ARTICLE

How Comparable are India's Labour Market Surveys?

Rosa Abraham[1] · **Anand Shrivastava**[1]

Abstract

The Centre for Monitoring Indian Economy's (CMIE) Consumer Pyramid Household surveys have emerged as an important source of regular labour market data for India. Given the differences in methods in data collection between the CMIE and official employment sources, it becomes exceedingly important to establish some comparability between the government and the CMIE datasets. With the release of the official Periodic Labour Force Surveys for 2017–18, we now have an overlap between the official datasets and CMIE datasets. In this paper, we examine the extent of comparability of labour force estimates from these two datasets. We find that employment estimates for men are broadly comparable. However, for women, there is a consistent divergence, with CMIE estimates of women's workforce participation lower than that of NSS-PLFS. We find that irrespective of the reference period used in the PLFS estimation of employment statuses, there is no convergence with the CMIE employment estimate for women's employment. Moreover, the mismatch in CMIE-PLFS estimates occurs across all types of women's employment and irrespective of what reference period of employment (in official data) is used.

Keywords Labour statistics · Women's employment · Measurement of work · PLFS · CMIE

JEL Classification J16 · J21 · J00

1 Introduction

Regular, frequent and real-time labour market data is crucial to understand the status of an economy. Such data becomes all the more important during the time of an economic shock. With the COVID-19 crisis and the consequent economic lockdowns imposed across the country, data on employment and other labour market impacts during this time has become crucial for researchers and policymakers alike.

✉ Rosa Abraham
   rosa.abraham@apu.edu.in

1 Azim Premji University, Bengaluru, India

ISLE 🌀 Springer

The Centre for Monitoring Indian Economy's (CMIE) Consumer Pyramids House-hold Survey (CMIE-CPHS henceforth) dataset is the only nationally representative survey that was carried out prior to the lockdown, during the lockdown (partially, as a phone survey), as well as in its aftermath. It has emerged as an important source of evidence on the impacts of the crisis on employment and labour force participation. It is therefore pertinent to examine to what extent CMIE estimates are comparable with the National Sample Survey estimates.

Several questions have been raised on CMIE's household survey including the sampling strategy being used, representativeness as well as unusual trends in cer-tain outcomes over time. Somanchi and Dreze (2021) find huge variations in cer-tain measures such as literacy rate and asset ownership between the CMIE and other national statistics such as the National Family Health Survey (NFHS). The sampling strategy of beginning at the main street, according to Dreze and Somanchi, may result in a biased sample with poor households being underrepresented. Vyas (2021) in his response clarifies that although the sampling starts at the main street, the size of the sample mandates that enumerators move beyond just the main street and into the outskirts. Somanchi (2021) finds that besides literacy and asset ownership, there are significant differences between CMIE and other nationally representative datasets including in age distribution of the population, sex ratio and consumption expenditure.

In this paper, we focus on the comparability of employment-specific estimates from CMIE with those from official labour force surveys. Therefore, the question is not about the population at large, but rather on the labour force and how the two surveys compare in terms of different labour market estimates.

Specifically, in this paper we attempt to establish comparability between the NSS-PLFS and CMIE-CPHS employment estimates. With the release of unit-level data from the National Sample Survey's (NSS) Periodic Labour Force Survey for 2017–18, we now have data that directly overlap with each other from the two dif-ferent data sources, i.e. NSS and CMIE. Ex ante, one would expect some difference in estimates as the definition of employment as well as the method of conducting the survey, are different in CMIE-CPHS as compared to the government surveys. We discuss these differences in methods in detail in a later section.

Table 1 provides the estimates of labour force participation rate (LFPR), work-force participation rate (WPR) and unemployment rate (UR) overall and for men and women for four surveys—the 2015–16 Labour Bureau Employment Unemployment Survey (2015–16 LB), the three waves each of CMIE-CPHS from 2016 and 2017 (2016 CMIE and 2017 CMIE), the 2017–18 round of NSS-PLFS (2017–18 PLFS) and the three waves of CMIE-CPHS from 2017 (2017 CMIE). In terms of LFPR and WPR, there is a broad alignment between the CMIE and official statistics for men. For estimates of female LFPR and WPR, however, we see a much larger divergence between the two estimates. Women's WPR as per official statistics is twice that of the estimates from CMIE.

All four surveys claim national representativeness with the use of appropriate weights. One of the potential sources of difference in estimates could be definitional differences in the concept of employment being adopted including differences in how questions are asked, and the reference period used. In this paper, we investigate

Table 1 Comparing labour force aggregates from CMIE and official estimates

	LFPR			WPR			UR		
	Overall	Female	Male	Overall	Female	Male	Overall	Female	Male
2015–16 LB	52.4	27.4	75.5	49.9	25.4	72.6	3.7	5.8	2.9
2016 CMIE	46.8	15.6	74.8	43.0	12.1	70.7	8.2	22.4	5.5
2017–18 PLFS	49.8	23.3	75.8	46.8	21.9	71.0	6.0	5.6	6.1
2017 CMIE	44.5	13.0	72.7	41.3	10.2	69.3	7.3	21.4	5.0

All estimates are for individuals 15 years and above. Estimates from LB and PLFS are for Usual Principal and Subsidiary Activity (UPSS). The CMIE samples refer to January-December calendar year, PLFS to a July-June year, and the Labour Bureau survey was conducted between April-December

these differences using two methods to compare between the 2017–18 NSS-PLFS and 2017 CMIE-CPHS.

First, since one of the most important uses of individual-level employment data is to model the correlates of employment status, we see how similarly a given model predicts employment status across the two different datasets. We first fit a model of correlates of employment status on CMIE-CPHS 2017, and then use the model to generate predictions in PLFS 2017. We find that the model correctly predicts around 80% of the observations in PLFS which is the same rate of success that the model has in the original data. This shows that a model of individual employment status behaves very similarly in CMIE-CPHS as it does in PLFS. We do, however, see a systematic divergence of predictions in the case of women, far more than among men. We explore how different predictions vary across other subgroups within the male and female population to identify if there are particular groups that are systematically being assigned a different employment status across the two surveys.

Second, we see how closely state-level estimates of Labour Force Participation Rate (LFPR) and Workforce Participation Rate (WPR) from CMIE-CPHS 2017 compare with those obtained from PLFS 2017. We find that while for men the estimates map quite well, implying that the bias and the variance are low, the same is not the case for women. This implies that the effect of the difference in the definition of employment and/or the method of surveying is largely observed in the responses for women.

Given the focus on non-official data sources like the CMIE in understanding the Indian labour market, this paper contributes to this discussion in two ways. First, it provides a useful and simple framework for comparing across two distinct datasets measuring similar outcomes. Second, in the particular context of the CMIE, it highlights the extent of overlap and mismatch between official and this private data source which we believe will be useful for researchers working with CMIE's data.

The rest of the paper is organised as follows. The next section describes the different definitions of employment used in the two surveys. Section 3 briefly describes the data construction process. The fourth section presents the methodology and the results of comparing the two surveys using a model to predict employment status. Section 5 presents a comparison of state-level employment estimates from the two surveys and Section 6 concludes.

2 Labour Market Surveys and Definitions of Employment Used

From 2017–18, the PLFS replaced the Labour Bureau's Employment and Unemployment surveys (EUS) which had replaced the earlier quinquennial NSS-EUS surveys. The PLFS 2017–18 covered 1,02,113 households and 4,33,339 individuals and is nationally representative. The survey is conducted annually. Since 2017 there have been three more rounds of PLFS in 2018–19, 2019–20 and 2020–21.

