The Indian Journal of Labour Economics (2023) 66:181–201
https://doi.org/10.1007/s41027-023-00427-8

ARTICLE

Labour Incomes in India: A Comparison of Two National Household Surveys

Mrinalini Jha[1] · Amit Basole[2]

Accepted: 5 January 2023 / Published online: 25 February 2023

Abstract

The COVID-19 pandemic created a need for high-frequency employment and income data. Policy-makers and researchers of developing countries typically have not had access to such data. In India, a new private high-frequency panel dataset has recently emerged as the dataset of choice for analysis of the economic impact of COVID-19. This is the Consumer Pyramids Household Survey (CPHS) conducted by the Centre for Monitoring the Indian Economy (CMIE). But the CPHS has also been criticised for being inadequately representative nationally by missing poor and vulnerable households in its sample. We examine the comparability of monthly labour income estimates for the pre-pandemic year (2018–19) for CPHS and the official Periodic Labour Force Survey (PLFS). Across different methods and assumptions, as well as rural/urban locations, CPHS mean monthly labour earnings are anywhere between 5 percent and 50 percent higher than corresponding PLFS estimates. In addition to the sampling concerns raised in the literature, we point to differences in the way employment and income are captured in the two surveys as possible causes of these differences. While CPHS estimates are always higher, it should also be emphasised that the two surveys agree on some stylised facts regarding the Indian workforce. An individual earning ₹50,000 per month lies in the top 5 percent of the income distribution in India as per both surveys. Second, both PLFS and CPHS show that half the Indian workforce earns below the recommended National Minimum Wage.

Keywords Income data · Labour income · Income distribution · Household survey data · India

✉ Amit Basole
amit.basole@apu.edu.in

1 O.P. Jindal Global University, New Delhi, India

2 Azim Premji University, Bengaluru, India

JEL Classification D31 · J31

1 Introduction

The COVID-19 pandemic devastated livelihoods worldwide. Formulating and implementing an effective public response to the pandemic requires high-frequency data. In the Indian case, the only national-level household survey data on incomes, employment and consumption, that was available throughout the pandemic at a lag of a few months, was the private Consumer Pyramids Household Survey (CPHS) of the Centre for Monitoring the Indian Economy (CMIE). This dataset has been used extensively to analyse the economic impact of the pandemic as well as the reach of relief measures (Abraham, Basole & Kesar, 2021; Bhattacharya & Sinha Roy 2021; Bussolo et al. 2021; Deshpande 2020, Deshpande 2022; Dhingra & Ghatak 2021; Gupta et al., 2021; Gupta et al. 2022; State of Working India, 2021; Vyas 2020). More recently, this dataset has also been used by World Bank researchers to estimate extreme poverty in India prior to the pandemic (Sinha Roy & van der Weide, 2022). However, the CPHS is a complex and relatively new dataset, and several questions remain unanswered as to its comparability with official surveys released by the National Statistical Office (NSO). In this paper, we examine the comparability of labour income estimates from the CPHS and India's official labour force survey, the Periodic Labour Force Survey (PLFS).

In the past year, there has been a debate about the sampling and representativeness of the CPHS. Somanchi (2021) compares CPHS with the National Family Health Survey (NFHS) on age structure, sex ratio, educational distribution, and asset ownership. He finds that the former underrepresents women and young children, and overrepresents well-educated households while underrepresenting the poor. Pais and Rawal (2021) critique the sampling strategy of the survey. On comparing data with other national surveys on some key employment and occupation variables, they identify problems with survey design and implementation which could result in exclusion of households belonging to marginalised communities. Vyas (2021a, 2021b) counters these claims and maintains that even with differences in the sampling techniques when compared with official surveys, CPHS is representative at the national level. Abraham and Shrivastava (2022) compare the employment status for men and women across CPHS, Labour Bureau and PLFS data to find that even with the differences in methodologies, the estimates for labour market indicators for men coming from CPHS are comparable with those from other surveys. They find results to vary for women with measures of participation in the labour force being sensitive to the way questions were asked in the different surveys. Sinha Roy and van der Weide (2022) have recently developed an alternative set of household weights for the CPHS sample that brings survey averages in closer agreement with NSO estimates.

We build on this literature by comparing estimates of labour income from the CPHS and PLFS. But our aim is not to "benchmark" CPHS estimates to PLFS (implying that PLFS numbers are true estimates which other surveys must be able to reproduce). Rather we aim to improve our understanding of CPHS data by comparing it to a more familiar data source, and in the process also to increase our confidence in estimated labour income levels by placing two independent sources side by side. Our exercise is also likely to be of interest to policy-makers and researchers in other developing countries since it reveals the extent of agreement or disagreement between independent household surveys in a developing country context.

We employ two distinct approaches to calculate labour incomes in CPHS. First, we combine employment and income data, and calculate labour incomes only for those individuals who report being employed. While this is the most straightforward approach, it suffers from some problems. One problem is that the definition of employment in CPHS is different from that in PLFS. The PLFS follows the standard NSO approach and considers any individual who engages in an income-generating activity for at least one hour of the past week as employed (Current Weekly Status). CPHS does not have a time criterion and instead only asks if the person was employed on the day of the survey or the day prior, or failing both, whether they expected to return to work in the near future. While definitional differences are to be expected across surveys, a more significant problem is that employment and income data are collected independently in CPHS, without reference to one another. This is unlike the NSO approach where income questions are fielded only to those who report being employed. Lastly, the structure of the CPHS data forces us to combine same-month recall data for employment with 4-month recall for income, in order to obtain both for the same month. As we discuss below, these factors make the CPHS income data potentially noisier when combined with employment information. For example, 15 percent of employed individuals report zero incomes, while 7 percent of individuals with positive incomes are not employed.

To circumvent this problem, our second approach uses only the CPHS income module to identify employed workers in the survey. That is, an individual is considered employed if they report earning a non-zero labour income. We provide estimates for three different definitions—a person is considered employed (and thus part of the estimate) if they report earning an income for at least six months, at least four months or at least one month of the preceding year. This method is simpler to execute and allows the user to make full use of the income data (all 12 months as opposed to only 3 months for which employment data are available). But it suffers from two limitations. First, it is not a conventional way to define employment. Second, these data cannot be used for any purpose that also requires information on the type of employment or occupation.

Our main findings are as follows. At the all-India level, mean labour income in 2018–19 as per the PLFS was INR 11,233 per month.[1] Using the employment-income linked approach, the CPHS estimate for 2018–19 stood at INR 14,531 with zero incomes included and at INR 17,354 without zero incomes. That is, monthly labour income as per CPHS was 29 percent higher than that reported in PLFS (55%

[1] All estimates for PLFS and CPHS are weighted population estimates. All values are in Indian National Rupees.

if we drop zero incomes). Under the six months income approach, the estimates come closer, with monthly labour income being estimated at INR 13,547 for CPHS, i.e. 19 percent higher than the PLFS estimate.

