Association between Occupational Diversity Distribution and Rural Livelihood Pattern: A Discrete Choice Study in Purulia District of West Bengal

Mrinmoy Kumar Paul[1] & Dr. Subhasis Bhattacharya[2]

1 Research Scholar, Department of Economics, SidhoKanhoBirsha University, Purulia, West Bengal

2 Professor, Department of Economics, SidhoKanhoBirsha University, Purulia, West Bengal

Abstract

Objective: Occupational diversity is one of the key way-out to the rural people to struggle against poverty. Different literatures suggest that scope of occupational diversity varies between poor and non-poor on the basis of different social and physical factors like skill, education and experience. Present study has some intention to investigate the nature of occupational diversity in a rural agricultural frame like Purulia of Western West Bengal.
Methods: The study is based on primary data which are systematically chosen by the selection of blocks as per Indices of Human Development and villages as per population engaged in agriculture and households as per probability proportional to size from the categories of landless agricultural labourers, marginal cultivators and marginal agricultural labourers. The study considers occupational diversity as a binary discrete choice variable and uses binomial logistic regression model over some selected independent covariates.
Results: The study uses Jamovi-2.3.3.0 for the analysis of the binomial logistic regression results. The independent covariates densities over occupational diversity and non-diversity are significantly explained by the study and indicates their influences over it.

Key Words: Occupational Diversity, Livelihood, Discrete Choice, Purulia

Address for Correspondence: Prof. Subhasis Bhattacharyya, Professor, Department of Economics, Sidho-Kanho-Birsha University, Sainik School, Ranchi Road, Purulia-723104; e-mail:suva69eco@gmail.com

Association between Occupational Diversity Distribution and Rural Livelihood Pattern: A Discrete Choice Study in Purulia District of West Bengal

Introduction:

The literatures of development economics make two perspectives regarding the procedure of changing rural occupational structure in the developing countries. The first perception is based on variety of rural linkages which exists under certain assumptions and allows us to make overall development of the rural sector. For example, due to some affiliation of agricultural growth supplementary incomes and new demands will generate and for this the use of local resources and skills will be accelerated. Such factors will initiate occupational diversification in rural sector. ***Kuznets (1966)*** identified such growth as agricultural transformation which can be achieved through mechanisation of agriculture attended by an increase in the productivity of agricultural labourers and create surplus. Ultimately the demand for labour in the non-agricultural sector was promoted by the generations of agricultural surplus and an altering pattern of consumption demand (***Unni, 1994, 1996, 1998***). Thus, in the succeeding stage of development growth of tertiary sector is leaded by such enhancement in the demand for services in both the primary and secondary sectors. As per Kuznets' such spreading out of non-farm sector is termed as prosperity. ***Chandrasekhar, (1993)*** pointed out that Kuznets' concept of prosperity is not constrained to the rural sector of an economy or region because it is founded on the association between modern economic growth and the diversification to more productive activities depended on the national or international market. Earlier studies confirmed that such rural sector based development approach fabricated high average rates of returns to investment in agriculture (***Mellor, 1976***). As a result small scale industries received capital straightforwardly from cultivators and constructing rapid growth. The price associations for industrial consumer goods formulate potential elevated profits and investment. The income support in agriculture becomes enough to maintain taxes to self-finance various infrastructural requirements. Thus, a twofold story is functioning simultaneously -- one is considerable net outpouring of resources from agriculture that will encourage growth in other sectors of the rural economy, and the other is agriculture itself which is intensifying speedily with profitable way. But ***Chandrasekhar, (1993)*** criticizes Mellor's observant, which is applicable for a diagnostic narrative of a particular experiential certainty. As per Chandrasekhar's version, Mellor study was conditional to the subsistence of a number of inclinations which authorize the understanding of a monotonic affiliation between agricultural growth and an augmentation in rural non-agricultural employment. A lot of literatures criticize such growth linkage approach which is mainly founded on some impractical assumptions about the receptiveness of local non-farm output to increasing demand by farmers (***Harriss, 1987; Hart, 1989; 1993***).

McGee (1971) studied such growth prospect of the rural sector manage to pay for agricultural labourers and industrial renovation. He identified such surplus agricultural labour as an outcome of agrarian distress and termed it as distressed induced. Due to limited labour absorption capacity of the industrial sector, it may not be ready to absorb the surplus labour produced by the agricultural sector. The surplus labour is then brazen out with no alternative but to reconcile into a low productive tertiary sector. The example of such distressed induced diversification is seasonal migration of the poorer households towards urban areas in exploration of unskilled employment in informal sector. ***Vaidyanathan, (1986)*** identified such distressed induced diversification as residual sector hypothesis. The residual sector hypothesis will transpire when the labour is not completely absorbed in the agricultural sector and the non-agricultural sector proceeds as a mop for the surplus labour. ***Chandrasekhar, (1993)*** in his study over India argued that agricultural stagnation and substitution of labour by capital during post green revolution period augmented such diversification into non-farm activities. As a result, the subsistence of an excess labour supply incapable to join in protected agricultural jobs might have been a stronger incentive to non-farm investment than farm growth. In alike stratum, the position and dynamism of rural non-farm activity may be due to its own cost and distinctive competitiveness characteristics rather than farm output growth (***Ellis, 1998***). But other studies display that in India most dynamic growth areas in the rural non-farm sector depends profoundly on urban and export demand, and acquire insignificant associations to agriculture (***Fisher et al., 1997***). Thus, diversification as a livelihood strategy at individuals and households level is often alienated into two overarching considerations of necessity or choice (***Ellis, 2000***). Such consideration of necessity or choice is occasionally pretense as a difference between survival and choice (***Davies, 1996***). Migration literature termed it as push versus pull reasons to migrate (***Brenstein, H., 1992***). In the literature, there is an inference, in part, with respect to this dichotomy that diversification for distraction reasons is a bad thing. Consideration of necessity refers to automatic and distraction motivations for diversifying and considerations of choice, refers to intended and practical motivations for diversifying. Such diversification for distraction household members' responsibility is casual and leads to low productivity activities with poor prospect (***Ellis, 2000***). In other way, it is a last resort rather than a striking substitute livelihood (***Bhagat, 2011***). It may also show the way to households approving a more vulnerable livelihood system than they possessed previously (***Davies, et al2007***). Some studies shows that occupation diversification is found as a strategy of dispersal risk to reduce vulnerability to random crises such as floods, droughts, and illness as well as the seasonal fluctuations of natural resources are termed as distress occupational diversification (***Papola, 2013, 1994; Reddy, D.N. and Venkatanarayan, M. 2013***). Compared to the poor, the non-poor are more able of financing this diversification if it is costly with high entry barriers, and is originally risky (***Biggs, S., et al., 2014***). Some study results specified that livelihood diversification at the household level is correlated with advanced wealth status and possession of a range of assets as part of a progressive, accumulation livelihood strategy for those with fewer constraints (***Martin S.M. and Lorenzen K., 2013)***. Present study deals with the association between occupational diversity and pattern of rural livelihood in terms of discrete choice models. Livelihood of people of Purulia district is mainly influenced by agriculture. Economic sources of their earnings majorly depends on agriculture, and to consider the issues of occupational diversity and food security, the study has some plan to observe the levels occupational diversity and food security of the people majorly linked with agriculture. The rural temperament of Purulia with topographic impediment like a steady runoff is one of the major obstacles to agriculture. During summer a small amount of land only receive the irrigation facilities as per their close proximities with the connected water sources. So the crisis over agriculture is continuous and people's reliance over agriculture is the important aspect in terms of occupational diversity (***Loison, S. A., 2019***).

Data and Methods:

The study is based on collection of primary level data which is a juxtaposition of scientific selection of blocks, villages and households. The block selection methodology depends upon the level of development levels of the blocks as per district Human Development Indices (HDI) which is quite old (***DHDR, 2012***) and no recent data structure is available. The descriptive statistics of HDI score of all the 20 blocks of the district shows that average HDI score is 0.371 with standard deviation 0.0497. The respective 25th, 50th and 75th percentile scores of HDI are 0.347, 0.380 and 0.392. On the basis of such percentile scores the study makes a levelling of blocks in categories of highly developed, moderately developed, poorly developed and least poor developed blocks. The maximum and minimum values of such HDI scores are 0.49 and 0.28. As per the constructed percentile HDI score of all twenty blocks shows that there are five blocks whose score is greater than 0.392 which are termed as highly developed blocks; the six blocks whose HDI score is less than 0.392 and more than or equal to 0.380 are termed as moderately developed blocks; the four blocks whose HDI score is less than 0.380 but more than equal to 0.347 are termed as poorly developed blocks and lastly, there are five blocks whose HDI score is less than 0.347 are termed as least poor developed blocks. In figure-1 the third column shows such identification of blocks in terms of four identified categories of development. Now from each category two blocks are chosen on the basis of best representation in terms of HDI. In each mentioned four formed categories there are different number of blocks and the study estimates the $Mean - SD$ like standard normalisation process of the given HDI values for each categories. The study observes that for highly, moderately, poorly and least poor developed blocks the $Mean - SD$ differences are 0.393258, 0.379523, 0.347926, and 0.286321 respectively. By this way from each of four HDI categories two blocks are chosen (Table-1) and in total eight blocks are selected from the list of twenty blocks.

After the selection of the respective blocks in terms of their performance in HDI score, the study considers village selection as the next significant task. The village selection process is a two stage purposive sampling and two variables are identified here to enumerate the inclusion of the villages. These two variables are (i) number of household and (ii) representative of the people who are linked with the agriculture without land holding, and that should be proxied by the study by the number of landless agricultural labourers, marginal cultivators and marginal agricultural labourers from the Census-2011 data set. In the second stage of village selection methodology, the study assembles the village wise information of landless agricultural labourers, marginal farmers and marginal agricultural labourers of each village whose household size is greater than 150 for each blocks. The study imposes the number of these three classifications as the prime significant determinant to get suitable information regarding occupational diversity. So the sum of these three categories population including male and female considered as the final basis of village selection. Then the villages are ranked in terms of such participation aggregates under these categories. The maximum value obtain the relative importance and the village percentage of landless agricultural labourers, marginal cultivators and marginal agricultural labourers household sum is selected for the intervention. To consider the issue of occupational diversity, larger size of the household is adopted by the study and it ranked the villages under each block in terms of descending number of households. Then from such descending order list the study assembles the villages where the highest number of sum of landless agricultural labourers, marginal cultivators and marginal agricultural labourers exist. From each block top performing villages is selected and by this way from eight blocks eight villages are selected. Figure-2 and figure-3 indentify the selection of eight villages from eight blocks.

