

Indian urban workers' labour market transitions

Jyotirmoy Bhattacharya *

November 13,2021

Abstract

This paper studies gross labour market flows and determinants of labour market transitions for urban Indian workers using a panel dataset constructed from Indian Periodic Labour Force Survey (PLFS) data for the period 2017–18 to 2019–20. Longitudinal studies based on the PLFS have been hampered by data problems that prevent a straightforward merging of the 2017–18 and 2018–19 data releases. In this paper, we propose and validate a matching procedure based on individual and household characteristics that can successfully link almost all records across these two years. We use the constructed data set to document a number of stylised facts about gross worker flows and to estimate the effects of different individual characteristics and work histories on probabilities of job gain and loss.

1 Introduction

The movement of workers from job to job and from employment to non-employment and back, and the simultaneous creation and destruction of jobs by firms, are the mechanisms through which labour markets reallocate resources in response secular changes and shocks. How well they do this — in terms of efficiently matching workers to jobs with the least amount of resources lost in waiting and search — is an important determinant of individual welfare and aggregate productivity.

For this reason, the study of labour market transitions and gross worker and job flows has been an important part of empirical labour economics for long. Some of the key contributions in this literature are [Abowd and Zellner \(1985\)](#), [Blanchard et al. \(1990\)](#), [Davis and Haltiwanger \(1992\)](#), [Davis et al. \(1998\)](#), [Shimer \(2005\)](#), [Hall \(2005\)](#), [Fujita and Ramey \(2009\)](#), [Shimer \(2012\)](#), [Elsby et al. \(2013\)](#), [Hall and Kudlyak \(2019\)](#) and [Ahn and Hamilton \(2021\)](#). A recent review of evidence is provided by [Davis et al. \(2006\)](#). This empirical literature developed in dialogue with theoretical work on search and matching models of the labour market, surveyed for example in [Rogerson and Shimer \(2011\)](#).

*Ambedkar University Delhi, Email: jyotirmoy@jyotirmoy.net. I am grateful to seminar participants at Centre for Sustainable Employment (Azim Premji University) and Ambedkar University Delhi and to Ishan Anand and Anamitra Roychowdhury for very helpful comments. Replication code for the paper is available at: <https://github.com/jmoy/plfs-jb>

The picture of the labour market that emerges from these studies is one of a constant churning in which firms create and destroy jobs in response to growth or shrinkage in their activity and workers move from job to joblessness and other jobs either because their old job has been destroyed, or because of individual-level shocks, or in hopes of better opportunities. The changes observed in aggregate employment rates are just the small net surface result of this incessant process of reallocations at the microeconomic level. The wealth of empirical evidence on these flows that has been available for developed economies informs current research on the business cycle behaviour of the labour market and the normative and welfare impacts of different labour market institutions and policies.

Unfortunately, for long it was not possible to study these processes for the Indian economy because of the unavailability of any sources of panel data on workers. However, in just the last couple of years, such data has finally become available in the form of the official Periodic Labour Force Survey (PLFS) conducted by the Indian National Sample Survey Office (NSSO) as well as the private Consumer Pyramids Household Survey (CPHS) of the Centre for Monitoring Indian Economy (CMIE).

The present study is part of the emerging literature that uses these sources to study labour market transitions in India. We use data from all three available annual micro-data releases of the PLFS, viz. from 2017–18, 2018–19 and 2019–20, to look at gross worker flows for urban India. The paper is restricted to urban workers since the PLFS has a panel structure only for urban workers.

We begin by constructing a combined panel dataset from the three annual data releases. This is a non-trivial task because of an undocumented change in sampling unit identifiers that makes it impossible to use the provided household and individual identifiers to merge the data from the 2017–18 and 2018–19 files. This is perhaps the reason why PLFS data has not been used much for longitudinal studies so far. We overcome this difficulty by developing a matching procedure based on individual and household characteristics that can match records between these two files with a high degree of accuracy. This procedure is described in section 3.3.

In section 4 we use this constructed dataset to compute gross flow rates across different labour market states and record a set of stylised facts that emerge. These facts mostly match what is observed in the case of other countries, but with some specificities as well, the most important being a large difference between the patterns for men and women and the existence of large flows between salaried employment, casual employment and self-employment.

Section 5 documents large differences in gross flow rates between different industries, different forms of employment and different states (provinces). It finds that while most of the differences between states can be explained by differences in the composition of economic activity, there still exists for some states an unexplained state-specific effect.

Finally, section 6 looks at the determinants of job gains and losses at the individual

level. The results once again show a vast difference between men and women in the probabilities of job gains and losses and in the impact of marriage and child-bearing on these probabilities. The results also show differences in labour market outcomes by age and level of education and a significant impact of employment histories on transition probabilities.

2 Related work

As we have mentioned, due to the paucity of data, there is very limited prior research on gross flows in Indian labour markets.

[Sarkar et al. \(2019\)](#) used Indian Human Development Survey (IHDS) panel data from 2005 and 2012 to look at labour market transitions of Indian women in the context of the debate around the low and decreasing Labour Force Participation Rate (LFPR) of Indian women. They found that women have higher exit probabilities from and lower entry probability into employment compared to men. They also studied different determinants of entry and exit. The key limitation of their paper, imposed by their data source, was the long gap of seven years between the two observations. As a result they faced serious attrition problems with 20% of women in the 2005 sample not being present in the 2012 sample. Also, they could not observe labour market movements at higher frequencies, which we find to be significant.

[Deshpande and Singh \(2021\)](#) use Consumer Pyramids Households Survey data from 2016–2019 to look at women’s labour market transitions, once again in the context of the women’s LFPR debate. The panel design of the CPHS allows them to track individuals over the entire study period and one of their key findings, corroborated by the present study, is that the low cross-sectional labour participation rate of women does not truly capture the work histories of women. They find large gross movements of women into and out of work at a quarterly frequency. Though the the average cross-sectional labour force-participation rate of women was only 14.6% in their data, the percentage of women who were observed to be employed at least once in the entire study period was much higher at 44%.

The paper closest to the present one is that of [Menon and Nath \(2021\)](#) who look at urban labour market transitions and gross flows using the same Periodic Labour Force Survey data that we use . However they only work with the 2017–18 and 2018–19 data releases and do not address the coding change between these two data releases. They are therefore forced to split worker histories at the boundary of these two years. From each year, they consider only one panel — that which started in the first quarter in that year — and only look at transitions at an annual frequency. They also look at only three broad labour market states — employment, unemployment and non-participation. In contrast our recoding algorithm allows us to merge data from 2017–18 and 2018–19 files, which we further combine with the 2019-20 files to create one combined data set that includes all the PLFS data released so far. As a result we work with much larger

sample sizes. We look at all the panels and calculate transition rates at a quarterly frequency with a finer set of labour market states which distinguishes between salaried work, casual work and self-employment. As a result we are able to pick out gross flows at a higher granularity that brings out more clearly the dynamic nature of the urban Indian labour market.

