Journal of Statistical Research *ISSN 0256 - 422 X*
2011, Vol. 45, No. 1, pp. 37-47
Bangladesh

JOINT MODELLING OF MIXED CORRELATED NOMINAL, ORDINAL AND CONTINUOUS RESPONSES

E. BAHRAMI SAMANI
Department of Statistics, Shahid Beheshti University, Tehran, Iran
Email: ehsan_bahrami_samani@yahoo.com

ZH. TAHMASEBINEJAD
Department of Statistics, Elm o Sanat University, Tehran, Iran
Email: jtahmasebinejad@yahoo.com

SUMMARY

This paper develop a new class of joint modelling of mixed correlated nominal, ordinal and continuous responses with elliptically contoured errors is presented. A full likelihood-based approach is used to obtain maximum likelihood estimates of the model parameters. To illustrate the application of such modelling the data corresponding to the British Household Panel Survey (BHPS) is analyzed.

Keywords and phrases: Joint Modelling; Latent Variable; Life Satisfaction; Maximum Likelihood; Elliptically Contoured Distribution.

1 Introduction

We describe a model for the joint distribution of outcomes related to statistical problems arising from the analysis of mixed correlated nominal, ordinal and continuous data, a common enough occurrence in practice. The data used in this paper is extracted from wave 15 of the British Household Panel Survey (BHPS), a survey of adult Britons, being carried out annually since 1991 by the ESRC UK Studies Center with the Institute for Social and Economical Research at the University of Essex. In these data, the current economic activity (nominal response), life satisfaction status (ordinal response) and Income (continuous response) are the mixed correlated responses and the effect of explanatory variables on these responses should be investigated simultaneously. Consequently, we need to consider a method in which these variables can be modelled jointly.

For joint modelling of responses, one method is to use the general location model of Olkin and Tate (1961), where the joint distribution of the continuous and categorical variables is decomposed into a marginal multinomial distribution for the categorical variables and a conditional multivariate normal distribution for the continuous variables, given the categorical variables for a mixed Poisson and continuous response where Olkin and Tate's (1961) method is used see Yang et al. (2007). A second method for joint modelling is to decompose the joint distribution as a multivariate marginal distribution for the continuous responses and a conditional distribution for categorical variables given the continuous variables. Cox and Wermuth (1992) empirically examined the choice between these two methods. The third method uses simultaneous modelling of categorical and continuous variables to take into account the association between the responses by the correlation between errors in the model for responses. For more details of this approach see, for example, Heckman (1978) in which a general model for simultaneously analyzing two mixed correlated responses is introduced and Catalano and Ryan (1992) who extended and used the model for a cluster of discrete and continuous outcomes (see also, Fitzmaurice and Laird, 1997). All the above references consider correlated nominal and continuous responses. A model for ordinal and continuous responses without considering any covariate effect is also presented by Poon and Lee (1989). Bahrami Samani et al. (2008) who extended and used a model for mixed ordinal and continuous responses with considering any covariate effect is presented. In this model ordinal response can be dependent on the continuous response. With this model, the dependence between responses can be taken into account by the correlation between errors in the models for continuous and ordinal responses. We also take into account the ordinality information available in categorical response by a cumulative latent variable model for ordinal response.

The aim of this paper is to use and extend an approach similar to that of Bahrami Samani et al. (2008), which jointly models a ordinal and a continuous variable, for joint modelling of multivariate nominal, ordinal and continuous outcomes. In this model nominal and ordinal responses can be dependent on the continuous responses. With this model, the dependence between responses can be taken into account by the correlation between errors in the models for continuous, ordinal and nominal responses.

Section 2 develops the general mixed correlated-data model. In this section outlines maximum likelihood estimation for the model. In Section 3, simulation study. the next section, the proposed methodology is applied on the BHPS data. Finally, concluding remarks are given.

