Nuffield College
UNIVERSITY OF OXFORD

ECONOMICS DISCUSSION PAPERS

Paper No: 2022-W03

Two-sample age-period-cohort models with an application to Swiss suicide rates

By Bent Nielsen

October 2022

Two-sample age-period-cohort models with an application to Swiss suicide rates

B. Nielsen[1]
17 October 2022

Abstract: A two-sample age-period-cohort model is analyzed. Levels and linear trends are not fully identified. The identification problem is characterized and the model is reparametrized in terms of parameters that are invariant to the identification problem. It is shown how the identification problem changes when restricting some of the time effects to be common across samples. The analysis allows for mixed frequencies where age and period scales are grouped in different ways and can be implemented in a generalized linear model framework. The methodology is applied to Swiss suicide rates.
Keywords: Age-period-cohort model, Canonical parametrization, Identification, Invariance, Mixed-frequency data, Two-samples.

1 Introduction

We revisit a study of Swiss suicide rates (Riebler et al., 2012). The data consists of rates by age, period and gender. Suicide rates tend to increase with age and to be higher for men than for women. The rates vary by period and by birth cohort, quite possibly due to time-varying socio-economic factors. Disentangling time and gender effects could be helpful in work to prevent suicides. A two-sample age-period-cohort model is well-suited for this. However, one has to take care when separating the time effects. It is well-known that linear time effects are not identifiable in age-period-cohort models. This issue becomes more involved when two samples involved. A further complication is that age and period are measured at different frequencies. Thus, the purpose of this paper is to clarify what can be identified in this context.

A very interesting aspect of two sample models is the ability to compare variations across samples. A few examples are as follows. Riebler & Held (2010) compared female mortality in two countries, Denmark and Norway. Dinas & Stoker (2014) compared male and female participation rates in US presidential elections after the introduction of universal suffrage. Cairns et al. (2011) investigated selection effects in life insurance by comparing the mortality of assured lifes in England and Wales with the mortality of the general population. In all these examples it is of interest to formulate and investigate hypotheses about common age, period or cohort effects. For this we need to know exactly what the hypotheses entail and what the associated degrees of freedom are.

In standard one-sample age-period-cohort models, the main identifiable objects are double-differenced time effects (Fienberg & Mason, 1979, Holford, 1983, Clayton & Schifflers, 1987, McKenzie, 2006, Kuang et al., 2008b). The double differences are interpretable as log-odds-ratios or differences-in-differences. This carries over to the two-sample situation. We will show that double differences are identified within each

[1]Nuffield College & Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk. This research is support by ERC grant DisCont 694262.

sample. We also consider cross-sample double differences and show that they are not identified even when restricting the period effect to be common across samples. In the mixed-frequency situation the double differences have to be formed with some care by differencing in 'macro' steps.

The traditional age-period-cohort problem is as follows. The model has a predictor that combines age, period and cohort effects and possibly an intercept. Because the age, period and cohort time scales are linearly related, one finds that linear trends in age, period and cohort are not separately identifiable. There is a lively debate about what to do about that. Broadly speaking, there appears to be three approaches to the traditional age-period-cohort identification problem. *First*, one can impose constraints on the time effects to achieve identification. This has been done in many ways. One can set a subset of the time effect parameters to zero as discussed by Fienberg & Mason (1979). More elaborately, one can choose a particular generalized inverse as in the intrinsic estimator (Yang & Land, 2013, Fu, 2018), or apply monotonicity constraints (Fosse & Winship, 2019, O'Brien, 2022). None of these choices are universally agreable and this has attracted considerable discussion, see for instance Luo (2013), O'Brien (2011), Bell & Jones (2015), Reither et al. (2015), Keiding & Andersen (2016). *Second*, as suggested by Fienberg & Mason (1979), Holford (1983), Clayton & Schifflers (1987), Chauvel & Schröder (2014), Rosenberg (2019), one can apply the first approach for estimation and then extract estimates for estimable functions. These are functions of the time effects that do not change when moving linear trends between the age, period and cohort effects. As an example, double differences are estimable because double differencing removes linear effects. A related idea is to detrend the time effects. *Third*, by analyzing the mapping from the time effects to the predictor in detail, one can characterize all possible estimable functions or, in a different parlor, all functions that are invariant to linear transformations of the time effects. By reparametrizing the predictor exclusively in terms of freely varying, invariant parameters one can avoid the identification problem in the statistical analysis. (Kuang et al., 2008b). See also Smith & Wakefield (2016), Fannon & Nielsen (2019) for reviews.

In the two-sample situation, the immediate issue is to generalize the description of the identification problem. This is done by describing the full set of transformations that can be applied to the time effects without altering the predictor (Carstensen, 2007). From this, we can reparametrize the predictor in terms of estimable or invariant quantities. The analysis is first done for an unrestricted model where the two samples have separate age-period-cohort models and subsequently restricted to the case where time effects are common across samples. The analysis is done in a context of mixed-frequency data arrays building on Nielsen (2022).

There is a relatively small literature on two-sample age-period-cohort methods. Riebler & Held (2010) consider the identification problem although without giving a full characterization. Estimation is done by Bayesian methods, which can hide identification issues (Nielsen & Nielsen, 2014). Recently, Fu et al. (2021) have suggested a two-sample approach based on the intrinsic estimator. This relies on an asymptotic framework where the period dimension increases (Fu, 2016) and permits inference on age effects but not on period or cohort effects.

With the invariant parametrization, the age-period-cohort problem can be approached as any other regression problem whether it is in the context of least squares estimation or a generalized linear model. One can then focus on classical statistical issues such as implementing hypotheses through zero restrictions, assessing the model fit and choosing the data generating structure that motivates asymptotic inference. The identification problem only shows up when seeking to plot the original time effects. We show how this can be done with minimal impact from decisions with respect to the non-identifiable linear trends. The approach is related to that of Chauvel & Schröder (2014).

The inferential framework for the Swiss data is as follows. The data has a large information content in each age-period cohort cell with a population exposure in the order of 100,000 individuals. At the same time the number of parameters is large relative to the number of observations. Thus, it seems reasonable to apply an asymptotic theory where the age-period dimensions are kept fixed while the dispersion of the statistical errors is assumed to be small in each cell. For this, we rely on a small-dispersion asymptotic theory (Harnau & Nielsen, 2018).

In the analysis of the Swiss data, estimation is done by generalized least squares to account for differences in the dispersion for the two samples. It is found that the age-period-cohort models for the individual samples cannot be reduced, yet, the period effects appear to be common. It is investigated to which extent period effects can be replaced by a family integration index formed from marriage and divorce rates.

2 Motivation: the Swiss suicide data

We consider the Swiss suicide data presented and analyzed by Riebler et al. (2012). In short, the data consists of suicide mortality counts and mid-year population data organized as two mixed-frequency age-period arrays for women and men. Age is grouped in five year intervals $15-19, \ldots, 75-79$ while period is annual for $1950-2007$. Thus, there are 13 age groups covering a 65 year range and 58 annual periods.