In the employment-linked approach, income inequality is higher in CPHS, as expected, if zero income values are included. The Gini is 0.52 in CPHS as compared to 0.44 in PLFS. If we exclude zero incomes, the Ginis come very close to one another (0.42 in CPHS, 0.44 in PLFS), but the means move further apart since the bottom decile in CPHS, which was zero, is now higher than the PLFS bottom decile. To circumvent the problem of zero incomes in CPHS, we also measure the p90/p50 ratio. This ratio is 3.1 for CPHS with zero incomes, 2.9 without them, and 2.7 for PLFS at the all-India level. By the six-month income approach, the story is flipped with the Gini falling to 0.37 in CPHS (and 0.44 as before in PLFS). The p90/p50 ratio is 2.5 as per CPHS, by this approach.

Across different methods and assumptions, as well as rural/urban locations, CPHS mean monthly labour earnings are anywhere between 5 percent (income definition for urban workers) and 50 percent (employment-linked approach, excluding zero incomes at all-India level) higher than corresponding PLFS estimates. The results presented here are consistent with the concern that CPHS misses poorer households compared to NSO surveys. But it is also possible that the differences arise from the way employment and income are captured differently in the two surveys. Or both may be the case. The current study cannot distinguish between these possibilities.

It is also worth noting that even though the estimates from the two surveys diverge significantly in proportionate terms, certain stylised facts regarding absolute levels of labour incomes in India are supported by both surveys. The 95th percentile in both PLFS and CPHS is less than INR 50,000 a month. That is, an individual earning INR 50,000 per month lies in the top 5 percent of the income distribution in India. Second, both PLFS and CPHS show that the median worker earns around INR 10,000 a month or less. Recall that the Expert Committee on Determining the Methodology for fixing the National Minimum Wage (Ministry of Labour and Employment 2019) proposed INR 375 per day (INR 9,375 per month) for rural areas and INR 430 per day (INR 10,750 per month) for urban areas as the minimum wage. Thus, half of the Indian workforce earns below the recommended National Minimum Wage, as per both surveys.

The remainder of the paper is organised as follows. Section 2 briefly describes the two data sources. Section 3 presents results from an approach that matches employment and income data within CPHS. Section 4 presents results for an alternative approach that uses only income data. Section 5 concludes.

2 Data Sources for Labour Incomes in India

Various surveys carried out by the NSO have traditionally been the principal sources of information on incomes in India. These include the Employment-Unemployment Survey (EUS) and its latest avatar, the PLFS. The last quinquennial round of the EUS was released in 2011–12. The EUS did not report incomes for the self-employed who constitute more than half the workforce. The new avatar of the EUS is the PLFS that reports labour incomes for salaried and wage workers as well as for the self-employed. and has been coming out annually since 2017–18. Most recent unit-level data are available for the fourth round (2020–21).[2] In addition, specialised surveys of agricultural households or informal firms are also periodically released and used by researchers to estimate income trends, inequality, and so on (Abraham 2019; Chakravorty et al. 2019). Since household surveys typically miss the top of the income distribution, NSO survey data have also been combined with income tax data to construct a more complete distribution of income for India (Chancel & Piketty 2019). In addition, household and individual incomes from labour and non-labour sources are available in the India Human Development Survey (IHDS) for 2004–05 and 2011–12. The third IHDS round is currently underway. To this existing landscape of nationally representative household surveys, the privately conducted CPHS was added in 2014.[3] This is an ongoing national panel survey of around 200,000 households, where each sample household is interviewed three times a year.

As of today, PLFS and CPHS constitute the two main regularly released sources of data on labour incomes in India. We describe the two in more detail below.

2.1 Periodic Labour Force Survey

The PLFS is the new, annual version of the quinquennial Employment-Unemployment surveys conducted until 2011–12 by the NSO. This is a repeated cross section for rural areas and a rotating panel for urban areas, covering just over 100,000 households across the country. PLFS collects income information for every household in the rural sample once a year and every urban one four times a year (once in each quarter). Data for four rounds have been released thus far (2017–18, 2018–19, 2019–2020, and 2020–21). Information is collected between July and June for each survey round. So, PLFS 2018–19 covers the period June 2018 to July 2019. It should be noted that while the earlier EUS was representative at the district level, PLFS estimates are not representative at the district level, but only at the NSS region level.

Following the NSO approach, PLFS defines employment in four different ways—usual principal activity status (UPAS), usual subsidiary activity status (USAS), current weekly activity status (CWS), and current daily activity status (CDS). Individuals are considered to be employed as a principal activity, if they report being employed for the major part (six months or more) of the preceding year. Individuals are classified as employed under USAS if they have worked for at least 30 days in the 365 days preceding the date of the survey. Notably, income information for all employment types is collected on the basis of an individual's CWS. If a person worked (or had employment even if they did not work due to illness, vacation, etc.) for at least one hour on at least one day during the seven days preceding the survey, they are considered to be employed under the CWS definition. Having decided

[2] In the interim (2015–16), the Labour Bureau Employment-Unemployment Survey reported incomes in categories.

[3] Both NSSO and IHDS survey data are available free of cost. CPHS is only available to subscribing customers.

the employment status on the basis of CWS, questions are then asked about their incomes.

Income is reported from three sources of employment—regular wage or salary, casual/daily wage, and self-employment. Income from regular wage employment is asked for the last month, and that from self-employment is asked for the last 30 days. Income from casual wage employment is asked for the past week. For income from casual/daily wage employment at the monthly level, we multiply weekly reported earnings by four to get their monthly equivalent. We then combine incomes of individuals from these three potential sources to get a value of labour income at the monthly level for each individual.

A special note must be made of two differences in the PLFS approach as compared to CPHS. First, in the PLFS, income questions are asked depending on the employment status of workers, i.e., conditional on a worker's employment. Thus, whenever there is a non-zero labour income for a worker, it indicates the worker is employed and conversely employed workers very rarely have zero incomes in this dataset.

Second, income coming from agricultural and other seasonal sources is difficult to measure on a monthly basis. As a result, while documenting incomes for workers in the agricultural sector, PLFS does not capture the last month's income (which could be zero even though the worker is employed). Instead, recognising that the value of the output has to be spread over the entire year, output is 'distributed over the entire production process based on past experience'.[4] This is to say, incomes are smoothed and reflect an average month's income, not necessarily last month's income.

2.2 Consumer Pyramids Household Survey

This is a panel survey in which each household is interviewed three times a year, in three "waves". Wave one runs from January to April, wave two from May to August, and wave three from September to December. Thus, a household that is interviewed in January will once again be interviewed in May and then a third time in September. Running since 2014, the panel has around 2 lakh households (9 lakh individuals) across the country. The CPHS does not drop households deliberately. In case of attrition, new households are added to maintain sample size. In 2016, questions on employment status were added, and in 2018, more details on sector and type of occupation were included.[5]

To make the time period covered by CPHS comparable to that in PLFS while not cutting across waves, we use data from the last wave of 2018 (September–December 2018) and the first two waves of 2019 (January–August 2019), covering income data between September 2018 and August 2019.

[4] Government of India (2016), p. A-24.

[5] Details are available at the official site: https://consumerpyramidsdx.cmie.com/. More detailed information on CPHS modules is also available in State of Working India (2021).

