Ann. Inst. Statist. Math.
Vol. 40, No. 3, 451–465 (1988)

A CLASS OF ESTIMABLE CONTRASTS IN AN AGE-PERIOD-COHORT MODEL

C. HIROTSU

Faculty of Engineering, University of Tokyo, 7-3-1 Hongō, Bunkyō-ku, Tokyo 113, Japan

(Received February 24, 1987; revised December 1, 1987)

Abstract. A class of estimable contrasts is defined for cohort effects in an age-period-cohort model. It is useful for detecting a systematic change in cohort effects without suffering from a short term deviation. This together with the follow-up analysis of residuals will give a good insight into the data. Numerical examples are given to illustrate how the method applies.

Key words and phrases: Age-period-cohort model, age-period interaction effects, analysis of residuals, outlier type contrasts, second order differential type contrasts, slippage type contrasts.

1. Introduction

The age-period-cohort analysis has been a popular epidemiological tool since Frost (1939) employed it in his study of mortality rate from tuberculosis. The procedure he developed was mainly descriptive. While a plot provides generally a useful first step in analyzing data, it does not provide a simple summary of the results and what is worse it can be sometimes misleading when only estimable are the second order differences of ordered parameters as it is the case in the age-period-cohort model. Identifiability problems in the simultaneous estimation of all of the three factors have been discussed by many authors (see Fienberg and Mason (1979), for example). Holford (1983) proposes to concentrate discussions only on the estimate functions which are invariant as to the particular constraint applied. Fienberg and Mason (1985) and Heckman and Robb (1985) discuss new types of model specification beyond the usual three effects cohort model.

In Section 2 of the present paper we characterize some popular classes of contrasts in the context of the one-way analysis of variance model. In particular we propose a class of contrasts which works for detecting a more systematic change than Holford's curvature component aims at when only estimable are the second order differences of the ordered parameters. In Section 3 it is extended to cohort effects in the age-period-cohort model still under the classical linear model. In Section 4 a comparison of the standardized contrasts is given a likelihood interpretation. An extention to a generalized linear model is discussed in Section 5. In Section 6 the analysis of residuals after fitting the age-period-cohort model is described. The examples are given in Section 7 to illustrate some of the ideas.

2. A characterization of classes of contrasts in the one-way analysis of variance model

Suppose that we are given a one-way layout model

$$y_i = \mu_i + \varepsilon_i \qquad (i = 1,\dots,a)\,,$$

for a treatments, where the ε_i are the uncorrelated experimental errors with the same variance. Then some popular classes of contrasts are

(a) $\mu_i - \mu_{i+1} \quad (i = 1,\dots,a-1)\,,$

(b) $\mu_i - \bar{\mu} \quad (i = 1,\dots,a)\,,$

(c) $\bar{\mu}_{(i)} - \bar{\mu}_{(\bar{i})} \quad (i = 1,\dots,a-1)\,,$

where $\bar{\mu}$ denotes the overall average of the μ_i's and $\bar{\mu}_{(i)}$ and $\bar{\mu}_{(\bar{i})}$ the averages from $j = 1$ to i and from $j = i+1$ to a, respectively. We shall call (a), (b) and (c) as differential, outlier and slippage type contrasts, respectively. The differential and the slippage type contrasts are directional, namely the former cuts off a systematic trend (low cut) and the other cuts off a short term deviation (low pass), whereas the outlier type is omnibus.

If the interests are concentrated in the second order differences which are orthogonal to the linear trend in μ_i's, the differential and the outlier type contrasts may be easily extended to (a′) $\mu_i - 2\mu_{i+1} + \mu_{i+2}$ $(i = 1,\dots,a-2)$ and (b′)$\{I - \Pi(B_a)\}\boldsymbol{\mu}$, where $\boldsymbol{\mu} = (\mu_1,\dots,\mu_a)'$, I an identity matrix and $\Pi(B_a) = B_a(B_a' B_a)^{-1}B_a'$ an orthogonal projector onto the column space of $B_a = \begin{bmatrix} 1 & 1 \cdots 1 \\ 1 & 2 \cdots a \end{bmatrix}'$. The subscript a of B_a will be omitted when there is no anxiety of confusion. The vector $\{I - \Pi(B)\}\boldsymbol{\mu}$ has been called a curvature component by Holford (1983) in the context of the age-period-cohort analysis.

To extend the slippage type contrasts we interpret the sample versions of the three types of the first order contrasts as estimating the differences $\delta_i = \mu_i - \mu_{i+1}$ $(i = 1,\dots,a-1)$ under the respective restrictions. First, without any restriction imposed on $\boldsymbol{\mu}$, the least squares estimator should be $\hat{\delta}_{di} = y_i - y_{i+1}$ $(i = 1,\dots,a-1)$. Assuming that the μ_j's are homogeneous excepting the μ_i, namely assuming

$$\mu_1 = \cdots = \mu_{i-1} = \mu_{i+1} = \cdots = \mu_a \neq \mu_i \,, \tag{2.1}$$

we obtain the least squares estimator $\hat{\delta}_{oi} = \{a/(a-1)\}(y_i - \bar{y})$. Note that the assumption (2.1) is equivalent to the assumption that $\boldsymbol{\mu}$ can be expressed as $\boldsymbol{\mu}_{(i)} = (\boldsymbol{j}\ \boldsymbol{e}_i)\boldsymbol{\eta}$ with some regression coefficient $\boldsymbol{\eta} = (\eta_0, \eta_1)'$, where $\boldsymbol{j}$ is the unit vector and $\boldsymbol{e}_i$ the vector with the i-th element unity and all the other elements zero.