The Centre for Monitoring Indian Economy (CMIE), a private business information organisation, has been collecting data relating to employment and unemployment status since 2016. The Consumer Pyramid Household Survey as it is called (henceforth referred to as CMIE-CPHS) covers about 1,60,000 households and 5,22,000 individuals. The survey is conducted in three 'waves' with each wave spanning four months, beginning from January. Each individual is surveyed in every wave, so that for every year, their employment and unemployment status is available for three points in the year. Alongside employment status, the survey collects information on gender and age of the respondent and other demographic characteristics. While CMIE-CPHS is a household survey, the enumerator does not go to the household with a questionnaire to be filled point by point. Instead, the enumerator has a free-ranging discussion with the household head where all the questions from the survey are woven in. Additionally, the questions used to discern employment status differ between the government surveys and the CMIE survey, in particular in terms of the reference period used.

In the CMIE-CPHS, employment status is discerned from one question. This categorises an individual into one of four possible statuses, i.e. (i) Employed, (ii) Unemployed, willing and looking for a job, (iii) Unemployed, willing but not looking for a job, (iv) Unemployed, not willing and not looking for a job. The CMIE identifies an individual as employed if he/she 'is engaged in any economic activity either on the day of the survey or on the day preceding the survey, or is generally regularly engaged in an economic activity'. Individuals who were in some form of employment but were not at work on that day of the survey due to various reasons such as illness, leave or holiday are still considered as employed when there is a reasonable surety of them going back to work. Therefore, the reference period used by the CMIE-CPHS refers to the day of the survey, or the previous day, while also allowing for a broader period, in the case of those people who may not have worked on those specific days of the survey, but have employment, in general.

The NSS-EUS schedule uses four different reference periods (and hence different questions) to arrive at four possible activity statuses for an individual. These reference periods are one year, one month, one week and each day of the reference week. The Usual Principal Activity status (UPA) identifies a person as in the labour force if they have spent the majority of their time in the 365 days preceding the survey either working or looking for work. If a person is Not Employed nor looking for work for the majority of the year but working for at least a month in the 365-day reference period (i.e. subsidiary status), then they are identified as employed as per Usual Subsidiary Activity status (USA). The Usual Principal and

Subsidiary Status (UPSS) is a combined definition where a person is considered to be in the labour force if they were employed either in the UPA or the USA definition.

Under the Current Weekly Status (CWS), a person is identified as working if they worked for at least an hour during the seven days preceding the survey. Finally, the Current Daily Status uses the day as the unit of analysis. A person's activity status on each day of the reference week determines the Current Daily Status (CDS). Under CDS, a person is considered as working, if they actually worked for at least one hour in the day or had work for one hour but did not do the work.[1]

Firstly, by identifying an individual's status as on the day of the survey, or on the day preceding it, at first glance, the CMIE definition seems to be closest to the NSS' CDS interpretation of employment. Secondly, by allowing for individuals who are 'generally regularly employed' to be also identified as employed, the CMIE interpretation of employment is similar to the NSS UPA/UPSS approach. Therefore, if we were trying to find the extent of definitional alignment between CMIE and NSS, we can conclude that there is no one interpretation of employment that the NSS/LB uses that is perfectly equivalent to the CMIE-CPHS definition. Instead, the CMIE definition of employment is, in a sense, a combination of two or more definitions of employment as identified under NSS. In the next section, we look at how employment, as identified under each of the existing NSS-EUS measures, compares with CMIE. We also construct a measure within PLFS that closely approximates the one used in CMIE and do the similar comparison.

3 Data

The CMIE-CPHS dataset is conducted in waves with each wave spanning four months. The latest wave for which data is available is January to April 2022. In order to compare with PLFS 2017–18, we consider data for the year 2017–18 for the same months for which the PLFS survey was conducted, i.e. July 2017 to June 2018. Since the CMIE-CPHS interviews households thrice in a year, for every individual we have employment status at three times in a year. The PLFS, on the other hand, is conducted across the year, with any individual being interviewed only once at some point in the year.

To make the CMIE data as closely comparable to the PLFS in terms of sample design, we take only the months of July, August from the second wave, all four months of the third wave and the first wave of the subsequent year, and only the months of May and June from the second wave of that year. Therefore, we have a June–July sample, identical to the PLFS time period.[2]

---

[1] This is a departure from the earlier CDS measurement under NSS−EU surveys. There, a person was considered as working a "full day" if they were engaged in work for four hours or more, or as working "half day" if they worked less than four hours but at least one hour.

[2] This manner of construction of the data breaks a wave in CMIE. However, we have also done the analysis using a January−December sample of CMIE, against the standard PLFS sample and our findings are broadly unchanged.

Since the CMIE would have information on an individual's activity status at three points of time (unlike the PLFS which has information only at a single point of time), we further modify this data. We construct a pseudo-cross section simulating a sampling scheme where an individual would be randomly allocated to one of the waves. This was done by randomly choosing and retaining only one of the two or three possible employment statuses for an individual in a year. All estimates for CMIE-CPHS are derived from this pseudo-cross section dataset.[3]

4 Comparing Surveys by Fitting a Model of Employment Status

Since any two surveys will always interview different individuals, the feasible way to compare between the two is to look at individuals with similar demographic characteristics in the two surveys and see if the employment status assigned to these individuals by the two surveys are similar, on an average. The assumption here is that if one of the surveys systematically differ in the classification of the employment status of some demographic subsection of society, then this would reflect as a difference in the relationship between the employment status variable and those demographic variables across the two surveys.

We model the relationship between employment and demographic variables by constructing an individual-level econometric model using one of the datasets. The dependent variable in this econometric model is the individual's employment status and the independent variables are demographic characteristics of the individual. This estimated model is then applied on the second dataset to see to what extent the same model can accurately predict the employment status in this dataset. This allows us to establish to what extent the surveys diverge in their attribution of employment status to an individual, and if there are any systematic patterns in this divergence, i.e. identify if there are certain kinds of individuals or subgroups more likely to be differentially classified between the two surveys. Therefore, given the difference in definitions used across surveys, we wanted to see whether a person identified as employed by the CMIE-CPHS definition would be similarly identified in the NSS-PLFS. We do this exercise for all definitions used under the NSS-PLFS, i.e. UPA/UPSS/CWS. We also construct a definition in PLFS that closely approximates that of the CMIE-CPHS. Since the CDS is a measure of person day, we modified it so as to approximate the CMIE definition. CDS, as used here, identifies an individual as employed if they reported as working on the day of the survey, or the day prior to the survey, or for the majority of the year (i.e. by UPA status). This definition, therefore, uses information from the UPA status and from daily activity status to arrive at an employment categorisation for an individual that is closest definitionally to what CMIE uses.

---

[3] Our results do not change substantially if we use a pooled cross section, i.e. a dataset with all three observations available for an individual, but with each treated as a separate entity, instead of a pseudo–cross section.

4.1 A model of Employment Status

Constructing an econometric model of demographic correlates of employment status is also instructive because one of the most important uses of individual-level employment data is to study the factors affecting labour market outcomes (Kingdon & Unni, 2001; Klasen & Pieters, 2012; Srivastava & Srivastava, 2010). Establishing the comparability of such a model across datasets would enable researchers to map the changes in the effects of determinants like age and gender across time.

The approach we take is to first estimate a multinomial logit model of activity status on CMIE-CPHS and then use the model to predict outcomes in NSS-PLFS 2017. We then see to what extent the estimated predictions from this model (developed on CMIE-CPHS) align with the actual statuses as in NSS-PLFS.