Age grouping is commonly used by data providers when counts are small. In the present data, the counts range from 5 to 54 for women with a median of 27. For men, the range is 18 to 133 with a median of 66.

Figure 1 shows crude suicide rates by age and period for women and men. The crude rates are found as the sum of all mortality counts for a given age or period divided by the sum of the population data and standardized as rates per 100,000 people. The rates increase with age while female rates are about a third of the male rates. The rates per period fall through the 1950s and 1960s, then increase through the 1970s, after which they fall back again. The peak could match socio-economic trends.

Following Riebler et al., we will consider replacing the period effect with a family integration index composed from marriage and divorce rates as shown in Figure 2(d) below. This choice is motivated by Durkheim's theory from the late 1800s and numerous subsequent work considering the relationship between marital status and suicide.

Riebler et al. applied a Bayesian, two-sample, age-period-cohort, over-dispersed Poisson model. They found a common period effect for the two samples and investigated the extent to which the period effects could be replaced by socio-economic time series.

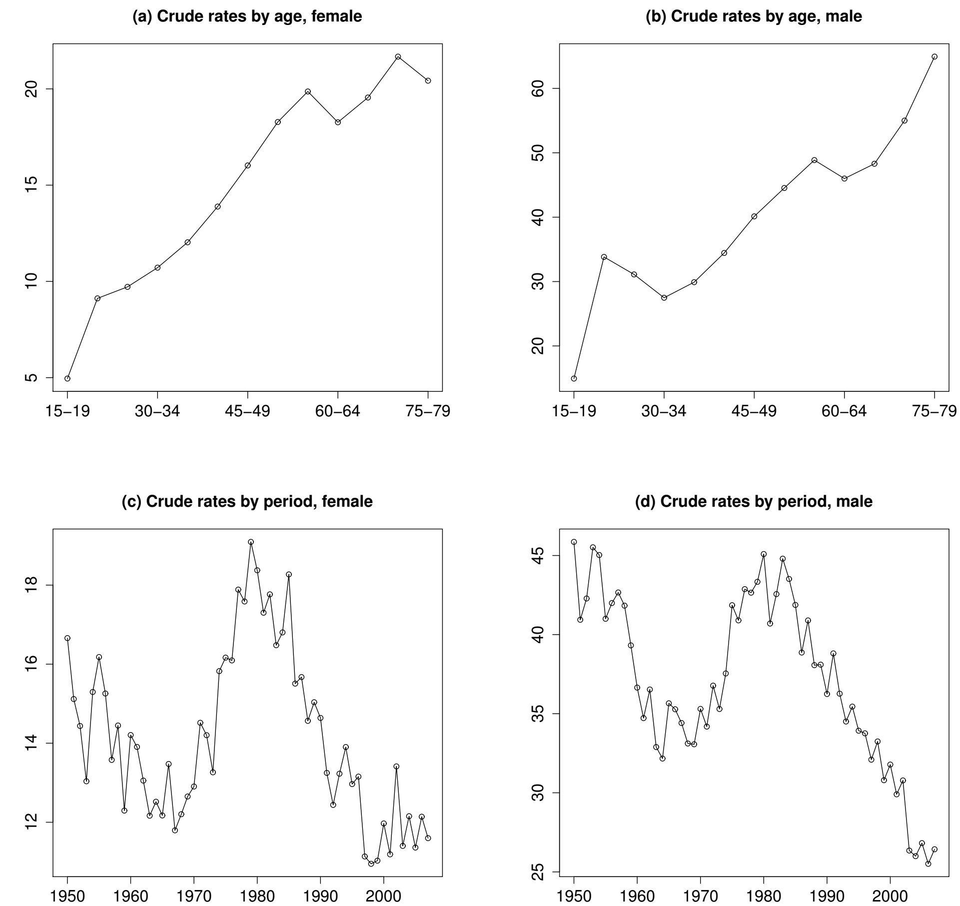

Figure 1: Crude rates per 100,000.

3 The 2-sample model

Suppose we have two samples of data in the form of rates, counts, or doses and responses. Each sample is organized in the same way in terms of mixed-frequency age-period arrays. We will first review the organization of the data, then introduce the unrestricted age-period-cohort model and finally turn to restrictions on the model. We will use the recent theory for mixed-frequency models in Nielsen (2022), henceforth N22.

We will consider the situation where age is grouped over G ages and period is annual. The general case where both age and period are grouped has more complicated notation and is left to the appendix.

3.1 Data structure

The two data arrays have the same mixed-frequency structure with A_G age groups of length G covering $A = A_G G$ ages while period is annual covering P years. The data arrays are regular when $G = 1$. In the Swiss data, $G = 5$, $A_G = 13$ and $P = 58$.

Refering to each observation by the highest relevant age, we get the index set

$$\mathcal{I}_{age,per} = \big\{age = A - gG \text{ for } g = 0, 1, \dots, A_G - 1 \text{ and } 1 \leq per \leq P\big\}. \tag{1}$$

Cohorts are defined according to the convention

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

so that the lowest and highest possible cohort values are 1 and $C = A + P - G$. The possible cohorts over the set $\mathcal{I}_{age,per}$ are

$$\mathcal{I}_{coh} = (1, \dots, C). \tag{3}$$

3.2 The unrestricted model

A two-sample age-period-cohort model has predictor

$$\mu_{age,per,s} = \alpha_{age,s} + \beta_{per,s} + \gamma_{coh,s} + \delta_s \quad \text{for } age, per \in \mathcal{I}_{age,per} \text{ and } s = 1, 2. \tag{4}$$

The time effects on the right hand side of the model equation have dimension

$$q = 2(A_G + A + 2P - G + 1), \tag{5}$$

as each sample has A_G ages, P periods, $A+P-G$ cohorts and 1 intercept. In particular, for regular index arrays, we have $q = 4A + 4P$.

The time effects on the right hand side of (4) are not fully identified. There are two contributions to this under-identification. First, we have the standard age-period-cohort problem that only one level and two linear slopes are identifiable. Second, the mixed-frequency indexation results in additional constraints as noted by Fienberg & Mason (1979). N22 describes the full set of constraints for one mixed-frequency sample. This carries immediately over to the two-sample model.

To write down the mixed-frequency constraints, recall that two integers i, j are congruent modulo G, if G divides their difference $i-j$ and we write $i \equiv j \bmod G$. Combining the assumption that G divides age and A with the relation $coh = per + A - age$ from (2) gives the congruence $coh \equiv per \bmod G$.