If the assumption

$$\mu_1 = \cdots = \mu_i \neq \mu_{i+1} = \cdots = \mu_a \tag{2.2}$$

is made, then we obtain the least squares estimator $\hat{\delta}_{si} = \bar{y}_{(i)} - \bar{y}_{(\bar{i})}$. The assumption can be expressed again in the form $\boldsymbol{\mu}_{(i)} = (\boldsymbol{j}\ \boldsymbol{b}_i)\boldsymbol{\eta}$ with $\boldsymbol{b}_i$ the vector with zeros as its first i elements and unities for the rest.

Now we proceed to the second order differences $\gamma_i = \boldsymbol{l}_i'\boldsymbol{\mu} = \mu_i - 2\mu_{i+1} + \mu_{i+2}$ $(i = 1,\dots, a-2)$. Consider a model $\boldsymbol{\mu}_{(i)} = (B\ \boldsymbol{c}_{ai})(\eta_0, \eta_1, \eta_2)'$ with $\boldsymbol{c}_{ai} = (0,\dots,0,\ 1,\ 2,\dots, a-i-1)'$, which implies a slope change at the $(i+1)$-th point, and obtain the least squares estimator of the i-th second order difference $\hat{\gamma}_{si} = \boldsymbol{l}_i'\hat{\boldsymbol{\mu}}_{(i)} = \boldsymbol{l}_i'\,(B\ \boldsymbol{c}_{ai})\{(B\ \boldsymbol{c}_{ai})'(B\ \boldsymbol{c}_{ai})\}^{-1}(B\ \boldsymbol{c}_{ai})'\boldsymbol{y}$. If this is done for $i = 1,\dots, a-2$, then the vector $\hat{\boldsymbol{\gamma}}_s = (\hat{\gamma}_{s1},\dots,\hat{\gamma}_{sa-2})'$ can be written as

$$\hat{\boldsymbol{\gamma}}_s = L_a' \sum_{i=1}^{a-2} (B\ \boldsymbol{c}_i)\{(B\ \boldsymbol{c}_i)'(B\ \boldsymbol{c}_i)\}^{-1}\,(B\ \boldsymbol{c}_i)'\boldsymbol{y} \,, \tag{2.3}$$

by virtue of the relation $\boldsymbol{l}_k'(B\ \boldsymbol{c}_i) = (0, 0, \delta_{i,k})$, where $L_a = (\boldsymbol{l}_1,\dots, \boldsymbol{l}_{a-2})$, $\delta_{i,k}$ the Kronecker delta and the subscript a of $\boldsymbol{c}_{ai}$ omitted. As shown in Appendix 1 the $\hat{\boldsymbol{\gamma}}_s$ can be transformed into a very simple form

$$\hat{\boldsymbol{\gamma}}_s = \Gamma_s(L_a' L_a)^{-1} L_a' \boldsymbol{y} \,, \tag{2.4}$$

with Γ_s a diagonal matrix given in (A.1) of Appendix 1. This should be a natural extension of the slippage type first order contrasts $\hat{\boldsymbol{\delta}}_s = (\hat{\delta}_{s1},\dots, \hat{\delta}_{sa-1})'$ since the latter can be expressed as $\hat{\boldsymbol{\delta}}_s = \operatorname{diag}[a/\{i(a-i)\}](D_a' D_a)^{-1} D_a'\boldsymbol{y}$, where $D_a = (\boldsymbol{d}_1 \cdots \boldsymbol{d}_{a-1})$ with $\boldsymbol{d}_i = (0,\dots,0,\ 1,-1,\ 0,\dots,0)'$ being the vector yielding the i-th first order difference δ_i (see Appendix 2). The vector $\hat{\boldsymbol{\gamma}}_s$ has been introduced in Hirotsu (1986) to form a statistic testing for linearity against convexity and an explicit form of the matrix $(L_a' L_a)^{-1} L_a'$ is given there.

These interpretations lead to quite different forms of contrasts from those obtained above if a general covariance matrix Ω is assumed for the experimental error. The general case is required in the application of these contrasts to the three effects cohort model.

Assuming Ω is known, the least squares estimator of $\delta_i = \boldsymbol{d}_i'\boldsymbol{\mu}$ for the model (2.1) is obtained as

$$(2.5)\qquad \hat{\delta}_{oi} = \boldsymbol{d}_i'(\boldsymbol{j}\ \boldsymbol{e}_i)\{(\boldsymbol{j}\ \boldsymbol{e}_i)'\Omega^{-1}(\boldsymbol{j}\ \boldsymbol{e}_i)\}^{-1}(\boldsymbol{j}\ \boldsymbol{e}_i)'\Omega^{-1}\boldsymbol{y}\,.$$

After some calculations similar to those for obtaining the cross validation statistic in a multiple regression analysis we get an expression $\hat{\delta}_{oi} = \{\boldsymbol{e}_i'\Omega^{-1}\boldsymbol{e}_i - (\boldsymbol{j}'\Omega^{-1}\boldsymbol{e}_i)^2/(\boldsymbol{j}'\Omega^{-1}\boldsymbol{j})\}^{-1}\boldsymbol{e}_i'\Omega^{-1}[\boldsymbol{y} - \boldsymbol{j}\{\boldsymbol{j}'\Omega^{-1}\boldsymbol{y}/(\boldsymbol{j}'\Omega^{-1}\boldsymbol{j})\}]$. This shows that $\hat{\boldsymbol{\delta}}_o = (\hat{\delta}_{o1},\ldots,\hat{\delta}_{oa-1})'$ is simply given by

$$(2.6)\qquad \hat{\boldsymbol{\delta}}_o = \Delta_o(\Omega)(I\ \ 0)\Omega^{-1}\{\boldsymbol{y} - \hat{E}_0(\boldsymbol{y})\}\,,$$