The model that we estimate is given below. The dependent variable takes one of three of values representing an individual's activity status—Employed, Unemployed and Out of the Labour Force (OOLF). We classify the fourth category in CMIE—unemployed and willing but not looking for work—as out of the labour force, and the unemployed includes those willing and looking for a job as well as those willing but not looking for a job.

$$\frac{P(y_{ij} = k)}{P(y_{ij} = 0)} = exp(\alpha + \beta_{1k}age_{ij} + \beta_{2k}age_{ij}^2 + \beta_{3k}education\,status_{ij}$$
$$+ \beta_{4k}childphh_{ij} + \gamma_{jk} + \varepsilon_{ijk}),$$
$$k \in \{1, 2\}$$

Here, the subscript i indicates an individual and j refers to the state that they belong to. The employment status of an individual is denoted by y_{ij}. It can be 0, 1 or 2 indicating OOLF, Unemployed and Employed, respectively. The variable *age* is a continuous variable indicating age in years. We also include a squared age term to incorporate nonlinear relation between employment and age, i.e. while older individuals are more likely to be employed compared to younger, beyond a certain age, this relation may no longer hold as older workers withdraw from the labour market. E*ducation status* is a categorical variable representing levels of education, going from illiterate, primary, middle, secondary, higher secondary, to graduate and above. *Childphh* captures the number of children per adult member in the household. γ_{jk} indicates state fixed effects. These variables are standard individual-specific demographic features that are likely to affect employment decisions. Given that the intention of this model is to compare across similar individuals, these characteristics are pertinent to this estimation. We do not include demand-side variables. Other variables like household consumption expenditure are likely to be correlated with employment outcomes, and hence these are not incorporated.

The model is estimated separately for men and women, as well as rural and urban areas, to allow for the relation between these covariates and employment status to vary by gender and region. The model is first estimated on CMIE-CPHS, and the estimated coefficients are then used to predict the employment status in NSS-PLFS. Being a multinomial model, every individual will have three probabilities

Table 2 Distribution of prediction matches, overall

	CMIE–CMIE	CMIE–PLFS	PLFS–PLFS	PLFS–CMIE
Matched	86.39	80.19	81.67	84.07
Employment overpredicted	0.99	1.52	1.67	0.2
Employment Underpredicted	0.02	0.03	0.06	0.03
LFP overpredicted	3.54	2.96	4.86	6.04
LFP underpredicted	9.07	15.3	11.76	9.66

representing the likelihood of that individual being in each of the three employment outcomes. The status that has the highest predicted probability is chosen as the final predicted employment status for that individual. These predictions are then compared with the actual observed employment status. The Usual Principal and Subsidiary Status (UPSS) definition of employment is used for the baseline exercise. However, this is later extended to all other definitions available in NSS-PLFS to see how well these compare with the CMIE-CPHS model-based predictions.

An observation is identified as 'Matched' if the predicted employment status for that individual is the same as their actual employment status. If the predicted and actual status are not the same then it may be classified in one of four categories. If the predicted outcome is Employed but the actual status is Unemployed, then the observation is classified as 'Employment Overpredicted', and vice versa for 'Employment Underpredicted'. An observation is classified as 'LFP overpredicted' if the prediction is that the individual is in the labour force, i.e. they are either Employed or Unemployed, but the actual employment status is Out Of the Labour Force, and vice versa for 'LFP Underpredicted'. Table 15 in the Appendix describes the possible outcomes for every combination of predicted and actual economic status.

4.2 Results

First, we match the predicted outcomes on the PLFS data from the CMIE-CPHS model against actual outcomes in PLFS based on the UPSS definition.

The model estimation results are given in the appendix (Table 16). The findings of interest for this paper are the results of the matching exercise. We find that overall, the model estimated on CMIE-CPHS is able to predict the activity status of approximately 80% of individuals correctly in the NSS-PLFS data (Table 2).

Is an accurate estimation for 80 per cent of individuals indicative of a good match between the two surveys? To answer this we use the same model to predict outcomes within the CMIE-CPHS data itself. We also do the opposite exercise, i.e. estimating the model on NSS-PLFS and then using it to predict outcomes in CMIE-CPHS. Table 2 provides the results of this estimation. We can see that the rate of correct prediction is similar across all four cases. This implies that the model is as good at predicting employment statuses within the data it is estimated on, as it is at predicting employment status on another dataset.

Table 3 Distribution of prediction matches, women

	CMIE-CMIE	CMIE-PLFS	PLFS-PLFS	PLFS-CMIE
Matched	87.4	77.03	78.75	84.07
Employment overpredicted	0.06	0.04	0.08	0.2
Employment Underpredicted	0.01	0.01	0.05	0.03
LFP overpredicted	0.62	0.75	3.7	6.04
LFP underpredicted	11.9	22.17	17.41	9.66

Table 4 Distribution of prediction matches, men

Male (Table 4)	CMIE-CMIE	CMIE-PLFS	PLFS-PLFS	PLFS-CMIE
Matched	85.48	83.3	84.51	85.49
Employment overpredicted	1.82	3.0	3.22	1.92
Employment Underpredicted	0.03	0.1	0.07	0.08
LFP overpredicted	6.16	5.1	5.99	8.6
LFP underpredicted	6.51	8.6	6.21	3.91

This provides a very strong argument that the two surveys, despite definitional differences, classify approximately 80 per cent of the population in the same way in terms of their employment status.[4] As for the rest of the 20 per cent, the model fails to predict their actual status, and it is quite likely that this is because of important factors that are not measured.

It is instructive to see who the unmatched are and what is the nature of the mismatch.

Most of the mismatch comes from LFP underpredictions, i.e. individuals who are in the labour force as per NSS-PLFS but are identified by the CMIE model as being out of the labour force. We further estimated the predictions separately for men and women to investigate if there were gender-specific differences in the extent of under/overprediction.[5]

In the women-only model, the CMIE is able to predict correctly for only 77 per cent of the sample (Table 3). For 22 per cent of women, the CMIE model identifies them as being out of the labour force whereas according to the NSS-PLFS employment status, they are employed. The male–female difference in accurately capturing employment status holds true for all four estimation–prediction combinations, i.e. estimating on CMIE data and predicting on PLFS data, and vice versa (Table 4).

---

[4] It could be the case that there is a section of population that the model predicts correctly in the CMIE−CPHS data, but it gets replaced by another section in NSS−PLFS. This means that although the two surveys classified them differently, they both got predicted correctly within each dataset. However, going by the consistency of the results across slices along various variables, this seems unlikely.

[5] Slicing the data along other lines, including estimating separately for different states, did not show significant differences across other categories.

For men, the predictions by the CMIE-based model on PLFS data are correct for about 83% of the predictions. Similar outcomes are seen when a PLFS-based model is used to predict male employment outcomes in the CMIE data. Further the mismatches are more or less equally distributed across the LFP overpredicted and underpredicted categories.

Additionally, while for men the LFP under or overprediction are roughly equal suggesting errors being equally likely to occur both ways, in women almost all the mismatch comes from LFP underprediction. This shows that the model *consistently* predicts that women are out of the labour force when they are actually in it. Hence, we can conclude that the effects of differences in the definition of employment between CMIE-CPHS and NSS-PLFS are likely to show up in differences in classification of women who are in the labour force. Hence, one would expect that the workforce participation rate and unemployment rate for women could be very different between the two surveys (we confirm this in Sect. 5).