The under-identification of the time effects can be characterized as follows. The predictor is invariant to transformations of the time effects of the form

$$\begin{aligned}\mu_{age,per,s} = &\Big\{\alpha_{age,s} + a_s + d_s \times (A - age)\Big\} \\ &+ \Big\{\beta_{per,s} + b_s + d_s \times per + \sum_{j=1}^{G-1} f_{j,s} 1_{(per \equiv j \bmod G)}\Big\} \\ &+ \Big\{\gamma_{coh,s} + c_s - d_s \times coh - \sum_{j=1}^{G-1} f_{j,s} 1_{(coh \equiv j \bmod G)}\Big\} + (\delta_s - a_s - b_s - c_s),\end{aligned} \tag{6}$$

for any values of $a_s, b_s, c_s, d_s, f_{j,s}$ for $1 \leq j < G$ and $s = 1, 2$. Here, d_s indicates arbitrary slopes, a_s, b_s, c_s indicate arbitrary macro levels appearing at time intervals of G steps, while $f_{j,s}$ indicate arbitary micro levels appearing inbetween the macro steps. The latter relate to the seasonal micro pattern noted by Holford (2006) and Riebler & Held (2010).

The transformations in (6) describe the identification problem completely as proved in N22. For the regular, one-sample, regular case, these transformations simplify to those described by Carstensen (2007). The dimension of the transformations is $2(G+3)$. Subtracting this from the time effect dimension q defined in (5) shows that the dimension of the variation of the predictor is

$$p = q - 2(G+3) = 2\{A_G + A + 2P - 2(G+1)\}. \tag{7}$$

For regular index arrays, we have $p = 4A + 4P - 8$ corresponding to eight constraints, which are three level- and one slope-constraint for each sample.

Confronted with the under-identification, it is convenient to reparametrize the predictor in terms of freely varying, invariant parameters (Kuang et al., 2008b). We will refer to the vector of these as the canonical parameter following terminology from exponential family theory (Sundberg, 2019). The canonical parameter is defined in terms of two linear planes and a set of double differences.

We define the double differences as follows. Let Δ_G indicate a macro difference operator over G periods, giving first differences $\Delta_G\beta_{per,s} = \beta_{per,s} - \beta_{per-G,s}$. Correspondingly, let Δ_1 indicate a 1-period difference operator. In combination, we get second differences

$$\Delta_G\Delta_1\beta_{per,s} = \Delta_G\beta_{per,s} - \Delta_G\beta_{per-1,s} = \beta_{per,s} - \beta_{per-1,s} - \beta_{per-G,s} + \beta_{per-G-1,s}. \tag{8}$$

The double differences $\Delta_G\Delta_G\alpha_{age,s}$, $\Delta_G\Delta_1\beta_{per,s}$ and $\Delta_G\Delta_1\gamma_{coh,s}$ are invariant to the transformations in (6). However, the simpler one-period double differences $\Delta_1\Delta_1\beta_{per}$ are only invariant to (6) in the regular case as pointed out by Gascoigne & Smith (2021). Why is that? The one-period double difference of the period effect is constructed from the period effect at consecutive periods per, $per-1$ and $per-2$. The transformations in (6) involve sums of arbitrary coefficients multiplied with indicators defined in terms of modulos. The arbitrary parameters are only eliminated when the three consecutive periods are congruent modulo G. This is only possible for $G = 1$.

Identified and invariant linear planes for the two samples can be expressed in terms of identified and invariant levels

$$\mu_{A,1,s} = \alpha_{A,s} + \beta_{1,s} + \gamma_{1,s} + \delta_s, \tag{9}$$

along with identified and invariant slopes chosen as

$$\lambda_{1,s} = \mu_{A-G,1,s} - \mu_{A,1,s} = \Delta_G\gamma_{1+G,s} - \Delta_G\alpha_{A,s} \tag{10}$$

$$\nu_{G,s} = \mu_{A,1+G,s} - \mu_{A,1,s} = \Delta_G\gamma_{1+G,s} + \Delta_G\beta_{1+G,s}, \tag{11}$$

as well as micro levels defined as

$$\nu_{h,s} = \mu_{A,1+h,s} - \mu_{A,1,s} = \Delta_h\gamma_{1+h,s} + \Delta_h\beta_{1+h,s} \qquad \text{for } h = 1, \dots, G-1. \tag{12}$$

We note that the slopes and micro levels in (10)-(12) cannot be separated into individual slopes for age, period, cohort.

By imposing additional constraints such as $\Delta_G \beta_{1+G,s} = 0$, we get a unique value for $\Delta_G \gamma_{1+G,s}$ from (11). It is common to refer to this uniqueness as identification. However, this type of identification destroys the invariance property as it depends on the choice that $\Delta_G \beta_{1+G,s} = 0$, which is not invariant to the transformations in (6).

We can now write down the canonical parameter (N22) as $\xi = (\xi_1', \xi_2')'$, where

$$\xi_s = \big(\mu_{A,1,s}; \quad \lambda_{1,s}; \quad \nu_{1,s}, \dots, \nu_{G,s}; \quad \Delta_G \Delta_G \alpha_{A-gG,s} \text{ for } 0 \leq g \leq A_G - 3; \\ \Delta_G \Delta_1 \beta_{per,s} \text{ for } G+2 \leq per \leq P; \quad \Delta_G \Delta_1 \gamma_{coh,s} \text{ for } G+2 \leq coh \leq C\big). \tag{13}$$

The canonical parameter has dimension p as given in (24).

Finally, we must write the predictor as a function of the canonical parameter in order to reparametrize the model. A representation of the predictor in terms of the canonical parameter is given in the Appendix for a general mixed-frequency setup. For simplicity, we focus on the case of regular index sets here. That is, when $G = 1$, we have

$$\mu_{age,per,s} = \mu_{A,1,s} + (A - age)\lambda_{1,s} + (per - 1)\nu_{1,s} \\ + \sum_{t=age}^{A-2} \sum_{u=t}^{A-2} \Delta_1 \Delta_1 \alpha_{u+2,s} + \sum_{t=3}^{per} \sum_{u=3}^{t} \Delta_1 \Delta_1 \beta_{u,s} + \sum_{t=3}^{coh} \sum_{u=3}^{t} \Delta_1 \Delta_1 \gamma_{u,s}, \tag{14}$$

with the convention that empty sums are zero. This arises by applying the one-sample representation of Martínez Miranda et al. (2015) to each sample. The first three terms of the representation define a linear plane. Since the cohort is entangled with age and period, then $\lambda_{1,s}$ should be interpreted as an age-cohort slope for fixed period while $\nu_{1,s}$ is a period-cohort slope for fixed age. The representation implies that we can write the predictor in terms of a common design vector for each samples so that

$$\mu_{age,per,s} = \xi_1' x_{age,per} 1_{(s=1)} + \xi_2' x_{age,per} 1_{(s=2)}. \tag{15}$$

One interpretation of the double sums of double differences in (14) is that they represent detrended versions of the original time effects where, for instance, the two first elements of the period effects are constrained to zero. The levels and linear slopes for the three time effects are collected in an identified, invariant linear plane. Typically, plots of the double sums will be trending and not offer appealing interpretation. We will return to a more useful way of detrending in the empirical application, but see also Chauvel & Schröder (2014). Thus, for now, the value of the representation in (14) is that it defines a design matrix that can be used for estimation.