Such kind of methods of selection of villages from eight blocks identified that there exist a variation in the percentage of participation of the above formulated three categories of labour. The variations can be easily understood from the figure-4. The study found that in Barabazar community development block more than 60 percent of the total population are under the above formulated categories of labour and the same for Raghunathpur-I such percentage is only 15.77. It is one of the intension of the village selection methods of the study to allow such variation in number of such formulated categories of labour because it will outbreak the block level heterogeneity in terms of occupation diversity, which helps us to explain the adopted model in terms of intra-block level diversity and gender wise occupational diversity.

The final stage is household selection from the categories of landless agricultural labourers, marginal cultivators and marginal agricultural labourers, and here the study uses random sampling procedure. Then from the list of selected villages 45 households were taken from each village at the rate of 15 from each category of landless agricultural labourers, landless marginal cultivators and landless marginal agricultural labourers. So total of 360 landless agricultural labour households were interviewed for collection of data regarding their income sources, food security and occupational diversity during 2018-19. For comparative analysis purpose the study also collects the information of landed agricultural farmers and from each selected village the information of 20 landed agricultural labourer households were collected by the study. Thus from all selected villages 160 number of landed agricultural labourer households were taken. Aggregately study collects information of 520 households from eight selected villages over four differently developed blocks collectively from landless and landed agricultural labourers.

Nature of Dependent and Independent Factors:

The non-farm activity in the rural sector emerges as the amalgamation of growth process of the economy and significant policy intervention of the state (***Viadyanathan, 1986***). Different studies decorated that distress diversification may be one noteworthy reason of participation of the rural landless agricultural labourer to engage in the non-farm activity. If there exists a dynamic workable non-farm sector that also implicated large number of landless agricultural labourer in to the non-farm activity. Such relationship is well-known as push & pull factors of dynamic non-agricultural system of the rural sector. The study considers a logistic model, where a set of 13 independent covariates are taken to explain the occupational diversity. The choice of independent covariates was taken from a huge set of earlier literature which specifies mainly on the rural setting. Table-2 explains such choices of dependent and independent covariates.

As the dependent variable is categorical, the study considers Binomial logistic regression technique to explain the relationships between the dependent and independent covariates. The estimates of the logistic regression identified the importance of the selected variable to explain the situation of occupational diversity. All the respective variables like family size, importance of nonfarm asset, literacy rate, operated area, monthly per-capita consumption expenditure; dependencies and location of the household are to be tested under the fitted binomial logistic specification about their significant influence on the occupational diversification at the household level.

The number of economic activities per household (OCCDV) is taken as the dependent variable. Such dependent variable is a binary one and as constructed by the purpose of the study. If the average household level involvement in the occupation is equal to or less than one then the value of od treated as 0, and if such value is greater than one then od treated as one.

$$od_i = \begin{cases} 0, if\ the\ household\ level\ occupational\ involvement\ \leq 1 \\ 1, if\ the\ household\ level\ occupational\ involvement > 1 \end{cases}$$

The dependent variable is occupational diversity, which is binary in nature as constructed by the study. Here, 0 means not occupationally diverse and the vice versa for 1. The study observes that 61 percent of the households are occupationally diverse. The mentioned table-3 and figure-5 identify the same facts. The binomial test of occupational diversity shows that both the levels are significant at 5 percent level, the density function for both the proportion of 0 and 1 level within the prior and posterior likelihood. Such levels of occupationally not diverse (0) lies between 0.25 and 0.50, and the levels for occupationally diverse (1) lies between 0.50 to 0.75.

The first independent variable is the family size ($FMSZ$) which appeared as numerical value. The study considers the average household size of district Purulia 5.16 as a standard from the Census-2011 data. The study treats $FMSZ$ as a binary independent variable on the basis of the Census-2011 specific average value.

$$FMSZ = \begin{cases} 0, if\ the\ size\ of\ household \leq 3 \\ 1, if\ 3 < \ size\ of\ household \leq 5 \\ 2, if\ 5 < \ size\ of\ household \leq 8 \\ 4, if\ the\ size\ of\ household > 8 \end{cases}$$

Among such categories the study observes that majority of the households are within the category 2 (43.65 percent) followed by category 3 (23.65 percent). Cross tabulation between occupational diversity and family size shows that between occupationally not diverse and occupationally diverse groups majority of households are under category 2 (50.25 percent and 39.5 percent) followed by category 1 (26.37 percent and 30.09 percent). The density distribution of the family size over occupational diversity categories shows that for the 0 or the first categories of family size majority of the families are under occupationally not diverse group and the same thing holds for very large family size (group 3), but for family size of group 1 and 2, majority of the households are under occupationally diverse category. The box-plot within the violin with jittered data and the vertical bar diagrams confirm such presentation (figure-6). The ANOVA between occupational diversity and family size shows that family size is a significant covariate (table-4) and the Levene's test to check the assumptions for homogeneity of variance is also found significant by the study. The normality test is also found significant with the help of Shapiro-Wilk test. Since both the dependent and independent variables in case of such relationship are dichotomous type it is always better to determine non-parametric one way ANOVA between occupational diversity and family size. The Kruskal-Wallis one way non-parametric ANOVA is also found significant with effect size 0.0493. The Dwass-Steel-Critchlow-Fligner (DSCF) (table-5) method to compare between different pairs shows that the comparison between family group 1 and 3 is found significant over the domain of occupational diversity.

Sex of the head is considered as the independent covariate to influence the occupational diversity. In the context the study observes that 87.1 percent household is headed by a male and identified as 0. In case of distribution between occupational diversity and sex of the head, the study observes that 34 percent are not occupationally diverse and 53.1 percent are occupationally diverse of the male headed families. The same relative percentages for the female headed families are 4.6 percent and 8.3 percent respectively. From the density distribution it is clear to us that majority of the female headed families are occupationally diverse in comparison to male headed families (figure-7). The parametric test of sex of the head of the households in terms of one way ANOVA shows that it is significant and the Levene's test to check the assumptions for homogeneity of variance is also found significant by the study. The normality test is also found significant with the help of Shapiro-Wilk test. The non-parametric ANOVA (Table-6 & 7) between sexes of the head under the settings of occupational diversity shows that Kruskal-Wallis one way non-parametric ANOVA is also found significant with effect size 0.181. DSCF method to compare between two sexes is found significant over the domain of occupational diversity.

Religion is considered as a covariate of the study and such religions have been categorised as Hindu (0), Muslims (1), Christen (2), Jain (3) and Other (4). Majority of the household is under Hindu religion (70.58 percent) followed by Muslims (16.35 percent). The religion wise density distribution of the occupational diverse households shows that Hindus and Christen families are equally distributed between two categories of occupational diversity, whereas majority of the Muslims households are occupationally diverse and just reverse is for the Jain households (figure-8). The Kruskal-Wallis one way non-parametric ANOVA shows that chi-square is not significant with the effect size 0.0294. The DSCF Pair-wise comparison between different religions categories over the two categories of occupational diversity are not found significant for any of the considered pairs. The probable cause of such insignificance is that Hindus occupy the major portion of the sample size (over 70 percent) and all other four categories hold only 30 percent. So the Pair-wise comparison is found insignificant over the domain of occupational diversity.

The study considers Caste of the household as an important covariate with five categories. 0 stands for general category, 1 stands for Schedule Castes, 2 stands for Schedule Tribes, 3 for Other Caste category and 4 for Other Backward Class (B). Purulia has some historical legacy for the presence of such huge households in category 4 (36.9 percent). The General category is the second highest occupying 31 percent. The density based distribution of caste over occupational diversity shows that majority of the General castes are in occupationally not diverse, but the SCs, Other Castes, and OBC-Bs are majorly occupationally diverse, whereas STs are found with symmetrical distribution between occupational diversity and non-diversity. The median derived from box-plot also supports such information (figure-9). The ANOVA between categories of Occupational diversity and caste categories shows that F-statistics is significant satisfying the Levene's homogeneity of variance test and the normality of the relation holds as per Shapiro-Wilk test for normality. Considering the categorical dependent and independent variables, the study also checks the non-parametric one way ANOVA test (Table-8). The Kruskal-Wallis test shows that chi-square is significant at 5 percent level with effect size 0.0639. The DSCF Pair-wise comparisons for non-parametric ANOVA of different caste categories under the sphere of occupational diversity & non-diversity identify that comparisons with general case with all other caste categories are found significant except the Other Caste categories. But such between groups comparisons are found insignificant for other caste categories.

Whether House Condition gives us some interpretation of occupational diversity or not may be a significant question (figure-10). The study categorises houses in four categories. These are Pucca or concrete House (3); Semi-Pucca houses (2), which is basically of concrete wall and tin / asbestos roofed; Mud-asbestos based (1), which is basically of mud wall and tin / asbestos roofed; and completely thatched (0). The distribution of house condition shows that majority of the households are living under category 1 followed by category 2 and then category 0. The interaction between house conditions with respect to occupational diversity is identified by the density distribution of different categories of house conditions. The household which are living in thatched house (0) are found completely occupationally diverse in nature and close same pattern is also followed by the distribution of households who are living in mud-asbestos (1) based houses. People living under Semi-Pucca houses (2) are found closely equal between occupationally diverse and occupationally not-diverse groups and people in Pucca houses are majorly occupationally non-diverse in nature. The box-plot of the following distribution supports such findings. The one way ANOVA between house condition and occupational diversity categories found significant F values, which also established Levene's homogeneity of variance test and the Shapiro-Wilk normality is also verified (Table-9). The non-parametric one way ANOVA by Kruskal-Wallis test is also found significant like parametric test and the effect size is 0.133. The DSCF Pair-wise comparison of different house conditions under the categories of occupational diversity and non-diversity shows that interaction between thatched houses and other houses are found significant except mud-asbestos based houses. Similarly interaction between mud-asbestos based houses and other houses are found significant except thatched houses.