The lower prediction success of the models for women's labour force participation, in particular, the systematic underprediction of women's labour force participation could mean that some important factors that determine women's employment status are not captured in the surveys and are omitted from the model. This fits well with the literature on women's labour force participation which says that surveys do not adequately capture women's work (Hussmanns et al., 1990) and that a significant reason for low labour force participation of women are often factors on labour demand side, including the nature of work available rather than the woman's or her household's characteristics (Fletcher et al., 2017; Verick, 2014). In particular, women's work is particularly sensitive to the reference period as well as phraseology of questions (Deshmukh et al. 2019). It is likely that the differences in reference period as well as the broader way in which the question is asked are more likely to affect the measurement of women's work. Non-inclusion of the labour demand-side factors in the model and higher sensitivity of women towards these factors could explain the lower prediction rate of the model for women.

The presence of other systematic patterns in the demographic composition of the unmatched sample was also examined. In particular, one would want to see if particular kinds of individuals were more likely to be matched or mismatched between the two surveys. Table 5 provides the distribution of different demographic groups in the overall sample of NSS-PLFS, as well as within each prediction category. Only the matched and the LFP underpredicted groups are examined since the latter accounted for most of the mismatched sample.

So, the 50 per cent against 'Women' in the first column indicates that they were half the population in PLFS. However, as the subsequent columns show, they accounted for 71 per cent of the LFP underpredicted sample and only 48 per cent of the Matched sample. This means that women are overrepresented in that sample of individuals, who are in the labour force as per NSS-PLFS but are identified by the CMIE model as being out of the labour force. The share of uneducated individuals is also much higher in the LFP underpredicted group in comparison with the overall sample, indicating that for these individuals, CMIE is likely to identify them as out of the labour force, whereas PLFS identifies them as employed.

Table 5 Share of each category in each group (%), overall

	Overall sample (NSS-PLFS)	LFP under-predicted	Matched
Female	50	71	48
Rural	69	71	69
By educational category			
No education	27	35	26
1st–5th Standard	17	18	17
6th–8th Standard	20	17	21
9th–10th Standard	14	10	15
11th–12th Standard	11	8	11
Graduate/Diploma	9	9	8
Postgraduate +	2	3	2
By activity status			
Own Account Worker	17	21	17
Employer	1	1	1
Unpaid Helper	6	26	3
Regular Salaried	11	19	10
Casual Wage	11	23	10
Unemployed	3	10	0
Student	13	0	15
Domestic Work	30	0	37
Others	8	0	8
By age group			
15–20 years	17	19	17
21–25 years	12	12	11
26–30 years	12	11	12
31–40 years	20	21	21
41 years and above	38	36	39

This means that women are overrepresented in that sample of individuals who are in the labour force as per NSS-PLFS but are identified by the CMIE model as being out of the labour force. The share of uneducated individuals is also much higher in the LFP underpredicted group in comparison with the overall sample, indicating that for these individuals, CMIE is likely to identify them as out of the labour force, whereas PLFS identifies them as employed.

Comparing employment statuses, there are notable deviations in the sample of LFP underpredictions. Unpaid family workers, identified as such by the PLFS, are vastly overrepresented in the LFP underpredicted group. Unpaid family workers are family members who work in the household enterprise and do not receive any regular salary or wages in return. Typically, they constitute about 6 per cent of the population. However, in the LFP underpredicted sample, they account for about 26 per cent. Regular salaried workers are also overrepresented in the LFP

underpredicted group, indicating that CMIE is likely to categorise such individuals as out of the labour force. Therefore, at either ends of the.

In terms of age distribution, we see significant differences in either ends of the age spectrum–younger individuals constitute a relatively larger share in the LFP underpredicted category than their share in the PLFS labour force, while older individuals are a smaller share. This indicates that CMIE is likely to classify younger individuals as out of the labour force compared to PLFS. In general in the Matched sample, the distribution of groups by location, education and gender is broadly similar to their distribution in the overall sample. In the LFP underpredicted sample, however, there are significant departures in the distribution of certain categories from their distribution in the overall population.

We do the same exercise separately for men and women, comparing the distribution of these categories within NSS-PLFS with that in the Matched sample and LFP underpredicted sample to see if these discrepancies manifest similarly for men and women. Table 6 shows the distribution of educational and age groups for men and women, and Table 7 shows the distribution of employment status for men and women. The tables also show the representation index (RI) of each category in the LFP underpredicted and Matched sample. The representation index is the ratio of the share of each category in a particular estimated sample to the share in the overall NSS-PLFS sample. So, an RI of 1 indicates that particular category is equally represented in a particular sample as they are in the overall population.

A comparison of the educational and age distributions in the predicted samples against the overall PLFS sample indicates that for men, there is a similar representation of each age and educational group in the matched and the LFP underpredicted sample. The only exception is with respect to men between 26 and 40 years. This group is not present in the LFP underpredicted category, indicating that for them, the CMIE model is able to accurately predict their employment status.

For women, we see considerable variation in the RI of certain groups in the LFP underpredicted sample. Less educated and highly educated women are overrepresented in the LFP underpredicted sample. For instance, women with no education constituted about 35 per cent of the PLFS working age sample but were 41 per cent of the LFP underpredicted sample. So for these women, the CMIE model was likely to peg them as out of the labour force while the PLFS identified them as workers. In the age categories too, we find that middle-aged women were more likely to fall into the LFP underpredicted sample.

For instance, among women, OAWs constitute about 4 per cent of the working age population and 19 per cent of the labour force in NSS-PLFS. At the same time, they constitute 19 per cent of the LFP underpredicted sample, and 0.2 per cent of the Matched sample. This gives an RI of one in the LFP underpredicted sample—the OAWs, as per PLFS, who CMIE-CPHS categorise as not in the labour force, are equally represented in this mismatched category as they are in the overall labour force. On the other hand, they are severely underrepresented in the matched sample. This indicates that for those women categorised as OAWs in PLFS, CMIE is highly unlikely to recognise them as employed. We see a similar underrepresentation of salaried workers in the Matched sample. All employment statuses are more or less similarly represented in the LFP underpredicted

Table 6 Distribution of educational and age groups status across predicted sample, men and women

	Overall sample (NSS-PLFS)	LFP under-predicted	Matched	RI in LFP under-predicted sample	RI in Matched sample
WOMEN					
No education	35.4	40.9	33.5	1.2	0.9
1st–5th Standard	16.7	18.2	16.3	1.1	1.0
6th–8th Standard	18.0	14.3	19.2	0.8	1.1
9th–10th Standard	11.5	7.6	12.8	0.7	1.1
11th–12th Standard	9.1	5.5	10.3	0.6	1.1
Graduate/Diploma	7.1	9.6	6.4	1.4	0.9
Postgraduate+	2.2	4.0	1.6	1.8	0.7
By age group					
15–20 years	15.6	5.0	18.9	0.3	1.2
21–25 years	12.4	11.2	12.7	0.9	1.0
26–30 years	12.6	15.7	11.8	1.3	0.9
31–40 years	21.1	29.5	18.6	1.4	0.9
41 years and above	38.3	38.5	38.1	1.0	1.0
MEN					
By educational category					
No education	18.5	21.7	18.3	1.2	1.0
1st–5th Standard	16.7	17.7	17.1	1.1	1.0
6th–8th Standard	22.6	23.6	23.1	1.0	1.0
9th–10th Standard	16.0	14.5	16.7	0.9	1.0
11th–12th Standard	12.5	13.6	12.3	1.1	1.0
Graduate/Diploma	11.0	8.4	9.9	0.8	0.9
Postgraduate+	2.7	0.5	2.7	0.2	1.0
By age group					
15–20 years	17.8	54.9	15.4	3.1	0.9
21–25 years	12.3	14.9	9.7	1.2	0.8
26–30 years	11.9		12.7	0.0	1.1
31–40 years	19.8		23.0	0.0	1.2
41 years and above	38.3	30.2	39.2	0.8	1.0

sample for women. This indicates that irrespective of employment arrangement as assigned by PLFS, any kind of female worker is equally likely to be identified as being out of the labour force in CMIE-CPHS. And, across all employment categories, there is a similar underrepresentation of women in the Matched sample.

