**Two-sample age-period-cohort models**
**Online Supplement**
**Appendix with technical results**

Bent Nielsen[1]

13 August 2025

# A Appendix: Technical results

## A.1 Representation and design vector, regular case

Recall the representation of the predictor in terms of the canonical parameter as given in (16). We will rewrite this on the form (18).

Combine (16) and (17). Insert $age = A - g$, $per = 1 + h$, so that $g = A - age$ and $h = per - 1$ while $coh = A - age + per = 1 + g + h$ with ranges $g = 0, 1, \ldots, A - 1$ and $h = 0, 1, \ldots, P - 1$. Swap summationg order of the double sums to get

$$\mu_{age,per,s} = \mu_{A,1,s} + g\lambda_s + h\nu_s + \sum_{t=0}^{g-2}(g-1-t)\Delta_1^2\alpha_{A-t,s} + \sum_{t=0}^{h-2}(h-1-t)\Delta_1^2\beta_{3+t,s} + \sum_{t=0}^{g+h-2}(g+h-1-t)\Delta_1^2\gamma_{3+t,s}. \quad (A.1)$$

This expresses the predictor linearly in terms of the canonical parameter with the same structure for the two samples. The common design vector is

$$x_{age,per} = \big\{1, g, h, m(g, A), \ldots, m(g, 2), m(h, 2), \ldots, m(h, P - 1), m(g + h, 2), \ldots, m(g + h, C)\big\}', \quad (A.2)$$

where $m(x, y) = \max(x - y + 1, 0)$. With this definition $\mu_{age,per,s} = \xi_s' x_{age,per}$. This is equivalent to $\mu_{age,per,s} = \xi_1' x_{age,per} 1_{(s=1)} + \xi_2' x_{age,per} 1_{(s=2)}$ as in (18).

## A.2 Common period effect, regular case

Consider the hypothesis of common period effects. We argue that this hypothesis does not imply that the linear age and cohort effects are identified. This is equivalent to arguing that the linear planes are not constrained by the common period effect hypothesis (21). To see that this is the case, consider the simpler linear plane model where non-linear time effects are left out (Fannon and Nielsen, 2019). For simplicity, we do this for a single sample representing the cross-sample difference. That is

$$\mu_{age,per} = \{\alpha_c + \alpha_\ell \times (A - age)\} + \{\beta_c + \beta_\ell \times per\} + \{\gamma_c + \gamma_\ell \times coh\} + \delta. \quad (A.3)$$

[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 was supported by ERC grant DisCont 694262.

This linear plane model is subject to the usual identification problem expressed in terms of the transformations (7), which are

$$\mu_{age,per} = \{(\alpha_c + a) + (\alpha_\ell + d) \times (A - age)\} + \{(\beta_c + b) + (\beta_\ell + d) \times per\} \\ + \{(\gamma_c + c) + (\gamma_\ell + d) \times coh\} + (\delta - a - b - c), \quad \text{(A.4)}$$

for any $a, b, c, d$ due to the identity $per = (age - A) + coh$ as per (2). In particular, if $b = -\beta_c$ and $d = \beta_\ell$ while $a = c = 0$, we get the observationally equivalent model

$$\mu_{age,per} = \{\alpha_c + \alpha_\ell^* \times (A - age)\} + \{\gamma_c + \gamma_\ell^* \times coh\} + \delta^*, \quad \text{(A.5)}$$

where $\delta^* = \delta + \beta_c$ as well as $\alpha_\ell^* = \alpha_\ell + \beta_\ell$ and $\gamma_\ell^* = \gamma_\ell + \beta_\ell$. Now, in this simplified setup, the common period effects hypothesis (21) corresponds to zero constraints on the linear period effects in (A.3), that is $\beta_c = \beta_\ell = 0$ yielding the 'restricted' model (A.5). As noted, the 'restricted' model (A.5) is observationally equivalent to the original model (A.3), so that the hypothesis does not constrain the model. Of course, if it was true that $\beta_\ell = 0$ then the age slope $\alpha_\ell$ would be identified. This truism cannot be confirmed statistically due to the observational equivalence of (A.3) and (A.5). This point was made by Clayton and Schifflers (1987). Keiding and 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.

