
	Tehang	34	0.00	515	0.01	22	0.09	77	55.20
Rajasthan	25F Gulabewala	29	0.00	112	0.02	1	0.00	62	328.13
	Rewasi	3	0.00	58	0.09	60	0.08	98	4.64
West Bengal	Amarsinghi	4	0.00	118	0.05	5	0.02	NA	NA
	Kalmandasguri	7	0.00	126	0.06	12	0.06	2	0.06
	Panahar	27	0.00	199	0.06	17	0.06	5	3.34

 Table 6
 Within-strata variance in landholdings in the study villages

Notes: N=number of holdings in each class; Var=variance.

Source: Survey data



9.75

2.6

39

Figure 1 Distribution of landholdings in SSS 4 (all villages) Note: Excludes Amarsinghi in West Bengal, as there are no households in SSS 4.

higher than for other substrata in each village, and is as high as 328 for 25F Gulabewala, 58 for Hakamwala, 55 for Tehang, and 38 for Zhapur. This is so because SSS 4 is an open-ended stratum that includes all holdings larger than 2 hectares. We recall that, in contrast to the method used in the 48th and 59th Rounds, when stratification was done on the basis of a measure based on the regional distributions of landholdings, the NSS 70th Round employed a common stratification design for all regions, whether characterised by small holdings or large. In addition to variance, the shape of the distribution is also of interest to us. The figures below show us how landholdings in SSS 4 were distributed in our study villages.

The histograms for the distribution of landholdings in SSS 4 of our study villages (Figure 1) affirm what would be intuitive – the number of landholdings decrease with increases in extent within the strata. In other words, the distribution of landholdings in the topmost stratum is positively skewed. This skewness in the distribution, along with the high variance noted before, has significant implications for the sampling process.

The LHS employs the method of simple random sampling (SRS) from within each stratum. By this method, each landholding has an equal chance of being picked. For a sample to be representative of the landholdings of the whole stratum, households that possess land equivalent to the mean extent in each strata, on average, ought to be picked. This is so because the mean holdings of sampled households are multiplied by the proportion of the sample size to the population in the strata to estimate features of the land ownership of the stratum as a whole.

However, due to the skewness of the distributions, if we randomly pick one household from SSS 4, there is a high probability, for any given draw, that this household will possess land that is smaller than the average extent (Table 7). This is so because the number of households whose extent is smaller than the mean is greater than the

Village	Less than mean	Greater/equal to mean
25F Gulabewala	0.65	0.35
Alabujanahalli	0.67	0.33
Hakamwala	0.71	0.29
Kalmandasguri	0.50	0.50
Panahar	0.60	0.40
Rewasi	0.65	0.35
Siresandra	0.80	0.20
Tehang	0.66	0.34
Zhapur	0.61	0.39

 Table 7 Probability of picking households relative to the mean

Source: Survey data

number whose extent is greater than the mean. This tendency is stronger in some villages than others, but is present in nearly all our study villages.

This, of course, is a problem that arises when sampling from any skewed (e.g. non-normal) distribution. This is usually counteracted in sampling design by



Figure 2 Sampling distribution of households picked from SSS 1 across all study villages (10, 25, 50, 100, 250, and 500 iterations)

picking a relatively large sample, thus ensuring that the sampling distribution (distribution of sampled means) is nearly normally distributed, and that the expected mean is not only equal to the population mean but close to it in a large proportion of samples.

To check the robustness of the LHS sampling design, I used the Monte Carlo method to repeatedly sample households from the study villages based on the LHS methodology. In other words, repeated samples of two households were drawn for each SSS from each of the study villages. The sampling iteration was done 10, 25, 50, 100, 250, and 500 times for each SSS. The number of iterations is chosen purely for illustrative reasons. The determining factor for the estimated survey statistics is the (small) sample size, and not the number of iterations as it is the former that makes the distribution of the sample mean potentially non-normal. (The highest number of iterations – 500 – provides the best approximation to the true and unknown distribution of the small sample survey statistics, which we are attempting to characterise through the Monte Carlo exercise.) The sampled values were then aggregated based on the sampling weights derived from the ratio of the sampled households to the household population in each SSS for each village. The results have been presented separately for each SSS, aggregated across all villages.

