Nuffield College
UNIVERSITY OF OXFORD

ECONOMICS DISCUSSION PAPERS

Paper No: 2022-W02

Age-period-cohort analysis of mixed frequency data

By Bent Nielsen

September 2022
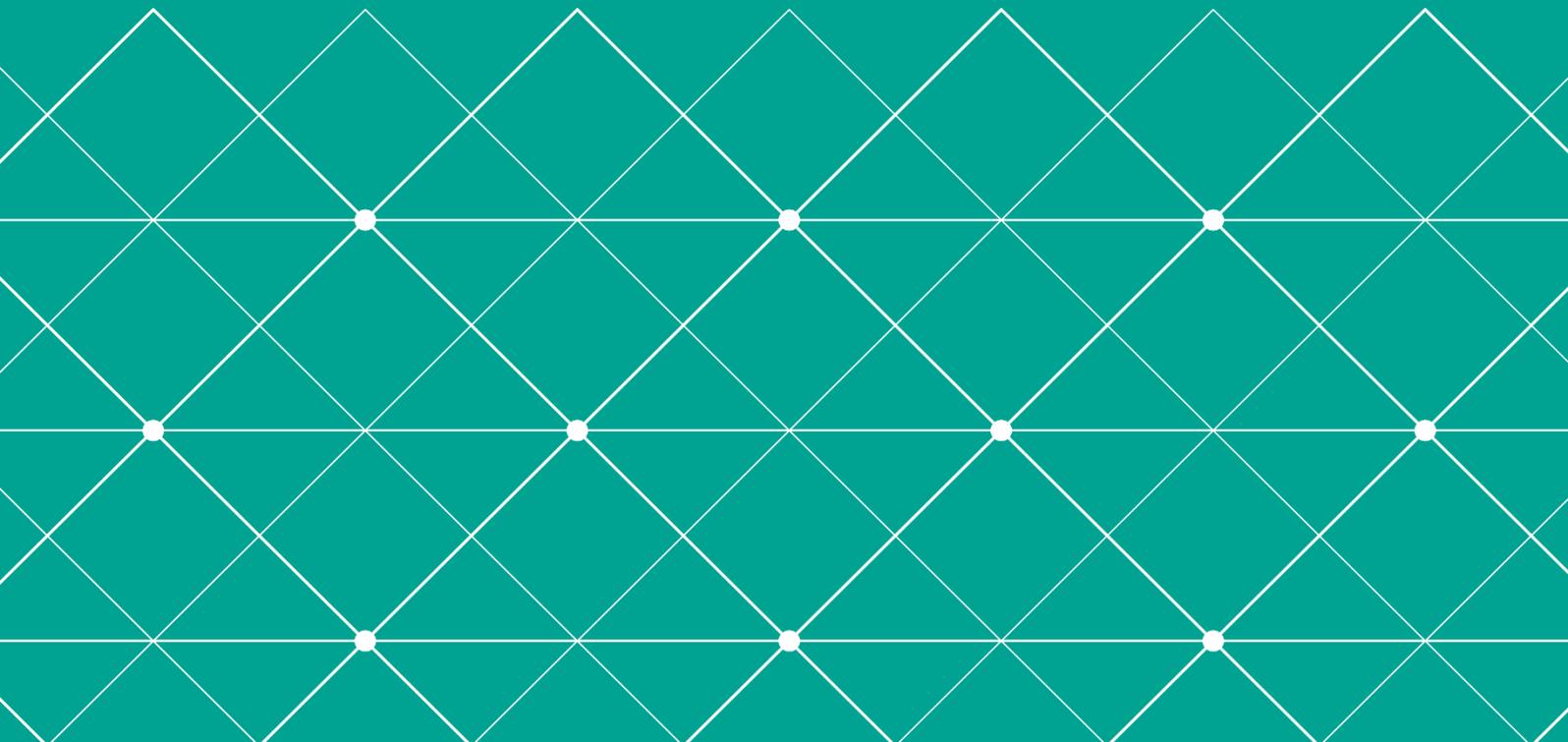
Age-period-cohort analysis of mixed frequency data

B. Nielsen[1]
28 September 2022

Abstract: The age-period-cohort model for mixed frequency data is analyzed. Mixed frequencies arise when the age and period scales have different frequencies. The implied cohort sequence then skips certain values. This is accounted for by reference to the coin problem in number theory. Further, the standard age-period-cohort identification problem is extended by modulo congruences for age-cohort and for period-cohort combinations. An invariant parametrization is found. Using that, standard methods for inference and forecasting apply. The analysis is illustrated with mesothelioma data.
Keywords: Age-period-cohort model, Canonical parametrization, Identification, Mixed-frequency data, Unequal intervals.

1   Introduction

Cohort data often have mixed frequencies, where outcomes are recorded by age and period at different frequencies. For instance, five-year age groups could be combined with one-year or three-year periods. Such groupings, or unequal intervals, influence the identification problem for age-period-cohort models as well as the sequence of possible cohort indices. Both problems are addressed. First, using features of the coin problem in number theory, the sequence of possible cohort indices is characterized. Second, the identification problem for mixed frequency data is described. Taken together, these findings allow a reparametrization of the age-period-cohort predictor in terms of freely varying identified parameters that are invariant to the identification problem. With this, it is possible to apply standard generalized linear models for inference and to match degrees of freedom with the dimension of the design matrix.

Mixed frequency data are common. The empirical illustration in this paper uses data on UK mesothelioma mortality which are published using five-year age groups and a one-year period. Holford (2006) considered lung cancer mortality among Californian women with five-year age and three-year period groups. Dinas & Stoker (2014) considered US presidential voting participation with ten-year age and four-year period groups. Riebler & Held (2010) considered data on chronic obstructive pulmonary disease in England and Wales with five-year age and one-year period groups.

There is a considerable literature on age-period-cohort models for regular data where age and period have the same units or frequency. As a starting point, these models have predictors written as linear combinations of age, period and cohort effects and an intercept. Such predictors are over-parametrized in that the time effects have four levels and three linear slopes, but only one level and two linear slopes can be identified. This means that the time effects can be changed by an arbitrary four-dimensional transformation involving three levels and a linear slope without changing the overall predictor

---

(Carstensen, 2007). Functions of the time effects that are not changed by such transformations are said to be invariant to the transformations.

The approaches to the identification problem are of three types. First, a popular approach is to introduce identifying but non-invariant constraints on the time effects (Yang & Land, 2013; Fu, 2018). Second, to apply the first approach for initial estimation and then focus on 'estimable', invariant parameters (Holford, 1983, 2006). Third, to reparametrize the model in terms of freely varying, invariant parameters (Kuang et al., 2008b). This reparametrization combines the unidentified, non-invariant levels and linear trends into an identified, invariant linear plane while the non-linear effects of the identification problem are captured through identified, invariant double-differences. This allows standard statistical methods to be used for estimation and inference without regard to the identification problem. The third approach is followed here. For reviews, see Smith & Wakefield (2016), Fannon & Nielsen (2019).

There is a much smaller literature on mixed frequency age-period-cohort models. Fienberg & Mason (1979) studied an example where the period unit is twice the age unit and noted that an additional identification constraint arises relatively to models for regular data. Holford (1983) revisited the small example and rewrote the age and cohort indices to have two components, a bi-annual index matching period and a year-within-two-year-period index. Moreover, the age effect was decomposed into four types of components, denoted group, linear, parallelism and curvature. Holford (2006) looked at a more complicated case with five-year age groups and and three-year period groups. Holford noted that this gives the additional problem that the cohort sequence skips some values. Gascoigne & Smith (2021) noted that quantities that are invariant for regular data arrays may not be invariant for mixed frequency arrays. However, as yet, neither the skipping in the cohort sequence nor the identification problem have been fully analyzed for mixed frequency data. Both problems are addressed here.

The skipping problem is similar to the coin problem in number theory: suppose we have two types of coins of different denomination, for instance 5 and 3. Then the monetary amounts 1, 2, 4 and 7 cannot be obtained. The numbers 1, 2, 4 and 7 match the gaps observed by Holford (2006). The number of skips and algorithms for finding their location are discussed. The present analysis builds on Ramírez Alfonsín (2005).

The identification problem is characterized. It turns out that relevant transformations are the traditional linear transformations for regular data along with modulo congruences for age-cohort and for period-cohort combinations. Thus, in the example of five-year age groups and three-year period groups we have the usual 4 linear constraints as well as $5 - 1 = 4$ binding congruences for period-cohort combinations and $3 - 1 = 2$ binding congruences for age-cohort combinations. This gives a total of 10 constraints.

The dimensions of the skipping and the transformations determine the degrees of freedom. The estimable parameters are those combinations of the time effects that are invariant to the transformations. With this, the age-period-cohort predictor is reparametrized in terms of freely varying, identified, invariant parameters. This generalizes the reparametrization of Martínez Miranda et al. (2015) for regular age-period arrays and builds on the theory of Kuang et al. (2008b).

Other recent developments include Luo & Hodges (2016) who argued that identification by constraints should be done cautiously for mixed frequency data. Riebler & Held

(2010) suggested a Bayesian smoothing procedure, where identification issues are less obvious (Nielsen & Nielsen, 2014). Fu (2018) suggested the Intend-to-Collapse method. When period are observed more frequently than age, the cohort is defined by tentatively collapsing period to the same frequency as age.

The mixed frequency problem is superficially related to the overlapping cohort problem. This is the problem that a person, who is age 10 at some point during the year 2015 could have been born in 2004 or in 2005. In its detail, that problem appears to be distinct from the present discussion (Osmond & Gardner, 1989, Carstensen, 2007).

The empirical illustration uses UK mesothelioma mortality. Mesothelioma is a disease with long latency and typically caused by exposure to asbestos. Only little is known about the magnitude of the future burden. It is therefore of considerable interest to forecast future case numbers. The publically available data have five year age-groups and one-year periods. Data at annual frequency can been modelled using a Poisson model (Martínez Miranda et al., 2015). The grouping of the data appears to introduce some over-dispersion. Using asymptotic theory of Harnau & Nielsen (2018) it is possible to conduct inference and derive distribution forecasts. The statistical tests indicate that an age-cohort model suffices in line with earlier studies of regular data. For illustration, forecasting from a sub-sample with triennial periods is also considered. Despite using relatively little data, the mixed-frequency age-cohort model holds up quite well compared with methods that use annual data.