The equations (14), (15) parametrize the age-period-cohort predictor in terms of the canonical parameter. As the aim is to use this for estimation, we must be assured that the information content is the same as with the original expression for the predictor in terms of the time effects. Thus we need to know that

(a) *ξ is a linear function of μ that is invariant to the transformations in (6)*;

(b) *μ is a linear function of ξ given by (14) or (31)*;

(c) *The parameter ξ is exactly identified in that $\xi^\dagger \neq \xi^\ddagger$ implies $\mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$.*

In other words, in the original model (4), the time effects generate a certain variation in the predictor, which we will match with the variation in the data. The identification problem means that different time effects can generate the same predictor. This redundancy is eliminated when parametrizing the predictor in terms of the canonical parameter. As the canonical parameter is invariant to the transformations (6), the identification problem is no longer of any consequence for estimation.

The properties (a)-(c) follow from Theorem 1 of N22, which builds on the analysis in Kuang et al. (2008b). That result gives such a summary for one-sample, mixed-frequency arrays and can be applied to each sample in the two-sample situation.

3.3 The hypothesis of common period effects

With a two-sample set-up it is natural to ask if any of the time effects are common across samples. Here, we look at the period effects.

The hypothesis that the non-linear part of the period effects is common is that

$$\Delta_G \Delta_1 \beta_{per,1} = \Delta_G \Delta_1 \beta_{per,2} \qquad \text{for } G+2 \leq per \leq P. \tag{16}$$

The degrees of freedom of the hypothesis is $P-G-1$. The interpretation of this hypothesis follows from the representation (14) for the regular-frequency case and (31) for the mixed-frequency case. We see that the double sum of double differenced period effects will be common while there are no other constraints. In particular, there are no constraints to the linear planes.

For estimation under the hypothesis it is convenient to rewrite the design vector expression for the predictor in (15). By adding and subtracting $\xi_1' x_{age,per} 1_{(s=2)}/2$ and $\xi_2' x_{age,per} 1_{(s=1)}/2$ we get

$$\mu_{age,per,s} = \Big(\frac{\xi_1+\xi_2}{2}\Big)' x_{age,per} + \Big(\frac{\xi_1-\xi_2}{2}\Big)' x_{age,per} \big\{1_{(s=1)} - 1_{(s=2)}\big\}. \tag{17}$$

Thus, restricting the period double differences to be common is equivalent to a zero restriction on the second parameter $(\xi_1-\xi_2)/2$. Indeed, if we were to impose the restriction that all parameters are common for the two samples, so that $\xi_1 = \xi_2$, then two regression parameters in (17) reduce to ξ_1 and zero, respectively.

Alternatively, we could formulate the hypothesis of common period effects in terms of the unidentified period effect, so that

$$\beta_{per,1} = \beta_{per,2} \qquad \text{for } 1 \leq per \leq P. \tag{18}$$

This formulation implies the $P-G-1$ restrictions for the double differences in (16). In addition, it gives $G+1$ constraints to the macro and micro levels and the slope of the cross sample difference of the period effects. This parametrizes the cross-sample differences of the linear planes in terms of age and cohort effects without actually restricting the planes. Thus, it would be observationally equivalent to parametrize the linear planes in terms of age and period effects or in terms of period and cohort effects. This is the standard age-period-cohort identification problem. Of course, if it the linear period effects did truly satisfy (18), then the cross-sample differences of the age and cohort slopes would be identified. This truism cannot be confirmed by statistical analysis. This point was made for the one-sample, regular-frequency case by Clayton & Schifflers (1987). Keiding & Andersen (2016) raised the point in a comment on a paper concerned with the question whether delaying childbearing to older ages might be associated with more positive educational and health outcomes for the children. See also Fannon & Nielsen (2019) for a review.

Would the hypothesis (18) formulated for the period effects in levels then imply that cross-sample differences are invariant? For instance, are the cross-sample age difference $\alpha_{age,2} - \alpha_{age,1}$ or at least the cross-sample double age difference $\Delta_G \alpha_{age,2} - \Delta_G \alpha_{age,1}$ invariant? From a statistical viewpoint the answer is negative since the hypothesis does not constrain the linear planes and we cannot therefore not move away from age-period-cohort problem that different parametrizations of the linear planes are observationally equivalent. In any case, the levels of the cross-sample age differences would remain entwined with those of the cross-sample cohort differences.

To conclude, from a statistical viewpoint, the hypothesis of common period effects concerns their non-linear part as outlined in (16). Treated as such, the restriction can be imposed as a zero constraint on the cross-sample differenced canonical parameter through (17). The degrees of freedom of the hypothesis then matches the dimension of the constraint.

3.4 Replacing period effect with time series

Could it be that socio-economic effects explain the period movements that we see in the data? This can be investigated by imposing the time series as a restriction on the period effect. When doing so, the identification problem should be taken into account.

We replace the double differenced period parameter $\Delta_G \Delta_1 \beta_{per}$ by the double differences of the external time series multiplied by a free scalar parameter, that is $\psi \Delta_G \Delta_1 T_{per}$ say. The degrees of freedom will then be the number of period double differences, $P - G - 1$, minus one free parameter.

In a two-sample age-period-cohort model the external time series restriction can be done in various ways. It can be imposed on the cross-sample differenced parameter $\xi_1 - \xi_2$, or on the common parameter $\xi_1 + \xi_2$, or on the individual parameters ξ_1, ξ_2.

3.5 Further sub-models

Other sub-models may be relevant. We give an overview, but see also N22.

A *period-cohort* model for the cross-sample differenced predictors arises when the non-linear part of the age effects is common, that is $\Delta_G \Delta_G \alpha_{age,1} = \Delta_G \Delta_G \alpha_{age,2}$. The degrees of freedom of the hypothesis is $A_G - 2$.

An *age-period* model for the cross-sample differenced predictors arises with the restriction $\Delta_G \Delta_1 \gamma_{coh,1} = \Delta_G \Delta_1 \gamma_{coh,2}$. The degrees of freedom of the hypothesis is $C - G - 1$.

An *age-drift* model arises when both the period and cohort double differences are restricted to be zero. That is $\Delta_G \Delta_1 \beta_{perh,1} = \Delta_G \Delta_1 \beta_{per,2}$ and $\Delta_G \Delta_1 \gamma_{coh,1} = \Delta_G \Delta_1 \gamma_{coh,2}$. The degrees of freedom is the count of parameters $P + C - 2(G + 1)$. This hypothesis does not restrict the linear plane. For further discussion, see Clayton & Schifflers (1987).

A pure *age* model occurs when restricting the age-drift model further by requiring the cross-sample period-cohort slope and micro effects through $\nu_{h,1} = \nu_{h,2}$. This gives further G constraints. Now the linear slope only varies with age and can be attributed as an age effect.

These restrictions are imposed on the cross-sample differenced predictor. The same type of restrictions could be imposed on the predictors for the individual samples or on the cross-sample common predictor.