2 General mixed correlated-data model

Let S_{ij} for $j = 1, \ldots, M_1$, be nominal response with K_j levels, let $S^*_{ij,r}$ for $j = 1, \ldots, M_1$ be the latent variable associated with category r and assume a latent Gaussian regression model

$$S^*_{ij,r} = \beta'_{j,r} X_i + \varepsilon_{ij,r}, \;\; r = 1, \ldots, K_j - 1,$$

for the reference category K_j, no linear predictor is assumed, $S^*_{ij,K_j} = \varepsilon_{ij,K_j}$, to ensure identifiability. The principle of random utility postulates

$$S_{ij} = r \;\; \Leftrightarrow \;\; S^*_{ij,r} = \max_{l \in \{1,\ldots,K_j\}} S^*_{ij,l}.$$

For ordinal response Y_{ij} for $j = M_1+1, \ldots, M_2$, with c_j levels. Let Y^*_{ij} for $j = M_1+1, \ldots, M_2$ denote the underlying latent variable for the jth ordinal response of the ith individual, which leads to have Y_{ij} as:

$$Y_{ij} = \begin{cases} 1 & Y^*_{ij} \leq \theta_{j1}, \\ l+1 & \theta_{jl} < Y^*_{ij} \leq \theta_{j,\,l+1},\; l = 1, \ldots, c_j - 2, \\ c_j & Y^*_{ij} > \theta_{j,c_j-1}, \end{cases}$$

where $\theta_j = (\theta_{j1}, \ldots, \theta_{j,c_j-1})$ is the vector of cut-point parameters for the jth ordinal response. $Var(Y^*_{ij}) = 1$ for $j = M_1 + 1, \ldots, M_2$ to ensure identifiability.

Z_{ij} for $j = M_2 + 1, \ldots, M$, represent continuous responses. All of these responses may be correlated. The joint model takes the form:

$$\left.\begin{array}{lll} S^*_{ij,r} & = \beta'_{j,r} X_i + \varepsilon^{(1)}_{ij,r} & j = 1, \ldots, M_1, r = 1, \ldots, K_j - 1 \\ Y^*_{ij} & = \beta'_j X_i + \varepsilon^{(2)}_{ij} & j = M_1 + 1, \ldots, M_2 \\ Z_{ij} & = \beta'_j X_i + \varepsilon^{(3)}_{ij} & j = M_2 + 1, \ldots, M \end{array}\right\} \tag{2.1}$$

Let

$$\varepsilon_i = (\varepsilon^{(1)}_i, \varepsilon^{(2)}_i, \varepsilon^{(3)}_i)' \overset{iid}{\sim} EC_M(0, \Sigma; g)$$

where $\varepsilon^{(1)}_i = (\varepsilon^{(1)}_{i,1}, \ldots, \varepsilon^{(1)}_{i,K_{M_1}-1})'$, $\varepsilon^{(1)}_{i,r} = (\varepsilon^{(1)}_{i1,r}, \ldots, \varepsilon^{(1)}_{iM_1,r})'$, $\varepsilon^{(2)}_i = (\varepsilon^{(2)}_{i(M_1+1)}, \ldots, \varepsilon^{(2)}_{iM_2})'$, $\varepsilon^{(3)}_i = (\varepsilon^{(3)}_{i(M_2+1)}, \ldots, \varepsilon^{(3)}_{iM})'$ and $\theta_j = (\theta_{1,j}, \ldots, \theta_{c_j-1,j})'$, for $j = 1, \ldots, M_1$, is the vector of cut-point parameters for the jth ordinal response and X_i is the vector of explanatory variables for the i^{th} individual and Σ is the $M \times M$ covariance matrix which for illustration, has the following structure,

