INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 7, Issue - 11, Nov – 2021
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date:04/11/2021 Acceptance Date: 19/11/2021 Publication Date: 30/11/2021

Can number of husbands be more than their wives: Validity of NSS Data

Apoorva Gupta

[1]Assistant Professor, Department of Economics, Ramjas College, University of Delhi, Delhi
Email - apoorvagupta@ramjas.du.ac.in

Abstract: *This paper compares five special rounds of National Sample Survey (NSS) on household social consumption on education, spreading across three decades [NSS 42 (1986-87), NSS 52 (1995-96), NSS 64 (2007-08), NSS 71 (2014), and NSS 75 (2017-18)] and explains the trends in education indicators over time. In this process, the paper compares the data across the rounds and tries to bring out the discrepancies in one of the rounds, NSS 64 (2007-08). Since academic research depends upon the validation of data, results based on out of trend data, without any concrete reason, may have serious implications on policy formulation process. This paper brings out the discrepancy in three variables of this round: sex, relationship to head and marital status. A comparison of the results is drawn with the census data at the aggregate level, wherever possible. Since it's only one NSS round, i.e. NSS 64 (2007-08) which is not going in line with other rounds, as an additional exercise, a comparison of results for this round is conducted with two other nationally representative data sets. One, we use two rounds of Indian Human Development Survey (IHDS), Round 1 (2004-05) and Round 2 (2011-12); and two, four rounds of National Family Health Survey (1992-93, 1998-99, 2005-06, 2015-16). The comparisons highlight that NSS 64 (2007-08) data not only deviates from the usual trend of NSS education data, it also deviates from the national level trends, as depicted by other comparative surveys. The paper thus brings out discrepancy in NSS 64 (2007-08) round and suggests that there is a strong reason to be cautious about using data for this round for policy implications. A solution to this discrepancy problem is also suggested.*

Key Words: *Validation of data, Education, Gender.*
JEL Classification: *C81, I20, J10*

1. INTRODUCTION:

The National Sample Survey (NSS) Organization has been conducting nation-wide socio-economic sample surveys since 1950s. These surveys are household surveys, land and livestock surveys, enterprise and establishment surveys and village surveys. A variety of topics have been covered throughout these surveys, ranging from demographics, employment & unemployment, agriculture and rural labour, household consumer expenditure, utilization of public services in health, education and other sectors, availability of infrastructure facility in Indian villages, to surveys on medium and small industrial establishments. Different rounds, corresponds to different years and hence, constitute different issues dealt in NSS surveys. Given that NSS is the oldest and nationally representative sample survey, the data from NSS rounds has been widely used for research purposes and policy formulation. However, the validity of NSS data has always been under question.

There are studies which compare NSS data with other nationally representative data sources to ensure external validity of the results presented by NSS. Most of the comparisons have usually been done for NSS employment-unemployment surveys and consumption surveys, majorly to correctly estimate unemployment and poverty rates, respectively. For instance, Suryanarayana and Iyengar (1986) highlight the various possible sources of bias in NSS data and validity of its comparison with Central Statistical Organization (CSO) – National Accounts Statistics (NAS) data. Vaidyanathan (1986) reviews possible errors in NSS consumption data and checks on the validity of the data. Nirad (2019) compares employment-unemployment surveys of NSS with another nation-wide sample survey: Indian Human Development Survey (IHDS), Round 2 (2011-12) and find that stock of person's unemployment rate for women is different across the two types of surveys. Hirway (2002) highlights the problems in the employment-unemployment survey of NSS in capturing the details of the workforce, especially for those who are involved in "difficult to measure sectors", like subsistence work. Comparing NSS results with pilot time use survey, she finds that time use surveys are better able to capture workforce than NSS surveys. However, there are some other studies, which find consistency in NSS estimates. For instance, Sundaram and Tendulkar (2003) compare data from Consumer Expenditure Survey (CES) and employment-unemployment surveys of 1990s, and find a decline in poverty estimates, for both the surveys. Moreover, these validation checks are questionable, given differences in the data collection methods, sampling strategies, estimation procedures; differences in coverage and time period of the surveys; and the possibility of non-sampling biases (Minhas, 1988). Thus, seemingly comparable estimates may not be comparable at the first place.

This paper is different from the earlier literature on two accounts. First, this paper is conducting validity checks of the data for NSS Round 64 (2007-08) on three accounts: (a) by comparing NSS data across rounds; (b) by comparing NSS data with the census data of relevant years; and (c) by comparing NSS results with two other nation-wide sample survey: Indian Human Development Survey (IHDS), Round 1 (2004-05) and Round 2 (2011-12); and four rounds of

National Family Health Survey: NFHS 1 (1992-93), NFHS 2 (1998-99), NFHS 3 (2005-06), NFHS 4 (2015-16). Second, in this paper, we are neither checking the validity of employment-unemployment surveys nor consumption data. In this paper, we are focusing on special rounds of NSS, which are rounds on social consumption surveys. These are special surveys conducted to collect data from the households on the expenditure on health and education. In this paper, we will be using data only from the education surveys and from five different rounds. Analyzing data from these rounds highlights that the data for one of the rounds, corresponding to the year 2007-08 is misleading. The details are given in Section 3. Our results thus highlight that the data for NSS Round 64 of the year 2007-08 has to be used with caution for policy formulation purposes.

The paper is organized as follows. The next section describes the data sources used for this paper. Section 3 compares results from NSS data, IHDS data and NFHS data and highlights the discrepancy in the data. Section 4 presents recoding solution for NSS 64 (2007-08) data. Section 5 concludes the paper.

2. DATA DESCRIPTION:

The paper uses five social consumption survey rounds of NSS on Education, schedule 25.2. These are NSS 42 (1986-87), NSS 52 (1995-96), NSS 64 (2007-08), NSS 71 (2014) and NSS 75 (2017-18). Apart from standard demographic characteristics, like gender, age, marital status, education, place of residence, social group and religion, these rounds ask specific questions on the education profile of the members in the age group of 3 to 35 years.[1] These are related to special modules on education characteristics of the children who are currently enrolled like, type of school, medium of instruction, and education expenditure incurred; and special module on children who are not currently enrolled, like reasons for and level of dropouts. The second data source that we will be using for this paper is the Indian Human Development Survey (IHDS) Round 1 (2004-05) and Round 2 (2011-12). IHDS is again nationally representative household survey, similar to NSS. Apart from standard demographic characteristics, like gender, age, marital status, education, place of residence, social group and religion, IHDS also asks specific questions on education of children, which allows us to compare the two data sources with each other. We compare NSS with IHDS data because IHDS has two rounds so far. Round 1 (2004-05) was conducted before NSS 64 (2007-08), while IHDS Round 2 (2011-12) was conducted after NSS 64 (2007-08). Thus, it is much closer to 2007-08, compared to NSS survey years of 1995-96 and 2014, before and after, respectively. The third data source that we use is National Family Health Survey (NFHS), four rounds: NFHS 1 (1992-93), NFHS 2 (1998-99), NFHS 3 (2005-06), NFHS 4 (2015-16). We compare NSS data with NFHS data for two reasons. One, NFHS 3 and NFHS 4 are much closer to NSS 64 (2007-08) data, compared to other NSS rounds that are considered under study. Two, both NSS and NFHS are cross-sectional data sources, unlike IHDS which is a panel data. The number of households and individuals covered in different rounds for the three data sets are given in Table 1.