Outline of the paper: Section 2 shows the mixed-frequency data structure and links to the coin problem. Section 3 analyzes the age-period-cohort predictor, describes the over-parametrization and gives an invariant parametrization. Section 4 has the empirical illustration. Section 5 concludes. Technical derivations are given in an appendix.

2 Data structure and time scales

The data can be an age-period array of rates, counts, or doses and responses. We will introduce notation for age, period and cohort indices that allows for grouping of age and period and as a consequence also of cohort.

2.1 The age and period time scales

For mixed-frequency data arrays, we have A_G age groups of length G and P_H period groups of length H. The data array is regular when $G = H = 1$. We will assume that the largest common divisor of G and H is unity. If we have groups with a common divisor larger than unity, such as 10 and 4, we can scale by the common divisor of 2. We will refer to each observation by the highest relevant age and period. We denote the highest values of age and period by $A = A_G G$ and $P = P_H H$. Indexing age backwards and period forwards, we get the index set

$$\mathcal{I}_{age,per}: \quad \begin{cases} age = A - gG & \text{where} \quad g = 0, 1, \ldots, A_G - 1, \\ per = H + hH & \text{where} \quad h = 0, 1, \ldots, P_H - 1. \end{cases} \quad (1)$$

Typically, age data are aggregated by age. Aggregation is often due to a small sample size. Period data are typically either aggregates over a period interval or cross

3

period		age									
real	per	real A-age	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89
		40	*35*	*30*	*25*	*20*	*15*	*10*	*5*	*0*	
1985-87	*3*		43	38	33	28	23	18	13	8	3
1988-90	*6*		46	41	36	31	26	21	16	11	6
1991-93	*9*		49	44	39	34	29	24	19	14	9
1994-96	*12*		52	47	42	37	32	27	22	17	12
1997-99	*15*		55	50	45	40	35	30	25	20	15
2000-02	*18*		58	53	48	43	38	33	28	23	18
2003-05	*21*		61	56	51	46	41	36	31	26	21
2006-08	*24*		64	59	54	49	44	39	34	29	24
2009-11	*27*		67	62	57	52	47	42	37	32	27
2012-14	*30*		70	65	60	55	50	45	40	35	30
2015-17	*33*		73	68	63	58	53	48	43	38	33
2018-20	*36*		76	71	66	61	56	51	46	41	36

Table 1: Cohort indices for $G = 5$ year age groups and $H = 3$ year period groups.

sections at particular points in time. While this has implications for interpretation of the adopted indices, it does not change any technical aspects.

Table 1 illustrates the mixed frequency data structure. This is a modified version of the Holford (2006) example. As indicated in italics, age runs from 45 to 89 in groups of 5, while period runs from 1985 to 2020 in groups of 3. Thus, the index limits are $A = 45$, $G = 5$ and $A_G = 9$ for age and $P = 36$, $H = 3$ and $P_H = 12$ for period. Macro-blocks of dimension $GH = 15$ years are indicated with dashed lines.

2.2 The cohort indices

Cohorts are defined according to the convention

$$coh = per + A - age. \tag{2}$$

With the notation (1), we can write $coh = H + gG + hH$ for $g = 0, 1, \ldots, A_G - 1$ and $h = 0, 1, \ldots, P_H - 1$. Thus, the smallest and largest cohort values over $\mathcal{I}_{age,per}$ are H and $C = A + P - G$.

We return to Table 1. Here, the age and period values are shown in italics. The possible cohort values range from $H = 3$ to $A + P - G = 76$. We note that the cohort values 4, 5, 7, 10 and 69, 72, 74, 75 are skipped. These values are 3 plus 1, 2, 4, 7 and 76 minus 1, 2, 4, 7. The skipping leaves 66 possible cohort values.

The skipping problem is similar to coin problem in number theory. If we have coins of two denominations, G, H and one of the coins is unity, we can form any positive monetary amount. But, suppose the coins satisfy $G, H \geq 2$ with largest common denominator of unity. Which are the monetary amounts that can be obtained from those coins? That is, which combinations $gG + hH$ can be formed for non-negative integers g, h? We will need the following results, see the Appendix for details.

The Frobenius number $F_{G,H} = GH - G - H$ is the largest number that cannot be represented as a non-negative integer combination. Sylvester pointed out that the

number of non-representable numbers is $S_{G,H} = (G-1)(H-1)/2$. The set $\mathcal{N}_{G,H}$ of non-representable numbers is described in (32) in the Appendix.

Translating the coin problem into the mixed frequency problem, we see that the set of possible cohort values is

$$\mathcal{I}_{coh} = (H, \ldots, C) \backslash (H + c, C - c : c \in \mathcal{N}_{G,H}). \tag{3}$$

This set includes an unbroken cohort sequence from $H + F_{G,H} + 1$ to $C - F_{G,H} - 1$. The total number of possible cohorts is $C - (H - 1) - 2S_{G,H}$, which equals $A + P - GH$.

When we later identify the cohort parameters, we will look at those cohorts in the array $\mathcal{I}_{age,per}$ that remain when dropping indices with age A and period H. This corresponds to dropping the first row and last column in Table 1. The smallest possible cohort is then $H + G + H$, while the largest possible cohort is C as before. The Frobenius and Sylvester numbers are not changed. Thus, this smaller set of cohorts is

$$\mathcal{I}_{coh}^{\circ} = (G + 2H, \ldots, C) \backslash (G + 2H + c, C - c : c \in \mathcal{N}_{G,H}). \tag{4}$$

The set $\mathcal{I}_{coh}^{\circ}$ has $C - (2H + G - 1) - 2S_{G,H} = A + P + 1 - (G + 1)(H + 1)$ elements.

Table 1 illustrates these results. Here $G = 5$, $H = 3$, $A = 45$, $P = 36$. The coin problem gives $F_{5,3} = 7$ and $S_{5,3} = 4$, while $\mathcal{N}_{5,3} = (1, 2, 4, 7)$. The sets $\mathcal{I}_{coh}$ and $\mathcal{I}_{coh}^{\circ}$ have unbroken sequence from 11 to 68 and from 19 to 68, respectively. For instance,

$$\mathcal{I}_{coh} = (3, 6, 8, 9, 11, \ldots, 68, 70, 71, 73, 76). \tag{5}$$

3   Age-period-cohort predictor

We now consider the mixed frequency age-period-cohort predictor of the form

$$\mu_{age,per} = \alpha_{age} + \beta_{per} + \gamma_{coh} + \delta \qquad \text{for } age, per \in \mathcal{I}_{age,per}. \tag{6}$$

It is well-known that the model is over-parametrized in the regular case. On the right hand side of (6), we have A_G ages, P_G periods, $A + P - GH$ cohorts and the intercept. Thus, the dimension of these over-parametrized time effects is

$$q_{G,H} = A_G + P_H + A + P - GH + 1. \tag{7}$$

3.1   Characterizing the over-parametrization

The over-parametrization of the age-period-cohort equation (6) has two sources. First, we have the usual constraints from the regular case. Second, Fienberg & Mason (1979), point to additional constraints with mixed frequencies. We find the full set of constraints.

For regular data arrays, it is known that the time effects on the right hand side of (6) indicate 4 levels and 3 linear slopes, but only one level and two linear slopes can be identified. This comes about from the identity (2) linking the age, period and cohort indices, so that the predictor on the left hand side of (6) is unchanged when making certain linear transformations of the time effects on the right hand side of (6). This

can be summarized as an invariance property of the predictor with respect to a four dimensional transformation of the time effects of the form

$$\mu_{age,per} = \{\alpha_{age} + a + d \times (A - age)\} + (\beta_{per} + b + d \times per)$$
$$+ (\gamma_{coh} + c - d \times coh) + (\delta - a - b - c), \quad (8)$$

for any values of a, b, c, d (Carstensen, 2007). For the regular case, this characterizes all possible transformations of the time effects on the right hand side that do not change the predictor on the left hand side (Kuang et al., 2008b). For the mixed frequency case, these constraints continue to apply, but this is no longer the full set of constraints.

The additional transformations can be characterized through modulo operations. Recall that two integers i, j are congruent modulo m, if m divides their difference $i - j$ and we write $i \equiv j \mod m$. By assumption, H divides per and G divides age and A. Thus, the relation $coh = per + A - age$ from (2) gives the congruences

$$coh \equiv per \mod G, \quad (9)$$
$$coh \equiv A - age \mod H. \quad (10)$$

The first relation divides cohorts and periods into G classes, thus giving $G-1$ constraints. Similarly, the second relation gives $H - 1$ constraints. The congruences are only non-constraining when $G = H = 1$. When cohort and period are congruent, a constant can be added to the cohort effect and subtracted from the period effect, while leaving their sum unaffected. This implies unidentified seasonal patterns in the age, period and cohort effects. The seasonal pattern relates to the micro-effects noted by Holford (2006). Riebler & Held (2010) have the construction (9) for the case where $H = 1$.

We summarize the above constraints. That is, in the mixed frequency model, the predictor is invariant to time effect transformations of the form

$$\mu_{age,per} = \left\{\alpha_{age} + a + d \times (A - age) + \sum_{i=1}^{H-1} e_i 1_{(A-age \equiv i \mod H)}\right\}$$
$$+ \left\{\beta_{per} + b + d \times per + \sum_{j=1}^{G-1} f_j 1_{(per \equiv j \mod G)}\right\}$$
$$+ \left\{\gamma_{coh} + c - d \times coh - \sum_{i=1}^{H-1} e_i 1_{(coh \equiv i \mod H)} - \sum_{j=1}^{G-1} f_j 1_{(coh \equiv j \mod G)}\right\}$$
$$+ (\delta - a - b - c), \quad (11)$$

for any values of $a, b, c, d, e_1, \ldots, e_{H-1}, f_1, \ldots, f_{G-1}$. As yet, it has not been argued that this gives a complete description of the over-parametrization.

The dimension of the transformations in (11) is $G + H + 2$. Subtracting this from $q_{G,H}$ defined in (7) indicates that the dimension of the parameter space is

$$p_{G,H} = q_{G,H} - G - H - 2 = A_G + P_H + A + P - (G+1)(H+1). \quad (12)$$

6

3.2 Double differenced time effects

Double differenced time effects have a log-odds-ratio interpretation (Fienberg & Mason, 1979; Martínez Miranda et al., 2015). For regular data arrays such double differences are invariant to the transformations in (8). It is therefore possible to parametrize the predictor invariantly in terms of a linear plane combined with double differences of the time effects (Kuang et al., 2008b; Fannon & Nielsen, 2019). Gascoigne & Smith (2021) note that the double differences are not in general invariant in the mixed frequency situation. Using the mixed frequency transformations in (11), it is possible to explain their observation in more detail and to derive an alternative invariant parametrization.

