

# Deviance analysis of age-period-cohort models

B. NIELSEN<sup>1</sup>

3 September 2014

**Abstract:** Age-period-cohort models for Lexis diagrams are considered. The identification problem is addressed by reparametrizing the models in terms of freely varying parameters. The reparametrisation covers all three Lexis diagrams. A new plot of the unidentified time effects is suggested and interpreted. Deviance analysis covering a wide range of sub-models is presented. The methods are applied to Belgian lung cancer data.

**Keywords:** Age-period-cohort model, Canonical parametrization, Generalized Linear Models, Identification, Lexis diagrams

## 1 Introduction

Age-period-cohort models are extensively used in actuarial sciences, demography, epidemiology and social sciences. The model predictor combines time effects for age, period and cohort, but these time effects cannot be fully recovered from the predictor. The proposal of this paper is to focus on the parts of the time effects that actually can be recovered. This is done by reparametrising the predictor in terms of freely varying parameters along the lines of Kuang, Nielsen and Nielsen (2008a). The paper makes three contributions. First, a new reparametrisation is suggested that unifies analysis for the three Lexis diagrams of age-period data arrays, age-cohort data arrays and cohort-period data arrays. Secondly, a new suggestion is given for a graphical representation of the unidentified time effects. Thirdly, the parametrisation allows for the formulation of 14 sub-models of the age-period-cohort model including two-factor models such as the age-cohort model and the age-drift model and one-factor models such as an age-model. In applications, a deviance table gives a quick overview of the likelihood of these sub-models.

The linear age-period-cohort model has predictor of the form

$$\mu_{ik} = \alpha_i + \beta_j + \gamma_k + \delta, \quad (1.1)$$

where  $i, j, k$  are indices for age, period and cohort, which are linked so that  $j = i + k - 1$ . The classical identification problem of the model is that the levels and linear slopes of the individual age,  $\alpha_i$ , period,  $\beta_j$  and cohort time effects,  $\gamma_k$ , are not identifiable. Broadly speaking, there are two types of solutions: either to work with restricted versions of time effects  $\alpha_i, \beta_j, \gamma_k$  on the right hand side of the equation (1.1), or to find a more parsimonious parametrisation of the predictor  $\mu_{ik}$  on the left hand side of the equation (1.1) in terms of freely varying parameters. The first approach of working with the time effects requires an ad hoc identification of the time effects that is motivated by mathematical convenience and the substantive context. A number of methods exist, including the dynamic Bayesian method of Berzuini and Clayton (1994), the proposal of

---

<sup>1</sup>Nuffield College & Department of Economics, University of Oxford & Programme on Economic Modelling, INET, Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk.

Osmond and Gardner (1982), and the intrinsic estimator of Yang, Fu and Land (2004). With this approach it is often challenging to track how initial identification choices influence the inferences drawn from the analysis, see Holford (1985, 2006), Clayton and Schifflers (1987b), O'Brien (2011), Luo (2013), Nielsen and Nielsen (2014). The second approach is to reparametrise the age-period-cohort model in terms of freely varying parameters, see Kuang, Nielsen and Nielsen (2008a). This leads to a statistical model that is a regular exponential family and statistical analysis that conforms with standard statistical theory. From this viewpoint the formulation (1.1) is therefore a motivation or a symbolic description of the model in the tradition of Wilkinson and Rogers (1973).

Age-period-cohort data are collected in a variety of ways. Indeed, Keiding (1990) distinguishes between three principal types of Lexis diagrams: age-period rectangular arrays, age-cohort rectangular arrays, and period-cohort rectangular arrays. The first contribution is a unified parametrisation of (1.1), that is valid for all three Lexis diagrams. The solution addresses the difficulty that age-period-cohort analysis includes three time scales both in the parametrisation and in the structure of the data array. The idea is to embed the Lexis diagrams in a age-cohort coordinated system. This generalizes the solution for rectangular age-cohort arrays by Kuang, Nielsen and Nielsen (2008a).

The second contribution is a new graphical representation of the unidentified age, period and cohort time effects. The time effects are only identified up to an arbitrary linear trend. A graphical representation should emphasize the identifiable variation while suppressing the impact of the linear trend on the visual impression. The proposal is to choose the linear trends so that the time effects start and end in zero. This plot highlights the non-linear part of the time effects over and above the arbitrary linear effect. Interpretations of particular patterns are discussed.

The third contribution is a deviance table for the age-period-cohort model and its sub-models. The age-period-cohort model has three time effects, two linear trends and a constant level. Interesting sub-models arise by dropping one or two of the three time effects, by dropping two time effects and a linear trend, by dropping all three time effects, and by dropping all factors but the constant. A deviance table gives a quick overview of which factors are relevant, and which are not.

Throughout the analysis is illustrated using the Belgian lung cancer data of Clayton and Schifflers (1987a). It has a modest dimension and an interesting age-period-cohort structure that illustrates the ideas well. It is dose-response data organised in an age-period array with 11 age groups and 4 period groups. The data are modelled using a Poisson regression. The analysis is done using the R package `apc`, see Nielsen (2014).

The paper is organized as follows. Typical data arrays are discussed in §2. A dose-response Poisson model, the identifications problem and the reparametrization are presented in §3. The plot of the unidentified time effects follows in §4, while the deviance table is presented in §5. Variations of the distributional assumptions are discussed in §6. §7 concludes. Technical details are collected in an Appendix.

	1955-1959		1960-1964		1965-1969		1970-1974	
25-29	0.19	(3)	0.13	(2)	0.50	(7)	0.19	(3)
30-34	0.66	(11)	0.98	(16)	0.72	(11)	0.71	(10)
35-39	0.78	(11)	1.32	(22)	1.47	(24)	1.64	(25)
40-44	2.67	(36)	3.16	(44)	2.53	(42)	3.38	(53)
45-49	4.84	(77)	5.60	(74)	4.93	(68)	6.05	(99)
50-54	6.60	(106)	8.50	(131)	7.65	(99)	10.59	(142)
55-59	10.36	(157)	12.00	(184)	12.68	(189)	14.34	(180)
60-64	14.76	(193)	16.37	(232)	18.00	(262)	17.60	(249)
65-69	20.53	(219)	22.60	(267)	24.90	(323)	24.33	(325)
70-74	26.24	(223)	27.70	(250)	30.47	(308)	36.94	(412)
75-79	33.47	(198)	33.61	(214)	36.77	(253)	43.69	(338)

Table 1: Belgian lung cancer data. Rates with number of responses in paranthesis. Age groups in rows, Period groups in columns.

## 2 Data structure

Mortality tables typically have one of 3 standard formats. These are age-period arrays, age-cohort arrays and period-cohort arrays. In 1875 Lexis referred to these as the 3 principle sets of death, see Keiding (1990). Another common format are the reserving triangles from general insurance, which are triangular age-cohort arrays, see England and Verrall (2002). These data arrays are special cases of generalized trapezoid arrays.

### 2.1 Generalized trapezoid arrays

Lexis diagrams are unified in an age-cohort coordinate system. This is to exploit the symmetry that age  $i$  and cohort  $k$  add up to the period  $j$ , in that  $i + k = j + 1$ . The Lexis diagrams then become special cases of the generalized trapezoid data array with index set

$$\mathcal{I} = \{i, k : 1 \leq i \leq I, 1 \leq k \leq K, L + 1 \leq j \leq L + J\}, \quad (2.1)$$

where  $I$ ,  $J$  and  $K$  are the numbers of age, period and cohort indices, while  $L + 1$  is the lower period index, see Kuang, Nielsen and Nielsen (2008a).

A rectangular age-cohort array arises when  $L = 0$  and  $J = I + K - 1$ . A triangular age-cohort array arises when  $L = 0$  and  $I = J = K$ . A rectangular age-period array arises when  $L = I - 1$  and  $K = I + J - 1$ . This is visualized through a data example.

### 2.2 Illustration: Belgian lung cancer data

Table 1 shows the Belgian lung cancer mortality data arranged as an age-period array. It is taken from Clayton and Schifflers (1987a). It presents mortality rates  $Y_{ik}/Z_{ik}$  and responses  $Y_{ik}$ , the actual number of deaths, organised by 5-year age and period groups.

