Empirical Economic Review (EER)
Volume 6 Issue 1, Spring 2023
ISSN$_{(P)}$: 2415-0304 ISSN$_{(E)}$: 2522-2465
Homepage: https://ojs.umt.edu.pk/index.php/eer

Article QR

Title:	**Testing For Homogenous or Heterogenous Doers in Longitudinal Latent Class Regression Framework**
Author (s):	Saima Ishaq , Ahsan-ul-Haq Satti
Affiliation (s):	Pakistan Institute of Development Economics (PIDE), Islamabad, Pakistan
DOI:	https://doi.org/10.29145/eer.61.02
History:	Received: September 7, 2022, Revised: April 15, 2023, Accepted: March 15, 2023, Published: June 30, 2023
Citation:	Ishaq, S. (2023). Testing for homogenous or heterogenous doers in Longitudinal latent class regression framework. *Empirical Economic Review*, *6*(1), 00–00. https://doi.org/10.29145/eer.61.02
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
Department of Economics and Statistics, Dr. Hasan Murad School of Management
University of Management and Technology, Lahore, Pakistan

Testing for Homogenous or Heterogenous Doers in Longitudinal Latent Class Regression Framework

Saima Ishaq[1] and Ahsan-ul-Haq Satti

Pakistan Institute of Development Economics (PIDE), Islamabad

Abstract

Latent class regression analysis is applied in context of conditional and unconditional analysis. The empirical analysis is conducted in novel way for exploratory and confirmatory perspective utilizing longitudinal British household data of Understanding Society. The study aims to explore the profile differences for subjective satisfaction towards work and confirms the absence of differential effects of job-related variables across the explored broad classes of satisfied and non satisfied job doers. For further insights into behaviour of selected classes, conditional models are employed. Step 3 approach is utilized in this regard for investigating the contribution of background variables such as gender, age, occupation and quality of life for shaping their response of being satisfied or non-satisfied with their jobs. This study overall tests and confirms the absence of heterogenous triggers for job satisfaction in British society.

Keywords: mixture models, latent class regression analysis, Step-3 analysis, unobserved heterogeneity, differential effects, job satisfaction.

Introduction

"Mixture models" were primarily presented in order to better accommodate for the overlooked heterogeneity in the population (Agresti et al., 2000). In context of mixed mode data, the assumption of data emerging from various distributions is more relevant and the task of un-mixing the distributions is theoretically more comprehensible since the sampled population is inherently measured at different scales. With the help of generalized linear models we can model dependent variables as a function of explanatory variables using link functions for various distributions of exponential family Skrondal and Rabe-Hesketh (2004) but for varied kind of variables one aggregate solution may mask the diversified behaviour of the entities.The

[1]Corresponding Author: saimaishaq_16@pide.edu.pk

aggregate regression solution is also inadequate for the mix of population if we hypothesize unknown classes in data.

Contrary to standard regression framework we can suppose a mix of subgroups to explain the structure of the overall population under mixture regression framework which have the advantages of relaxing traditional ordinary least squares assumptions. Units of measurement under these models are flexible enough to incorporate mixed nature of structure. Regression mixture approach allows for simultaneous estimation of regression equations after classifying individuals into distinctive classes. The alternative approach "Regression mixtures" is the subfamily of finite mixtures (McLachlan et al., 2019; Wedel & Kamakura, 2000).These mixture models involve continuous latent variables for finding unknown groups. The advantage of utilizing these methods is dual; they serve as an exploratory exercise for finding subgroups or classes in heterogeneous data and after class enumeration the parameters can serve to find across classes meaningful differences. As a by-product of this modelling scheme different regression relations for each class can be described for focused analysis. Regression mixture (RM) is a nonparametric random-effects model (Simonoff, 2003; Skrondal & Rabe-Hesketh, 2004; Tuma & Decker, 2013). These models adhere to the complexities inherent in real life data measurements (Qu et al., 1996).When we treat categorical data or similarly encounter discrete distributions the RM can be restricted with discrete latent variable for finding unknown classes in the given data. In that case the version of RM is known as latent class regression model (LCRM). LCRM is of great use within the economic context to incorporate qualitative differences in the effects of a predictor variable on an outcome and vice versa to measure heterogeneity in class effects.

Mixture of Regressions/ latent class regression (LCR) are among the most widely used approaches for dealing with heterogeneity in regression problems. For applications in closely related field to economics see market segmentation case study by Tuma and Decker (2013), and Wedel and Kamakura (2000) present a nice introduction to this version of mixture models. They highlight the strength of regression mixtures as a special case of finite mixtures to accommodate hypothesis testing within statistical standard theory. Also, the flexibility to accommodate dependent variables of scale types other than nominal or ordinal, and the possibility of estimating conditional and unconditional models is discussed by them.

In Hartzel et al. (2001), we find nice and non-technical introduction to RM and their great potential for use in socials sciences. The paper presents theoretical claims of standard regression models with a list of applications scope in marketing research. Extensions of the basic structure influenced by background variables is also described with two empirical applications. One is for trade performance show, and another application is of conjoint based study. The writers discussed different classes in terms of their different attributes and provide preference of discrete latent variable for classifying groups over the continuous latent variable in such cases. Extending to Hartzel et al. (2001) for various discrete and categorical variables. Vermunt (2005), and Skrondal and Hesketh (2004) presented mixed-effects applied cases priorly extensively. The writers presented an extension of mixed effects logistic regression model for cases in which the dependent variable is a discrete latent variable measured with multiple indicators. The writers offered improved maximum- likelihood based solution for the multivariate case and adapted the E step of the EM algorithm making use of the conditional independence assumptions implied by the model. The model was illustrated with an example from organizational research in which they built classes based on latent task-variety differences. After controlling for individual-level covariates they found significant indication for betweencluster variation in the latent class distribution of different clusters.