Period double differences are defined as follows. Let Δ_s denote an s-step difference operator, giving first differences $\Delta_H \beta_{per} = \beta_{per} - \beta_{per-H}$ and second differences $\Delta_H^2 \beta_{per} = \Delta_H \beta_{per} - \Delta_H \beta_{per-H}$, so that $\Delta_H^2 \beta_{per} = \beta_{per} - 2\beta_{per-H} + \beta_{per-2H}$.

Why is the period double difference $\Delta_H^2 \beta_{per}$ only invariant to (11) if $G = 1$? First, we note, that the double difference is invariant to the linear transformations in (8) as for regular data. Second, the double difference is invariant to the micro effect stemming from the age grouping in (9), if the three time points per, $per - H$ and $per - 2H$ are congruent modulo G. Since 1 is the largest common divisor of G and H, this can only happen when $G = 1$. This matches the observation of Gascoigne & Smith (2021).

In the mixed frequency case, we must form the double difference in another way. We will be able to identify the following period double differences:

$$\Delta_{GH}\Delta_H \beta_{per} = \beta_{per} - \beta_{per-H} - \beta_{per-GH} - \beta_{per-H-GH}.$$

These double differences reduce to the quantity $\Delta_H^2 \beta_{per}$ considered above when $G = 1$. We argue for invariance with respect to transformations in (11) as follows. The linear transformation (8) is eliminated by the double differencing. Now, the micro effect in (9) is eliminated since we have congruence modulo G of the time points per and $per - GH$ as well as of $per - H$ and $per - GH - H$. Alternatively, we can argue by expressing the double differences in terms of the invariant predictor through

$$\Delta_{GH}\Delta_H \beta_{per} = \mu_{age,per} - \mu_{age,per-H} - \mu_{age-GH,per-GH} + \mu_{age-GH,per-GH-H} \qquad (13)$$

for $per = H + hH$ with $h = G + 1, \dots, P_H - 1$ and arbitrary age.

In a similar fashion, we can identify the age double differences

$$\Delta_{GH}\Delta_G \alpha_{age} = \alpha_{age} - \alpha_{age-G} - \alpha_{age-GH} - \alpha_{age-G-GH} \qquad (14)$$

for $age = A - gG$ with $g = 0, 1, \dots A_G - H - 2$. For the cohorts, we can identify

$$\Delta_G \Delta_H \gamma_{coh} = \gamma_{coh} - \gamma_{coh-G} - \gamma_{coh-H} - \gamma_{coh-G-H}. \qquad (15)$$

for $coh \in \mathcal{I}_{coh}^\circ$ as defined in (4).

3.3 Invariant parametrization of linear planes

In the mixed frequency case, we will need $G + H + 1$ points to anchor a set of linear planes. Equivalently, we can choose an overall level and a set of slopes. A convenient choice of the overall level is

$$\mu_{A,H} = \alpha_A + \beta_H + \gamma_H + \delta, . \qquad (16)$$

7

The slopes can be chosen as

$$\lambda_g = \mu_{A-gG,H} - \mu_{A,H} = \Delta_{gG}\gamma_{H+gG} - \Delta_{gG}\alpha_A \qquad \text{for } g = 1, \ldots H, \qquad (17)$$

$$\nu_h = \mu_{A,H+hH} - \mu_{A,H} = \Delta_{hH}\gamma_{H+hH} + \Delta_{hH}\beta_{H+hH} \qquad \text{for } h = 1, \ldots G. \qquad (18)$$

The above parameters are invariant functions of the original time effects since they are functions of the invariant predictor on the left hand side of (11).

We note that it is not possible to separate the slopes in (17), (18) into individual slopes for age, period and cohort. The cohort slope is unavoidably entangled with the age slope through (17) and with the period slope through (18). This property mimicks what is known for regular data arrays. Thus, λ_g are age-cohort slopes while ν_h are period-cohort slopes.

We collect the above identified double differences and plane parameters in the vector

$$\xi = \big(\mu_{A,H}; \ \lambda_1, \ldots, \lambda_H; \ \nu_1, \ldots, \nu_G; \ \Delta_{GH}\Delta_G\alpha_{A-gG} \text{ for } g = 0, 1, \ldots, A_G - H - 2;$$
$$\Delta_{GH}\Delta_H\beta_{hH} \text{ for } h = G + 1, \ldots, P_H - 1; \ \Delta_G\Delta_H\gamma_{coh} \text{ for } coh \in \mathcal{I}_{coh}^\circ\big). \quad (19)$$

We will represent the predictor as a linear function of this invariant parameter. We will refer to this invariant parameter as the canonical parameter, borrowing a terminology from exponential family theory (Sundberg, 2019). It has dimension $p_{G,H}$ matching (12).

3.4 Representation

We can now parametrize the predictor in terms of the invariant canonical parameter. To write down the representation, we follow Holford (2006) and express the age, period and cohort time scales through the Euclidean representations $g = q_g H + r_g$ and $h = q_h G + r_h$ for $q_g, q_h \geq 0$ while $0 \leq r_g < H$ and $0 \leq r_h < G$, see (33), (34). Thus,

$$age = A - q_g GH - r_g G, \qquad per = H + q_h GH + r_h H. \qquad (20)$$

The implies that the cohort satisfies

$$coh = H + (q_g + q_h)GH + r_g G + r_h H. \qquad (21)$$

We first present an intermediate result, where, for each value of the Euclidean remainders, the representation corresponds to that of Martínez Miranda et al. (2015). Next, in Theorem 3.1 below, we show that the predictor can be expressed a linear function of the canonical parater. For each pair of Euclidean remainders, the intermediate representation involves a linear plane and double sums of double differences of the three age-period-cohort time effects. It is derived in the appendix and has the form

$$\mu_{A-q_g GH - r_g G, H+q_h GH + r_h H}$$
$$= M_{r_g, r_h}^{intercept} + q_g M_{r_g, r_h}^{age/coh} + q_h M_{r_g, r_h}^{per/coh} + S_{q_g, r_g}^{age} + S_{q_h, r_h}^{per} + S_{q_g+q_h, r_g, r_h}^{coh}. \quad (22)$$

Here, the intercept and slopes can be expressed as follows

$$M_{r_g, r_h}^{intercept} = \mu_{A-r_g G, H+r_h H}, \qquad (23)$$

$$M_{r_g, r_h}^{age/coh} = \mu_{A-GH-r_g G, H+r_h H} - \mu_{A-r_g G, H+r_h H}, \qquad (24)$$

$$M_{r_g, r_h}^{per/coh} = \mu_{A-r_g G, H+GH+r_h H} - \mu_{A-r_g G, H+r_h H}. \qquad (25)$$

Thus, for each pair of Euclidean remainders, the intercept and slopes are expressed in terms of three anchoring points. The slopes also satisfy

$$M_{r_g,r_h}^{age/coh} = \Delta_{GH}(\gamma_{H+GH+r_gG+r_hH} - \alpha_{A-r_gG}), \tag{26}$$

$$M_{r_g,r_h}^{per/coh} = \Delta_{GH}(\gamma_{H+GH+r_gG+r_hH} + \beta_{H+GH+r_hH}). \tag{27}$$

In particular, $M_{r_g,r_h}^{age/coh}$ involves a combination of first differences of the age and cohort effects. These first differences are not individually identified. Thus, $M_{r_g,r_h}^{age/coh}$ should be interpreted as a combined age-cohort slope. In a similar fashion, $M_{r_g,r_h}^{per/coh}$ is a period-cohort slope. The S-terms are sums of double differences

$$S_{q_g,r_g}^{age} = 1_{(q_g \geq 2)} \sum_{t=1}^{q_g-1} \sum_{s=0}^{tH-1} \Delta_{GH}\Delta_G \alpha_{A-(r_g+s)G}, \tag{28}$$

$$S_{q_h,r_h}^{per} = 1_{(q_h \geq 2)} \sum_{t=1}^{q_h-1} \sum_{s=1}^{tG} \Delta_{GH}\Delta_H \beta_{H+(G+r_h+s)H}, \tag{29}$$

$$S_{q_g+q_h,r_g,r_h}^{coh} = 1_{(q_g+g_h \geq 2)} \sum_{t=1}^{q_g+q_h-1} \sum_{s=1}^{tG} \sum_{u=1}^{H} \Delta_G\Delta_H \gamma_{H+(r_g+u)G+(r_h+s)H}. \tag{30}$$

We note that the representation (22) has intercept and slope expressed in terms of the predictor at particular points as well as double sums of double differences. Thus, this representation is identified and invariant to the transformations in (11). However, this representation is not expressed directly in terms of the canonical parameter. At this point, we can therefore not assess the dimension of the variation of the predictor and therefore degrees of freedom.

The representation (22) depends on the particular choice of Euclidean remainders. Each representation has increments on a grid of 'macro-steps' that are GH steps apart. If we set both remainders to, for instance, zero, we get a reference representation corresponding to the 'macro-categories' of Holford (2006). For other values of the Euclidean remainders one could take differences to the reference value to describe micro-steps.

The representation in Martínez Miranda et al. (2015, Theorem 1) for the regular case arises from (22) with $G = H = 1$ so that the Euclidean remainders are zero.

We now proceed to represent the predictor in terms of the canonical parameter.

Theorem 3.1. *Let μ be the collection of age-period-cohort predictors $\mu_{age,per}$ of the form (6) over the mixed frequency array, $age, per \in \mathcal{I}_{age,per}$. Then*
(a) ξ is a linear function of μ that is invariant to the transformations in (11);
(b) μ is a linear function of ξ given by (22) combined with (48), (49) and (50);
(c) The parameter ξ is exactly identified in that $\xi^\dagger \neq \xi^\ddagger$ implies $\mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$.

Theorem 3.1 confirms that the variation of the predictor over the full index set has the same dimension as the canonical predictor. This dimension is $p_{G,H}$ as given in (12). Moreover, the set of transformations in (11) restricts the dimension q_{GH} of the time effects precisely by $q_{GH} - p_{GH}$ constraints resulting in a p_{GH} dimensional variation. We have that the canonical parameter is on the one hand invariant to these transformations

and on the other hand in a one-one relation with the predictor. Thus, the canonical parameter is a maximal invariant function of the time effects under the transformations in (11). This generalizes the findings for regular arrays in Kuang et al. (2008b), see also Cox & Hinkley (1974, §5.3) for a general reference.