Table 1: Number of Household covered

NSS Rounds	No. of Households Covered
NSS 42 (1986-87)	77,035
NSS 52 (1995-96)	72,883
NSS 64 (2007-08)	1,00,581
NSS 71 (2014)	65,926
NSS 75 (2017-18)	1,13,757

IHDS Rounds	No. of Households Covered
IHDS 1 (2004-05)	41,554
IHDS 2 (2011-12)	42,152

NFHS Rounds	No. of Households Covered
NFHS 1 (1992-93)	88,558
NFHS 2 (1998-99)	92,437
NFHS 3 (2005-06)	1,09,041
NFHS 4 (2015-16)	6,01,509

Source: Author's calculations based on NSS, IHDS and NFHS data

[1]NSS 75 (2017-18) focused on 3-35 years; NSS 71 (2014) and NSS 64 (2007-08) focused on the age group 5-29 years; NSS 52 on the age group 5-24 and NSS 42 did not specify any age. For overall comparison, we will focus on all individuals. However, while comparing education indicators, to remain consistent over time, we will be focusing on the children in the age group of 5-18 years.

There are two things to note from Table 1. One, the number of households covered by NSS varies across rounds, however, given that it is nationally representative data, and the rounds should be comparable to each other, what matters more is the average of the variables of interest. Two, the number of households covered by NSS is more than that covered by IHDS and far less than covered by NFHS, however, on an average, the result should match. Moreover, the discrepancy that we want to highlight in NSS Round 64 (2007-08) is not really because of different sampling strategy followed.[2] The descriptive statistics on the distribution of households across rounds, for the three data sources show similar picture. However, the trends become very different, when the data is taken to individual level. And this is where we answer why there is a discrepancy in NSS 64 (2007-08) and why we compare it with IHDS and NFHS rounds. Table 2 below shows the percentage of males and females covered across various rounds of NSS.[3] A comparison is drawn with census data, for the respective years.[4]

Table 2: Sex composition across NSS rounds

NSS Rounds	Male	Female	Sex Ratio	Census Year	Sex Ratio
NSS 42 (1986-87)	51.79	48.21	905	1981	934
NSS 52 (1995-96)	51.82	48.18	910	1991	927
NSS 64 (2007-08)	48.52	51.48	1059	2001	933
NSS 71 (2014)	51.45	48.55	936	2011	943
NSS 75 (2017-18)[#]	51.87	48.12	916	2021[$]	945

Source: NSS figures are based on author's calculations based on NSS data. Census numbers are taken from various census documents over time. [#]: Round 75 (2017-18) has a third category for sex: Transgender. These are not included in the analysis [$]: For the year 2021, it is projected number

In India, since the first census, after independence, conducted in 1951, till the latest available census of 2011, the sex ratio as always been in favour of men. That is, there has always been more number of men than women in India. Table 2 shows two things. One, every round of NSS underestimate the true population sex ratio, except NSS 64 (2007-08). Two, NSS 64 (2007-08) is an outlier. It not only over estimates the census sex ratio; it shows more number of females than males in India, which has never happened in independent India so far. One may believe that it is a sample survey, and the estimates are based on sample weights, and hence the numbers might be correct. To answer this, we compare the sex ratio from NSS data with IHDS and NFHS data sources. Table 3 shows the percentage of males and females covered across two rounds of IHDS and four rounds of NFHS.[5]

Table 3: Sex composition across IHDS and NFHS rounds

IHDS Rounds	Male	Female	Sex Ratio
IHDS 1 (2004-05)	50.74	49.26	965
IHDS 2 (2011-12)	49.72	50.28	1004

NFHS Rounds	Male	Female	Sex Ratio
NFHS 1 (1992-93)	50.79	49.18	966
NFHS 2 (1998-99)	50.78	49.22	965
NFHS 3 (2005-06)	49.94	50.06	984
NFHS 4 (2015-16)	50.23	49.77	989

Source: Author's calculations based on IHDS and NFHS data sets

[2] We have analyzed the sampling strategy for the NSS rounds under considerations. It is similar for all rounds. Moroever, the distribution of households covered across states, sector (rural and urban), social group and religion; across different rounds of NSS is same. Similarly, the distribution of households across states, region, social group and religion is same for IHDS and NFHS rounds, respectively. Thus, there is no particular reason in any round, for any dataset to be different when the distribution of households surveyed is concerned. The details are available from the author upon request.

[3] Sample weights have been used in all the calculations.

[4] NSS 75 (2017-18) has a third category for sex: Transgender. These are not included in the analysis for this paper. Hence, the numbers for males and females for this round, may not add up to 100.

[5] Sample weights have been used in all the calculations.

Table 3 clearly shows that IHDS data over estimate the number of females in India. However, this overestimation is insignificant (compared to significant overestimation in NSS 64 (2007-08)). Moreover, for NFHS rounds, the sex ratios have always been in favour of males. For NFHS 3, though the percentage of females (using sample weight) is greater than that of males, yet it is insignificant. This is where the paper contributes to the literature and highlights the discrepancy in NSS 64 (2007-08). Moreover, even if females are over estimated in any data, in any round, it should not affect the variables of interest for any researcher conducting the study. However, this is not true with NSS 64 Round data. This paper, thus, discusses these datasets at length and highlights how three variables, sex of a person, relationship to head and marital status, collected in NSS 64 (2007-08) are wrongly coded and hence, can make the data for that round non-comparable to other NSS rounds. It is important to highlight this discrepancy in the data as these are special education survey rounds and usually the government policies related to reducing gender gaps in education are made and implemented on the basis of research conducted using such data sources. In the next section, we highlight this discrepancy, and suggest a solution on re-coding these variables. We suggest how re-coding these variables can make NSS 64 round data comparable to other rounds.

