AGE-PERIOD-ENVIRONMENT MODEL FOR ANALYZING CONSUMER SURVEY DATA FOR MARKETING RESEARCH

Hanayama, Nobutane
Shobi university, Information technology
Shimomatsubara 655
Kawagoe 3501153, Japan
E-mail: nob-hanayama@jcom.home.ne.jp

1. Introduction

Age-period-cohort analysis is a method of research developed primarily by demographers and it has been adapted to the study of various attitudinal and behavioral phenomena (Glenn 1977). Also in the field of marketing research, since the seminal work of Reynolds and Rentz (1981), cohort analysis has gained recognition. As for data on habitual or behavioral phenomena for Japanese consumers, Mori (2004) and Fukuda (2010), for example, successfully applied the age-period-cohort model to them and obtained significant results.

The term cohort originally referred to a Roman military unit. In technical language of demography, however, a cohort is defined as those people within a geographically or otherwise delineated population who experienced the same significant life event within a given period of time (Glenn, 1977). And, in recent statistical study, the age-period-cohort analysis indicates a method for projecting age, period and cohort effects on human attitude, behavior, morbidity, or death etc. from data given by age and period which is occasionally called cohort table (Fienberg and Mason, 1978).

Though the age-period-cohort model has been successfully fitted to those data, there exists an exact linear dependency among those three factors as described by Fienberg and Mason (1978), hence even the first-order differences of age, period or cohort effects are not estimable (Scheffé, 1959) in the most likelihood estimation for the model. So far several researchers have been suggested measures for addressing the problem which include a measure focusing estimable functions (Holford, 1983), a measure using Bayesian method (Nakamura, 1982) and nonlinear models (Moolgavkar et al., 1979 and Hanayama, 2007).

In the field of marketing research, prior to those measures, the constrained multiple regression model has been applied by Rentz et al. (1983) and Rentz and Reynolds (1991) for addressing the identifiability problem in the model. Fukuda (2010), however, pointed out that the estimates of parameters in the model are sensitive to the arbitrary choice of the identifying constraint, and applied the Bayesian cohort model developed by Nakamura (1982, 1986) to the data on household vehicle expenditure in the U.S. and Japan.

Though the Bayesian age-period model has been successfully applied to data on marketing research, it needs a special tool for estimating the parameters in the model so that versatile tools for regression analysis are not available. So our aim in this study is to introduce a model which is equivalent to the original age-period-cohort model in the meaning that they have the same space spanned by column vectors in their design matrix and show that the first-order trends as well as higher-order trends those three factors are estimable (see Section 2). Further the introduced model is applied to the data on consumer preference obtained from JNN Data Bank conducted by Japanese 28 TV stations including Tokyo Broadcasting System, Inc. (TBS) as their key station (see Section 3). Finally concluding remarks are described in Section 4.

2. Age-period-environment model

Consider groups of N_{ij} people who are in the $[A_{i-1}, A_i)$ age group at the time (year) P_j, where $i = 1,\cdots,I$; $j = 1,\cdots,J$ and $A_i - A_{i-1} = P_j - P_{j-1} = \tau$. Let $Y_{ij}^{(n)}$ ($n = 1,\cdots N_{ij}$) be a random variable which indicates whether the n_{ij} th member of N_{ij} people has a certain property ($Y_{ij}^{(n)} = 1$) or not ($Y_{ij}^{(n)} = 0$), where "a certain property" means, for example, being alive, having a certain attitudinal or behavioral feature, or saying "yes" on questionnaire. The response variables $Y_{ij}^{(n)}$ are illustrated on the Lexis diagram (Keiding, 1990) in Figure 1.

In age-period-cohort the following model is assumed on $Y_{ij}^{(n)}$:

$$\eta\left(E[Y_{ij}^{(n)}]\right) \equiv \eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{j-i+1}, \tag{2.1}$$

where $\eta(\cdot)$ is a Link function used in the generalized linear model (McCullagh and Nelder, 1989), μ is the total mean, α_i and β_j are the age and period effects associated with $[A_{i-1}, A_i)$ and P_j, and γ_{j-i+1} is the cohort effect associated with the group of people who were in the $[A_0, A_1)$ age group at the year P_{j-i}. As long as the idea of birth cohort γ_{j-i+1} indicates the effect of exposure to the environment at P_{j-i} which affect only the $[A_0, A_1)$ age group of people, that is, babies or children.