Number of room within the existing house is considered as good covariate by the study (figure-11). The four categories of houses as per number of rooms can be formed as 1 room, 2 rooms, 3 rooms and more than 3 rooms. The study observes that majority of the houses are 1 room house securing 61 percent followed by 2 rooms (30.6 percent). Out of the 1 room houses the study observes that 12.5 percent of the total households are occupationally non-diverse and 48.5 percent households of total households of such houses are occupationally diverse in nature. The corresponding percentage for 2 rooms houses are 20.2 and 10.4 percent. The cross tabulation between occupational diversity and non-diversity with number of rooms shows that household living in 1 room and household living in more than 3 rooms are more occupationally diverse in comparison to households living in 2 and 3 rooms. The probable cause behind such distribution is the huge difference between the sample sizes belonging to 1 room and more than 3 rooms. The number of households under the more than 3 rooms is found very low (only 7). The parametric and non-parametric ANOVA are found significant with significant Levene's homogeneity of variance and significant Shapiro-Wilk normality (table-10 & 11). The DSCF Pair-wise comparison of different categories of rooms under the specifications of occupational diversity and non-diversity shows that except the intersection between 1 room and 2 rooms all other intersections are found significant.

Type of fuel used by the households is considered as one important covariate of occupational diversity. The houses are separated as use of firewood (0), coal (1) and gas (2). The study observes that 68.87 percent of the total households uses firewood for cooking fuel, where as 19.4 percent uses gas and 11.7 percent uses coal (figure-12). The density of distribution of fuel use shows that gas user households are specifically occupationally diverse, where as firewood and coal users are found with more or less equal amount between occupational diverse and occupational non-diverse groups. From the box-plot jittered data it is easily understood that majority households concentrate on use of firewood. The parametric ANOVA shows that type of fuel used by the houses is found significant (table-12). The Levene's homogeneity of variances and Shapiro-Wilk test of normality are also found significant. The nonparametric ANOVA is also found significant in terms of Kruskal-Wallis Chi-square with effect size 0.0709. The DSCF Pair-wise comparison between groups of different fuel user households are found significant between firewood users and gas users and also between coal and gas users.

Value of non-farm assets is considered as the independent variable to influence the occupational diversity and non-diversity. The cardinal money value of the non-farm asset per household ($VNFA$) is measured by the study. Study collects the information of households based non-farm assets and converted them in monetary values with respect to the price level data of 2018 from the retail price data of CPI provided by MOSPI with base price of 2012. Thus, there will be huge chance of manipulation error generation if such raw data is directly used in the regression. Hence, the study nomalised such household level data at the village level and for such nomalisation the study uses most common technique as given below:

$$VNFA_i^j = \frac{V_i^j - V_{min}^j}{V_{max}^j - V_{min}^j}$$

Here, ${nf^a}_i^j$ is the normalised value of the non-farm asset for the $i - th$ household in the $j - th$ village, V_i^j is the money value of the non-farm assets belonging to $i - th$ household in the $j - th$ village, V_{min}^j is the minimum value of the non-farm asset of any household in the $j - th$ village, and V_{max}^j is the maximum value of the non-farm asset of any household in the $j - th$ village. Then, the normalized values of $VNFA$ now distributed between four categories on the basis of total distribution of such score between 0 to 1. The, households whose $0 \leq VNFA < 0.25$ is identified as 0; the households whose $0.25 \leq VNFA < 0.50$ is as 1; the households whose $0.50 \leq VNFA < 0.75$ is as 2 and the households whose $VNFA \geq 0.75$ is identified as 3. The study observes that 32.9 percent households of the total are belongs to category 0 and 30 percent belongs to category 2.The cross tabulation based density distribution between occupational diversity and non-diversity with VNFA categories shows that holding of VNFA for categories 0 and 1 are majorly occupationally diverse and the holding of VNFA for categories 2 and 3 are majorly occupationally non-diverse (figure-13). From such distribution it can be strongly recommended that holding of VNFA is a good indicator of occupational diversity in study area. Such claim is also supported by the parametric and non-parametric ANOVA test produced significant results between VNFA categories and Occupational diversity categories. For parametric ANOVA the F value is found significant with significant Levene's homogeneity of variance and significant Shapiro-Wilk normality test (Table-13). For non-parametric ANOVA Kruskal-Wallis is also found significant Chi-square with effect size 0.793. The DSCF Pair-wise comparisons between different groups VNFA shows that except between groups 0 and 1, and groups 2 and 3, the VNFA Pair-wise movements are found significant by the study.

Literacy level of the head of the household (LIT) is also considered as categorical in terms of the study reference level. As per Census, 2011 data the average literacy rate of Purulia is 64.48 percent which is measured in terms of the adult population level, but in this study we express the literacy of the head of the household as follows.

$$LIT_i = \begin{cases} 1\ if\ illiterate \\ 2, if\ illiterate\ \&\ can\ do\ sign \\ 3, if\ time\ spends\ in\ school < 4\ years \\ 4, if\ 4 \leq time\ spends\ in\ school < 8\ years \\ 5, if\ 8 \leq time\ spends\ in\ school < 10\ years \\ 6, if\ time\ spends\ in\ school \geq 10\ years \end{cases}$$

The study observes that 24.2 percent of the total head of the households are illiterate and 30.6 percent are illiterate and can do signature. 26.7 percent of the head of the households has schooling less than 4 years (figure-14). The study observes a clear tendency in favour of occupational diversity as the time spends in school increases. For illiterate and illiterate with ability to do signature, the study observes that more or less same frequency is found between occupational diversity and non-diversity. The box-plot jittered data shows that majority of the households are scattered within the first three categories of the created variable LIT. The parametric tests between occupational diversity categories and literacy categories are found significant in terms of F value and the Levene's homogeneity of variance and the Shapiro-Wilk normality tests are found significant (Table-14). The non-parametric Kruskal- Wallis test is found significant with effect size 0.122. The DSCF Pair-wise comparisons between different literacy categories within the purview of occupational diversity and non-diversity shows that intersection between 1 and 3; 1 and 4; 2 and 3; 2 and 4 are found significant.

The operational area or the landholding in acre per household ($LANDHOL$) may be significant variable for the study like occupational diversity. As per the statistics given by Ministry of Agriculture and Farmers Welfare, Government of India, the least category of agricultural labour is chosen from the category of marginal farmer household whose land holding is less than 0.05 acre. Considering this as standard, the study identified LANDHOL area in acre per household as a categorical variable.

$$LANDHOL = \begin{cases} 1; if\ the\ operational\ area \leq 0.05\ acre \\ 2, if\ 0.05 < the\ operational\ area \leq 0.5\ acre \\ 3, if\ 0.5 < the\ operational\ area \leq 1\ acre \\ 4, if\ the\ operational\ area > 1\ acre \end{cases}$$

The study observes that 29.8 percent households are under 1 category of landholding, and 50.6 percent are under 2 category of landholdings. Thus nearly 80 percent of the studied households are either under category of marginal farmers or small farmers. The density based distribution of landholding categories intersect with occupational diversity categories which shows that first two categories of landholding has almost equal share between occupationally diverse and non-diverse groups, but the large land holders are found occupationally diverse (Figure-15). The major cause behind such outcome is that large landholder lease out their land and they are interested in various high earning business or jobs and as a result the occupational diversity for them is high. It is also important to note that the number of last two categories are found relatively small, which is 20 percent of the total. The parametric ANOVA between landholding categories and occupational categories is found with significant F value (Table-15). The Levene's homogeneity of variance test and Shapiro-Wilk normality tests are found significant also. The non-parametric ANOVA of the same is found significant in terms of Kruskal-Wallis test with effect size 0.118 and the DSCF Pair-wise comparisons between different landholding categories under two occupational diversity scenarios show that comparisons between categories 1 and 3, 1 and 4, 2 and 3 are found significant.

The development and scope of the living place is a significant determinant of occupational diversity. In our study we choose the blocks in terms of human development index values. Such composition identifies that there are four categories like high, moderate, poor and least poor developed blocks as per the HDI value. The study treat locational factors ($LOCT$) as binary also.

$$lc = \begin{cases} 0, if\ the\ household\ is\ in\ the\ Poor\ or\ Least\ poor\ developed\ blocks \\ 1, if\ the\ household\ is\ in\ the\ High\ or\ Moderate\ developed\ blocks \end{cases}$$

The study observes that almost equal representations of locations are found under the created categories. Such representation is influenced by the choice methodologies of the study design. The density based distribution of the two locational categories show that both the between groups have almost equal proportion in occupational diversity. The within groups distribution shows that occupational diversity proportion is quite high in comparison to non-diversity (Figure-16).Both the parametric and non-parametric ANOVA between locational categories and occupational diversity categories are found insignificant, though the normality and homogeneity of variances are supported by the Shapiro-Wilk and Levene's test. The effect size of Kruskal-Wallis is very small (0.0059) and the DSCF Pair-wise comparisons are also found insignificant.

Dependency ratio may have some significant influences over occupational diversity. The dependency ratio of the i-th household (dd_i) will be estimated by

$$DEPDN_i = \frac{Child_i + Old_i}{Adults_i}$$

It is the ratio between the number of child and old aged in some specific family divided by the number of adults of that specific household. The families with $DEPDN_i = 0$ has the minimum dependency and it can be treated as the terminal conditions and such categories marked as 1. For the families with $1 > DEPDN_i > 0$ shows the scenario of some low level of dependencies and this is categorized as 2. $DEPDN_i = 1$ is another case for the families with equal number of adults and child plus old aged, which is categorized as 3. The $DEPDN_i > 1$ is another category adopted by the study and it is identified as category 4. The study observes that dependency as formed by the study, where majority of the households are within category 3 and followed by category 2. In other categories of dependencies secures only 12 percent of the total households. The density wise distribution of different categories of dependencies over two categories of occupational diversities shows that except the first category, every category is found with greater share in favour of occupational diversity than non-diversity (Figure-17). The parametric ANOVA and non-parametric ANOVA of dependencies categories over occupational diversity categories are shown insignificant. But the normality by the Shapiro-Wilk is satisfied. As per Kruskal-Wallis ANOVA, the pair wise DSCF is found not-significant between any cases.

