9/19/2016

Assignment-1

Analysis of Female Labour Participation

GROUP-5

Bhupesh Joshi
Amitesh Saha
Kiran Soibam
Roopam Chabara

The State to be removed for the analysis is Delhi (07)

Q1. Estimate a probit model for probability of participation in the labour market.

(a) The explanatory variables are to be used as similar to that in Models A and C of Table 3 of Chatterjee et al paper. Identify the 'common' variables available between your data set and that in their study to carry out this analysis. The 'rank' variable in the paper has to be substituted with URBAN4_2011 qualitative variable with four categories. Let these two models be called as variant 1 and variant 2 respectively. Present the results for the estimated coefficients (Table 2) as well as the marginal effects (Table 3). Before this table of results present a table of mean and standard deviation of the variables used in the model (Table 1).

Sol. The list of variable used in the model along with their description and summary are as follows are as follows:

Table 1. Description of the variables used in the empirical analysis

Variable	*Description*	*Mean*	*SD*
Total number of observations = 75543			
Individual Level			
age	Age	36.39163	19.9032
age_sq	Squared of age	1720.483	1600.1
eduyrs	Numbers of years of schooling	4.2542	4.875634
eduyrs_sq	Schooling Years Squared	61.97215	61.97215
martlst	Marrital status	1.59525	0.809569
Household level			
log_NPERSONS	Log of Household Size	1.653694	0.483565
nchild0_5_sh	Share of Children below 6 years in the household	0.111884	0.144742
nchild6_15_sh	Share of Children 6 years or more in the household	0.163379	0.183373
nadlfdep_sh	Share of female dependents (aged 60+) in the household	0.400617	0.164692
nadlmdep_sh	Share of male dependents (aged 60+) in the household	0.329272	0.163656
eduadlt_cd	Maximum schooling of household	3.773295	1.659074
incsohh	Occupation of head of household	3.430072	2.081966
dst	Household belong to Scheduled Tribes or not	0.088294	0.283724
dsc	Household belong to Scheduled Castes or not	0.206849	0.405049
dobc	Household belong to Other Backward Castes or not	0.413857	0.492527
duch	Household belong to Bhramin caste	0.049098	0.216074
dosgr	Household belong to Others	0.241902	0.428238
dhindu	Household belongs to Hindu religion	0.80525	0.396011
dmuslim	Household belongs to Muslim religion	0.132362	0.338886
dchrstn	Household belongs to Christian religion	0.026157	0.159604
djain dorelg dsikh	Household belongs to Jain , other , sikh religion		
Location			

URBAN4_2011	Population from urban or rural	2.003177	0.91265
Employment			
psu_sh_wkfarm	Share of Agricultural Self-employed and Regular workers aged 15+ in working age population	0.206152	0.174569
psu_sh_wkagwag	Share of farmed labor aged 15+ in working age population	0.109882	0.128965
psu_sh_wkbsns	Share of Non-Farm Self Employed aged 15+ in working age population (business)	0.123669	0.152466
psu_sh_wknonag	Share of Non-Farm Regular Wage aged 15+ in working age population	0.174431	0.144335
psu_sh_wknrega	Share of All Casual workers aged 15+ in working age population (NAREGA)	0.024981	0.046415
psu_sh_wksalry	Share of Salaried workers aged 15+ in working age population	0.184287	0.217865
psu_sh_wkanim	Share of participation for animals workers aged 15+ in working age population	0.176598	0.121921
ci_terc	Confidence in institution	1.989529	0.813031
ai_index	Economic status	0.096829	2.568423

For estimation of probit model for probability of participation of women in the labor market we use the following two variants of the model.

$$LFP = f_1(Individual, Household)$$

And

$$LFP = f_2(Individual, Household, \text{Location}, \text{Employment})$$

Upon estimation we get the coefficients of dependent variable as following:

<u>Table 2. Coefficient's for working-age women based Different models</u>

Dependent Variable: Labor force participation				
Independent Variable	***Variant 1***	***Variant 2***	***Variant 3***	***Note***
Pseudo R2	0.1676	0.2382	0.2402	
Constant	-3.209451	-5.074999	-5.084108	
age	0.1350759	.1537483	.1530292	
age_sq	-0.0016495	-0.00183	-.0018239	
eduyrs	-0.0647286	-0.0599		
eduyrs_sq	0.0019499	0.003501		
martlst2	-0.0675577	-0.04304	-.0441904	martlst1 = married as base
martlst3	0.4246627	0.536099	.5283247	
log_NPERSONS	-0.2486973	-0.28566	-.2839304	
nchild0_5_sh	0.7961358	0.757835	.7639516	
nchild6_15_sh	1.312407	1.337841	1.337738	

