Modelling Satisfaction with Leisure Activities Over Time Using Graphical Chain Models

Salgueiro, Maria de Fátima
Instituto Universitário de Lisboa (ISCTE-IUL), Department of Quantitative Methods and UNIDE
Av. Forças Armadas,
1649-026 Lisboa, Portugal
E-mail: fatima.salgueiro@iscte.pt

Smith, Peter W.F.
University of Southampton, Southampton Statistical Sciences Research Institute
Highfield
Southampton, SO17 1BJ, United Kingdom
E-mail: P.W.Smith@soton.ac.uk

Introduction

In recent years subjective well-being (SWB) has been a subject of great interest by psychologists, economists, social scientists and policy makers, and an area of on-going research. Statistical approaches adopted in the literature to model SWB often consider an overall measure or a derived total score. Ordered probit models, and fixed and random effects models have often been used, mostly with cross-sectional data.

Panel survey data allow the investigation of stability and change over time. Recent approaches to modelling longitudinal survey data include structural equation models and graphical chain models (GCMs) - see Berrington et al. (2008). Graphical modelling is a statistical technique explicitly based on the concept of conditional independence: the emphasis is on the assessment of the possible relationships between pairs of variables, controlling for other variables under analysis. The independence structure of the variables in the model is represented in a graph. To model longitudinal data, with a natural ordering between the variables, chain graphs with ordered blocks are required, and GCMs have to be considered – see Cox and Wermuth (1996).

The British Household Panel Survey (BHPS) is a national representative survey of individuals in private domiciles, conducted in Great Britain, since 1991, on an annual basis (see Taylor et al., 2008). The BHPS aims at providing information about social and economic change, both at the individual and at the household level. Several SWB measures are available in the BHPS. Since 1996 respondents have been asked to rate their satisfaction levels with several domain dimensions of life satisfaction, namely amount of leisure time, use of leisure time and social life.

The current paper discusses the use of GCM to model longitudinal perceptions of satisfaction with leisure activities, and their possible predictors. BHPS data from years 2002 to 2005 are used. The sample includes 1970 employees, who were original sample members and fully answered all questions under analysis. The proposed GCM has five blocks. Blocks 2 to 5 include the repeated measures of the satisfaction with leisure activities derived score (computed as a sum of three items). Block 1 includes the 2002 measures of the possible determinants of satisfaction: age, gender, marital status, perceived health status, qualifications, social class, number of children in the household, number of hours worked per week and household income.

The capabilities of the software MIM (Edwards, 2000), available for performing model selection in GCMs, are discussed.

The BHPS sample under analysis

This paper analyses data from the British Household Panel. BHPS respondents are asked to rate their satisfaction levels, namely with the amount of leisure time they have; the way they spend their leisure time; and their social life, using a seven-point Likert-type scale from 1 (not satisfied at all) to 7 (completely satisfied). Preliminary results from factor analysis suggest these three observed variables can be considered indicators of single factor of satisfaction with leisure activities. Thus, for the analysis conducted in this paper a derived score variable, the satisfaction with leisure activities score, was computed as a sum of three items. The score variable ranges from 3 (the lowest satisfaction value) to 21 (the highest satisfaction value). The repeated measures of the score variable are modelled in this paper as continuous variables.

The BHPS subsample of 1970 individuals considered in this paper is now characterized. In 2002, the age of the respondents ranges from 16 to 77 years old, with a mean value of 40.82. Of the respondents, 51.4% are females. Regarding marital status, 74.4% of the respondents are married or living as a couple, and 25.6% are widowed, divorced or separated. In terms of education, 67.3% of the respondents have A levels or above and 32.7% have less than A levels. Of the respondents, 42% have a professional or intermediate social class, 41.3% are skilled and the remainder are partly skilled or unskilled. In 61.1% of the households there are no children under 12 years old, and in 38.9% of the households there is at least one child. The health status is perceived as excellent by 27.3% of the respondents, 50.8% rate it as good and the remainder perceive a fair, poor or very poor health status. The number of hours normally worked per week has a maximum value of 78 hours, with a mean of 33.88 hours. The income of the household ranges from 244.7 to 33 690.5 pounds, with a mean value of 3 403.1 pounds. The logarithm of the income was computed and considered for the analysis conducted in this paper. The mean satisfaction levels have remained quite stable over the period under analysis: 14.08 (in 2002); 14.05 (in 2003); 14.13 (in 2004) and 13.59 (in 2005).