3 Data

3.1 Survey design

The Periodic Labour Force Survey (PLFS) is a quarterly survey on employment issues conducted by the National Sample Survey Office (NSSO) of the Government of India. Started in 2017-18, it replaces the earlier quinquennial employment-unemployment surveys of the NSSO. Annual reports and microdata for the survey are published in a July-June cycle. So far data has been released for the years 2017–18, 2018–19 and 2019–20. The design of the survey is documented in [NSSO \(2016\)](#).

The PLFS covers both rural and urban areas. In urban areas it has a rotational panel design, with a new panel starting every quarter and being visited for four successive quarters. This is the first official household survey in India with a panel structure and the primary aim of this paper is to use this structure to study transition processes in the labour market that cannot be captured by purely cross-sectional data. Since rural households are visited only once, this paper is restricted to urban households.

The PLFS is a stratified, multi-stage survey. We describe the sample design only for the urban areas. The first-stage units (FSUs) are urban blocks from the Urban Frame Survey of the NSSO. This sampling frame changes every two years, so in the period of our study one sampling frame is used for 2017–18 and 2018–19 and another for 2019–20.

Table 1 gives the number of households in each panel visited in each quarter. The staggered pattern of the table demonstrates the rotating nature of the panels. The table starts with 2017-Q3 since the PLFS follows a July-June cycle. Following the official PLFS documentation, the panels are identified as P_{ij} where i refers to the sampling frame used and j to the serial number of the panel.

The broad outline of the survey design is as follows. A first stage of stratification is carried out by dividing towns and cities on the basis of population. From these strata, blocks, which are the first-stage units, are selected using probability proportional to size with replacement. Larger blocks are divided into sub-blocks and only two sub-blocks are selected from these blocks for further sampling.

Households in blocks are stratified on the basis of the general education level of their members. From these second-stage strata households are selected for panels on the basis of simple random sampling without replacement.

Table 1: Number of households visited in each panel in each quarter

No. of households visited (thousands)										
Year-Quarter	Panel identifier									
	P11	P12	P13	P14	P15	P16	P17	P18	P21	P22
2017-Q3	11.4									
2017-Q4	11.1	11.5								
2018-Q1	11.0	11.2	11.5							
2018-Q2	10.8	11.0	11.2	11.5						
2018-Q3		11.0	11.1	11.3	11.5					
2018-Q4			11.1	11.1	11.3	11.5				
2019-Q1				11.0	11.2	11.3	11.4			
2019-Q2					11.2	11.3	11.4	11.1		
2019-Q3						10.9	10.9	10.8	11.4	
2019-Q4							11.3	11.1	11.3	11.5

3.2 Initial validation

We merge the data files for all the three years and run the following validation checks to ensure that distinct households or persons are not merged together due to data errors in the original data.

1. The religion and social group of the household should not have changed.
2. The size of the household should not have changed by more than 3.
3. A person's gender and relation to the head of the household should not have changed.
4. A person's age should not have changed by more than 4 years.¹

While some of these variables can in fact change for households and persons, the frequency of such changes is rare enough that most reported changes are likely to be due to data errors. We therefore remove all the records of households which fail any of these tests, removing the records of the entire household in case any of its members fails any of the person-level tests. 634 households are removed from the sample as a result.

¹Given that each person is followed over four quarters, age in years should not change by more than 1. However, examination of the reported age distribution shows significant bunching around multiples of 2 and 5 which suggests imprecise reporting of age. Since the purpose of this step is only to prevent incorrect linking, we keep a wider acceptance range for reported age.

3.3 Constructing the panel dataset

In the microdata files, each household is supposed to be uniquely identified by the combination of panel number, first-stage unit (FSU) number, sub-block number, second stage stratum number and sample household number. Each person within a household has a person number which remains the same across visits. In principle this should lead to a simple linking of households and individuals across different years' microdata files. However, a number of problems emerge in actually carrying out this process using the files provided by the NSSO. These data issues have also been documented by [Menon and Nath \(2021\)](#) and [Abdul-Razak and Sahoo \(2021\)](#).

First, the panel identifier is not provided in the released data. We work around this by inferring panel numbers on the basis of the documented panel schedule, visit number and the year and quarter of the visit.

Second, in the 2017–18 revisit file for persons, the quarter numbers are recorded as 3, 4 or 5 instead of the expected 2, 3 and 4. Tabulating the visit numbers against the quarter numbers shows no third visits in the reported quarter 3 and no fourth visits in the reported quarter 4. Thus the most plausible explanation is that quarter 2 has been incorrectly coded as quarter 3 and so on. We correct the records with this assumption.

The third, and gravest, difficulty is that an entirely different sets of first-stage unit (FSU) numbers are used in the 2017–18 and the 2018–19 data releases even though both years are supposed to use the same sampling frame. This change is not explained in the PLFS documentation. Because of this change it is not possible to directly link data across these two years.

We develop a matching procedure to overcome this problem. We consider the development of this procedure to be a major contribution of this paper since in the absence of such a procedure it would be impossible to link together panels that began in 2017-18 and were continued into 2018-19. For these panels we would lose the opportunity to track households over the entire four-quarter span of observation.

Our procedure is based on the hypothesis, eventually confirmed, that the change in the FSU number is a simple renumbering and not a reorganization of the FSU and that within the renumbered FSUs the sub-block serial number, second-stage stratum number and sample household numbers still point to the same household.

For each state/union territory (province) and district we consider all possible (FSU number in 2017–18, FSU number in 2018–19) pairs as renumbering candidates. For each such pair we compare households identified by the same sub-block, second-stage stratum and household number and count households of size greater than two that have matching panel number, religion and social group and whose corresponding members have the same sex, relation to the head of the household and general education level. The number of households which successfully match on these criteria becomes the score of the candidate renumbering pair. We exclude households of size one or two to avoid spurious matches. The variables on which we match, specially the general

education level, could presumably change over time, but the proportion of genuinely matching households with such changes is likely to be very small at a quarterly frequency. To minimize the possibility of genuine changes causing failed matches, we consider for each household only the last visit in 2017–18 and the first visit in 2018–19.

For each FSU number in 2017-18 we pick the FSU number in 2018-19 with the highest score obtained in the previous step as the best match. In case a 2018–19 FSU number is selected as the best match for more than one 2017–18 FSU number we drop all the pairs involved.

At the end of this process we are able to find best matches for 4272 out of the 4320 FSUs from 2017–18 that were to be revisited in 2018–19.