3. COMPARING DEMOGRAPHIC CHARACTERISTICS: AGE AND FEMALE HEADED HOUSEHOLDS

This section compares the demographic characteristics of individuals covered across the five rounds of NSS under study. The main variable of interest for this section to study is the sex of a person. Table 4 presents the average age of individuals across different rounds of NSS, IHDS and NFHS.

Table 4: Average age (in years)

Rounds	Male	Female
NSS 42 (1986-87)	24.05	24.55
NSS 52 (1995-96)	25.18	25.24
NSS 64 (2007-08)	27.30	26.74
NSS 71 (2014)	28.26	29.14
NSS 75 (2017-18)	29.60	30.58
IHDS 1 (2004-05)	27.05	27.29
IHDS 2 (2011-12)	29.35	30.19
NFHS 1 (1992-93)	24.79	24.67
NFHS 2 (1998-99)	25.39	25.38
NFHS 3 (2005-06)	26.22	26.54
NFHS 4 (2015-16)	29.11	29.62

Source: Author's calculations based on NSS, IHDS and NFHS datasets.

Table 4 shows that across all NSS rounds (except NSS 64), and for different rounds of IHDS and NFHS, females are significantly older than males.[6] Thus, even if IHDS dataset overestimate the number of females compared to the census data (as shown in Table 3), the trend for age is similar to that of NSS data. It is NSS 64 (2007-08) which is showing off track results, wherein females are significantly younger than males.

The biggest and the most crucial affect of this discrepancy in sex composition presented in NSS 64 (2007-08) become visible in the data for head of the household. Usually, the head of the household is the prime decision maker of the family. The sex of the household head can significantly affect many aspects of a household, especially through the head's education and occupation, like child's education (Saha, 2013), child's health (Mukherjee et al., 2011), household food demand (Abdulai et al., 1999), inequality (Cain et al., 2009), access to microfinance and poverty levels (Imai et al., 2010), and their overall survival strategies (Mencher, 1993). Table 5 shows the percentage of male and female headed households across different rounds of three datasets.

[6] Though for NFHS1 and NFHS2, females are younger than males, however the difference is insignificant. For NFHS3 and NFHS4, females are significantly older than males.

Table 5: Percentage of male and female headed households

Rounds	Household Head	
	Male	Female
NSS 42 (1986-87)	90	10
NSS 52 (1995-96)	91.11	8.89
NSS 64 (2007-08)	11.37	88.63
NSS 71 (2014)	87.98	12.02
NSS 75 (2017-18)	87.79	12.2
IHDS 1 (2004-05)	90.04	9.86
IHDS 2 (2011-12)	85.22	14.78
NFHS 1 (1992-93)	90.75	9.24
NFHS 2 (1998-99)	89.69	10.31
NFHS 3 (2005-06)	85.64	14.36
NFHS 4 (2015-16)	85.37	14.63

Source: Author's calculations based on NSS, IHDS and NFHS datasets

As per census data, female headed households have increased from about 19 million in 2001 to more than 26 million in 2011, showing an increase from about 9 percent to 14 percent, a 5 percentage point increase. However, the percentage of such households has usually remained below 15 percent.[7] This is shown in the trends of various rounds of NSS, IHDS and NFHS under study. It is only for NSS 64 (2007-08); the percentage of female headed households has surged drastically, indicating some discrepancy in data for this round. As mentioned earlier, since household decisions related to child's education are influenced by the gender of household head, it may have some serious implications on the household level decisions taken, and thereby on the results observed. It may also affect the government policies that are specially geared towards female headed households.[8]

Given this discrepancy in the data, one solution that is suggested is to re-code the sex variable for NSS 64 (2007-08). Post recoding, i.e. when the codes for males and females are reversed for this particular round, the numbers are matching with the trend of other NSS rounds and with the IHDS and NFHS rounds. Table 6 shows the percentage of males and females headed households, post recoding sex variable.

Table 6: Percentage of male and female headed households, post recoding

Rounds	Household Head	
	Male	Female
NSS 42 (1986-87)	90	10
NSS 52 (1995-96)	91.11	8.89
NSS 64 (2007-08)	88.63	11.37
NSS 71 (2014)	87.98	12.02
NSS 75 (2017-18)	87.79	12.2
IHDS 1 (2004-05)	90.04	9.86
IHDS 2 (2011-12)	85.22	14.78
NFHS 1 (1992-93)	90.75	9.24
NFHS 2 (1998-99)	89.69	10.31
NFHS 3 (2005-06)	85.64	14.36
NFHS 4 (2015-16)	85.37	14.63

Source: Author's calculations based on NSS, IHDS and NFHS datasets

Thus, recoding the variables is a potential solution to the problem that is highlighted in the dataset. This problem is not just in sex variable for NSS 64 (2007-08). It is for two more variables: Relationship to head variable and marital status variable. The next section describes the problem and the recoding solution for these variables.

[7] Data collected from various census website for the year 2001 and 2011. Web links are mentioned in the References.

[8] Government of India has a scheme for supporting training and employment programme for women who are in female headed households. For more details, refer to Support to Training and Employment Programme for Women | National Portal of India (accessed on June 14, 2021)

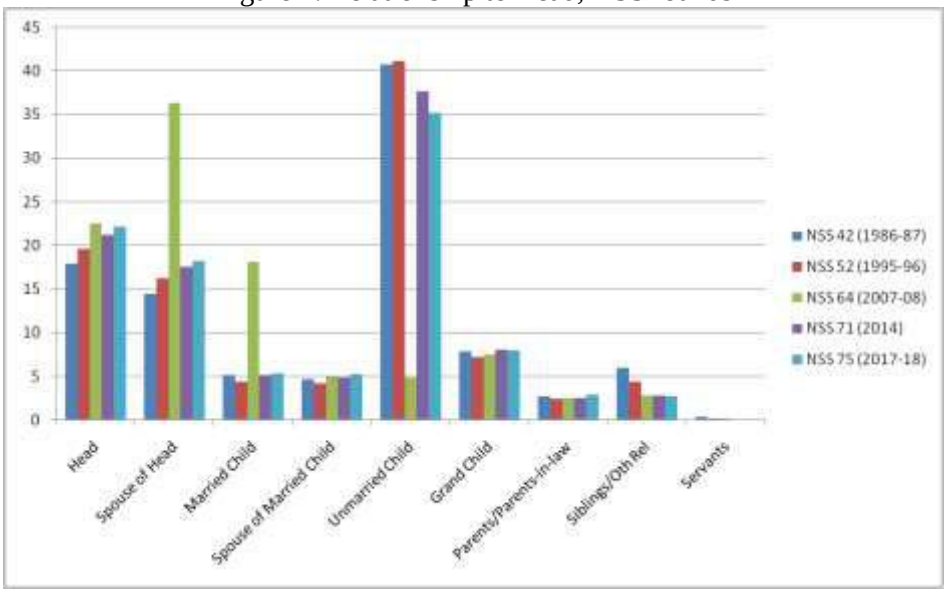
4. COMPARISON OF NSS WITH IHDS AND NFHS DATASET: CAN RE-CODING HELP

This section brings out the discrepancy in other variables for NSS 64 (2007-08): relationship to head and marital status. It gives the details of these variables, the problem in coding and the recoding solution thereof. In describing the problem and the solution, I have not just compared the numbers for this round with the numbers for other NSS rounds, and with IHDS and NFHS rounds, but also compared bivariate trends for NSS data.