Now, return to the hypothesis (21) of common period effects across the two samples. This results in an age-cohort model for the cross-sample difference. By the argument above, the cross-sample age difference $\alpha_{age,2} - \alpha_{age,1}$ is not invariant. But, of course, if it was true that $\beta_{per,2} - \beta_{per,1} = 0$ and that $\sum_{age} \alpha_{age,s} = 0$, then we would have the truism that the cross-sample age difference $\alpha_{age,2} - \alpha_{age,1}$ would be identified.

Equally, under the restriction of common age effects across the samples and if it was true that $\alpha_{1,2} - \alpha_{1,1} = \alpha_{2,2} - \alpha_{2,1} = 0$ and that $\sum_{per} \beta_{per,s} = 0$, we would have that $\beta_{per,2} - \beta_{per,1}$ is identified as argued by Riebler and Held (2010). Again, this is a truism that is not testable.

## A.3 Detailed representation of predictor, regular case

N22 represents the predictor as

$$\mu_{A-q_g GH - r_g, H + q_h GH + r_h, s} \\ = M^{intercept}_{r_g,r_h,s} + q_g M^{age/coh}_{r_g,r_h,s} + q_h M^{per/coh}_{r_g,r_h,s} + R^{age}_{q_g,r_g,s} + R^{per}_{q_h,r_h,s} + R^{coh}_{q_g+q_h,r_g,r_h,s}. \quad \text{(A.6)}$$

The levels and slopes are

$$M^{intercept}_{r_g,r_h} = \mu_{A,H} + 1_{(r_g>0)}\lambda_{r_g,s} + 1_{(r_h>0)}\nu_{r_h,s} \\ + 1_{(r_g>0)}1_{(r_h>0)}\Delta_{r_g G}\Delta_{r_h H}\gamma_{H+r_g G+r_h H,s}, \quad \text{(A.7)}$$

$$M^{age/coh}_{r_g,r_h,s} = \lambda_{H,s} + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_h H}\gamma_{H+GH+r_g G+r_h H,s} \\ + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_g G}\big(\gamma_{H+GH+r_g G,s} + \alpha_{A,s}\big), \quad \text{(A.8)}$$

$$M^{per/coh}_{r_g,r_h,s} = \nu_{G,s} + 1_{(r_g>0)}\Delta_{GH}\Delta_{r_g G}\gamma_{H+GH+r_g G+r_h H,s} \\ + 1_{(r_h>0)}\Delta_{GH}\Delta_{r_h H}\big(\gamma_{H+GH+r_h H,s} + \beta_{H+GH+r_h H,s}\big), \quad \text{(A.9)}$$

The macro step double differences can be written as sums of double differences. The idea is that for a sequence $x_i$ then $x_n - x_m = \sum_{i=m+1}^{n}(x_i - x_{i-1})$. We get, for instance, that $\Delta_{r_g G}\gamma_{H+r_g G+r_h H} = \sum_{u=1}^{r_g}\Delta_G\gamma_{H+uG+r_h H}$ and $\Delta_{r_h H}\gamma_{H+r_g G+r_h H} = \sum_{v=1}^{r_h}\Delta_G\gamma_{H+r_g G+vH}$. We include these double sums in the $S$-terms in (A.13), (A.15), (A.16).

The non-linear terms are

$$R^{age}_{q_g,r_g,s} = 1_{(q_g\geq 2)}\sum_{v=1}^{q_g-1}\sum_{u=0}^{vH-1}\Delta_{GH}\Delta_G\alpha_{A-(r_g+u)G,s}, \tag{A.10}$$

$$R^{per}_{q_h,r_h,s} = 1_{(q_h\geq 2)}\sum_{v=1}^{q_h-1}\sum_{u=1}^{vG}\Delta_{GH}\Delta_H\beta_{H+GH+(r_h+u)H,s}, \tag{A.11}$$

$$R^{coh}_{q_g+q_h,r_g,r_h,s} = 1_{(q\geq 2)}\sum_{v=1}^{q-1}\sum_{u=1}^{vG}\Delta_{GH}\Delta_H\gamma_{H+GH+r_g G+(r_h+u)H,s}. \tag{A.12}$$