nadlfdep_sh	1.207782	1.159895	1.168479	
nadlmdep_sh	0.3793375	0.375284	.3778551	
dst	0.3942046	0.171344	.1725641	dsc as base dummy
dobc	-0.0272393**	-0.10055	-.1033091	
duch	-0.3565557	-0.30511	-.3154585	
dosgr	-0.1608891	-0.2246	-.2296929	
dhindu	0.2873151	0.140603	.140375	djain, dorelg dsikh as base dummy
dmuslim	-0.2324797	-0.15582	-.1571761	
dchrstn	0.1283121	0.126141	.1214068	
URBAN4_20112		0.061219***	.0586 **	URBAN4_20111= Urban metro as base dummy
URBAN4_20113		-0.02592*	-.025331 *	
URBAN4_20114		-0.0212*	-.0229254 *	
psu_sh_wkfarm		3.44794	3.445323	
psu_sh_wkagwag		2.316132	2.303227	
psu_sh_wkbsns		1.135435	1.11375	
psu_sh_wknonag		1.235872	1.229981	
psu_sh_wknrega		4.355172	4.349915	
psu_sh_wksalry		0.813235	.8058674	
psu_sh_wkanim		Removed because of collinearity		
edu_cd2			.0270003 *	eduadlt_cd1 = Not Literate as base dummy
edu_cd3			-.1429303	
edu_cd4			-.358584	
edu_cd5			-.2641858	
edu_cd6			.0287482*	
* insignificant at 10% level of significance ** insignificant at 5% level of significance *** insignificant at 1% level of significance				

Table 3. Marginal probability effects for working-age women based on Usual Status

Dependent Variable: Labor force participation				
Independent Variable	***Model 1***	***Model 2a***	***Model 2b***	***Note***
Pseudo R2	0.2402	0.2403	0.2407	
age	.0496821	.0497059	.049751	
age_sq	-.0005922	-.0005923	-.0005925	
martlst2	-.0142154	- .0141993	-.0144272	martlst1 = married as base
martlst3	.1855298	.1857363	.185712	
log_NPERSONS	-.0921801	- .0922725	-.0919465	
nchild0_5_sh	.2480226	.2478434	.2455033	
nchild6_15_sh	.4343067	.4337885	.4317105	
nadlfdep_sh	.3793554	.3790096	.378305	
nadlmdep_sh	.1226735	.1222927	.122187	

dst	.0584418	.0585869	.0601613	dsc as base dummy
dobc	-.0333391	- .0338082	-.0335338	
duch	-.0923898	-.0932435	-.0940326	
dosgr	-.0715808	- .0720184	-.0712115	
dhindu	.0442773	.0440155	.045821	djain, dorelg dsikh as base dummy
dmuslim	-.0490624	-.0492369	-.0464275	
dchrstn	.0408143	.0421223	.0499276	
Location Dummy				
Other Metro	.0191863	.0195858*	.0231773	URBAN4_20111= Urban metro as base dummy
More dev. Village	-.0081985	-.0075083***	-.0029012**	
Less dev. Village	-.0074286	- .0068188***	-.0028058**	
psu_sh_wkfarm	1.11855	1.119913	1.118696	
psu_sh_wkagwag	.7477598	.7507863	.7601442	
psu_sh_wkbsns	.3615875	.3622278	.3564842	
psu_sh_wknonag	.3993224	.4027787	.4089087	
psu_sh_wknrega	1.412233	1.417257	1.421005	
psu_sh_wksalry	.2616309	.2642774	.2725031	
Education categories (Illiterate as base)				
Primary	.0088307	.0090604***	.0096399***	edu_cd1 = Not Literate as base dummy
Middle	-.0450894	- .0446576	-.0437717	
Secondary	-.1057243	- .1050931	-.1030949	
Higher Secondary	-.07890	- .0781912	-.0765463	
Post Hi. Secondary	.0094092	.0100943***	.0120294***	
Quartile of Confidence in Institution (low being the base category) citerc_pcag2				
Medium	-	.001204***	-	Low confidence as base category
High	-	- .0088942	-	
Quartiles of Confidence in Institution aggregated at PSU level				
psu_sh_cig1	-	-	.0160693	
psu_sh_cig2	-	-	.0666991	
psu_sh_cig3	Removed because of collinearity			
* insignificant at 10% level of significance ** insignificant at 5% level of significance *** insignificant at 1% level of significance				

Q 1.a. How will you conclude which variant fits your data better? What are the changes in the results between these two versions of the model?

Sol. Pseudo R^2 is one of the measure to know the goodness of fit of a model. As clear from Table 2 Variant 2 fits the data better. The Pseudo R^2 increase to 0.24 form 0.17 on addition for new variables of employment and location. The overall explanatory power of Variant 2 is much higher than that of Variant 1 suggesting that the local structure of employment matters.

This all signifies that though the new model fits the data better but the participation rate of women along the rural-urban gradation are not significantly different (at 1% LoS). The coefficients of

URBAN4_20113 and URBAN4_20114 are highly insignificant means that with local employment structure being taken into account there is virtually no difference in the participation rates between more developed villages, less developed village and urban metro cities.

Marginal effect is defined as, how much the predicted probability of labor force participation changes for a unit increase in the corresponding explanatory variable.