a. Relationship to Head

In all the five rounds of NSS that are under study, the codes of relationship to head variable are as follows: *self -1, spouse of head -2, married child -3, spouse of married child -4, unmarried child -5, grandchild -6, father/mother/father-in-law/mother-in-law -7, brother/sister/brother-in-law/sister-in-law/other relatives -8, servants/ employees/ other non-relatives -9*. The percentages of all the household members are shown in Figure 1.

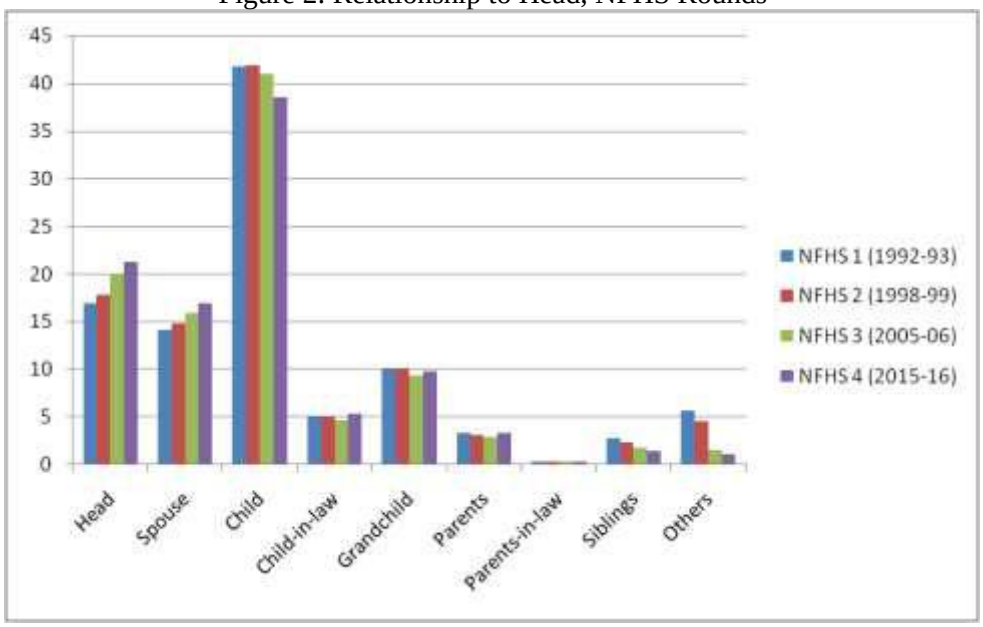
Figure 1: Relationship to Head, NSS rounds



Source: Author's calculations based on NSS rounds

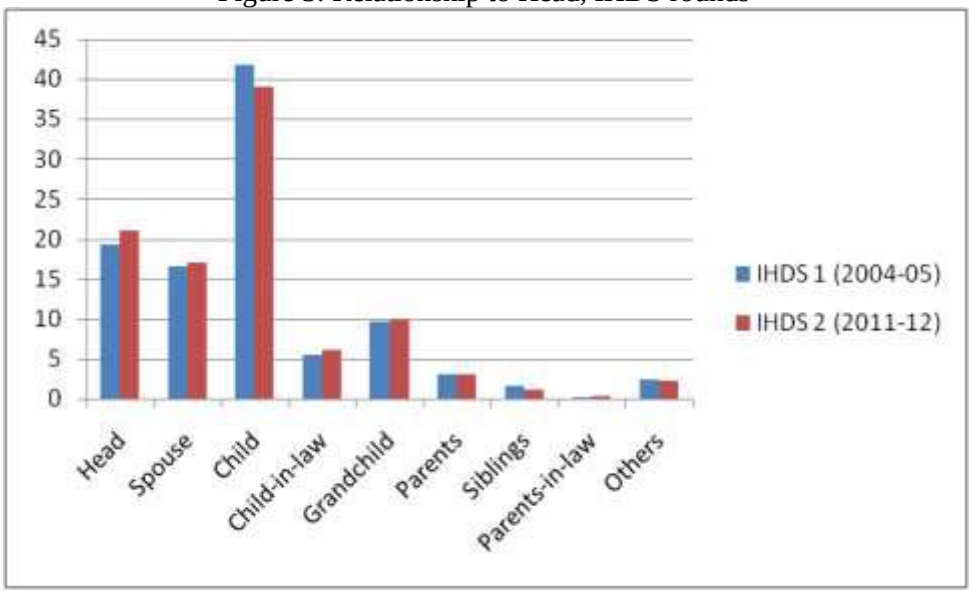
The graph shows that 36 percent of household members claim to be "spouse of head" in data for NSS 64 (2007-08); while this percentage for other rounds ranges from 14 percent to 18 percent. Similarly, the percentage of married and unmarried children differs significantly from all other rounds. The percentages of different members in a household for IHDS and NFHS rounds are given in Figure 2 and 3, respectively.