Yamaguchi (2000) has also discussed mixed effect model for Japanese women for measuring gender biases in case of vote and support for female participation in workforce. The data is taken from the national base for measuring attitudes of Japanese people towards social norms in society. The models opted to formalize the relationship between indicators for various possible groups are regression extensions of log-linear latent-class models with group variables. The application focuses on predictors of three latent classes of gender role attitudes among Japanese women. Three class regression solutions emerged in their case as the best solution, and the classes were labeled as “anti-work gender equality supporters” “traditional gender role supporters,” “pro-work gender-equality supporters,” following their probabilistic response patterns differences. Each class had different characteristics which were explained in terms of distinctive parameters.

Kim et al. (2016) have guided in their article for the model building process for including a latent class predictor in regression mixture models. For this purpose, they designed low to high separated mixtures. First, they estimated an 'unconditional model' which included no latent class predictors. The model was explained in terms of differences between classes in regression weights. Then the writers included latent class predictor into the model and compared whether there were any substantive differences in the model results. Contribution of predictors in class predictors was questioned by comparing different models and the difference was compared in terms of differential effects in class structure to the baseline no predictors case. Exclusion of predictors may lead to bias in class sizes was also addressed in that study. Finally, they used an applied dataset to show the effects of omitting the direct effect from the latent class predictors to an outcome variable. Latent class Regressions (LCR) models have been particularly popular in various other fields (see Andrews & Currim, 2003; Bierbrauer et al., 2004). Still their Application in economics is not much in tradition. One study by Sánchez and Puente (2015) is of peculiar interest since it sets first example for the differential effects of mismatch inherent in educational and skills of workers related to job satisfaction outcome.

The paper studied Spanish labor market by employing regression mixtures and explained significant differential effects of relevant labor market indicators on job satisfaction. The related job quality indicators included in this paper were salary, promotion chances, number of working hours and kind of tasks performed. Using Survey of Quality of Working Life, the writers found that highly educated individuals show higher levels of dissatisfaction than those with low qualification. Separate consideration of educational and skill mismatches was emphasized in conclusion. Extending to Sánchez and Puente (2015) we have evaluated unconditional and conditional models (see section 2 for details) for longitudinal sample of more than 8600 individuals. How are specific dimensions of job quality linked to job satisfaction in various classes of workers? To explore this query, we employed LCRM to find the best combination of classes for explaining differential impacts of work-related features such as job nature (full time or part time), work arrangements (standard or nonstandard), job size, working hours and workplace size.

Job satisfaction is generally perceived to be influenced by core job quality feature. In regressing modelling environment, we have tried to explore the links of subjective job quality indicators (job satisfaction level) with some objective job characteristics. To the best of our knowledge this study would contribute to empirical testing of heterogenous job satisfaction segments in British society by a novel approach. The variants of conditional models will further validate for the assumed differences in labor class due to socio economic differences.

In next section we provide some background for data and models employed. Section 3 presents unrestricted unconditional latent class regression models followed by restricted variants and conditional step 3 case. Lastly, the article is concluded.

Data and Models

The empirical data employed to find differential effects of work features on job quality is longitudinal and consists of 9 waves over the years. It is taken from a national representative sample of British households' data source labelled understanding society. For further technical details of data scheme look into report by Understanding Society (2021).We have picked some crucial indicators of work for exploring their impact on job satisfaction, which is regarded as subjective indicator of job quality in this context (Clark, 2005). For measuring the differential, the repeated sample of over 9 years (2010 to 2018) is taken from specific adult survey Teaching longitudinal dataset of British households. The specific data set is chosen for ease of handling since this is designed to facilitate longitudinal data uses in class rooms and academic research. For sample size , data collection techniques and other technical queries (Understanding Society, 2021).

The rationale for choosing this sample is also linked to availability of variables of interest over the time for a specific age group of adults. The specific longitudinal design of the survey allows to access individuals with maximum response over the years. The variables of interest include age, degree level, occupation categories, working hours' length, organization size and work schedules. Latent class regression mixtures (LCRM) are applied to test for any possible source of unseen heterogeneity in job quality within the sampled observations. The data set spans over 9 years and some subjective and objective indicators of work nature are picked to address the extent of difference in their impact on job quality. Additional important predictors included for measuring their varying effects on job satisfaction were hours worked in week and company size. We have not taken the lengthy list of predictors in this modelling scenario since the data consists of repeated measures and there were distinctive groups with stark age and working status differences, therefore there were quite missing information when we attempted to include more work-related features. We limited the scope of variables also to measure the differences for better understanding the extent and source of differences and for avoiding complicated cross classifications in multivariable analysis.

All of the variants are applied on longitudinal employment data in results section; additional consideration sin analysis were inclusion of sampling weight and complex sampling standard errors calculations since the sample chosen was clustered sample (see Table A in Appendix).