The study considers the household size-wise MPCE of the study area and judges Rural Uniform Reference Period (URP) method estimated values Rs 1278.94 as a bench mark for that. Then the study estimates the average inflation rate for the year 2013 to 2018 (six years) and the inflated value of URP identified for rural areas is Rs 1364.34. Considering such bench mark for individual level and transforming it towards household level, the study converts MPCE into a categorical variable, where 0 means the household below the standard MPCE norms and 1 means the household equal and above the standard MPCE. The study observes that out of 520 households

61.7 percent are below the standard MPCE (Figure-18). The density based distribution between MPCE categories and occupational diversity categories shows that 0 group as per MPCE categories is more occupationally diverse as compared to other MPCE Groups. In case of MPCE 1 group, the distribution between occupationally diverse and non-diverse group, the percentage allocation is almost equal. The parametric ANOVA between MPCE and occupational diversity shows significant F value with significant Levene's homogeneity of variance and satisfying the Shapiro-Wilk normality. The non-parametric ANOVA of the same is found significant .Kruskal-Wallis Chi-square with effect size 0.09. The DSCF test to observe the variation between two MPCE groups over occupational diversity categories is found significant. From such result it is also observed that the choice covariate MPCE is a good indicator over occupational diversity (Table-16 & 17).

Chi-Square & Likelihood Ratio Test of Independent Covariates

The influences of the selected independent variable to explain the occupational diversity is the essential part of the study. It is also important to understand the reliability of the model. After selection of such thirteen covariates to understand their influences over occupational diversity, the study checked the significance of all independent covariates through Chi-square test and likelihood ratio test. Study observes that family size is a significant covariate in terms of likelihood ratio test, and in case of Chi-square test it is found insignificant, though the p value of 0.001. The list of variables like house condition, number of rooms available in the house, fuel used for cooking, value of non-farm assets, educational years spent in school, operational land holding, dependency level of the families and MPCE are found significant both under the Chi-square as well as under likelihood ratio test. Thus, among the 13 selected variable the study observes that nine found significant as per likelihood ratio test (Table-18).

Result Analysis

Binomial Logistic Model Specification of Occupational Diversity

The study exercises logistic regression to comprehend the determinants of occupational diversification among agricultural labour households and most of the study variables are measured in terms of per household marking.

Thus the functional relationship is

$$OCCDV_i = f(FMSZ_i, SEX_i, RLG_i, CASTE_i, HC_i, ROOM_i, FUEL_i,$$
$$VNFA_i, LIT_i, LANDHOL_i, LOCT_i, DEPDN_i, MPCE_i)$$

The logistic model can be expressed as

$$p(OCCDV_i = j) = \frac{e^{\sum \gamma_i x_i}}{1 + e^{\sum \gamma_i x_i}}, for\ j = 0,1$$

The equation used to estimate the coefficients is as the form,

$$\log[\frac{p_i}{1 - p_i}] = \gamma_0 + \gamma_1 FMSZ_i + \gamma_2 SEX_i + \gamma_3 . RLG_i + \gamma_4 . CASTE_i + \gamma_5 . HC_i + \gamma_6 . ROOM_i + \gamma_7 . FUEL_i + \gamma_8 VNFA_i + \gamma_9 LIT_i + \gamma_{10} LANDHOL_i + \gamma_{11} LOCT_i + \gamma_{12} DEPDN_i + \gamma_{13} MPCE_i$$

Here, the dependent variable is the occupational diversity at the household level and $od = 0$ means the insignificant occupational diversity, and $od = 1$ means a substantial presence of occupational diversity. The ratio $\left(\frac{P_i}{1-P_i}\right)$ is termed as odds-ratio which is nothing but the ratio of probability of occupational diversity and non-diversity. In the above model the study considers the logarithm of such odds ratio is a linear function of the set of independent variables. This equation authorizes for the clarification of the logistic weights for variables in the same way as in linear regressions. For example, $e^{\gamma x}$is the multiplicative factor by which the odds ratio would adjust if x changes by one unit. An odds ratio of 1 signifies no effect of an occupational diversity variable. When the odds ratio is greater than 1, it indicates that the characteristic increases the odds of that occupational diversity category compared to base, and a ratio less than 1 indicates that it diminishes the odds. For instance, an odds-ratio of 0.75 signifies that the chance that households with the given characteristic get a given occupational diversity is 75 percent of the chance of being occupationally diverse. An odds ratio of 1.5 implies a 1.5 times greater chance of being occupationally diverse in the given category than the other. The higher the odds ratio, the stronger the association, so these present an implicit ordering of occupational diversity with different independent variables.

Binomial Logistic Model Fit Measures of Occupational Diversity

AIC, McFadden R² and Over all model Test

The logistic regression fitted by the study first considers the model fit measures (Table-19). The overall model fit measures show that the Chi-square is found significant. This is a good fit that also can be justified by McFadden R^2 value 0.0459 which is constructed from the likelihood ratio index covering all predictors. Another Pseudo R^2 formed by Cox and Snell is also found valid. The Pseudo R^2 of Nagelkerke's is found quite large in comparison to Cox and Snell. To compute estimates of logistic model maximum likelihood estimation process is used. Through iteration the difference between observed response and predicted response is minimised. The measure of such discrepancy is known as deviance. If the deviance is larger, discrepancy becomes quite high. In the fitted model, the study found the deviance as 3.35. The Akaike Information Criterion (AIC) is not very high (10.1) to reject the model. Thus, the overall model fitness is satisfied as per the purpose of study.

Tests of Significance and Odds Ratio's of Independent Covariates over Occupational Diversity

The responsiveness of all selected independent covariates with their respective levels has been tested and the result is shown in table-20. For the covariate like family size the study observes that odds of occupational diversity of the households belong to family size 1 is 357 percent greater than the family size 0. The same between family size 2 is 294 percent greater than the family size 0. For the family size belonging to 3 occupational diversity is 555 percent greater than the family size 0. From such results, the study opines that family size is a very significant variable for measuring occupational diversity and that is confirmed from the p-values for all levels of family size. Sex of the head of the household is found an insignificant determinant for measuring occupational diversity. May be the lower presence of female heads over male heads in the total sample size is one of the major causes of such insignificance. The results of religion show that if one moves from Hindu religion to any other religion category, occupational diversity increases by more than and close to 100 percent level. For example for Muslims it increases by 301 percent, for Christen it increases by 156 percent, for Jains it increases by 93 percent and for other religion it is by 109 percent. But no one of such different religion level is found significant and the study observes that religion is quietly low significant to explain the occupational diversity.

Among the five caste categories, the study considers general caste as the base or reference level and found the value of odds ratios of all other caste categories subject to that base is more than 100 percent in terms of occupational diversity. For example the value of odds of occupational diversity of the households belonging to Scheduled caste is 136 percent greater than the General caste. Scheduled Tribes households' occupational diversity is 164 percent higher than the base level. For other Caste category the value of odds of occupational diversity is 203 percent greater than the base and for OBC-B category, the value of odds of occupational diversity is 431 percent higher than the General Caste category. But among such different caste level, the study observes that movement from General caste to Other Caste and to OBC-B is found significant. Thus, partially some of the caste categories in terms of logistic specification found that occupational diversity is found significant. In case of house condition the four categories are framed by the study. The 0 category is the people living under thatched house and this category is considered as the base or reference level under logistic specification. The people living under category 0 is under the most vulnerable position and so why all the odds ratios of the households living in other house conditions are found less than 1 in terms of occupational diversity. For example the odds ratio for occupational diversity between people living under mud based tin / asbestos-roofed houses is found 0.2618 with respect to base households living in thatched houses. To understand such facts in more better way, we can say that people living in thatched houses are occupationally diverse in terms of people living in mud based tin / asbestos roofed houses is $\frac{1}{0.2618} = 3.819$. Thus the occupational diversity of the household living in thatched house is 281 percent higher than the household living under mud based tin / asbestos-roofed houses. Similarly, the occupational diversity of the household living in thatched house is 219 percent higher than the household living under semi-pucca tin/ asbestos-roofed houses. The occupational diversity between thatched living households is 376 percent higher than the Pucca house living households. Regarding house condition, the study observes that movement from thatched house to mud based tin / asbestos-roofed houses and to semi-pucca tin/ asbestos-roofed houses are found significant in terms of p-value. The study concludes that under such logistic specification house condition of the households is a significant determinant of occupational diversity.

Same as house condition, number of room exists in the house is functioning becomes a significant variable to explain occupational diversity. Like house condition, most vulnerable group is living in a house with one room and that is considered as the base group. So, why all the odds ratio's are found less than one. For example if the individual moving from a house with 2 rooms to a house with one room, the odds ratio for such occupational diversity becomes $\frac{1}{0.3592} = 2.7843$. That means the occupational diversity increases by 178 percent. Similarly, if the household moves from a house with 3 rooms to 1 room, occupational diversity enhances by 221 percent and if the household moves from a house with more than 3 rooms to 1 room, occupational diversity enhances by 289 percent. The study found that such movement from 3 rooms to 1 room and 2 rooms to 1 room are found significant p-value. Thus, number of rooms within the house is found a significant determinant for occupational diversity. Study considers fuel uses for cooking is another covariates for levelling occupational diversity. The base group is the firewood uses, and like before that is the reason for the odds ratio's value less than unity. A household who is using coal if started to use firewood, the odds ratio for occupational diversity moves to $\frac{1}{0.3073} =$ 3.2544, that means the occupational diversity of that household increases by 225 percent. Similarly, if the movement is from gas to firewood, the occupational diversity enhances remarkably by 1910 percent. The study observes that the movement between firewood users and coal users is found significant in terms of p-value at 5 percent level.