Figure 2: Relationship to Head, NFHS Rounds



Source: Author's calculations based on NFHS rounds

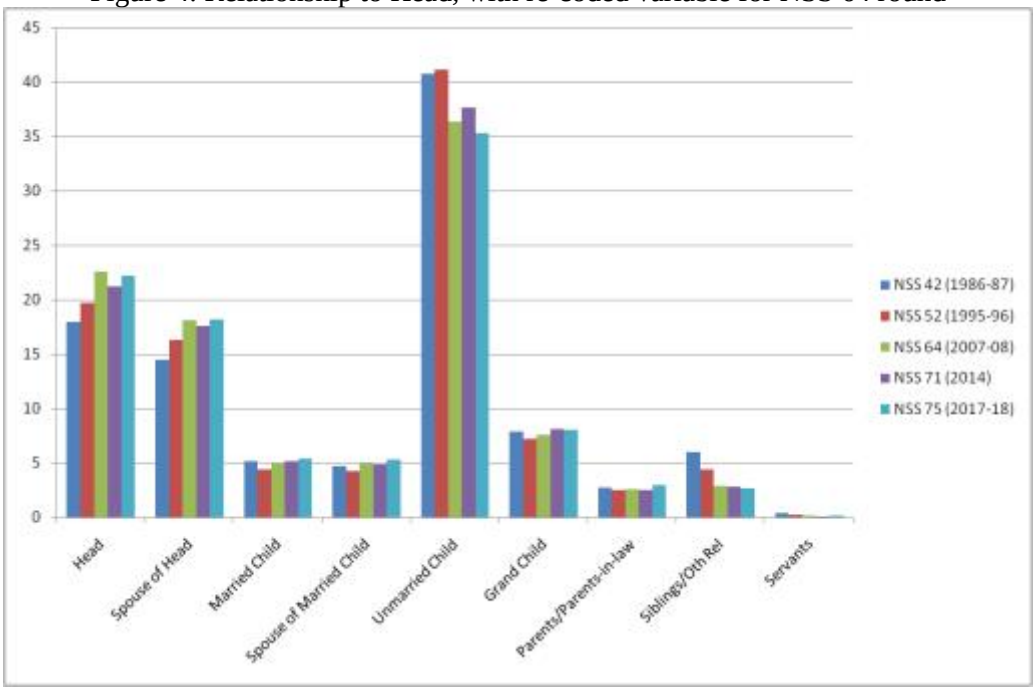
Figure 3: Relationship to Head, IHDS rounds



Source: Author's calculations based on IHDS rounds

Comparing the data for NSS 64 (2007-08) with other rounds of NSS and with IHDS and NFHS, it is clear that there is a problem in relationship to head variable for this round. As mentioned earlier, since research is based on the validity of data, study based on NSS 64 (2007-08) using this variable may give misleading results. As a suggested solution, recoding the "relation to head" variable only for this particular round makes the data in sync with other rounds. If recoding is done in the following manner: *self -1, spouse of head -5, married child -2, spouse of married child -3, unmarried child -4, grandchild -6, father/mother/father-in-law/mother-in-law -7, brother/sister/brother-in-law/sister-in-law/other relatives -8, servants/ employees/ other non-relatives -9,* the percentages for this round for this variable will be in sync with other rounds. Thus, the person who claims to be "spouse of head" and should be coded as "2", was wrongly given the code as "3". Similarly, the person who claims to be "married child" and should be coded as "3", was wrongly given the code as "4". The person, who claims to be "spouse of married child" and should be coded as "4", was wrongly given the code as "5". And the person who claims to be "unmarried child" and should be coded as "5", was given the code as "2". Recoding the household members accordingly makes the data for this round in harmony with other rounds. Figure 4 presents the results.

Figure 4: Relationship to Head, with re-coded variable for NSS 64 round



Source: Author's calculations based on NSS rounds

Given that for solving the discrepancy in the data for NSS 64 (2007-08) we have re-coded the two variables: sex and relationship to head; to bring the data in sync with other variables, marital status has also to be re-coded. Going by the original codes for sex and relationship to head, it would appear that there are more husbands to a single wife in NSS 64, which given the Indian marriage system, is very less likely to occur. Thus, this raises the question on validity of NSS 64, and hence we suggest a solution for recoding these variables.

b. Marital Status

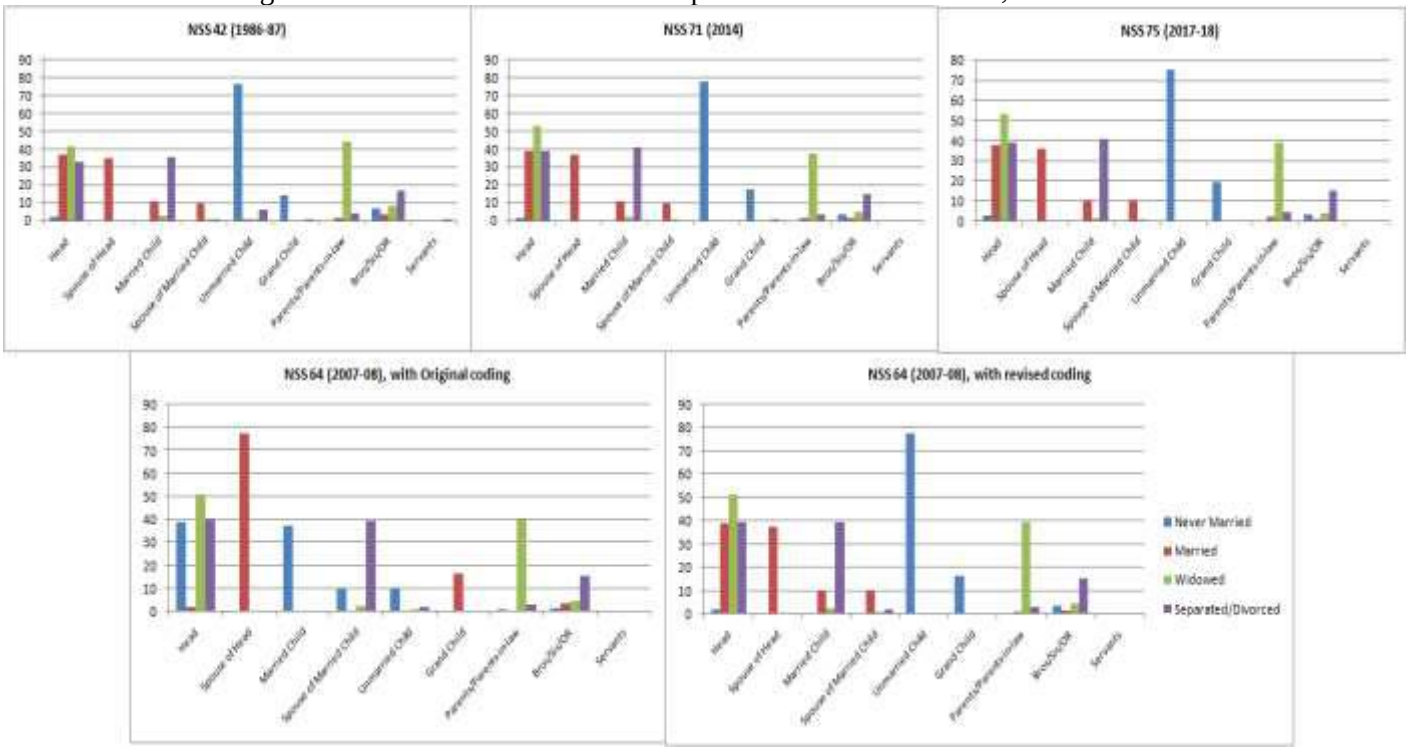
NSS reports the marital status of a person as: *never married – 1, currently married – 2, widowed – 3, divorced/separated – 4.* Marital status was not asked in NSS 52 (1995-96), however, for all other rounds, data is available. Table 7 gives the percentages of individuals across NSS rounds, by their marital status. Though at first instance, it could be argued that this variable is correctly coded and there is no need to recode this variable. However, when "relationship to head" variable is recoded, there is a need to recode "marital status" variable accordingly. Otherwise, the data will not be in synchronization with the data of other rounds. To substantiate this, Figure 5 plots the percentages of household members as per their marital status and as per their relation with the head of the household. With the original coding (as given in the data and codebook), the percentages for NSS 64 (2007-08) is showing a completely different picture. However, when the two variables are recoded, these percentages get synchronized with other rounds.[9]