Latent Class Regression Models

The technical frame work of regression mixtures is adopted from the study of Khalili & Chen, (2007). Eq 1 stated below serves as the base structure based on generalized linear models discussed in Skrondal & Hesketh (2004), and Vermunt & Magidson (2013). Here $llll$ serves as a latent variable which relates exogenous (covariates or predictors) depicted as ex, dependent variables are indicated as y over the index I, the relation between these variables is described as

$$ff(\mathrm{y}_{ii} \mid \mathrm{e}_{ii}) = \sum^{KK}_{xx=1} PP(llll \mid ee_{ii}) ff(\mathrm{y}_{ii} \mid llll, eeee_{ii}) = \sum^{KK}_{xx=1} PP(llll \mid \mathrm{ex}_{ii}) \prod^{HH}_{h=1} ff(\mathrm{y}_{iih} \mid llll, eeee_{ii}) \quad (1)$$

Here $ff(\mathrm{y}_{ii} \mid \mathrm{e}_{ii})$ is taken as the probability density function estimated from $ff(\mathrm{y}_{ii})$values conditional on (ee_{ii}) values. the latent variable aids between the $\mathrm{e}_{ii}$and the y_{ii} variables whereas $PP(llll \mid eeee_{ii})$ is the probability of belonging to a certain latent class given an individual's realized covariate values. $ff(\mathrm{y}_{ii} \mid llll, eeee_{ii})$ is the probability density of y_{ii} conditional and ex_{ii}

(the mixture densities) which implies that latent variable can be influenced by exogenous variables, and response variables can possibly be affected by both exogenous and latent variables. The last $ff(y_{iih} \mid llll, eeee_{ii})$ part indicates that response variables of various segments are mutually independent given the latent and exogenous variables.

LCRM is a special case of Eq. 1 where a distinction between covariates and predictors can be made .In this mixture model variant different numbers of replications per case are allowed and the conditional densities

$ffrr_{iiii} \mid ll, e^{pp}_{iiii}$ are restricted to the same form for each time t. Given latent variable $llll$ and repeated observations of job satisfaction, single case total replication are denoted byr_{ii}, with TT_{ii} denoting as total replications.The unconditional model assumes the role of external variables is not taken into account and class formation is done solely based on predictors.

The restricted unconditional model involving only predictors becomes.

$$ffr_{ii} \mid e^{pp}_{ii1}, e^{pp}_{ii2} = \sum^{LL}_{ll=1} PP(ll) ffrr_{ii} \mid ll, {ee_{ii1}}^{pp}, {ee_{ii2}}^{pp} \quad (2)$$

Conditional Latent Class Regression Models

An important extension of conditional models incorporating covariates in basic model are described (Hagenaars & McCutcheon, 2002; Shockey, 1988).In this case the latent class Regression model for repeated observations can be obtained by making class membership dependent on covariates (Kamakura & Agrawal, 1994). The analysis conducted this way is the standard way to incorporate role of covariates in class formation also known as step 1 analysis. In such conditional model, it is assumed that the probability of belonging to latent class x depends on the values of ${ee_{ii1}}^{cc}$, ${ee_{ii2}}^{cc}$. For P predictors e^{pp}_{ii} affecting r_{ii}, and using R numeric or nominal covariates e^{cc}_{ii1}affecting $llll$.

The most general probability structure takes on the following form:

$$ffr_{ii} \mid e^{cc}_{ii1}, e^{pp}_{ii} = \sum^{LL}_{ll=1} PP(l \mid e^{c}_{i}) \prod_{ii=1}{}^{TT_{ii}} ffrr_{iiii} \mid ll, e^{pp}_{iiii} \quad (3)$$

Or

LL TT_{ii}

$$f(r_i \mid e_{i1c}, e_{i2c}, e_{pi1}, e_{pi2}) = \sum_{l=1} P(l \mid e_{i1c}, e_{i2c}) \prod_{i=1} f(r_{ii} \mid l, e_{ii1p}, e_{ii2p}) \quad (4)$$

Such a LC Regression model for repeated measures is very similar to multilevel (two-level), mixed, or random-coefficients models, in which random effects are included to deal with the dependent observations problem (Agresti et al., 2000; Muthén & Asparouhov, 2002).

Analysis

Step 3 analysis is the 3-step strategy for incorporating role of covariates or distal /dependent outcomes in mixture models. The basic procedure known as step 1 approach implies inclusion of covariates as active or inactive exogenous in same step of class formation and follows the same steps as done in unconditional models discussed above, for that reason we have opted for more sophisticated 3 step approach for doing conditional analysis in this study. The step 3 approaches are based on 3 steps and explained in further. The core structure of doing three steps is adopted from (Vermunt et al, 2013).

1. In step 1, a latent class regression model is built for the set of predictors and dependent variable. Here we decide regarding the significant predictors for the number of distinctive classes and other model features.
2. Using the final model from step 1 subjects are assigned to latent classes based on their posterior class membership probabilities and the class assignments are appended to basic data file. Class assignment can be modal (to the class for which the posterior membership probability is largest) or proportional (to each class with a weight equal to the posterior membership probability for that class).
3. Using the assigned class memberships from previous step 2, the association between the class membership and exogenous variables is examined with multinomial logistic regression analysis. Or simple cross-tabulations. The external variables can be (distal) outcomes influenced by class membership or both predictors of class membership. In case of applying proportional assignment in step 3 analysis, adjusted step-three maximum likelihood-based analysis requires expanding the data set to hold M records per entity having weights equal to the

posterior membership probabilities. To incorporate these weights more efficiently BCH adjustment based robust standard errors are used. In the BCH adjustment instead of estimating a latent class model one may perform the logistic regression analysis or can compute the crosstabulations in the standard way with the modification of an expanded data file with M records per entity (Vermunt, 2010; Vermunt & Magidson, 2021). ML adjustment is the preferred approach when the external variables are covariates or categorical dependent variables. Also, the suggested adjustments utilized were of BCH or ML based with modal assignment and proportional assignment of data values in step 1. The paper tested for basic step 3 with adjustments discussed above with model and proportional assignments.

Results

In order to meet following study objectives relevant to regression framework we will build and test empirical economic models with repeated measured labour market data:

1. To compare and evaluate different models for finding differential effects of job-related features impacting job quality.
2. To verify the differing impact of auxiliary variables across different classes by testing for the possible sources of heterogeneities in given data.

Class 2 solution was identified compared to 3-class solution by bootstrapping and other relative fit criteria(AIC3,BIC) suggested for particularly regression mixtures (Wedel & Kamakura, 2000).