Table 2. presents that data as a trapezoid in an an age-cohort coordinate system. It is seen that there are  $I = 11$  age groups,  $J = 4$  period groups, and  $K = I + J - 1 = 14$  cohort groups. The first observations appear in the 11th period diagonal, so the lower period index is given by  $L + 1$  where  $L = \max(I, J) - 1 = 10$ .

	1880	1885	1890	1895	1900	1905	1910	1915	1920	1925	1930	1935	1940	1945
25											0.2	0.1	0.5	0.2
30										0.7	1.0	0.7	0.7	
35									0.8	1.3	1.5	1.6		
40								2.7	3.2	2.5	3.4			
45							4.8	5.6	4.9	6.1				
50						<i>6.6</i>	<i>8.5</i>	<i>7.7</i>	10.6					
55					10.4	<i>12.0</i>	12.7	14.3						
60				14.8	16.4	18.0	17.6							
65			20.5	22.6	24.9	24.3								
70		26.2	27.7	30.6	36.9									
75	33.5	33.6	36.8	43.7										

Table 2: Belgian lung cancer data. Rates shown in an age-cohort coordinate system.

Three entries in Table 2 are shown in italic. The age 50 entry for the 1905 cohort is the central point for the first diagonal. This, along with the points to the right and below is subsequently used for the reparametrization of the age-period-cohort model.

### 3 Model & identificaton

Age-period-cohort models are often modelled using generalized linear models. An example is the dose-response Poisson model, which is presented here, while other models are discussed in §6. The data consists of responses, such as counts of deaths,  $Y_{ik}$  and doses or exposures  $Z_{ik}$ , where age  $i$  and cohort  $k$  vary in a generalized trapezoid  $\mathcal{I}$ . The responses are assumed independent Poisson distributed conditional on the doses. The expectation is

$$\mathbb{E}(Y_{ik}|Z_{ik}) = \exp(\mu_{ik})Z_{ik} = \exp(\mu_{ik} + \log Z_{ik}), \quad (3.1)$$

with predictor  $\mu_{ik}$  satisfying (1.1) and offset  $\log Z_{ik}$ . The log likelihood is

$$\ell(\xi; Y|Z) = \sum_{i,k \in \mathcal{I}} Y_{ik}(\mu_{ik} + \log Z_{ik}) - \sum_{i,k \in \mathcal{I}} Z_{ik} \exp(\mu_{ik}). \quad (3.2)$$

Replacing the predictor  $\mu_{ik}$  with the time effects on the right hand side of (1.1) leads to an unidentified likelihood where the design matrix has reduced rank.

The identification problem can be addressed in two ways. The first approach is to work with restricted versions of the time effects. While this gives an appearance of preserving a connection to the time effects it leads to challenges in terms of interpretation, inference, graphical representation and forecasting, see Nielsen and Nielsen (2014) for a detailed discussion. The second approach is to reparametrise the model in terms of freely varying parameters. This is done in the following.

#### 3.1 Representation of the trend model

We start by considering the trend model, which is the simplest model where the identification problem occurs. The trend model arises when the age, period, and cohort time

effects are all linear trends, so that, with  $j = i + k - 1$ ,

$$\alpha_i = \alpha_c + \alpha_\ell(i - 1), \quad \beta_j = \beta_c + \beta_\ell(j - 1), \quad \gamma_k = \gamma_c + \gamma_\ell(k - 1). \quad (3.3)$$

The linear predictor of the age-period-cohort model in (1.1) then reduces to

$$\mu_{ik} = \alpha_c + \alpha_\ell(i - 1) + \beta_c + \beta_\ell(j - 1) + \gamma_c + \gamma_\ell(k - 1) + \delta. \quad (3.4)$$

The predictor  $\mu_{ik}$  is therefore parametrised in terms of three slope parameters  $\alpha_\ell, \beta_\ell, \gamma_\ell$ , and four level parameters  $\alpha_c, \beta_c, \gamma_c, \delta$ . The graph of the predictor  $\mu_{ik}$  in (3.4) is a linear plane. A parsimonious parametrisation of the linear plane requires just three pieces of information. From those three pieces of information it is not possible to identify all the seven time effect parameters  $\alpha_\ell, \beta_\ell, \gamma_\ell, \alpha_c, \beta_c, \gamma_c, \delta$ . The consequence is that the equations (1.1) and (3.4) for the predictor in terms of the time effects are motivations for the analysis in the tradition of symbolic description of factorial models of Wilkinson and Rogers (1973). Since the aim of the age-period-cohort model is to perform time series analysis of its components it is useful to work with an explicitly identified parametrisation. We will therefore represent the predictor in terms of identified parameters.

The plane can be represented by any three points in the generalized trapezoid index set  $\mathcal{I}$  that are not linearly related. Equally, the plane can be represented by an point in  $\mathcal{I}$  combined with slopes in two directions. The choice is completely up to the user. From an identification view point it is important to note whatever choice is made the three pieces of information are functions of the predictor on the left hand side of (3.4) so they are fully identifiable.

In the age-period-cohort model it is useful to identify the plane in terms of a single reference point along with two slopes. This is useful for instance in a Poisson model where conditioning on an overall level gives a multinomial model, see §6.1. The reference point is chosen symmetrically in age and cohort, so that it is the middle point on the first period diagonal with an odd number of observations. That is

$$\nu_0 = \mu_{UU} \quad \text{with} \quad U = \text{integer}\left(\frac{L+3}{2}\right). \quad (3.5)$$

The linear slopes are measured as 1-step increments in the age and cohort directions

$$\nu_a = \mu_{U+1,U} - \mu_{UU}, \quad \nu_c = \mu_{U,U+1} - \mu_{UU}. \quad (3.6)$$

This way, the linear surface (3.4) has representation

$$\mu_{ik} = \nu_0 + (i - U)\nu_a + (k - U)\nu_c. \quad (3.7)$$

For the Belgian lung cancer data, the lower period index is  $L + 1 = 11$  so that  $U$  is the integer part of  $13/2$  so  $U = 6$ . The points  $\mu_{UU}, \mu_{U+1,U}, \mu_{U,U+1}$  are illustrated with italic font in Table 2 for the Belgian lung cancer data.

## 3.2 Decomposing the time effects

The above analysis implies that the time effects appearing on the right hand side of (1.1) are not fully identifiable from the predictor on the left hand side of (1.1). The variation in

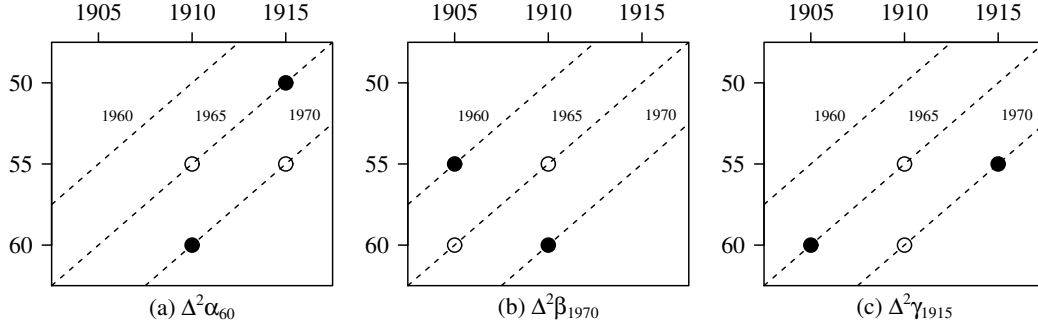


Figure 1: Illustration of log-odds-ratio interpretation of  $\Delta^2\alpha_{60}$ ,  $\Delta^2\beta_{1970}$  and  $\Delta^2\gamma_{1915}$ . The vertical/horizontal axes are for age/cohort. Period is on the diagonals. Filled/unfilled symbols indicate that log mortality is taken with positive/negative sign.

the time effects is, however, identified up to a linear trend. Indeed the second differences of the time effects are identifiable as remarked by Osmond and Gardner (1982), Clayton and Schiffers (1987b). To explore this the age effects is decomposed into linear and non-linear terms following the argument of Kuang, Nielsen and Nielsen (2008a).