To check on the quality of this mapping we use it to rewrite all the FSU numbers in the 2017–18 data and rerun the validation tests from section 3.2. Only 62 out of 34,452 households that were to be matched fail these validation tests. This confirms our hypothesis that FSU were renumbered and not reorganized and demonstrates the procedure’s success in inferring that renumbering.

Note that it is appropriate to validate the procedure by comparing variables like religion or relation to head which were used in the scoring function for FSU renumbering candidate pairs. This is because the validation is run on the entire set of households and not just on the set of matching households which were used for the scoring. Had we been mistaken in our assumption that the FSUs had only been renumbered and not reorganized or had we not correctly inferred the renumbering, the set of matching households would be very small during the scoring stage and subsequently linking all the other households using the FSU renumbering inferred would have led to a large number of households failing the validation tests.

As a further check on this procedure we run it on the 2018-19 and 2019-20 data sets in which no change in FSU numbers actually occurred. Our procedure correctly matches each FSU number in 2018-19 to itself in 2019-20, further confirming its validity.

The rest of this paper uses the data set with the 2017–18 FSU numbers replaced by the inferred 2018–19 renumbering. Households belonging to the few FSUs which could not be matched as taken as attrited at the end of 2017-18.

3.4 Age group and study period

All our analyses are restricted to individuals in the working age group of 15 to 65 years.

We only use data from 2017-Q3 to 2019-Q4 in order to exclude the impact of the COVID pandemic.

3.5 Employment status

The PLFS collects information regarding employment status for three different reference periods—the usual principal/subsidiary status with a reference period of 365 days

Table 2: Employment states

State	Description	PLFS codes
slf-emp	Self-employed in household enterprise or helper in household enterprise	11,12,21
csl-emp	Casual wage labour	41, 42, 51
sal-emp	Regular salaried/wage employee	31
sck-emp	Had work but did not work due to sickness	61, 71
nwrk	Had work but did not work due to other reasons	62, 72
unemp	Unemployed (not engaged in work but available for work)	81, 82
nopart	Not in labour force (not available for work)	91, 92, 93, 94, 95, 97, 98, 99
attrit	Attrited from the sample	

Table 3: Percentage of households attrited in each visit after first

Visit	% of households attrited								
	Panel								
	P11	P12	P13	P14	P15	P16	P17	P18	P21
2	2.6	2.1	2.5	1.8	1.5	1.7	0.0	2.5	0.7
3	4.0	3.6	3.1	2.9	2.7	1.6	3.9	0.2	
4	5.3	4.0	3.8	3.8	2.6	4.7	0.4		

prior to the visit (collected only on the first visit), daily statuses for each day of the week prior to the visit and a current weekly status derived from these daily statuses.

All our analyses are based on the current weekly status. The usual status is not collected on revisits and is therefore not useful in a longitudinal study.

The labour market states I use in the tabulations are obtained by combining PLFS's two-digit status codes (NSSO, 2016) into a smaller number of categories. The list of states and their description is given in Table 2.

3.6 Attrition

Table 3 gives the attrition rates. The attrition rates are comparable to other surveys of a similar nature. In fact some of the attrition rates of less than 1% are quite unusual and surprising.

Still, even these low attrition rates are a cause for concern since their magnitude will

turn out to be comparable to the magnitude of the gross flows between labour market states which are the focus of this paper. Attrition would not make a difference if the histories of those attrited were to be similar to those not attrited. But this is unlikely to be true, since at least one source of attrition is individuals moving from their previous address and the labour market trajectories of these movers are likely to be different from those who continue to stay at their older address.

There exists a literature on adjusting for attrition in longitudinal surveys, and in the context of gross labour flows [Abowd and Zellner \(1985\)](#) and [Feng and Hu \(2013\)](#) for example use such adjustments. But these adjustments require strong assumptions about the unobserved attrition process which we are not comfortable making. We therefore report attrition rates side-by-side with the estimated flow rates so that readers may judge for themselves the worst case bounds on the effects of attrition on the estimates on the lines of [Manski \(2009\)](#).

4 Gross Flows

4.1 Method

In this section, we look at pairs of observations of the same individual over successive quarters to compute gross flow rates between the labour market states defined in [Table 2](#).

For each gender and each period we form matrices of gross flows where the entry in the i -th row and the j -th column in the quarter t matrix is the estimated number of individuals in labour market state i in quarter t who moved to labour market state j in quarter $t + 1$ as a percentage of the total population of that gender in that quarter. Sampling weights provided by the NSSO are used while calculating these percentages. We then average these matrices over all quarters t in our study period to produce [Table 4](#). The numbers in parentheses in the table are transition probabilities expressed as percentages, where the parenthesised entry in the i -th row and j -th column is number of persons in state i who go to state j in the next quarter, as a percentage of the total number of persons in state i in the initial quarter. These probabilities are also averaged over quarters.

Both the row totals and column totals in the table are a measure of the percentage of population in each labour market state, the row totals being the measure of individuals in the state in quarter t and the column totals being a measure of individuals in quarter $t + 1$. The two totals do not exactly match for two reasons. First, if we number our quarters from 1 to N then the row totals are an average over quarters 1 to $N - 1$ while the column totals are an average over quarters 2 to N . Second, the set of individuals included in the calculation between quarter t to $t + 1$ is not the same as the set of individuals included in the calculation between quarter $t + 1$ and $t + 2$ because of households entering and leaving the survey and because of attrition.

Table 4: Occupancy of and gross flow between employment states, numbers in parentheses are transition probabilities as percentages