Table1

Summary of Unconditional Regression Models

LCRM CASES	LLHD	BIC(LLH)	AIC(LLH)	AIC3(LLH)	npr	df	C.err.	Entrp R^2
Basic2clu	-91172.6	182553.6	182391.1	182414.1	23	8630	0.0904	0.698
basic 3clu	-89185.1	178687.6	178440.3	178475.3	35	8618	0.1393	0.6828
2clurestrictd	-91177.2	182544.7	182396.4	182417.4	21	8632	0.0904	0.698
3clurestrictd	-89189.5	178669	178442.9	178474.9	32	8621	0.1393	0.6828

Initial diagnostic of data reflects that responses for satisfaction score were skewed, and more responses were concentrated towards majorly last four categories on ordinal scale (variable details given in Appendix D).

From the Table 1 we can see that basic and restricted version of 2 and 3 class are presented. First 2 models show base line 2 classes and 3 classes cases. The selected model should improve in terms of higher log likelihood and lower information loss criteria. Further classes were not tested since beyond 3 since class separation indicator (classification error) were performing low. Though the case for 2 classes (Basic 2 cluster model) was most parsimonious with 23 parameters and 0.1 % classification error and relative highest score 69 % for entropy R^2 but the reason for choosing it as baseline model was its higher theoretical interpretability compared to additional class case (Basic 3 cluster model).

We inspected the parameters effects across both classes followed by profiles divisions across job satisfaction categories. Since the dependent indicator had 7 categories where first 5 categories implied somehow being unsatisfied or neutral with job and the last three categories clearly indicated for being satisfied and fully satisfied. Around 40% of subjects belonged to varying levels of being unsatisfied led by moderate satisfiers and most job satisfiers with work. The high level satisfiers were scattered over remaining 60% division of data. Initially with no covariates class 2 models was intuitively more appealing since the pattern of being satisfied or unsatisfied were broadly divided clearly for this model. The results were further endorsed by lowest classification errors for this model.

After choosing class 2 model (Basic 2 Cluster) we found few predictors insignificant (jbft_dv) so we dropped that from further analysis and estimated the restricted versions of 2 class solution and 3 class solutions also) with further restrictions driven by economic theory. Restricted version of 2 classes was chosen as a final model for classification and validation (details of classification in next section and validation table is provided in Appendix A).

From the table given below we can see that the parameters across the two classes are significantly different except for jbft_dv(full time job or part time job).Though Insignificant predictors are the general issue with the mixtures of regressions (Vermunt, 2001). In our case, for other each job satisfaction predictor the p-value is found to be less than .05 implying the null hypothesis stating effects associated with that predictor are zero would be rejected. Thus, for each predictor, information of the response for that predictor contributes significantly to differentiate between the job satisfaction classes.

Table 2

Parameters of Unconditional Regression Model

Classes	Indicators			Coeff	*SE*	*z*-val	*p*
Class (1)	1			0.02	0.03	0.695	0.49
Class(2)	1			-0.02	0.03	-0.695	0.49
jbsat(completely dissatisfied)		|	Class(1)	-1.2	0.26	-4.518	6.20E-06
jbsat(mostly dissatisfied)		|	Class(1)	-0.5132	0.1746	-2.9399	0.0033
jbsat(somewhat dissatisfied)		|	Class(1)	0.28	0.0902	3.184	0.0015
jbsat(neither satisfied or dissatisfied)		|	Class(1)	0.33	0.02	11.41	3.60E-30
jbsat(somewhat satisfied)		|	Class(1)	1.13	0.08	12.76	2.70E-37
jbsat(mostly satisfied)	1	|	Class(1)	0.9998	0.1839	5.4357	5.50E-08
jbsat(completely satisfied)	1	|	Class(1)	-1.0341	0.2633	-3.927	8.60E-05
Classes	Indicators			Coeff	*SE*	*z*-val	*p*
jbsat(completely dissatisfied)	1	|	Class(2)	-2.4876	0.2778	-8.9532	3.50E-19
jbsat(mostly dissatisfied)	1	|	Class(2)	-2.1701	0.1955	-11.098	1.30E-28
jbsat(somewhat dissatisfied)	1	|	Class(2)	-1.1408	0.1144	-9.9702	2.10E-23
jbsat(neither satisfied or dissatisfied)	1	|	Class(2)	-0.6041	0.0563	-10.7242	7.80E-27
jbsat(somewhat satisfied)	1	|	Class(2)	1.2	0.0994	12.0749	1.40E-33
jbsat(mostly satisfied)	1	|	Class(2)	2.8047	0.1821	15.402	1.60E-53

jbsat(completely satisfied)	1	|	Class(2)	2.3978	0.2782	8.618	6.80E-18
jbsat	jbsize	|	Class(1)	-0.0068	0.0037	-1.8222	0.068
jbsat	jbsize	|	Class(2)	-0.0226	0.0041	-5.5589	2.70E-08
jbsat	jbterm_dv	|	Class(1)	-0.0031	0.009	-0.3427	0.73
jbsat	jbterm_dv	|	Class(2)	-0.0474	0.0112	-4.225	2.40E-05
jbsat	jbhrs	|	Class(1)	0.0015	0.0014	1.0134	0.31
jbsat	jbhrs	|	Class(2)	-0.0042	0.0014	-2.9078	0.0037
jbsat	jbft_dv	|	Class(1)	0.0052	0.0305	0.1714	0.86
jbsat	jbft_dv	|	Class(2)	0.0095	0.0368	0.2587	0.8
jbsat	hiqual_dv	|	Class(1)	-0.0107	0.0052	-2.0664	0.039
jbsat	hiqual_dv	|	Class(2)	0.0302	0.0079	3.8096	0.00014