Table 7: Marital Status for NSS rounds

	Panel A: Original Codes for NSS 64 (2007-08)				Panel B: Recodes for NSS 64 (2007-08)			
	NSS 42 (1986-87)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)	NSS 42 (1986-87)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)
Never Married	53.24	48.04	47.91	45.69	53.24	46.7	47.91	45.69
Married	41.65	46.7	47.24	49.4	41.65	48.04	47.24	49.4
Widowed	4.77	4.92	4.55	4.66	4.77	4.92	4.55	4.66
Separated/Divorced	0.34	0.35	0.3	0.25	0.34	0.35	0.3	0.25
Total	100	100	100	100	100	100	100	100

Source: Author's calculations based on NSS rounds

Figure 5: Marital status and relationship to head of the household, NSS Rounds



Source: Author's calculations based on NSS rounds

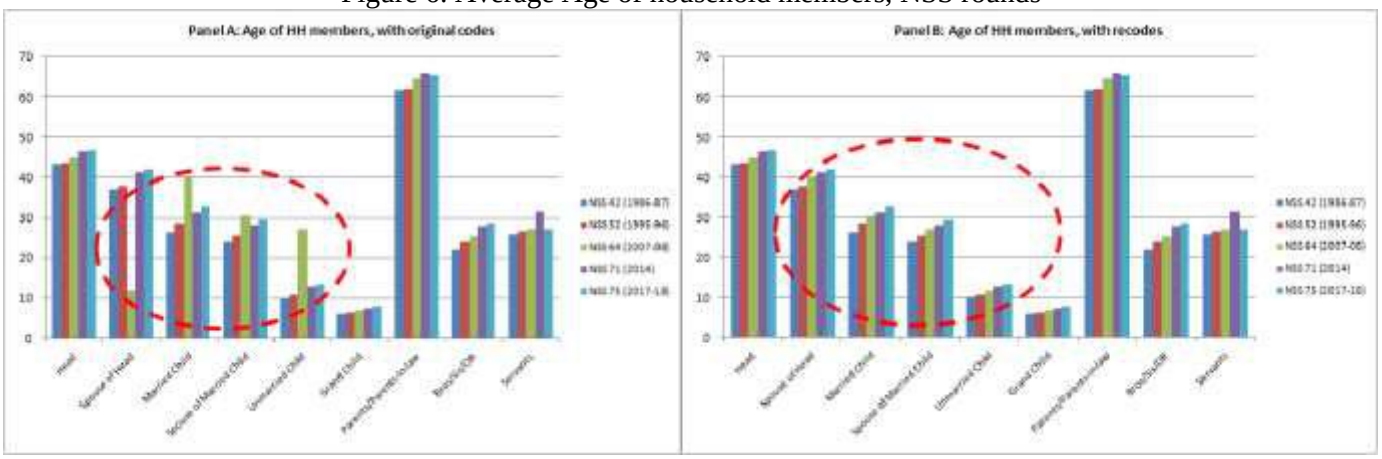
[9]I have tried similar graphs with re-coding the two variables one by one. The results are same.

Thus, Figure 5 shows that using original codes for marital status for NSS 64 (2007-08) versus the re-codes bring in huge differences when this is plotted with other variables of interest. The following sub-sections depict some of the variables that have got affected because of the differences in coding issues in NSS 64 (2007-08).

c. Relationship to Head and Age

Figure 6 gives the average age of households members with respect to their relationship with the head of the household. Panel A gives with original codes, while Panel B gives when the recoded variable for NSS 64 (2007-08) is used. Spouse of the head was significantly shown to be under aged, while the married child, spouse of married child and unmarried child were shown to be significantly over aged in relation to their counterparts in other rounds. Post recoding, the data is in sync with other rounds.