Emp. status [t]	percentage								
	Emp. status [t+1]								
	slf-emp	csl-emp	sal-emp	unemp	nopart	sck-emp	nwrk	attrit	ALL
<i>Female</i>									
slf-emp	4.66 (79.97)	0.04 (0.66)	0.09 (1.52)	0.05 (0.92)	0.79 (13.53)	0.02 (0.29)	0.09 (1.58)	0.09 (1.52)	5.82 (100.00)
csl-emp	0.06 (3.44)	1.20 (67.27)	0.09 (5.11)	0.08 (4.25)	0.31 (17.66)	0.00 (0.07)	0.00 (0.28)	0.03 (1.91)	1.78 (100.00)
sal-emp	0.08 (0.77)	0.05 (0.52)	8.75 (86.98)	0.10 (1.02)	0.58 (5.72)	0.02 (0.17)	0.17 (1.74)	0.31 (3.09)	10.06 (100.00)
unemp	0.04 (1.78)	0.08 (3.31)	0.10 (4.10)	1.57 (63.60)	0.56 (22.92)	0.00 (0.01)	0.00 (0.12)	0.10 (4.16)	2.46 (100.00)
nopart	0.76 (0.96)	0.26 (0.32)	0.56 (0.71)	0.50 (0.63)	75.56 (95.12)	0.01 (0.01)	0.03 (0.04)	1.76 (2.22)	79.43 (100.00)
sck-emp	0.01 (23.93)	0.00 (0.00)	0.02 (38.05)	0.00 (1.82)	0.01 (17.68)	0.01 (11.50)	0.00 (4.99)	0.00 (2.03)	0.06 (100.00)
nwrk	0.10 (25.87)	0.00 (1.19)	0.16 (42.27)	0.00 (1.29)	0.05 (11.76)	0.00 (0.76)	0.05 (13.71)	0.01 (3.15)	0.39 (100.00)
ALL	5.71 (5.71)	1.63 (1.63)	9.78 (9.78)	2.31 (2.31)	77.85 (77.85)	0.05 (0.05)	0.36 (0.36)	2.31 (2.31)	100.00 (100.00)
<i>Male</i>									
slf-emp	24.13 (90.27)	0.32 (1.22)	0.58 (2.16)	0.36 (1.34)	0.35 (1.31)	0.07 (0.25)	0.32 (1.21)	0.60 (2.23)	26.73 (100.00)
csl-emp	0.35 (3.63)	7.90 (82.06)	0.35 (3.60)	0.48 (4.98)	0.23 (2.39)	0.01 (0.06)	0.04 (0.38)	0.28 (2.90)	9.62 (100.00)
sal-emp	0.56 (1.69)	0.29 (0.87)	30.07 (90.69)	0.51 (1.52)	0.35 (1.04)	0.05 (0.14)	0.28 (0.86)	1.05 (3.18)	33.16 (100.00)
unemp	0.37 (5.50)	0.54 (7.93)	0.49 (7.22)	4.44 (65.29)	0.59 (8.70)	0.00 (0.04)	0.03 (0.37)	0.34 (4.96)	6.80 (100.00)
nopart	0.32 (1.44)	0.21 (0.93)	0.30 (1.33)	0.55 (2.45)	20.49 (90.97)	0.00 (0.02)	0.01 (0.06)	0.63 (2.82)	22.52 (100.00)
sck-emp	0.09 (42.38)	0.01 (4.68)	0.05 (27.04)	0.01 (3.01)	0.01 (5.94)	0.02 (10.76)	0.01 (3.25)	0.01 (2.94)	0.20 (100.00)
nwrk	0.41 (42.18)	0.05 (5.01)	0.29 (30.61)	0.04 (4.40)	0.02 (2.47)	0.01 (0.59)	0.11 (11.16)	0.03 (3.59)	0.96 (100.00)
ALL	26.23 (26.23)	9.31 (9.31)	32.14 (32.14)	6.38 (6.38)	22.04 (22.04)	0.15 (0.15)	0.80 (0.80)	2.94 (2.94)	100.00 (100.00)

4.2 Stylised facts

Table 4 presents us with a number of stylised facts.

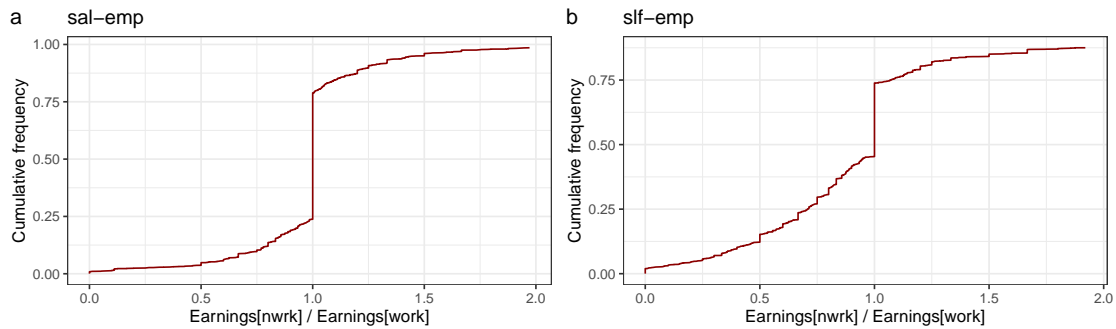
1. *Changes of state are infrequent.* In the two tables most of the flows are concentrated along the diagonal. On an average, in each quarter, only 6.32% of women and 10.68% of men change their labour market state. This has a very important implication for measurement. Sources of measurement error or attrition bias which may be small with respect to the overall sample size may be large enough to swamp the small number of observations of state transitions. Therefore a much higher level of accuracy of observation is required for surveys meant to study transitions compared to surveys meant to study cross-sectional features. It is therefore worrying that the PLFS does not include mechanisms like reinterviews of a subset of respondents a short while after the original interview to quantify measurement errors.
2. *Gross flows are much larger than net flows.* Even though gross flows are small relative to the size of the population, they are much larger than net flows. This indeed is the basic fact motivating the study of gross labour market flows, reconfirmed in our case. For example, on an average in every quarter 0.10% of women moved from salaried employment to unemployment while at the same time 0.10% of women moved back from unemployment to salaried employment. Thus on the net there is no flow from salaried employment to unemployment, though every quarter there is a significant number of women moving between these two employment states. These changes at the individual level which have obvious welfare implications would be completely missed by a study of net flows.
3. *Labour market conditions are very different for men and women.* The low occupancy in employment states for women reconfirms the low labour force participation rate for women in India on which there is already an extensive literature. But the tables bring out the fact that this substantial difference in stocks is accompanied also by a substantial difference in flows in the sense that women experience much large inflows and outflows into employment. For example, the transition probability for women in salaried employment to move into unemployment or nonparticipation is $1.02 + 5.72 = 6.74\%$ whereas for men the corresponding transition probability is $1.52 + 1.04 = 2.56\%$. These high proportionate rates of outflows for women are matched by equally high rates of inflows. Similar comparisons hold for casual work and self-employment. Thus, while women participate less in employment at any given moment of time, there is a much greater flux of women moving into and out of employment. This matches the results of [Sarkar et al. \(2019\)](#) and [Deshpande and Singh \(2021\)](#) who also find that Indian women move into and out of employment at a significantly higher rate than Indian men.

4. *Unemployment and non-participation overlap.* For women, flows into all three categories of employment is higher from non-participation than from unemployment. This could be explained by the very high percentage of women in the non-participation state. However, even for men the flow from non-participation to employment is substantial. For example, 0.30% of the male population moved from non-participation to salaried employment which is not negligible compare to the 0.49% of the male population that moved from unemployment to salaried employment.