From this exercise we note that the sampling distribution of these repeated samples provides reasonable estimates for SSS 1 (Figure 2, Table 8), SSS 2 (Figure 3, Table 9), and SSS 3 (Figure 4, Table 10), in particular as the number of iterations increase, reflected, for example, in the mean value of the sampling mean being close to the true value. At 500 iterations, the moments of the sampling distribution show little difference between the mean sample mean and the mean sample median for SSS 1, SSS 2, and SSS 3, as well as between the mean sample values and the true values for the underlying collection of villages. The standard error, that is, the standard deviation of the sample means, is also reasonably low. These features together

		SSS 1		
Iterations	Mean	Median	Standard Error	Skewness
10	0.001	0.001	0.000	0.320
25	0.001	0.001	0.000	-0.269
50	0.001	0.001	0.000	-0.003
100	0.001	0.001	0.000	0.289
250	0.001	0.001	0.000	0.265
500	0.001	0.001	0.000	0.000
Population	0.001			1.854

Table 8 Moments of randomly picked samples of households from SSS 1 over repeatediterations (all villages)



Figure 3 Sampling distribution of households picked from SSS 2 across all study villages (10, 25, 50, 100, 250, and 500 iterations)

suggest that the cost of using a small sample and a single draw to estimate features of the first 3 strata for our study villages is tolerable.

In contrast, the cost of employing a small sample and a single draw for SSS 4 (Figure 5, Table 11) is much greater than for other strata, perhaps because of the greater skewness

		SSS 2		
Iterations	Mean	Median	Standard Error	Skewness
10	0.159	0.158	0.034	0.000
25	0.180	0.176	0.035	0.592
50	0.192	0.184	0.059	1.746
100	0.179	0.183	0.040	0.030
250	0.182	0.175	0.050	0.864
500	0.187	0.181	0.050	0.868
Population	0.186			1.497

Table 9 Moments of randomly picked samples of households from SSS 2 over repeatediterations (all villages)

of the true distribution of landholdings within the stratum. We see that there is a significant difference between the mean and the median, implying non-normality. It can be seen that the sampling distribution is positively skewed. This means that over repeated samples, a larger number of the samples comprised households that together held a mean extent of land smaller than the mean of the sampling distribution (and than the true value in the underlying population from which sampling was done). The skewness of the population in each stratum is generally captured poorly by the samples, but this, is especially marked for SSS 4. The sampling error was of the order of 2.6 ha on a mean of 7.4 ha. The erratic sampling of a sample combined with the skewness of the distribution from which sampling is taking place results in a situation in which the chance of selecting unrepresentative households is significant. The reason that SSS 4 exhibits these results so markedly is the large variance and skewness of the underlying distribution, which is not adequately accommodated by the LHS sampling design. The sample size for SSS 4 is too small to ensure that the sampling mean provides a reasonable estimate of the true mean with sufficient confidence in any given draw (or, as we have seen, even over a large number of draws).

There are significant implications of this shortcoming in the LHS methodology. While this exercise relies on repeated sampling to understand the shortcoming, by attempting to characterise the distribution of small sample statistics, in practice the LHS samples households from each FSU only once. Given the variance and the skewed distribution, a sample size of 2 proves inadequate to provide representative samples for the largest substrata with reasonable confidence. A fairly large number of times, the random sampling will lead to picking households that are smaller than the expected value of landholdings in the largest stratum. Consequently, on these occasions, the area covered by the largest size-classes will be underestimated – as will be inequality in the distribution of landholdings.