Graphical chain models

Graphical modelling is a form of multivariate analysis that uses mathematical graphs to represent models. It has emerged as a statistical technique explicitly based on the concept of conditional independence: the emphasis is on the assessment of the relationships between pairs of variables, conditioning on (controlling for) other variables under analysis. A the key tool in graphical modelling is the graph of the model. Vertices in the graph correspond to variables and edges correspond to associations between variables. The Markov properties can be used to read the association structure between the variables in the model directly from the graph. Edge exclusion tests allow the testing of conditional independencies between pairs of variables, given other variables in the model. For an introduction to graphical models see Whittaker (1990); for a more general discussion see Cox and Wermuth (1996).

When there is a natural ordering of the variables in the model, as is the case when longitudinal data is used and variables are observed over time, graphical chain models (GCM) should be considered. The chain graph of the model consists of a set of blocks of variables separated by time and the chain structure is supplied from subject-matter knowledge about responses and potential influences. The relationship between variables in the same block is represented by undirected edges, whereas the relationship between variables in two different blocks is represented by directed edges.

MIM (M*ixed* I*nteraction* M*odelling* – www.hypergraph.dk) is a program developed by Edwards (2000), currently available for fitting graphical chain models with both continuous and discrete variables. MIM offers various options for model selection, including automatically implemented procedures, such as forward selection and backwards elimination. The data analyst can also perform model selection manually, by specifying the models to be tested, and the edges to be added or to be deleted from the model at each step.

The proposed GCM

The dependence chain graph proposed in this paper to model satisfaction with leisure activities over time, and its determinants, contains five blocks. Satisfaction with leisure activities is measured by a derived score, computed as the sum of the responses of the individuals to three dimensions: satisfaction with the amount of leisure time, satisfaction with the use of leisure time and satisfaction with social life. The first block (b1) includes the 2002 measures of the explanatory variables considered as possible determinants of satisfaction with leisure activities: age, gender, marital status, perceived health status, qualifications, social class, number of children in the household, number of hours worked per week and household income. Blocks 2 to 5 (b2 to b5) include the repeated measures of the leisure activities satisfaction score, from 2002 to 2005 respectively. It is expected that the repeated measures of satisfaction with leisure activities are not conditionally independent, given all variables in past blocks, and therefore all possible directed edges between blocks 2 to 5 are not ruled out a priori.

The absence of an undirected edge between two variables in b1 means that the two possible determinants of job satisfaction are conditionally independent, given the remaining determinants. The absence of a directed edge between a variable in b1 and, for example, the 2003 score of satisfaction with leisure activities (in b3), would indicate a conditional independence between satisfaction with leisure activities in 2003 and the covariate, given the 2002 score of satisfaction with leisure activities (in b2) and the remaining covariates (in b1). From a substantive point of view all possible directed edges between variables in b1 and variables in b2 to b5 are postulated to be present in the initial chain graph.

Modelling satisfaction with leisure activities over time: methodological options and results

The derived score is assumed to follow a normal distribution. Conditional-Gaussian distributions are used, allowing the simultaneous consideration of quantitative and qualitative variables in the model.

The purpose of the modelling selection strategy is to detect conditional independencies, thus providing the analyst with a final model where all edges present correspondent to statistically significant associations. Alternative options available in MIM to perform model selection in GCM are now discussed and used.