4 Empirical illustration

We now consider the Swiss suicide data reviewed earlier. A two sample age-period-cohort predictor will be applied in the context of a log-normal model using least squares for the log rates. We will first analyze the two samples separately using one-sample age-period-cohort analysis. All time effects appear significant and the scale parameters are found to be different. Thus, it is necessary to correct for this difference through generalized least squares regression combined with a small-dispersion asymptotic theory. We will then find that we cannot reject the hypothesis of a common period effect. We will also explore the use of a marriage-divorce index as period effect.

The data analysis was done in R (R Core Team, 2022) using code building on the apc package (Nielsen, 2015).

4.1 Initial one-sample analyses

At first, we model the two samples separately. We assume that the log suicide rate is normal, where the expectation has a linear age-period-cohort structure as in (4) and constant variance. At this point, the parameters for expectation and variance are allowed to depend on the sample. The models are estimated separately for the two samples using the least squares method. For inference we can, at this point, rely on the exact distribution theory for the normal model.

model	$-2\log L$	df	F vs. apc	df vs. apc	p_F	$\hat{\sigma}$
APC	-363.02	572				0.218
AP	-157.38	684	1.60	112	0.0003	0.229
AC	-68.82	624	5.25	52	0.0000	0.254
PC	-18.55	583	30.11	11	0.0000	0.272
$\chi^2_{normality}(2) = 4.50$ ($p = 0.105$).						

Table 1: Analysis of variance, women.

model	$-2\log L$	df	F vs. apc	df vs. apc	p_F	$\hat{\sigma}$
APC	-841.70	572				0.159
AP	-605.47	684	1.88	112	0.0000	0.170
AC	-434.27	624	7.88	52	0.0000	0.199
PC	-210.52	583	68.10	11	0.0000	0.239
$\chi^2_{normality}(2) = 0.31$ ($p = 0.579$).						

Table 2: Analysis of variance, men.

Tables 1, 2 show separate analyses of variance for women and for men. Each table consider full APC models and reductions to sub-models AP, AC and PC where, respectively, the cohort, the period and the age effects are omitted. The restrictions are all strongly rejected. We note that the residual standard deviation, $\hat{\sigma}$, is somewhat smaller for women than for men, matching the lower number of cases for women.

Tables 1, 2 also show tests for the normality assumption. The tests are standard skewness and kurtosis based tests, see for instance Hendry & Nielsen (2007). In both cases, the p-values are large and the assumption of normality cannot be rejected.

4.2 Two-sample analysis by least squares

model	$-2\log L_{OLS}$	df	F vs. apc	df vs. apc	p_F	$\hat{\sigma}_{OLS}$	$\hat{\sigma}_{GLS}$
APC	-1129.99	1144				0.191	0.159
AP	-836.72	1256	2.19	112	0.0000	0.201	0.167
AC	-1052.62	1196	1.16	52	0.2092	0.192	0.160
PC	-1021.68	1155	7.74	11	0.0000	0.197	0.164
OLS: $\chi^2_{normality}(2) = 15.82$ $[p = 0.0004]$							
GLS: $\chi^2_{normality}(2) = 2.13$ $[p = 0.3440]$							

Table 3: Analysis of variance, both samples. Submodels for cross-sample differenced predictor. For GLS, data for women are scaled to have same dispersion as men.

We apply the proposed two-sample age-period-cohort analysis, where the common predictor is unrestricted while the cross-sample differenced predictor is restricted in various ways. The log normal specification is maintained. We consider both the case with common scale parameter so that least squares estimation can be used and the case of different scale parameters in the two sample so that generalized least squares estimation is needed. Table 3 summarizes the results.

Asymptotic inference relies on asymptotic arguments. We must specify the assumptions underlying the asymptotics. One approach is to let the size of the data array and therefore the size of the parameter vector increase as in Fu (2016). Another approach is to use small-dispersion asymptotics, where the size of the data array is kept fixed while the scale, or dispersion, parameters are shrinking in the asymptotic experiment. The second approach seems particular useful here as the dose, or exposure, is large. The dose is the entire Swiss population of the relevant age and gender. It is in the order of 100,000 for each cell. With the second approach it is possible to justify use of F distributions as limiting distributions. Formal analysis is given by Jørgensen (1987) in the context of exponential dispersion models whereas Harnau & Nielsen (2018) develop a Central Limit Theorem for this situation. Both setups allow log normal distribution, but the latter has some flexibility in allowing approximate log normal distributions and some types of over-dispersed Poisson models. The implementation in Kuang & Nielsen (2020) is suited for the present situation. Thus, the idea of the asymptotic experiment is that log suicide rates converge to constant parameters for large values of the dose appearing as denominator of the rates. Convergence of the rates to constant parameters corresponds to a shrinking dispersion. More specifically, we apply a log normal model where we hold fixed the dimension of the index array, the canonical age-period-cohort parameter, and the ratio of the scales in the sample, while the scale parameters shrink in the asymptotic experiment.

The common variance assumption can be tested by comparing the least squares log likelihoods for the APC model in Table 3 and in Tables 1, 2 to get the Bartlett test statistic (Harnau, 2018; Kuang & Nielsen, 2020). The test statistic is $363.02 + 841.70 - 1129.99 = 78.73$ multiplied by a Bartlett correction factor which is very close to unity in this case. The test statistic is large compared to a χ^2_1-distribution. In addition, normality is rejected. This is quite possibly a consequence of mixing two samples with different scale. Ordinary least squares will therefore not give reliable inference. Instead, generalized least squares is needed. This is done by scaling the log rates and the design matrix for women by the ratio of the scale estimates for men and women as reported in Tables 1, 2. This results in a scaling factor of 0.159/0.218. We see that normality can no longer be rejected.

We now consider the restrictions on the age-period-cohort structure for the differenced predictor. The F-statistics for the three sub-models considered in Table 3, turn out to be exactly identical when estimating by ordinary least squares and by generalized least squares. This is a consequence of the particular block structure of the design and covariance matrices and can be checked through a somewhat detailed derivation. However, given the difference in dispersion for women and men, we must apply the F-statistics under the generalized least squares setup. In that case the F-statistics are asymptotically F-distributed under the small-dispersion setup. Thus, from Table 3 we learn that we can reduced the model for the cross-sample differenced predictor to an age-cohort model, whereas the age-period and period-cohort models are strongly rejected. The interpretation of the age-cohort model is that the period effect is common across samples. This matches conclusions by Riebler et al. (2012).

4.3 Plots of time effects

Figure 2 shows detrended time effects estimated by the generalized least squares methods from the unrestricted age-period-cohort models. The detrending serves three purposes. First, to emphasize non-linearity, which is the only part of the time effects that is invariant to the identification problem. Second, to disentangle the plots, so that they can be viewed separately. Indeed, with fewer constraints the plots are linked inextricably and must be viewed jointly (Carstensen, 2007). Third, to ensure that the degrees of freedom shown in the plot matches the degrees of freedom associated with the time effects. Detrending can be done in various ways. Here it is done by restricting the time effects to be zero at a time close to the beginning and a time close to the end. There is no estimation uncertainty at those two points and the degrees of fredom can be appreciated. Chauvel & Schröder (2014) suggest to set the averaged time effect to zero and then eliminate the time trend. This achieves the first two objectives above.