For the unrestricted model classification, we imposed certain order restrictions to compare more parsimonious model to baseline case. Since role of hours and pay is described in economic literature positive for boosting employees moral (Malik et al. 2012; Wanger, 2017), so we imposed the increasing restrictions on both predictors. That implied to test for the hypothesis; People with higher earnings and full-time work situations are more satisfied with their jobs job satisfaction. The restricted versions for 2 and 3 classes (see table 1 2clurestrictd and 3clurestrictd) were tested. Results for these models in terms of lower value of information criteria (BIC and AIC3) remained same compared to basic cases also no mark-able change in classification error and entropy R2 was observed. Since the restricted model did not improve the effect sizes (see given table 3) therefore for finalizing base model we employed parametric bootstrapping on basic 2cluster case. The results supported for the basic 2 class case fits the model well (see the bootstrapped *p* value is insignificant for the absolute fit statistic).

Table 3

Parametric Bootstrapping Result

2-Class Ordinal Regression Model					
Chi-squared Statistics			Bootstrap		
df	L^2	p-value	*p*-value	*SE*	CV

8630	181847.5	3.3e-31906	1	0.0000	176039.7
Number of cases			8653		
Number of replications			59855.83		
Number of parameters (Npar)			23		
Random Seed			30060		
Best Start Seed			30060		
Monte Carlo Seed			499531		

From the given table we can see that the intercept for each category of job satisfaction for both classes are significantly different. For negative feedback categories consisting of ‘completely dissatisfied’ and ‘mostly dissatisfied’ is negative for both classes and for the categories of somewhat dissatisfied ‘neither satisfied or satisfied’ and ‘somewhat satisfied’ it stands out to be positive for class 1 and negative for class 2 implying initially class 1 contains more respondents with low satisfaction levels compared to class 2 which consists of more respondents with highest satisfaction on job levels. For extreme positive categories of satisfaction with job for class 2 we have highly significant and high size of initial response for these categories compared to somewhat contrary response for extreme satisfaction level for class 1 individuals.

The beta parameter for each predictor is a measure of the influence of that predictor on jobs satisfaction. The beta effect estimates under the column labelled class 1 suggest that class 1 is less likely to be influenced by organization size (job size) working hours and qualification. Class 1 and class 2 both are not influenced by job size (beta is approximately 0). Job size appears to be insignificant predictor for job satisfaction in case of class 1 and significant for class 2). Why we have reported the predictors which were somehow not significant, and not very much impactful on the levels of job satisfaction for replying this we take a point of departure comparing to general tradition of regression results reporting in which only significant implies good results. Interestingly though the predictors of job satisfaction were significantly different across both groups. But the effect sizes had explanatory power negligible indicating the exercise done to be futile at first glance.

Table 4

Regression Parameters of Restricted 2 Class Model

Term				Coef	*SE*	*z*-value	*p*-value
Class(1)	1			0.0264	0.039	0.6816	0.5
Class(2)	1			-0.0264	0.039	-0.6816	0.5
jbsat(completely dissatisfied)	1	|	Class(1)	-1.2366	0.163	-7.6073	2.80E-14
jbsat(mostly dissatisfied)	1	|	Class(1)	-0.5351	0.107	-4.9944	5.90E-07
jbsat(somewhat dissatisfied)	1	|	Class(1)	0.2767	0.057	4.8699	1.10E-06
jbsat(neither satisfied or dissatisfied)	1	|	Class(1)	0.3319	0.029	11.4131	3.60E-30
jbsat(somewhat satisfied)	1	|	Class(1)	1.1432	0.054	21.3722	2.40E101
jbsat(mostly satisfied)	1	|	Class(1)	1.0218	0.116	8.8245	1.10E-18
jbsat(completely satisfied)	1	|	Class(1)	-1.0017	0.163	-6.1411	8.20E-10
jbsat(completely dissatisfied)	1	|	Class(2)	-2.5438	0.178	-14.2502	4.50E-46
jbsat(mostly dissatisfied)	1	|	Class(2)	-2.2081	0.123	-17.9425	5.50E-72
jbsat(somewhat dissatisfied)	1	|	Class(2)	-1.1595	0.083	-14.0439	8.40E-45
jbsat(neither satisfied or dissatisfied)	1	|	Class(2)	-0.6029	0.056	-10.7358	6.90E-27
jbsat(somewhat satisfied)	1	|	Class(2)	1.2185	0.071	17.1488	6.40E-66
jbsat(mostly satisfied)	1	|	Class(2)	2.8419	0.104	27.214	4.40E163

Term				Coef	*SE*	*z*-value	*p*-value
jbsat(completely satisfied)	1	|	Class(2)	2.4539	0.159	15.4375	9.20E-54
jbsat	jbsize	|	Class(1)	-0.0068	0.004	-1.8309	0.067
jbsat	jbsize	|	Class(2)	-0.0227	0.004	-5.5816	2.40E-08
jbsat	jbterm _dv	|	Class(1)	-0.003	0.009	-0.3368	0.74
jbsat	jbterm _dv	|	Class(2)	-0.0471	0.011	-4.222	2.40E-05
jbsat	jbhrs	|	Class(1)	0.0013	0.001	1.242	0.21
jbsat	jbhrs	|	Class(2)	-0.0044	0.001	-4.17	3.00E-05
jbsat	hiqual _dv	|	Class(1)	-0.0106	0.005	-2.056	0.04
jbsat	hiqual _dv	|	Class(2)	0.0303	0.008	3.843	0.00012

Since the objective was exploratory where things could turn as expected or contrary. The general hypothesis of the differential impact of chosen job features was negated in this case study implying the homogenous impact of chosen features exists across both groups of satisfied doers and nonsatisfied doers. The somehow similar effects from both unrestricted and restricted unconditional models urged us to look further for the possible source of difference for both classes. Since background variables or covariates come to play their role for finding the source of latent class membership in mixture model therefore we estimated conditional models by adding subjective/ background variables in basic unconditional model

Before getting into conditional models in the following we briefly discuss unconditional profiles for comparing to conditional profiles in next section.