The period double sum in (A.11) has the form $\sum_{v=1}^{q_h-1}\sum_{u=1}^{vG}f(u)$. Rewrite the inner sum as $\sum_{v=1}^{q_h-1}\sum_{t=0}^{v-1}\sum_{w=1}^{G}f(tG+w)$. Swap outer sums to get $\sum_{t=0}^{q_h-2}\sum_{v=t+1}^{q_h-1}\sum_{w=1}^{G}f(tG+w)$. The $v$-sum has $q_h - 1 - t$ elements, so we get $\sum_{t=0}^{q_h-2}(q_h-1-t)\sum_{w=1}^{G}f(tG+w)$. Rename $w$ as $v$ to get $\sum_{t=0}^{q_h-2}(q_h-1-t)\sum_{v=1}^{G}f(tG+v)$. This can be written out as

$$\begin{aligned} S^{per}_{q_h,r_h,s} &= 1_{(q_h\geq 2)}\sum_{t=0}^{q_h-2}(q_h-1-t)\sum_{v=1}^{G}\Delta_{GH}\Delta_H\beta_{H+GH+tGH+(v+r_h)H,s} \\ &\quad + 1_{(r_h>0)}q_h\sum_{v=1}^{r_h}\Delta_{GH}\Delta_H\beta_{H+GH+vH,s}. \end{aligned} \tag{A.13}$$

Whenever $r_h = 0$, the second term falls away. Moreover, since a sum of $G$ micro $\Delta_H$-differences is a macro $\Delta_{GH}$-difference, we can write the expression as

$$S^{per}_{q_h,0,s} = 1_{(q_h\geq 2)}\sum_{t=0}^{q_h-2}(q_h-1-t)\sum_{v=1}^{G}\Delta_{GH}\Delta_{GH}\beta_{H+(t+2)GH,s}. \tag{A.14}$$

In the regular case where $G = H = 1$ so that $h = q_h$, we recognize the expression from (A.1). Inspired by the Holford (2006) analysis, we will refer to $S^{per}_{q_h,0,s}$ in (A.14) as the macro period effect. For general $r_h$, the expression in (A.13) involves all period double differences with index $hH$ where $G+2 \leq h \leq H+q_h GH+r_h H = per$ and with declining weights. We will refer to the differences $S^{per}_{q_h,r_h,s} - S^{per}_{q_h,0,s}$ as micro period effects.

The age double sum of double difference in (A.10) is of the form $\sum_{v=1}^{q_g-1}\sum_{u=0}^{vH-1}f(u)$. Write the inner sum in macro and micro steps to get $\sum_{v=1}^{q_g-1}\sum_{t=0}^{v-1}\sum_{w=0}^{H-1}f(tH+w)$. Then swap the first two sums to get $\sum_{t=0}^{q_g-2}\sum_{v=t+1}^{q_g-1}\sum_{w=0}^{H-1}f(tH+w)$. Noting that the $v$-sum has $q_g - 1 - t$ elements, but that the $v$-index is not appearing elsewhere and renaming $w$ as $u$ gives $\sum_{t=0}^{q_g-2}(q_g-1-t)\sum_{u=0}^{H-1}f(tH+u)$. This can be written out as

$$\begin{aligned} S^{age}_{q_g,r_g,s} &= 1_{(q_g\geq 2)}\sum_{t=0}^{q_g-2}(q_g-1-t)\sum_{u=0}^{H-1}\Delta_{GH}\Delta_G\alpha_{A-tGH-(u+r_g)G,s} \\ &\quad + 1_{(r_g>0)}q_g\sum_{u=0}^{r_g-1}\Delta_{GH}\Delta_G\alpha_{A-uG,s}. \end{aligned} \tag{A.15}$$

The expression for $r_g = 0$ is the macro age effect. In the regular case where $G = H = 1$ this reduces to the expression in (A.1).