Introduce first differences  $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$  and second differences  $\Delta^2\alpha_i = \Delta\alpha_i - \Delta\alpha_{i-1}$ . The second differences are identifiable from the predictors since

$$\Delta^2\alpha_i = \mu_{ik} - \mu_{i-1,k+1} - \mu_{i-1,k} + \mu_{i-2,k+1}.$$

Thus, the age second differences have interpretation as growth rates of the time effects as well as log odds ratios of the mortality. Figure 1 illustrates the interpretation of  $\Delta^2\alpha_{60}$ ,  $\Delta^2\beta_{1970}$  and  $\Delta^2\gamma_{1915}$  with the Belgian lung cancer data in mind.

In turn, the level of the age effect decomposes in terms of a linear trend and a double sum of double differences. With a view to cover generalized trapezoid data indices  $\mathcal{I}$  the linear trend is anchored at  $i = U, U + 1$ , so that

$$\alpha_i = \alpha_c + \alpha_\ell(i - U) + 1_{(i < U)} \sum_{t=i+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2\alpha_s + 1_{(i > U+1)} \sum_{t=U+2}^i \sum_{s=U+2}^t \Delta^2\alpha_s. \quad (3.8)$$

The level  $\alpha_c = \alpha_U$  and linear slope  $\alpha_\ell = \Delta\alpha_{U+1}$  are not identifiable from the predictor.

### 3.3 Representation of the age-period-cohort model

The representation for the trend model in §3.1 and the decomposition of the time effects in §3.2 are now combined into a representation for the age-period-cohort model. From that expression a design matrix can be constructed.

The analysis of the trend model and the time effects suggests that the age-period-cohort model can be parametrised parsimoniously by the parameter

$$\xi = (\nu_0, \nu_a, \nu_c, \Delta^2\alpha_3, \dots, \Delta^2\alpha_I, \Delta^2\beta_{L+2}, \dots, \Delta^2\beta_{L+J}, \Delta^2\gamma_3, \dots, \Delta^2\gamma_K)', \quad (3.9)$$

of dimension  $p = I + J + K - 3$ . Indeed, Theorem 2 of Kuang, Nielsen and Nielsen (2008a) shows that the parameter  $\xi$  uniquely identifies the age-period-cohort model for any generalized trapezoid index set, see also Theorem 3.1 below.

A representation for the predictor arises by combining the linear trend representation (3.7) and the age decomposition (3.8) with similar decompositions for period and cohort. This results in the representation

$$\mu_{ik} = \nu_0 + (i - U)\nu_a + (k - U)\nu_c + A_i + B_{i+k-1} + C_k, \quad (3.10)$$

where the linear trend parameters  $\nu_0, \nu_a, \nu_c$  are defined in (3.5), (3.6). The terms  $A_i, B_j, C_k$  are double sums of differences cumulated forward if  $i$  and  $k$  are larger than  $U$  backward otherwise so that

$$A_i = 1_{(i < U)} \sum_{t=i+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2 \alpha_s + 1_{(i > U+1)} \sum_{t=U+2}^i \sum_{s=U+2}^t \Delta^2 \alpha_s \quad (3.11)$$

$$B_j = 1_{(L \text{ odd} \ \& \ j=2U-2)} \Delta^2 \beta_{2U} + 1_{(j > 2U)} \sum_{t=2U+1}^j \sum_{s=2U+1}^t \Delta^2 \beta_s \quad (3.12)$$

$$C_k = 1_{(k < U)} \sum_{t=k+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2 \gamma_s + 1_{(k > U+1)} \sum_{t=U+2}^k \sum_{s=U+2}^t \Delta^2 \gamma_s. \quad (3.13)$$

The properties of the representation are summarized in the following theorem, which is proved in the Appendix.

**Theorem 3.1.** *Let the predictor  $\mu_{ik}$  satisfy (1.1) for  $i, k \in \mathcal{I}$ . Consider the parameter  $\xi$  of (3.9). Then*

- (i)  $\mu$  is a function of the time effects  $\alpha_i, \beta_j, \gamma_k$  through  $\xi$  as given by (3.10);
- (ii) the parametrisation of  $\mu$  by  $\xi$  is exactly identified in that  $\xi^\dagger \neq \xi^\ddagger \Rightarrow \mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$ .

Theorem 3.1 generalizes the result for age-cohort arrays in Kuang, Nielsen and Nielsen (2008a). For age-period arrays the representation is equivalent, yet different in appearance, to that of Martínez Miranda, Nielsen and Nielsen (2014). That representation is formulated for an age-period coordinate system and therefore lacks the symmetry coming with an age-cohort coordinate system.

An important implication of Theorem 3.1 is that it allows us to draw on the theory of exponential families. The Poisson model is an exponential family with a predictor that is linear in  $\xi$  due to (3.10). Theorem 3.1 shows that the parameter  $\xi$  is freely varying. The exponential family is then regular with  $\xi$  as canonical parameter, see Barndorff-Nielsen (1978, p. 116).

The result could also be written in terms of group-theoretic arguments. The idea is to note that the equation (1.1) is a mapping from the time effects  $\alpha_i, \beta_j, \gamma_k$  to the predictor  $\mu_{ik}$ . The mapping is invariant to a change of levels and slopes in that

$$\mu_{ik} = (\alpha_i + a + id) + (\beta_j + b - jd) + (\gamma_k + c + kd) + (\delta - a - b - c - d), \quad (3.14)$$

for arbitrary constants  $a, b, c, d$ . We can think of this as a group acting on the time effects. Theorem 3.1 then shows that  $\xi$  is a maximal invariant. Further details are given in Kuang, Nielsen and Nielsen (2008a) and Nielsen and Nielsen (2014).

### 3.4 Design matrix

The representation (3.10) shows that the predictors satisfy a linear relation  $\mu_{ik} = X'_{ik}\xi$ , for some vector  $X_{ik}$ . Stacking the row vectors  $X'_{ik}$  gives a design matrix. By swapping the summation order of the double sums in (3.11)-(3.13) the design matrix is seen to have row vectors

$$X'_{ik} = (1, i - U, j - U, X'_{ik,a}, X'_{ik,p}, X'_{ik,c}) \quad (3.15)$$

where  $X_{ik,a}$ ,  $X_{ik,p}$ ,  $X_{ik,c}$  are vectors of dimension  $I - 2$ ,  $J - 2$  and  $K - 2$  given by

$$X'_{ik,a} = \{\dots, 1_{(i < U)}(s - i - 1) + 1_{(i > U+1)}(i - s + 1), \dots\} \quad \text{for } s = 3, \dots, I, \quad (3.16)$$

$$X'_{ik,p} = \{\dots, 1_{(L \text{ odd \& } j=L)} + 1_{(j > 2U)}(j - s + 1), \dots\} \quad \text{for } s = 3, \dots, J, \quad (3.17)$$

$$X'_{ik,c} = \{\dots, 1_{(k < U)}(s - k - 1) + 1_{(k > U+1)}(k - s + 1), \dots\} \quad \text{for } s = 3, \dots, K. \quad (3.18)$$

### 3.5 Likelihood

The log likelihood (3.2) for the Poisson model can now be written in a computationally convenient way. Inserting that the predictor satisfies  $\mu_{ik} = X'_{ik}\xi$  gives

$$\ell(\xi; Y|Z) = \sum_{i,k \in \mathcal{I}} Y_{ik}(X'_{ik}\xi + \log Z_{ik}) - \sum_{i,k \in \mathcal{I}} Z_{ik} \exp(X'_{ik}\xi), \quad (3.19)$$

where  $\xi$  is a freely varying vector of dimension  $p = I + J + K - 3$ . This is the likelihood of a Poisson regression with a log link, design matrix given by (3.15) and offset  $\log Z_{ik}$ .