Classification of Unconditional Model

The given classification/Profile output contains information on the class sizes, the class-specific marginal probabilities and means of the job satisfaction variable. It is clear from the first row that class 1 contains about 50% of the subjects (.5135), segment 2 contains about .4865%. Examination of class-specific probabilities shows that overall, segment 1 is least likely be completely satisfied with their work only 0.03% are completely satisfied compared to segment 2 who are most likely 28.% completely satisfied followed by 52 % likely of mostly satisfied level to 13% for somewhat satisfied level. The first four lowest levels of dissatisfaction are least reported in this class around 5 % in total compared to class 1 which has around 29 % likely cases reporting first four low score on satisfaction scale followed by highest likely cases of somewhat satisfied. Later in conditional models we will show how to classify each case into the most appropriate segment.

Table 5
Classification Probabilities of LCRM

classes	Non-satisfied doers	Satisfied doers	Total size
Size of class	0.5132	0.4868	
Satisfaction with job			
Completely dissatisfied	0.0323	0.0088	0.0209
Mostly dissatisfied	0.0635	0.0093	0.0371
Somewhat dissatisfied	0.1392	0.0202	0.0812
Neither satisfied or dissatisfied	0.1433	0.0271	0.0868
Somewhat satisfied	0.3146	0.1307	0.2251
Mostly satisfied	0.272	0.5219	0.3936
Completely satisfied	0.0351	0.2821	0.1553

Conditional Models

The objective of this particular study was to look for possible differences in two segments of given classes at various levels of occupational choices, age, gender and at various levels of overall quality of life. For meeting this objective, we separately examined 2 sets of models under step 3 analysis.

The variants in regression context were covariate proportional maximum likelihood based for individual and model case assignment followed by BCH corrections. In case of BCH corrections, the data does not follow chi2 distribution therefore criteria of L^2 does not provides fit statistics in the given table 5. In the given case (Maximum likelihood (ML) based both corrections proved significantly fit. Considering lowest value of relative information tools, we opted for covariate proportional maximum likelihood based (CPML) case for 2 classes. We conducted step 3 analysis for 2 sets of covariates; first included role of occupational segregation for making satisfaction level choices. The same models also measured gendered differences across classes. The covariates included in second step 3 analysis were age and satisfaction with life to check the hypothesis of overall quality of life as a covariate for satisfactions with jobs (see in Appendix specification Table B).

Table6

Step 3 Regression Specification A

VARIANT	LLH	BIC(LLH)	AIC(LLH)	AIC3(LLH)	*L²*	*df*	*p*	C.Err.
CPML 2Class	-40806	81755.42	81638.54	81651.54				0.4542
CPBCH 2-Class	-40722	81587.55	81470.68	81483.68	11.2076	11	0.43	0.4542
CMML 2Class	-40850	81843.36	81726.48	81739.48				0.4541
CMBCH 2-Class	-40461	81065.05	80948.17	80961.17	14.3696	11	.21	0.4541

From the given Table it is observed that Gender differences in this case were present in making satisfaction choices. Females had relatively more chance to belong to class 2 (satisfiers) compared to males (61 % to 39%). This difference of proportions might be because of sample size difference of both genders (58 vs. 42%). Age categories in sample are scattered from youth prime ages to very old people (16 years to 88 years). Total proportion is evenly divided in age groups and they are somehow nearly distributed across classes. The interesting finding is similar pattern of life satisfaction reporting over the years as reporting job satisfaction. The respondents for highest life satisfaction are more like to belong to job satisfiers class compared to those who are neutral and dissatisfied with life have more chances to fall in class 1. The proportion of mostly satisfied and somewhat satisfied with life is higher compared to basic model of job satisfaction presented earlier but likely patterns are somehow similar.

Table 7
Conditional Classification for Case A

Occupations	Gender	Non-satisfied doers	Satisfied doers
Self employed	Male	0.4342	0.5658
Self employed	Female	0.3861	0.6139
Occupations	Gender	Non-satisfied doers	Satisfied doers
Paid employment(ft/pt)	Male	0.5718	0.4282
Paid employment(ft/pt)	Female	0.5225	0.4775
Unemployed	Male	0.6524	0.3476
Unemployed	Female	0.606	0.394
Retired	Male	0.2791	0.7209
Retired	Female	0.2409	0.7591
Family care or home	Male	0.539	0.461
Family care or home	Female	0.4893	0.5107
full-time student	Male	0.7388	0.2612
full-time student	Female	0.6986	0.3014
LT sick or disabled	Male	0.7234	0.2766
LT sick or disabled	Female	0.6819	0.3181
Govt training scheme	Male	0.9538	0.0462
Govt training scheme	Female	0.9442	0.0558
Unpaid, family business	Male	0	1
Unpaid, family business	Female	0	1
On apprenticeship	Male	0.444	0.556
On apprenticeship	Female	0.3955	0.6045

Doing something else	Male	0.5054	0.4946
Doing something else	Female	0.4558	0.5442

In the following from the conditional effects of occupation categories and gender we can cross examine the likely distribution. Since the identification constraints in effect coding impose certain restriction on the sum of parameters for categorical variables therefore, we have sum of parameters equal to zero in this case. We can see that people doing family business are most likely to belong satisfied class followed by retired and self -employed and on apprenticeship. Quite naturally the categories including of those individuals who are not working actively are more likely to belong to unsatisfied group. The class effects could be temporary for these individuals when they get back to work if want to and definitely needs further investigation in future.