The CPHS definition of employment differs substantially from the NSO method. An individual is considered employed if they worked or had work on the day of the survey or the day prior, or were not working but returning to work in the near future (e.g. sickness or holiday). The respondent is asked regarding incomes earned from various sources in the past four months (since the last interview) for each member of the household regardless of whether they are employed. So, with households being interviewed ideally three times in a year, once in each wave, they are asked about their incomes for each month, but with a lag of one, two, three, or four months. Thus, unlike the PLFS, in CPHS income and employment data are collected without reference to one another. Employment information is collected in the People of India module; income information is collected as part of the Income module. We discuss the implications of this for estimating labour incomes in Section 3.1.

The survey collects income information at the household level as well as for every individual in the household at the monthly level. Different sources of income such as wage and salary, business, rent, pension, and dividends are reported. Incomes that are not accruing to specific individuals are reported collectively at the household level. Wage and salary income is reported for every individual in the member income module of CPHS. Business income is available at the household level in the household income module of CPHS. Since the present study is concerned with labour incomes, we exclude income from rent, pensions and dividends from our analysis. Business income is discussed separately in Sections 3.2 and 4.2.

Three points are worth noting. First, CPHS does not smooth incomes. The reported income for every month is directly recorded. Second, employment and income information for the same month can be obtained, but only as a cost of different recall periods (same month for employment, four month for income). Third, employment and income information are asked for each member independently with no reference to each other.

3 Comparing Labour Incomes in CPHS and PLFS using Employment Information

3.1 Combining Employment and Income Data in CPHS

Traditionally, in surveys conducted by the NSO, such as the EUS (till 2011–12) or the more recent PLFS, labour incomes are only asked to individuals who report themselves as employed, as per the CWS definition. The comparable approach in CPHS is then to use the employment status of individuals as reported in the People of India (POI) module and match this with their incomes reported in the Income module. Setting aside the question of comparability of employment definitions (for which see Abraham & Shrivastava 2022), this exercise still has some limitations.

To begin with, owing to the fact that income information for individuals is available for every month in CPHS while employment information is available for only three points in time in a year, on the day of the interview in each wave, matching of income with employment status leads to a drop in the number of observations for every individual from potentially 12 to a maximum of 3. But since usual labour datasets such as PLFS only give one observation per year, this is not, in itself, a large problem.

A more substantive issue is that in CPHS, income and employment data for any given month are asked at different times with no reference to each other. Thus, even if one works with a perfectly matched income-employment sample for a maximum of three times in a year per individual, the fact remains that two noisy data generating processes are being combined when we condition income statistics on employment status in CPHS.

In addition, there is a problem of difference in recall periods. For example, for an individual who was interviewed in the month of May, their employment status is available with a one day recall period. But the income information they provide during this interview pertains to the previous four months (January, February, March, and April). If we want income information for May (i.e., the same month as the employment information), we need to access the same person's interview conducted in the next wave, in September (when they provide income for the months of May, June, July, and August). So in a matched sample, employment is available with a one day lag, while income is available with a four-month lag.

Finally, as has been shown by Abraham and Shrivastava (2022), the definition of employment in CPHS leaves out some individuals who are actually employed. In keeping with their findings, we see that 7 percent of individuals in the working age population reported as 'not working' in CPHS have positive incomes. This possibility does not exist in PLFS for reasons mentioned above.

A second indication of the noise in the data generation process is that there is a substantial fraction of zero incomes in CPHS for employed individuals (14%). The corresponding number in PLFS is only 0.75 percent. One simple explanation for zero incomes reported by employed individuals is that unlike PLFS, CPHS does not smooth seasonal incomes. But this can explain the occurrence of zero values mainly for self-employed workers in agriculture who may have "lumpy" income streams. Indeed, as much as 31 percent of monthly incomes for individuals who report being self-employed in agriculture are zero. As expected, this falls significantly when we average monthly incomes for an individual (but still remains at 10%). More surprisingly, however, CPHS also reports a large percentage of zero incomes for those in salaried or daily wage jobs (11% falling to 4% on averaging). The corresponding number for PLFS is less than 1 percent. It is not clear why one in ten wage workers should report no income in the same month in which they say they had wage employment. Perhaps the recall period difference alluded to earlier is partly responsible for this.

The high prevalence of zero incomes for employed individuals presents a significant problem in analysing labour incomes in CPHS. Note that this problem does not exist if we use household-level income data—i.e., incomes of all individuals from all sources summed together.

3.2 Treatment of Self-employment Income

Since the present paper is concerned with labour incomes and nearly half of the Indian labour force is self-employed, we need to contend with the question of mixed income (capital and labour) obtained from self-employment. In PLFS, the vast majority of self-employed workers are either own-account workers or unpaid family workers. Employers constitute less than 2 percent of the workforce. Hence, it is reasonable to consider self-employment income as principally labour income.

In CPHS, there are two key differences. First, self-employed workers are not distinguished as own-account workers, unpaid family workers and employers. Second, income from self-employment is reported at the household level as "business income" and not attributed to any individual. Given this, we have the following options. Our main approach is to take household-level business income and distribute it among individuals. In this scenario, all business income is treated as labour income.[6] We distribute business income among members of the household using the variable on employment which identifies workers as either salaried, casual, or self-employed. Business income is distributed among all the self-employed members of the household.[7]

As a robustness check, we also present some estimates from a second approach. This is to include business income only for those households where self-employed individuals are running microenterprises where such income can justifiably be considered largely as labour income. For this, we need information that allows us to distinguish different kinds of self-employment. Some information is available in a variable called "Nature of occupation". We look at the nature of occupations for the individuals in households reporting a positive business income. In such households, the most frequently reported occupation is "organised farmer" (37%), followed by "businessman" (24%), and "small farmers" (11%). The remaining occupation types constitute less than 10 percent each. The descriptions of these occupations available on the CPHS website suggest that the category "businessman" comes closest to what the PLFS would call "employer" (NSS activity status 12).[8] For all households reporting a business income and with at least one individual's occupation listed as "businessman" who also reports a positive wage income, we exclude business income from the calculation of labour income.[9]

[6] That is, as in PLFS, we assume that "business income" in the vast majority of cases is income from self-employment in microenterprises, and as such is largely labour income and not capital income.

[7] Note that, while treating all self-employed income as labour income and ignoring returns to capital could be considered a problematic assumption, we are making the same assumption for both data sources. Hence, the comparison of estimates from both sources remains valid.

[8] The descriptions are available after free registration at this site: https://consumerpyramidsdx.cmie.com/kommon/bin/sr.php?kall=wkbquest&id=1697&k=master_cue. The occupational categories are: Self-Employed Entrepreneur, Wage Labourer, Small Farmer, Organised Farmer, Industrial Workers, Businessman, White-Collar Professional Employees and Other Employees, Support Staff, Non-Industrial Technical Employee, Agricultural Labourer, White Collar Clerical Employees, Small Trader/Hawker/Businessman without Fixed Premises, Home-based Worker, Qualified Self Employed Professionals, Manager, Legislator/Social Worker/Activists. In addition there are out-of-labour-force categories such as Unoccupied, Non-schooling child, Student, Homemaker and Retired/Aged.