where $\Delta_o(\Omega)$ is a diagonal matrix with $\{\boldsymbol{e}_i'\Omega^{-1}\boldsymbol{e}_i - (\boldsymbol{j}'\Omega^{-1}\boldsymbol{e}_i)^2/(\boldsymbol{j}'\Omega^{-1}\boldsymbol{j})\}^{-1}$ as its i-th diagonal element and $\hat{E}_0(\boldsymbol{y}) = \boldsymbol{j}\{\boldsymbol{j}'\Omega^{-1}\boldsymbol{y}/(\boldsymbol{j}'\Omega^{-1}\boldsymbol{j})\}$ the least squares estimator of the mean vector under the null model $E_0(\boldsymbol{y}) = \boldsymbol{j}\eta_0$. Unless $\Omega = I$, this generally differs from $\hat{\boldsymbol{\delta}} = \{a/(a-1)\}(\boldsymbol{y} - \boldsymbol{j}\bar{y})$, which is estimating the parametric functions $\mu_i - \bar{\mu}_{-i}$ $(i = 1,\ldots,a)$, $\bar{\mu}_{-i}$ being the average of the μ_j's excepting μ_i. The function $\mu_i - \bar{\mu}_{-i}$ itself should not be so interesting without such an assumption as (2.1).

Assuming (2.2), the least squares estimator $\hat{\delta}_{si}$ of the δ_i is formally given by (2.5) with $\boldsymbol{e}_i$ replaced by $\boldsymbol{b}_i$. In this case $\hat{\boldsymbol{\delta}}_s = (\hat{\delta}_{s1},\ldots,\hat{\delta}_{sa-1})$ can be written as

$$(2.7)\qquad \hat{\boldsymbol{\delta}}_s = D_a'\sum_{i=1}^{a-1}(\boldsymbol{j}\ \boldsymbol{b}_i)\{(\boldsymbol{j}\ \boldsymbol{b}_i)'\Omega^{-1}(\boldsymbol{j}\ \boldsymbol{b}_i)\}^{-1}(\boldsymbol{j}\ \boldsymbol{b}_i)'\Omega^{-1}\boldsymbol{y}\,,$$

by virtue of $\boldsymbol{d}_k'(\boldsymbol{j}\ \boldsymbol{b}_i) = (0,\ -\delta_{i,k})$. Again we have a simple expression

$$(2.8)\qquad \hat{\boldsymbol{\delta}}_s = \Delta_s(\Omega)(D_a'D_a)^{-1}D_a'\Omega^{-1}\{\boldsymbol{y} - \hat{E}_0(\boldsymbol{y})\}\,,$$

with $\Delta_s(\Omega)$ a diagonal matrix given in (A.2), the derivation of which is deferred to Appendix 3. The i-th component of the $\hat{\boldsymbol{\delta}}_s$ of course reduces to $\bar{y}_{(i)} - \bar{y}_{(\bar{i})}$ when $\Omega = \sigma^2 I$.

For the sample version of the second order outlier type contrasts in a general covariance matrix case we get, after some calculations, an expression similar to (2.6) as $\hat{\boldsymbol{\gamma}}_o = -2\Gamma_o(\Omega)(I\ \ 0)\Omega^{-1}\{\boldsymbol{y} - \hat{E}_B(\boldsymbol{y})\}$, where $\Gamma_o(\Omega)$ is the diagonal matrix of $\{\boldsymbol{e}_{i+1}'\Omega^{-1}\boldsymbol{e}_{i+1} - \boldsymbol{e}_{i+1}'\Omega^{-1}B(B'\Omega^{-1}B)^{-1}B'\Omega^{-1}\boldsymbol{e}_{i+1}\}^{-1}$ $(i = 1,\ldots, a-2)$ and $\hat{E}_B(\boldsymbol{y}) = B(B'\Omega^{-1}B)^{-1}B'\Omega^{-1}\boldsymbol{y}$ the least squares estimator of the mean vector under the null model $E_B(\boldsymbol{y}) = B(\eta_0, \eta_1)'$. The $\hat{\boldsymbol{\gamma}}_o$ coincides with Holford's curvature component when $\Omega = \sigma^2 I$.

Finally we can extend the result (2.4) of the slope change model to the general covariance matrix case as

(2.9) $$\hat{\gamma}_s = \Gamma_s(\Omega)(L_a' L_a)^{-1} L_a' \Omega^{-1}\{y - \hat{E}_B(y)\},$$

with $\Gamma_s(\Omega)$ a diagonal matrix of $\{c_i'\Omega^{-1}c_i - c_i'\Omega^{-1}B(B'\Omega^{-1}B)^{-1}B'\Omega^{-1}c_i\}^{-1}$ $(i = 1,\dots, a-2)$. The derivation of (2.9) is very similar to that of (2.8) and so is omitted.

Now the characteristics of the three types of the second order contrasts are obvious. The naive estimator $\hat{\gamma}_d$ is low cut and should be too noisy to interpret. The outlier type contrasts $\hat{\gamma}_o$ will be more stable but are detecting only a projecting change. The slippage type contrasts $\hat{\gamma}_s$ are low pass and trace a very systematic change in $\boldsymbol{\mu}$ without suffering from a short term derivation. This follows because each element of the $\hat{\gamma}_s$ is the least squares estimator of the corresponding element of γ with a low dimensional model assumed in $\boldsymbol{\mu}$. The high correlation between the subsequent elements of $\hat{\gamma}_s$ suggests that only one or two separate peaks should be given a practical interpretation and large terms shortly before or after the peak need not necessarily be given another interpretation than that given to the peak. This idea is partly supported by the result that the degrees of freedom of the approximated chi-squared distribution remains around two when a goes to infinity for the sum of squares of the slippage type contrasts whereas that for the outlier type goes to infinity with a (Hirotsu (1986)).