Generally in the generalized linear model it is needed that, for every set of η_{ij} which satisfy the model, values of parameters in the model are uniquely determined for getting estimates of the parameters. If not it is needed to assume plausible constrains on the parameters under which values of parameters are uniquely determined for every set of η_{ij}. As for the age-period-cohort model, it is seen that, for every set of η_{ij} satisfying the model (2.1),

$$\eta_{ij} = \mu^* + \alpha_i^* + \beta_j^* + \gamma_{j-i}^*, \tag{2.2}$$

holds for any $(a,b,c,d)' \in R^4$, where $\mu^* = \mu + a + b$, $\alpha_i^* = \alpha_i - a + id$, $\beta_j^* = \beta_j - b - jd$ and $\gamma_{j-i+1}^* = \gamma_{j-i+1} - c + (j-i)d$, hence there remains an additional identifiability problem even when plausible constraints such that $\alpha_1 = \beta_1 = \gamma_{J-1} = 0$, which are equivalent to $a = \alpha_i$, $b = \beta_j$ and $c = \gamma_{J-1}$, are assumed on the parameters.

For addressing the above problem Rentz et al. (1983) and Rentz and Reynolds (1991) assume that $\alpha_{I-1} = \alpha_I$ or $\beta_1 = \beta_2$ or $\gamma_{1-I} = \gamma_{2-I}$ in addition to $\alpha_1 = \beta_1 = \gamma_J = 0$. Though the values of parameters are uniquely determined under one of those constraints in addition to $\alpha_1 = \beta_1 = \gamma_{J-1} = 0$, the first-order (linear) trend of age or period or cohort factor is or changed depending on choice of constraint. Considering the above problem we introduce the following model:

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=j-i+1}^{j-1} \xi_k, \tag{2.3}$$

where μ, α_i and β_j are common with ones in the original age-period-cohort model and ξ_k indicates the effect associated with the environment at the year P_k ($k = 2-I,\cdots,J$) which comes out τ years after exposure to it. The idea of environmental effect and the deference form the period effect are illustrated in Figure 2.

Because of the words "environment at the year" and "exposure to the environment" used in its definition we may consider that the environmental effect is similar to the period effect or cohort effect. The environmental effect is, however, different from the period effect in the meaning that the period effect is one on all people being alive at the time and does not maintain effectiveness in the next period while the environmental effect maintain effectiveness continuously after exposure to the environment. Besides the environmental effect is different from the cohort effect in the meaning that the cohort effect in the meaning that the cohort effect is just on the $[A_0, A_1)$ age group of people while the environmental effect is on all of people being alive at the period.

Using the vector matrix expression the age-period-environment model is expressed like

$$\boldsymbol{\eta} = X\boldsymbol{\theta} \tag{2.4}$$

where $\boldsymbol{\eta} = (\eta_{11}, \eta_{12}, \cdots, \eta_{IJ})'$, $X = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \cdots, \boldsymbol{x}_{IJ})'$, $\boldsymbol{x}_{ij} = (1, \mathbf{0}_{i-1}, 1, \mathbf{0}_{I-i}, \mathbf{0}_{j-1}, 1, \mathbf{0}_{J-j}, \mathbf{0}_{I-i+j-1}, \mathbf{1}_{i-1}, \mathbf{0}_{J-j})$ and $\boldsymbol{\theta} = (\mu, \alpha_1, \cdots, \alpha_I, \beta_1, \cdots, \beta_J, \xi_{2-I}, \cdots, \xi_{J-1})'$, where $\mathbf{0}_k$ and $\mathbf{1}_k$ are k dimensional row vectors and $\mathbf{0}_k = (0, 0, \cdots, 0)$ and $\mathbf{1}_k = (1, 1, \cdots, 1)$; $k = 1, 2, \cdots$. In a special case that $I = 5$ and $J = 3$ the model $\boldsymbol{\eta} = X\boldsymbol{\theta}$ concretely expressed like:

$$\begin{pmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \\ \eta_{21} \\ \eta_{22} \\ \eta_{23} \\ \eta_{31} \\ \eta_{32} \\ \eta_{33} \\ \eta_{41} \\ \eta_{42} \\ \eta_{43} \\ \eta_{51} \\ \eta_{52} \\ \eta_{53} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \xi_{-3} \\ \xi_{-2} \\ \xi_{-1} \\ \xi_0 \\ \xi_1 \\ \xi_2 \end{pmatrix}. \tag{2.5}$$

As seen from the above formula (2.5) the values of $\boldsymbol{\theta}$ are uniquely determined under one of those constraints in addition to the constraints $\alpha_1 = \beta_1 = \xi_{J-1} = 0$.

3. Result

In this section the results of fitting the age-period-environment model to the data on consumer preference obtained from JNN Data Bank (http://www.tbs.co.jp/research/index-j.html) conducted by Japanese 28 TV stations including Tokyo Broadcasting System, Inc. (TBS) as their key station are shown. The data is given in Table 1 which indicates the rates of numbers of people who reply "yes" to the question "do you like pizza?" given in every 5-year by 5-year age group.

The estimates of age, period and environmental effects are indicated in Figure 3, 4 and 5 respectively.

4. Conclusion

In this study the age-period-environment model has been introduced as one alternative to the age-period cohort model and fitted to data on rates of Japanese people with a taste for pizza.

Subtitle

AGE-PERIOD-ENVIRONMENT MODEL AGE-PERIOD-ENVIRONMENT MODEL

Figure or Table Title

Table 1. Rates of numbers of people who reply “yes” to the question “do you like pizza?” in percent

Year / Age group	1977	1982	1987	1992	1997	2002	2007
16-19	56.95	64.3	75.25	69.9	68.55	67.4	69.85
20-24	56.3	53.1	55	50.7	59.3	62.6	64.8
25-29	36.4	45.4	56.7	55	49.3	54	59.6
30-34	23.7	35.3	47.9	52.2	54.7	55.7	53.7
35-39	19.4	25.7	38	51.2	54.8	52.4	50.8
40-44	14.2	17.8	33.2	36.2	42.1	52.5	47.1
45-49	15.3	16.8	18.8	26.4	32.4	49.4	48.2
50-54	17.8	14.7	19.6	13.9	21.2	36.8	49
55-59	11.4	13.1	13.8	19.2	18.4	24.4	39.9