[9] The mean (median) household business income for those with occupation category 'businessman' is INR 36,500 (INR 27,500).

3.3 Comparison of the Two Datasets

State of Working India (2021) provided some estimates for labour incomes from the two surveys (see Box 4.2 in the report). Here, the analysis is extended. The CPHS numbers are averages from pooled cross sections, i.e., each individual contributes up to three times to the average. Since we do not know to what extent zero values represent legitimate seasonal or other fluctuations in income, and to what extent they are spurious, we report means and median with and without zero incomes.

At the all-India level, estimated mean labour income in 2018–19 as per the PLFS was INR 11,233 per month (Table 1). The CPHS value stood at INR 14,531 with zero incomes included and at INR 17,354 without zero incomes. That is, monthly labour income as per CPHS was 29 percent higher than that reported in PLFS (55% if we drop zero incomes). The medians are somewhat closer at INR 8,038 for PLFS and INR 10,215 for CPHS including zero incomes (a difference of 27%) and INR 12,155 without them (a difference of 51%). The extent of discrepancy is higher for rural as compared to urban incomes, as can be expected due to differences in how seasonal incomes are recorded in both surveys (see Section 2).

Figure 1 gives the entire income distribution for both surveys at the rural, urban and all-India levels. The mean incomes at various quantiles as well as other comparisons for demographic and employment groups are available in Appendix Table S1. The CPHS averages are higher for the entire distribution with the exception of the bottom 10 percent, which is zero if zero incomes are included.

If business incomes are excluded for those households where an individual reports "businessman" as an occupation and also has a positive wage income, the CPHS average drops slightly to INR 14,424. The difference between the two

Table 1 Comparison of PLFS and CPHS labour incomes in 2018–19 with matched CPHS employment and income data

	Source	Mean (INR per month)	Median (INR per month)	CPHS mean to PLFS mean ratio	CPHS med to PLFS med ratio	Gini	p90/p50
All India	CPHS-zero	14,531	10,215	1.29	1.27	0.52	3.09
	CPHS-no zero	17,354	12,155	1.54	1.51	0.42	2.91
	PLFS	11,225	8,038	–	–	0.44	2.71
Rural	CPHS-zero	12,286	8,971	1.46	1.28	0.53	2.78
	CPHS- no zero	15,147	10,289	1.8	1.47	0.42	2.73
	PLFS	8,413	6.986	–	–	0.38	2.17
Urban	CPHS-zero	19,207	14,562	1.13	1.3	0.45	2.87
	CPHS-no zero	21,533	15,721	1.27	1.4	0.39	2.79
	PLFS	17,021	11,237	–	–	0.45	3.14

Source and notes: Author's calculations using CMIE-CPHS and PLFS. The CPHS numbers are averages from pooled cross sections, i.e., each individual contributes up to three times to the average. All values are in January 2020 real terms scenarios, viz., including business incomes as labour incomes versus excluding them for those households which are operating a larger business enterprise, is 1 percent or less for all-India, rural, and urban workforce. Thus, it is likely that profit incomes generally form a very small fraction of self-employment incomes in CPHS, as is also the case with PLFS.

One indication that the problem of zero incomes is connected to problems of type of employment as well as accurately measuring employment, comes from a comparison of incomes of female versus male workers. The bottom 25 percent of women workers report zero incomes in CPHS. That is, for a given month, a quarter of women workers reported to be employed on the day of the interview were reported to have zero earnings for that month (when the same household was interviewed four months later). But the rest of the distribution is similar to PLFS. Since women tend to be over-represented in unpaid family work (as helpers in own-account businesses for example) and in other part-time and irregular work, it is possible that they get recorded as being employed but do not earn an income on a regular basis. PLFS does not ask income questions to unpaid family workers. In part due to the higher prevalence of zero incomes among women workers, the female to male earnings ratio is much lower in CPHS (0.47 with zero, 0.55 without zero) as compared to PLFS (0.63).

As expected, income inequality is higher in CPHS if zero income values are included. The overall labour income Gini is 0.52 in CPHS and 0.44 in PLFS. If we exclude zero incomes, the Ginis come very close to one another (0.42 in CPHS, 0.44 in PLFS), but the means move further apart since the bottom decile in CPHS, which was zero, is now higher than the PLFS bottom decile. In such cases, a measure of inequality such as the p90/p50 ratio is a good option since it is less sensitive to the problem of zero incomes. This ratio is 3.1 for CPHS with zero incomes, 2.9 without them, and 2.7 for PLFS at the all-India level (Table 1). Thus, the two surveys are well-agreed on the extent of labour income inequality, by this measure.

The foregoing exercise shows that labour income estimates from the two surveys are systematically different from each other with the entire CPHS distribution significantly right-shifted compared to PLFS. One may be tempted to ask here whether one of the surveys is closer to "true" incomes. As noted earlier, recent criticisms of CPHS include the charge that it misses poorer households compared to NSO surveys. The results presented here are consistent with this claim. But since many other factors, such as definitions, survey method and so on, also differ across the two, based only on the work presented here, we are not in a position to take a firm stand on this question. Our modest aim is to acquaint readers with how these two independent sources stackup in relation to one another so that results based on one or the other can be interpreted appropriately. In this context, it is worth noting that even though a 30–50 percent discrepancy is large in proportionate terms and has significant implications for analysis of income poverty as well as inequality, certain basic stylised facts regarding labour incomes in India are supported by both surveys. We discuss these in Section 5.