$$\begin{aligned} var(\epsilon^{(3)}_{ij}) &= \sigma^2, \quad j = M_2 + 1, \ldots, M, \\ cov(\epsilon^{(1)}_{ij,r}, \epsilon^{(1)}_{ij',r}) &= \rho^{(1)}_{jj',r}, \qquad j, j' = 1, \ldots, M_1,\;\; r = 1, \ldots, K_j - 1,\;\; j \neq j', \\ cov(\epsilon^{(1)}_{ij,r}, \epsilon^{(2)}_{ij'}) &= \rho^{(2)}_{jj',r}, \quad j = 1, \ldots, M_1,\;\; r = 1, \ldots, K_j - 1,\;\; j' = M_1 + 1, \ldots, M_2, \\ cov(\epsilon^{(1)}_{ij,r}, \epsilon^{(3)}_{ij'}) &= \sigma\rho^{(3)}_{jj',r}, \quad j = 1, \ldots, M_1,\;\; r = 1, \ldots, K_j - 1,\;\; j' = M_2 + 1, \ldots, M, \\ cov(\epsilon^{(2)}_{ij}, \epsilon^{(2)}_{ij'}) &= \rho^{(4)}_{jj'}, \qquad j, j' = M_1 + 1, \ldots, M_2,\;\; j \neq j' \\ cov(\epsilon^{(3)}_{ij}, \epsilon^{(3)}_{ij'}) &= \sigma^2\rho^{(5)}_{jj'}, \qquad j, j' = M_2 + 1, \ldots, M,\;\; j \neq j' \\ cov(\epsilon^{(2)}_{ij}, \epsilon^{(3)}_{ij'}) &= \sigma\rho^{(6)}_{jj'}, \quad j = M_1 + 1, \ldots, M_2,\;\; j' = M_2 + 1, \ldots, M. \end{aligned}$$

Because of identifiability problem we have to assume:

$$\begin{aligned} var(\epsilon^{(1)}_{ij,r}) &= 1, \quad j = 1, \ldots, M_1,\;\; r = 1, \ldots, k_j, \\ var(\epsilon^{(2)}_{ij}) &= 1, \quad j = M_1 + 1, \ldots, M_2. \end{aligned}$$

The vector of coefficients $\beta = (\beta_{1,1}, \ldots, \beta_{M_1,K_{M_1}-1}, \beta_{M_1+1}, \ldots, \beta_M)'$, cut-points parameters θ_j for $j = 1, \ldots, M_1$ and Σ should be estimated. The parameter vector, β_j for $j = M_2 + 1, \ldots, M$, includes an intercept parameter but β_j, for $j = M_1 + 1, \ldots, M_2$ and $\beta_{j,r}$ for $j = 1, \ldots, M_1, \quad r =$

$1,\ldots,K_j-1$, due to having cut-points parameters, are assumed not to include any intercept. In this model any multivariate distribution can be assumed for the errors in the model. Here, a multivariate normal distribution is used. The likelihood for this model is given:

Let $y = (y'_1,\ldots,y'_n)'$, $s = (s'_1,\ldots,s'_n)'$, $z = (z'_1,\ldots,z'_n)'$ and $x = (x'_1,\ldots,x'_n)'$ where $s_i = (s_{i1,1},\ldots,s_{iM_1,K_{M_1}-1})'$, $y_i = (y_{i(M_1+1)},\ldots,y_{iM_2})'$ $z_i = (z_{i(M_2+1)},\ldots,z_{iM})'$ and $x_i = (x_{i1},\ldots,x_{iv})'$, and v is the number of explanatory variables for the ith individual.

The joint likelihood function for the parameters and latent variables is:

$$
\begin{aligned}
P(\eta,\Sigma\,|y,s,z\ ,x) &= \prod_{i=1}^{n} f(z_i,\ y_i,s_i|x_i,\eta,\Sigma)\\
&= \prod_{i=1}^{n} P(S_{i1}=r_{i1},\ldots,S_{iM_1}=s_{iM_1},\\
&\qquad Y_{i,M1+1}=y_{i,M1+1},\ldots,Y_{iM_2}=y_{iM_2}\,|z_i\ ,x_i)\ f(z_i|x_i)\\
&= \prod_{i=1}^{n} P(S^*_{ij,r_{ij}} = \max_{l\in\{1,\ldots,K_j\}} S^*_{ij,l}, j=1,\ldots,M_1, \theta_{j,y_{ij}-1} < Y^*_{ij} \le \theta_{j,y_{ij}},\\
&\qquad j = M_1+1,\ldots,M_2\,|z_i,x_i\,)f(z_i|x_i).
\end{aligned}
$$

This likelihood can be maximized by function "`nlminb`" in software **R**.This function may be used for minimization of a function of parameters.