The representation of the predictors in terms of the canonical parameters reveal a perfect fit property for the two macro blocks formed from the smallest H ages and largest G periods and from the largest H ages and smallest G periods. This is argued in the appendix and was pointed out by Riebler & Held (2010) for the case where $H = 1$. For general $G, H \geq 2$, the cohort values of the perfectly fitted macro blocks are not in sequence because of the coin problem feature.

3.5   Submodels

In practice, not all of the age-period-cohort components may be needed. This can be investigated by imposing parameter restrictions. We can formulate the restrictions in two ways. First, the restrictions can be formulated in terms of the time effects entering the original formulation of the predictor in (6). The time effects are subject to the constraints in (11). This must be considered when computing degrees of freedom. Alternatively, the restrictions can be formulated in terms of the canonical parameters in (19). As these parameters are freely varying the degrees of freedom count simply matches the dimension of the restriction.

As an example, consider the *age-cohort* model $\mu_{age,per} = \alpha_{age} + \gamma_{coh} + \delta$. In terms of the time effect formulation (6), the restriction is $\beta_{H+hH} = 0$ for $h = 0, 1, \ldots, P_H - 1$. At first glance this points to a P_H-dimensional restriction. However, the time effects are not individually identified due to the $G + 1$ constraints shown in (11). The restriction is therefore equivalent to a restriction on the identified, freely-varying non-linear effects $\Delta_{GH}\Delta_H \beta_{H+hH} = 0$ for $h = G+1, \ldots, P_H - 1$ given in (13). The argument is the same as applied for regular data, see Holford (1983), Nielsen & Nielsen (2014), Martínez Miranda et al. (2015), Fannon & Nielsen (2019).

In a similar way, the *period-cohort* model $\mu_{age,per} = \beta_{per} + \gamma_{coh} + \delta$ arises by restricting $\Delta_{GH}\Delta_G \alpha_{A-gG} = 0$ for $g = 0, 1, \ldots, A_G - H - 2$. And, the *age-period* model $\mu_{age,per} = \alpha_{age} + \beta_{per} + \delta$ corresponds to restricting $\Delta_G \Delta_H \gamma_{cov} = 0$ for $coh \in \mathcal{I}_{cov}^\circ$, which has dimension $A + P + 1 - (G+1)(H+1)$.

An *age-drift* model arises when setting both the period and cohort double differences to zero, that is $\Delta_{GH}\Delta_H \beta_{H+hH} = 0$ for $h = G + 1, \ldots, P_H - 1$ and $\Delta_G \Delta_H \gamma_{cov} = 0$ for $coh \in \mathcal{I}_{cov}^\circ$. The dimension of the restriction is then the sum of $P_H - 1 - G$ and $A + P + 1 - (G+1)(H+1)$. This does not identify the linear age slope as the linear planes remain unrestricted. For further discussion, see Clayton & Schifflers (1987).

A pure *age* model $\mu_{age,per} = \alpha_{age} + \delta$ arises when setting period and cohort time effects to zero. The age model is a sub-model of the age-drift model where we restrict the period-cohort slopes through $\nu_h = 0$ for $h = 1, \ldots, G$. The degrees of freedom relative to the full age-period-cohort model is then $A + P + P_H - (G+1)(H+1)$.

4   Analysis of the UK mesothelioma data

Mesothelioma is a cancer that is typically caused by exposure to asbestos fibres. It has a long latency period and it is rapidly fatal once discovered. The number of male deaths in Great Britain appear to have peaked in 2016 with 2205 death of all ages and 2117 deaths at age under 90. Forecasting the future burden of mesothelioma deaths remains of interest from a public health viewpoint and for general insurers.

4.1   The data

The Health and Safety Executive (HSE), Britain's regulator for workplace health and safety, provides data on https://www.hse.gov.uk/statistics/tables/#lung, with annual updates. The publically available data has mixed frequency with age in 5 year groups while period is annual. The July 2022 data release covers the years 1968–2020. Here, we consider data for males with age between 25 and 89.

Mesothelioma has a long latency period. This can be seen from the distribution of responses: (0.3%, 6.4%, 19.4%, 52.7%, 19.3%, 1.9%) for ages (0-39, 40-54, 55-64, 65-79, 80-89, 90+). The average age by period has been gradually increasing with period in line with improved health regulation and health care. This implies that there are only few responses for ages below 40 and for the oldest cohorts.

Previous studies had access to regular data with a one-year frequency for both age and period (Isakson et al., 2021, Mammen et al., 2021, Martínez Miranda et al., 2015, 2016). Here, the publically available mixed-frequency data is used.

The HSE publishes distribution forecasts by year until 2050. The most recent forecasts use data until 2017. These forecasts use a multinomial dose-response model where the doses are based on an epidemiological model for exposure (Peto et al., 1995; Hodgson et al., 2005; Tan et al., 2010). The latter paper presents distribution forecasts using a Bayesian implementation. Here we use a response-only model as this removes the uncertainty around epidemiological modelling of doses and suffices for forecasting.

4.2   Analysis with annual period

We consider data with five year age groups, $G = 5$, and annual period, $H = 1$.

The baseline model is a Poisson model with an age-period-cohort predictor, but no measure for exposure. Inference and distribution forecasts can then be based on a multinomial sampling scheme obtained by conditioning on the total number of deaths and an asymptotic distribution theory arising for large expected values of the total number of deaths (Martínez Miranda et al., 2015). This approach appears to be successful for modelling UK mesothelioma when annual data are available.

The aggregation appears to give a slight over-dispersion. Thus, we use methods for inference and distribution forecasting for an over-dispersed Poisson model as developed in Harnau & Nielsen (2018). The counts of cases, $Y_{age,per}$, are assumed independent over age and period with non-negative, infinitely divisible distributions with finite third moment. Examples of such distributions include Poisson, compound Poisson, negative binomial, log normal, gamma and generalized gamma convolutions. The log expectation $\log \mathsf{E}(Y_{age,per}) = \mu_{age,per}$ is assumed to have age-period-cohort structure, while variance

and expectation are proportional so that $\mathsf{Var}(Y_{age,per})/\mathsf{E}(Y_{age,per}) = \sigma^2 > 0$. The Poisson model has $\sigma^2 = 1$. For the asymptotic argument, we assume $\tau = \sum_{age,per} \mathsf{E}(Y_{age,per})$ is large while $\mathsf{E}(Y_{age,per})/\tau$ is fixed and the skewness of $Y_{age,per}$ vanishes.

model	deviance	df	dev/df	F vs. apc	df vs. apc	p_F
apc	725.08	517	1.40			
ap	11009.69	624	17.64	68.53	107	0.000
ac	801.30	564	1.42	1.16	47	0.228
pc	9994.26	528	18.93	600.83	11	0.000

Table 2: Deviance analysis of mesothelioma data with annual period.

Table 2 reports a deviance analysis. We consider the full age-period-cohort (apc) model and sub-models (ap, ac, pc), where one of the three effects is dropped. There are $A_G = 13$ age groups and $P = P_H = 53$ periods, giving $n = 689$ observations. In light of (12), the APC model has 172 parameters and thus 517 degrees of freedom.

We allow over-dispersion because the deviance for the apc model is large relative to a χ^2_{517} distribution. The deviance divided by the degrees of freedom is an estimate of the over-dispersion. The sub-models can be tested through an F-statistic of the form $F_{ap} = \{(dev_{ap} - dev_{apc})/(df_{ap} - df_{apc})\}/\{dev_{apc}/(n - df_{apc})\}$, which are asymptotically $\mathsf{F}_{df_{ap}-df_{apc},n-df_{apc}}$ distributed (Harnau & Nielsen, 2018). We see that reduction to the age-cohort specification cannot be rejected while the age-period and period-cohort models are rejected. This conclusion is in line with previous studies using regular data (Martínez Miranda et al., 2015, 2016).

Figure 1 shows detrended time effects. Detrending has been done with three objectives in mind: to emphasize non-linearity, to ensure invertibility of the covariance of the estimators and to disentangle the plots. With fewer constraints, such as only imposing slope constraints on one of the plots, the three plots become linked inextricably and should be considered jointly (Carstensen, 2007). Thus, more constraints are used here (Nielsen, 2015; Fannon & Nielsen, 2019).

Figure 1(a) shows the detrended age effect. The transformation (11) with $H = 1$ shows that the age effect has an arbitrary level and linear slope. Thus, the age effect has two constraints in that it starts and ends in zero, so as to emphasize the non-linearity. The first observation at age 29 is for the interval 25-29. The age effect is nearly the same for the two lower age groups and then it takes a clear concave pattern. This is common for epidemiological studies. The dashed lines indicate twice the point-wise standard errors. This is zero at points where constraints are imposed. The age effects are larger than their standard errors supporting the rejection that the non-linear age effect is zero.

Figure 1(b) shows the detrended period effect. The effect has an arbitrary level, linear slope and $G - 1 = 4$ micro levels according to (11). The macro trend, shown with bullets at $GH = 15$ year intervals is detrended to start and end in zero. The micro levels starting $3, 6, 9, 12$ are demeaned to start on the line between the first two macro estimates as indicated with crosses. The periods effects are smaller than two standard errors supporting that the setting the non-linear period effect to zero cannot be rejected.