3. Three types of contrasts in cohort effects

Suppose that we have a three effects cohort model

(3.1) $$y_{ij} = \alpha_i + \beta_j + \mu_{i+j-1} + \varepsilon_{ij} \quad (i = 1,\dots, a;\, j = 1,\dots, b),$$

where α_i denotes the i-th age effect in antichronological order, β_j the j-th period effect and μ_k the k-th cohort effect. Since the cohort effects belong to the interaction space of age and period, only the contrasts $\mu_k - 2\mu_{k+1} + \mu_{k+2}$ $(k = 1,\dots, a+b-3)$ are estimable as are all linear combinations of them. Note that for a usual two-way interaction model only estimable interactions are the $(\alpha\beta)_{ij} - (\alpha\beta)_{i+1\,j} - (\alpha\beta)_{i\,j+1} + (\alpha\beta)_{i+1\,j+1}$ $(i = 1,\dots, a-1;$ $j = 1,\dots, b-1)$ and all the linear combinations of them. Without loss of generality we can put $\alpha_{a-1} = \alpha_a = \beta_b = 0$ in (3.1) and express the model in matrix form as

(3.2) $$y = [X_0 \ X_1](\boldsymbol{\theta}' \ \boldsymbol{\mu}')' + \varepsilon,$$

where $\boldsymbol{\theta} = (\alpha_1,\dots, \alpha_{a-2}, \beta_1,\dots, \beta_{b-1})'$, $\boldsymbol{\mu} = (\mu_1,\dots, \mu_{a+b-1})'$ and the coefficient matrices X_0 and X_1 are of full ranks.

The naive estimator of the second order contrast $\gamma = L_{a+b-1}'\boldsymbol{\mu}$ is obtained as $L_{a+b-1}'\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}} = \{X_1'\Omega^{-1}X_1 - X_1'\Omega^{-1}X_0(X_0'\Omega^{-1}X_0)^{-1}X_0'\Omega^{-1}X_1\}^{-1}$ $\cdot\{X_1' - X_1'\Omega^{-1}X_0(X_0'\Omega^{-1}X_0)^{-1}X_0'\}\Omega^{-1}y$ is the least squares estimator of $\boldsymbol{\mu}$.

The outlier type contrasts are obtained assuming a regression model $\boldsymbol{\mu} = \boldsymbol{\mu}_{(i)} = (B_{a+b-1}\ \ e_{i+1})(\eta_1, \eta_2, \eta_3)$ in (3.2) and forming $\hat{\gamma}_{oi} = \boldsymbol{l}_i'\hat{\boldsymbol{\mu}}_{(i)} = \boldsymbol{l}_i'(B\ \ e_{i+1})\hat{\boldsymbol{\eta}}$ $(i = 1, \ldots, a + b - 3)$, where $\hat{\boldsymbol{\eta}}$ is the least squares estimator of $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)'$. More conveniently it is equivalent to applying the formula of Section 2 for an assumed one-way layout model $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}^*$ with the covariance matrix

$$\mathrm{var}\,(\boldsymbol{\varepsilon}^*) = \Omega^* = \{X_1'\Omega^{-1}X_1 - X_1'\Omega^{-1}X_0(X_0'\Omega^{-1}X_0)^{-1}X_0'\Omega^{-1}X_1\}^{-1}\,.$$

Thus we get $\hat{\gamma}_o = -\,2\Gamma_o(\Omega^*)(I\ \ 0)\Omega^{*-1}\{\hat{\boldsymbol{\mu}} - \hat{E}_B(\hat{\boldsymbol{\mu}})\}$, where $\hat{E}_B(\hat{\boldsymbol{\mu}}) = B(B'\Omega^{*-1}B)^{-1}\cdot B'\Omega^{*-1}\hat{\boldsymbol{\mu}}$. In this case the $\hat{\gamma}_o$ differs from Holford's curvature component even when $\Omega = \sigma^2 I$ since then Ω^* is not a multiple of an identity matrix.

Similarly we obtain the slippage type contrasts from (2.9) as $\hat{\gamma}_s = \Gamma_s(\Omega^*)(L_{a+b-1}'L_{a+b-1})^{-1}L_{a+b-1}'\Omega^{*-1}\{\hat{\boldsymbol{\mu}} - \hat{E}_B(\hat{\boldsymbol{\mu}})\}$.

4. Comparing the standardized contrasts

In a cohort analysis it is often desired to compare two models $E(\hat{\boldsymbol{\mu}}) = (B\ \ \boldsymbol{e}_{i+1})\boldsymbol{\eta}$ and $E(\hat{\boldsymbol{\mu}}) = (B\ \ \boldsymbol{c}_i)\boldsymbol{\eta}$. The former implies an outlier at the $(i + 1)$-th cohort and the other a slope change at the $(i + 1)$-th cohort. For the purpose the standardized contrasts are more appropriate than the raw estimators $\hat{\gamma}_{oi}$ and $\hat{\gamma}_{si}$.

Suppose that $\hat{\boldsymbol{\mu}}$ is distributed as a multivariate normal with mean $(B\ \ \boldsymbol{e}_{i+1})\boldsymbol{\eta}$ and known covariance matrix Ω^*. Then the minus two times log likelihood evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\eta}}$ is, after some calculations, obtained as

$$-\,2\log(L_o) = c + \{\boldsymbol{e}_{i+1}'\Omega^{*-1}\Pi^{\perp}(B, \Omega^*)\hat{\boldsymbol{\mu}}\}^2/\ \mathrm{var}\,\{\boldsymbol{e}_{i+1}'\Omega^{*-1}\Pi^{\perp}(B, \Omega^*)\hat{\boldsymbol{\mu}}\}\,,$$

where

$$c = \log|2\pi\Omega^*| + \hat{\boldsymbol{\mu}}'\Omega^{*-1}B(B'\Omega^{*-1}B)^{-1}B'\Omega^{*-1}\hat{\boldsymbol{\mu}}$$