For maximization of a likelihood function one may minimize minus log likelihood function. The function "`nlminb`" uses optimization method of port routine which is given in " http://netlib.bell-labs.com/cm/cs/cstr/153.pdf.". The function "nlminb" uses a sequential quadratic programming (SQP) method to minimize the requested function. The details of this method can be find in Fletcher (2000). The observed Hessian matrix may be obtained by "nlminb" function or may be provided by function "fdHess".

3 Simulation study

We consider four continuous variables S^*_1, S^*_2, Y^*_3 and Z_4. The nominal variable S with three levels is defined as

$$S = r \ \Leftrightarrow\ S^*_r = \max_{l\in\{1,2,3\}} S^*_l, \quad r = 1,2.$$

The ordinal variable Y_2 with three levels is defined as

$$Y_2 = \begin{cases} 1 & Y^* < \theta_1,\\ 2 & \theta_1 \le Y^* < \theta_2,\\ 3 & Y^* \ge \theta_2.\end{cases}$$

The variables S^*_1, S^*_2, Y^*_3 and Z are generated by a multivariate normal distribution with zero mean and correlation matrix Σ whose all off-diagonal elements are 0.5. For this we consider 3 values for n (50, 100, and 1000). In this analysis we use 1000 sets of simulation.

3.1 Simulation study without covariates

In this subsection, we analyze the following simple model

$$S_r^* = \varepsilon_r, \ \ Y_3^* = \varepsilon_3, \ \ Z_4 = \mu_z + \varepsilon_4, \ r = 1, 2.$$

Table 1 contains the average estimated values of μ_z, σ_z^2, $\rho_{12,1}^{(2)}$ (the correlation between S_1^* and Y_3^*), $\rho_{12,2}^{(2)}$ (the correlation between S_2^* and Y_3^*), $\rho_{13,1}^{(3)}$ (the correlation between S_1^* and Z_4), $\rho_{13,2}^{(3)}$ (the correlation between S_2^* and Z_4) and $\rho_{23}^{(6)}$ (the correlation between Y_3^* and Z_4), θ_1, θ_2 for $n = 50$, $n = 100$, and $n = 1000$. The parameter estimates by the model (for $n = 50$, $n = 100$ and $n = 1000$) are close to the true values of the parameters. Of course, the more the value of n the better the estimates.

Table 1: Results of the simulation study

		$n = 50$		$n = 100$		$n = 1000$	
Parameter	True value	Est.	S.E.	Est.	S.E.	Est.	S.E.
μ_z	0.000	0.060	0.141	0.022	0.112	0.001	0.016
σ_z^2	1.000	1.141	0.124	1.025	0.064	1.002	0.023
$\rho_{12,1}^{(2)}$	0.500	0.450	0.136	0.495	0.070	0.501	0.019
$\rho_{12,2}^{(2)}$	0.500	0.420	0.127	0.494	0.095	0.504	0.021
$\rho_{13,1}^{(3)}$	0.500	0.447	0.124	0.497	0.091	0.508	0.012
$\rho_{13,2}^{(3)}$	0.500	0.480	0.178	0.495	0.071	0.502	0.011
$\rho_{23}^{(6)}$	0.500	0.455	0.139	0.499	0.061	0.501	0.014
θ_1	−1.000	−1.161	0.257	−0.995	0.164	−0.998	0.027
θ_2	1.000	1.073	0.290	1.020	0.160	0.997	0.012

3.2 Simulation study with covariates

In this subsection, simulation study is used to illustrate the application of our proposed model.

$$S_r^* = \beta_1 X + \varepsilon_r, \ \ Y_3^* = \beta_2 X + \varepsilon_3, \ \ Z_4 = \mu_z + \beta_3 X + \varepsilon_4, \qquad r = 1, 2.$$

In the simulation study, we set the true values of the parameters to be $\beta = (\beta_1, \beta_2, \beta_3) = (0, 0, 0)$. The response X was obtained by generating a uniform (0, 1) variable. Table 2 gives the results.

The parameter estimates by the model (for $n = 50$, n=100 and n=1000) are close to the true values of the parameters. Of course, the more the value of n the better the estimates.