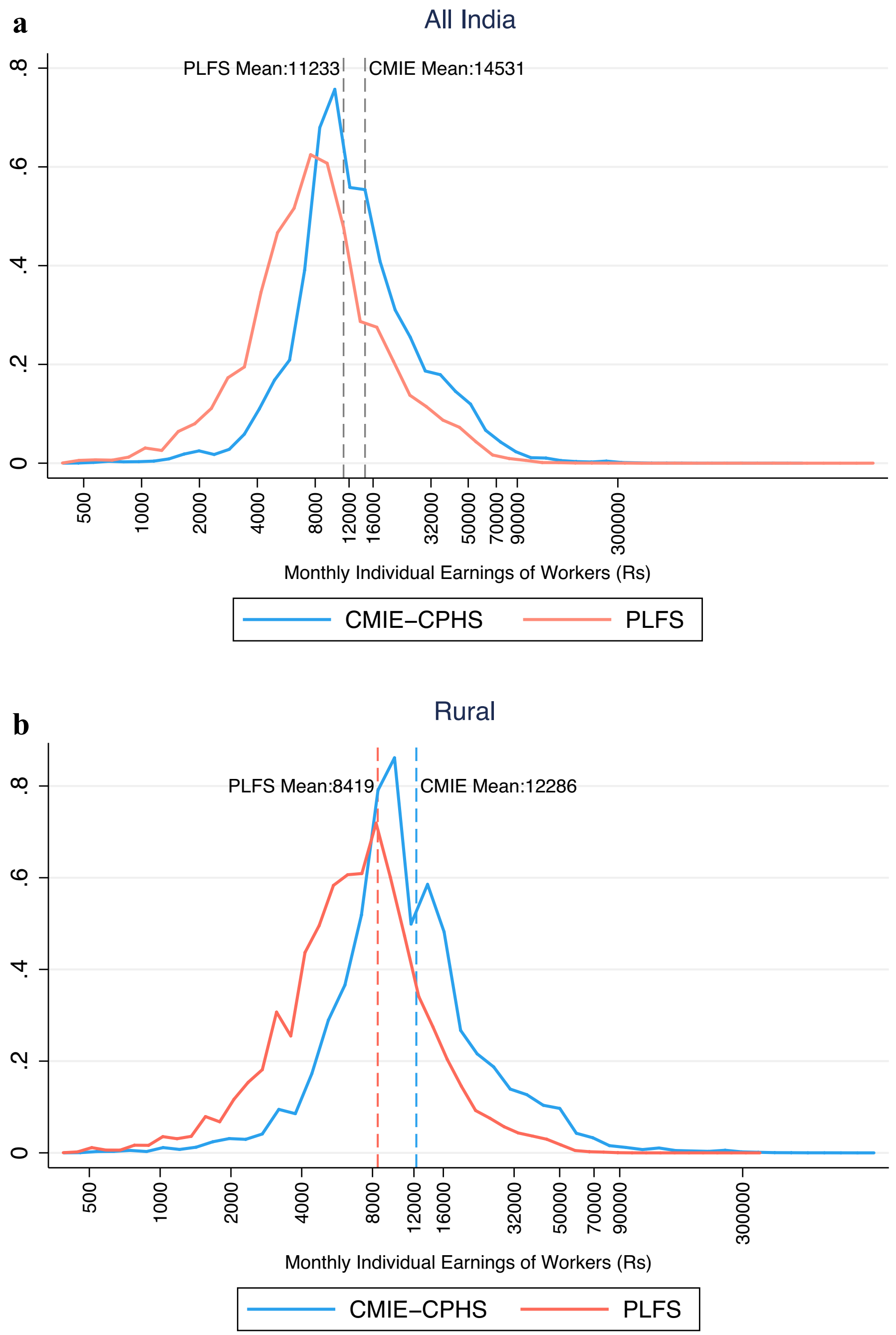

Fig. 1 Distribution of labour incomes as per PLFS and CMIE-CPHS for rural, urban and all-India in 2018–19 using combined employment-income approach. *Source and notes*: Author's calculations using CMIE-CPHS and PLFS. The CPHS numbers are averages from pooled cross sections, i.e., each individual contributes up to three times to the average. All values are in January 2020 real terms. The X-axis gives actual income values for the corresponding log value for ease of reference

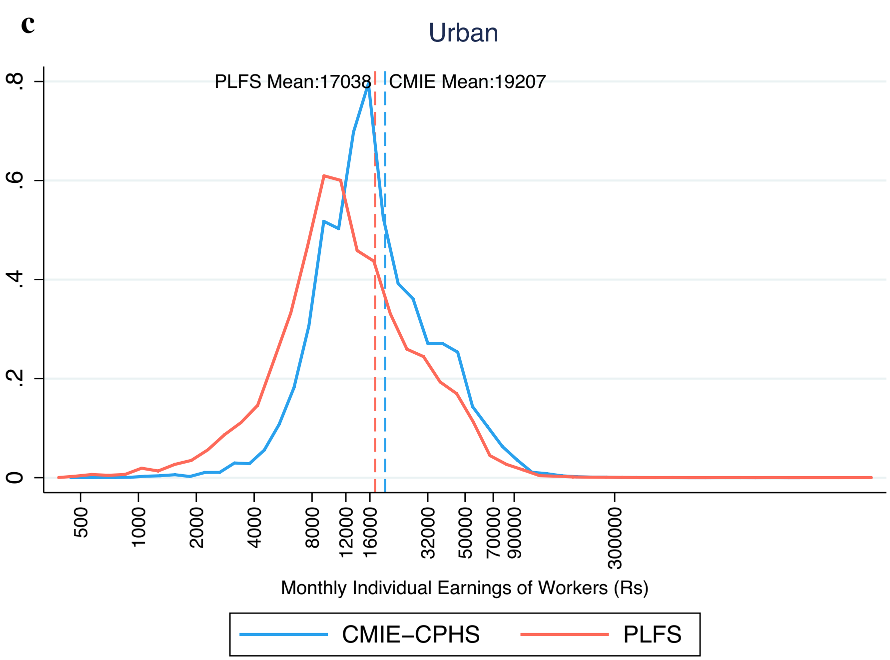

Fig. 1 (continued)

4 Comparing Labour Incomes in CPHS and PLFS without using Employment Information

Combining employment and income information has the obvious advantage that analysis requiring details on the type of employment becomes possible. However, combining these two modules comes at the cost of introducing noise in the income data by restricting it to only those three months for which employment data are available and forcing us to rely only on the four-month income recall (see Section 3). As we have seen, the result is a divergence, at times going up to 50 percent, between CPHS and PLFS estimates. For analysis of labour incomes that does not require employment information (such as inequality trends, or gender/caste/religion earnings gaps), we now propose another method that avoids combining employment and income information. This approach makes full use of the CPHS income data and achieves closer agreement to PLFS estimates.

4.1 Identification of Employed Workers using Income Information in CPHS

If we do not merge income data with employment status, the entire 12 months of income for an individual can potentially be used. But the problem that arises is, how do we know if a person is employed in a given month. At one extreme, equating employment with income, one could say that any individual with zero income in any month is unemployed or out of the labour force in that month. In this case, the employed sample consists of only those individuals who report positive incomes and their average monthly income is given by averaging their non-zero incomes in a year.[10] This method minimises errors of inclusion while maximizing the errors of exclusion. That is to say, it minimizes the possibility of including those without employment in the analysis, but maximizes the possibility of excluding employed persons (e.g. self-employed who are earning zero incomes in some months).

A more reasonable way to use income data in order to define employment is to adapt two definitions of employment from PLFS—usual principal and usual subsidiary status—in income terms. For the principal status definition, individuals earning a positive income for at least 6 months in a year can be considered employed. We can then calculate the average annual income for these individuals. Note that, in this approach, a zero income in a given month does not imply that the individual is not employed in that month. That is, we cannot average only the non-zero incomes to get an average monthly income for the employed workers, but rather we average all monthly incomes (zero and non-zero) as long as non-zero incomes are reported for at least six months.

The second approach is inspired by the subsidiary status definition in the PLFS. Those who have a positive income for at least one month in a year are analogous to those who have worked for at least 30 days in one year and hence are employed. Having identified the employed workers, we use their reported incomes for all months to get an average annual income and allot that as the labour income for these individuals. This method is likely to lower the average incomes of individuals—we elaborate on this while discussing the descriptive results in the next section.

One caveat is important to note here. As we have already discussed, for those who earn incomes at the end of a season or for some other reasons are paid once in several months, being employed in a month is not the same as earning an income in that month. Thus, the method outlined above will miss workers who have worked for six months or more, but report incomes for less than those number of months. The subsidiary status-inspired method on the other hand is too low a bar and is likely to yield low average incomes (since a lot of zeros potentially get averaged). Since income data are collected in PLFS conditional on being employed as per the CWS definition (and not UPAS or USAS), it is reasonable to suppose that workers employed in a subsidiary capacity are less likely to be captured in the income module.[11]

In order to give readers a sense of how sensitive incomes are to our definitions, in addition to the UPAS and USAS-inspired approaches we discussed above, we

[10] Note that this approach will not produce results identical to the "No zero income" approach in Section 3.3. This is because there we were forced to rely on only those months where employment data were available. Here, we would be averaging all non-zero incomes in a year.