The statistic  $T = \sum_{i,k \in \mathcal{I}} Y_{ik} X'_{ik}$  is minimal sufficient. Exponential family theory shows that the likelihood equation  $T = E_{\xi} T$  has a unique solution when the sufficient statistic  $T$  is interior to its convex support, see Barndorff-Nielsen (1978). A sufficient condition is that all elements of the sufficient statistic  $T$  are positive. Kuang, Nielsen and Nielsen (2009) give a formal analysis of the condition for an age-cohort model.

The likelihood can be analysed by any software for generalized linear models. The R package `apc` has the design matrix pre-coded.

### 3.6 Illustration: Belgian lung cancer data

Fitting the model to the Belgian lung cancer data gives a deviance of 20.2 with  $p = 0.32$  when compared to a  $\chi^2_{18}$ -distribution. Here, the deviance is the log likelihood ratio statistic between an age-period-cohort model with  $\mu_{ik} = X'_{ik}\xi$  and  $p = I + J + K - 3$  parameters and a saturated model with unrestricted  $\mu_{ik}$  and therefore  $IJ$  parameters, where  $I = 11$ ,  $J = 4$ ,  $K = I + J - 1$ . To complement this Figure 2 shows probability transforms of the observations given the fitted age-period-cohort model. This confirms that the deviance with its many degrees of freedom does not hide outliers.

Figure 3 (a)-(c) shows the estimated second difference parameters  $\Delta^2 \alpha_i$ ,  $\Delta^2 \beta_j$ ,  $\Delta^2 \gamma_k$ . The estimates are plotted with pointwise confidence bands, for which the derivation is discussed below in §5.5. The second differences are volatile, in particular for youngest age and cohort groups which have fewer responses. Most estimates are close to zero judged by their pointwise confidence bands. The pointwise confidence bands are, however, not



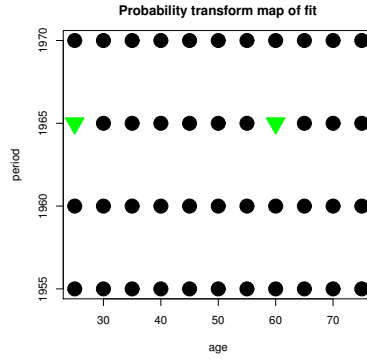


Figure 2: Probability transform map. A circle/triangle indicates that the observation between the central 10%-90% and the left 5%-10% quantiles of the fitted distribution.

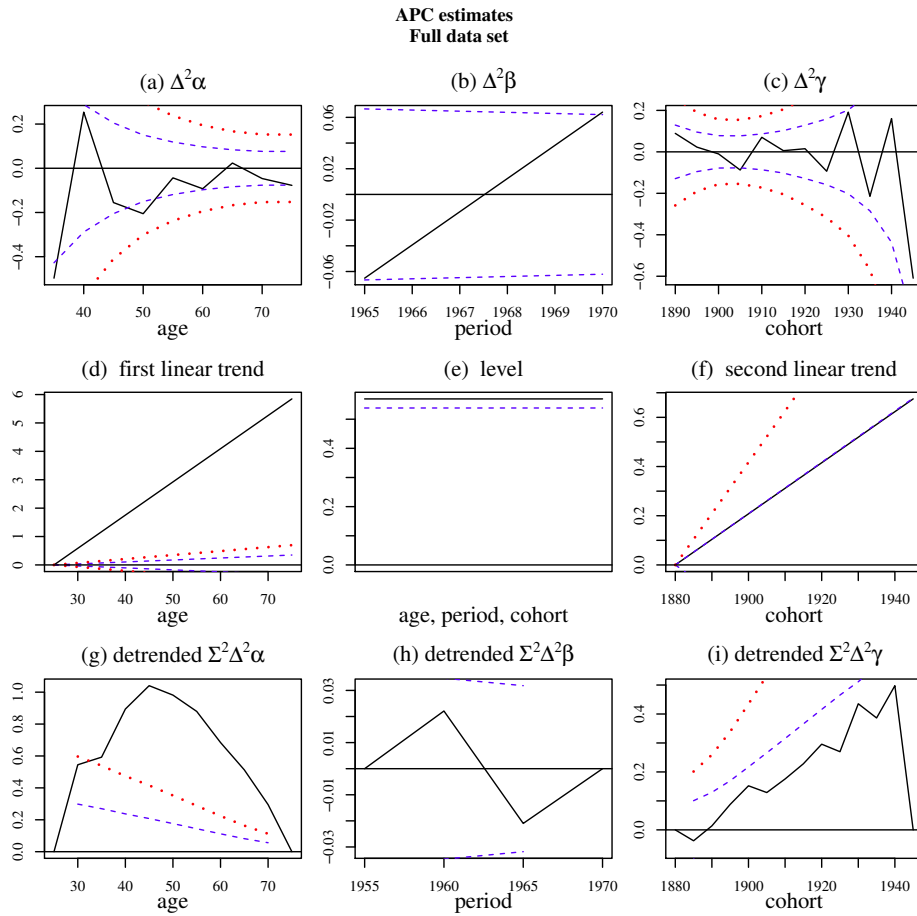


Figure 3: Estimates for age-period-cohort model fitted to Belgian lung cancer data. Shown with pointwise standard deviations centered at zero and multiplied by 1 (dashed) or 2 (dotted). Panels (a)-(c) show canonical parameters. Panels (d)-(i) show ad hoc identified time effects.

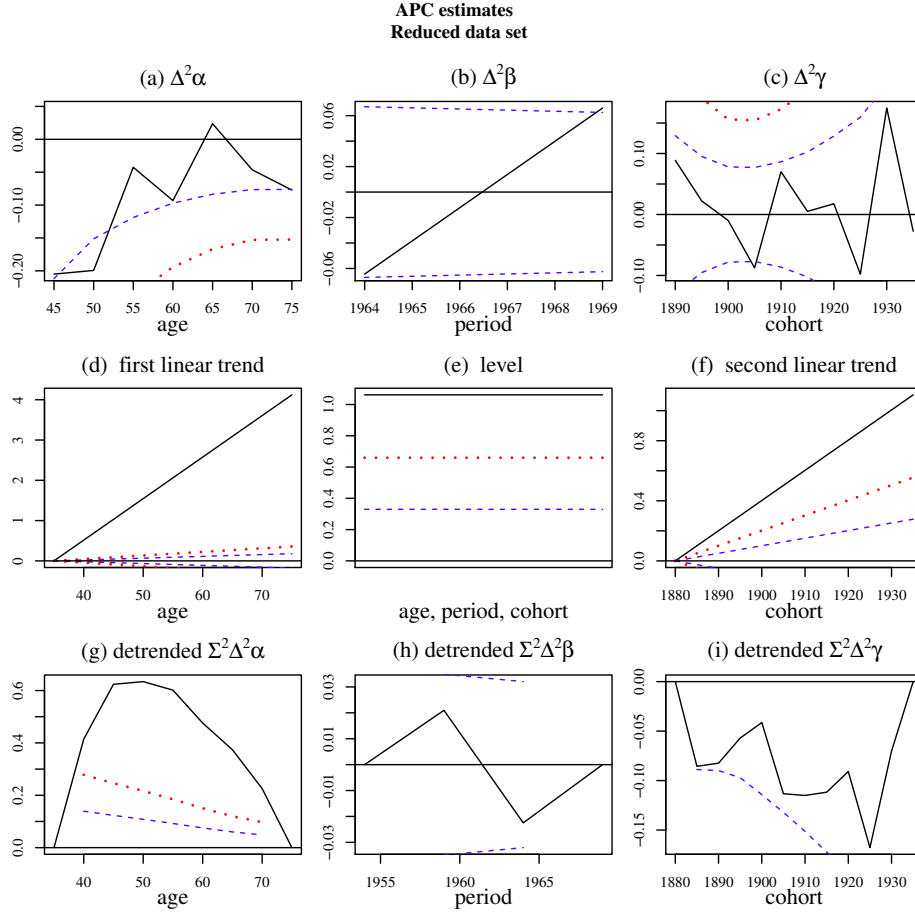


Figure 4: Estimates for age-period-cohort model fitted to Belgian lung cancer data, excluding the first two age groups. Shown with pointwise standard deviations centered at zero and multiplied by 1 (dashed) or 2 (dash-dotted). Panels (a)-(c) show canonical parameters. Panels (d)-(i) show ad hoc identified time effects.

fully informative about joint hypotheses, such as the hypothesis of absence of age effect so that  $\Delta^2\alpha_i = 0$  for all  $i$ . Formal test for such joint hypotheses follow in §5.6.