For men, in contrast, there is equal representation of all kinds of employment categories in the Matched sample. In terms of the LFP underpredicted sample, for men, we find that unpaid helpers, as identified by PLFS, are likely to categorised as out of the labour force by CMIE, while OAWs and salaried workers are under-captured as workers by CMIE.

Table 7 Comparison of employment patterns between CMIE & PLFS, with representation index

	Overall sample (UPSS, NSS-PLFS)	Labour force sample (NSS PLFS	LFP under-predicted	Matched	RI in LFP under-predicted sample	RI in Matched sample
	FEMALE					
Own Account worker	4.3	18.6	19.0	0.2	1.0	0.03
Employer	0.1	0.5	0.5	0.0	1.0	0.08
Unpaid Helper	7.0	29.9	30.0	0.4	1.0	0.06
Regular Salaried	4.6	19.9	20.2	0.2	1.0	0.04
Casual Wage	6.0	25.5	24.6	0.7	1.0	0.11
Unemployed	1.3	5.6	5.7	0.0	1.0	0.02
Student	11.1		0.0	14.4	0.0	1.3
Domestic Work	58.3		0.0	74.8	0.0	1.3
	MALE					
Own Account Worker	29.8	39.3	27.1	32.9	0.7	1.1
Employer	1.7	2.3	1.9	1.9	0.8	1.1
Unpaid Helper	5.7	7.6	16.1	5.1	2.1	0.9
Regular Salaried	16.7	22.0	16.1	18.3	0.7	1.1
Casual Wage	17.3	22.8	19.3	18.7	0.8	1.1
Unemployed	4.6	6.1	19.6	0.1	3.2	0.0
Student	14.7		0.0	15.9	0.0	1.1
Domestic Work	0.9		0.0	0.6	0.0	0.6

Table 8 Distribution of workforce by employment type

	PLFS (UPSS)			CMIE		Representation Index (CMIE/PLFS)	
	Male	Female		Male	Female	Male	Female
Own Account Worker	41.83	19.7					
Employer	2.41	0.5					
Unpaid Worker	8.05	31.63					
(Total self-employed)	*52.29*	*51.9*	Self-employed	47.6	25.2	0.9	0.5
Salaried	23.4	21.08	Salaried#	21.2	26.4	0.9	1.3
Casual wage worker	24.31	27.06	Casual wage worker	31.2	48.4	1.3	1.8

Salaried workers in CMIE include permanent and temporary salried

Comparing the distribution of the workforce across the two surveys throws further light on this mismatch. Table 8 compares the distribution of the working age population across activity statuses between the two surveys. The Representation Index in the last two columns calculates the factor by which a particular category is under/overrepresented in CMIE compared to PLFS. A factor close to one indicates that there is a similar share of that category of workers in both surveys. A factor less than one indicates that their share is lower in CMIE in comparison with PLFS.

Table 9 Distribution of men and women across industries

	PLFS (UPSS)		CMIE		Representation Index (CMIE/PLFS)	
	Male	Female	Male	Female	Male	Female
Agriculture	28.4	11.2	23.7	5.0	0.8	0.4
Mining&Utilities	0.9	0.1	0.5	0.0	0.6	0.4
Light Manufacturing	1.9	0.9	1.3	0.3	0.7	0.4
HeavyManufacturing	4.4	0.4	3.9	0.2	0.9	0.4
Construction	9.8	0.9	12.6	0.5	1.3	0.6
Textiles/Apparel	2.3	1.2	1.7	0.4	0.7	0.3
Trade	8.3	1.0	10.0	0.7	1.2	0.7
Health	0.6	0.6	0.4	0.3	0.7	0.6
Education	1.9	1.7	2.0	0.9	1.0	0.5
PublicAdmin	1.3	0.2	1.7	0.2	1.3	0.9
OtherServices	11.4	2.0	11.5	1.5	1.0	0.8
Out of Workforce	29.0	79.8	30.7	89.9	1.1	1.1

Light manufacturing refers to NIC code 10–17, heavy manufacturing refers to NIC code 18–33

For men, the distribution of the workforce across employment types is broadly similar across the two surveys (Table 8). For women, there is a much lower share of self-employed workers in CMIE as compared to PLFS. Since CMIE does not distinguish between the categories of self-employment (Own Account Worker (OAW), unpaid worker, employer), we cannot distinguish for certain what this source of mismatch is. The very low representation index for self-employed indicates the particular underrepresentation of these groups of workers in the CMIE female workforce.

We also examined whether activity in a particular industry for men or women was being differently captured across the two surveys. Table 9 compares the distribution of the working age population in different industries across the two surveys along with the representation index.

For men, as can be seen from the representation index, there is a fairly similar distribution of men across major industries between the two surveys. However, for women, there is a systematic underrepresentation of women across all industries, and in particular in textiles/apparel. Here too, given the predominance of women in this kind of work, as per PLFS, this results in the overall lower estimations of women in the workforce in CMIE.

4.3 Comparison Across Different Definitions of Employment

As described in Sect. 2, the NSS-PLFS collects employment information using four reference periods, and therefore has four variations of employment statuses for any individual. In the above analysis, we used the UPSS definition to estimate the model and make inter-data comparisons. If using the UPSS definition increases (or decreases) the prediction success rate, this could imply that the UPSS definition is closer to the CMIE-CPHS definition of employment.

Table 10 Comparison across different definitions of workforce, men

CMIE model run on PLFS data	UPA	UPSS	CWS	CDS
Matched	81.22	80.1	80.22	82.13
Employment overpredicted	1.59	1.59	2.36	0
Employment underpredicted	0.02	0.02	0.02	0.02
LFP overpredicted	3.37	3.3	3.57	4.91
LFP underpredicted	13.8	14.49	13.84	12.94

Table 11 Comparison across different definitions of workforce, women

CMIE model run on PLFS data (female)	UPA	UPSS	CWS	CDS
Matched	79.16	77.04	79.05	79.36
Employment overpredicted	0.04	0.04	0.09	0
Employment underpredicted	0.01	0.01	0.01	0.01
LFP overpredicted	0.95	0.9	0.99	0.97
LFP underpredicted	19.85	22.01	19.87	19.66

We make similar comparisons with the status as identified under the UPA, CWS or CDS definitions. We needed to adapt the CDS estimate under PLFS to make it comparable with the CMIE definition. The CDS definition identifies an activity status for an individual for every seven days prior to the date of the survey. It also accounts for people who are generally employed but happened to be not working on that day on account of sickness or other reasons. For the purpose of this analysis, we identify a person as working if they worked in the day preceding the survey, or were in general working, i.e. were recorded as working under UPA, but happened to be not in the workforce on the day preceding the survey. We believe that this most approximates the CMIE definition of employment as those who were working on the day of, or day preceding the survey, or have some surety of employment.

The multinomial logit model is estimated on the CMIE-CPHS data as earlier, and the predictions made on the NSS-PLFS data. Then, predictions are matched to the actual employment status according to the various definitions. The results are in Tables 10 and 11.