Figure 6: Average Age of household members, NSS rounds

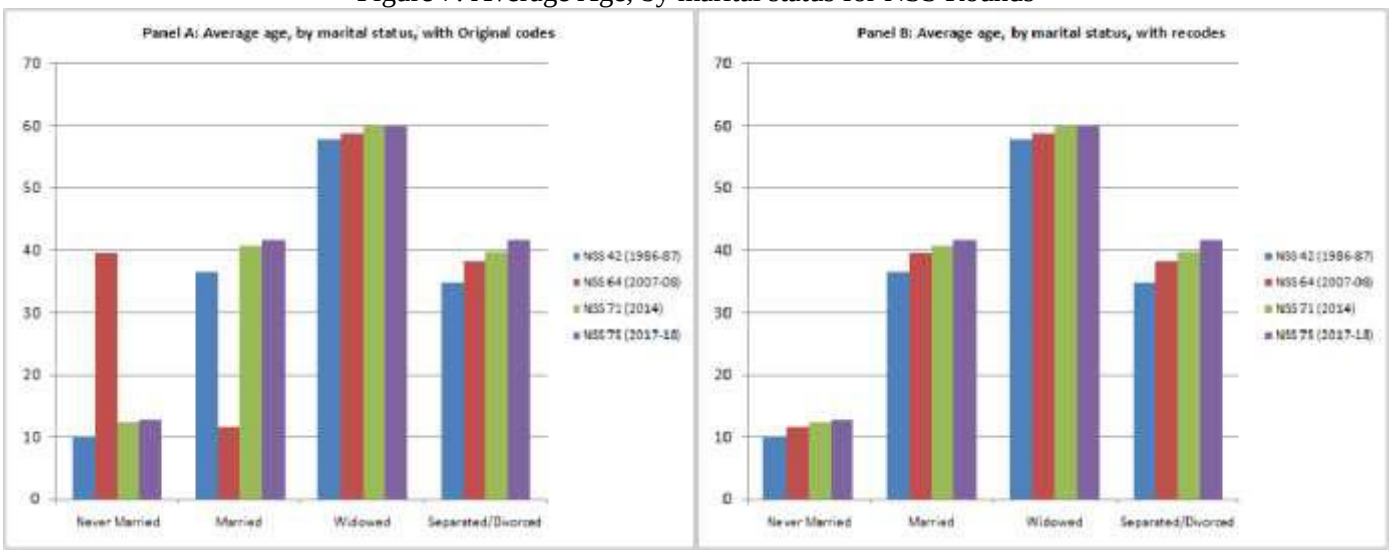


Source: Author's calculations based on NSS rounds

d. Marital Status and Age

Figure 7 gives the average age of individuals by their marital status for different NSS rounds. Panel A gives with original codes, while Panel B gives when the recoded variable for NSS 64 (2007-08) is being used. Recoding the marital status variable for NSS 64 makes the data in sync with other rounds.

Figure 7: Average Age, by marital status for NSS Rounds



Source: Author's calculations based on NSS rounds

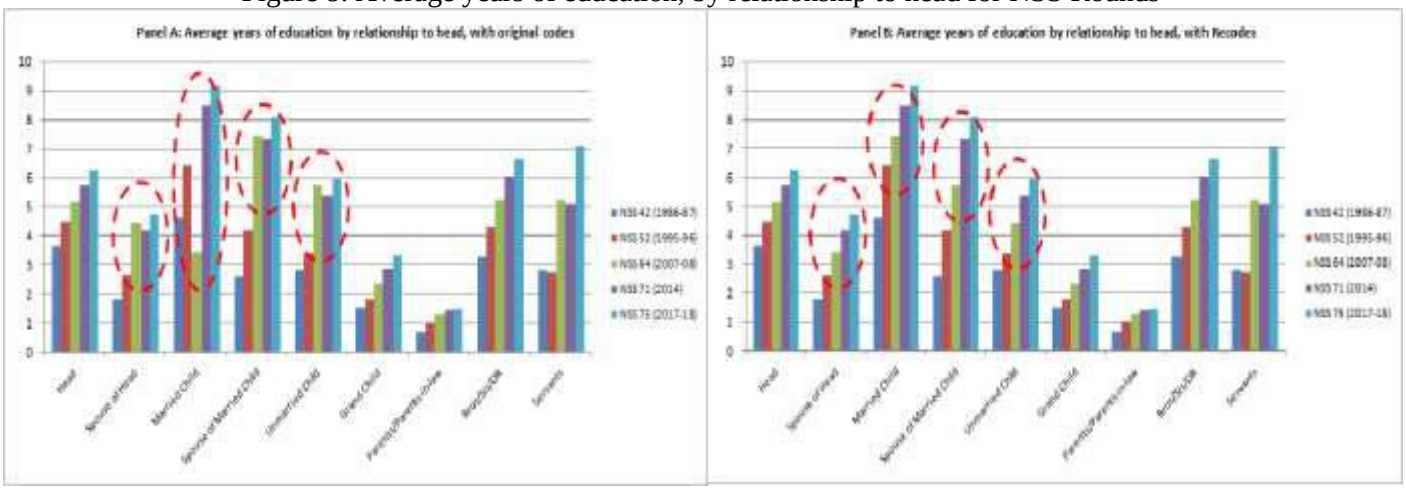
e. Years of Education

NSS data collects information on the education level that any person has completed. However, the coding for these levels of education was different in different years. We have made these categories consistent across rounds. Following Deshpande and Ramachandran (2016a), we have recoded them into years of education. These are: illiterate – 0 years of schooling; without formal schooling or below primary – 3 years of schooling; primary – 5 years of schooling; middle –

8 years of schooling; secondary – 10 years of schooling; higher secondary – 12 years of schooling; graduation – 15 years of schooling; above graduation – 17 years of schooling.

We have calculated average years of education of individuals, by their relationship to the head of the household, sex and marital status for all the five rounds. This section shows how the coding problem of NSS 64 (2007-08) data has made years of education of individuals to get significantly different from the trend which other rounds are showing. Figure 8 plots average years of education with relationship to head. Married child is shown to have significantly lower average years of education, which gets in line with other rounds, post recoding of the relationship to head variable. Table 8 shows average years of education with respect to sex. Males, in all other rounds are having significantly higher years of education than females, except for NSS 64. Table 9 shows similar numbers with respect to marital status of individuals. Married individuals, in all other rounds are having significantly higher years of education compared to unmarried individuals, except NSS 64. Post recoding, the average years of education of individuals for NSS 64 (2007-08) are in line with the trends depicted by other rounds.

Figure 8: Average years of education, by relationship to head for NSS Rounds



Source: Author's calculations based on NSS rounds

Table 8: Average years of education, by sex for NSS Rounds

	Panel A: Original Codes for NSS 64 (2007-08)					Panel B: Recodes for NSS 64 (2007-08)				
	NSS 42 (1986-87)	NSS 52 (1995-96)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)	NSS 42 (1986-87)	NSS 52 (1995-96)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)
Male	3.52	4.26	3.68	5.85	6.38	3.52	4.26	5.09	5.85	6.38
Female	1.97	2.73	5.09	4.53	5.04	1.97	2.73	3.68	4.53	5.04
Transgender	.	.	.	.	3.95	.	.	.	.	3.95
Total	2.78	3.52	4.41	5.21	5.74	2.78	3.52	4.41	5.21	5.74

Source: Author's calculations based on NSS rounds

Table 9: Average years of education, by marital status for NSS Rounds

	Panel A: Original Codes for NSS 64 (2007-08)				Panel B: Recodes for NSS 64 (2007-08)			
	NSS 42 (1986-87)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)	NSS 42 (1986-87)	NSS 64 (2007-08)	NSS 71 (2014)	NSS 75 (2017-18)
Never Married	2.74	4.89	4.97	5.54	2.74	4.18	4.97	5.54
Married	3.05	4.18	5.74	6.23	3.05	4.89	5.74	6.23
Widowed	1.00	1.89	2.15	2.36	1.00	1.89	2.15	2.36
Separated/Divorced	1.76	3.69	5.21	5.83	1.76	3.69	5.21	5.83
Total	2.78	4.41	5.21	5.74	2.78	4.41	5.21	5.74

Source: Author's calculations based on NSS rounds

There could be other such trends which can be represented by using these three variables: relationship to head, gender and marital status, especially when this is education specific round. For instance, number of currently enrolled students, number of students attending private schools, English medium schools, and education expenditure on students,

are few to name such variables.[10] Given that there is literature documenting gender bias and hence gender gaps in education in India (Lancaster et al., 2008, Himaz, 2009; Azam and Kingdon; 2013; Azam, 2016; Saha, 2013; Deshpande and Gupta, 2020), it is important to get clarification on coding of these variables to make sure that the data for this round can be correctly used by the researchers to get meaningful results, which are comparable to other rounds.