The estimates for the level and slopes are

$$\widehat{\nu}_0 = \widehat{\mu}_{50,1905} = 1.96, \quad \widehat{\nu}_a = 0.50, \quad \widehat{\nu}_c = 0.12, \quad (3.20)$$

(0.06)                      (0.08)                      (0.07)

with standard deviations reported below coefficients. The slope in the cohort direction is only marginally significant. However, it is probably advisable to sort out the age, period and cohort effects over and above the linear plane before imposing such a hypothesis.

While the estimated double differences have a volatile appearance their population versions are invariant to the choice of data array. To illustrate this, the model was estimated from a reduced data set not using the information in the first two age groups. This also removes the last two cohort groups, but leaves the number of period groups unaltered. Results are plotted in Figure 4. Up to sampling error, the double differences are the same in the panels (a,b,c) of Figures 3 and 4.

The estimates for the level and slopes based on the reduced data set are

$$\widehat{\nu}_0 = \widehat{\mu}_{55,1900} = \underset{(0.06)}{2.41}, \quad \widehat{\nu}_a = \underset{(0.07)}{0.41}, \quad \widehat{\nu}_c = \underset{(0.06)}{0.05}. \quad (3.21)$$

The estimates in (3.21) has an anchoring point different from the full sample estimates in (3.20). Corrected for that the estimates are roughly in line. Indeed the new levels estimate of 2.41 is close to a corrected version of the old  $1.96 + 0.50 - 0.12 = 2.34$ .

## 4 An ad hoc identified plot of time effects

Given the original model formulation in (1.1) it is of interest to represent the time effects for age, period and cohort somehow. Plotting the time effects will inevitably involve an ad hoc identification of the linear trends. However, the visual impression of such ad hoc identified time effects will be different, possibly because the human mind tends to focus on deviations from a horizontal line. Table II of Clayton and Schifflers (1987b) illustrates this point, whereas a more formal discussion is given in Nielsen and Nielsen (2014). It is therefore necessary to be clear about the purpose of such plots and the consequences of the ad hoc identification. It is worth noting that the estimation outline in the previous section is done without any need for ad hoc identification. The ad hoc identification is only needed for illustration.

In the following a particular plot is suggested and some interpretations are given. This is then illustrated using the Belgian lung cancer data.

### 4.1 The plot

The double differences give a very volatile impression of the deviation from the pure trend model of §3.1. It is very hard to see if there are tendencies in the time effects over and above the arbitrary linear trend. With that in mind the linear trend could be chosen so as to tie the ad hoc identified time effects down to start and end in zero - just as a Brownian bridge ties down a Brownian motion. Visually, this will emphasize variation over and above arbitrary linear trends. For the age effect this is done as follows.

Expand the age effect  $\alpha_i$  as in (3.8), but with an anchoring point  $U = 1$ . Insert the estimated double differences and choose the level and slopes so that the series starts and ends in zero. This gives the detrended age effect

$$\widehat{\alpha}_i^{detrend} = a_c + a_\ell(i - 1) + 1_{(i>2)} \sum_{t=3}^i \sum_{s=3}^t \widehat{\Delta^2 \alpha}_s, \quad i = 1, \dots, I, \quad (4.1)$$

which satisfies  $\widehat{\alpha}_1^{detrend} = \widehat{\alpha}_I^{detrend} = 0$  when  $a_c = 0$  and  $a_\ell = (I - 1)^{-1} \sum_{t=3}^I \sum_{s=3}^t \widehat{\Delta^2 \alpha}_s$ . The detrended age effect has  $I - 2$  non-zero values in line with the dimension of the age double differences. It does, however, not share the invariance to the choice of data array, see §4.3 for further discussion.

## 4.2 Interpretation

In some situations the time effect has a concave appearance. This will be particularly visible in the detrended plot. Examples include the age effects in the Belgian lung cancer data discussed in §4.3, the lung cancer study for Californian women by Holford (2006) and the mesothelioma study by Martínez Miranda, Nielsen and Nielsen (2014). Holford also finds a cyclical shape for the period effect. Stylized interpretations of these shapes are explored in the following.

A quadratic concave shape arises when the double differences are equal,  $\Delta^2\alpha_i = a$ . In that case the detrended age effect satisfies, as shown in Appendix A.2,

$$\alpha_i^{detrend} = \frac{1}{2}(i-1)(i-I)a, \quad i = 1, \dots, I. \quad (4.2)$$

This quadratic curve will be concave when the double differences are negative. If  $a = 0$  the age effect is log linear. When  $a < 0$  the age effect is sub log linear and the increase in mortality is levelling off with age. This applies to the cancer studies mentioned above.

An S shape, which is concave at first and then convex, can arise when the double differences are linear,  $\Delta^2\alpha_i = a + b(i-2)$ . The age effects is then cubic and its detrended version satisfies, see Appendix A.2 for details,

$$\alpha_i^{detrend} = \frac{1}{2}(i-1)(i-I)\left\{a + \frac{b}{3}(i+I-2)\right\}, \quad i = 1, \dots, I. \quad (4.3)$$

The S shape occurs when the term in curly brackets is zero for some  $i$  so  $1 < i < I$ . This happens if  $(I-1) < -3a/b < 2(I-1)$ . In particular, if  $a = -1$  and  $-3a/b$  is in the middle of the indicated interval then, approximately,  $\Delta^2\alpha_i$  increases from  $-1$  to  $+1$  while  $\alpha_i^{detrend}$  is concave for  $i \leq I/2$  and then convex. The interpretation is that the age effect is a first sub log linear and then super log linear, so that the increase in the age effect falls less and less and then start increasing. If instead  $a = -1$  and  $b$  is chosen so that  $\Delta^2\alpha_i$  increases, approximately, from  $-1$  to  $0$  then  $-3a/b$  is outside the indicated interval and  $\alpha_i^{detrend}$  is concave, although not quadratic.

## 4.3 Illustration: Belgian lung cancer data

Figure 3 (g)-(i) show the estimated time effects that are ad hoc identified by the detrending method. These plots start and end in zero in order to focus on variation over and above the arbitrary linear trends. In §5.6 the period and cohort effect are found not to be significant.

The plot (g) of the detrended age effect shows a concave appearance. The age double differences in (a) are very volatily, but have a tendency to increase linearly from negative towards zero, roughly in line with the cubic interpretation offered above. This will be tested formally in §5.6.

The plot (h) of the detrended period effect shows an S-shape over the just 4 observed periods. The period double differences in (b) rise linearly from approximately  $-0.06$  to  $+0.06$ . So the levelling off of the increase in the period effect becomes less and less.

The plot (i) of the detrended cohort effect also shows an S-shape, or, rather, an approximate linear shape until 1940 and then a sharp drop.

Figure 3 (d)-(f) indicate the linear plane that arises with the identification choice made for time effects. The slopes of the plane in the age and the cohort directions are indicated in panels (d), (f), while panel (e) shows the corresponding level. The normalisation of the implied linear plane is chosen to match the non-linear effects in panels (g)-(i). Thus, the mortality rate for people of age 25 born in 1945 and observed in 1970 is found by combining the relevant values from the panels (d)-(i). The linear plane increases in cohort and even more in age. Consequently, it is also increasing in period. In terms of magnitudes, the linear growth in particular in the age direction seems to dominated the non-linear effects seen in panels (g)-(i).

Figure 4 shows the estimates from a truncated data set where the youngest two age groups and, consequently, the also last two cohort groups are removed. The ad hoc identification is now different for age and cohort so that the graphs in panels (d)-(i) are somewhat different from those in Figure 4(d)-(i). This illustrates the wellknown point that some caution is needed when interpreting add hoc identified time trends, yet many features are similar in the two figures.