and

$$\Pi^{\perp}(B, \Omega^*) = I - B(B'\Omega^{*-1}B)^{-1}B'\Omega^{*-1}\,.$$

Similarly assuming $E(\hat{\boldsymbol{\mu}}) = (B\ \ \boldsymbol{c}_i)\boldsymbol{\eta}$, we obtain $-\,2\log(L_s) = c + \{\boldsymbol{c}_i'\Omega^{*-1}\cdot \Pi^{\perp}(B, \Omega^*)\hat{\boldsymbol{\mu}}\}^2/\ \mathrm{var}\,\{\boldsymbol{c}_i'\Omega^{*-1}\Pi^{\perp}(B, \Omega^*)\hat{\boldsymbol{\mu}}\}$. Now it is simple algebra to show that $F = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{a+b-3})'\Omega^{*-1}\Pi^{\perp}(B, \Omega^*)$ is equal to $(L_{a+b-1}'\Omega^{*-1}L_{a+b-1})^{-1}L_{a+b-1}'$, which gives in turn a simpler form (2.9) of the $\hat{\gamma}_s$. Thus comparing likelihoods from the outlier and the slope change models reduces to comparing the standardized components of $\hat{\gamma}_s$ and $\hat{\gamma}_o$. It should be noted that the nonestimable component cancels out in both models.

5. Generalized linear models

We can extend the previous results easily to the generalized linear models. Suppose, for example, that the y_{ij} in Section 3 are distributed as independent Poisson variables with the mean m_{ij}. If the person year T_{ij} are known, a log linear model

$$\log(m_{ij}/T_{ij}) = \alpha_i + \beta_j + \mu_{i+j-1} \quad (i = 1,\dots,a; j = 1,\dots,b)\,, \tag{5.1}$$

is often assumed (see for example, Holford (1983)). We again employ the identifiability conditions $\alpha_{a-1} = \alpha_a = \beta_b = 0$ and express (5.1) in a matrix form as

$$\log(\boldsymbol{m}) = \log(\boldsymbol{T}) + (X_0 \;\; X_1)(\boldsymbol{\theta}' \;\; \boldsymbol{\mu}')'\,, \tag{5.2}$$

where $\log(\boldsymbol{m})$ and $\log(\boldsymbol{T})$ denote the vector of $\log(m_{ij})$ and $\log(T_{ij})$, respectively. Then the naive estimator of the second order difference $\gamma = L'_{a+b-1}\boldsymbol{\mu}$ is obtained as $\hat{\gamma} = L'_{a+b-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ is the maximum likelihood estimator of $\boldsymbol{\mu}$ under the model (5.2).

The outlier type contrasts can be obtained assuming $\boldsymbol{\mu} = (B \;\; \boldsymbol{e}_{i+1})\boldsymbol{\eta}$ in (5.2). It is, however, easier to get the asymptotically equivalent solution by assuming a linear model $\hat{\boldsymbol{\mu}} = (B \;\; \boldsymbol{e}_{i+1})\boldsymbol{\eta} + \boldsymbol{\varepsilon}^*$, with $\mathrm{var}(\boldsymbol{\varepsilon}^*) = \Omega^* = \{X_1'\Omega^{-1}X_1 - X_1'\Omega^{-1}X_0(X_0'\Omega^{-1}X_0)^{-1}X_0'\Omega^{-1}X_1\}^{-1}$, where $\Omega = \mathrm{diag}(m_{ij}^{-1})$. The m_{ij} in Ω may be replaced by $\hat{m}_{ijB}$, the maximum likelihood estimator of m_{ij} under the null model $\log(m) = \log(T) + (X_0 \;\; X_1 B)(\boldsymbol{\theta}' \;\; \boldsymbol{\eta}')'$. The $\hat{m}_{ijB}$ can be obtained by an alternate scaling of the initial two-way table of T_{ij} by the marginal totals $y_{i\cdot}$ and $y_{\cdot j}$. Thus we obtain

$$\hat{\gamma}_o \propto (I \;\; 0)\Omega^{*-1}\{\hat{\boldsymbol{\mu}} - \hat{E}_B(\hat{\boldsymbol{\mu}})\}\,. \tag{5.3}$$

In (5.3) we dropped the diagonal coefficient matrix which should be cancelled out by the standardization.

Similarly we obtain the slippage type contrasts

$$\hat{\gamma}_s \propto (L'_{a+b-1}L_{a+b-1})^{-1}L'_{a+b-1}\Omega^{*-1}\{\hat{\boldsymbol{\mu}} - \hat{E}_B(\hat{\boldsymbol{\mu}})\}\,. \tag{5.4}$$

As stated in Hirotsu (1982) we can replace $\Omega^{*-1}\{\hat{\boldsymbol{\mu}} - \hat{E}_B(\hat{\boldsymbol{\mu}})\}$ in (5.3) and (5.4) by an asymptotically equivalent statistic $X_1'(\boldsymbol{y} - \hat{\boldsymbol{m}}_B)$, where $\hat{\boldsymbol{m}}_B$ is the vector of $\hat{m}_{ijB}$. This version will be convenient for applications. For the standardization we exploit the asymptotic variance $X_1'\hat{M}_B X_1 - X_1'\hat{M}_B(X_0 \;\; X_1 B) \cdot\{(X_0 \;\; X_1 B)'\hat{M}_B(X_0 \;\; X_1 B)\}^{-1}(X_0 \;\; X_1 B)'\hat{M}_B X_1$, where $\hat{M}_B = \mathrm{diag}(\hat{m}_{ijB})$.

6. Analysis of residuals

The cohort effects are in some sense very special interactions between age and period which remain constant throughout the ages and periods observed. They must therefore be formed before entering the cohort table. On the other hand we cannot deny a possibility of the age and period interaction in the usual sense. In this line various interaction models including a polynomial type and also Johnson and Graybill's (1972) type are discussed in Fienberg and Mason (1985) and Heckman and Robb (1985). An informal but useful procedure for detecting such an interaction is the analysis of residuals after fitting the age-period-cohort model. The procedure is explained only for the Poisson model of Section 5 since the extensions to the other models are straightforward.