Between the two variants of the model most of the variables have similar effect with insignificant changes. The marginal effects of age and education represent an Inverted-U shaped. The predicted probability of female labor participation first decreases with an increasing rate with years of schooling (increases for age) and then decreases at a decreasing rate. The marginal effect of marriage is positive and significant. Married women are less likely to participate according to both the models. Having more children or more dependent people in household does effect the participation rate of women in labor force.
There is a positive marginal effect of caste for ST when compared to SC's, however negative for others. The marginal effect of OBC's becomes significant in Variant 2. Muslim women are less likely to participate in workforce as compared to other religions.

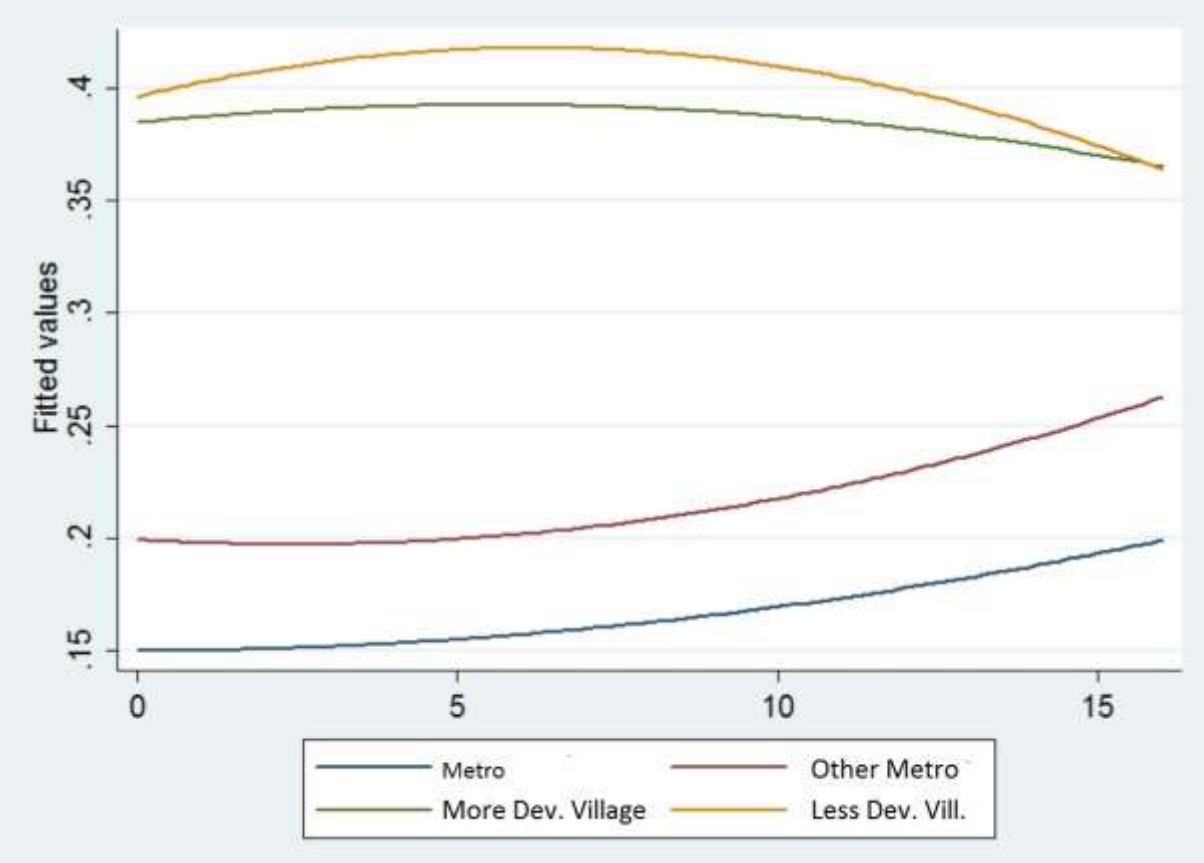
Q.1.b. Choose the best fitting version and instead of education in years and its square as continuous variables, replace it with the dummy variables for level of education. Let this be variant 3. Compare the results for marginal effects (add another column in Table 3 to make this comparison).

i) Assuming that variant 2 fits better than variant 1 (or else use variant 1 for comparison) how will you interpret the effect of education on female work participation? Will the interpretation be the same in both variants?

Sol. Replacing the education in years and its square with the dummy variables for level of education, improves the overall explanatory power of the model only marginally. The change in marginal effects of all the explanatory variables is very small and similar (in direction).
The rate of marginal effect of education increases till secondary education and again falls after that. This means that with an increase in level of education from primary to middle to secondary the predicted probability of female labor participation in the workforce goes decreasing at an increasing rate (till secondary education), however the rate of decrease of predicted probability of female labor participation with an increase in level of education falls after the secondary education. This confirms the inverted U- shaped curve of effect of education on the labor participation of women.
Yes, the interpretation will be similar for both the variants (Variant 3 and Variant 2). As evident from Variant 2 the variable square of years of education (eduyrs_sq) which tells us the rate of change of marginal effect is positive which means that the predicted probability of female labor participation first decreases with an increasing rate with years of schooling (increases for age) and then decreases at a decreasing rate. However, it would be difficult to tell the level of education post which the direction of rate of change of marginal effect changes in Variant 2.