These direct movements from non-participation to employment could have two sources. One, it would include workers who complete their job searches between two observations and hence are never observed in the searching state. But, equally important is the possibility of response or measurement errors. Because of the stigma associated with unemployment some of the unemployed may respond that they are not looking for work. Or the responses of some of the non-employed may be incorrectly coded. In any case, our results show that it would be inappropriate to focus excessively on the unemployment rate as a measure of the functioning of the labour market. Among those recorded as non-participating there are many who do wish to participate.

5. *Job-to-job flows are significant.* There are significant gross flows between the three employment states of salaried employment, casual employment and self-employment. So, for example, compared to 0.49% of the male population flowing from unemployment to salaried work each quarter, 0.35% flow from casual work and 0.58% flow from self-employment each quarter. Thus these categories of work do not exist in ironclad compartments and employment-to-employment flows are as important as non-employment-to-employment flows. In fact we know from the studies in countries where matched employer-employee records are available that there exists substantial job-to-job flows. We cannot capture these flows in our data since we have no employer identifiers, but movement between the different kinds of employment gives us a glimpse of these flows. Also to the extent that salaried work is associated more with the formal sector and self-employment and casual work are associated more with the informal sector, these flows show that there exists significant worker porosity between these sectors of the urban Indian economy.
6. *'nwrk' requires further disaggregation* The PLFS uses the employment status codes 62 and 72 for individuals who had work in household enterprises or salaried employment respectively but did not work due to reasons other than sickness. We aggregate these two codes into our 'nwrk' state. There are significant flows to this state from employment states, specially from salaried employment. For example, for women the flow from salaried employment to 'nwrk' is 0.17% of women while the corresponding flow from salaried employment to unemployment is lower at

Figure 1: Earnings before and during ‘nwrk’



0.1%.

While part of these flows would be persons taking a break or vacation, it is quite possible that they also include persons temporarily laid off or not being able to work due to disruptions in production.

To unpack this further we look at individuals who transition from either salaried employment or self-employment to the state ‘nwrk’ and compare their earnings before and after the transition. Figure 1 presents the empirical cumulative distribution function of the ratio of earnings after to earnings before entering the ‘nwrk’ state. There is a large jump in both cases at 1, showing that for a large fraction—around 50% for salaried employment and 25% for self-employed—there is no change in earnings on moving to this state, so the transition is likely to be a voluntary break from work. For around 25% in both cases there is actually a reported increase in income.² And for around 25% in the case of salaried employees and around 50% in the case of the self-employed there is a decrease in income which might indicate some sort of involuntary inactivity.

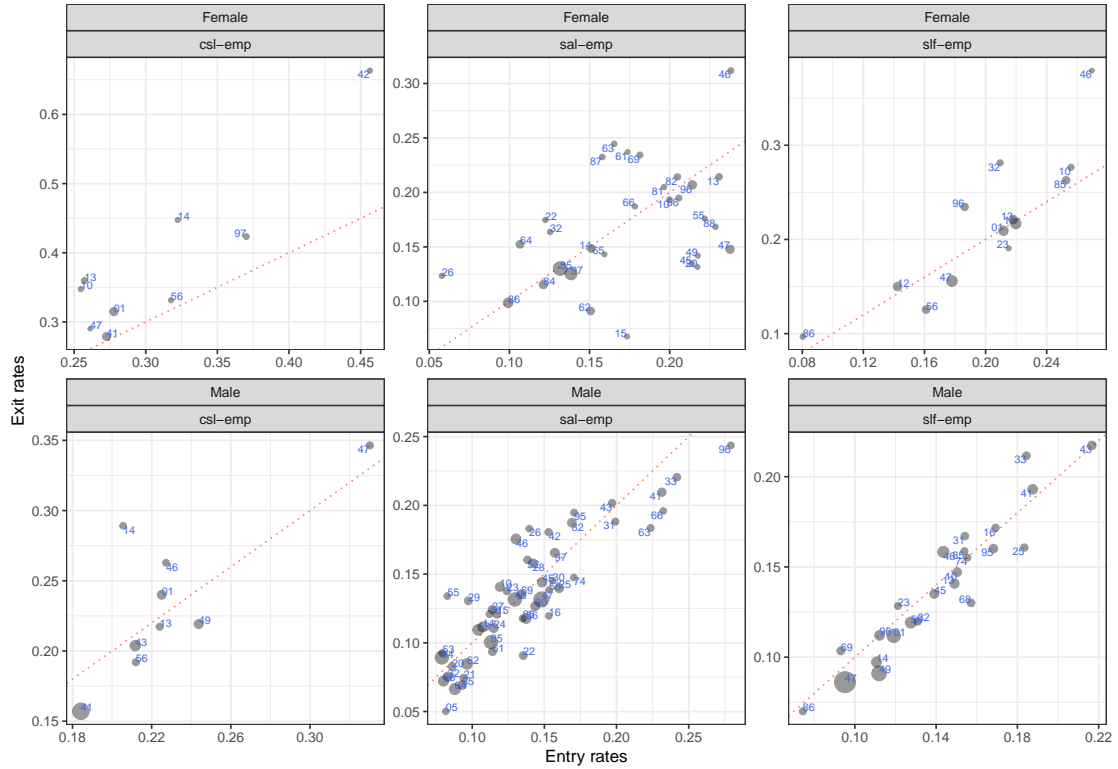
A look at the PLFS data for the post-COVID period shows a substantial increase in flows into ‘nwrk’ and in the proportion of individuals earning nothing after entering this state. Given these indications of heterogeneity it would be useful if future versions of PLFS were to unpack this category further.

5 Diversity in flow rates by industry and region

To compare the labour market transition rates by industry and regions, we use the following definitions, following [Davis and Haltiwanger \(1992\)](#),

²The self-employment category includes unpaid family workers whose incomes are recorded as zero and whose income can only increase when they move to another state.