[11] Since incomes are reported on the basis of CWS in PLFS, when we restrict our sample to workers employed under either UPAS or USAS, PLFS misses incomes for workers who are employed under UPAS but unemployed under CWS. Similarly for USAS. However, we find that the proportion of such workers in PLFS is small—5 percent of workers are employed under UPAS but not under CWS. It is the same for USAS. With the exception of this small percentage of workers, both PLFS and CPHS capture incomes for the same set of employed individuals under the two definitions of employment, respectively.

also present numbers for a four-month criterion (i.e. an individual is considered employed and is part of the sample is they report non-zero incomes for at least four months of the year).

While this method cannot be used to compare incomes for workers in different types of jobs, it allows us to compare incomes for the total sample of employed workers by various demographic characteristics, as will be discussed later.

4.2 Treatment of Self-employment Income

We distribute business income among members of the household using the following rule. If a household reports business income and there are some individuals in the working age group in the household reporting no wage income, we randomly select one individual and attribute the business income to that individual. The assumption here is that if a household reports business income and some members of the household earn wages while others do not, at least one person not earning wage income is likely to be involved in the family business. For those households who report a business income but where everyone in the household also has a wage income, we distribute business income across all individuals, since there is no way to identify which individuals are likely to have contributed to the business.[12] Note, these rules are followed under the sample where we work only with income data, and no employment information is available for individuals. Total labour income is then given by the sum of both wage income and business income for every member of the household.

As a robustness check, we have also tried leaving out business incomes entirely in the calculation of labour incomes. This approach makes the strong assumption that all business income is profit. This does not seem like a good assumption. But we provide some estimates based on this approach to give readers a sense of how sensitive the numbers are to such manipulations, as we take into account the two extreme possibilities of all business income being labour income on the one hand, and all business income being capital income on the other.

4.3 Using the Income-based Definition of Employment to Generate Basic Comparisons Between CPHS and PLFS

Under the UPAS-inspired definition of employment, the average all-India labour income is INR 11,350 as per PLFS and INR 13,547 as per CPHS (Table 2).[13] That is, the mean monthly income in CPHS is around 19 percent higher than that in PLFS. In general, the differences are smaller than what we obtained using

[12] Note that it is possible for more than one person in the household to be contributing to the family business, as either a paid or an unpaid worker. If the former is the case, our approach will inflate the average labour earnings. However, PLFS data reveal that generally only one person from a household is classified as own-account worker (employment status 11) with others being classified as unpaid family workers (status 12).

[13] The difference in PLFS numbers across the two approaches is due to small differences in sample construction.

Table 2 Comparison of PLFS and CPHS labour incomes in 2018–19 using only income data in CPHS to define employment

	Source	Mean (INR per month)	Median (INR per month)	CPHS mean to PLFS mean ratio	CPHS median to PLFS median ratio	Gini	p90/p50
All India	CPHS	13,547	10,289	1.19	1.27	0.37	2.5
	PLFS	11,358	8112	–	–	0.44	2.7
Rural	CPHS	10,849	8945	1.27	1.27	0.33	2.13
	PLFS	8520	7014	–	–	0.37	2.17
Urban	CPHS	17,992	13,636	1.05	1.18	0.38	2.58
	PLFS	17,172	11,553	–	–	0.45	3.1

Source and notes: Author's calculations using CMIE-CPHS and PLFS. In this approach, employment information is not used, rather a person is defined as employed if they report a positive income in at least six months of the previous year. CPHS reports self-employment income ("business income") at the household level. We attribute this income to one randomly chosen individual in the household. See text for details. All values are in January 2020 real terms

combined employment-income data. Figure 2 shows the complete income distributions from both surveys. Unlike with the first method that combined the employment and income data in CPHS, this approach does not suffer from the problem of zero incomes.

The average income in rural areas is around INR 10,849 in CPHS, 27 percent higher than the average in PLFS at INR 8500. Urban incomes match much more closely, with CPHS reporting a mean of INR 17,992 and PLFS of INR 17,172 (a discrepancy of just under 5%). The complete distributions are given in Fig. 2 and Appendix Table S2. The fact, that urban incomes match closely while rural incomes do not, is consistent with the differences in the method of capturing seasonal self-employment incomes in the two surveys.

Inequality in the income distribution as measured by the Gini coefficient is lower in CPHS for all categories. For the all-India sample, the Gini coefficient is 0.37 in CPHS and 0.44 in PLFS. Looking at earnings by gender, we find that incomes for male workers are significantly higher in CPHS, while those for women workers match quite closely. But the estimates from the two surveys are close enough for both sexes such that the gender earnings gaps are closer together than in our first approach—0.54 for CPHS and 0.65 for PLFS. Detailed results by caste, religion, age and education are available in Appendix Table S2.

If the analysis is done only using wage incomes and dropping business incomes entirely from the labour earnings of individuals, the CPHS average falls to INR 12,433. Note, this is an extreme situation, but it offers us a sense of the range of the distribution of average incomes in CPHS in the absence of employment information.

Finally, we offer some estimates of mean labour income at the all-India, rural and urban levels, under two more definitions of employment. These give a sense of how sensitive estimates are to such changes in definition. If we adopt the USAS-inspired definition of employment, the all-India monthly average drops to INR 9523 (Table 3). This is expected since individuals reporting a positive income for even one month get identified as employed in CPHS under this definition. The average monthly income for the employed individuals is calculated by taking an annual average of their incomes. The possibility of zero incomes in the remaining eleven months for such employed individuals implies that their average monthly income can get pulled down.

The intermediate approach that defines an individual as employed and hence, part of our sample if they report at least four months of non-zero income in CPHS yields results shown in Table 3. The all-India mean income for this sample is INR 11,032, around 18 percent lower than the mean under the six-month definition. Note that this comes very close to the PLFS estimate.

As we have seen, the income-based approach described above produces a closer agreement between the two surveys. These estimates go against the conclusion that CPHS will always produce higher averages. But this comes at the cost of departing from a more conventional approach that uses direct employment-related questions to identify employed individuals. Hence, we do not suggest that this method be the only one used to produce CPHS-based income estimates. Rather, deploying both approaches (employment-linked and only income-based) will give a more robust sense of the results in any particular application.

5 Discussion and Conclusion

The CPHS has been the most widely used dataset to document the impact of COVID-19 on various indicators of economic well-being. However, concerns have been raised about the representativeness of this survey, particularly with respect to missing poor and vulnerable households. To gain some insights into this issue as well as to improve our understanding of labour incomes in India, in this paper we have presented a comparison of income estimates from the CPHS with those obtained from the PLFS, for a pre-pandemic year (2018–19).