(ii) If you have to compare the effect of education, for differences in place of residence (URBAN4_2011 variable), which variant will you choose. Illustrate your justification using the graphical approach.

Sol.



For Urban areas (Metro and other Urban) the effect of education is U-Shaped in nature which means that with an increase in years of education the predicted probability of participation of female in the labor force first decreases with increase in years in education and then increases with an increase in years of education. This is because in urban areas there are more skilled jobs available which leads to increased participation of women with increase in education.

For rural areas (More and less developed villages) the effect of education is an inverted U shaped curve. This means that the predicted probability of participation of female in the labor force first increases with increase in years in education and then decreases with an increase in years of education. This is because in the rural areas there are more agricultural and daily wage job which does not requires much education, therefore an increase with increase in years of education. But since there is less supply of educated jobs (or white collared jobs) therefor with an increase in education after a certain limit the predicted probability of female labor participation decreases. Hence an inverted U shaped curve.

Q. 2) After you have chosen the 'best fitting' model from the results in 1 above, keep that as the benchmark model and call it say model 1.

Sol. The best fitted model based on Pseudo R^2 comes out to be Variant 3. Let it be called Model 1

$$LFP = m_1 (Individual, Household, \text{Location, Employment})$$

a) In model 1 add the variables that capture the perception of the confidence of institutions (ci) at the individual level, let this be model 2a.

Sol. Let the model be:

$$LFP = \mathrm{m}_{2a}\ (Individual, Household, \text{Location, Employment, CI})$$

For model 2a if the confidence in institution increases from low to medium there is no change in marginal effect i.e. there is no change in predicted probability of labor force participation for women does not changes with an increased confidence in institution from low to medium. However, with the increase in confidence in institutions from low to high there is a decrease in predicted probability of labor force participation for women.

b) Estimate another model 2b with the perceptions of the confidence of institutions aggregated at the 'psu' level (see variable names above) instead of the (ci) in 2a. What is your conclusion about the choice between the models 2a and 2b? Why do you think, the results are different even though you are using one or the other form of 'ci'?

Sol. Let the model 2b be:

$$LFP = \mathrm{m}_{2b}\ (Individual, Household, \text{Location, Employment, PSUCI})$$

The estimated marginal effects are presented in Table 4. Model 2b is a better fit for the data based on pseudo R^2

In contrast with the result presented by the dummy variable for confidence in institution the variable which aggregates the confidence at "PSU" level gives a clear result. The PSU which has higher confidence in institutions has a greater positive marginal effect. This means that the increase predicted probability of labor force participation for women will be more for a similar increase in share of confidence at PSU level for a higher confidence level.

i.e.

$$\frac{\partial Pr}{\partial PSUCI(medium)} > \frac{\partial Pr}{\partial PSUCI(low)}$$

This can be because a safe village or locality provides a greater sense of security to women and hence an incentive to move out to participate in the work force as compared as compared to the women in unsafe locality.

The results are different in case of individual confidence in institution and the one aggregated at the PSU level because

1. The decision of participation also depends on society; hence a safe locality is will provide more incentive for women to work.
2. The individual confidence in institution will not be good explanatory variable because ours being a patriarchal society a lot depends on how the society/community feels as a whole.

Q. 3) Test for equality of coefficients of the share of workers in the different sectors at the psu level using the final model in 2b above. What is your conclusion?

Sol. Let the hypothesis to be tested be

H0: psu_sh_wkagwag=psu_sh_wkbsns=psu_sh_wkfarm=psu_sh_wknonag=psu_sh_wknrega=psu_sh_wksalry
H1: not H0

The result are as follows:

$Chi^2(6) = 3956.79$

$Prob > chi^2 = 0.0000$

Since we can reject the null hypothesis base on the result of the test we can safely say that coefficients of the share of workers in the different sectors at the psu level is not equal.

Q.4) (a) Add state level dummies in model 2b, how do the results change and what do you conclude from that?

Sol. Let model to be estimated be

$$LFP = \mathrm{m}_{2b.1}\ (Individual, Household, \text{Location, Employment, PSUCI, STATEDUMMY})$$

Since the model is unable to converge using a probit model we use logit model to estimate the data.

Table 4. Marginal probability effects for working-age women based on different models

	Dependent Variable: Labor force participation				
Independent Variable	***Model 2b.1***	***Model 2b.2***	***Model 2b.3***	***Model 2b.4***	***Logit Model***
Pseudo R^2	0.2591	0.2653	0.2601	0.2678	0.2693
age	.0506998	.050869	.0506884	.0507291	.0511995
age_sq	-.0006025	-.0006	-.0006019	-.0005998	-.0006086
martlst2	-.0162759	-.0170265*	-.0175071*	-.0050065*	-.0039017*
martlst3	.1953009	.1824494	.1974978	.1946193	.206005
log_NPERSONS	-.0925065	-.0647019	-.0916705	-.1047042	-.102549
nchild0_5_sh	.2408004	.1939735	.2510177	.2569281	.2550769
nchild6_15_sh	.4295367	.3817568	.4412756	.4303025	.4118138
nadlfdep_sh	.3568446	.3393623	.3578733	.3685746	.3565581
nadlmdep_sh	.1122072	.1013628	.1077252	.0901879	.0837081
dst	.0552894	.048238	.0601329	.0516874	.0484026
dobc	-.0401047	-.027812	-.0398724	-.0381104	-.0363249
duch	-.1053187	-.0858105	-.1056498	-.0951357	-.0882425
dosgr	-.0721413	-.0525928	-.0726242	-.0711827	-.0687053
dhindu	.0158285*	.0115654*	.0160747*	.0170536*	.0154739*