Table 8

Conditional Parameters for Case 1

Covariates	Unsatisfied	Satisfied	Wald	*p*-value
Intercept	0.0506	-0.0506	0.1681	0.68
Job satisfaction				
Self employed	-0.2155	0.2155	163.5145	2.50E-29
Paid employment(ft/pt)	0.0446	-0.0446		
Unemployed	0.3046	-0.3046		
Retired	-0.5807	0.5807		
Family care or home	-0.0371	0.0371		
Full-time student	0.392	-0.392		
LT sick or disabled	0.3996	-0.3996		
Govt. training scheme	1.1328	-1.1328		
Unpaid, family business	-1.341	1.341		
On apprenticeship	-0.1952	0.1952		
Doing something else	-0.027	0.027		

Gender				
Male	0.0552	-0.0552	130.0866	3.90E-30
Female	-0.0552	0.0552		

Conclusion

In this article we have conducted conditional and unconditional analysis for the British longitudinal data featuring job satisfaction level conditional to some job related intrinsic and extrinsic features. The diversified sample is chosen to explore the possible presence of heterogeneous sub populations within the larger group of individuals. We were interested to find the differences in effect sizes of some important job satisfaction indicators. The data becomes sparse when latent framework is applied to tabulate cross relations of included 5 to 8 indicators for 8000 plus individuals therefore absolute fit diagnostics become invalid in this case.

Considering such asymptotic limitation of data, the solution was decided by absolute, relative and bootstrapping model selection techniques simultaneously. Though we could find prevalence of job satisfiers and nonsatisfiers in British household's sample. Interestingly, the predictors of job satisfaction were significantly different across both groups but the effect sizes had explanatory power negligible indicating the exercise done to be futile at first glance. It was expected since the objective was exploratory where things could turn as expected or contrary. The exercise did not support the presence of heterogonous segments with response to chosen indicators. Also, the general hypothesis of the differential impact of chosen job features was negated in this case study implying the homogenous impact of chosen job features existed across both classes of so-called satisfied doers and non-satisfied doers. The results urged us to look further for the source of difference could be addressed by subjective background variables since background variables/ covariates come to play their role for finding the source of latent class membership. Therefore, we did conditional analysis with step 3 approaches. Through variants of Step 3 models in regression case we found occupational and gendered differences in both classes. For the particular case, occupation and subjective scores on quality of life turned to be the best indicative of current standing of individuals on job satisfaction ladder.

References

Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, *30*(1), 27–80. https://doi.org/10.1111/0081-1750.t01-1-00075

Andrews, R. L., & Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, *20*(4), 315–321. https://doi.org/10.1016/j.ijresmar.2003.04.001

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using biasadjusted three-step approaches. *Sociological methodology*, *43*(1), 272–311. https://doi.org/10.1177/0081175012470644

Bierbrauer, M., Trück, S., & Weron, R. (2004, June 6–9). *Modeling electricity prices with regime switching models* (Paper presentation). Proceedings of 4th International Conference, Kraków, Poland.

Clark, A. E. (2005). Your money or your life: Changing job quality in OECD countries. *British Journal of Industrial Relations*, *43*(3), 377–400. https://doi.org/10.1111/j.1467-8543.2005.00361.x

Gudicha, D. W., & Vermunt, J. K. (2013). Mixture model clustering with covariates using adjusted three-step approaches. In *Algorithms from and for nature and life* (pp. 87-94). Springer.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*: Cambridge University Press.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*(2), 81–102. https://doi.org/10.1177/1471082X0100100201

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing*, *11*(5), 451–464. https://doi.org/10.1016/01678116(94)00004-2

Khalili, A., & Chen, J. (2007). Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, *102*(479), 1025–1038.

Kim, M., Vermunt, J., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(4), 601–614. https://doi.org/10.1080/10705511.2016.1158655

Malik, M. E., Danish, R. Q., & Munir, Y. J. A. J. o. e. (2012). The impact of pay and promotion on job satisfaction: Evidence from higher education institutes of Pakistan. *American Journal of Economics*, *2*(Special Issue), 6–9. https://doi.org/10.5923/j.economics.20120001.02

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and its Application*, *6*, 355–378. https://doi.org/10.1146/annurev-statistics-031017-100325

Muthén, B., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus* (Mplus Web Notes No. 4). https://www.statmodel.com/download/webnotes/CatMGLong.pdf

Qu, Y., Tan, M., & Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52(3), 797–810. https://doi.org/10.2307/2533043

Sánchez, N. S., & Puente, A. C. F. (2015). Educational and skill mismatches: Differential effects on job satisfaction. A study applied to the Spanish job market. *Estudios de Economía*, *41*(2), 261–281.

Shockey, J. W. (1988). Latent class analysis: An introduction to discrete data models with unobserved variables. In *Common problems/proper solutions: Avoiding error in quantitative research* (pp. 288–315). Sage.

Simonoff, J. S. (2003). *Analyzing categorical data* (Vol. 496). Springer.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*: Chapman and Hall.

Tuma, M., & Decker, R. (2013). Finite mixture models in market segmentation: a review and suggestions for best practices. *Journal of Business Research Methods*, *11*(1), 215–215.

Understanding Society. (2021). Waves 1-13, 2008-2020. Institute for Social and Economic Research. https://www.understandingsociety.ac.uk/about/about-the-study

Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed discrete outcome variables. *Multivariate Behavioral Research*, *40*(3), 281–301. https://doi.org/10.1207/s15327906mbr4003_1

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2013). Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax. *Statistical Innovations Inc*. https://www.statisticalinnovations.com/wpcontent/uploads/LGtechnical.pdf

Vermunt, J. K., & Magidson, J. (2021). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 356–364. https://doi.org/10.1080/10705511.2020.1818084

Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric randomcoefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, *13*(2), 6–13.