Here too, we find that irrespective of the definition of employment used, the CMIE-PLFS predictions deviate more in the case of women, than in the overall sample. Therefore, the divergence between the two estimates for women's employment status continue irrespective of the definition of employment used.

4.4 Robustness Check: CMIE-CPHS as a Pooled Cross Section

The CMIE-CPHS survey is a panel survey, which we convert to a pseudo-cross section by dropping random observations per individual so that every individual is represented only once in the dataset. Doing so made the CMIE-CPHS data most comparable to the NSS-PLFS data structure. As a robustness check, we treated the

CMIE-CPHS survey as a pooled cross section, with every observation representing a separate individual. Therefore, although any individual had up to three observations for him/her, for the analysis we treated these as separate individuals. This was akin to treating the panel structure as a pooled cross section. We find that the results did not vary significantly when estimated on a pooled cross section. So, the results are robust to all definitions of employment status and are not affected by the differences in sampling methodology across surveys.

5 Comparing State-Level Estimates

In this section, we compare the aggregate estimates of employment status from the CMIE-CPHS survey with the NSS-PLFS for the same year. For most policy and discussion purposes, these aggregate estimates are the ones that are used. Hence, the most important question is of the kind: 'If the CMIE-CPHS estimate of the unemployment rate is 6%, how much is it according to the NSS definition?' While we do not attempt to provide a direct answer to this question, we try to provide a framework for thinking about the differences in the aggregate estimates in the two surveys.

There are a couple of ways one could think about how the differences in the definition and survey method could translate into differences in aggregate measures. One is to assume that there are certain sections of the population that are going to be classified differently in the two surveys. States that have a higher proportion of these populations are going to show more divergence between estimates from the two surveys.

This follows from the approach we took in the previous section, and allows us to make some predictions based on our results there. We established there, for instance, that women and unpaid workers are more likely to be attributed a status of being out of the labour force in the CMIE compared to the NSS-PLFS. We had found that for women the differences in classification are likely to be for those who are in the workforce as per the NSS-PLFS, but for men there was no such pattern. Hence, we can predict that states that have a higher labour force participation for women are going to see more mismatch between the workforce participation rates in the two surveys, while this will not be the case for men. We find exactly this result in the data (Table 10).

While this is instructive, it does not bring us any closer to answering the question posed at the beginning of this section, i.e. is there a consistent divergence between the CMIE-CPHS estimate and the NSS-PLFS estimate? To compare observations at the state level, and examine the differences in methods and definitions, we can model the classification as a stochastic process. Imagine the ideal case where the sample of the two surveys is exactly the same–by doing this we abstract from any sampling errors that may contribute to differences in estimates.[6] Now every individual i is classified according to the NSS-PLFS

---

[6] As we are looking at state level aggregates, the sample sizes are still large enough to make this abstraction.

definition as either Employed or Not Employed. The proportion of individuals classified as Employed gives us the NSS-PLFS estimate of the WPR (Workforce Participation Rate). For an individual classified as Employed in the NSS-PLFS data, let the probability of him/her being classified as Employed in the CMIE-CPHS data be p_1. Similarly, for an individual classified as Not Employed in the NSS-PLFS data, let the probability of him/her being classified as Employed in the CMIE-CPHS data be p_2. If we define binary variables x_i and y_i that take the value 1 if individual i is classified as Employed in the NSS-PLFS and CMIE data, respectively, then we can write

$$Pr(y_i = 1) = p_1^{x_i} p_2^{1-x_i}, x_i \in \{0, 1\}$$

Hence if the sample size is n out of which there are k individuals classified as Employed in the NSS-PLFS data, then the probability distribution of the number of Employed in the CMIE-CPHS data is a Poisson Binomial distribution, with its expected value given by $E[k'] = kp_1 + (n-k)p_2$. This would be the expected number of people classified as Employed in the CMIE-CPHS data. Hence, the expected value of the WPR in the CMIE-CPHS data for a state j is related to the WPR in the PLFS data through the equation

$$E\left[\text{WPR}_j^{\text{CMIE}}\right] = E[\frac{k'_j}{n_j}] = p_{2j} + (p_{1j} - p_{2j})\text{WPR}_j^{\text{PLFS}}$$

This can be rewritten as follows.

$$\text{WPR}_j^{\text{CMIE}} = p_{2j} + (p_{1j} - p_{2j})\text{WPR}_j^{\text{PLFS}} + e_j \qquad (1)$$

In this general form, the probabilities are allowed to be different for each state, aligning to the idea of states having different proportions of populations that are prone to being differently classified. But this cannot be estimated from the data we have. In order to estimate this and get some useful interpretation out of the estimation, we make two assumptions.

- We assume that the value of p_1 is the same for all individuals. $p_{1j} = p_1 \forall j$.
- We assume $p_2 = 0$. We know that CMIE, on average underpredicts both LFPR.

and WPR. Hence, we make this assumption to easily interpret the regression results.

Hence, the regression model becomes.

$$\text{WPR}_j^{\text{CMIE}} = p\text{WPR}_j^{\text{PLFS}} + e_j \qquad (2)$$

We estimate this using aggregate estimates for WPR and LFPR for 24 states obtained from the NSS-PLFS 2015–16 data and the pseudo-cross section constructed from the CMIE 2016 data.

5.1 Unemployment Rate

The question we started this section with was about unemployment rate. In an attempt to answer the question, let us consider the LFPR counterpart of Eq. 2.[7]

$$LFPR_j^{CMIE} = p' LFPR_j^{PLFS} + v_j \tag{3}$$

Now, we can get an expression for the unemployment rate (UR) as estimated from.

CMIE data in terms of that estimated from LB data.

$$UR_j^{CMIE} = 1 - \frac{WPR_j^{CMIE}}{LFPR_j^{CMIE}} = 1 - \frac{p\, WPR_j^{PLFS} + e_j}{p' LFPR_j^{PLFS} + v_j} \tag{4}$$

From 4, we can see that there is no way of deriving a linear relationship between UR_j^{CMIE} and $UR_j^{PLFS} = 1 - \frac{WPR_j^{PLFS}}{LFPR_j^{PLFS}}$. Hence, running a linear regression between state-level unemployment rate estimates from CMIE-CPHS and NSS-PLFS would not make any sense as the coefficient would not have any interpretation under this model. Hence, we restrict ourselves to LFPR and WPR.

5.2 Regression Results

Figure 1 compares the state-level LFPR and WPR estimates from the NSS-PLFS and the CMIE-CPHS with the NSS-PLFS estimates on the horizontal axis and the CMIE estimates on the vertical axis. The results from the regression are presented in Table 11.

We can see that our estimates of the coefficient p are 0.89 for both LFPR and WPR. This can be interpreted as an underreporting of LFPR and WPR in the CMIE data compared to the NSS-PLFS data, of around 11 per cent (Table 12).

Here, we have assumed that the p is constant for all individuals. But from our previous exercise, we know that it is likely to be different for men and women. Hence, we estimate the models separately for men and women. The results are presented in Figs. 2 and 3 and Tables 12, 13 and 14.

We find that while the underreporting is estimated as significantly higher for women, for men it is close to zero. For men, we cannot statistically reject the hypothesis that $p = 1$. The most likely explanation is that for men the differences in classification may either be very small or be in different directions that get averaged, while for women they are systematically in one direction, which shows up in the graph and regression.