For the cohort double sum in (A.12), swap summation order as above and write out the macro difference $\Delta_{GH}$. Thus, the sum has the form $\sum_{v=1}^{q-1}\sum_{u=1}^{vG} f(u)$, where $f(u) = \Delta_{GH}k(GH+uH)$. As for the period sum, we get $\sum_{t=0}^{q-2}(q-1-t)\sum_{v=1}^{G} f(tG+v)$. The summand is $f(tG+v) = \Delta_{GH}k(tGH+GH+vH)$. Write out $\Delta_{GH}$ to get $f(tG+v) = \sum_{u=1}^{H} \Delta_G k(tGH + uG + vH)$. Insert in the previous sum to get, as in (A.16), that $\sum_{t=0}^{q-2}(q - 1 - t)\sum_{v=1}^{G}\sum_{u=1}^{H} \Delta_G k(tGH + uG + vH)$. This can be written out as

$$\begin{aligned} S^{coh}_{q_g,q_h,r_g,r_h,s} = {} & 1_{(q_g+q_h\geq 2)}\sum_{t=0}^{q-2}(q_g + q_h - 1 - t)\sum_{u=1}^{H}\sum_{v=1}^{G}\Delta_G\Delta_H\gamma_{H+tGH+(u+r_g)G+(v+r_h)H} \\ & + 1_{(r_h>0)}\sum_{u=1}^{H}\sum_{v=1}^{r_h}\Big(q_g\Delta_G\Delta_H\gamma_{H+uG+vH+r_gG,s} + q_h\Delta_G\Delta_H\gamma_{H+uG+vH,s}\Big) \\ & + 1_{(r_g>0)}\sum_{u=1}^{r_g}\sum_{v=1}^{G}\Big(q_g\Delta_G\Delta_H\gamma_{H+uG+vH,s} + q_h\Delta_G\Delta_H\gamma_{H+uG+vH+r_hH,s}\Big) \\ & + 1_{(r_g>0)}1_{(r_h>0)}\sum_{u=1}^{r_g}\sum_{v=1}^{r_h}\Delta_G\Delta_H\gamma_{H+uG+vH,s}. \end{aligned} \tag{A.16}$$

Again, when $r_g = r_h = 0$ only the first term remains and this is the macro cohort effect. In the regular case $G = H = 1$ this reduces to the expression in (A.1).

## A.4 Some remarks on Bayesian age-period-cohort analysis

A frequentist regression approach was used in the above analysis. Since the original analysis of Riebler and Held (2010) used a Bayesian approach, it is useful to highlight some consequences of the age-period-cohort problem in the Bayesian context.

Bayesian analysis tends to give a smooth impression of the parameters. A feature of the Bayesian analysis is that it will give an answer regardless of whether the likelihood is identified or not. This comes about as follows. The Bayesian approach seeks to update a prior distribution of the parameters through the likelihood to get the posterior distribution of the parameters. In more detail, suppose the likelihood is $p(x \mid \xi, \psi)$ for data $x$ and (vector) parameters $\xi, \psi$. To mimic the age-period-cohort problem, suppose $\xi$ is identified, but $\psi$ is not, so that the likelihood satisfies $p(x \mid \xi, \psi) = p(x \mid \xi)$ for all values of $\psi$. In this situation, the posterior distribution of the parameters given the data satisfies, see Nielsen and Nielsen (2014) for a derivation,

$$p(\xi, \psi \mid x) = p(\xi \mid x)p(\psi \mid \xi). \tag{A.17}$$

Thus, if one proceeds with the likelihood $p(x \mid \xi, \psi)$ without realizing that $\psi$ is not identified, then the Bayesian analysis will produce a posterior of the indicated form whereas maximum likelihood would stall and thus prompt further analysis.

The consequence of this posterior result for age-period-cohort modelling is as follows. If on the one hand, the investigator has full information about which parameters are identified and invariant, one can simply put priors on those and ignore the remaining parameters just as we have done for the above frequentist approach. This is the approach advocated by Smith and Wakefield (2016). If on the other hand the invariance structure is not fully known, as in the original analysis of the Swiss data, then one may inadvertently end up making probabilitic statements about parameters that are non-invariant.

For a single-sample of regular data, Smith and Wakefield (2016) suggested that priors should be put on the regular spaced double differences following Berzuini and Clayton (1994) and also on the parameters of the linear plane. This idea extends to the regular two-sample situation. When working with mixed frequency data the double differenced parameters are no longer invariant, so Gascoigne and Smith (2023) opted to use a spline approach. Alternatively, priors could be put on the invariant parameters discussed here.

# References


Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine*, 13:823–838.

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

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

Gascoigne, C. and Smith, T. (2023). Penalized smoothing splines resolve the curvature identifiability problem in age-period-cohort models with unequal intervals. *Statistics in Medicine*, 42:1888–1908.

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

Keiding, N. and 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.

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

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

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