Figure 2(a) shows the two detrended age effects. The age effects are subject to an arbitrary level and linear slope due to the invariance transformations (6). Thus, the plotted age effects are constrained to start and end in zero in order to highlight the non-linearity. The detrended age effect first rises considerably relative to an overall

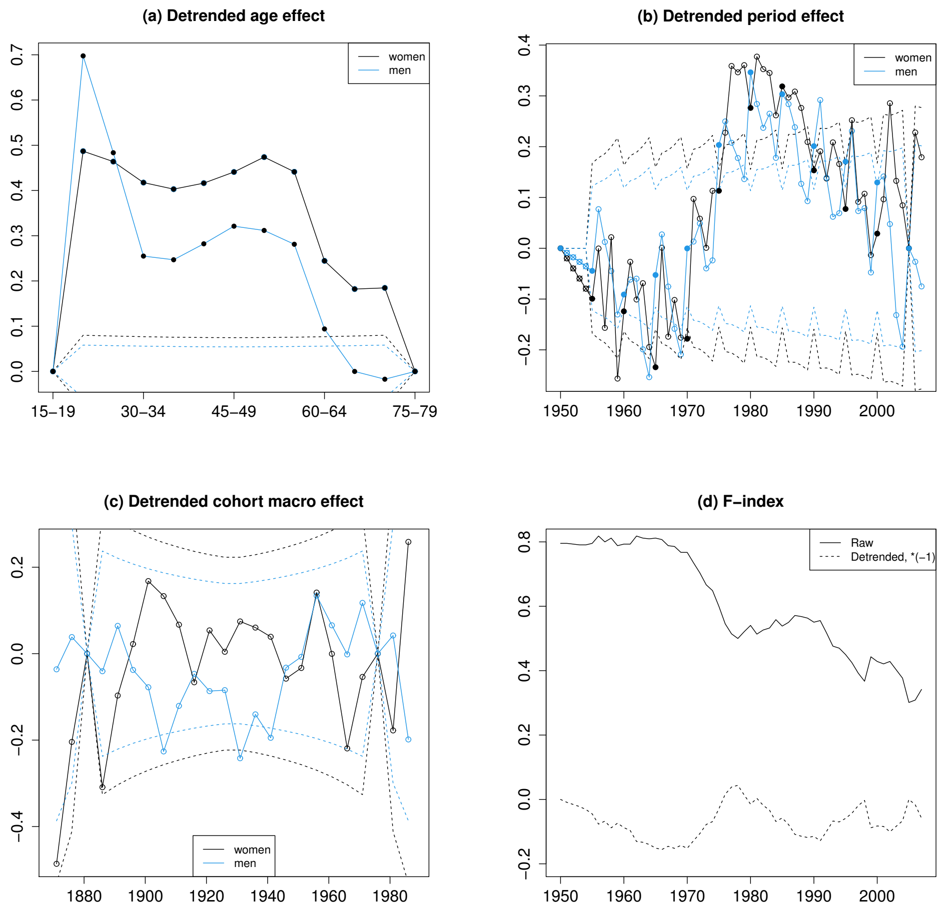

Figure 2: (a)-(c): Detrended estimates of time effects. (d): F-index

age-period-cohort linear plane. This corresponds to a relatively large increase in suicide rates for people in their twenties. The curves then decline gradually corresponding to relative declines in rates. There are small local peaks around the age of 50 and for men also at age 75. The dotted lines show two standard errors from zero. By construction, the dotted lines are exactly proportional for the two samples reflecting the different dispersion for women and men. We conclude that the age-trends are significant for both men and women in line the individual tests for the period-cohort models in Tables 1, 2. The age-trends do also appear to be quite different for women and men in line with the rejection of period-cohort model reported in Table 3.

Figure 2(b) shows detrended period effects. Now, we need to adjust for an arbitrary level, arbitrary slope and four arbitrary micro levels. Thus, the macro trend, shown with bullets at five year intervals is detrended to start and end in zero. The micro trends starting at 1, 2, 3, 4 years after 1950 are demeaned to start on the line between the first two macro estimates in 1950 and 1955 as marked with crosses. The idea here is to minimize the apparent seasonality in a way that is common for the two samples. Once again, we see that the period effects are individually significant, although somewhat less than for ages, in line with the tests for the age-cohort models reported in Tables 1, 2.

The period-trends do appear to follow each other in line with the non-rejectance of the age-cohort restriction in Table 3.

Figure 2(b) shows detrended cohort macro effects. The transformations (6) show that the cohort effect is subject to the same under-identification as the period effect, but the cohort macro effects are only subject to an arbitrary level and slope. Thus, we apply two zero constraints. The cohort effect is inevitably quite noisy because of the mixed-frequency setup with 112 degrees of freedom associated with cohorts among 754 observations. The estimated cohort parameters are particularly noisy at end points. Thus, the zero constraints are chosen three macro steps from the end points. The micro effects starting 1, 2, 3, 4 years after the macro effect are noisy and not shown in this plot of macro effects. Overall, the individual cohort effects are less significant than age and period effects – compare with the age-period tests in Tables 1, 2. The cohort effects are somewhat different across samples in line with Table 3.

4.4 Replacing the period effect with an external time series

model				Test vs APC			
common	diff.	$-2\log L_{GLS}$	df	F	df	p_F	$\hat{\sigma}_{GLS}$
APC	APC	3862.93	1144				0.159
APC	AC	3940.30	1196	1.16	52	0.209	0.160
APC	AC+F	3934.12	1195	1.08	51	0.321	0.159
AC+F	APC	4340.08	1195	8.35	51	0.000	0.182
AC+F	AC+F	4432.71	1246	5.15	102	0.000	0.184

Table 4: Analysis of variance, both. Submodels for common and differenced predictors.

We now investigate if the period effect can be replaced by an external time series. Following Riebler et al. (2012), we consider a familiy integration index computed as $(m_{per} - d_{per})/(m_{per} + d_{per})$ where m_{per} and d_{per} are the counts of marriages and divorces in a given period. The F-index is shown in Figure 2(d). A high value of this measure indicates better integration which could be associated with lower suicide rate.

The period effect is restricted to follow the F-index, by substituting the double differences of the period effect with those of the F-index. Table 4 gives an analysis of variance. Here, the common predictor and the cross-sample differenced predictor are restricted either individually or jointly. The first model has an age-period-cohort structure for both predictors, while the second model has an age-cohort structure for the cross-sample differenced predictor. The remaining models have one or both of the predictors following an age-cohort model combined with the F-index as period effect.