Figure 2: Industry-level diversity (Entry and exit from industry)



$E_{i,t}$ = set of workers in labour market state i at time t

$\#S$ = number of elements in the set S

$$\text{Entry rate}_{i,t} = \frac{\#(E_{i,t+1} \cap E_{i,t}^c)}{(\#E_{i,t} + \#E_{i,t+1})/2}$$

$$\text{Exit rate}_{i,t} = \frac{\#(E_{i,t} \cap E_{i,t+1}^c)}{(\#E_{i,t} + \#E_{i,t+1})/2}$$

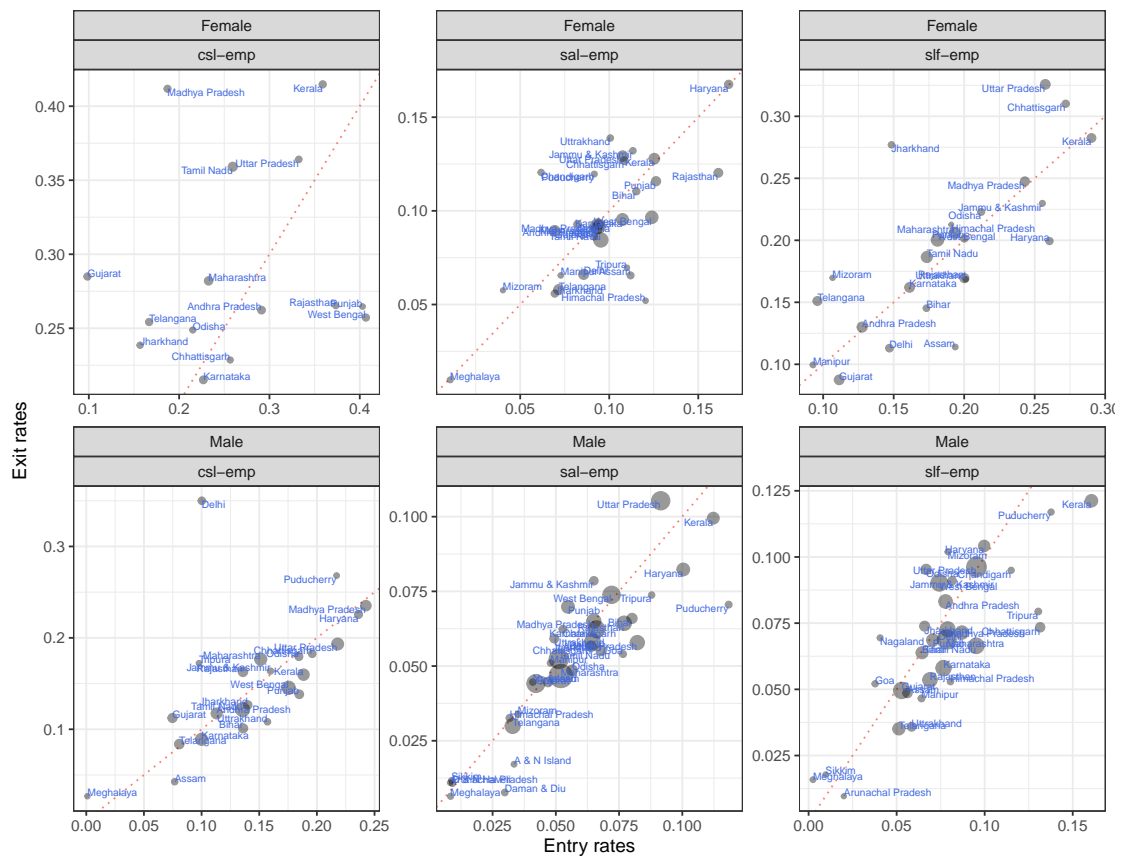
$$\text{Gross flow}_{i,t} = \text{Entry rate}_{i,t} + \text{Exit rate}_{i,t}$$

The definitions of entry rate and exit rates differ from the usual definition of growth rates only in using the average of the occupancy in time t and $t + 1$ in the denominator, instead of the occupancy only at time t . The advantage of these alternative definition is that they produce numbers bounded between 0 and 2, making comparisons easier.

Figure 2 shows the average all-India entry and exit rates for two-digit NIC 2008 industry codes (CSO, 2008). Figure 3 shows the average entry and exit rates from employment for the different states (provinces) of India. In both cases data for men and women and for casual-, salaried- and self-employment are plotted separately.

The first observation from these figures is the high degree of clustering in most cases around the dashed 45-degree line. Industries and states with high entry rates also tend to have high exit rates. Thus, the variation of entry and exit rates across states and

Figure 3: Regional diversity (Entry and exit from work)



industries is not due to the net expansion or shrinkage of industries. Rather it is due to the intrinsic labour market conditions in each industry.

The second observation is that the variation between industries and states is quite large. Even if one looks at a narrowly defined group such as salaried male workers, entry and exit rates range from 0.1 to 0.25 across 2-digit industries and 0.01 to 0.1 for entry and exit from work in different states.

One difficulty in interpreting the two figures separately is that the industry-wise composition of employment differs from state to state. To the extent that labour market transition rates are influenced by both state-specific and industry-specific factors, looking at each separately is subject to confounding. An industry might show high turnover because it is primarily situated in a state with weak labour market protections. On the other hand a state might show a high turnover because most of its workers are employed in industries where hiring and firing is cheap.

To try and decompose the industry and state effects, we estimate the following linear model which imposes an additive structure on these effects:

$$\begin{aligned} [\text{Gross flow rate}] = & \sum_{i,j,k} \beta_{ijk} D_i^I D_j^T D_k^G \\ & + \sum_{l,k} \gamma_{lk} D_l^S D_k^G + [\text{Period dummies}] \end{aligned}$$

Here we use the gross flow rate—the sum of entry and exit rates—as a dependent variable that captures the rate of labour turnover. On the right-hand side are a range of dummy variables

1. D_i^I : industry dummies, where i ranges over 2-digit industry codes.
2. D_j^T : employment type dummies, where j ranges over {slf-emp, csl-emp, sal-emp}.
3. D_k^G : gender dummies, where k ranges over {Male, Female}.
4. D_l^S : state dummies, where l ranges over the states of India

The regression is fit to observations of gross flow rates in each quarter for each state, industry, employment type, and gender combination. To avoid noisy data from combinations with very few sample observations, we only keep those combinations whose total sampling multiplier is greater than 5×10^7 for women and 10^8 for men.

The regression is estimated without an intercept and therefore coefficients for all possible industry- employment type-gender combinations can be estimated. For the state-gender combinations, Maharashtra-Male is taken as the baseline.

Tables 5 and 6 give the statistically significant coefficients for industry, employment type, gender combinations and state, gender combinations respectively.