Now we look at this separately for men and women, knowing that much of the deviation from the two surveys comes in the capturing of women's work has been difficult. Figures 2 and 3 show the comparison for men and women

---

[7] Note that here we take the Usual Principal Activity Status (UPS) numbers for the official survey, and we define LFPR for the CMIE−CPHS data to not include people who are unemployed and. willing to work but not actively looking for work.
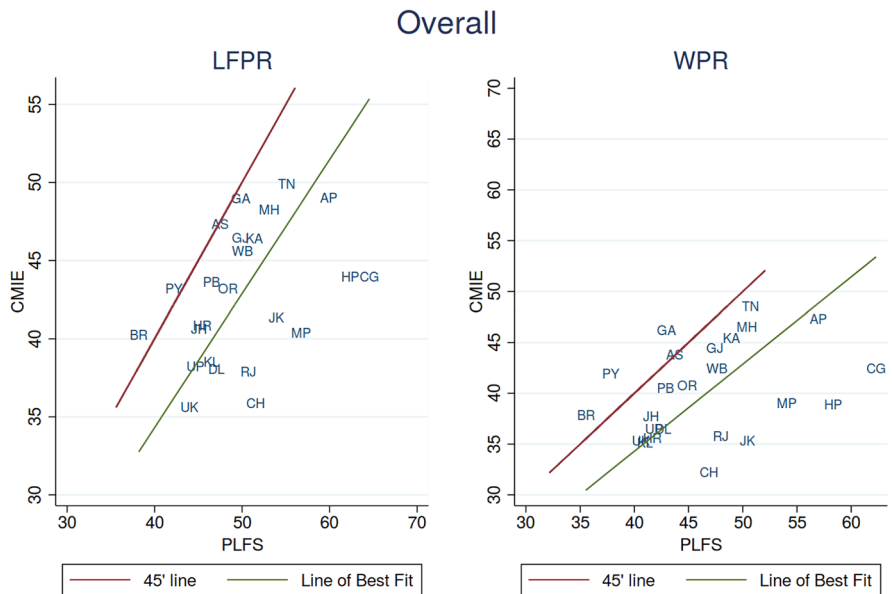
Fig. 1 Comparison of state-level estimates of overall LFPR and WPR obtained from PLFS 2017–18 and concurrent CMIE waves

Table 12 Regression estimates, overall

	Cmie_Lfpr	Cmie_Wfpr
plfs_lfpr_upa	0.881***	
	(−0.0232)	
plfs_wfpr_upa		0.883***
		(−0.0236)
Observations	26	26
R-Squared	98	98

Table 13 Regression estimates, women

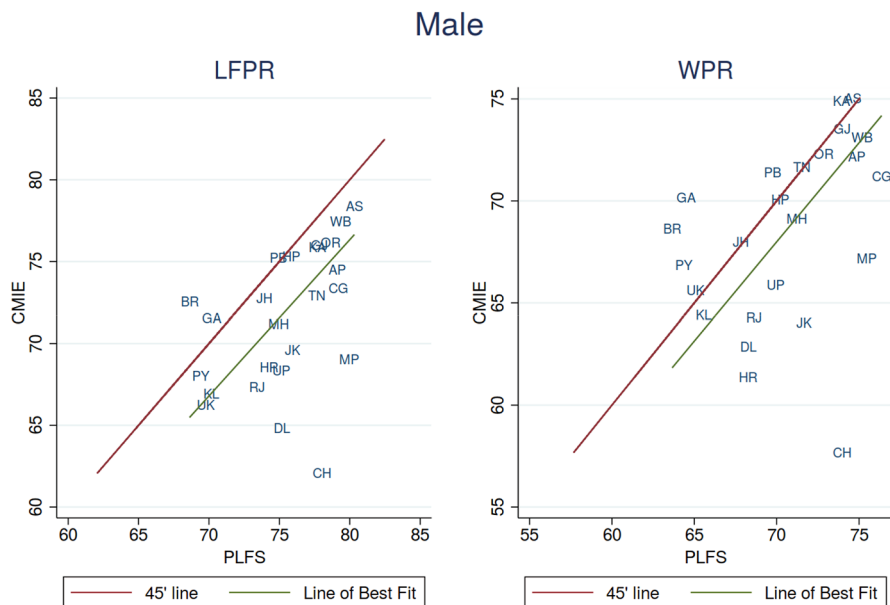|	−1	
	LFPR CMIE (Female)	WFPR_CMIE
LFPR PLFS (Female)	0.529***	
	(−0.061)	
WFPR_PLFS (Female)		0.477***
		(−0.0598)
Observations	26	2%6

Fig. 2 Comparison of state-level estimates of LFPR and WPR for men, obtained from PLFS 2017–18 and concurrent CMIE waves
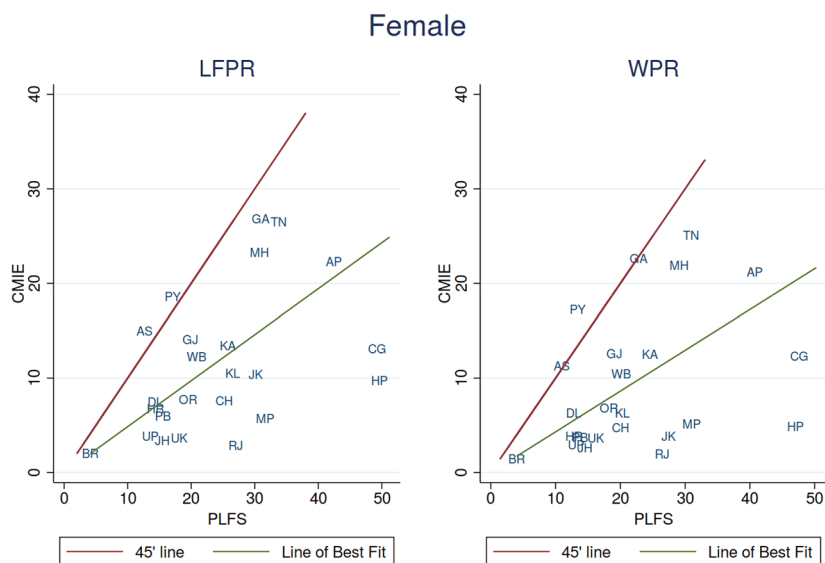

Fig. 3 Comparison of state-level estimates of LFPR and WPR for women, obtained from PLFS 2017–18 and concurrent CMIE waves

Table 14 Regression estimates, men

	LFPR CMIE (Male)	WFPR CMIE
LFPR PLFS (Male)	0.959*** (−0.0118)	
WFPR PLFS (Male)		0.980*** (−0.0125)
Observations	26	26

separately. As we can see the estimated LFPR and WPR for men from the two surveys are very close with a difference of less than 5 per cent. For women, the LFPR estimates are close but WPR estimates show a huge bias where CMIE underpredicts WPR by 40 per cent.

An additional point to note is that for men not only is the estimate close to 1 but the standard errors are also very small. This shows that not only is p close to 1 on average, it is likely to be close to 1 for most states too. This implies that whatever section of the male working age population gets differently classified in the two surveys, the misclassification averages out in most states resulting in close matching between the male LFPR and WPR numbers obtained from the two surveys.

The low WPR for women in CMIE may be due to multiple reasons. Some modes of doing the survey—which in the case of CMIE is an extended conversation with the household head—may lend themselves to more bias (Bardasi et al., 2011). Additionally, the difference in the way in which employment status is ascertained could mean that some people who would be classified as employed according to the principal status question would be classified as unemployed in the CMIE definition. This is likely to be particularly the case for people who are in irregular employment, and women are more likely to be in this situation.