Table 2: Results of the simulation study with Covariates

		$n = 50$		$n = 100$		$n = 1000$	
Parameter	True value	Est.	S.E.	Est.	S.E.	Est.	S.E.
β_1	0.000	0.068	0.120	0.032	0.112	0.003	0.023
β_2	0.000	0.073	0.142	0.024	0.107	0.005	0.013
β_3	0.000	0.050	0.201	0.028	0.133	0.002	0.076
μ_z	0.000	0.081	0.245	0.034	0.182	0.006	0.057
σ_z^2	1.000	1.067	0.213	1.035	0.071	1.008	0.022
$\rho_{12,1}^{(2)}$	0.500	0.487	0.106	0.485	0.069	0.509	0.033
$\rho_{12,2}^{(2)}$	0.500	0.431	0.115	0.490	0.081	0.507	0.040
$\rho_{13,1}^{(3)}$	0.500	0.468	0.157	0.479	0.063	0.503	0.025
$\rho_{13,2}^{(3)}$	0.500	0.454	0.159	0.497	0.068	0.501	0.031
$\rho_{23}^{(6)}$	0.500	0.461	0.143	0.495	0.084	0.504	0.021
θ_1	-1.000	-1.134	0.276	-0.983	0.084	-0.999	0.031
θ_2	1.000	1.099	0.281	1.073	0.186	0.995	0.044

4 Application

4.1 Data

The data used in this paper is excerpted from the 15th wave (2005) of the British Household Panel Survey (BHPS); a longitudinal survey of adult Britons, being carried out annually since 1991 by the ESRC UK Longitudinal Studies Center with the Institute for Social and Economical Research at the University of Essex. These data are recorded for 11251 individuals. The selected variables which will be used in this application are explaind in the following.

One of the responses is the life Satisfaction (LS), [where the related question is QA: “How dissatisfied or satisfied are you with your life overall?”] which is measured by directly asking the level of an individual’s satisfaction with life overall, resulting in a three categories ordinal variable [1: Not satisfied at all (10.300%). 2: Not satis/dissat (45.400%) and 3: Completely satisfied (44.300 %)]. The current economic activity (CEA) is also measured by a three-category nominal variable [where the related question is QB: ”Please look at this card and tell me which best describe your current situation?”] which is measured by directly asking the level of an individual’s CEA, resulting in a three categories nominal variable [1: self-employed or employed (63.110%), 2: unemployed (2.970%) and 3: Other (retired from paid work altogether, on maternity leave, family care, full-time student/at school, long term sick or disabled, on a government training scheme, other) (33.920%)]. Moreover, the exact amount of an individuals annual income (INC) in the past year in thousand pounds, considered here in the logarithmic scale, is also excerpted as a continuous response variable (mean: 4.068). As some values of annual income in thousand pounds are between 0 and 1, some of the logarithms of incomes are less than 0.

These three responses, LS, CEA and logarithm of income are endogenous correlated variables and should be modelled as a multivariate vector of responses.
Socio-demographic characteristics, namely: Gender [male: 44.200% and female: 55.800%], Marital Status (MS)[married or living as couple: 68.500%, widowed: 8.300%, divorced or separated: 8.400% and never married: 14.800%], Age (mean: 49.180) and number of people in the household (NPH) [less than four members: 67.27% and four or more than four members: 32.73%] are also included in the model as covariates. The vector of explanatory variables is

$$X = (Gender, Age, MS_1, MS_2, MS_3, NPH)$$

where MS_1, MS_2 and MS_3 are dummy variables for married or living as couple, widowed and divorced or separated.

4.2 Models for BHPS data

We apply the model described in section 2 to evaluate the effect of Age, Gender, NPH and MS simultaneously on LS, $CEA_r, \ \ r = 1, 2$ and Income. We shall also try to find answers for some questions, including (1) do a significant correlations exist between three responses? (2) what would be the consequence of not considering these correlations?