dmuslim	-.0650167	-.0740976	-.0638513	-.0593223	-.0592168
dchrstn	.03651***	.037085***	.03552***	.0378898***	.035713***
Location dummy (Urban metro as base)- URBAN4_20111					
Other Metro	.0134764*	.0099355*	.0135267*	.013751*	.0173958*
More Developed Vill.	.03093***	.02159***	.03179***	.0325046***	.0365994***
Less Developed Vill.	.0498154	.0330958	.0522624	.0492381	.0534801
psu_sh_wkfarm	1.013565	.9988665	1.01626	.9663617	.938784
psu_sh_wkagwag	.6787889	.6440338	.6792125	.6055075	.5890367
psu_sh_wkbsns	.3556934	.3861295	.3488376	.375717	.3667733
psu_sh_wknonag	.4686801	.4488457	.4614254	.4772153	.4677489
psu_sh_wknrega	.9010827	.8794441	.911859	.8395714	.8093843
psu_sh_wksalry	.2678482	.3113495	.262725	.3209503	.3191584
Education categories (Illiterate as base)- educd1					
Primary	.0200431	.0330932	.0195625	0239522	.01987
Middle	-.0467072	-.0260219	-.0478043	-.0377675	-.0386227
Secondary	-.0974718	-.0669919	-.098723	-.0854138	-.0826185
Higher Secondary	-.0775924	-.0364117	-.0788721	-.0605033	-.060796
Post Hi. Secondary	.0157374	.0802976	.0154873	.0434116	.0392568
psu_sh_cig1	-.0315084	-.0314829	-.0309457	-.0376236	-.0364568
psu_sh_cig2	.0239036	.0222513	.0261267	.0279023	.0266681***
Per Capita Income Quartiles Dummy (lowest as base and rest in increasing order)- pcincq1					
pcincq2	-	-	-.0341024	-	-
pcincq3	-	-	-.0000855*	-	-
pcincq4	-	-	.01563***	-	-
pcincq5	-	-	.013513***	-	-
Occupation of House hold head (Agri.&Alld as base category)					
Ag. Labor	-	-	-	.03407	.0320949
Non Ag. Labor	-	-	-	-.0346531	-.0334315
Artisan	-	-	-	-.0790852	-.0773272
Business	-	-	-	-.0591741	-.058477
Salaried	-	-	-	-.0960455	-.0918807
Others	-	-	-	-.1670373	-.15667
Asset index quartiles (lowest as base and rest in increasing order)					
quin_g2	-	-.0444987	-	-	-
quin_g3	-	-.0559558	-	-	-
quin_g4	-	-.1099939	-	-	-
quin_g5	-	-.1567401	-	-	-
State level dummy (J & K as base dummy)					
Himanchal	.1754141	.1942371	.1746561	.187446	.1912209
Punjab	-.0738569	-.0443969	-.0736912	-.0870787	-.0795959
Chandigarh	-.0120676*	-.0073789*	-.0069464*	-.0216977*	-.0235818*

Uttarakhand	.061655***	.066272***	.065222***	.0477892***	.0471613
Haryana	-.0483456	-.0292576	-.0479025	-.0630038	-.0593347
Rajhasthan	.04694***	.04569***	.04892***	.0357682***	.03606***
UP	-.0515701	-.0733313	-.0427309	-.0674088	-.0651153
Bihar	-.115539	-.1380208	-.1071808	-.1234662	-.1160353
Sikkim	.0352862*	.0142124*	.0366216*	.0224212*	.0182623*
Arunachal Pradesh	-.05799***	-.05578***	-.05939***	-.069971***	-.0629537***
Nagaland	-.055763**	-.022046**	-.052814**	-.0576933**	-.0555479*
Manipur	-.1426348	-.1574723	-.1417866	-.1405473	-.1377183
Mizoram	.120944***	.184541***	.119435***	.1282774***	.1359614***
Tripura	-.1434532	-.1547056	-.1432161	-.14656	-.1357531
Meghalaya	-.0165953*	-.026733*	-.0184691*	-.0352754*	-.0248714*
Assam	-.1461169	-.1595601	-.1442356	-.1557109	-.1458377
West Bengal	-.108297	-.1267054	-.1033035	-.1228257	-.1162541
Jharkhand	-.1219358	-.1364067	-.1150724	-.1291118	-.1199253
Orissa	-.1337112	-.1489734	-.1254781	-.1444297	-.1328142
Chhattisgarh	.0836342	.0749253	.097234	.0623845	.0620643
MP	.0455215**	.0274796**	.0552737**	.0264273**	.0258247**
Gujarat	-.01028***	-.00793***	-.00490***	-.032081***	-.0305251***
Daman & Diu	-.07309***	-.05628***	-.07451***	-.088201***	-.0821211***
Dadar and Nagar Haveli	-.0174334*	-.0103622*	-.0141857*	-.0255093*	-.020662*
Maharashtra	-.006987**	-.007113**	-.004763**	-.0279324**	-.0241001**
Andhra	.0545934**	.0507432**	.0590095**	.0291592**	.0302057**
Karnataka	.025579*	.0195057*	.0287632*	-.0029059*	-.0011622*
Goa	-.1600978	-.1613169	-.1586516	-.1658217	-.1560763
Kerala	-.0779566	-.0614025	-.0750886	-.0936765	-.0841097
Tamil Nadu	-.0042681*	.0109011*	-.002452*	-.0255256*	-.0232058*
Pondicherry	-.0678024	-.0404916	-.06743	-.0930748	-.0810222