The study considers value of non-farm assets (VNFA) as designed before. The data set collects the information of household based non-farm assets and converted them in monetary values with respect to the price level data of 2018 from the retail price data of CPI provided by MOSPI with base price of 2012. After normalisation of such money value of VNFA, study partitioned the whole data into four quartiles. Then the study turned the formed four groups in terms of categorical variables. The base group is considered as 0, who are in the first quartile in terms of VNFA. The logistic regression shows that odds ratio of all such movements from the base 0 group is less than unity. If the household with VNFA under category of 2nd quartile moves to first quartile the occupational diversity odds ratio becomes $\frac{1}{0.1557} = 6.4237$, which means for such movements occupational diversity enhances by 542 percent which is also with significant p-value at 5 percent level of significance. If the movement is from 3rd quartiles to 1st quartiles of VNFA, the occupational diversity enhances by 400 percent which is also with significant p-value at 5 percent level of significance, and if it is from 4th quartile to 1st quartile, the occupational diversity enhances by 385 percent, but it is found insignificant. In overall sense, the VNFA asset holding is a very important determinant of occupational diversity. Next, the study considers education status (LIT) of the head of households as an independent covariates for measuring occupational diversity. The base or reference group is completely illiterate head of the household. Now, as education proceeds the occupational diversity also increases and the results is also verified by our earlier discussion. If the individual is moving from illiteracy to the category of literate with ability to make signature occupational diversity enhances by 331 percent and it is found significant for 5 percent level. The same thing happens if the head of the household moves from illiteracy to a category of level of education less than four years, less than eight years, and less than 10 years, the occupational diversity increases by 487 percent, 429 percent and 177 percent respectively. Both of such first two movements are found significant at 5 percent level. Thus from illiteracy any improvements in education level of the head up to 10 years, the study observes that occupational diversity increases. Here, any improvement in education may enhance the chance of participation of the head in versatile occupation in compare to the illiterate, who are mainly strenuous in some low profile job. May be improvement in education up to 10 years improves the basic skills or knowledge of different non-farm activity and which helps the head of households to expand the chances of occupational diversity. But if the education level of the head is found above 10 years, the study observes that odds ratio of the movement from illiteracy is less than one. This actually means that adopting education by more than 10 years, the head of the households gets some specific options to engage himself / herself in some specific standard jobs to maintain his livelihood and it reduces the chances of occupational diversity at such education level.

Landholding or operational area in acre is considered as another significant variable by the study. The marginal farmers who hold the land by less than 0.05 acre are considered as the base or reference group. The study found that improvement of landholding from marginal farmer to small farmer as constructed by the study, the odds ratio of the occupational diversity increases by 139 percent and this is also significant at 5 percent level. But if it changes from marginal farmers to medium size farmers, the study observes that odds ratio is less than unity, which means the occupational diversity reduces. The same thing also happens for large size farmers also. Thus, holding of specific land size is a significant determinant of occupational diversity. If the land holding is greater than 0.5 acre, the chance of occupational diversity is quite low for the households of the study region. Land possession of such amount may open chances of greater income and that curtails the possibility of occupational diversity in Purulia study region. The study in its methodology considers location of the household as a significant variable to explain the occupational diversity. The households living in a poor and least poor developed blocks are identified as the base and reference group and the household living in high or moderate developed blocks are judged with respect to that base or reference group.

The study observes that under such comparison odds ratio is 381 percent and this is also significant at 5 percent level. Such outcome confirms that high and moderate developed blocks give the higher opportunity of participation in diversified occupation compared to poor and least poor developed blocks.

Dependency is created by the study as a ratio of sum of number of child and old age and number of adults in a family. The base group of dependency is 0 which actually means no dependency. Surely, for such variable for its higher value more income is required to maintain the livelihood of the households. The study observes that any movement for dependency 0 to 1 or 0 to 2 or 0 to 3 and 0 to 4, the odds ratio increases by 344, 549, 282 and 381 percent respectively. All such movements are found significant at 5 percent level. Such outcome confirms that dependency is a strong variable to determine the occupational diversity in the study area. The monthly per-capita consumption expenditure is considered as the last variable of the study. Here, the household with MPCE 0 means the vulnerable group which has the MPCE less than the study specified bench mark. Such MPCE below such bench mark is considered as base or reference group of the study. The study observes that the movement from 0 to 1, the odds ratio is less than unity. The above outcome identifies that if the household moves from group 1 to group 0, the odds ratio becomes $\frac{1}{0.3135} = 3.1899$ which is also found significant at 5 percent level. Thus, if any household moves from secure MPCE to vulnerable MPCE, it enhances the chance of occupational diversity by 219 percent.

Conclusion:

Predictiveness of the Fitted Model Over Occupational Diversity

Now, the study considers the predictiveness of such fitted model in terms of specificity and sensitivity. The predictive behaviour of the set model will help us to determine how reliable the fitted model is. Two important characters of such reliability measurements are specificity and sensitivity. By sensitivity we mean true positive rate, that may be the case of occupational diverse

(OCCDV=1). It measures the proportion of the occupational diversity that is correctly classified as such and it is complementary to the false negative rate i.e. occupationally non-diverse (OCCDV=0). Thus sensitivity is measured as :

$$Sensitivity = \frac{true\ positive\ rate}{true\ positive\ rates + false\ negative\ rates}$$

Similarly, the specificity means true negative rate i.e. here the case of occupationally non-diverse (OCCDV=0). It means proportion of occupationally non-diverse that is correctly classified as such and it is complementary to the false positive rate i.e occupationally diverse (OCCDV=1).

$$Specificity = \frac{true\ negative\ rate}{true\ negative\ rates + false\ positive\ rates}$$

Estimating the probability of each household and comparing their estimated probability with actual position in terms of either occupationally diversity or occupationally non-diversity, the study classified the households in below table with cut-off point equal to 0.5 (Figure-19).Firstly, by default the software (Jamovi-2.3.3.0) considers the cut-off points equal to 0.5. When the cut-off value is set at 0.5, then the predicted model clarifies that 97.5 percent of occupationally non-diverse is correctly specified and 99.1 percent of occupationally diverse is correctly specified. The study also added that over all accuracy of such predicted model is 98.5 percent with the value of specificity 97.5 percent and sensitivity 99.1 percent. The area under the Relative Operating Characteristics (ROC) curve (AUC) is an aggregate metric which evaluates how good a logistic regression model classifies occupational diversity and non-diversity outcomes at all possible cut-offs. At cut-off levels 0.5, the AUC is 99.9 percent. The possibility of such cut-off points generating Type-I error is equal to $0.9 = (100 - 99.1)$ percent and Type-II error is $2.5 = (100 - 97.5)$ percent. The cut-off curve identifies the differences between intersection of the two curves, like specificity and sensitivity curves with respect to the identified bench mark at 0.5 level. The difference between these two gives us some prior responsibility to minimise the overall error (Table-21). Then we have to modify the sensitivity which actually means proportion of occupationally diverse (OCCDV=1) is correctly specified and specificity identifies as proportions of occupationally non-diverse (OCCDVS=0) (Figure-20).

So, now the study changes to cut-off points at 0.55 percent to match the intersection between specificity and sensitivity with the cut-off points. Thus, for changing of such cut-offs, the Type-I error increases to $1.6 = (100 - 98.4)$ and Type-II error reduces to $1 = (100 - 99)$. For such change in cut-offs from 05 to 0.55, the overall accuracy of the predictive model increases to 98.7 percent which is greater than 98.5 percent of the earlier model. The AUC curve area of newly fitted model remains same (0.999) as before (Figure-21& 22). Thus, when the cut-off value is set at 0.55, then the predicted model clarifies that 99 percent of occupationally non-diverse are correctly specified and 98.4 percent of occupationally diverse are correctly specified (Table-22).The study discusses vividly the nature of occupational diversity in the study area. On the basis of existing literature, the study sets a number of numerous social and economic variables which directly as well as indirectly influences occupational diversity. The study fitted a binomial logistic regression model on the basis of such independent covariates over the discrete occupational diversity choice. Most of the independent variables assumed by the study significantly explain the occupational diversity character in terms of overall model test and also by their different individual levels. The study observes that a numerous factors like family size, caste categories, house condition, number of rooms in the house, fuel use, value of non-farm assets holdings, education years, land holdings, location of the households, dependency factors and monthly per-capita consumption expenditure play a critical role to explain the occupational diversity. The study receives quite satisfactory model fitting results and the predictiveness of the model explained by the classification tables and in terms of ROC curves.

References:

1. Kuznets, S., 1966, Modern Economic Growth, New Haven, CT, Yale University Press

2. Unni, J., 1994, Labour Participation Decisions of Married Women in Rural India, Indian Economic Review, Vol. 29(2), pp-177-193.

3. Unni, J., 1998, Non-Agricultural Employment and Poverty in Rural India: A Review of Evidence. Economic and Political Weekly, 33(13):A36 – A44.

4. Unni J., 1996, Non Agricultural Employment and Rural Livelihoods: Macro vis a vis Micro view. The Indian Journal of Labour Economics, 39(4):795-807.

3. Chandrasekhar, B., 1993, Sociology of Development and Change, Orient Longman, New Delhi

4. Mellor, J. W., 1976, The New Economics of Growth: A Strategy for India and Developing World, Cornel University Press, New York

5. Hariss, B., 1987, Regional Growth Linkages from Agriculture, Journal of Development Studies, Vol. 23(2), pp-275-89

6. Hart, G., 1989, The Growth Linkages Controversy: Some lessons from MUDA case, Journal of Development Studies, Vol. 25(4), pp-571-5

7. Hart, G., 1993, Regional Growth Linkages in the Era of Liberalisation: A critique of New Agrarian Optimism, World Employment Programme Research Working Paper No. 37, Geneva, ILO

8. McGee, T. G., 1971, The Urbanisation Process in a Third World: Explorations in Search of a Theory, G Bell & Sons Ltd., London.