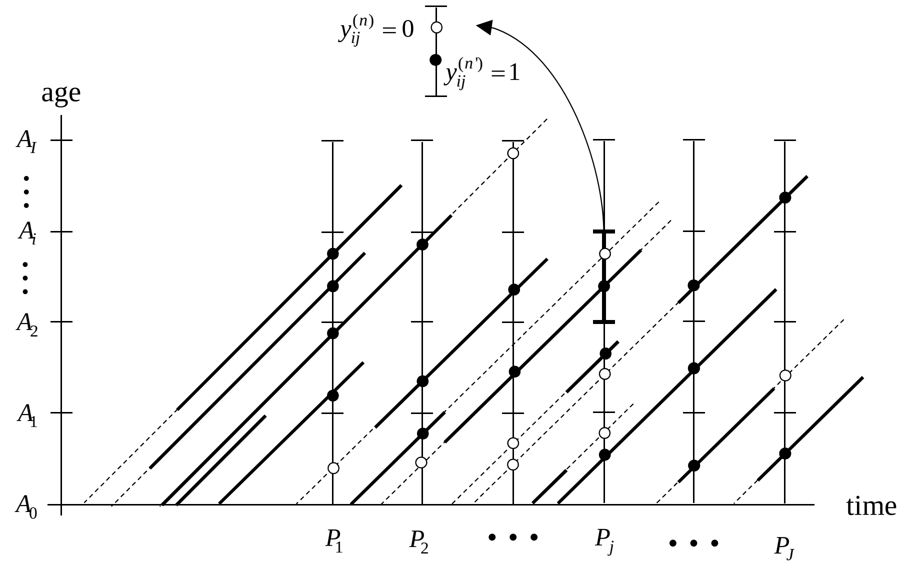

Figure 1. Response variables on the Lexis diagram

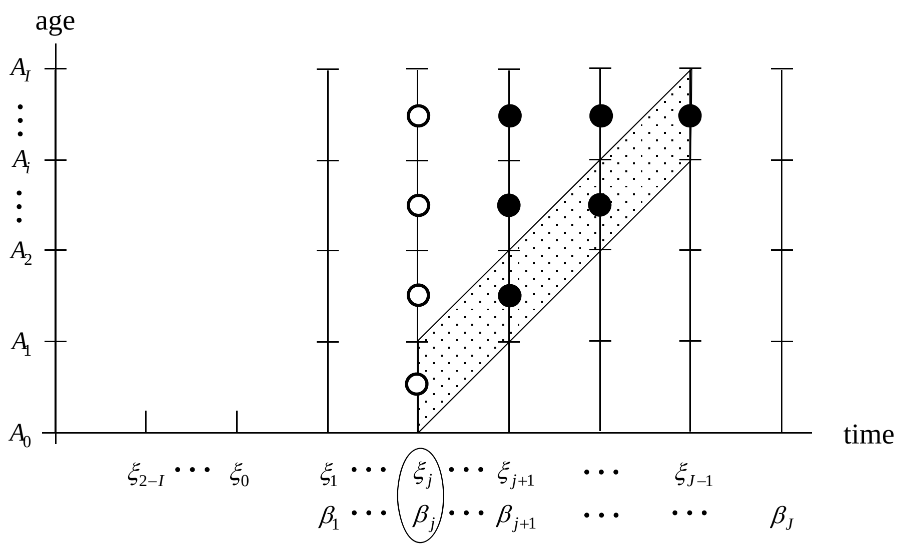

Figure 2. Environmental, period and cohort effect on the Lexis diagram

Figure 3. Estimates of the age effect

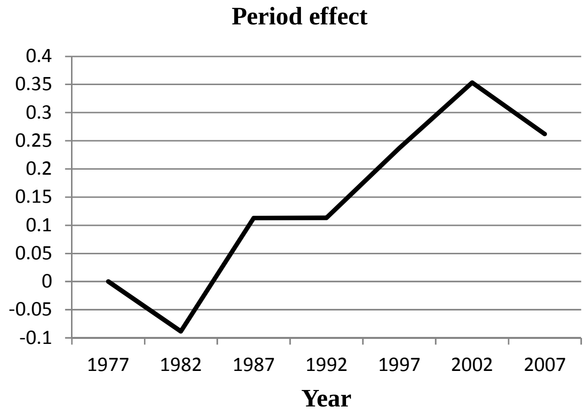

Figure 4. Estimates of the period effect

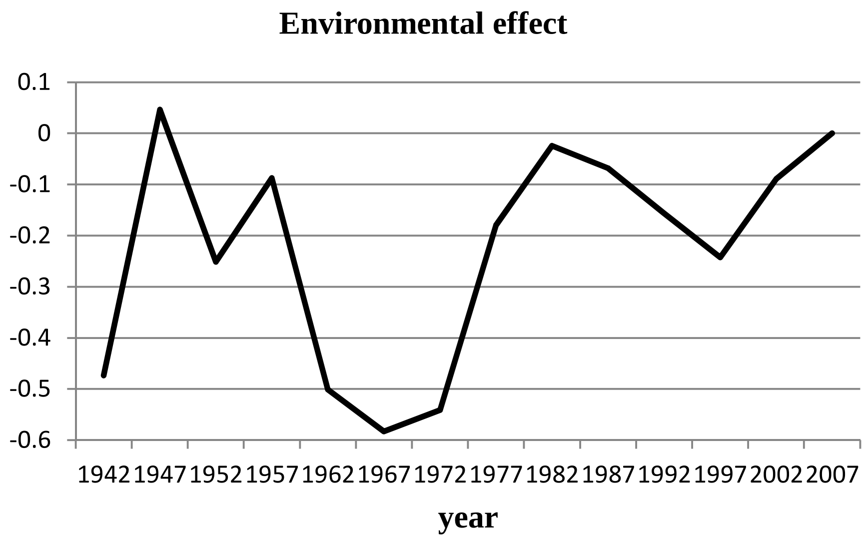

Figure 5. Estimates of the environment effect

REFERENCES

Glenn, ND. Cohort analysis. London: Sage Publications, 1977; p 8.