We discuss two approaches to generating labour income estimates from CPHS. First, using employment status as captured in the People of India module to define the sample of employed workers and the second using only the Income module, where earning an income itself is a criterion of employment. We also highlight some problems with matching income and employment data for individuals within CPHS. Our second method enables one to overcome those problems and still work with labour incomes. This approach works if the type of employment or occupation (data only available in the People of India module) is not necessary for analysis.

Our analysis shows that labour income estimates from the two datasets diverge to a greater extent if we use employment information within CPHS to select employed individuals (30–50%). The CPHS income distribution is systematically right shifted compared to the PLFS. On the other hand, if earning a positive income for at least six months of the year is taken to be the definition of employment, we achieve a closer agreement (all-India difference of 20%, with almost identical estimates for urban incomes) between the two datasets. Possibly a significant source of noise is removed when we do not combine employment with incomes.

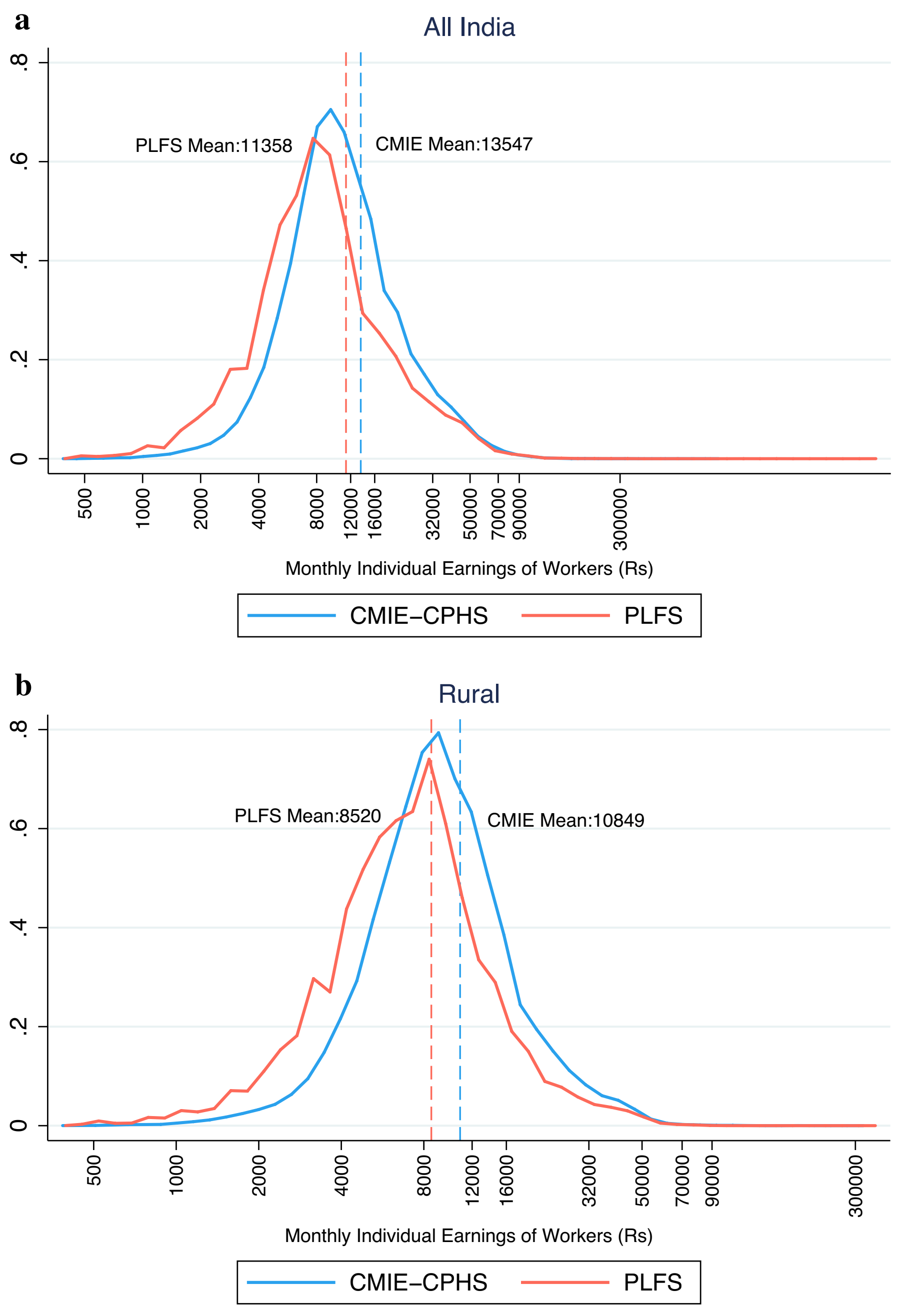

Fig. 2 Distribution of labour incomes as per PLFS and CMIE-CPHS for rural, urban and all-India, in 2018–19, using income-based (6 months) definition of employment. *Source and notes*: Author's calculations using CMIE-CPHS and PLFS. In this approach, employment information is not used, rather a person is defined as employed if they report a positive income in at least six months of the previous year. See text for details. All values are in January 2020 real terms. The X-axis gives actual income values for the corresponding log value for ease of reference

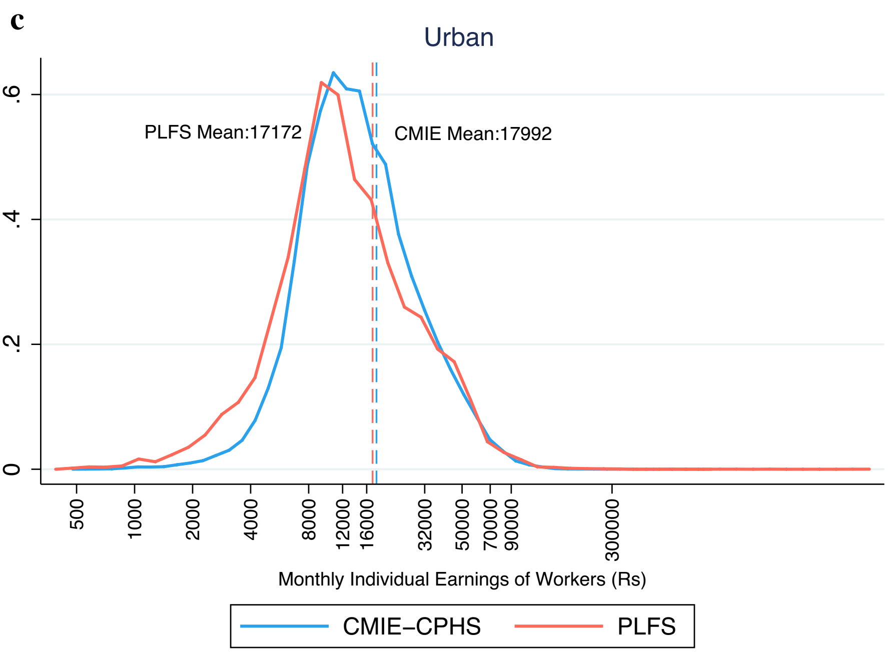

Fig. 2 (continued)

Table 3 Mean monthly labour incomes under different definitions of employment (INR per month)

Definition of Employment	All India	Rural	Urban
6-month definition	13,547	10,849	17,992
4-month definition	11,339	9078	15,095
1-month definition	9523	8089	11,962

Source and notes: Author's calculations using CMIE-CPHS and PLFS. We use CPHS income data to define a person as employed if they earn a positive income in at least six months, at least four months, or at least one month of the preceding year. See text for details. All values are in January 2020 real terms.