Let $\hat{\boldsymbol{m}}$ be the maximum likelihood estimator under the age-period-cohort model and define the residual by $\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{m}}$. Then by the general theory for the multinomial distribution the asymptotic variance of $\boldsymbol{r}$ is consistently estimated by

$$\{v_{kl}\} = \hat{M} - \hat{M}(X_0\ X_1)\{(X_0\ X_1)'\hat{M}(X_0\ X_1)\}^{-1}(X_0\ X_1)'\hat{M}\ ,$$

where $\hat{M} = \mathrm{diag}\,(\hat{m}_{ij})$. The standardized residual is then defined by $\boldsymbol{r}^* = \mathrm{diag}\,(v_{kk}^{-1/2})\boldsymbol{r}$.

7. Examples

To illustrate some of the ideas we consider the data from Tango (1985) on liver cirrhosis mortality and suicides, both of males in Japan from 1955 to 1975.

Example 7.1. Liver cirrhosis mortality: The data presented in Table 1 give the number of liver cirrhosis deaths. The estimated person years are given in Table 2.

The standardized versions of three types of contrasts from a Poisson model are shown in Table 3. In the slippage type contrasts there is observed a large and systematic pattern, namely a long term convexity followed by a short term concavity, whereas no systematic pattern is detected by the naive or the outlier type contrasts. In particular the naive estimator is seen to be too noisy to interpret. The large outlier type contrasts occurred at the upper extreme cohorts are somewhat unreliable for the small population sizes and should not be overly interpreted. A large component observed at the 11th cohort, which corresponds to the birth at the early Showa era (1925–1935), might well be interpreted as a turning point from the long term convexity to the short term concavity. In interpretation it should be noted that the convexity may mean deceleration of

Table 1. Number of liver cirrhosis deaths of males in Japan.

Age	Period				
	1955–59	1960–64	1965–69	1970–74	1975–79
75–79	2520	2848	3119	3648	4391
70–74	3280	3843	3966	5129	5493
65–69	3615	3995	4776	5728	6400
60–64	3364	4101	4824	6214	6378
55–59	3195	3805	4763	5619	6056
50–54	2720	3225	3773	4728	7160
45–49	1977	2349	2800	5012	8218
40–44	1230	1463	2522	5002	5832
35–39	646	1011	2077	2962	2426
30–34	478	670	890	986	826
25–29	258	279	266	260	218
20–24	159	131	116	98	62

Table 2. Estimated person years ($\times$ 5000).

Age	Period				
	1955–59	1960–64	1965–69	1970–74	1975–79
80–84	148	176	209	268	352
75–79	356	407	484	594	752
70–74	634	732	858	1036	1216
65–69	962	1104	1291	1468	1640
60–64	1311	1513	1677	1828	1940
55–59	1685	1853	1975	2054	2248
50–54	1974	2094	2167	2341	2989
45–49	2184	2244	2406	3070	3807
40–44	2305	2456	3104	3850	4139
35–39	2497	3157	3897	4158	4365
30–34	3177	3907	4161	4359	4943
25–29	3903	4120	4301	4881	5074
20–24	4168	4274	4823	5013	4322
15–19	4476	4998	5116	4359	4114

the downward trend or the acceleration of the upward trend and that those two patterns are undistinguishable since a linear trend is nonestimable.

The standardized residuals after fitting the age-period-cohort model are given in Table 4. Zero entries at the cells $(1, 1)$ and $(12, 5)$ are due to the fact that those cells are estimated by the respective observed cell frequencies only. As compared with the normal theory cases there are observed somewhat large elements probably due to the over-dispersion often experienced in those epidemiological researches. There appears, however, no systematic pattern. The Pearson's chi-squared value 91.43 for residuals is somewhat large at 30 degrees of freedom but small enough as compared with the overall chi-squared value 4226 due to the cohort effects with the degrees of freedom 14, or in other words the latter is highly significant as

Table 3. Standardized contrasts for cohort effects.

Change point	Differential	Outlier	Slippage
2	− 0.23	− 21.62	16.42
3	− 1.74	− 19.20	28.84
4	6.37	− 0.91	40.41
5	− 0.03	9.24	46.18
6	5.85	14.82	49.91
7	− 1.33	9.46	49.66
8	0.12	12.43	48.59
9	9.93	17.51	43.91
10	6.04	− 0.12	28.24
11	− 8.53	− 29.45	3.26
12	− 10.28	− 26.80	− 14.71
13	− 2.23	− 2.84	− 17.45
14	0.05	8.71	− 12.83
15	− 0.65	8.94	− 7.60

Table 4. Standardized residuals.

Age	Period				
	1955–59	1960–64	1965–69	1970–74	1975–79
75–79	0	1.20	− 2.32	− 1.51	2.18
70–74	− 1.24	3.62	− 0.12	− 0.69	− 0.67
65–69	− 1.06	2.86	1.09	− 1.01	− 1.32
60–64	− 0.89	− 0.67	0.92	0.99	− 0.56
55–59	− 2.39	− 0.29	2.03	0.12	0.45
50–54	1.38	− 0.82	− 0.27	− 0.28	− 0.04
45–49	− 0.56	0.61	0.02	0.53	− 0.80
40–44	0.10	− 0.48	− 0.27	1.27	− 0.43
35–39	0.38	− 2.80	0.32	3.10	0.24
30–34	4.67	− 3.00	− 4.20	0.30	1.22
25–29	3.22	− 1.18	− 1.30	− 1.19	1.03
20–24	3.09	0.39	− 0.47	− 1.17	0

compared with the former. It should be also noted that the largest component 49.91 in Table 3 explains nearly 59% of the overall cohort effects by only a single element.