5. CONCLUDING REMARKS:

The aim of this paper is to bring out the discrepancy in NSS 64 (2007-08) data, which is not only out of trend of other education specific rounds, but is also apart from census data, IHDS and NFHS data. Though there are studies that have been conducted using data from this round, but they haven't compared the results with other NSS rounds. For instance, Saha (2013) find gender disparity in intra-household educational expenses. She has used Blinder-Oaxaca Decomposition strategy to estimate gender gap in education expenditure and has controlled for other characteristics as well, like social group, sector of residence, household type, household size, and state. However, given that there is discrepancy in the data, brings this study under question. Thus, a deeper analysis of gender gaps in education indicators is required. Moreover, it is crucial to study whether this gap has increased or decreased over time, or whether this gap is more at the enrolment stage, and or at the expenditure incurring stage or both.[11] Similarly, Sankar (2011) studies the participation rates of adolescents in secondary education, by gender and finds gender disparities in age specific enrolment rate, and gross and attendance ratios. Using simple descriptive statistics, our paper also presents gender gaps in years of education and trends in these gender gaps over time. However, our paper goes one step ahead, in explaining why using NSS 64 (2007-08) data can give misleading results on gender gaps in education indicators, and hence can have implications on the results of the research studies based on data for this specific round. It not just highlights the coding problem, it also suggests a solution to the problem and presents the results post recoding the variables. Thus, it is crucial that data for the research studies have to be correct, on an average, as it can have implications on policies that both central and state governments have been undertaking for quite a long time for reducing gender disparities and hence gender gaps in education indicators.

REFERENCES:

1. Abdulai, A., Jain, D. K., & Sharma, A. K. (1999). Household food demand analysis in India. *Journal of Agricultural Economics*, *50*(2), 316-327
2. Azam, M (2016), "Private tutoring: evidence from India. *Review of Development Economics*, *20*(4), 739-761
3. _____, & Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, *42*, 143-164
4. Cain, J. S., Hasan, R., Magsombol, R., & Tandon, A. (2010). Accounting for inequality in India: Evidence from household expenditures. *World Development*, *38*(3), 282-297.
5. Deshpande, A & Ramachandran, R. (2016), "The changing contours of intergroup disparities and the role of preferential policies in a globalizing world: Evidence from India", *CDE WP, (267)*
6. _____, & Gupta, A. (2019). *Nakusha? Son Preference, Resource Concentration and Gender Gaps in Education* (DP No. 07/19). Ashoka University, Department of Economics.
7. Himaz, Rozana (2009), "Is there a boy bias in household education expenditure: The case of Andhra Pradesh in India", MPRA Paper No. 21889
8. Hirway, I. (2002). Employment and unemployment situation in 1990s: How good are NSS data?. *Economic and Political Weekly*, 2027-2036.
9. Imai, K. S., Arun, T., & Annim, S. K. (2010). Microfinance and household poverty reduction: New evidence from India. *World Development*, *38*(12), 1760-1774.
10. Kingdon, G.G. (2005), "Where has all the bias gone? Detecting gender bias in the intrahousehold allocation of educational expenditure", *Economic Developmentand Cultural Change*, *53*(2), 409-451
11. Lancaster, Geoffrey, Pushkar Maitra & Ranjan Ray (2008), "Household Expenditure Patterns and Gender Bias: Evidence from Selected Indian States, Oxford Development Studies, 36:2, 133-157
12. Mencher, J.T. (1993), "Female-headed, Female-supported Households in India: Who are They and What are the Survival Strategies? In where did all the men go?" J.P Mencher and A. Okongwu (Eds) Westview Press, 203 – 231

[10]We have made similar graphs for these variables as well, by gender, relationship to head and marital status. Details are available from the author.

[11] Kingdon (2005) uses Hurdle Model to show that once the decision to enroll a girl child to a school has been taken, there is no gender gap in education expenditure.

13. Minhas, B. S. (1988). Validation of large scale sample survey data Case of NSS estimates of household consumption expenditure. *Sankhyā: The Indian Journal of Statistics, Series B*, 279-326.
14. Mukherjee, S., Haddad, S., & Narayana, D. (2011). Social class related inequalities in household health expenditure and economic burden: evidence from Kerala, south India. *International Journal for Equity in Health*, *10*(1), 1-13.
15. Nirad, I. (2019). An Exercise in Cross-Validation of Data on Employment and Unemployment. *Annals of the University Dunarea de Jos of Galati: Fascicle: I, Economics & Applied Informatics*, *25*(3).
16. Saha, A. (2013), "An assessment of gender discrimination in household expenditure on education in India", *Oxford Development Studies*, *41*(2), 220-238
17. Sankar, D. (2011). *Participation in India: An analysis of NSS 64th round data*. World Bank.
18. Sekher, T. V. (2012), "Ladlis and Lakshmis: fifinancial incentive schemes for the girl child", *Economic and Political Weekly*, 58-65
19. Sundaram, K., & Tendulkar, S. D. (2003). Poverty has declined in the 1990s: A resolution of comparability problems in NSS consumer expenditure data. *Economic and Political Weekly*, 327-337.
20. Suryanarayana, M. H., & Iyengar, N. S. (1986). On the reliability of NSS data. *Economic and Political Weekly*, 261-264.
21. Vaidyanathan, A. (1986). On the validity of NSS consumption data. *Economic and political Weekly*, 129-137.

WEB LINKS

Census data has been accessed through various census documents. Mentioned below are the web links for the same. All these links have been accessed during June-July, 2021

- Census of India - Male & Female headed households (censusindia.gov.in)
- Census of India Website : Office of the Registrar General & Census Commissioner, India (censusindia.gov.in)
- literates-and-literacy-rates-by-sex-2011-census.pdf (educationforallinindia.com)

ACKNOWLEDGEMENTS

The author would like to thank conference participants at Department of Economics, MSU, Udaipur and the anonymous referees for their useful suggestions on the previous draft of this paper.