Figure 1(c) shows the detrended cohort macro effect. The detrending is done so that third estimates from the beginning and from the end are set to zero. Thus, the standard
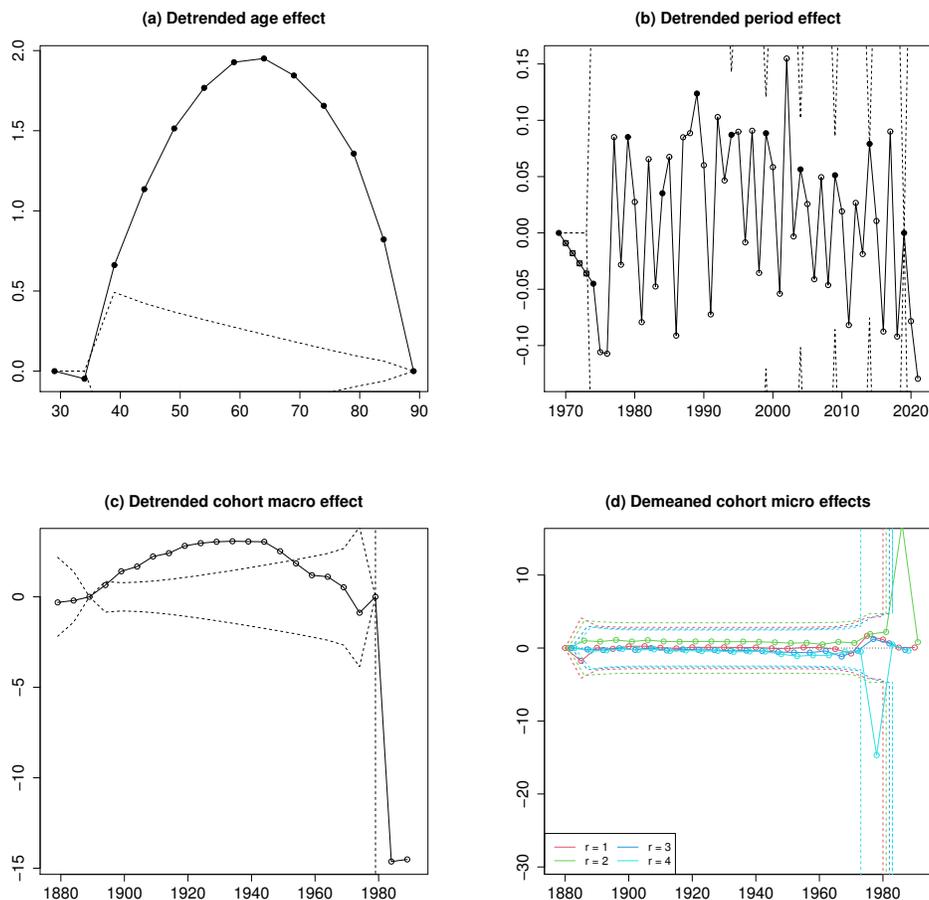
Figure 1: Detrended time effects for annual period data.

errors are zero in those points. The reason for this is that the extreme cohort values are associated with the top right and bottom left macro block of the data matrix as illustrated in Table 1. Those macro blocks are only repeated once in the data and are therefore subject to small sample issues. These issues are worsened by very low and some volatile responses. Indeed, it is seen that the standard errors become very large for the youngest cohorts. If the detrending had been done by setting the extremes to zero this volatility would dominate the figure. The cohort macro effect follows a concave pattern. This is consistent with asbestos being increasingly used for early cohorts and then less used as health legislation was brought in.

Figure 1(d) shows the demeaned cohort micro effect. Four micro effects arise by looking at time points shifted by $r = 1, 2, 3, 4$ years relative to those for the macro effects. The micro effects are found by taking difference between the representations at times $qG + r$ and at times $qG + 0$. These micro effects are then demeaned to start in zero. We see that the micro effects is relatively small throughout. This indicates that cohort effect on an annual scale would be fairly smooth with only a modest seasonality as in Holford (2006). By focusing on the macro effects, smoothing is avoided. For a smoothing approach, see Gascoigne & Smith (2021). Again the standard errors indicate

model	deviance	df	dev/df	F vs. apc	df vs. apc	p_F
apc	116.77	66	1.77			
ap	3615.16	144	25.11	25.35	78	0.000
ac	129.83	77	1.69	0.67	11	0.760
pc	2120.57	72	29.45	188.76	6	0.000

Table 3: Deviance analysis of mesothelioma data with triennial period.

volatility for the recent cohorts with sparse responses.

4.3 Analysis with triennial period

To explore the more complicated situation where both frequencies are larger than one, we set $H = 3$ and consider the mestheolioma data for the periods 2020, 2017, 2014, etc. We then have $G = 5$ and $H = 3$. The low number of responses for those aged below 40 generates some instability when analyzing the data with triennial period. Thus, we drop the the three lower age groups. Thus, there are $A_G = 10$ age groups and $P_H = 17$ periods, giving $n = 170$ observations. In light of (12), the APC model has 104 parameters and thus 66 degrees of freedom.

Table 3 reports a deviance analysis for the mesothelioma data. The over-dispersion is now stronger. As before, the age-cohort specification cannot be rejected.

Figure 2 is set up as before. With $G = 5$ and $H = 3$, the macro steps are now $GH = 15$, marked with bullets in panels (a,b). We identify $H - 1 = 2$ micro levels for age and, as before, $G - 1 = 4$ micro levels for period. These are marked with crosses in panels (a,b). For the cohort, each combination of $r_g G + r_h H$ generates as different trend. The value for $r_g = r_h = 0$ is taken as the macro effect in (c), whereas demeaned differences are shown for the remaining 14 combinations in (d).

In panels (a,b) the detrended age and cohort macro effects have significant, concave shapes. In panel (b) the detrended period effect has an insignificant, concave shape. The concavity can also be discerned from Figure 2(b). In panel (d) the demeaned micro effects are drifting downwards as $r = r_g G + r_h H$ increases. This is to be expected as the difference to the references macro effect of $r = 0$ is expected to be larger for larger values of r. For some larger values of r, the micro effect is significant for the more recent cohort values, which is consistent with sparse, volatile responses.

Overall, the impression for the triennial data is similar although noisier than for the annual period data. However, data had to be trimmed as estimates and standard errors are more influenced by sparse, volatile responses. This could be due to the relatively more parameters in the triennial model.

4.4 Forecasting the UK mesothelioma burden

We proceed with forecasting the UK mesothelioma burden using the age-cohort models. The periods were found not to be significant and, indeed, when forecasting, parsimonious models are often preferable. This is in line with earlier studies (Martínez Miranda et al., 2015, 2016). We will focus on forecasting those cohorts that are observed in the sample.
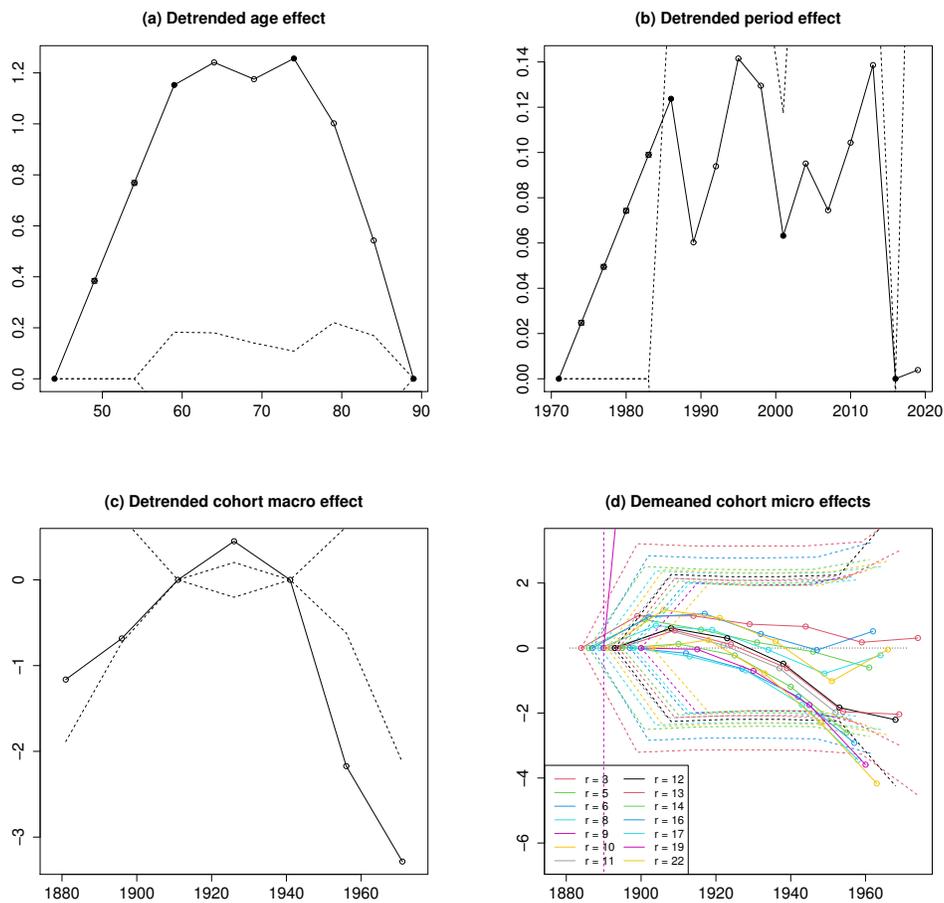
Figure 2: Detrended time effects for triennial period data.

With an age-cohort model, parameters need not be extrapolated. This contrasts with an age-period-cohort model where the period parameters would have to be extrapolated.

Forecasts are done using four different age-cohort models. All models have 10 age groups covering ages 40-89 and period starting in 1982. By truncating the data for the lower age groups and lower periods observations with low counts are avoided. The counts for ages 40-89 cover 98-99% of counts for those of age less than 90. The four variations of the models are that period ends in either 2017 or 2020 and the period frequency H is either one or three.

The forecasts are done 15 periods ahead for the ages 55-89. This is to include only in-sample cohorts when working with the triennial data. For that situation Table 4 indicates how the macro blocks of cohorts are replicated in the cohort period. The macro blocks have dimension $GH = 15$.

Point forecasts are constructed by rolling the representation in (22) forward in time with period parameters set to zero. As such they are functions of the canonical parameter and therefore invariant to the transformations in (11), see Kuang et al. (2008a). Intercept corrections are used (Martínez Miranda et al., 2015). Distribution forecasts for the over-dispersed Poisson model can be added following Harnau & Nielsen (2018).

	age									
	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89
period	*45*	*40*	*35*	*30*	*25*	*20*	*15*	*10*	*5*	*0*
2002 *33*	78	73	68	63	58	53	48	43	38	33
2005 *36*	81	76	71	66	61	56	51	46	41	36
2008 *39*	84	79	74	69	64	59	54	49	44	39
2011 *42*	87	82	77	72	67	62	57	52	47	42
2014 *45*	90	85	80	75	70	65	60	55	50	45
2017 *48*	93	88	83	78	73	68	63	58	53	48
2020 *51*	96	91	86	81	76	71	66	61	56	51
2023 *54*				84	79	74	69	64	59	54
2026 *57*				87	82	77	72	67	62	57
2029 *60*				90	85	80	75	70	65	60
2032 *63*				93	88	83	78	73	68	63
2035 *63*				96	91	86	81	76	71	66
2038 *63*							84	79	74	69

Table 4: Cohort indices for $G = 5$ year age groups and $H = 3$ year period groups.