Example 7.2. Suicides: The data presented in Table 5 give the number of suicides. Standardized contrasts are shown in Table 6. Again a long term convexity is observed in the slippage type contrasts, in particular a strong acceleration of the upward trend, or deceleration of the downward trend, around the 13th cohort. The cohort might look outlying by the outlier type contrasts. Both interpretations, however, turn out to be misleading by the inspection of the residuals given in Table 7. There are observed some large deviations. In particular along with the 13th cohort there are one particularly large positive element and three moderately large

Table 5. Number of suicides of males in Japan.

Age	Period				
	1955–59	1960–64	1965–69	1970–74	1975–79
80–84	940	3962	2393	2372	2511
75–79	1916	8732	4822	5978	5539
70–74	2591	7136	4962	4299	7028
65–69	3280	4026	3753	4487	5726
60–64	3541	2692	3138	4387	5632
55–59	4033	1979	2422	3964	6083
50–54	3415	2308	2105	3260	5807
45–49	3176	2765	2444	2636	4369
40–44	2552	3089	3003	3018	3103
35–39	2800	3349	3134	3270	3250
30–34	4443	3004	3115	3278	3256
25–29	10029	2388	2429	2992	3083
20–24	16033	1670	1731	2168	2642
15–19	7054	890	937	1222	1495

Table 6. Standardized contrasts for cohort effects.

Change point	Differential	Outlier	Slippage
2	− 1.21	4.77	− 5.58
3	0.17	3.24	− 7.92
4	− 2.10	− 1.41	− 9.34
5	2.08	− 1.72	− 8.94
6	− 5.37	− 7.21	− 8.29
7	3.19	3.64	− 5.38
8	− 0.30	8.86	− 4.67
9	1.69	17.62	− 7.20
10	8.41	23.86	− 15.24
11	2.87	8.22	− 30.45
12	− 3.39	− 20.40	− 48.96
13	− 19.60	− 52.74	− 63.33
14	− 13.63	− 31.07	− 61.12
15	20.32	25.82	− 43.08
16	− 7.93	19.61	− 29.49
17	3.55	27.31	− 11.36

negative elements. This suggests that the cohort effect should have been highly overestimated by the unduly large entry at the (13, 1) cell of Table 5 and that such an outlying value should not have been observed in Table 6 if a parameter was assigned to the (13, 1) cell. This sort of modeling is also suggested in Fienberg and Mason (1985). Further there is observed a clear age by period interaction pattern in Table 7, that is, relatively high observations in younger generations in periods 1955–1959 and 1975–1979, and in older generations in the period 1960–1974. Pearson's chi-squared statistic for residuals amounts to 2881 and makes the overall chi-squared statistic 6925 due to the cohort effects variations less significant. Tango

Table 7. Standardized residuals.

Age	Period				
	1955–59	1960–64	1965–69	1970–74	1975–79
80–84	0	2.83	6.44	2.70	− 8.90
75–79	− 2.55	3.40	7.26	5.29	− 10.99
70–74	− 6.76	3.39	11.26	6.07	− 10.31
65–69	− 9.39	6.12	13.61	7.70	− 13.00
60–64	− 11.26	3.20	14.49	6.79	− 9.07
55–59	− 8.87	1.67	13.79	7.26	− 10.39
50–54	− 9.17	2.15	9.02	3.35	− 3.07
45–49	− 4.53	0.73	4.93	1.53	− 1.79
40–44	− 3.77	0.29	2.80	0.37	2.41
35–39	1.63	− 1.86	− 4.78	− 5.95	10.57
30–34	0.49	− 11.89	− 16.90	− 2.59	25.55
25–29	15.81	− 10.51	− 11.09	− 11.82	17.26
20–24	27.77	− 0.27	− 12.12	− 6.34	− 0.41
15–19	6.28	7.18	− 12.38	0.47	0

also noted that the residual chi-squared can be drastically lessened by introducing 3 by 5 age-period interaction into the model, where three generations, younger, middle and older, are chosen somewhat arbitrarily. The residual analysis here suggests a 2 by 3 interaction model. The relation between this sort of block interaction models and the Johnson and Graybill's type model is discussed in Hirotsu (1983).

8. Discussion

The second order differences of data are taken usually for the purpose of low cut as to be seen in the nearest neighbour analysis, for example. On the contrary it is intended in the cohort analysis to detect a systematic change in the cohort effect when only estimable are the second order differences. For the purpose the slippage type and the outlier type contrasts may be preferred to the naive and noisy second order differences. The outlier type contrasts here differ from Holford's curvature component in assuming three dimensional models $\boldsymbol{\mu}_{(i)} = (B\ \boldsymbol{e}_{i+1})\boldsymbol{\eta}$ rather than forming linear combinations $\{I - \Pi(B)\}\hat{\boldsymbol{\mu}}$.

The model basis approach for the slippage type contrasts can be modified to assuming linear trends for some given numbers of cohorts before and after the change point, including the naive and the slippage types here as its two extremes. The applicability of this approach is now under investigation.

Although we focussed our attention on the cohort effects here, all the discussions can be applied to the age or the period effects almost as it is.

Acknowledgements

The author thanks Messrs H. Tsubaki and Y. Maruyama for their help in preparing numerical examples.