Because there are several nominal variables in b1, and to avoid higher order interaction terms, we start with the main effects model and perform *stepwise forward selection*, using either *unrestricted models* or *decomposable models* (those with explicit formulae for the maximum likelihood estimators). MIM tests for the inclusion of each of the 36 possible undirected edges among the nine variables in b1. Figure 1 displays the independence graph for the variables in b1, with the 12 edges added by MIM when just decomposable models are considered. It is possible to conclude that, in 2002, marital status, age, income and having children in the household are conditionally independent from gender, number of hours worked, education and perceived health status, given social class.

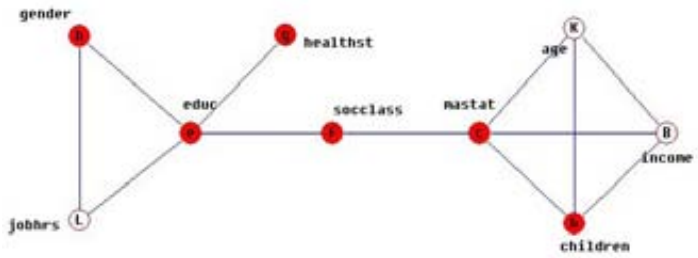

Figure 1

Conditional independence structure among the possible determinants of satisfaction with leisure activities (in 2002), when only decomposable models are considered in the model selection procedure.

Stepwise forward selection considering unrestricted models was also used, resulting in a chosen model with many more edges present. A stepwise backwards elimination procedure was then applied to the obtained model, in order to test for possible edge removal. Figure 2 displays the independence graph of the resulting model. It is possible to conclude that the conditional independence structure among variables in b1 is not as simple as the one obtained when just decomposable models are considered – four of the edges present in Figure 1 have been removed and 12 new edges are now present in the graph of the chosen model for b1 depicted in Figure 2.

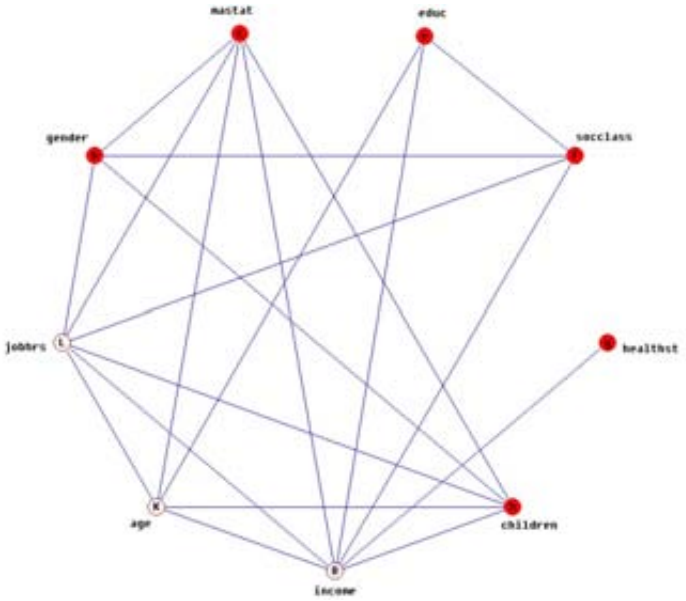

Figure 2

Conditional independence structure among the possible determinants of satisfaction with leisure activities (in 2002), when unrestricted models are considered when performing model selection.

Once the model for variables in b1 has been selected, in the second stage of the model selection procedure MIM tests for the inclusion of all possible directed edges between the nine determinants of satisfaction with leisure activities (in b1) and the 2002 measure of the satisfaction score (in b2). When just decomposable models are considered, MIM adds seven edges between variables in b1 and the satisfaction score in b2 (L02). These edges are displayed in Figure 3. Two edges are not added, corresponding to conditional independences between L02 and income of the household, and between L02 and age of the respondent, given the remaining determinants of satisfaction with leisure activities in b1.

In the third stage of the model selection procedure MIM tests for the inclusion of all possible directed edges between variables in b1 (and b2) and the 2003 measure of the satisfaction score (L03, in b3). The directed edge between L02 and L03 is included in the model, suggesting that the 2003 measure of satisfaction with leisure activities is not conditionally independent of the 2002 measure, given all nine determinants of satisfaction. Three out of the nine possible directed edges between variables in b1 and b3 are not included in the model, corresponding to conditional independences between L03 and income of the household, age of the respondent and number of hours worked per week, given the remaining determinants of satisfaction with leisure activities in b1 and the 2002 satisfaction score.