	AC-2020		AC-2017		HSE-2017
	$H = 1$	$H = 3$	$H = 1$	$H = 3$	
2021	1918 [1808,2027]		1914 [1806,2022]		1943 [1829,2073]
2022	1881 [1774,1989]		1877 [1771,1984]		1885 [1767,2022]
2023	1811 [1706,1917]	1836 [1617,2055]	1872 [1753,1992]	1865 [1669,2061]	1822 [1700,1965]
2024	1739 [1636,1842]		1787 [1787,1902]		1753 [1629,1902]
2025	1771 [1668,1875]		1804 [1671,1919]		1678 [1551,1832]
2026	1720 [1607,1833]	1641 [1440,1842]	1721 [1688,1833]	1665 [1485,1844]	1599 [1470,1758]

Table 5: Forecasts with 90% confidence bands.

Table 5 shows the four age-cohort forecasts. The fifth forecast is produced by the Health and Safety Executive (HSE) using annual age-period data until 2017. The forecasts are accompanied by 90% forecast standard errors.

Figure 3 shows the same forecasts taken up to 2032. The figure shows all five types of point forecasts. The two age-cohort forecasts using data until 2020 are accompanied by 90% forecast standard errors shown as shaded regions. The age-cohort forecasts broadly follow the same trajectory. The annual period, 2017 forecast tracks the first three out-of-sample observations quite well. The triennial, 2017 forecast is a bit off for 2020. This is possibly because it uses very little information from the period after the peak in 2016. Figure 3 indicates that the triennial forecast has must wider forecasts standard errors than the annual forecast. The annual forecasts are somewhat volatile. The triennial forecast appears smoother, but this is an artefact of having only three out-of-sample forecasts in the figure and they happen to lie more or less on a line.
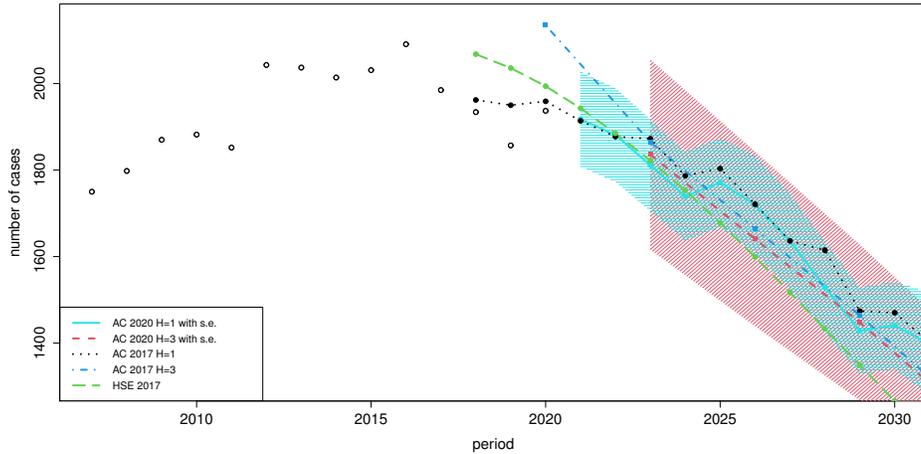
Figure 3: Forecasts of annual numbers of mesothelioma deaths.

The HSE forecast is based on the Hodgson et al. (2005), which uses a dose-response age-cohort model, where doses are constructed from epidemiological knowledge with respect to time since exposure, age at exposure and lung clearance. Tan et al. (2010) embedded this in a Bayesian set-up to achieve distribution forecasts. This forecast is more smooth than the age-cohort forecasts as it uses annual age and period data. It uses observations for all ages less than 90 and should perhaps be scaled down by 1-2% to be fully comparable with the age-cohort forecasts. It is interesting to note that the Bayesian forecast bands are wider than the annual-period age-cohort bands and nearly as wide at the triennial-period age-cohort bands despite using far more data.

Overall, the forecasts follow broadly the same trajectory. There is a tendency that case numbers come down more steeply with the HSE forecast than the age-cohort forecats. It is remarkable that the forecasts using five-year age groups and three-year period groups gives forecasts that are very similar to the other forecasts, even though it uses far fewer observations. For evaluation of historical forecasts by the HSE and the age-cohort method see Martínez-Miranda et al. (2016).

5   Conclusion

The age-period-cohort model for mixed frequency data was analyzed. The skipping in the cohort is linked to the coin problem in number theory. The identification problem was fully described. The linear relation $coh = per + A - age$ gives the usual problem of identifying levels and linear trends in an age-period-model remains. In addition, frequency related congruences between cohort and period and between cohort and age give additional problems of identifying 'micro' levels. A parametrization that is invariant to both identification problems was suggested. With that parametrization, standard statistical methods can be used.

The method was applied to counts of mesothelioma deaths with five-year age groups and one or three-year period groups. Inferences and forecasts are similar to those ob-

tained from annual, regular data in earlier studies. The grouping of the data appears to give some over-dispersion. The age-period-cohort model reduces to an age-cohort model, which is convenient for forecasting. The grouping does not give seem to give particular biases when forecasting, but grouping does increase forecast variance.

The identification problem implies that standard double differences of the time effects are in general not identified. For instance, if the data have 5-year age groups and annual periods groups, then the five-year difference of the annual difference of the period effect is identified, whereas the annual double difference of the period effect is not. If a Bayesian approach is adopted it may not be a good idea to put priors on the latter (Smith & Wakefield, 2016; Gascoigne & Smith, 2021). With the present analysis a more informed Bayesian analysis can be done if so desired.

A   The Coin problem and technical details

The coin problem Suppose there are coins of two denominations, $G, H \geq 2$ with largest common denominator of unity. Which combinations $gG + hH$ can be formed for non-negative integers g, h?

The Frobenius number $F_{G,H} = GH - G - H$ is the largest number that cannot be expressed as a non-negative integer combination of G, H (Ramírez Alfonsín, 2005, Theorem 2.1.1). The number of positive numbers that cannot be expressed as a non-negative integer combination of G, H is $S_{G,H} = (G-1)(H-1)/2$, as pointed out by Sylvester in 1882 (Ramírez Alfonsín, 2005, Theorem 5.1.1).

The set of non-representable numbers is the complement of the representable numbers not exceeding the Frobenius number. Those representable numbers can be expressed as a triangular array with non-repeated, non-ordered elements. If $\lfloor x \rfloor$ denotes the floor of a real x, that is the largest integer not exceeding x, we get

$$\mathcal{R}_{G,H} = \left(gG + hH : 0 \leq g \leq \left\lfloor \frac{F_{G,H} - 1}{G} \right\rfloor, 0 \leq h \leq \left\lfloor \frac{F_{G,H} - 1 - gG}{H} \right\rfloor \right). \qquad (31)$$

Details are given below. Recently, Binner (2021) has given an alternative representation. The set of non-representable numbers is then

$$\mathcal{N}_{G,H} = (0, 1, 2, \ldots, F_{G,H}) \backslash \mathcal{R}_{G,H}. \qquad (32)$$

Thus, the Frobenius number is the largest number in $\mathcal{N}_{G,H}$, while the Sylvester result gives the number of elements in $\mathcal{N}_{G,H}$. If $G = 1$ or $H = 1$ then $\mathcal{N}_{G,H}$ is the empty set.

As an illustration, let $G = 5$ and $H = 3$. The largest skipped number is the Frobenius number $F_{5,3} = 7$. The number of skipped values is $S_{5,3} = 4$. The set of representable numbers less than $F_{5,3}$ is $\mathcal{R}_{5,3} = (0, 3, 6, 5)$. Thus, the set of non-representable numbers is $\mathcal{N}_{5,3} = (1, 2, 4, 7)$.

Next, how many ways can a monetary amount $m = gG + hH$ be represented? Given $g, h \geq 0$ we apply Euclidean division to get

$$g = q_g H + r_g \quad \text{where} \quad q_g = \lfloor g/H \rfloor \quad \text{so that} \quad 0 \leq q_g \text{ and } 0 \leq r_g < H, \qquad (33)$$
$$h = q_h G + r_h \quad \text{where} \quad q_h = \lfloor h/G \rfloor \quad \text{so that} \quad 0 \leq q_h \text{ and } 0 \leq r_h < G. \qquad (34)$$

Insert these representations into $m = gG + hH$ to get

$$m = (q_g H + r_g)G + (q_h G + r_h)H = (q_g + q_h)GH + r_g G + r_h H. \tag{35}$$

Here, the remainders $0 \le r_g < H$ and $0 \le r_h < G$ and the total number $q = q_g + q_h$ of GH macro-coins are unique. The q macro-coins can be decomposed in $q + 1$ ways in terms of G and H coins, corresponding to the macro-blocks in Table 1 that are repeated along diagonals from top left to bottom right. The top right block has elements of the form $r_g G + r_h H$. Details are given below.

Algorithm for representing a representable number m in terms of coins. From (35), we know that there is a unique representation

$$m = qGH + r_g G + r_h H. \tag{36}$$

We will have that either $0 \le r_g G + r_h H < GH$ or $GH < r_g G + r_h H < 2GH$. We first check the former case. Euclidean division gives $m = \tilde{q}GH + \tilde{r}$ with $\tilde{q} = \lfloor m/(GH) \rfloor$ and $\tilde{r} = m - \tilde{q}GH$. We then check if $\tilde{r} \in \mathcal{S}_{G,H}$ where

$$\mathcal{S}_{G,H} = (gG + hH : 0 \le g < H, 0 \le h < G) \tag{37}$$

by running through all options. If $\tilde{r} \in \mathcal{S}_{G,H}$, we get $\tilde{r} = r_g G + r_h H$ and $q = \tilde{q}$. If $\tilde{r} \notin \mathcal{S}_{G,H}$ then we must have $\tilde{r} + GH \in \mathcal{S}_{G,H}$. Running through all options, we get $\tilde{r} + GH = r_g G + r_h H$ and $q = \tilde{q} - 1$. Details are given below.

Derivation of the set of representable numbers in (31). We argue that the representable numbers $x = iG + jH$ satisfying $0 \le x < GH$ can be expressed as a unique combination of i, j so that $0 \le i < H$ and $0 \le j < G$. First, note that neither of $i = H$ and $j = G$ is possible, since then $x \ge GH$. Second, suppose $(i', j') \ne (i, j)$ exists so that $x = i'G + j'H$. Subtracting the two representations of x we get $(i - i')G + (j - j')H = 0$. Since G divides 0 but not H, then G must divide $j - j'$. When $0 \le j, j' < G$ this is only possible if $j = j'$. Divide by H to conclude $i = i'$.