Reynolds, F. D., & Rentz, J. O. (1981). Cohort analysis: an aid to strategic planning. Journal of Marketing, 45, 62–70. doi:10.2307/1251542

Fukuda, Kosei (2010). A cohort analysis of household vehicle expenditure in the U.S. and Japan: A possibility of generational marketing. Marketing Letters Volume 21, Number 1, 53-64, DOI: 10.1007/s11002-009-9077-2

Mori, Hiroshi and Clason, Dennis L. (2004). A Cohort Approach for Predicting Future Eating Habits: The Case of At-Home Consumption of Fresh Fish and Meat in an Aging Japanese Society. International Food and Agribusiness Management Review, Volume (Year): 07; Issue (Month): 01.

Fienberg, S.E. and Mason, W.M. (1978). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, in Schuessler, K.F. (ed.) Sociological Methodology. Josey-Bass: San. Francisco, pp 1-67.

Scheffé, H. (1959). The Analysis of Variance. New York: John Wiley & Sons: pp.14-19.

Nakamura, T. (1982). A Bayesian cohort model for standard cohort table analysis. Proceedings of the Institute of Statistical Mathematics, 29, 77–97. In Japanese.

Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rate. Biometrics 39, 311-324.

Moolgavkar S.H., Stevens R.G. and Lee J.A. (1979). Effect of age on incidence of breast cancer in females. Journal of National Cancer Institute 62, 493-501.

Hanayama, N. (2007). An extended age period cohort model for analysing (age, period)-tabulated data. Statistics in Medicine 26, 3459-3475.

Rentz, J. O., & Reynolds, F. D. (1991). Forecasting the effects of an aging population on product consumption: an age–period–cohort framework. JMR, Journal of Marketing Research, 28, 355–360. doi:10.2307/3172871.

Rentz, J. O., Reynolds, F. D., & Stout, R. G. (1983). Analyzing changing consumption patterns with cohort analysis. JMR, Journal of Marketing Research, 20, 12–20. doi:10.2307/3151407.

Nakamura, T. (1982). A Bayesian cohort model for standard cohort table analysis. Proceedings of the Insti tute of Statistical Mathematics, 29, 77–97. In Japanese.

Nakamura, T. (1986). Bayesian cohort models for general cohort table analyses. Annals of the Institute of Statistical Mathematics, 38B, 353–370. doi:10.1007/BF02482523.

McCullagh, P, Nelder, JA. Generalized Linear Model 2nd ed., London: Chapman & Hall, 1989.

Keiding, N. Statistical inference in the Lexis diagram. Philosophical Transactions of the Royal Society of London Series A 1990; 332: 487-509.

ABSTRACT

The age, period and cohort (APC) model has received a considerable popularity for statistical analysis of data on consumer behaviors over the long term. In APC model, however, there exists an exact linear dependency among the three factors; hence even the first-order (linear) trend of each factor cannot be identified in the model. In this study, as an alternative one to APC model, a model which has age, period and environmental (APE) effects as its parameters is introduced for analyzing (age, period)-tabulated data on consumer survey for marketing research. It is shown that APE model is free from the non-identifiability problem from which APC model suffers though it is equivalent to APC model in terms of space spanned by column vectors in its design matrix. Further the results of fitting APE model to the data on consumer preference obtained from JNN Data Bank conducted by Japanese 28 TV stations including Tokyo Broadcasting System, Inc. (TBS) as their key station are shown. It is figured out, from those results, how Japanese food culture has been changed in the past several decades with respect to age, period and environmental (social) factors.