The conclusion from the table is that we cannot reject replacing the period effect of the cross-sample differenced parameter with the F-index. Eliminating the F-index from that model to get an age-cohort structure adds another degree of freedom. The relevant F-statistic is 4.906 with p-value 2.7%, which gives a marginal decision. Thus, there is slight evidence that it is better to replace the period effect in the difference parameter with the F-index than eliminating it all together. The coefficient ψ for the F-index is estimated by -0.217 with standard error 0.098. The first sample is chosen as women.

Thus, changes in family integration has a bigger impact for men than for women. An increase in the F-index by 1% implies that the suicide rate for men decreases by 0.2% more than that of women. This is in line with conclusions by Riebler et al.

5 Conclusion

We considered the two-sample age-period-cohort model and clarified what can be identified invariantly and the associated degrees of freedom. The hypothesis of a common period effects was analyzed in a similar fashion. Replacing the period effect with an external time series was discussed.

The two-sample age-period-cohort model was applied to the Swiss suicide data previously analyzed by Riebler et al. (2012). Inference was conducted for a log-normal model using an asymptotic framework where the age-period dimensions are fixed while dispersions shrink. A generalized least squares approach was used to account for differences in dispersion for women and men. The hypothesis of a common period effect could not be rejected. This was done by setting the cross-sample difference of the period effects to zero. There was weak evidence that it would be better to replace the cross-sample period difference with an external family integration index instead of removing it altogether. The interpretation is that an increase in the family integration index gives a larger improvement in suicide rates for men than for women.

For the present empirical analysis forecasting is not of particular interest. If forecasting were of interest, it would be desirable to work with invariant forecasts (Kuang et al., 2008a). The small-dispersion asymptotics allow the construction of distribution forecasts (Harnau & Nielsen, 2018; Nielsen, 2022). This asymptotic theory applies not only for log normal models as here but also for over-dispersed Poisson models.

A General mixed-frequency arrays

The two-sample age-period-cohort model is generalized to general mixed-frequency arrays building on Nielsen (2022), henceforth N22.

Data structure, The general mixed-frequency setup has A_G age groups of length G covering $A = A_G G$ ages and P_H period groups of length H covering $P = P_H H$ periods. We assume that the largest common divisor of G and H is unity. If groups have common divisor larger than unity, such as 10 and 4, we can scale by the common divisor of 2.

An age-period data array has index set

$$\mathcal{I}_{age,per}: \quad \begin{cases} age = A - gG, & \text{where} \quad g = 0, 1, \ldots, A_G - 1, \\ per = H + hH & \text{where} \quad h = 0, 1, \ldots, P_H - 1. \end{cases} \tag{19}$$

When, $G, H \geq 2$, certain cohort values will be skipped as noted by Holford (2006). Table 5 illustrates this for a case with $G = 5$ and $H = 3$. Macro blocks of dimension GH are indicated with dashed lines. Note that top left and bottom right macro blocks are identical apart from trimming. The cohort values $4, 5, 7, 10$ and $42, 45, 47, 48$ are skipped. This corresponds to 3 plus the values $1, 2, 4, 7$ and 49 minus the same values.

The skipping problem is akin to the coin problem (N22): If we have coins of denominations G, H, which monetary amounts can we form? Let $\mathcal{N}_{G,H}$ denote the non-representable monetary amounts. The Frobenius number $F_{G,H} = GH - G - H$ is the largest non-representable number, while Sylvester pointed out that the number of non-representable numbers is $S_{G,H} = (G-1)(H-1)/2$ (Ramírez Alfonsín, 2005). Algorithms for finding $\mathcal{N}_{G,H}$ are discussed in N22. As an example, if we have coins $G = 5$ and $H = 3$ then $\mathcal{N}_{5,3} = (1, 2, 4, 7)$ while $F_{G,H} = 7$ and $S_{G,H} = 4$.

The possible cohorts over the set $\mathcal{I}_{age,per}$ are given by

$$\mathcal{I}_{coh} = (H, \ldots, C) \backslash (H + c, C - c : c \in \mathcal{N}_{G,H}). \tag{20}$$

The number of possible cohorts is then $C - (H - 1) - 2S_{G,H}$, which equals $A + P - GH$. We will also consider the subset of cohort values arising when dropping age A and period

period		age						
		real	50-54	55-59	60-64	65-69	70-74	75-79
real	*per*	*A-age*	*40*	*20*	*15*	*10*	*5*	*0*
1984-86	*3*		28	23	18	13	8	3
1987-89	*6*		31	26	21	16	11	6
1990-92	*9*		34	29	24	19	14	9
1993-95	*12*		37	32	27	22	17	12
1996-98	*15*		40	35	30	25	20	15
1999-01	*18*		43	38	33	28	23	18
2002-04	*21*		46	41	36	31	26	21
2005-07	*24*		49	44	39	34	29	24

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

H from $\mathcal{I}_{age,per}$. This smaller set of cohorts is

$$\mathcal{I}_{coh}^{\circ} = (G+2H,\ldots,C)\backslash(G+2H+c, C-c : c \in \mathcal{N}_{G,H}). \tag{21}$$

Unrestricted age-period-cohort model. Consider the predictor given in (4). The time effects on the right hand side of the model equation have dimension

$$q = 2(A_G + P_H + A + P - GH + 1), \tag{22}$$

as each sample has A_G ages, P_H periods, $A+P-GH$ cohorts and 1 intercept.

The invariance transformations in (6) generalize as

$$\begin{aligned}\mu_{age,per,s} = &\Big\{\alpha_{age,s} + a_s + d_s \times (A - age) + \sum_{i=1}^{H-1} e_{i,s} 1_{(A-age \equiv i \bmod H)}\Big\} \\ &+ \Big\{\beta_{per,s} + b_s + d_s \times per + \sum_{j=1}^{G-1} f_{j,s} 1_{(per \equiv j \bmod G)}\Big\} \\ &+ \Big\{\gamma_{coh,s} + c_s - d_s \times coh - \sum_{i=1}^{H-1} e_{i,s} 1_{(coh \equiv i \bmod H)} - \sum_{j=1}^{G-1} f_{j,s} 1_{(coh \equiv j \bmod G)}\Big\} \\ &+ (\delta_s - a_s - b_s - c_s),\end{aligned} \tag{23}$$

for any values of $a_s, b_s, c_s, d_s, e_{i,s}, f_{j,s}$ for $1 \le i < H$, $1 \le j < G$ and $s = 1, 2$. The transformations in (23) have dimension $2(G+H+2)$. Subtracting this from q defined in (22) shows that the dimension of the variation of the predictor is

$$p = q - 2(G+H+2) = 2\{A_G + P_H + A + P - (G+1)(H+1)\}. \tag{24}$$

The canonical parameter has dimension p. It consists of invariant double differences

$$\Delta_{GH}\Delta_G \alpha_{age,s} \quad \text{for} \quad age = A - gG \text{ with } 0 \le g \le A_G - H - 2, \tag{25}$$

$$\Delta_{GH}\Delta_H \beta_{per,s} \quad \text{for} \quad per = H + hH \text{ with } G+1 \le h \le P_H - 1, \tag{26}$$

$$\Delta_G \Delta_H \gamma_{coh,s} \quad \text{for} \quad coh \in \mathcal{I}_{coh}^{\circ}, \tag{27}$$

and invariant linear plane parameters

$$\lambda_{g,s} = \Delta_{gG}\gamma_{H+gG,s} - \Delta_{gG}\alpha_{A,s} \qquad \text{for } g = 1,\ldots,H, \tag{28}$$

$$\nu_{h,s} = \Delta_{hH}\gamma_{H+hH,s} - \Delta_{hH}\beta_{H=hH,s} \qquad \text{for } h = 1,\ldots,G. \tag{29}$$