9. Vaidyanathan, A., 1986, Labour Use in Rural India- A Study on Spatial and Temporal Variations, EPW, Vol. 21 (52).

10. Ellis, F., 1998, Survey Article: Household Strategies and Rural Livelihood Diversification, Journal of Development Studies, Vol. 35(1), pp-1-38

11. Fisher T and Mahajan V. 1997. The Forgotten Sector. New Delhi: Oxford and IBH Publishing Co.

12. Ellis, F., 2000, Rural Livelihood and Diversity in Developing Countries, Oxford University Press

13. Davis B., Winters P., Covarrubias G.K., Quinones E., Zezza A., Stamoulis K., Bonomi G. and DiGuseppe S. 2007. Rural Income Generating Activities: Cross-Country Comparison. ESA Working Paper No. 07-17. Rome: FAO

14. Brenstein, H., Crow, B., &Jhonson, H., 1992, Rural Livelihood Crisis and Response, Oxford University Press

15. Bhagat, R. B., 2011, Emerging Pattern of Urbanization in India. Economic and Political Weekly Vol. 46(34), pp- 10-12.

16. Papola, T.S., 2013, Employment Growth during the Post-Reform Period. The Indian Journal of Labour Economics, 56(1):1-13.

17. Papola T.S., 1994, Structural Adjustment, Labor Market Flexibility and Employment. The Indian Journal of Labour Economics, 37(1):3-16.

18. Reddy, D.N., and Venkatanarayan, M., 2013, Declining Labour Use in Agriculture: A Case of Rice Cultivation in Andhra Pradesh, MPRA Working paper No. 49204. Hyderabad, Telangana: NIRD.

19. Biggs, S, Justice, S and Lewis, D., 2014, Patterns of Rural Mechanization, Energy and Employment in South Asia: Reopening the Debate. Economic and Political Weekly, XLVI(9): 78-82.

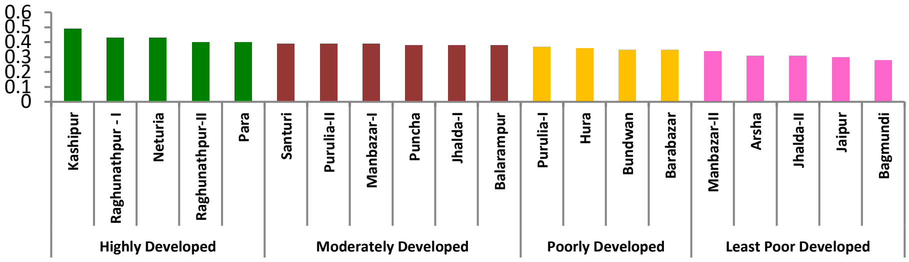
20. Martin, S. M., Lorenzen, K., &Bunnefeld, N. 2013, Fishing farmers: Fishing, livelihood diversification and poverty in rural Laos. Human Ecology, 41(5), 737–747.

21. Loison, S. A., 2019, Household livelihood diversification and gender: Panel evidence from rural Kenya. Journal of Rural Studies, 69, 156–172.