For comparative purposes, two models are considered. The first model (model I) is a marginal model which does not consider the correlation between three responses and can be presented as,

$$\begin{aligned} CEA_r^* &= \beta_{1,r}\ MS_1 + \beta_{2,r}\ MS_2 + \beta_{3,r}\ MS_3 + \beta_{4,r}NPH + \beta_{5,r}Gender + \beta_{6,r}AGE + \varepsilon_{1,r}^{(1)} \\ LS^* &= \beta_1\ MS_1 + \beta_2\ MS_2 + \beta_3\ MS_3 + \beta_4 NPH + \beta_5 Gender + \beta_6 AGE + \varepsilon_2 \\ \ln(INC) &= \beta_7 + \beta_8\ MS_1 + \beta_9\ MS_2 + \beta_{10}\ MS_3 + \beta_{11} NPH + \beta_{12} Gender + \beta_{13} AGE + \varepsilon_3^{(3)}, \end{aligned}$$

where $r = 1, 2$. The second model (model II) uses model I and takes into account the correlations between three errors. Here, influence of a small perturbation of the asymmetry parameter of the skew-normal distribution (λ_1 and λ_2) is also studied.

4.3 Results

This is confirmed by the curvature $C_{max} = 1.12$ computed from (4.2). This curvature does not indicate extreme local sensitivity. Here, a multivariate normal distribution is assumed for errors. Results of using two models are presented in Table 1. Deviance for testing model (I) against model (II) is equal to 294.180 with 5 degrees of freedom (P-value=0.000) which indicates that model (II) has a better fit to these data. As it can be seen, all correlation parameters are strongly significant. They show a positive correlation between LS and CEA_1 ($\hat{\rho}_{12,1}^{(2)}$ = 0.140) , a positive correlation between LS and CEA_2 ($\hat{\rho}_{12,2}^{(2)}$ = 0.137) , a positive correlation between $\log(INC)$ and CEA_1 ($\rho_{13,1}^{(3)}$ = 0.121) , a positive correlation between $\log(INC)$ and CEA_2 ($\rho_{13,2}^{(3)}$ = 0.141) a positive correlation between $\log(INC)$ and LS ($\rho_{23}^{(6)}$ = 0.136). Consideration of the responses associations yields more precise estimates as indicated by the smaller variance estimates and the smaller estimated variance of $\log(INC)$ and LS in model (II). So, we restrict our interpretation to the results of model (II).

Model (II) shows a significant effect of age (the older the individual the more the life satisfaction), MS (married people are more satisfied than never married people and divorced or separated people are less satisfied than never married people), NPH and gender on the life satisfaction status. All explanatory variables have significant effect on the nominal response of the current economic activity.

Also the effect of all explanatory variables are significant on the logarithm of income. Never married people have less logarithm of income than married people and divorced or separated people. Females have more logarithm of income than males and the older people earn less money than younger ones.

5 Discussion

In this paper a multivariate latent variable model is presented for simultaneously modelling of nominal, ordinal and continuous correlated responses. We assume a multivariate normal distribution for errors in the model. However, any other multivariate distribution such as t or logistic can be also used. Binary responses are a special case of ordinal responses. So, our model can also be used for mixed binary and continuous responses. Generalization of our model for nominal, ordinal and continuous responses with missing responses is an ongoing research on our part.

References

[1] Ash, R. B. and Dolens-Dade, C. A. (2000), *Probability and measure theory*: Academic Press.

[2] Samani, E.B., Ganjali, M. and Khodaddadi, A. (2008). A latent variable model for mixed continuous and ordinal responses. *Journal of Statistical Theory and Applications*, **7**(3), 337–348.

[3] Catalano, P. and Ryan, L. M. (1992), Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **50**(3), 1078–1095.

[4] Cox, D. R, and Wermuth, N. (1992), Response models for mixed binary and quantitative variables. *Biometrika*, **79**(3), 441–461 .

[5] Cook, R. D. (1986). Assessment of local influence. *Journal of Royal Statistical Society, Series B*, **48**, 133–169.

[6] Fitzmaurice, G. M. and Laird, N. M. (1997), Regression models for mix discrete and continuous responses with potentially missing values. *Biometrics*, **53**, 110–122 .

[7] Fletcher, R. (2000), *Practical Methods of optimization*. John Wiley.

[8] Heckman, J. J. D. (1978), Endogenous variable in a simultaneous Equation system. *Econometrica*, **46**(6), 931-959.

[9] Olkin L. and Tate R. F. (1961), Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*, **32**, 448-456.

[10] Poon, W. Y. and LEE, S. Y. (1987), Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika* **52**(3), 409–430.