* insignificant at 10% level of significance
** insignificant at 5% level of significance
*** insignificant at 1% level of significance

After addition of state level dummies, the pseudo R^2 increases. The location of an individual have an effect on the female labor participation rate.
The addition of state level dummy provides a better fit for the data.

Q.4 b) Now add the variables that capture household's economic status. How do the results change and why do they change? Does the paper have this variable?
Sol. Let us capture the economic status of the household using 5 quartiles.

LFP = $m_{2b.2}$ ($Individual$, $Household$, Location, Employment, PSUCI, STATEDUMMY, Economic Status (asset ownership))

The pseudo R^2 increases after addition of the variable explaining the effect of Economic Status of a household in model. This means that the model 2b.2 is better fit for data as compared to model 2b.1. There is no significant difference in the marginal effects of the various independent variables between the two model.
The marginal effect decreases (increases in magnitude) as the economic status of household increases. This means that the change predicted probability of participation of female in labor force further goes on decreasing as the economic status of the household increases from the lowest category.

$$mag\left(\frac{\partial Pr}{\partial Status(low)}\right) < mag\left(\frac{\partial Pr}{\partial Status(medium)}\right) < mag\left(\frac{\partial Pr}{\partial Status(high)}\right) < mag\left(\frac{\partial Pr}{\partial Status(V.High)}\right)$$

All the marginal effects have negative direction. As the economic status of the household increase the predicted probability of participation of female in labor force decreases because for a lower income group the need for participation in workforce is more. As the economic status increases the patriarchal society treats women as unequal's and does not favors its participation in the labor force.
Yes, the paper has this variable. The economic status of household is captured by the land size in the paper.

Q.4 c) Instead of quintiles of asset index which captures economic status, you want to use quintiles of per capita income (pcincq). Note that both these variables (asset or per capita income) are categorical in nature now. Is there any reason that the authors give on what they do not include 'income' variable in their model? How is the asset index variable different from per capita income variable that you can use to counter the arguments given in the paper to include it in the model?

Sol. Let call this model as model 2b.3

LFP = $m_{2b.3}$ ($Individual$, $Household$, Location, Employment, PSUCI, STATEDUMMY, Economic Status (per capita income))

As evident from the above Table the pseudo R^2 decreases if we replace the variable that captures the economic status by quintiles of asset index with quintiles of per capita income (pcincq). This means that per capita income is not a good measure to capture the economic status of the household. Further the rate of decrease of marginal effect decreases with an increase in per capita income. This means that as the per capita income increases the predicted probability of participation of female in labor force increases.
According to author income is not the best measure to know the economic status on an individual. The income-effect hypothesis is weakened by the fact that there is a large gap in female LFPR between rural and urban households with similar living standards. An additional reason to be skeptical about the income effect-hypothesis comes from short-term economic fluctuations due to natural disasters or unforeseen circumstances (also seasonal fluctuation).

The other reason can be that often respondent does not reveal their true income. The income of an individual is usually underreported because of psychological reasons as well fear of government. The biasness in the reporting of income variable creates unexpected results as evident from the model 2b.3.

However, ownership of assets is a better measure to derive the economic status of the household because it is accurately reported and is unbiased in nature. The asset ownership is not subjected to short- term fluctuation. The economic status of a household is aggregated using multiple assets owned by a household. This gives a fair estimate of the economic status of the household.

Q.4 d) What happens if you use the variable 'occupation of the head of the household' instead of the quintiles of asset index variables? Which of the two models will you choose between (b) and (d) based on your results? Will it help you better in supporting the author's contention as discussed in (c) and in the paper?
Sol. Let the estimated model 2b.4 be:

LFP = $\mathrm{m}_{2b.4}$ ($Individual$, $Household$, Location, Employment, PSUCI, STATEDUMMY, Occupation of the head of the household)

The marginal effects are presented in Table 4.