The largest difference is seen for cohort trends in panels (i). Since the last two cohort groups are removed in Figure 4 the sharp drop in 1945 is gone. The roughly linear shape until 1940 now dominates. Due to the detrending that linear shape now has a horizontal slope in Figure 4 (i). Note, that the detrending makes it easier to see from Figure 4 (i) than Figure 3 (i) that the linear shape actually wriggles quite a bit. However, that wiggly shape is not significant as will be discussed in §5.6.

The linear components in (d)-(f) are also changed from Figure 3 to Figure 4. This is a result of the different ad hoc identification and reconcile as in (3.20), (3.21).

Overall, the estimates are remarkably stable with respect to removing some of the observations. This is a further indication that the Poisson model with age-period-cohort describes the data well.

## 5 Inference

The age-period-cohort model has a number of interesting sub-models. Most of these have previously been discussed by Clayton and Schifflers (1987a,b). The advantage of the present approach is that the sub-models are directly seen to be linear restrictions on the canonical parameter  $\xi$  in (3.9). Since the restrictions are linear the sub-models inherit the regular exponential family property of the unrestricted model. Estimation, computation of likelihood ratio test statistics or deviances and counting degrees of freedom are therefore very simple. Table 3 shows the restrictions imposed by the sub-models, while Figure 5 gives an overview of their nesting structure. The sub-models are discussed in some detail.

### 5.1 Two factor models

A prominent sub-model is the age-cohort model, denoted AC in Table 3. This hypothesis arises by setting the period double differences to zero:

$$\mathbf{H}^{AC} : \quad \Delta^2\beta_3 = \dots = \Delta^2\beta_J = 0. \quad (5.1)$$

	$\nu_0$	$\nu_a$	$\nu_c$	$\Delta^2\alpha$	$\Delta^2\beta$	$\Delta^2\gamma$	df
APC	*	*	*	*	*	*	$I + J + K - 3$
AP	*	*	*	*	*	0	$I + J - 1$
AC	*	*	*	*	0	*	$I + K - 1$
PC	*	*	*	0	*	*	$J + K - 1$
Ad	*	*	*	*	0	0	$I + 1$
Pd	*	*	*	0	*	0	$J + 1$
Cd	*	*	*	0	0	*	$K + 1$
A	*	*	0	*	0	0	$I$
P	*	=	=	0	*	0	$J$
C	*	*	0	0	0	*	$K$
t	*	*	*	0	0	0	3
tA	*	*	0	0	0	0	2
tP	*	=	=	0	0	0	2
tC	*	0	*	0	0	0	2
1	*	0	0	0	0	0	1

Table 3: Overview of sub-models of the age-period-cohort model.

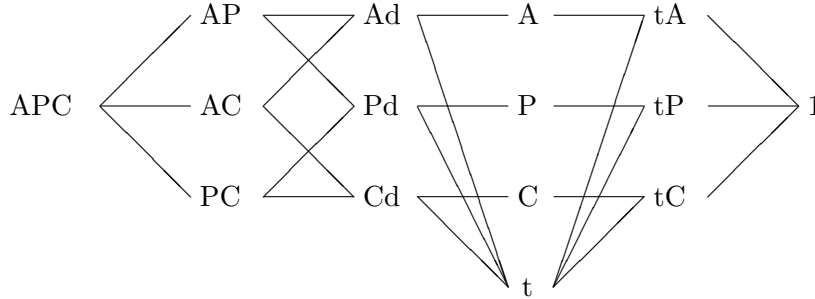


Figure 5: Diagram of the age-period-cohort model and its sub-models. The models are nested from right to left.

Expressed in terms of time effects the age-period model is

$$\mu_{ik}^{AC} = \alpha_i + \gamma_k + \delta, \quad (5.2)$$

so that the period effects are restricted to be zero:  $\beta_1 = \dots = \beta_J = 0$ . A formal proof for the equivalence of the two formulations is given by Nielsen and Nielsen (2014, §5.3). The degrees of freedom is readily seen to be  $J - 2$  when the restriction is formulated in terms of the canonical parameter as in (5.1).

The canonical parameter (3.9) subjected to the restriction (5.1) and the representation (3.10) continue to apply for the age-cohort model. However, the two linear trend can now be attributed in a unique way to the age and cohort time effects. Thus, the canonical parameter can equivalently be chosen in terms of an overall level,  $\mu_{UU}$  say, age differences  $\Delta\alpha_i = \mu_{ik} - \mu_{i-1,k}$  for  $i = 2, \dots, I$ , and the cohort differences  $\Delta\gamma_k = \mu_{ik} - \mu_{i,k-1}$

for  $k = 2, \dots, K$ . An alternative representation in terms of these parameters is

$$\begin{aligned} \mu_{ik}^{AC} = & \mu_{UU} - 1_{(i < U)} \sum_{t=i+1}^U \Delta\alpha_t + 1_{(i > U)} \sum_{t=U+1}^i \Delta\alpha_t \\ & - 1_{(k < U)} \sum_{t=k+1}^U \Delta\gamma_t + 1_{(k > U)} \sum_{t=U+1}^k \Delta\gamma_t. \end{aligned} \quad (5.3)$$

In a similar fashion the age-period model, denoted AP in Table 3, arises by eliminating the cohort double differences  $\Delta^2\gamma_k = 0$  for  $k = 3, \dots, K$ . The period-cohort model PC restricts the age double differences  $\Delta^2\alpha_i = 0$  for  $i = 3, \dots, I$ .

## 5.2 Drift models

The drift models of Clayton and Schifflers (1987a) arise by eliminating two time effects.

The age drift model, denoted Ad in Table 3, is given by the hypothesis

$$\mathbf{H}^{Ad} : \quad \Delta^2\beta_3 = \dots = \Delta^2\beta_J = 0, \quad \Delta^2\gamma_3 = \dots = \Delta^2\gamma_K = 0. \quad (5.4)$$

Now, there are two linear trends, but only one time effect, so the linear trends cannot be attributed to the time effect in a clear way. Further discussion of that issue is given in Clayton and Schifflers (1987a, 1987b). There is therefore no obvious alternative to the representation (3.10). We see that the identification problem is tied to the model specification in an intricate way: The trends are unidentified in the unrestricted age-period-cohort model, then identified under the age-cohort restriction (5.1), but unidentified under the further nested age-drift restriction (5.4).

The age-drift model is fiddly to express in terms of time effects as pointed out by Clayton and Schifflers (1987b). The two time effect models

$$\mu_{ik}^{Ad} = \alpha_i + \beta_c + \beta_\ell j + \delta \quad \text{and} \quad \mu_{ik}^{Ad} = \alpha'_i + \gamma_c + \gamma_\ell k + \delta, \quad (5.5)$$

are equivalent but neither is identified. The models are equivalent since  $\alpha'_i = \alpha_i + \beta_\ell$ ,  $\gamma_c = \beta_c - \beta_\ell$  and  $\gamma_\ell = \beta_\ell$ .