22. District Human Development Report, Purulia, 2012, Govt. of West Bengal.

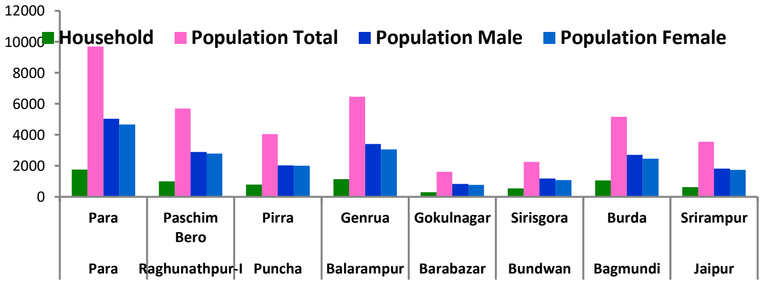
LIST OF FIGURES

Figure-1 : The Blocks of Purulia with HDI scores and Ranking



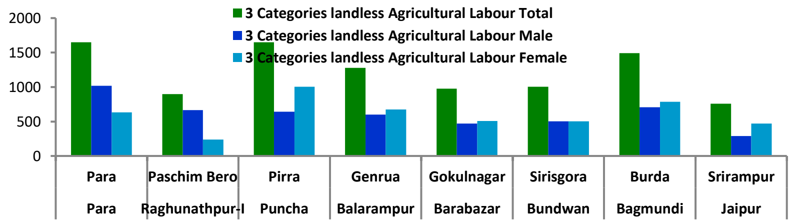
Source: DHDR, Purulia, 2012

Figure-2: Village selection on the basis of Household size



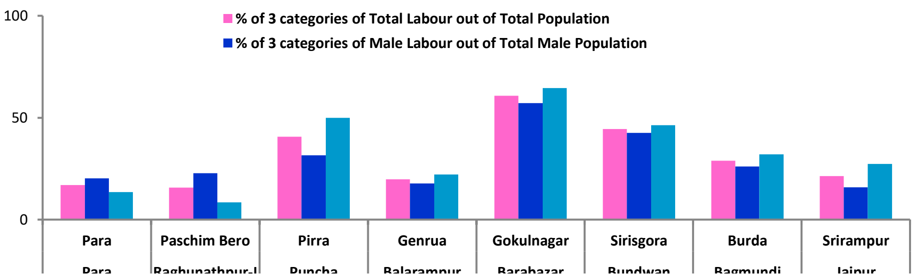
Source: Study Estimation, 2021

Figure-3: Village selection on the basis of 3 Categories of Selected Labour



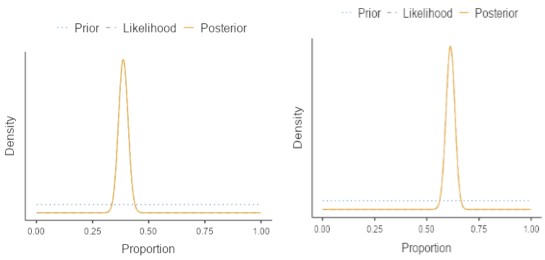
Source: Study Estimation, 2021

Figure-4: Variations in inter-village % of three categories of Labour



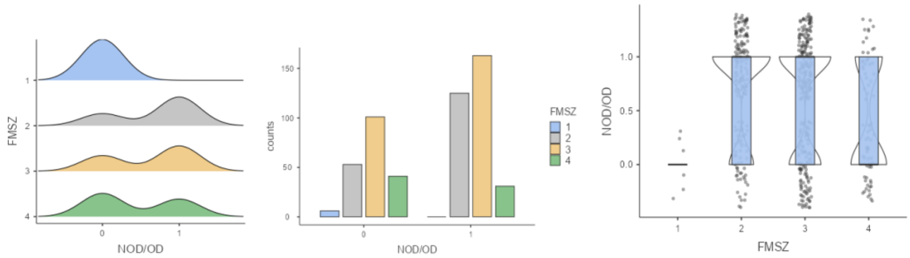
Source: Study Estimation, 2021

Figure-5 : Density based Distribution of Occupational Diversity



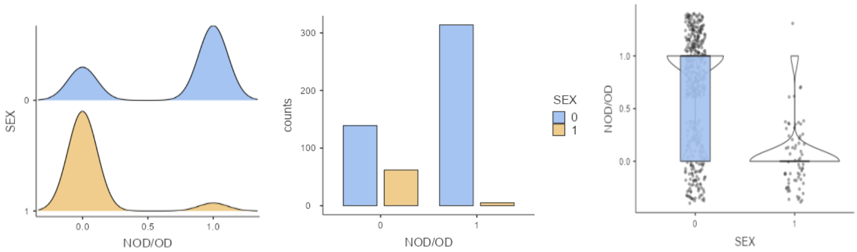
Source: Study Estimation, 2021

Figure-6 : Densities, Bars and Box-Plots of Family Size



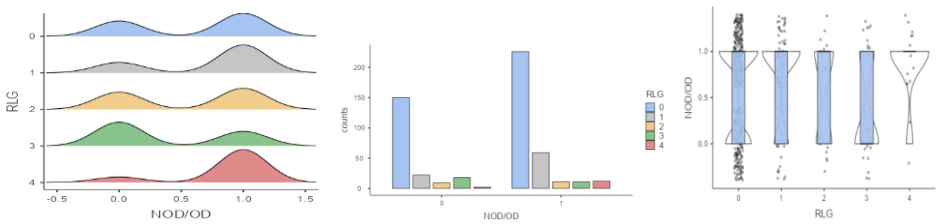
Source: Study Estimation, 2021

Figure-7 : Densities, Bars and Box-Plots of Sex of the Head



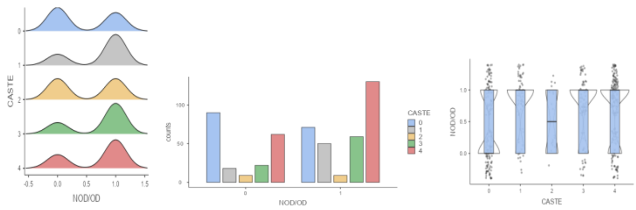
Source: Study Estimation, 2021

Figure-8 : Densities, Bars and Box-Plots of Religion of the Household



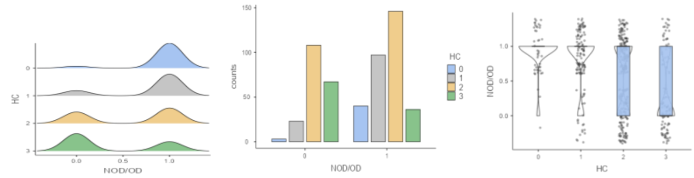
Source: Study Estimation, 2021

Figure-9 : Densities, Bars and Box-Plots of Caste of the Households



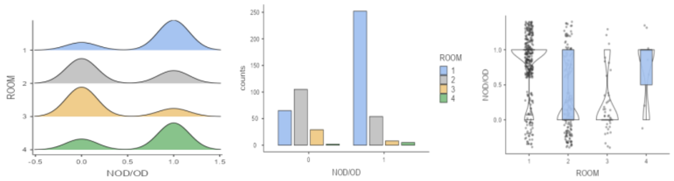
Source: Study Estimation, 2021

Figure-10 : Densities, Bars and Box-Plots of House Conditions



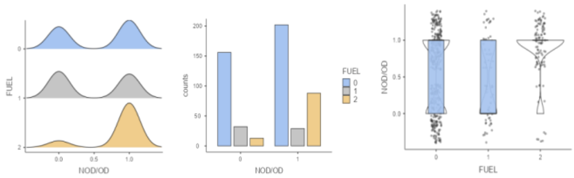
Source: Study Estimation, 2021

Figure-11 : Densities, Bars and Box-Plots of Number of Rooms



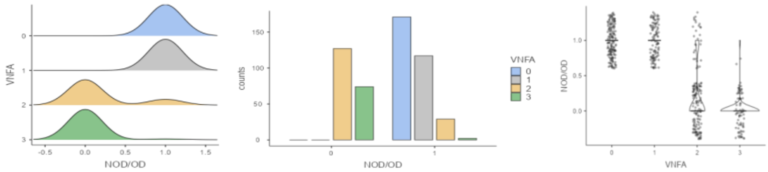
Source: Study Estimation, 2021

Figure-12 : Densities, Bars and Box-Plots of Fuel Type Uses



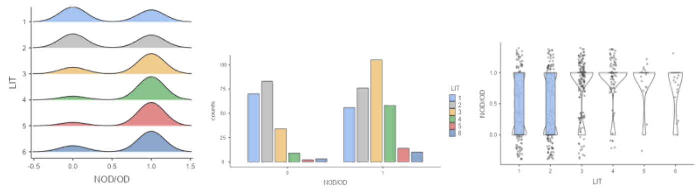
Source: Study Estimation, 2021

Figure-13 : Densities, Bars and Box-Plots of Value of Non-Farm Assets



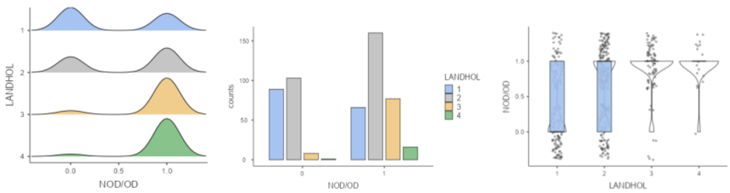
Source: Study Estimation, 2021

Figure-14 : Densities, Bars and Box-Plots of Education Status of the Head



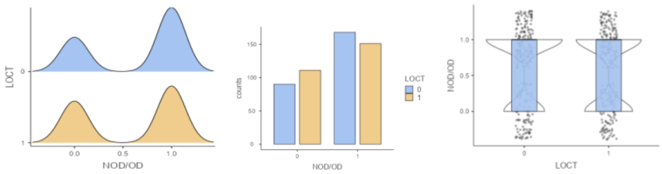
Source: Study Estimation, 2021

Figure-15 : Densities, Bars and Box-Plots of Landholding by the Households



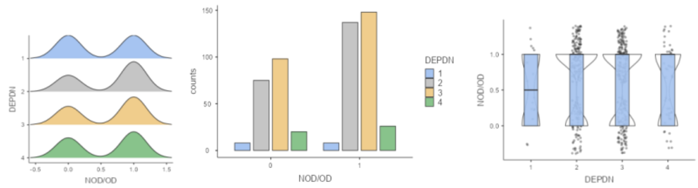
Source: Study Estimation, 2021

Figure-16 : Densities, Bars and Box-Plots of Locational Factors



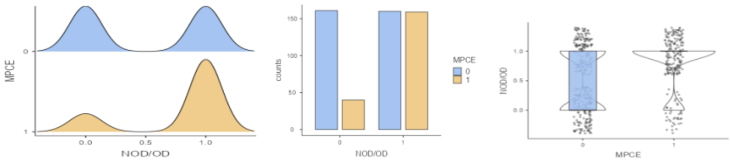
Source: Study Estimation, 2021

Figure-17 : Densities, Bars and Box-Plots of Dependency



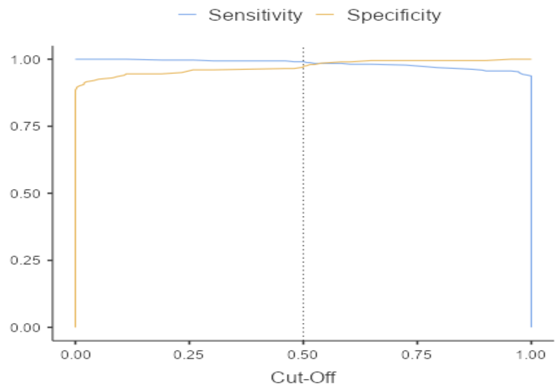
Source: Study Estimation, 2021

Figure-18 : Densities, Bars and Box-Plots of MPCE



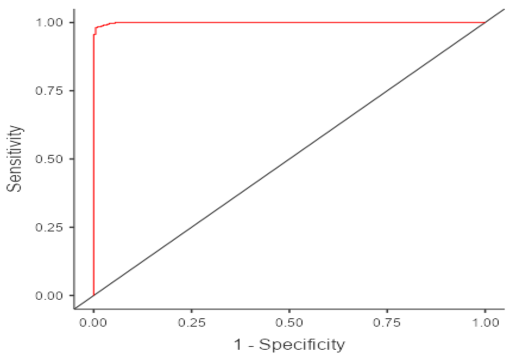
Source: Study Estimation, 2021

Figure-19: Specificity and Sensitivity Curves at the Cut-offs at 0.5



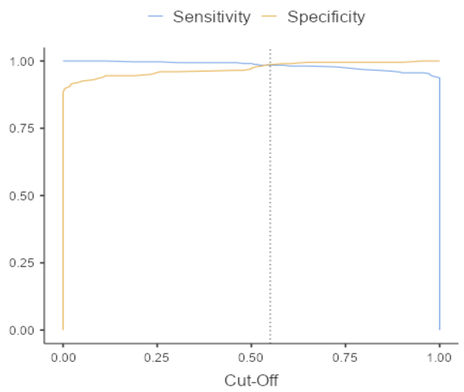
Source: Study Estimation, 2021

Figure-20: Area under Relative Operating Characteristics (ROC) Curve (AUC) at the Cut-off 0.5



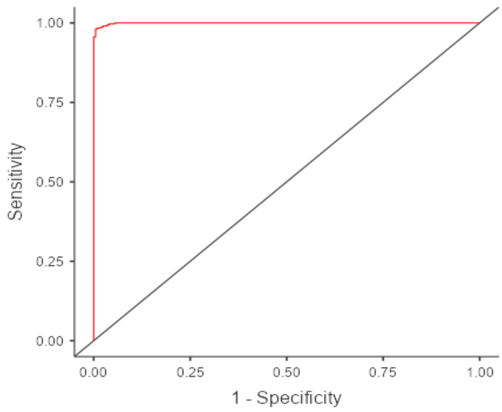
Source: Study Estimation, 2021

Figure-21: Specificity and Sensitivity Curves at the Cut-offs at 0.55



Source: Study Estimation, 2021

Figure-22: Area under Relative Operating Characteristics (ROC) Curve (AUC) at the Cut-off 0.55



Source: Study Estimation, 2021

LIST OF TABLES

Table-1: List of Sample selected Blocks

HDI	Blocks
Highly Developed	Raghunathpur-I, Para
Moderately developed	Puncha, Balarampur
Poorly developed	Bundwan, Barabazar
Least poor Developed	Jaipur, Bagmundi

Source: Study Estimation, 2021

Table-2: Set of Dependent and Independent variables

Variable	Code	Type	Nature	Levels
Occupational Diversity	OCCDV	Dependent	Binary	0, 1
Family Size	FMSZ	Independent	Categorical	0, 1, 2, 3
Sex	SEX	Independent	Dichotomous	0, 1
Religion	RLG	Independent	Categorical	0, 1, 2, 3, 4
Caste	CASTE	Independent	Categorical	0, 1, 2, 3, 4, 5
House Condition	HC	Independent	Categorical	0, 1, 2, 3
Room	ROOM	Independent	Categorical	1, 2, 3, 4
Fuel	FUEL	Independent	Categorical	0, 1, 2
Value of Non-Farm Assets	VNFA	Independent	Categorical	0, 1, 2, 3

Education	LIT	Independent	Categorical	0, 1, 2, 3, 4, 5
Land Holding	LNDHOL	Independent	Categorical	1, 2, 3, 4
Location	LOCT	Independent	Dichotomous	0, 1
Dependency	DEPDN	Independent	Categorical	0, 1, 2, 3, 4
Monthly Per-Capita consumption Expenditure	MPCE	Independent	Dichotomous	0, 1

Source: Study Estimation, 2021

Table-3 : Binomial Test Results of Occupational Diversity

	Level	Count	Total	Proportion	p	Bayes factor$_{10}$	95% Credible Interval	
							Lower	Upper
OCCDV	0	201	520	0.387	< .001	39216	0.346	0.429
	1	319	520	0.613	< .001	39216	0.571	0.654
Note. H_a is proportion ≠ 0.5								

Source: Study Estimation, 2021

Table-4 : Parametric ANOVA, Homogeneity of Variance & Normality of Family Size

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	p	Statistic	P
FMSZ	6.07	3	2.025	8.91	< .001	28.7	3	516	< .001	0.75	< .001
Residuals	117.23	516	0.227								

Source: Study Estimation, 2021

Table-5 : Non-Parametric ANOVA with DSCF of Family Size

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	25.6	3	< .001	0	1	5.11	0.002
				0	2	4.32	0.012
				0	3	2.91	0.168
				1	2	-2.59	0.258
				1	3	-5.67	< .001
				2	3	-4.02	0.023

Source: Study Estimation, 2021

Table-6 : Parametric ANOVA, Homogeneity of Variance & Normality of Sex of the Head

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	P	Statistic	P
SEX	22.3	1	22.33	115	< .001	141	1	518	< .001	0.688	< .001
Residuals	101	518	0.195								

Source: Study Estimation, 2021

Table-7 : Non-Parametric ANOVA with DSCF of Sex of the Head

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	94	1	< .001	0	1	- 13.7	< .001

Source: Study Estimation, 2021

Table-8 : Non-Parametric ANOVA with DSCF of Caste of the Households

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	33.2	4	< .001	0	1	5.752	< .001
				0	2	0.673	0.99
				0	3	5.972	< .001
				0	4	6.301	< .001
				1	2	- 2.689	0.317
				1	3	- 0.133	1
				1	4	- 1.262	0.9
				2	3	2.659	0.328
				2	4	2.142	0.553
				3	4	- 1.185	0.919