Figure 4 Sampling distribution of households picked from SSS 3 across all study villages (10, 25, 50, 100, 250, and 500 iterations)

6. Conclusions and Discussion

This note began by reviewing two important sources of data on landholdings in rural India, the Agricultural Census and the National Sample Survey of Land and Livestock Holdings (LHS). We reviewed the methodology and the concepts and definitions used in these surveys in order to identify possible shortcomings of method. In addition to definitional issues, we noted that there is a significant shortfall in the

		SSS 3		
Iterations	Mean	Median	Standard Error	Skewness
10	1.477	1.489	0.083	-0.251
25	1.447	1.453	0.065	0.385
50	1.422	1.425	0.074	0.245
100	1.424	1.432	0.066	-0.233
250	1.427	1.428	0.074	-0.032
500	1.426	1.426	0.069	0.085
Population	1.431			0.198

Table 10 Moments of randomly picked samples of households from SSS 3 over repeatediterations (all villages)

aggregate area of land as estimated by the LHS when compared to aggregates from the Agricultural Census and the Land Use Statistics. This tendency has been exacerbated over time.

To investigate the problem, I examined the sampling methodology used by the NSS for the LHS. For this, I highlighted some distributional characteristics of the underlying distribution of household landholdings in the study villages. I then employed the Monte Carlo method of sampling based on the LHS design from the surveyed villages for each of the Second Stage Strata. The result of this exercise showed that the distributional characteristics of possession of land of the village household population - in particular, the large variance combined with a skewed distribution - significantly affected the result of random sampling for the substratum with biggest landholdings, SSS 4. The sample size of two households allocated to this SSS does not result in a normal sampling distribution on account of its distributional characteristics. This means that if sampled many times over, a disproportionate number of samples included households whose mean landholdings were smaller in size than the expected value of the stratum. In the real life scenario of one-off sampling of households by the NSSO, this would mean a significantly high chance in any given survey of picking households that are insufficiently representative (in the sense of underestimating the mean of the largest substratum as well as inequalities within it). The LHS methodology lessens the chances of capturing the largest landholdings in the FSU. Consequently, the aggregate extent of land is likely to be underestimated as the observed mean from the selected households is likely to be lower than the actual mean of all households in SSS 4 in actual Land Holding Surveys.

Further, the underestimation of the extent of area held in household landholdings in SSS 4 is distributed over three size-classes in the post-enumeration size-class classification, namely, 2 to 4 ha, 4 to 10 ha, and more than 10 ha. While there will be



Figure 5 Sampling distribution of households picked from SSS 4 across all study villages (10, 25, 50, 100, 250, and 500 iterations)

an underestimation *collectively* for these size-classes, the *distribution of this underestimation* over these size-classes will depend on the pattern of land distribution in individual FSUs. Nevertheless, if the premise is that there are fewer households in the highest size-classes, with correspondingly smaller chances of being picked in small samples, then the resulting typical underestimation will be the most severe among the larger landholders – leading to substantial underestimation

		SSS 4		
Iterations	Mean	Median	Standard Error	Skewness
10	6.374	6.066	1.860	2.009
25	7.519	7.049	2.655	2.281
50	7.768	7.187	2.673	0.783
100	6.987	6.524	2.338	0.905
250	7.393	6.828	2.516	1.485
500	7.433	6.829	2.667	1.675
Population	7.419			5.810

Table 11 Moments of randomly picked samples of households from SSS 4 over repeatediterations (all villages)

of the area possessed by the largest landholders in the country as well as of the inequalities in agrarian landholding generally.

Acknowledgements: This paper has emerged from my work on a Project titled "Evaluating Official Statistics on Land and Livestock Holdings," undertaken by the Economic Analysis Unit of the Indian Statistical Institute, Bangalore Centre, under the supervision of V. K. Ramachandran and Madhura Swaminathan. The findings of the Project were first presented at a Consultation held at the Indian Statistical Institute, Bangalore Centre, on March 6, 2015. I am grateful to V. K. Ramachandran and Madhura Swaminathan, to participants at the Consultation, including T. J. Rao, K. P. Chandran, and Ramesh Chand, and to Abhijit Sen, Yoshifumi Usami, and Sanjay G. Reddy for comments. I am also grateful to my colleagues Aditi Dixit, Bheemeshwar Reddy, and Biplab Sarkar for discussions on the subject matter of this article.