The pseudo R^2 increases when we compare it with model 2b.2 (presented in Q4. b). This means that the model is better fit as compared to model 2b.2. Yes, this helps us in supporting the authors contention discussed in part c. As evident the occupation of household is able to better explain the economic status of the household as compared to per capita income. This is mainly fact that in most of the cases the head of the household is main source of earning. The occupation is not subjected to seasonal or short term fluctuation.
It can also be said that there is no biasness reporting of occupation status of the household. As evident the marginal effect is high and negative for high earning profession which depicts the higher economic status of household.
Therefor income is not the best measure to know the economic status of the household.

Q. 5) Why do you prefer to present the results as 'marginal effects' and not the estimated coefficients? Based on the analysis above given your data set, what do you conclude about the role of supply side and demand side variables in influencing women's participation in the workforce? The paper will give an idea of what are the supply side and demand side variables.

Sol. We know that for a model given by:

$$Y_i = F(x_i, \beta) + \epsilon_i$$

the marginal effect for a probit model is given by:

$$\frac{\partial Pr}{\partial x_{ij}} = f(x_i, \beta) * \beta$$

Since the marginal effects are not independent of x_i therefore we present the result of the models as marginal effects rather than the coefficients. The direction of marginal effect will be similar to that of the coefficient but not the magnitude.

The analysis of the data set shows a considerable effect of demand-side variables. With only supply side variables like age, education, economic status etc. the explanatory power of the model is low. I increase once we include the demand-side variables like local employment structure, location etc. In the analysis the variant 1 represent the standard supply side model whereas the variant 2 incorporates the demand side of the model as well. The marginal effect of demand side variables is high as compared to the supply side variable as evident from Table 3. This suggests that a unit change in the demand side variables changes the predicted probability more as compared to the supply side variables. Therefore, it is necessary to study the demand side variables along with the supply side variables affecting the participation of female in labor force.

Q. 6) Based on the final set of explanatory variables, estimate a logit model. Do you find that the logit model fits that data better than the probit model? Justify your answer with suitable results/analysis.

Sol. Yes, the Logit model does fit the data better than the probit model. As evident from Table 4 (Logit) the pseudo R^2 is more for logit model as compared to the probit model. Further as evident from Table 4 most of the coefficient of variables (and the marginal effect) are overestimated in case of probit model.

Q.7) Use the odds ratio to:

a) Compare between Muslim and 'other' religion women

Sol. Odds ratio is defined as

$$\Omega(x_i) = \frac{p_i}{1 - p_i}$$

For a discrete change in x the odds ratio is defined as

$$\frac{\Omega(x_i, x_j + \delta)}{\Omega(x_i, x_j)} = e^{\beta_j * \delta}$$

The odds for female participating in the labor force decreases by O.7070152 for a Muslim woman when compared to "other" religion women holding all other variables constant.

b) Compare between Muslim and Hindu women.

Sol. $$Odds\ ratio = \frac{Odds\ ratio\ of\ hindu\ given\ others}{Odds\ ratio\ of\ muslim\ given\ others}$$

$$Odds\ ratio = \frac{1.098}{0.707} = 1.55$$

The odd for female participating in the labor force is 1.55 times higher for a Hindu woman as compared to a Muslim woman holding all other variables constant.

c) Interpret the variable 'children in 0-5 age group'.

Sol. We know that the coefficient of the logit model for 'children in 0-5 age group' is 1.418353. It can be said that
For an additional child aged between 0-5 we expect the logit (log of Odds) to change by 1.418353.

Section II: Multiple Response Models

Q.1) Use the variable wrk_3cat to estimate a multinomial logit model with the explanatory variables chosen on the basis of the final results in the previous section. Choose 'NILF' as the omitted category. Interpret the results based on the age, education, 'ci' variables and 'psu share of workers in different occupation' using RRR.
Sol. The pseudo R^2 for the estimated model is 0.2468. Relative risk ratio (RRR) is defined as

$$RRR = \frac{\Omega_{m/n}(x_{ij})}{\Omega_{m/n}(x_{ij}+\delta)} = e^{\beta_{jm}-\beta_{in}}$$

Age: With an increase of age by one year the ratio of odds of an individual with part-time job relative to that of the one not in labor force is 1.309504, whereas for an individual with part-time job relative to that of the one not in labor force is 1.410664.

Education: 1. With a shift of education from illiterate to primary ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 1.123254, whereas for an individual being in full-time job relative to that of the one not in labor force is 1.153636.
2. With a shift of education from illiterate to middle school ratio of odds of an individual being in part-time job relative to that of the one not in labor force is .8252517, whereas for an individual being in full-time job relative to that of the one not in labor force is .7713139.
3. With a shift of education from illiterate to secondary ratio of odds of an individual being in part-time job relative to that of the one not in labor force is .6065889, whereas for an individual being in full-time job relative to that of the one not in labor force is .6521608.
4. With a shift of education from illiterate to higher secondary ratio of odds of an individual being in part-time job relative to that of the one not in labor force is .6065889, whereas for an individual being in full-time job relative to that of the one not in labor force is 1.04036.
5. With a shift of education from illiterate to post higher secondary ratio of odds of an individual being in part-time job relative to that of the one not in labor force is .7675157, whereas for an individual being in full-time job relative to that of the one not in labor force is 2.374404.