Even after controlling for state effects, the results of Table 5 shows a wide variation in gross flow rates among industries and forms of employment, with coefficients ranging from 0.26 for male salaried workers in chemical manufacturing to 1.0 for female

Table 5: Regression coefficients, industry and employment type, β_{ijk} (only coefficients significant at 5% level of significance, baseline 0, equation estimated without intercept)

Industry	NIC Code	Emp. Status	Coefficient	
			Female	Male
Manufacture of chemical and chemical products	20	sal-emp		0.26
Food and beverage service activities	56	sal-emp		0.27
Manufacture of food products	10	slf-emp		0.27
Wholesale trade, except of motor vehicles	46	sal-emp		0.30
Land transport and transport via pipelines	49	csl-emp		0.33
Other manufacturing	32	sal-emp		0.34
Land transport and transport via pipelines	49	slf-emp		0.34
Wholesale trade, except of motor vehicles	46	slf-emp		0.39
Computer programming etc	62	sal-emp	0.47	0.32
Security and investigation agencies	80	sal-emp		0.40
Public administration and defence	84	sal-emp	0.50	0.30
Education	85	sal-emp	0.47	0.34
Land transport and transport via pipelines	49	sal-emp		0.41
Manufacture of wearing apparel	14	sal-emp	0.55	0.29
Manufacture of textiles	13	sal-emp	0.49	0.34
Food and beverage service activities	56	slf-emp	0.49	0.36
Human health activities	86	sal-emp	0.43	
Activities of households as employers of domestic personnel	97	sal-emp	0.46	0.41
Other personal service activities	96	sal-emp	0.44	
Retail trade, except of motor vehicles	47	slf-emp	0.54	0.33
Crop and animal production	01	slf-emp	0.52	0.35
Manufacture of wearing apparel	14	slf-emp	0.65	0.27
Financial services, except insurance and pension	64	sal-emp	0.63	0.30
Manufacture of tobacco products	12	slf-emp	0.46	
Retail trade, except of motor vehicles	47	sal-emp	0.55	0.43
Specialized construction activities	43	csl-emp		0.50
Construction of buildings	41	slf-emp		0.51
Manufacture of textiles	13	slf-emp	0.57	0.44
Crop and animal production	01	csl-emp	0.72	0.30
Construction of buildings	41	csl-emp	0.60	0.49
Other personal service activities	96	slf-emp	0.86	0.33
Education	85	slf-emp	0.64	
Manufacture of textiles	13	csl-emp	1.00	
Manufacture of rubber and plastic products	22	sal-emp		1.06

Table 6: Regression coefficients, state γ_{lk} (only coefficients significant at 5% level of significance, baseline Maharashtra-Male)

State	Female	Male
Telangana	-0.16	-0.08
Karnataka	-0.11	
Gujarat	-0.09	-0.05
Maharashtra	-0.08	
Andhra Pradesh	-0.08	-0.06
Tamil Nadu	-0.07	
Madhya Pradesh		0.04
Uttar Pradesh		0.06
Haryana		0.09
Kerala		0.09

casual workers in the manufacture of textiles. Even within the same industry, workers with different forms of employment face different conditions. For example in education, salaried women have a gross flow estimate of 0.47 while self-employed women have a gross flow estimate of 0.64.

When it come to states, the most significant fact about Table 6 is the small number of entries. For most state-gender combinations the estimated coefficient is not statistically significant, confirming that most of the variation among states is driven by differences in industrial composition among states and not state-specific factors.

Unfortunately, there is no discernable pattern among the states whose coefficients are in fact statistically significant. We must conclude that these differences must be explained by idiosyncratic factors.

6 Determinants of job loss and gain

The analysis in the previous sections looked at gross flows between consecutive pairs of quarters. However, our data actually tracks workers for up to four quarters. In this section we use this additional information to look at how labour market states of workers are correlated over longer time spans. At the same time we extend our study of gender differences and see how other worker characteristics influence labour market transitions for men and women.

To keep the analysis tractable we collapse the labour market states into just two: employment and non-employment. We choose not to use the traditional three-way classification of employment, unemployment and non-participation in the light of our observation earlier on the possible conceptual and measurement overlap between unemployment and non-participation.

To track the transitions between these two states we define two dummy variables

Lost and Gained. Lost is 1 for observations in which the individual is employed in that quarter but is not employed in the next quarter, and 0 otherwise. Gained is 1 for observations in which the individual is not employed in that quarter but is employed in the next quarter. Thus Gained and Lost measure movements into and out of employment. For the interpretation of the results to follow it must be remembered that both these variables have the value 0 both for those who are out of employment and remain out of employment and for those who are in employment and remain in employment.

As explanatory variables we use the following worker characteristics:

1. Very.Young: Has the value 1 if the individual's age is less than 21 and 0 otherwise.
2. Young: Has the value 1 if the individual's age is between 21 and 30 and 0 otherwise.
3. Graduate: Has the value 1 if the individual's general education level is graduation or higher and 0 otherwise.
4. Has.Child: Has the value 1 if the individual belongs to a family which has a child less than 5 years of age and 0 otherwise. We have to use this proxy for child-bearing since the data does not provide any information to link parents to children.
5. Married: Has the value 1 if the individual is currently married and 0 otherwise.

We also include as explanatory variables the following two summaries of the individual's employment history:

1. E.Ratio: It is the net fraction of periods in which the individual has been observed to be employed. For worker i in quarter t , let E_{it} be the total number of quarters (not necessarily consecutive) up to and including t in which we have observed them to be employed and N_{it} be the total number of quarters (not necessarily consecutive) up to and including t in which we have observed them not to be employed. Let V_{it} be the visit number (between 1 and 4) for that individual in quarter t . Then we define:

$$\text{E.ratio}_{it} = \frac{E_{it} - N_{it}}{V_{it}}$$

2. EN.Streak For individual i in quarter t , this is the number of consecutive periods up to and including t for which they have been in the same state that they are in quarter t . So for an employed individual this is the number of consecutive periods we have observed them to be employed while for non-employed individuals this is the number of consecutive periods we have observed them not in employment. This is set up as a categorical variable.

Table 7: Determinants of job gain and loss. Average marginal effects from logit model.

	Gained (Female)	Gained (Male)	Lost (Female)	Lost (Male)
Very.Young	−0.015*** (0.001)	−0.054*** (0.004)	0.077*** (0.013)	0.027*** (0.003)
Young	0.001 (0.001)	0.019*** (0.004)	0.022*** (0.005)	0.007*** (0.002)
Graduate	−0.003** (0.001)	−0.028*** (0.002)	−0.043*** (0.004)	−0.012*** (0.001)
Has.Child	−0.005*** (0.001)	0.014*** (0.003)	0.009 (0.005)	0.002 (0.001)
Married	−0.007*** (0.001)	0.037*** (0.005)	0.036*** (0.004)	−0.026*** (0.002)
E.Ratio	0.052*** (0.003)	0.120*** (0.008)	−0.089*** (0.012)	−0.039*** (0.004)
EN.Streak=2	−0.014*** (0.003)	−0.030*** (0.007)	−0.065*** (0.012)	−0.021*** (0.005)
EN.Streak=3	−0.020*** (0.004)	−0.046*** (0.011)	−0.099*** (0.015)	−0.034*** (0.006)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Using these variables we estimate the following equations using logit.

$$\begin{aligned} \text{Lost/Gained} = & 1 + \text{Very.Young} + \text{Young} + \text{Graduate} \\ & + \text{Has.Child} + \text{Married} + \text{E.Ratio} + \text{EN.Streak} \\ & + [\text{Period dummies}] + [\text{Visit no. dummies}] \end{aligned}$$

The equation for Lost is run on the set of individuals in employment while the equation on Gained is run on the set of individuals not in of employment. For each individual, data from visits upto one less than the last recorded visit is used since the variables Lost and Gained are defined by looking one quarter ahead. Observations for all quarters and all workers are pooled together. Each model is estimated separately for men and women.