Appendix 1

Let H' be the coefficient matrix y in the equation (2.3). Then it is easy to verify the following two equations,

$$H' L_a = \Gamma_s = \mathrm{diag}\,[\{c_i' c_i - c_i' B(B'B)^{-1} B' c_i\}^{-1}]\,, \tag{A.1}$$

and $H'B = (a-2)L_a' B = 0$. Then we get

$$\begin{bmatrix} H' \\ B' \end{bmatrix} = \begin{bmatrix} \Gamma_s & 0 \\ 0 & B'B \end{bmatrix} \begin{bmatrix} L_a & B \end{bmatrix}^{-1} = \begin{bmatrix} \Gamma_s & 0 \\ 0 & B'B \end{bmatrix} \begin{bmatrix} (L_a' L_a)^{-1} L_a' \\ (B'B)^{-1} B' \end{bmatrix},$$

and therefore,

$$H' = \Gamma_s (L_a' L_a)^{-1} L_a' .$$

Appendix 2

It is easy to verify the equation

$$(j\ D_a)^{-1} = \begin{bmatrix} (j'j)^{-1} j' \\ (D_a' D_a)^{-1} D_a' \end{bmatrix},$$

and also

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ a-1 & -1 & \cdots & -1 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ a-i & a-i & \cdots & -i & -i \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & -(a-1) \end{bmatrix} (j\ D_a) = aI .$$

Then we get

$$(D_a' D_a)^{-1} D_a' = \mathrm{diag}\left\{ \frac{i(a-i)}{a} \right\} \begin{bmatrix} 1 & -(a-1)^{-1} & \ldots & -(a-1)^{-1} \\ & \vdots & & \\ i^{-1} & i^{-1} & \cdots & -(a-i)^{-1} \\ & \vdots & & \\ (a-1)^{-1} & (a-1)^{-1} & \cdots & -1 \end{bmatrix}.$$

Appendix 3

Derivation of the equation (2.8)

Let $P\Omega P' = \operatorname{diag}(\omega_i)$ be an orthogonal transformation of Ω, $\Omega^{1/2}$ be defined by $P' \operatorname{diag}(\omega_i^{1/2})$ and $\Omega^{-1/2}$ by $(\Omega^{1/2})^{-1}$. Then $\hat{\boldsymbol{\delta}}_s$ of (2.7) can be written as $\hat{\boldsymbol{\delta}}_s = C'\Omega^{-1/2}\boldsymbol{y}$ with

$$C' = D_a' \sum_{i=1}^{a-1} (\boldsymbol{j}\ \boldsymbol{b}_i)\{(\boldsymbol{j}\ \boldsymbol{b}_i)'\Omega^{-1}(\boldsymbol{j}\ \boldsymbol{b}_i)\}^{-1}(\boldsymbol{j}\ \boldsymbol{b}_i)'(\Omega^{-1/2})' .$$

It is easy to verify the following two equations

$$\text{(A.2)}\quad C'(\Omega^{1/2})' D_a = \Delta_s(\Omega) = \operatorname{diag}[\{b_i'\Omega^{-1}b_i - (j'\Omega^{-1}b_i)^2/(j'\Omega^{-1}j)\}^{-1}] ,$$

and

$$C'\Omega^{-1/2}\boldsymbol{j} = D_a'\Omega^{1/2}\sum_{i=1}^{a-1} \Pi_i\Omega^{-1/2}\boldsymbol{j} = (a-1)D_a'\Omega^{1/2}\Omega^{-1/2}\boldsymbol{j} = 0 ,$$

where Π_i is an orthogonal projector onto the column space of $\Omega^{-1/2}(\boldsymbol{j}\ \boldsymbol{b}_i)$ in each of which $\Omega^{-1/2}\boldsymbol{j}$ is contained. Thus we have

$$C'\{\Omega^{-1/2}\boldsymbol{j}\quad (\Omega^{1/2})' D_a\} = (0\quad \Delta_s(\Omega)) .$$

Noting the equality

$$\{\Omega^{-1/2}\boldsymbol{j}\quad (\Omega^{1/2})'D_a\}'\{\Omega^{-1/2}\boldsymbol{j}\quad (\Omega^{1/2})'D_a\} = \begin{bmatrix} \boldsymbol{j}'\Omega^{-1}\boldsymbol{j} & 0 \\ 0 & D_a'\Omega D_a \end{bmatrix},$$

we get $C' = \Delta_s(\Omega)(D_a'\Omega D_a)^{-1}D_a'\Omega^{1/2}$ and therefore $C'\Omega^{-1/2}\boldsymbol{y} = \Delta_s(\Omega)(D_a'\Omega D_a)^{-1} \cdot D_a'\boldsymbol{y}$. More convenient expression (2.8) follows by virtue of Lemma (2.1) of Hirotsu (1982).

REFERENCES

Fienberg, S. E. and Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology*, (ed. K. F. Schussler), 1–67, Jossey-Bass, San Francisco.

Fienberg, S. E. and Mason, W. M. (1985). Specification and implementation of age, period and cohort models, *Cohort Analysis in Social Research*, (eds. W. M. Mason and S. E. Fienberg), 45–88, Springer-Verlag, New York.

Frost, W. H. (1939). The age selection of mortality from tuberculosis in successive decade, *Amer. J. Hygiene A*, **30**, 91–96.

Heckman, J. and Robb, R. (1985). Using longitudinal data to estimating age, period and cohort effects in earnings equations, *Cohort Analysis in Social Research*, (eds. W. M. Mason and S. E. Fienberg), 137–150, Springer-Verlag, New York.

Hirotsu, C. (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models, *Biometrika*, **69**, 567–577.

Hirotsu, C. (1983). An approach to defining the pattern of interaction effects in a two-way layout, *Ann. Inst. Statist. Math.*, **35**, 77–90.

Hirotsu, C. (1986). Cumulative chi-squared statistic as a tool for testing goodness of fit, *Biometrika*, **73**, 165–173.

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

Johnson, D. E. and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication, *J. Amer. Statist. Assoc.*, **67**, 862–868.

Tango, T. (1985). Estimation of age, period and cohort effects—Decomposition into linear trend and curvature component—, *Japan J. Appl. Statist.*, **14**, 45–59 (in Japanese).