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 $N$ people who are in the $[A_{i-1}, A_i)$ age group at the time ($year$) $P_j$, where $i = 1, \ldots, I$; $j = 1, \ldots, J$ and $A_{i-1} - P_{j-1} = \tau$. Let $Y_{ij}^{(n)}$ ($n = 1, \ldots, N$) be a random variable which indicates whether the $ij$th member of $N$ 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 $Y_{ij}^{(n)}$:

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

(2.1)

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

Generally in the generalized linear model it is needed that, for every set of $\eta_{ij}$ 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 $\eta_{ij}$. As for the age-period-cohort model, it is seen that, for every set of $\eta_{ij}$ satisfying the model (2.1),

$$
\eta_{ij} = \mu^* + \alpha_i^* + \beta_j^* + \gamma_{j-i+1}^*,
$$

(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 $\alpha_i = \beta_j = \gamma_{j-i+1} = 0$, which are equivalent to $a = \alpha_i$, $b = \beta_j$ and $c = \gamma_{j-i+1}$, 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 $\beta_j = \beta_{j+1}$ or $\gamma_{j-i+1} = \gamma_{j-i+2}$ in addition to $\alpha_i = \beta_j = \gamma_{j-i+1} = 0$. Though the values of parameters are uniquely determined under one of those constraints in addition to $\alpha_i = \beta_j = \gamma_{j-i+1} = 0$, 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:

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

(2.3)

where $\mu$, $\alpha_i$, and $\beta_j$ are common with ones in the original age-period-cohort model and $\xi_k$ indicates the effect associated with the environment at the year $P_k$ ($k = 2 - I, \ldots, J$) which comes out $\tau$ 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 \([A_0, A_t]\) 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

\[ \eta = X \theta \]  

where \( \eta = (\eta_{ij1}, \eta_{ij2}, \cdots, \eta_{ijM})' \), \( X = (x_{i1}, x_{i2}, \cdots, x_{iM})' \), \( x_{ij} = (1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \) and \( \theta = (\mu, \alpha_1, \cdots, \alpha_j, \beta_1, \cdots, \beta_j, \xi_{j-1}, \cdots, \xi_{j-1})' \), where \( \beta_k \) and \( \alpha_k \) are \( k \) dimensional row vectors and \( \beta_k = (0, 0, \cdots, 0) \) and \( \alpha_k = (1, 1, \cdots, 1) \); \( k = 1, 2, \cdots \). In a special case that \( I = 5 \) and \( J = 3 \) the model \( \eta = X \theta \) concretely expressed like:

\[
\begin{pmatrix}
\eta_{i1} \\
\eta_{i2} \\
\eta_{i3} \\
\eta_{i4} \\
\eta_{i5}
\end{pmatrix}
= \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\mu \\
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\xi_{j-3} \\
\xi_{j-2} \\
\xi_{j-1} \\
\xi_0 \\
\xi_1 \\
\xi_2
\end{pmatrix}.
\]

As seen from the above formula (2.5) the values of \( \theta \) are uniquely determined under one of those constraints in addition to the constraints \( \alpha_i = \beta_i = \xi_{j-1} = 0 \).

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16-19</td>
<td>56.95</td>
<td>64.3</td>
<td>75.25</td>
<td>69.9</td>
<td>68.55</td>
<td>67.4</td>
<td>69.85</td>
</tr>
<tr>
<td>20-24</td>
<td>56.3</td>
<td>53.1</td>
<td>55</td>
<td>50.7</td>
<td>59.3</td>
<td>56</td>
<td>59.6</td>
</tr>
<tr>
<td>25-29</td>
<td>36.4</td>
<td>45.4</td>
<td>56.7</td>
<td>55</td>
<td>49.3</td>
<td>54</td>
<td>57.7</td>
</tr>
<tr>
<td>30-34</td>
<td>23.7</td>
<td>35.2</td>
<td>47.9</td>
<td>52.2</td>
<td>54.7</td>
<td>57.7</td>
<td>53.7</td>
</tr>
<tr>
<td>35-39</td>
<td>19.4</td>
<td>25.7</td>
<td>38</td>
<td>51.2</td>
<td>54.8</td>
<td>52.4</td>
<td>50.8</td>
</tr>
<tr>
<td>40-44</td>
<td>14.2</td>
<td>17.8</td>
<td>33.2</td>
<td>46.1</td>
<td>52.5</td>
<td>47.1</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>15.3</td>
<td>16.8</td>
<td>18.8</td>
<td>26.4</td>
<td>32.4</td>
<td>49.4</td>
<td>48.2</td>
</tr>
<tr>
<td>50-54</td>
<td>17.8</td>
<td>14.7</td>
<td>19.6</td>
<td>13.9</td>
<td>21.2</td>
<td>36.8</td>
<td>49</td>
</tr>
<tr>
<td>55-59</td>
<td>11.4</td>
<td>13.1</td>
<td>13.8</td>
<td>19.2</td>
<td>18.4</td>
<td>24.4</td>
<td>36.9</td>
</tr>
</tbody>
</table>

Figure 1. Response variables on the Lexis diagram

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


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.