## 5.3 One factor models

A single factor model arises by eliminating two sets of double differences and one time effect. An example is the age model, A, where the hypothesis is

$$\mathbf{H}^A : \quad \Delta^2\beta_3 = \dots = \Delta^2\beta_J = 0, \quad \Delta^2\gamma_3 = \dots = \Delta^2\gamma_K = 0, \quad \nu_c = 0. \quad (5.6)$$

With this restriction there is one level and one linear slope, so the time effect model is fully identified:

$$\mu_{ik}^A = \alpha_i. \quad (5.7)$$

The period and cohort models arise in a similar way by the restrictions

$$\mathbf{H}^P : \quad \Delta^2\alpha_3 = \dots = \Delta^2\alpha_I = 0, \quad \Delta^2\gamma_3 = \dots = \Delta^2\gamma_K = 0, \quad \nu_a = \nu_c, \quad (5.8)$$

$$\mathbf{H}^C : \quad \Delta^2\alpha_3 = \dots = \Delta^2\alpha_I = 0, \quad \Delta^2\beta_3 = \dots = \Delta^2\beta_J = 0, \quad \nu_a = 0. \quad (5.9)$$

## 5.4 The trend model

The trend model was introduced in §3.1. It arises from the restriction

$$\mathbf{H}^t : \quad \Delta^2\alpha_3 = \cdots = \Delta^2\alpha_I = \Delta^2\beta_3 = \cdots = \Delta^2\beta_J = \Delta^2\gamma_3 = \cdots = \Delta^2\gamma_K = 0. \quad (5.10)$$

The linear plane can be restricted further to slope only in the age, period or cohort direction, that is

$$\mathbf{H}^{tA} : \quad \mathbf{H}^t \cap (\nu_c = 0), \quad \mathbf{H}^{tP} : \quad \mathbf{H}^t \cap (\nu_a = \nu_c), \quad \mathbf{H}^{tC} : \quad \mathbf{H}^t \cap (\nu_a = 0). \quad (5.11)$$

If it has no slope in any direction the mortality is constant, that is

$$\mathbf{H}^1 : \quad \mathbf{H}^t \cap (\nu_a = \nu_c = 0). \quad (5.12)$$

## 5.5 Sampling theory

The Poisson model is a regular exponential family with canonical parameter  $\xi$  and sufficient statistic  $T = \sum_{i,k \in \mathcal{I}} Y_{ik} X'_{ik}$ . Conditionally on the doses the sufficient statistic has expectation and variance given by

$$\begin{aligned} \kappa_1(\xi) = \mathbf{E}_\xi(T|Z) &= \sum_{i,k \in \mathcal{I}} Z_{ik} \exp(X'_{ik}\xi) X'_{ik}, \\ \kappa_2(\xi) = \mathbf{Var}_\xi(T|Z) &= \sum_{i,k \in \mathcal{I}} Z_{ik} \exp(X'_{ik}\xi) X_{ik} X'_{ik}, \end{aligned}$$

where  $\kappa_2(\xi)$  is also the expected information.

The asymptotic sampling theory assumes a fixed index set  $\mathcal{I}$ , that the total dose  $Z_{..} = \sum_{i,k} Z_{ik}$  increases to infinity while  $Z_{ik}/Z_{..}$  converge to positive constants  $\Sigma_{ik} > 0$  for all  $i, k$ . The Central Limit Theorem shows  $Z_{..}^{-1/2}\{Y_{ik} - Z_{ik} \exp(\mu_{ik})\}$  are asymptotically independent normal with variance  $\Sigma_{ik} \exp(\mu_{ik})$ . Since the dimension of the index set  $\mathcal{I}$  is kept fixed then  $Z_{..}^{-1/2}\{T - \kappa_1(\xi)\}$  is asymptotically normal with variance  $\kappa_2(\xi)$ . Applying the inverse mapping  $\kappa_1^{-1}$  and noting that  $\kappa_1(\xi)$  has derivative  $\kappa_2(\xi)$  then

$$\{\kappa_2(\xi)\}^{1/2}(\hat{\xi} - \xi) \xrightarrow{D} \mathbf{N}(0, I_p).$$

As a consequence, the standard output of standard deviations from for instance R is applicable as long as the doses  $Z_{ik}$  are not too small. Moreover the likelihood ratio test statistics, or deviance test statistic, for any smooth hypothesis is asymptotically  $\chi^2$ .

## 5.6 Illustration: Belgian lung cancer data

Table 4 presents a deviance table for the Belgian lung cancer data. The table is organised in line with Table 3. Columns 2-4 show the deviance of the sub-model along with the associated degrees of freedom and the p-value when judged against a  $\chi^2$ -distribution. Columns 5-7 show the log likelihood ratio statistics, or relative deviance, for the sub-model against the age-period-cohort model along with degrees of freedom and p-value. Finally, column 8 shows the information criteria suggested by Akaike.



	deviance	$df_{\text{dev}}$	$p_{\text{dev}}$	LR	$df_{\text{LR}}$	$p_{\text{LR}}$	aic
APC	20.2	18	0.32				341.4
AP	25.6	30	0.70	5.3	12	0.95	322.7
AC	21.5	20	0.37	1.2	2	0.54	338.6
PC	99.2	27	0.00	79.0	9	0.00	402.4
Ad	26.6	32	0.74	6.4	14	0.96	319.8
Pd	253.6	39	0.00	233.3	21	0.00	532.7
Cd	100.7	29	0.00	80.5	11	0.00	399.9
A	85.6	33	0.00	65.4	15	0.00	376.7
t	254.5	41	0.00	234.3	23	0.00	529.7

Table 4: Deviance table for Belgian lung cancer data.

	deviance	$df_{\text{dev}}$	$p_{\text{dev}}$	LR	$df_{\text{LR}}$	$p_{\text{LR}}$	aic
Ad	26.6	32	0.74				319.8
Adc	31.5	39	0.80	5.0	7	0.66	310.7
Adq	39.4	40	0.50	12.9	8	0.11	316.6

Table 5: Further analysis of the age-drift model.

Among the two factor models, the age-cohort model,  $\mathbf{p} = 0.54$ , and the age-period model,  $\mathbf{p} = 0.95$ , cannot be rejected, while the period-cohort model,  $\mathbf{p} = 0.00$ , is rejected. This suggests the age effect is needed, while the period and cohort effects are not significant. A further reduction to the age-drift model,  $\mathbf{p} = 0.96$ , cannot be rejected. The age-drift model is nested in both the age-period model and the age-cohort model, see Figure 5. Further reductions are not supported. The conclusion is in line with Clayton and Schifflers (1987a), who used these data to illustrate the age-drift model. The estimates under the age-drift model are similar to those for the age-period-cohort model and therefore not reported here.

In the discussion of Figure 3(g) it was suggested that the age effect could be cubic or quadratic. Table 5 investigates this. Its structure is similar to that in Table 5 albeit the age-drift model is taken as reference point for the likelihood ratios statistics in column 5. The age drift model has 3 deterministic parameters and  $I - 2 = 9$  age parameters. Imposing that the age effect is cubic cannot be rejected,  $\mathbf{p} = 0.66$ . Imposing a quadratic age effect cannot be reject against the age-drift model,  $\mathbf{p} = 0.11$ , but it is rejected against the cubic model with a deviance of 7.9 and just one degree of freedom. With this information it seems reasonable to settle on the cubic age-drift model. It has five parameters, three for the linear plane and two for the cubic age effect, or rather for the linear age double differences, that is

$$\widehat{\nu}_0 = 1.97, \quad \widehat{\nu}_a = 0.49, \quad \widehat{\nu}_c = 0.088, \quad \widehat{\Delta^2\alpha}_i = -0.15 + 0.014(i - 2), \quad i = 3, \dots, I. \quad (5.13)$$

The linear plane estimates are in line with those reported in (3.20). The linear age double differences increase from  $-0.15$  to just about  $-0.02$  throughout the 9 periods, see §4.2 for interpretation.

## 6 Variations of the statistical model

A generalized linear model framework allows for a variation of the distributional assumptions of the statistical model. A few of those variations deserve mentioning.

### 6.1 Poisson response model

In some situations only responses  $Y_{ik}$  are observed while doses or exposure  $Z_{ik}$  are unobserved. We saw in (3.1) that a Poisson dose-response model gives an age-period-cohort structure to the log conditional expectation of responses given doses. In a similar fashion a Poisson response model arises by giving an age-period-cohort structure to the unconditional log expectation of responses.

A prominent example is the chain-ladder model using for reserving in general insurance. The data are typically triangular in age and cohort and the standard model in that literature is an age-cohort model. In that literature age and cohort are called development year and policy or accident year and the model is used for forecasting. England and Verrall (2002) give an overview. Kuang, Nielsen and Nielsen (2008b, 2009) discuss, respectively, a general forecasting theory for age-period-cohort models and the maximum likelihood estimation of the chain ladder model.

Martínez Miranda, Nielsen and Nielsen (2014) apply the Poisson response model to forecast the future burden of mesothelioma mortality. It is proposed to use a multinomial sampling scheme where the cumulated response  $Y_{..} = \sum_{i,k \in \mathcal{I}} Y_{ik}$  increases in the asymptotic argument while the frequencies  $Y_{ik}/Y_{..}$  are kept fixed. Asymptotic distribution theory is given for estimators and distribution forecasts.