Similar reasoning applies to the fourth and fifth stages of the model selection procedure, when MIM tests for the inclusion of all possible directed edges between variables in previous blocks and the satisfaction score measure in b4 and b5, respectively. From Figure 3 it is possible to conclude that both the 2004 and the 2005 measures of the satisfaction score are conditionally independent of the income of the household, of the

age of the respondent and of the number of hours worked per week, given all other variables in previous blocks. It is also possible to conclude that all lagged effects between the repeated measures of satisfaction with leisure activities are present in the model – represented by the directed edges between variables in b2 to b5 (see Figure 3).

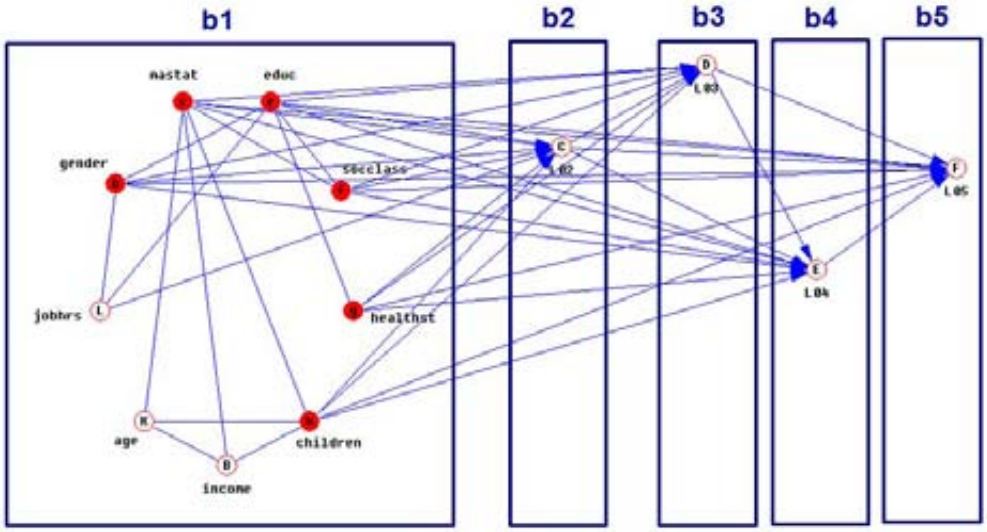

Figure 3

Chain graph of the chosen model for satisfaction with leisure activities over time (and possible determinants) when only decomposable models are considered in the model selection procedure.

Model selection was also conducted considering unrestricted models and performing a stepwise forward selection, starting with the main effects model and taking into account the specified chain structure. A backwards elimination procedure was then applied to the selected model, in order to test for possible edge removal. Figure 4 displays the chain graph of the final chosen model.