Next, we consider the set of representable numbers $\mathcal{R}_{G,H}$ less than the Frobenius number $F_{G,H} = GH - G - H$. For $j = 0$, we must have $0 \le iG \le F_{G,H} - 1$ or $0 \le i \le \lfloor (F_{G,H} - 1)/G \rfloor$. For any such i we must have $j \ge 0$ and $iG + jH \le F_{G,H} - 1$, so that $0 \le j \le \lfloor (F_{G,H} - 1 - iG)/H \rfloor$. The mentioned i, j combinations constitute the set $\mathcal{R}_{G,H}$ in (31). By the above argument, the elements in $\mathcal{R}_{G,H}$ are non-repeated as the i, j pairs are unique.

Uniqueness of the representation (35). Equation (35) is derived from the representations (33) and (34). In those representations, $r_g G$ and $r_h H$ are unique. As $r_g G$ cannot be represented in H coins and $r_h H$ cannot be represented in G coins, then r_g and r_h are unique in (35). However, the amount GH can be represented as G coins with value H or as H coins with value G. Thus, we can decompose $(q_g + q_h)GH$ in $q_g + q_h + 1$ different ways in terms of G and H coins.

Derivation of the representation (36). For a general representable number m, the representation $m = qGH + r_g G + r_h H$ in (35) applies with unique $q \ge 0$, $0 \le r_g < H$ and $0 \le r_h < G$. We will either have $0 \le r_g G + r_h H < GH$ or $GH < r_g G + r_h H < 2GH$,

while $r_g G + r_h H = GH$ is infeasible given the constraints to r_g, r_h (Ramírez Alfonsín, 2005, Remark 5.1.2). We can find the values of q, r_g, r_h as follows. Euclidian division gives $m = \tilde{q}GH + \tilde{r}$, where $\tilde{q} = \lfloor m/(GH) \rfloor$ and $\tilde{r} = m - \tilde{q}GH$ satisfying $0 \leq \tilde{r} < GH$. We check if $\tilde{r} \in \mathcal{S}_{G,H}$, see (37), which gives a representation $\tilde{r} = r_g G + r_h H$ and $q = \tilde{q}$. If $\tilde{r} \notin \mathcal{S}_{G,H}$, then we must have that $\tilde{r} + GH \in \mathcal{S}_{G,H}$ and we run through the options in $\mathcal{S}_{G,H}$ to find a representation $\tilde{r} + GH = r_g G + r_h H$ so that $q = \tilde{q} - 1$.

Derivation of the intermediate representation in (22). We modify the argument of Martínez Miranda et al. (2015, §A.1). We write, with $q = q_g + q_h$,

$$\mu_{A-q_g GH - r_g G, H + q_h GH + r_h H} = \alpha_{A-q_g GH - r_g G} + \beta_{H + q_h GH + r_h H} + \gamma_{H + qGH + r_g G + r_h H} + \delta. \quad (38)$$

We analyze the first three terms in turn and then combine.

For the period effect, let $H^* = H + r_h H$. Use telescoping sums to get $\beta_{H^* + q_h GH} = \beta_{H^*} + \sum_{t=0}^{q_h - 1} \Delta_{GH} \beta_{H^* + GH + tGH}$ for levels with $q_h \geq 1$ whereas $\Delta_{GH} \beta_{H^* + GH + tGH} = \Delta_{GH} \beta_{H^* + GH} + \sum_{s=1}^{tG} \Delta_{GH} \Delta_H \beta_{H^* + GH + sH}$ for differences with $t \geq 1$. Combine to get

$$\beta_{H^* + q_h GH} = \beta_{H^*} + q_h \Delta_{GH} \beta_{H^* + GH} + 1_{(q_h \geq 2)} \sum_{t=1}^{q_h - 1} \sum_{s=1}^{tG} \Delta_{GH} \Delta_H \beta_{H^* + GH + sH}. \quad (39)$$

For the age effect, let $A^* = A - r_g G$. Apply backward telescoping sums to get $\alpha_{A^* - q_g GH} = \alpha_{A^*} - \sum_{t=0}^{q_g - 1} \Delta_{GH} \alpha_{A^* - tGH}$ for levels with $q_g \geq 1$ and $\Delta_{GH} \alpha_{A^* - tGH} = \Delta_{GH} \alpha_{A^*} - \sum_{s=0}^{tH-1} \Delta_{GH} \Delta_G \alpha_{A^* - sG}$ for differences with $t \geq 1$. Combine to get

$$\alpha_{A^* - q_g GH} = \alpha_{A^*} - q_g \Delta_{GH} \alpha_{A^*} + 1_{(q_g \geq 2)} \sum_{t=1}^{q_g - 1} \sum_{s=0}^{tH-1} \Delta_{GH} \Delta_G \alpha_{A^* - sG}. \quad (40)$$

For the cohort effect, let $H^\dagger = H + r_g G + r_h H$. Proceed as for the period effect in (39) to get $\gamma_{H^\dagger + qGH} = \gamma_{H^\dagger} + q\Delta_{GH} \gamma_{H^\dagger + GH} + 1_{(q \geq 2)} \sum_{t=1}^{q-1} \sum_{s=1}^{tG} \Delta_{GH} \Delta_H \gamma_{H^\dagger + GH + sH}$. Write $\Delta_{GH} \Delta_H \gamma_{H^\dagger + GH + sH} = \sum_{u=1}^{H} \Delta_G \Delta_H \gamma_{H^\dagger + uG + sH}$. Combine to get

$$\gamma_{H^\dagger + qGH} = \gamma_{H^\dagger} + q\Delta_{GH} \gamma_{H^\dagger + GH} + 1_{(q \geq 2)} \sum_{t=1}^{q-1} \sum_{s=1}^{tG} \sum_{u=1}^{H} \Delta_G \Delta_H \gamma_{H^\dagger + uG + sH}. \quad (41)$$

Note that in the triple sum, the summands have indices in $\mathcal{I}_{coh}^\circ$.

Collecting the intercepts in (39), (40), (41) gives

$$M_{r_g, r_h}^{intercept} = \alpha_{A^*} + \beta_{H^*} + \gamma_{H^\dagger} + \delta = \mu_{A - r_g G, H + r_h H}. \quad (42)$$

Collecting the slopes in (39), (40), (41) noting that $q = q_g + g_h$ gives

$$M_{r_g, r_h}^{age/coh} = \Delta_{GH} \gamma_{H^\dagger + GH} - \Delta_{GH} \alpha_{A^*}$$

$$= \mu_{A - GH - r_g G, H + r_h H} - \mu_{A - r_g G, H + r_h H}, \quad (43)$$

$$M_{r_g, r_h}^{per/coh} = \Delta_{GH} \gamma_{H^\dagger + GH} + \Delta_{GH} \beta_{H^* + GH} \mu_{A - r_g G, H + r_h H}$$

$$= \mu_{A - r_g G, H + GH + r_h H} - \mu_{A - r_g G, H + r_h H}. \quad (44)$$

Thus, inserting all the above derivations into (38), we get the expression (22).

Perfect fit property. The macro block for the G smallest ages and H largest periods has unique cohort values. Thus, the design matrix X can be organized as

$$X = \begin{pmatrix} X_{11} & 0 \\ X_{21} & X_{22} \end{pmatrix}, \tag{45}$$

where X has $A_G P_H$ rows, the last GH rows $X_2 = (X_{21}, X_{22})$ correspond to the observations in the macro block and X_{22} is an invertible GH-dimensional matrix. Let

$$L = \begin{pmatrix} I_{A_G P_H - GH} & 0 \\ -X_{22}^{-1} X_{21} & I_{GH} \end{pmatrix}, \tag{46}$$

which has full rank. Then XL is block-diagonal with diagonal blocks X_{11} and X_{22}. This reveals the perfect fit property for the last GH observations.

The macro block for the G largest ages and H smallest periods is given by $M_{r_g,r_h}^{intercept}$ in (42). We can organize the design matrix to be of the form (45), but now with $X_{11} = M_{r_g,r_h}^{intercept}$. We need to show that the lower part $X_2 = (X_{21}, X_{22})$ has a column rank deficiency of GH so that we can post multiply X_2 with a lower triangular (2×2) block matrix L so that $X_2 L = (0, \tilde{X}_{22})$. Consider the representation (22), noting that for the rows in X_2 then $q_g + q_h \geq 1$. For each value of r_g, r_h, we demonstrate a linear dependence between the coefficient in four columns of X_2. The intercept $M_{g,h}^{intercept}$ and the slopes $M_{g,h}^{age/coh}$ and $M_{g,h}^{per/coh}$ have coefficients 1, q_g and q_h for $g = r_g$ and $h = r_h$ and a zero coefficient for $g \neq r_g$ with $0 \leq g < H$ and $h \neq r_h$ with $0 \leq h < G$. Further, reorganize the cohort double sum in (30) as

$$S_{q_g+q_h,r_g,r_h}^{coh} = 1_{(q_g+q_h \geq 2)} \sum_{t=1}^{q_g+q_h-1} \sum_{v=1}^{t} \Delta_{GH} \Delta_{GH} \gamma_{H+(v+1)GH+r_gG+r_hH}. \tag{47}$$

Choosing $v = 1$, we get that $\Delta_{GH} \Delta_{GH} \gamma_{H+2GH+gG+hH}$ has coefficient $q_g + q_h - 1$ for $g = r_g$ and $h = r_h$ and a zero coefficient otherwise. The coefficients 1, q_g, q_h, $q_g + q_h - 1$ multiplied by 1, -1, -1, 1 add up to zero.

Proof of Theorem 3.1.

(a) In §3.2, 3.3 it was argued that ξ is a function of μ that is invariant to (11).