Hence, the practical conclusion we can draw from this exercise is that while comparing LFPR and WPR numbers obtained from CMIE data to official statistics, it would be better to only use the estimates for men as the divergence in the estimates is quite significant for women. We can also conclude that whichever way we think about the process that generates the difference in estimates, the trends are going to be in the same direction, i.e. if the LFPR/WPR decreases according to CMIE, it will also decrease according to the government definition. This is because regardless of whether the difference is caused by sections of the population being classified differently, or by a stochastic process that generates the difference probabilistically, the parameters of these processes are not going to change with time, at least not over the period of a few years.

6 Conclusion

In the previous year, amidst the lack of government survey data to generate labour market estimates, CMIE-CPHS has emerged as the only source of household level nationally representative employment data until the PLFS 2020–21 round was released, nearly a year after the lockdown. In more recent times, CMIE-CPHS continues to be an important source of data being the only survey that collected data in the months prior to the lockdown, during the lockdown, and continues to do so. In this paper, making use of the overlap in the time periods of the official unit-level survey as well as CMIE survey, we compare key labour market indicators across the two surveys. Considering the variations in definitions of economic activity status and differing reference periods between the government surveys and CMIE-CPHS, we see to what extent the estimates from these surveys are comparable.

We first run an econometric model of employment status on a pseudo-cross section constructed from CMIE-CPHS 2017 data and find that 80% of the observations in the government surveys matched with the predicted status obtained from the model. Next, we develop a stochastic model of employment status classification and use it to compare state-level aggregate estimates from the CMIE-CPHS 2017 and PLFS 2017.

Taken together, both the econometric analysis and analysis of state-level variations indicate that measures of women's participation in the labour force seem particularly sensitive between the two surveys, and predictions of women's LFPR based on standard labour supply variables are vary considerably compared to those for men. When using data for men, the level of comparability is quite high and aggregate estimates like LFPR and WPR are found to match very closely. Further, we find that all kinds of economic activities of women are equally likely to not be captured by CMIE, and it is not one particular kind of women's employment that is being undercaptured. We find that it is not the reference period differences per se that is contributing to this difference as comparison of CMIE estimates with all reference periods under PLFS provide similar numbers.

Appendix

See Tables 15, 16.

Table 15 Categorising actual and predicted status combinations

Actual status	Predicted status	Category
OOLF	OOLF	Matched
OOLF	Unemployed	LFP overpredicted
OOLF	Employed	LFP overpredicted
Unemployed	OOLF	LFP underpredicted
Unemployed	Unemployed	Matched
Unemployed	Employed	Employment overpredicted
Employed	OOLF	LFP underpredicted
Employed	Unemployed	Employment underpredicted
Employed	Employed	Matched

Table 16 Estimates of multinomial logit model of employment status on CMIE-CPHS

Base outcome(OOLF)	Urban female		Rural female		Urban male		Rural male	
	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed
Age	0.0361***	0.229***	0.0440***	0.273***	0.375***	0.886***	0.295***	0.727***
	−0.0117	−0.00615	−0.0107	−0.00701	−0.0128	−0.00827	−0.0124	−0.00702
Age squared	−0.00128***	−0.00295***	−0.00106***	−0.00324***	−0.00543***	−0.0106***	−0.00418***	−0.00844***
	−0.000181	−0.0000781	−0.000153	−0.0000858	−0.00019	−0.000105	−0.00018	−0.0000879
Education categories (Base = No education)								
Primary	0.233	−0.311***	0.198	−0.242***	0.218	0.354***	0.145	0.538***
	−0.177	−0.0543	−0.134	−0.0486	−0.26	−0.128	−0.217	−0.0834
Middle	0.297*	−0.699***	0.0172	−0.514***	0.158	0.335***	0.141	0.656***
	−0.161	−0.0496	−0.13	−0.0454	−0.244	−0.116	−0.196	−0.0754
Secondary	0.314*	−0.922***	0.167	−0.567***	0.0328	0.281**	0.315	0.717***
	−0.163	−0.061	−0.141	−0.0614	−0.242	−0.116	−0.194	−0.0774
Higher secondary	0.481***	−1.182***	0.550***	−0.866***	0.239	−0.175	0.562***	0.269***
	−0.158	−0.0556	−0.133	−0.0604	−0.238	−0.113	−0.187	−0.0741
12th Standard Pass	0.984***	−1.086***	1.229***	−0.713***	0.606**	−0.720***	1.092***	−0.155**
	−0.158	−0.0625	−0.134	−0.0786	−0.236	−0.112	−0.186	−0.0753
Diploma/certificate	1.640***	−0.00872	2.102***	0.0587	1.331***	−0.430***	1.529***	0.138
	−0.246	−0.15	−0.293	−0.238	−0.266	−0.14	−0.254	−0.137
Graduate	1.659***	−0.245***	2.104***	−0.0644	1.332***	−0.483***	1.963***	−0.0037
	−0.156	−0.054	−0.137	−0.09	−0.237	−0.113	−0.188	−0.0811
Postgraduate	1.859***	0.639***	2.745***	1.095***	1.512***	−0.171	2.191***	0.268**
	−0.168	−0.0608	−0.174	−0.135	−0.249	−0.121	−0.229	−0.122
No. of children per household member	−1.367***	−1.805***	−1.110***	−0.690***	−1.772***	1.446***	−1.637***	1.961***
	−0.245	−0.195	−0.307	−0.158	−0.416	−0.191	−0.384	−0.161

Table 16 (continued)

Base outcome(OOLF)	Urban female		Rural female		Urban male		Rural male	
	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed
Constant	− 2.604***	− 6.041***	− 3.025***	− 8.012***	− 7.107***	− 14.62***	− 6.173***	− 12.07***
	− 0.238	− 0.157	− 0.228	− 0.196	− 0.314	− 0.192	− 0.286	− 0.149
Observations	5649	13,880	2759	10,614	7840	119,782	3910	72,983

Standard errors in parentheses

*$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Declarations

Conflict of interest Authors declare that they have no competing interests or conflict of interest related to the work submitted. This study was supported by funding from Azim Premji University Research Centre.

References

Somanchi, A. (2021). Missing the poor, big time: A critical assessment of the consumer pyramids household survey. Retrieved December 12, 2021 from https://osf.io/qmce9/

Bardasi, E., K. Beegle, A. Dillon, and P. Serneels. 2011. Do labor statistics depend on how and to whom the questions are asked? results from a survey experiment in Tanzania. *The World Bank Economic Review* 25 (3): 418–447.

Deshmukh, Neerad, Sonalde Desai, Santanu Pramanik and Dinesh Tiwari. 2019. 'Improving measurement of women's work participation: Sensitivity of work participation rates to questionnaire design.' *NCAER Data Innovation Centre Measurement Brief No. 2020–1.* NCAER, New Delhi.

Fletcher, E., R. Pande, and C.M.T. Moore. 2017. Women and work in India: Descriptive evidence and a review of potential policies. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3116310.

Hussmanns, R., F. Mehran, & V. Varma (1990). Surveys of economically active population, employment, unemployment, and underemployment: an ILO manual on concepts and methods. International Labour Organization

Kingdon, G.G., and J. Unni. 2001. Education and women's labour market outcomes in India. *Education Economics* 9 (2): 173–195.

Klasen, S. & J. Pieters (2012). Push or pull? drivers of female labor force participation during india's economic boom. IZA Discussion Paper No. 6395.

Somanchi, A & Dreze, J. (2021) New barometer of India's economy fails to reflect deprivations of poor households. Economic Times, 21st June 2021.

Verick, S. (2014). Women's labour force participation in India: Why is it so low?. International Labor Organization

Vyas, M. (2021). View: There are practical limitations in CMIE's CPHS sampling, but no bias. Economic Times 23rd June, 2021.