A more detailed analysis of the source of these differences is left for further research. But it is worth reiterating the relevant factors here. First, CPHS collects employment and income information for individuals independently, thereby introducing more noise in the process compared to the NSO approach of conditioning labour income on employment status. Second, the CPHS data structure entails combining very short period recall for employment (the day of the survey or the previous day) with long period recall for income (up to four months). Possibly due to these factors, a significant proportion of employed individuals in CPHS report zero incomes, even after averaging across months.

The two surveys diverge to a greater extent for rural and self-employed incomes as opposed to urban and salaried incomes. This could be explained by the fact that PLFS smooths seasonal incomes while CPHS does not.

Before concluding, it must be noted that, despite significant divergences in the two distributions, they still overlap substantially and certain stylised facts regarding income distribution in India are borne out by both data sources. For example, using any method, the 95th percentile for the all-India distribution, in both PLFS and CPHS, is less than INR 50,000 a month. That is, an individual earning INR 50,000 per month as income lies in the top 5 percent of the income distribution in India. Second, recall that the Expert Committee on Determining the Methodology for fixing the National Minimum Wage proposed a wage such that the expenditure on minimum recommended food intake, essential non-food items (namely clothing, fuel and light, house rent, education, medical, footwear, and transport) and other non-food items for the wage earner and their dependents can be met. The recommendation was INR 375 per day (INR 9375 per month) for rural areas and INR 430 per day (INR 10,750 per month) for urban areas as of July 2018. Both PLFS and CPHS show that the median worker earns just around this level or a little below. That is, fully half of the Indian workforce earns below the recommended national minimum wage, as per both surveys.

The foregoing comparative exercise has been carried out with the intention of providing a few different methods to calculate labour incomes using CPHS data. We also hope that the results presented will enable researchers and analysts to gain a more nuanced appreciation of these data and how they compare to the PLFS. In the process, we have also presented the broad distributional characteristics of the labour distribution in India, which we hope will be useful for policy design and analysis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41027-023-00427-8.

Acknowledgements This is a revised version of a working paper brought out by the Centre for Sustainable Employment, Azim Premji University (Jha and Basole 2022). We are thankful to Anand Shrivastava and Rosa Abraham for their comments. Authors are responsible for errors.

Funding No funding was received for this study.

Declarations

Conflict of interest The authors declare that they have no relevant financial or non-financial conflict of interest to disclose.

References

Abraham, R. 2019. Informal employment and the structure of wages in India: A review of trends. *Review of Income and Wealth* 65: S102–S122.

Abraham, R., A. Basole, and S. Kesar. 2022. Down and out? The gendered impact of the Covid-19 pandemic on India's labour market. *Economia Politica* 39 (1): 101–128.

Abraham, R., and Shrivastava, A. 2022. How Comparable are India's Labour Market Surveys, *Centre for Sustainable Employment Working Paper #45*, Azim Premji University, Bangalore.

Bhattacharya, S., and Sinha Roy, S. 2021. Intent to Implementation: Tracking India's Social Protection Response to Covid-19, *World Bank Social Protection and Jobs Working Paper*, 2107. https://openknowledge.worldbank.org/handle/10986/35746
Bussolo, M., Kotia, A., and Sharma, S. 2021. Workers at Risk: Panel Data Evidence on the COVID-19 Labor Market Crisis in India, *World Bank Policy Research Working Paper 9584*.
Chakravorty, S., S. Chandrasekhar, and K. Naraparaju. 2019. Land distribution, income generation and inequality in India's agricultural sector. *Review of Income and Wealth* 65: S182–S203.
Chancel, L., and T. Piketty. 2019. Indian income inequality, 1922–2015: From British Raj to Billionaire Raj? *Review of Income and Wealth* 65: S33–S62.
Deshpande, A. 2020. Early effects of lockdown in India: Gender gaps in job losses and domestic work. *The Indian Journal of Labour Economics* 63: 87–90. https://doi.org/10.1007/s41027-020-00261-2.
Deshpande, A. 2022. The Covid-19 pandemic and gendered division of paid work, domestic chores and leisure: Evidence from India's first wave. *Economia Politica* 39 (1): 75–100.
Dhingra S., and Ghatak, M. 2021, June 30. How has Covid-19 Affected India's Economy? *Economics Observatory*. Retrieved June 29, 2022 from https://www.economicsobservatory.com/how-has-covid-19-affected-indias-economy
Government of India. 2016. Instructions to Field Staff: Design, Concepts, Definitions and Procedures, Periodic Labour Force Survey, *National Statistical Office*, Government of India.
Gupta, A., Malani, A., and Woda, B. 2021. Inequality in India Declined During COVID. *National Bureau of Economic Research*, No. W29597.
Gupta, S., P. Seth, M. Abraham, and P. Pingali. 2022. COVID-19 and women's nutrition security: Panel data evidence from rural India. *Economia Politica* 39 (1): 157–184.
Jha, M., and Basole, A. 2022. Labour Incomes in India: A Comparison of PLFS and CMIE-CPHS Data. *Centre for Sustainable Employment Working Paper #46*, Azim Premji University, Bangalore.
Mohanan, P.C. 2021, July 6. Rebuilding India's Employment Statistics System. *Bloomberg Quint*. 2021. Retrieved June 28, 2022 from https://www.bloombergquint.com/economy-finance/jobs-data-rebuilding-indias-employment-statistics-system
Pais, Jesim, and Vikas Rawal. 2021, August 10. CMIE's Consumer Pyramids Household Surveys: An Assessment. *The India Forum*. Retrieved June 28, 2022 from https://www.theindiaforum.in/article/cmies-consumer-pyramids-household-surveys-assessment
Somanchi, Anmol. 2021. Missing the Poor, Big Time: A Critical Assessment of the Consumer Pyramids Household Survey. *SocArXiv*. https://osf.io/preprints/socarxiv/qmce9/
Vyas, M. 2020. Impact of lockdown on labour in India. *The Indian Journal of Labour Economics* 63: 73–77. https://doi.org/10.1007/s41027-020-00259-w.
Vyas Mahesh. 2021a. There are Practical Limitations in CMIE's CPHS Sampling, But No Bias, *The Economic Times*. Retrieved June 28, 2022 from https://economictimes.indiatimes.com/opinion/et-commentary/view-there-are-practical-limitations-in-cmies-cphs-sampling-but-no-bias/articleshow/83788605.cms
Vyas Mahesh. 2021b. Consumer Pyramids Household Survey: A Response and A Rejoinder. *The India Forum*. Retrieved June 28, 2022 from https://www.theindiaforum.in/letters/consumer-pyramids-household-survey-response-pais-and-rawal?utm_source=website&utm_medium=organic&utm_campaign=Letters&utm_content=Homepage