References

Bakshi, A., and Ramachandran, V. K. (2008), "A Note on Land Use and Crop Area Statistics in West Bengal," available at http://www.agrarianstudies.org/UserFiles/File/ S8_Bakshi_and_Ramachandran-_A_Note_on_Land_Use_and_Crop_Area_Statistics_ W_Bengal.pdf, viewed on May 22, 2016.

Basu, Ranjini (2015), "Land Tenures in Cooch Behar District, West Bengal: A Study of Kalmandasguri Village," Review of Agrarian Studies, vol. 5, no. 1, available at http://ras.org.in/land_tenures_in_cooch_behar_district_west_bengal, viewed on June 6, 2016.

Chandrasekhar, S., and Mehrotra, N. (2016), "What Would It Take? Doubling Farmers' Incomes by 2022," *Economic and Political Weekly*, vol. 51, no. 18, available at http://www.epw.in/journal/2016/18/commentary/doubling-farmers-incomes-2022.html, viewed on June 18, 2016.

Directorate of Economics and Statistics (DES) (2014), *Land Use Statistics At A Glance: State Wise*, Department of Agriculture, Cooperation, and Farmers Welfare, Ministry of Agriculture

and Farmers Welfare, Government of India, New Delhi, available at http://eands.dacnet.nic.in/LUS_1999_2004.htm, viewed on June 18, 2016.

Food and Agriculture Organisation (FAO) (2010), 2000 World Agricultural Census: Main Results and Metadata by Country (1996–2005), available at http://www.fao.org/docrep/013/i1595e/ i1595e.pdf, viewed on May 22, 2016.

Government of India (GoI) (2006a), *Household Ownership Holdings in India, 2003*, National Sample Survey 59th Round (January–December 2003), National Sample Survey Report No. 491(59/18.1/4), National Sample Survey Organisation, Ministry of Statistics and Programme Implementation, Nov.

Government of India (GoI) (2006b), *Some Aspects of Operational Land Holdings in India, 2002–03*, National Sample Survey 59th Round (January–December 2003), National Sample Survey Report No. 492(59/18.1/3), National Sample Survey Organisation, Ministry of Statistics and Programme Implementation, Aug.

Government of India (GoI) (2010), *Manual of Schedules and Instructions for Data Collection* (*Land Record States*), Ministry of Agriculture, Department of Agriculture and Cooperation (Agricultural Census Division).

Government of India (GoI) (2014), *Key Indicators of Land and Livestock Holdings in India*, Report KI (70/18.1), available at http://mospi.nic.in/Mospi_New/upload/KI_70_18.1_ 19dec14.pdf, viewed on May 22, 2016.

Government of India (GoI) (2015), *Household Ownership and Operational Holdings in India*, Report no. 571, Ministry of Statistics and Programme Implementation, National Sample Survey Office, New Delhi, available at http://mospi.nic.in/Mospi_New/upload/ Report_571_15dec15.pdf, viewed on June 18, 2016.

NITI Aayog (2016), *Report of the Expert Committee on Land Leasing*, Government of India, New Delhi, available at http://niti.gov.in/writereaddata/files/writereaddata/files/document_publication/Final_Report_Expert_Group_on_Land_Leasing.pdf, viewed on June 18, 2016.

Ramachandran, V. K. (1980), "A Note on the Sources of Official Data on Landholdings in Tamil Nadu," Madras Institute of Development Studies, *Data Series No. 1*, Chennai.

Ramachandran, V. K., Rawal, V., and Swaminathan, M. (eds.) (2010), *Socio-Economic Surveys of Three Villages in Andhra Pradesh: A Study of Agrarian Relations*, Tulika Books, New Delhi.

Rawal, V. (2008), "Ownership Holdings of Land in Rural India: Putting the Record Straight," *Economic and Political Weekly*, vol. 43, no. 10.

Rawal, V. (2014), "Changes in the Distribution of Operational Landholdings in Rural India: A Study of National Sample Survey Data," *Review of Agrarian Studies*, vol. 3 no. 2, pp 73–104.