When reparametrizing the predictor in terms of the canonical parameter, it is convenient to express the time scales through their Euclidean representations. For instance, age is given by $age = A - gG$. Let $q_g = \lfloor g/H \rfloor$ be the largest integer not exceeding g/H and let $r_g = g - q_g H$ so that $q_g \ge 0$ and $0 \le r_g < H$. Thus, we can write

$$age = A - q_g GH - r_g G, \quad per = H + q_h GH + r_h H, \quad coh = H + qGH + r_g G + r_h H, \tag{30}$$

where $q = q_g + q_h$. Each combination of $0 \le r_g < H$ and $0 \le r_h < G$ gives different micro effect and different representations

$$\mu_{A-q_g GH-r_g, H-r_h GH-r_h} = M^{intercept}_{r_g,r_h} + q_g M^{age/coh}_{r_g,r_h} + q_h M^{per/coh}_{r_g,r_h} + S^{age}_{q_g,r_g} + S^{per}_{q_h,r_h} + S^{coh}_{q_g+q_h,r_g,r_h}, \quad (31)$$

where the level and slopes are given by

$$M^{intercept}_{r_g,r_h} = \mu_{A,H} + 1_{(r_g>0)}\lambda_{r_g,s} + 1_{(r_h>0)}\nu_{r_h,s} + 1_{(r_g>0)}1_{(r_h>0)}\Delta_{r_g G}\Delta_{r_h H}\gamma_{H+r_g G+r_h H}, \quad (32)$$

$$M^{age/coh}_{r_g,r_h} = \lambda_H + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_h H}\gamma_{H+GH+r_g G+r_h H} + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_g G}\big(\gamma_{H+GH+r_g G}\alpha_A\big), \quad (33)$$

$$M^{per/coh}_{r_g,r_h} = \nu_G + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_g G}\gamma_{H+GH+r_g G+r_h H} + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_h H}\big(\gamma_{H+GH+r_h H}\beta_{H+GH+r_h H}\big), \quad (34)$$

while the double sums of double differences are

$$S^{age}_{q_g,r_g} = 1_{(q_g\ge 2)}\sum_{v=1}^{q_g-1}\sum_{u=0}^{vH-1}\Delta_{GH}\Delta_G\alpha_{A-(r_g+u)G,s}, \quad (35)$$

$$S^{per}_{q_h,r_h} = 1_{(q_h\ge 2)}\sum_{v=1}^{q_h-1}\sum_{u=1}^{vG}\Delta_{GH}\Delta_H\beta_{H+GH+(r_h+u)H,s}, \quad (36)$$

$$S^{coh}_{q_g+q_h,r_g,r_h} = 1_{(q\ge 2)}\sum_{v=1}^{q-1}\sum_{u=1}^{vG}\Delta_{GH}\Delta_H\gamma_{H+GH+r_g G+(r_h+u)H,s}. \quad (37)$$

For $G = H = 1$ one gets the expression (14) by changing summation indices.

Restrictions. The hypothesis of common period effects is now that $\beta_{per,1} = \beta_{per,2}$ for $per = H + hH$ with $0 \le h < P_H$. This is equivalent to restricting the double differences as $\Delta_{GH}\Delta_H\beta_{per,1} = \Delta_{GH}\Delta_H\beta_{per,2}$ for $per = H + hH$ with $G + 1 \le h < P_H$.

As in §3.3, we can formulate other restrictions on the time effects in terms of the double differences. The degrees of freedom are counted by counting the restrictions on the double differences.

References

Bell, A. & Jones, K. (2015). Should age-period-cohort analysts accept innovation without scrutiny? *Social Science & Medicine*, *128*, 331–333.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., & Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *Astin Bulletin*, *41*, 29–59.

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

Chauvel, L. & Schröder, M. (2014). Generational inequality and welfare regimes. *Social Forces*, *92*, 1259–1283.

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

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.

Fosse, E. & Winship, C. (2019). Bounding analyses of age-period-cohort effects. *Demography*, *56*, 1975–2004.

Fu, W. (2016). Constrained estimators and consistency of a regression model on a Lexis diagram. *Journal of the American Statistical Association*, *111*, 180–199.

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

Fu, W., Ding, J., Gao, K., Ma, S., & Tian, L. (2021). A likelihood ratio test on temporal trends in age-period-cohort models with application to the disparities of heart disease mortality among US populations and comparison with japan. *Statistics in Medicine*, *40*, 668–689.

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. (2018). Misspecification tests for log-normal and over-dispersed Poisson chain-ladder models. *risks*, *6*, 25.

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

Hendry, D. F. & Nielsen, B. (2007). *Econometric Modeling: A Likelihood Approach*. Princeton, NJ: Princeton University Press.

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.

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B*, *49*, 127–162.

Keiding, N. & Andersen, P. K. (2016). Linear effects of maternal age and period trends cannot be distinguished: Comment on an article by k. barclay and m. myrskylä. *Population and Development Review*, *42*, 711.

Kuang, D. & Nielsen, B. (2020). Generalized log-normal chain-ladder. *Scandinavian Actuarial Journal*, *2020*, 553–576.

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. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, *50*, 1945–1967.

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.

McKenzie, D. J. (2006). Disentangling age, cohort and time effects in the additive model. *Oxford Bulletin of Economics and Statistics*, *68*, 473–495.

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

Nielsen, B. (2022). Age-period-cohort analysis of mixed frequency data. Discussion paper 2022-w02, Nuffield College.

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

O'Brien, R. M. (2011). Constrained estimators and age-period-cohort models (with discussion). *Sociological Methods & Research*, *40*, 419–470.

O'Brien, R. M. (2022). Setting bounds on age, period, and cohort effects using observed data. https://doi.org/10.1007/s11135-022-01503-9.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Reither, E. N., Masters, R. K., Yang, Y. C., Powers, D. A., Zheng, H., & Land, K. C. (2015). Should age-period-cohort studies return to the methodologies of the 1970s? *Social Science & Medicine*, *128*, 365–365.

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

Riebler, A., Held, L., Rue, H., & Bopp, M. (2012). Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. *Journal of the Royal Statistical Society. Series A*, *175*, 479–490.

Rosenberg, P. S. (2019). A new age-period-cohort model for cancer surveillance research. *Statistical Methods in Medical Research*, *28*, 3363–3391.

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.

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