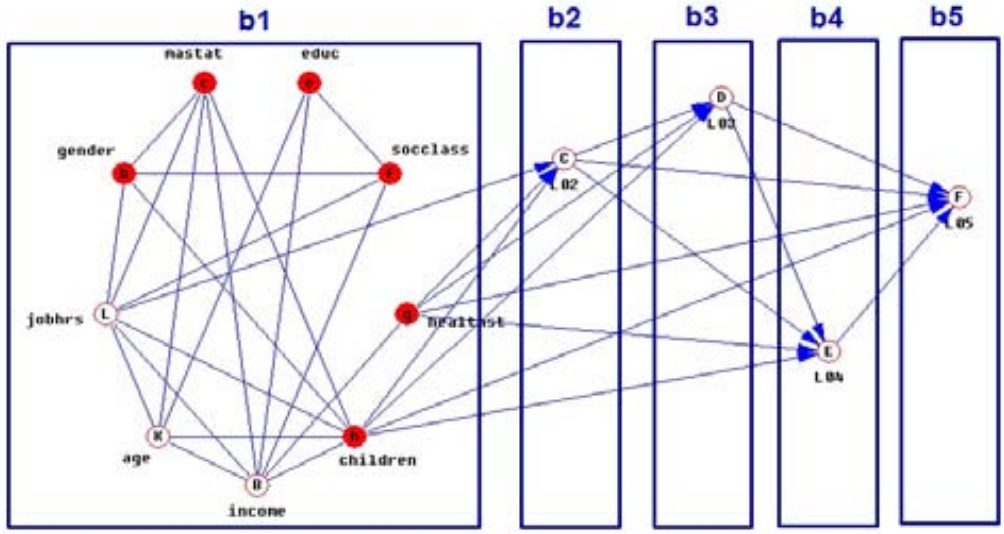

Figure 4

Chain graph of the chosen model for satisfaction with leisure activities over time (and possible determinants) when unrestricted models were considered in the model selection procedure.

From Figure 4 it is possible to conclude that, as in the case of just considering decomposable models (in Figure 3), all lagged effects between the repeated measures of satisfaction with leisure activities are present in the model. Thus, all directed edges between variables L02, L03, L04 and L05 are present in the chain graph (Figure 4).

It is also possible to conclude that fewer edges have been included in the model in stages two to five of the model selection procedure. Indeed, satisfaction with leisure activities in 2002 is only conditionally dependent on the number of children in the household, the perceived health status and the number of hours worked per week, given the remaining possible determinants of satisfaction in b1. These associations correspond to the three directed edges between variables in b1 and b2, present in Figure 4. Only two directed edges are present between variables in b1 and variables in b3 (and b4 and b5), suggesting that the repeated measures of satisfaction with leisure activities in 2003, 2004 and 2005 are not conditionally independent of perceived health status and number of children in the household, given all other variables in previous blocks.

Discussion

This paper has illustrated how to model satisfaction with leisure activities over time (and its determinants) using graphical chain models. The statistical software MIM has been used to perform model selection. Given the specified chain structure, and starting with the main effects model, a stepwise forward model selection procedure, followed by a backwards elimination procedure, have been proposed. Both decomposable and unrestricted models have been considered.

Results show that a simpler association structure, with more conditional independencies and fewer edges present in the graph, is obtained for variables in b1 when just decomposable models are considered. However, fewer edges are present between variables in b1 and variables in the subsequent blocks of the chain graph of the chosen model when unrestricted models are allowed. It might be expected that performing model selection using unrestricted models would result in a more complicated model, since just considering decomposable models does not allow the search across all possible models. However, this is not always the case with the models proposed in the current paper, and requires further research and explanation, both with this and other datasets.

Similar models can be fitted in other statistical packages (for example STATA), using a sequence of univariate regressions, when just one dependent variable is considered, as is the case in this paper. However, the graphical modeling framework is more flexible to model longitudinal data when there are repeated measures of two or more correlated variables.

One should note that a derived score of satisfaction was computed and used. Indeed, it is not yet possible to account for measurement error (i.e., to include latent variables in the model) when performing model selection in graphical chain models.

REFERENCES

Berrington, A., Hu, Y., Smith, P.W.F. and Sturgis, P. (2008): A graphical chain model for reciprocal relationships between women's gender role attitudes and labour force participation. Journal of the Royal Statistical Society, A, 171(1), 89-108.

Cox, D.R. and Wermuth, N. (1996): Multivariate dependencies – models, analysis and interpretation. London: Chapman and Hall.

Edwards, D. (2000): Introduction to Graphical Modelling. 2nd ed. New York: Springer-Verlag.

Taylor, M.F. (Ed.), with Brice, J., Buck, N., and Prentice-Lane, E. (2008): British household panel survey user manual volume A: Introduction, technical report and appendices. Colchester: University of Essex.

Whittaker, J. (1990): *Graphical Models in Applied Multivariate Statistics*, Chichester, John Wiley & Sons.