### 6.2 Logistic dose-response model

In the dose-response situation a logistic specification could be used as an alternative to the Poisson specification. In that case responses  $Y_{ik}$  are binomially distributed given doses  $Z_{ik}$  with count parameter  $Z_{ik}$  and success (mortality) probability  $\pi_{ik}$  and log odds satisfying the age-period-cohort specification in (1.1), that is

$$\log \frac{\pi_{ik}}{1 - \pi_{ik}} = \mu_{ik}.$$

Inference can be conducting using the same sampling scheme as in §5.5.

## 7 Discussion

The traditional formulation of the age-period-cohort model is viewed as a motivation or a symbolic description in the tradition of Wilkinson and Rogers (1973). The identification problem is addressed through a reparametrization in terms of the freely varying canonical parameter. A unified parametrization covering all three Lexis diagrams was suggested. A new plot of ad hoc identified time effects was suggested and interpretation was offered. A deviance table for a range of sub-models was suggested. The methods are implemented in the R package `apc`.

## A Technical details

### A.1 Proof of Theorem 3.1

(i) The expression (3.10) is seen to be a function of the parameter  $\xi$  in (3.9). Thus, it suffices to argue that (3.10) holds. First, we check the formula in the three anchoring points. From (3.11)-(3.13) it is seen that  $A_U = A_{U+1} = B_{2U-1} = B_{2U} = C_U = C_{U+1} = 0$ . Insert this in (3.10) along with the definitions of  $\nu_0, \nu_a, \nu_c$  from (3.5), (3.6) to see that the formula correctly gives  $\mu_{UU}, \mu_{U+1,U}, \mu_{U,U+1}$ . Secondly, we check that the representation (3.10) has the correct double differences. Taking double difference of  $\mu_{ik}$  as indicated in Figure 1 indicates that it suffices to check that  $\Delta^2 A_i = \Delta^2 \alpha_i$  for  $i = 3, \dots, I$ ,  $\Delta^2 B_j = \Delta^2 \beta_j$  for  $j = L + 1, \dots, L + J$  and  $\Delta^2 C_k = \Delta^2 \gamma_k$  for  $k = 3, \dots, K$ . In the case of  $A_i$  it is convenient to check the cases  $i > U + 3$ ,  $i = U + 3$ ,  $i = U + 2$ ,  $i = U + 1$ ,  $i = U$  and  $i < U$  separately. The case of  $C_k$  follows in the same way. In the case of  $B_j$  consider odd and even  $L$  separately. For odd  $L$  then  $2U = L + 3$  and consider cases  $j > 2U + 2$ ,  $j = 2U + 2$ ,  $j = 2U + 1$  and  $j = 2U$ . For even  $L$  then  $2U = L + 2$  and consider cases  $j > 2U + 2$ ,  $j = 2U + 2$  and  $j = 2U + 1$ .

(ii) is proved in the same way as Theorem 1 of Kuang, Nielsen and Nielsen (2008a): First, consider the case where  $\nu_0^\dagger \neq \nu_0^\ddagger$ . Then  $\mu_{UU}^\dagger \neq \mu_{UU}^\ddagger$  so  $\mu^\dagger \neq \mu^\ddagger$ . Secondly, consider the case where  $\nu_0^\dagger = \nu_0^\ddagger$  but  $\nu_a^\dagger \neq \nu_a^\ddagger$ . Then  $\mu_{UU}^\dagger = \mu_{UU}^\ddagger$  but  $\mu_{U+1,U}^\dagger \neq \mu_{U+1,U}^\ddagger$  so  $\mu^\dagger \neq \mu^\ddagger$ . Thirdly, consider in the same way the case where  $\nu_0^\dagger = \nu_0^\ddagger$   $\nu_a^\dagger = \nu_a^\ddagger$  but  $\nu_c^\dagger \neq \nu_c^\ddagger$ . Fourthly, consider the case where  $\nu_0^\dagger = \nu_0^\ddagger$   $\nu_a^\dagger = \nu_a^\ddagger$   $\nu_c^\dagger = \nu_c^\ddagger$  but  $\Delta^2 \beta_{2U+1}^\dagger \neq \Delta^2 \beta_{2U+1}^\ddagger$ . Following Figure 1(b) then  $\Delta^2 \mu_{U+1,U+1}^\dagger = \Delta^2 \beta_{2U+1}^\dagger \neq \Delta^2 \beta_{2U+1}^\ddagger = \Delta^2 \mu_{U+1,U+1}^\ddagger$ . Since  $\mu$  takes the same value at the points  $(U, U)$ ,  $(U + 1, U)$  and  $(U, U + 1)$  then  $\mu_{U+1,U+1}^\dagger \neq \mu_{U+1,U+1}^\ddagger$ . Continue in the same way with the remaining parameters of (3.10).  $\square$ .

### A.2 Quadratic and cubic time effects

The quadratic expression (4.2) is a special case of the cubic expression (4.3). The formula (4.3) is zero for  $i = 1, I$ . Thus it suffices to show that  $\Delta^2 \alpha_i^{detrend} = a + b(i - 2)$ . Write out the expression for  $\alpha_i^{detrend}$  in (4.3) as

$$\alpha_i^{detrend} = \frac{a}{2} \{i^2 - (I+1)i + I\} + \frac{b}{6} [i^3 + \{(I-2) - (I+1)\}i^2 + \{I - (I+1)(I-2)\}i + I(I-2)].$$

Take second differences noting that linear trends have second difference zero while

$$\Delta^2 i^2 = i^2 - 2(i-1)^2 + (i-2)^2 = 2, \quad (\text{A.1})$$

$$\Delta^2 i^3 = i^3 - 2(i-1)^3 + (i-2)^3 = 6(i-1). \quad (\text{A.2})$$

## References

- Barndorff-Nielsen, O.E. (1978) *Information and Exponential Families*. Chichester: John Wiley & Sons.
- Berzuini, C. and Clayton, D. (1994) Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* 13, 823–838.

- Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine* 26, 3018–3045.
- Clayton, D. and Schifflers, E. (1987a) Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Statistics in Medicine* 6, 449–467.
- Clayton, D. and Schifflers, E. (1987b). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine* 6, 469–481.
- England, P.D. and Verrall, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal* 8, 519–544.
- Holford, T.R. (1985) An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases* 38, 831–836.
- Holford, T.R. (2006) Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine* 25, 977–993.
- Kuang, D., Nielsen, B. and Nielsen, J.P. (2008a) Identification of the age-period-cohort model and the extended chain ladder model. *Biometrika* 95, 979–986.
- Kuang, D., Nielsen, B. and Nielsen, J.P. (2008b) Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* 95, 987–991.
- Kuang, D., Nielsen, B. and Nielsen, J.P. (2009) Chain-Ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* 4, 105–121.
- Keiding, N. (1990) Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society, London A* 332, 487–509.
- Luo, L. (2013) Assessing validity and application scope of the intrinsic estimator approach of the age-period-cohort problem. *Demography* 50, 1945–1967.
- Martínez Miranda, M.D., Nielsen, B. and Nielsen, J.P. (2013) Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society A*. Early view.
- Nielsen, B. (2014) apc: A package for age-period-cohort analysis. R package version 1.0, <http://CRAN.R-project.org/package=apc>.
- Nielsen, B. and Nielsen, J.P. (2014) Identification and forecasting in mortality models. *The Scientific World Journal*. vol. 2014, Article ID 347043, 24 pages. doi:10.1155/2014/347043.
- O’Brien, R.M. (2011a) Constrained estimators and age-period-cohort models. *Sociological Methods & Research* 40, 419–452.
- Osmond, C. and Gardner, M.J. (1982) Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine* 1, 245–259.

- Wilkinson, G.N. and Rogers, C.E. (1973) Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society C*, 22, 392–399.
- Yang, Y., Fu, W.J. and Land, K.C. (2004) A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological methodology* 34, 75–110.