Source: Study Estimation, 2021

Table-9 : Parametric ANOVA, Homogeneity of Variance & Normality of House Conditions

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	p	Statistic	P
HC	16.4	3	5.476	26.4	< .001	86.5	3	516	< .001	0.875	< .001
Residuals	106.9	516	0.207								

Source: Study Estimation, 2021

Table-10 : Parametric ANOVA, Homogeneity of Variance & Normality of Number of Rooms

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	p	Statistic	P
ROOM	28.3	3	9.425	51.2	< .001	11.7	3	516	< .001	0.865	< .001
Residuals	95	516	0.184								

Source: Study Estimation, 2021

Table-11 : Non-Parametric ANOVA with DSCF of Number of Rooms

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	69.1	3	< .001	1	2	-2.64	0.242
				1	3	-6.29	< .001
				1	4	-9.02	< .001
				2	3	-6.24	< .001
				2	4	-9.82	< .001
				3	4	-5.45	< .001

Source: Study Estimation, 2021

Table-12 : Non-Parametric ANOVA with DSCF of Fuel Type Uses

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	36.8	2	< .001	0	1	-1.82	0.402
				0	2	7.98	< .001
				1	2	7.68	< .001

Source: Study Estimation, 2021

Table-13 : Parametric ANOVA, Homogeneity of Variance & Normality of Value of Non-Farm Assets

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	p	Statistic	P
VNFA	97.7	3	32.5831	658	< .001	147	3	516	< .001	0.511	< .001
Residuals	25.6	516	0.0495								

Source: Study Estimation, 2021

Table-14 : Parametric ANOVA, Homogeneity of Variance & Normality of Education Status of the Head

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	p	F	df1	df2	p	Statistic	P
LIT	15	5	2.998	14.2	< .001	40.3	5	514	< .001	0.859	< .001
Residuals	108.3	514	0.211								

Source: Study Estimation, 2021

Table-15 : Non-Parametric ANOVA with DSCF of Landholding by the Households

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	P			W	P
NOD/OD	61.3	3	< .001	1	2	5.11	0.002
				1	3	10.229	< .001
				1	4	5.695	< .001
				2	3	7.225	< .001
				2	4	3.885	0.031
				3	4	0.659	0.967

Source: Study Estimation, 2021

Table-16 : Parametric ANOVA, Homogeneity of Variance & Normality of MPCE

ANOVA - NOD/OD						Homogeneity of Variances Test (Levene's)				Normality Test (Shapiro-Wilk)	
	Sum of Squares	df	Mean Square	F	P	F	df1	df2	p	Statistic	P
MPCE	11.1	1	11.097	51.2	< .001	178	1	518	< .001	0.805	< .001
Residuals	112.2	518	0.217								

Source: Study Estimation, 2021

Table-17 : Non-Parametric ANOVA with DSCF of MPCE

Kruskal-Wallis				Pairwise comparisons - NOD/OD			
	x^2	df	p			W	P
NOD/OD	46.7	1	< .001	0	1	9.67	< .001

Source: Study Estimation, 2021

Table-18: Chi-Square & Likelihood Ratio Test of Independent Covariates

	x^2 Tests			Likelihood ratio test		
	Value	Df	p	Value	df	P
OCCDV VS FMSZ	15.9	3	0.001	18.2	3	<.001
OCCDV VS SEX	0.26	1	0.61	0.262	1	0.608
OCCDV VS RLG	7.16	4	0.128	7.26	4	0.123
OCCDV VS CASTE	7.54	4	0.11	7.83	4	0.098
OCCDV VS HC	61.4	3	< .001	71.2	3	<.001
OCCDV VS ROOM	119	3	< .001	121	3	<.001
OCCDV VS FUEL	28.8	2	< .001	28.3	2	<.001
OCCDV VS VNFA	412	3	< .001	526	3	<.001
OCCDV VS LIT	33.6	5	< .001	340	**5**	<.001
OCCDV VS LNDHOL	146	3	< .001	199	**3**	<.001
OCCDV VS LOCT	3.07	1	0.08	3.7	1	0.08
OOCDV VS DEPDN	33.8	4	< .001	34.5	4	<.001
OCCDV VS MPCE	42.2	1	< .001	42.1	1	<.001

Source: Study Estimation, 2021

Table-19: Model Fit Measure Statistics of Occupational Diversity

Model Fit Measures								
						Overall Model Test		
Model	**Deviance**	**AIC**	**R^2_{McF}**	**R^2_{CS}**	**R^2_N**	**x^2**	**df**	**P**
1	3.35	10.1	0.0452	0.0719	0.0976	66.21	33	< .001

Source: Study Estimation, 2021

Table-20: Model Coefficients, Odds Ratio's of Independent Covariates

Model Coefficients - OCCDV							
						95% Confidence Interval	
Predictor	**Estimate**	**SE**	**Z**	**p**	**Odds ratio**	**Lower**	**Upper**
Intercept	114.4	146.21	0.7824	1	4.8227E+49	0	Inf
FMSZ:							
1 – 0	1.52	1.0954	1.3876	<.001	4.5722	0.0612	1.6871
2 – 0	1.37	0.8954	1.5300	<.001	3.9354	0.8739	1.9955
3 – 0	1.88	0.9942	1.8910	<.001	6.5535	1.2054	2.3465
SEX:							
1 – 0	-2.69	1.6874	-1.5942	0.996	0.06788	0.00247	1.69784
RLG:							
1 – 0	1.39	0.8645	1.6079	0.372	4.0149	0.08274	7.7341
2 – 0	0.94	0.8957	1.0495	0.999	2.5600	0.3254	3.5671
3 – 0	0.66	0.5724	1.1530	0.031	1.9348	0.9731	4.8177
4 – 0	0.74	0.6482	1.1416	0.092	2.0959	0.3363	3.34E+00
CASTE							
1 – 0	0.86	0.5642	1.5243	0.364	2.3632	1.0244	2.8671
2 – 0	0.97	0.7112	1.3639	0.564	2.6379	1.1387	2.9562
3 – 0	1.11	0.3642	3.0478	<.001	3.0344	1.1946	2.8677
4 – 0	1.67	0.8674	1.9253	<.001	5.3122	1.9967	3.6871
HC:							
1 – 0	-1.34	0.4628	-2.8954	0.079	0.26184567	0.0363	0.6984
2 – 0	-1.16	0.3044	-3.8108	<.001	0.31348618	0.6981	1.7854
3 – 0	-1.56	0.5124	-3.0445	<.001	0.21013607	0.7425	1.2897
ROOM:							
2 – 1	-1.024	0.2874	-3.5630	<.001	0.35915544	0.0061	0.0279
3 – 1	-1.167	0.4328	-2.6964	<.001	0.31129944	0.0094	0.0455
4 – 1	-1.357	0.6714	-2.0211	0.872	0.25743192	0.0096	0.3452
FUEL:							
1 – 0	-1.18	0.6749	-1.7484	<.001	0.30727874	0.0049	0.1058
2 – 0	-2.95	1.85	-1.5946	0.11	0.05233971	0.0093	0.1679
VNFA:							
1 – 0	-1.86	0.7358	-2.5279	<.001	0.15567263	0.0062	0.0591
2 – 0	-1.61	0.8664	-1.8583	<.001	0.19988761	0.0086	0.0943
3 – 0	-1.58	0.6717	-2.3522	0.995	0.2059751	0.0167	0.8672

LIT:							
1 – 0	1.46	0.8674	1.6832	<.001	4.30595953	1.4237	2.7618
2 – 0	1.77	0.6784	2.6091	<.001	5.87085336	1.9432	3.4472
3 – 0	1.66	0.5877	2.8246	<.001	5.25931084	2.0254	3.6671
4 – 0	1.02	0.8651	1.1791	0.347	2.77319476	1.6427	3.6651
5 – 0	-0.86	0.7489	-1.1484	0.442	0.42316208	0.1237	1.2779
LNDHOL:							
2 – 1	0.87	0.3241	2.6844	<.001	2.38691085	1.9628	3.4519
3 – 1	-0.92	0.5443	-1.6902	<.001	0.39851904	0.0068	0.0161
4 – 1	-1.86	0.8227	-2.2608	0.999	0.15567263	0.0087	0.4672
LOCT:							
1 – 0	1.57	0.8771	1.7900	<.001	4.80664819	1.04199	15.8051
DEPDN:							
1 – 0	1.49	0.3871	3.8491	<.001	4.43709552	1.6172	3.2281
2 – 0	1.87	0.4624	4.0441	<.001	6.4882964	2.4229	4.6652
3 – 0	1.34	0.3794	3.5319	<.001	3.81904351	1.2167	3.2433
4 – 0	1.57	0.4447	3.5305	<.001	4.80664819	1.7421	3.9649
MPCE:							
1 – 0	-1.16	0.8671	-1.3378	<.001	0.31348618	1.2476	2.6569
***Note.* Estimates represent the log odds of "OCCDV = 1" vs. "OCCDV = 0"**							

Source: Study Estimation, 2021

Table-21: Classification Tables of Occupational Diversity at Cut-off 0.5

Classification Table – OCCDV			
	Predicted		
Observed	**0**	**1**	**% Correct**
0	196	5	97.5
1	3	316	99.1
Note. The cut-off value is set to 0.5			
Predictive Measures			
Accuracy	**Specificity**	**Sensitivity**	**AUC**
0.985	0.975	0.991	0.999
Note. The cut-off value is set to 0.5			

Source: Study Estimation, 2021

Table-22: Classification Tables of Occupational Diversity at Cut-off 0.55

Classification Table – OCCDV			
	Predicted		
Observed	**0**	**1**	**% Correct**
0	199	2	99
1	5	314	98.4
Note. The cut-off value is set to 0.55			
Predictive Measures			
Accuracy	**Specificity**	**Sensitivity**	**AUC**
0.987	0.99	0.984	0.999
Note. The cut-off value is set to 0.55			

Source: Study Estimation, 2021