Wanger, S. (2017). *What makes employees satisfied with their working time? The role of working hours, time-sovereignty and working conditions for working time and job satisfaction* (IAB-Discussion Paper No. 20/2017). http://hdl.handle.net/10419/172884

Wedel, M., & Kamakura, W. A. (2000). Mixture regression models. In *Market segmentation* (pp. 101–124). Springer.

Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology*, *105*(6), 1702–1740.

Appendix

In this appendix we provide some additional results whether we should discard the findings or should take these as a compromising solution for the given complex units for this we tested the out of sample performance of the chosen model by cross validation and latent class classification. The diagonal entries in the given table indicate exact classification and off diagonal entries show miscalculations. Further various types of error rates are reported based on absolute, marginal and log likelihood differences of baseline and calculated model. The divergence in rates is not much and k cross validation was also tested to check the sample performance for prediction purpose. Results in table A also reveal 10-fold validation. The model sustained the level of good classification and prediction power (classification error rate (0.01) and entropy R2(around 70%).

Table A

Validation and Latent Classification

Latent Classification	Modal		
1	4361.1458		4834.463
2	370.1725	3448.365	3818.537
Total	4731.3183	3921.682	8653
Prediction Statistics			
Job-sat			
Error Type	Baseline	Model	R^2
Sq. Error	2.0315	1.4567	0.2829
LL	1.6136	1.4358	0.1102
Ab. Error	1.1243	0.9051	0.195
Pr. Error	0.6082	0.6059	0.0038
Classification Statistics		Class	
C.E		0.0975	
(Lambda)		0.7792	
Entropy R-sq		0.681	
Standard R-sq		0.7215	
		2	
Latent	1	473.3169	Total

Step 3 Regression Specification Case 2

In second specifications of step 3 analysis the modelling variants were same, but the covariates were chosen different. Here CPBCH (covariate proportional) turned to be best option based on relative fit criteria reporting lowest loss of information. We present this model 's parameters for further discussion.

Table B

Step 3 Regression Specification Case 2

	LL	BIC(LL)	AIC(LL)	AIC3(LL)	L^2	*df*	*p*	Class.Err.
CPML 2-Class	-36896.7	73935.76	73819.46	73832.46	1950.948	2171	1	0.3508
CPBCH 2-Class	-35204.7	70551.64	70435.34	70448.34				0.3507
CMML 2-Class	-35533.2	71208.72	71092.43	71105.43				0.3566
CMBCH 2-Class	-36722.4	73587.09	73470.8	73483.8	2601.163	2171	3.90E10	0.3565

The given table reports the impact of various levels of satisfaction with life categorical impact on both classes of satisfier doers and non-satisfier doers. We can see that for satisfiers the predicators are negligible to explain any differences whereas for non satisfiers these are for effective. Based on the 'Parameters' output we see that compared most satisfied last two categories the non-satisfied cases are less likely to be in class 2 than satisfied cases, this negligible effect to all categories of no or more satisfaction with life is present for class 1 . Age is though significant to shape class formation but negligible followed by education role which is not explaining the likely change in categorical scores of job satisfaction. For class 2 the pattern of change is more effected by more or most satisfied categories with life. We concluded here that response on choice for grading yours quality of life is more effective for class formation of second group.

Table C

Conditional Parameters for Case 2

Model for Classes	Cluster1	Cluster2	Wald	p-value

Intercept scaleofsatisfaction	0	-0.8056	61.325	4.80E-15

Covariates	Cluster1	Cluster2	Wald	p-value
completely dissatisfied	0.00	0	3907.84	5.1e-843
mostly dissatisfied	0.00	-0.5281		
somewhat dissatisfied	0.00	-1.621		
Neither Sat nor Dissat	0.00	-1.2234		
somewhat satisfied	0.00	-0.6632		
mostly satisfied	0.00	0.2658		
completely satisfied	0.00	1.3417		
age_	0.00	0.0191	404.4106	6.00E-90
qualification				
Degree	0.00	0	186.6861	2.00E-38
Other higher	0.00	0.0446		
A level etc	0.00	-0.1329		
GCSE etc	0.00	-0.3566		
Other qual	0.00	-0.2209		
No qual	0.00	-0.3452		

Table D

Data Variables Information /Survey

PSU ID	psu	3081
Stratum ID	strata	1599
Case ID	pidp	8653
Dependent		
jbsatis	Ord-Fix	7
cdissatis	1	1
mdissatis	2	2
somedissatis	3	3
neither sat or dissat	4	4
somewhat satisfied	5	5
mostly satisfied	6	6

completely satisfied	7	7
Independent		
jbsize	Num-Fix	11
1 - 2	1	1
3 - 9	2	2
10 - 24	3	3
25 - 49	4	4
50 - 99	5	5
100 - 199	6	6
200 - 499	7	7
500 - 999	8	8
1000 plus	9	9
fewer than 25	10	10
25 or more	11	11
jbterm_dv	Num-Fixed	6
permanentjob	1	1
seaswork	2	2
contractfixedt	3	3
agencyhiring	4	4
casual work	5	5
not permanent	6	6
jbhrs	Num-Fix	256
0	0	0
0.1	0.1	0.1
0.2	0.2	0.2
0.5	0.5	0.5
1	1	1
1.2	1.2	1.2
1.5	1.5	1.5
2	2	2
2.5	2.5	2.5
3	3	3
...		

88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
95	95	95
96	96	96
97	97	97
97.9	97.9	97.9

hiqual_dv	Num-Fixed	6
Degree	1	1
Other higher	2	2
A level etc	3	3
GCSE etc	4	4
Other qual	5	5
No qual	9	9