Confidence in Institutions: 1. With 1% increase in confidence in institutions of the "PSU" at low category the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is .8348129, whereas for an individual being in full-time job relative to that of the one not in labor force is .8153902.

2. With 1% increase in confidence in institutions of the “PSU” at medium category the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 1.250834, whereas for an individual being in full-time job relative to that of the one not in labor force is .8733464.

Psu share of workers in different occupation: 1. With 1% increase in the share of workers in farm labors at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 152.1737, whereas for an individual being in full-time job relative to that of the one not in labor force is 220.3746.
2. With 1% increase in the share of workers in agriculture wage labors at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 19.70777, whereas for an individual being in full-time job relative to that of the one not in labor force is 67.39558.
3. With 1% increase in the share of workers in business at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 5.80482, whereas for an individual being in full-time job relative to that of the one not in labor force is 19.39243.
4. With 1% increase in the share of workers in non-agriculture wage labors at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 10.5531, whereas for an individual being in full-time job relative to that of the one not in labor force is 32.64917.
5. With 1% increase in the share of workers in NEREGA labors at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 74.72612, whereas for an individual being in full-time job relative to that of the one not in labor force is 50.36668.
6. With 1% increase in the share of workers in Salaried labors at a PSU level the ratio of odds of an individual being in part-time job relative to that of the one not in labor force is 2.651867, whereas for an individual being in full-time job relative to that of the one not in labor force is 25.36086.

Q.2) Test the hypothesis that
(a) ‘education’ has the same effect for both part-time and full-time jobs.
Sol. The null and alternate hypothesis for testing are as follows:

$$\text{HO: } \beta_{partime(primary)} = \beta_{fulltime(primary)}$$
$$\text{H1: } \beta_{partime(primary)} \neq \beta_{fulltime(primary)}$$

Since Prob > chi2 = 0.6793 we fail to reject the null hypothesis

$$\text{HO: } \beta_{partime(middle)} = \beta_{fulltime(middle)}$$
$$\text{H1: } \beta_{partime(middle)} \neq \beta_{fulltime(middle)}$$

Since Prob > chi2 = 0.1764 we fail to reject the null hypothesis

$$\text{HO: } \beta_{partime(Secondary)} = \beta_{fulltime(Secondary)}$$
$$\text{H1: } \beta_{partime(Secondary)} \neq \beta_{fulltime(Secondary)}$$

Since Prob > chi2 = 0.2416 we fail to reject the null hypothesis

$$\text{HO: } \beta_{partime(HiSecondary)} = \beta_{fulltime(HiSecondary)}$$
$$\text{H1: } \beta_{partime(HiSecondary)} \neq \beta_{fulltime(HiSecondary)}$$

Since Prob > chi2 = 0.0000 we reject the null hypothesis

$$\text{HO: } \beta_{partime(PostHiSecondary)} = \beta_{fulltime(PostHiSecondary)}$$
$$\text{H1: } \beta_{partime(PostHiSecondary)} \neq \beta_{fulltime(PostHiSecondary)}$$

Since Prob > chi2 = 0.0000 we reject the null hypothesis

b) part-time and full-time work are indistinguishable
Sol. We test the following hypothesis

$$HO: \beta_{part\ time}(matrix\ of\ cofficients) = \beta_{full\ time}(matrix\ of\ cofficients)$$
$$H1: \beta_{part\ time}(matrix\ of\ cofficients) \neq \beta_{full\ time}(matrix\ of\ cofficients)$$

Since Prob > chi2 = 0.0000 we reject the null hypothesis.

c) IIA for part-time as the new category, keeping NILF and full-time as the original categories.

Sol. Independence of irrelevant alternative means that the new category added (part time) will have no effect on the odds ratio of other two.
We followed the following steps to test for IIA using Hausman test.
Step1: We estimate the full model and save the outcome.
Step 2: Estimate the restricted model i.e. model with removing part-time as a category and save the outcome.
Step 3: Then we use the Hausman test statistics which follows chi-square distribution to test the presence of IIA

The null hypothesis is as follows:

HO: Difference in coefficients not systematic (presence of IIA)
H1: Difference in coefficients is systematic (not IIA)

Since Prob > chi2 = 0.0000 we reject the null hypothesis i.e. the presence of IIA.

Q.2) Estimate an ordered logit model based on the three outcomes. Discuss how the 'ordering' will be formed along with some results to support that multinomial logit and not ordered logit is more suited for this analysis or vice-versa.

Sol. The ordering is formed with number of hours worked. We estimate the value of cut1 and cut2 as 7.640082 and 10.08987. We form the ordering as

$$Y_i = \begin{cases} 1\ (NILF) & 0 < no.of\ hours\ in\ labor\ force < 7.64 \\ 2(Part\ Time) & 7.64 < no.of\ hours\ in\ labor\ force < 10.09 \\ 3\ (Full\ time) & no.of\ hours\ in\ labor\ force > 10.09 \end{cases}$$

The pseudo R^2 is very low for ordered for the multinomial logit model (0.2468) than the Ordered logit model (0.1981). Therefore, multinomial logit is more suited for this analysis.