Table 7 gives the average marginal effects from the logit models. The numbers in parentheses are standard errors.

Being ‘very young’, i.e. being in the age group 15–20, increases the probability of job losses and decreases the probability of job gains. There are two potential explanations for this. One would be the precarity of the job market for young, inexperienced and less-educated workers. The other would be the fact that this is the age range for school and undergraduate education, and individuals in educational institutions in these age groups are more likely to continue education than join employment and those in em-

ployment are likely to leave employment to continue their education. Indeed, in our sample, 81.1% of the 'very young' who are not in employment report attending educational institutions. Of those leaving employment in this age group, 27.55% report attending an educational institution in the next quarter. Thus education does have a role to play. However, even if the equations for Lost are reestimated after dropping observations with employment to education transitions, being 'very young' still has a positive, significant effect, showing that precarity plays a role too.

For the young, i.e. the age group 21–30, the probabilities for both job loss and job gain are higher than the baseline. In this age group the outcomes for men and women diverge significantly. The additional loss probability for women is now more than three times that for men. The additional gain probability is positive and significant for men and very small and statistically insignificant for women. In this age group while men are more likely to enter employment, presumably after completing their education, women are losing employment.

Being a graduate reduces the probability of both job loss and job gain for both men and women. This is presumably because those with graduate and higher degrees have a more secure work life. Those in employment are less likely to lose jobs, and those out of employment are there by choice and not because of unwanted job loss, and hence are less likely to move into employment.

Having a child in the family does not have a statistically significant effect on the probability of job loss. The effect on job gain is significant and in opposite directions for men and women: negative (though small in magnitude) for women and positive for men.

The effect of marriage too is quite different for men and women. Being married reduces the probability of job loss and increases the probability of job gain for men. For women it is the opposite: the probability of job loss goes up and that of job gain goes down.

Overall, the picture that emerges is that of a gender- and age-differentiated labour market with women and the young facing greater employment instability.

When it comes to work histories the EN.Streak variables are negative and significant in all regressions. To recall, these measure that number of consecutive prior quarter in which the individual has been in the same state as in the quarter of observation. Thus this variable has a different meaning in the Gained and Lost regressions. For the Gained equation, estimated on the sample of individuals not in employment, it is the number of consecutive prior quarters they have not been in employment. For the Lost equation, estimated for the sample of individuals in employment, it is the number of consecutive prior quarters they have been in employment. Thus the results show the existence of inertia in job market states. Prolonged periods of non-employment perpetuate non-employment and prolonged periods of employment perpetuate employment. The magnitude of the estimated effects is larger for streaks of length three compared with streaks of length two, showing that this inertia becomes stronger the longer the

worker is in employment or non-employment.

Even after controlling for the streak variables, the variable `E.Ratio` remains statistically and economically significant: negative in job loss regressions and positive in job gain regressions. Thus it is not only consecutive spans of employment and non-employment which affect the probabilities of job gains and losses. Past occurrences of employment make getting back into employment more likely and getting out of employment less likely, even after controlling for the immediately prior history.

7 Conclusion

This study reveals a rich world of micro-level transitions in the urban Indian labour market with a great degree of diversity across genders, industries and regions. It is hoped that as more data becomes available we will be able to better tease out the institutional factors underlying this diversity. Essential to that endeavour would be panel data on jobs from the firm side and the ability to match workers and employers. The introduction of PLFS gives us hope that Indian official data would expand to cover these aspects as well.

References

- Abdul-Razak, S. and Sahoo, S. (2021). Using the rotational panel data from the Periodic Labour Force Survey: A cautionary tale. Available at SSRN.
- Abowd, J. M. and Zellner, A. (1985). Estimating gross labor-force flows. *Journal of Business & Economic Statistics*, 3(3):254–283.
- Ahn, H. J. and Hamilton, J. D. (2021). Measuring labor-force participation and the incidence and duration of unemployment. *Review of Economic Dynamics*.
- Blanchard, O. J., Diamond, P., Hall, R. E., and Murphy, K. (1990). The cyclical behavior of the gross flows of US workers. *Brookings Papers on Economic Activity*, 1990(2):85–155.
- CSO (2008). National industrial classification (all economic activities). Central Statistical Organisation, Government of India.
- Davis, S. J., Faberman, R. J., and Haltiwanger, J. (2006). The flow approach to labor markets: New data sources and micro-macro links. *Journal of Economic Perspectives*, 20(3):3–26.
- Davis, S. J. and Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3):819–863.
- Davis, S. J., Haltiwanger, J. C., and Schuh, S. (1998). *Job Creation and Destruction*. The MIT Press.
- Deshpande, A. and Singh, J. (2021). Dropping out, being pushed out or can't get in? decoding declining labour force participation of Indian women. Ashoka University Economics Discussion Paper 65.
- Elsby, M. W., Hobijn, B., and Şahin, A. (2013). Unemployment dynamics in the OECD. *Review of Economics and Statistics*, 95(2):530–548.
- Feng, S. and Hu, Y. (2013). Misclassification errors and the underestimation of the US unemployment rate. *American Economic Review*, 103(2):1054–70.
- Fujita, S. and Ramey, G. (2009). The cyclicity of separation and job finding rates. *International Economic Review*, 50(2):415–430.
- Hall, R. E. (2005). Job loss, job finding, and unemployment in the US economy over the past fifty years. *NBER Macroeconomics Annual*, 20:101–137.
- Hall, R. E. and Kudlyak, M. (2019). Job-finding and job-losing: A comprehensive model of heterogeneous individual labor-market dynamics. National Bureau of Economic Research, Working Paper No. 25625.

- Manski, C. F. (2009). *Identification for Prediction and Decision*. Harvard University Press.
- Menon, R. and Nath, P. (2021). Where do they come from, where do they go? labour market transitions in India. Azim Premji University, CSE Working Paper No. 42.
- NSSO (2016). Periodic Labour Force Survey: Instructions to field staff, vol I. National Sample Survey Office, Government of India.
- Rogerson, R. and Shimer, R. (2011). Search in macroeconomic models of the labor market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, pages 619–700. Elsevier.
- Sarkar, S., Sahoo, S., and Klasen, S. (2019). Employment transitions of women in India: A panel analysis. *World Development*, 115:291–309.
- Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95(1):25–49.
- Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, 15(2):127–148.