(b) We express μ as a linear function of ξ. Rewrite the intercept in (42) as

$$\begin{aligned} M_{r_g,r_h}^{intercept} &= \mu_{A-r_gG,H+r_hH} = \alpha_{A-r_gG} + \beta_{H+r_hH} + \gamma_{H+r_gG+r_hH} + \delta \\ &= \left\{ \alpha_A - 1_{(r_g>0)} \Delta_{r_gG} \alpha_A \right\} + \left\{ \beta_H + 1_{(r_h>0)} \Delta_{r_hH} \beta_{H+r_hH} \right\} \\ &\quad + \left\{ \gamma_H + 1_{(r_g>0)} \Delta_{r_gG} \gamma_{H+r_gG} + 1_{(r_h>0)} \Delta_{r_hH} \gamma_{H+r_hH} \right. \\ &\quad \left. + 1_{(r_g>0)} 1_{(r_h>0)} \Delta_{r_gG} \Delta_{r_hH} \gamma_{H+r_gG+r_hH} \right\} + \delta \end{aligned}$$

Collecting terms, noting that $\mu_{A,H} = \alpha_A + \beta_H + \gamma_H + \delta$, while $\lambda_{r_g} = \Delta_{r_gG}(\gamma_{H+r_gG} - \alpha_A)$ and $\nu_{r_h} = \Delta_{r_hH}(\gamma_{H+r_hH} + \beta_{H+r_hH})$, see (16), (17), (18), we get

$$M_{r_g,r_h}^{intercept} = \mu_{A,H} + 1_{(r_g>0)} \lambda_{r_g} + 1_{(r_h>0)} \nu_{r_g} + 1_{(r_g>0)} 1_{(r_h>0)} \Delta_{r_gG} \Delta_{r_hH} \gamma_{H+r_gG+r_hH}. \tag{48}$$

For slopes, note that $q = q_g + q_h$, so that the cohort slope can be associated in parts with the period and and in part with the age slopes. Combine the period slopes as

$$M_{r_g,r_h}^{per/coh} = \Delta_{GH}\gamma_{H+GH+r_gG+r_hH} + \Delta_{GH}\beta_{H+GH+r_hH}$$
$$= \Delta_{GH}\big(\gamma_{H+GH+r_hH} + \beta_{H+GH+r_hH}\big) + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_gG}\gamma_{H+GH+r_gG+r_hH}$$

Add and subtract $\nu_G = \Delta_{GH}(\beta_{H+GH} + \gamma_{H+GH})$ to get

$$M_{r_g,r_h}^{per/coh} = \nu_G + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_gG}\gamma_{H+GH+r_gG+r_hH}$$
$$+ 1_{(r_h>0)}\Delta_{GH}\Delta_{r_hH}\big(\gamma_{H+GH+r_hH} + \beta_{H+GH+r_hH}\big). \quad (49)$$

The age-cohort slopes combine similarly using $\lambda_H = \Delta_{GH}(\gamma_{H+GH} - \alpha_A)$ to get

$$M_{r_g,r_h}^{age/coh} = \Delta_{GH}\gamma_{H+GH+r_gG+r_hH} - \Delta_{GH}\alpha_{A-r_gG}$$
$$= \Delta_{GH}\big(\gamma_{H+GH+r_gG} - \alpha_{A-r_gG}\big) + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_hH}\gamma_{H+GH+r_gG+r_hH}$$
$$= \lambda_H + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_hH}\gamma_{H+GH+r_gG+r_hH}$$
$$+ 1_{(r_g>0)}\Delta_{GH}\Delta_{r_gG}\big(\gamma_{H+GH+r_gG} + \alpha_A\big). \quad (50)$$

The expressions for intercept and slopes in (48), (49) and (50) are expressed in terms of Δ_{r_gG}, Δ_{r_hH} and Δ_{GH} differences. Such differences can be rewritten as sums of Δ_G and Δ_H differences. For instance, in (48), we can write

$$\Delta_{r_gG}\Delta_{r_hH}\gamma_{H+r_gG+r_hH} = \sum_{s=1}^{r_gG}\sum_{t=1}^{r_hH}\Delta_G\Delta_H\gamma_{H+sG+tH}, \quad (51)$$

which is a linear transformation of G,H double differences of $\gamma_{H+sG+tH}$. Since $s,t \geq 1$, these double differences have index in $\mathcal{I}_{coh}^{\circ}$ and thus form a part of ξ in (19).

(c) We show that ξ is exactly identified along the lines of Kuang et al. (2008b). We iterate over the coordinates of ξ listed in (19).

First, consider the top-right macro block where $A - age = gG$ for $0 \leq g < G$ and $per = H + hH$ for $0 \leq h < H$. This has GH coordinates of μ defined from (48) through $\mu_{A-r_gG,H+r_hH} = M_{r_g,r_h}^{intercept}$ with $r_g = g$ and $r_h = h$. Now, $M_{r_g,r_h}^{intercept}$ is a bijective linear function of the GH-vector $\mu_{A,H}, \lambda_{r_g}, \nu_{r_h}, \Delta_{r_gG}\Delta_{r_hH}\gamma_{H+r_gG+r_hH}$ for $0 < g = r_g < G$ and $0 < h = r_h < H$. By appealing to (51), we can identify combinations $\Delta_G\Delta_H\gamma_{H+r_gG+r_hH}$ for $0 < g = r_g < G$ and $0 < h = r_h < H$. Thus, the vector $\mu_{A,H}, \lambda_{r_g}, \nu_{r_h}, \Delta_G\Delta_H\gamma_{H+r_gG+r_hH}$ is identified.

Second, consider the two macro blocks where $q_g + q_h = 1$. Given that $M_{r_g,r_h}^{intercept}$ is dealt with above, we can focus on the slopes. Here, ν_G can be identified from $M_{0,0}^{per/coh}$ in (49), while λ_H can be identified from $M_{0,0}^{age/coh}$ in (50). From (49), we get that $M_{r_g,r_h}^{per/coh} - M_{0,r_h}^{per/coh} = \Delta_{GH}\Delta_{r_gG}\gamma_{H+GH+r_gG+r_hH}$ for $q_g = 1$, $q_h = 0$ and $r_g > 0$. Similarly, (50) gives $M_{r_g,r_h}^{age/coh} - M_{r_g,0}^{age/coh} = \Delta_{GH}\Delta_{r_hH}\gamma_{H+GH+r_gG+r_hH}$ for $q_g = 0$, $q_h = 1$ and $r_h > 0$. By appealing to (51), we can identify combinations $\Delta_G\Delta_H\gamma_{H+GH+r_gG+r_hH}$ for $0 < g = r_g < G$ and $0 < h = r_h < H$. Now, (49) with $q_g = 1$, $q_h = r_g = 0$ gives $M_{0,r_h}^{per/coh} - M_{0,r_h-1}^{per/coh} = \Delta_{GH}\Delta_H(\gamma_{H+GH+H} + \beta_{H+GH+H})$ for $r_h = 1$ and $M_{0,r_h}^{per/coh} - M_{0,r_h-1}^{per/coh} = \Delta_{GH}\Delta_H\Delta_{r_hH}(\gamma_{H+GH+r_hH} + \beta_{H+GH+r_hH})$ for $r_h > 1$. This allows

identification of $\Delta_{GH}\Delta_H\beta_{H+GH+r_hH}$ for $0 < r_h < H$. Similarly, using (50) and $M_{r_g,r_h}^{age/coh}$ we can identify $\Delta_{GH}\Delta_G\alpha_{H+GH+r_gG}$ for $0 < r_g < G$.

Continue in a similar fashion by increasing $q = q_g + q_h$ by one step and use the associated macro blocks to identify the next set of parameters.

References

Binner, D. S. (2021). A complete list of all numbers not of the form ax+by. Technical report, arXiv:2110.01558v1.

Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, *26*, 3018–3045.

Clayton, D. & Schifflers, E. (1987). Models for tempral variation in cancer rates. II Age-period-cohort models. *Statistics in Medicine*, *6*, 469–481.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Dinas, E. & Stoker, L. (2014). Age-period-cohort analysis: A design-based approach. *Electoral Studies*, *33*, 28–40.

Fannon, Z. & Nielsen, B. (2019). Age-period-cohort models. In *Oxford Research Encyclopedia, Economics and Finance*. Oxford University Press. DOI: 10.1093/acrefore/9780190625979.013.495.

Fienberg, S. E. & Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology*, *10*, 1–67.

Fu, W. (2018). *A Practical Guide to Age-Period-Cohort Analysis*. Boca Raton, FL: CRC Press.

Gascoigne, C. & Smith, T. (2021). Using smoothing splines to resolve the curvature identifiability problem in age-period-cohort models with unequal intervals. Technical report, arXiv:2112.08299v1.

Harnau, J. & Nielsen, B. (2018). Over-dispersed age-period-cohort models. *Journal of the American Statistical Association*, *113*, 1722–1732.

Hodgson, J. T., McElvenny, D. M., Darnton, A. J., Price, M. J., & Peto, J. (2005). The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *British Journal of Cancer*, *92*, 587–593.

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

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine*, *25*, 977–993.

Isakson, A., Krummaker, S., Martínez-Miranda, M. D., & Rickayzen, B. (2021). Calendar effect and in-sample forecasting applied to mesothelioma mortality data. *Mathematics*, *9*, 1–17. Article 2260.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008a). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*, 987–991.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008b). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*, 979–986.

Luo, L. & Hodges, J. S. (2016). Block constraints in age-period-cohort models with unequal-width intervals. *Sociological Methods & Research*, *45*, 700–726.

Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P., & Vogt, M. (2021). Calendar effect and in-sample forecasting. *Insurance: Mathematics and Economics*, *96*, 31–52.

Martínez Miranda, M. D., Nielsen, B., & Nielsen, J. P. (2015). Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society, Series A*, *178*, 29–55.

Martínez-Miranda, M. D., Nielsen, B., & Nielsen, J. P. (2016). A simple benchmark for mesothelioma projection for great britain. *Occupational and Environmental Medicine*, *73*, 561–563.

Nielsen, B. (2015). apc: An R package for age-period-cohort analysis. *R Journal*, *7*, 52–64.

Nielsen, B. & Nielsen, J. P. (2014). Identification and forecasting in mortality models. *The Scientific World Journal*, *2014*, Article ID 347043, 24 pages.

Osmond, C. & Gardner, M. J. (1989). Age, period, and cohort models: Non-overlapping cohorts don't resolve the identification problem. *American Journal of Epidemiology*, *129*, 31–35.

Peto, J., Hodgson, J. T., Matthews, F. E., & Jones, J. R. (1995). Continuing increase in mesothelioma mortality in Britain. *Lancet*, *345*, 535–539.

Ramírez Alfonsín, J. L. (2005). *The diophantine Frobenius problem*. Oxford: Oxford University Press.

Riebler, A. & Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics*, *11*, 57–69.

Smith, T. R. & Wakefield, J. (2016). A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science*, *31*, 591–610.

Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge: Cambridge University Press.

Tan, E., Warren, N., Darnton, A. J., & Hodgson, J. T. (2010). Projection of mesothelioma mortality in Britain using bayesian methods. *British Journal of Cancer*, *103*, 430–436.

Yang, Y. & Land, K. C. (2013). *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. New York: CRC Press.