[11] Yang, Y., Kang, J., Mao, K. and Zhang, J. (2007), Regression models for mixed Poisson and continuous longitudinal data. *Statistics in Medicine*, **26**, 3782-3800.

Table 3: Results of using two models for BHPS data (LS: Life Satisfaction, CEA: The current economic activity, MS: Marital Statue and NPH: number of people in the household, parameter estimates highlighted in **bold** are significant at 5 % level.)

	Model I		Model II	
Parameter	Est.	S.E.	Est.	S.E.
Response: CEA_1 **(employed)**				
Marital Status (baseline: Never married)				
Married or Living as Couple	**1.668**	**0.035**	**1.262**	**0.044**
Widowed	**1.563**	**0.053**	**1.209**	**0.073**
Divorced or Separated	**0.431**	**0.089**	**0.268**	**0.037**
NHP (baseline: NPH$\leq$ 3)				
NPH ≥ 4	**-0.339**	**0.030**	**-0.211**	**0.045**
Gender(baseline: Male)				
Female	**-0.846**	**0.028**	**-0.711**	**0.037**
AGE	**-0.075**	**0.001**	**-0.061**	**0.001**
Response: CEA_2 (unemployed)				
Marital Status (baseline: Never married)				
Married or Living as Couple	**0.455**	**0.086**	**0.295**	**0.096**
Widowed	**1.390**	**0.122**	**1.220**	**0.130**
Divorced or Separated	**0.431**	**0.062**	**0.268**	**0.049**
NHP (baseline: NPH$\leq$ 3)				
NPH ≥ 4	**-0.398**	**0.080**	**-0.336**	**0.075**
Gender(baseline: Male)				
Female	**-1.057**	**0.065**	**-0.996**	**0.070**
AGE	**-0.094**	**0.001**	**-0.071**	**0.001**
Response: LS				
Marital Status (baseline: Never married)				
Married or Living as Couple	**-0.468**	**0.035**	**-0.300**	**0.006**
Widowed	**0.667**	**0.050**	**0.494**	**0.053**
Divorced or Separated	**-0.622**	**0.080**	**-0.358**	**0.043**
NHP (baseline: NPH$\leq$ 3)				

Table 3: Results of using two models for BHPS data (LS: Life Satisfaction, CEA: The current economic activity, MS: Marital Statue and NPH: number of people in the household, parameter estimates highlighted in **bold** are significant at 5 % level.)

	Model I		Model II	
Parameter	Est.	S.E.	Est.	S.E.
NPH$\geq$ 4	**0.215**	**0.029**	**0.184**	**0.029**
Gender(baseline: Male)				
Female	**0.069**	**0.037**	**0.039**	**0.021**
AGE	**0.009**	**0.001**	**0.005**	**0.001**
cut-point 1	**-1.659**	**0.370**	**-0.985**	**0.100**
cut-point 2	**0.785**	**0.377**	**0.448**	**0.117**
Response: $\log(INC)$				
Constant	**5.245**	**0.070**	**4.255**	**0.029**
Marital Status (baseline: Never married)				
Married or Living as Couple	**0.114**	**0.010**	**0.114**	**0.009**
Widowed	**0.215**	**0.019**	**0.216**	**0.019**
Divorced or Separated	**0.213**	**0.078**	**0.213**	**0.016**
NHP (baseline: NPH$\leq$ 3)				
NPH$\geq$ 4	**- 0.090**	**0.375**	**-0.063**	**0.096**
Gender(baseline: Male)				
Male	**-0.227**	**0.007**	**-0.227**	**0.007**
AGE	**-0.002**	**0.001**	**-0.002**	**0.001**
σ^2	**0.181**	**0.002**	**0.154**	**0.002**
Correlations				
Corr(CEA$_1^*$,LS*)	-	-	**0.140**	**0.011**
Corr (CEA$_1^*$,INC)	-	-	**0.141**	**0.002**
Corr(CEA$_2^*$,LS*)	-	-	**0.137**	**0.014**
Corr (CEA$_2^*$,INC)	-	-	**0.121**	**0.001**
Corr(LS*,INC)	-	-	**0.135**	**0.012**
-Loglike	39367.500		39220.410	