The identifiable trend in APC models is not unique

Bendix Carstensen   Steno Diabetes Center Copenhagen, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bcar0029@regionh.dk   b@bxc.dk
http://BendixCarstensen.com

Contents

1 Abstract

Extraction of the secular trend from age-period-cohort models has routinely been done by Holford's method [1], mostly with reference to the trend as being identifiable. This means identifiable in a particular parametrization derived by orthogonal projection of columns of the model matrix on the linear age, period and cohort. Orthogonality however requires an inner product, and in the age-period-cohort setting the usual inner product assigning equal weight to all observational units is hardly appropriate because records contain different amounts of information.

This commentary details the problems and devises a family of inner products with desirable invariance properties that should be used for definition of orthogonality when devising the identifiable trend. Examples using R are provided.

Keywords: Age-period-cohort model, identifiable trend, inner product, orthogonality

2 Introduction

Both in the literature on age-period-cohort (APC) methodology as well as in publications using the methodology there are frequent references to the "identifiable trend", almost invariably with reference to Holford's 1983 paper [1]. This note explains why the trend is only identifiable in a very narrow sense and proposes another way of identifying a trend, different from the one suggested by Holford (but still identifiable) and which meets desirable invariance criteria that Holford's proposal does not.

Parts of this was already pointed out by me more than 10 years ago [2], but the precise details of inner products of relevance and their invariance properties were not hammered out then. Now they are.

Unlike the vast majority of the literature on APC-modeling the results of that paper applies to *any* set of subsets of observation of events and risk time in a Lexis diagram, regardless of shape and size. The only requirement is that mean age and date of follow-up in each subset is defined.

2.1 Terminology

I shall use the term "drift" for the parameter estimated in the age-drift model as coined by Clayton & Schifflers [3, 4], and the term "trend" for the parameter extracted from an age-period-cohort model that supposedly represents some overall trend in rates, while I shall reserve the term "slope" for the slope of a specific linear relationship. Finally I consider two models identical if they produce the same fitted values — a given model may of course be *parametrized* in different ways.

3 A general APC model

The setting for APC models is commonly a tabulation of events and risk-time (population exposure) by age and date of follow-up; or more generally a set of subsets of a Lexis diagram with events and risk-time [5].

Suppose the mean age and period of follow-up for each observational record (Lexis subset; table entry) are a and p, respectively. Then a general model that encompasses both the traditional factor-based models as well as models based on parametric functions of a and p can be written:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) + h(p-a) \tag{1}$$

This can be taken either as a model specification with three (usually smooth) functions of the *quantitative* variables a and p, but also as a rendering of a classical factor based model by defining each of the three functions to have a separate value for each separate value of the function argument. The latter is normally using the notation $f(a) = f_a$ etc.

Note that the model is over-parametrized if the functions f, g and h contain the linear effects of a, p and $p-a$ respectively — which will normally be the case.

4 Identifiable trend

The model (1) can be reparametrized by extracting a slope from each of the terms:

$$\log\big(\lambda(a,p)\big) = \tilde{f}(a) + \tilde{g}(p) + \tilde{h}(p-a) + \gamma a + \pi p + \delta(p-a)$$

Even for fixed $\tilde{f}$, $\tilde{g}$ and $\tilde{h}$, the parameters γ, π and δ are not uniquely determined; for any κ we can add 0 to the equation in the form of $0 = (p - a - c)\kappa$:

$$\log\big(\lambda(a.p)\big) = \tilde{f}(a) + \tilde{g}(p) + \tilde{h}(p-a) + (\gamma - \kappa)a + (\pi + \kappa)p + (\delta - \kappa)(p-a)$$

Thus we see that the *sum* of the coefficients to p (period) and $p-a$ (cohort) is invariant under adding the κ terms. The quantity $\pi + \delta$ is what is normally referred to as the identifiable trend, because it is invariant under change of κ, but of course it depends on the choice of $\tilde{g}$ and $\tilde{h}$.

The size of $\pi + \delta$ depends on the choice of $\tilde{g}$ and $\tilde{h}$, that is on what we mean by the term "extracting a slope". Holford [1] devised a particular reparametrization by setting up a specific design matrix for $\tilde{f}$, $\tilde{g}$ and $\tilde{h}$.[1].

This is where the indeterminacy of the identifiable trend is hidden — the exercise proposed by Holford is a *projection* of the design matrix columns for period, respectively cohort on the orthogonal complement of the subspaces spanned by the intercept and the linear period (P_L) respectively cohort (C_L) (Holford's notation).

However, the inner product used to define orthogonality (see section A3) in Holford's paper is without mention taken to be the standard one, putting equal weight on all observed records. But since each observed record represents a table entry, records do not contain the same amount of information, so there is no obvious argument for the standard inner product here.

5 Model invariance and data format

Suppose we are using a Poisson likelihood to model rate data, where each record corresponds to a table entry of events (D) and risk-time (Y). Now take any (set of)

---

[1]Incidentally, in the table 1 of Holford's paper there is a typo, as the 2nd column of the C_L matrix is not orthogonal to the linear cohort term — the 8th element should be -2 not 2. See the last section A3.1 for details.

record(s) with response (D, Y) and replace (each) with, say, two records with identical values of all covariates (in this case a, p and all codings of age, period and cohort effects), but where the D and the Y are split so that the *sum* of the Ds and the Ys across the two records equals the original (D, Y). Then the model fit will be exactly the same, and the parameters in the fitted model will be the same as when using the original data.

This is because the log-likelihood contributions from the two records sum to the likelihood contribution from the single original one — the log-likelihood is $D \log(\lambda) - \lambda Y$ which is additive in D and Y. Specifically, if $D = d_1 + d_2$ and $Y = y_1 + y_2$, replicating the record's covariates (age, period and cohort) but using d_1 respectively d_2 as events and y_1 respectively y_1 as risk time in the two records, the contribution to the log-likelihood would be (identical covariates implies identical rates, λ):

$$d_1 \log(\lambda) - \lambda y_1 + d_2 \log(\lambda) - \lambda y_2 = (d_1 + d_2) \log(\lambda) - \lambda(y_1 + y_2) = D \log(\lambda) - \lambda Y$$

This is a feature of the Poisson likelihood and as such unrelated to APC modeling.

If the parametrization in an APC model were set up as proposed by Holford, the period/cohort columns for the non-linear effects would not be orthogonal to the linear effect of period/cohort after subdividing a subset of the observational record(s) as above. So the trend extracted by using the standard inner product is not only a function of the model and the observed data, but also of the particular data lay-out.

But if the inner product used to define orthogonality were induced by a diagonal matrix where the entries were proportional to either D or Y, we would actually have orthogonality both for the original data and for the data where records were split (and the model fit would still be the same). This is because the contribution to the inner product between any two vectors from the records with split data would be the same as the contribution from the aggregate record.

Extracting the trend would thus *only* be a function of the observed data and the model, *not* of the particular data layout.

The invariance under subdivision of the tabular entries is of relevance because the records actually are *aggregate* observations; in the extreme we may have data where each record were the follow-up of a single persons in a subset of the Lexis diagram, with age and period coded as in the corresponding aggregate record. And certainly we would like not only the model but also any parameter extracted to be the same regardless of the aggregation level of the data (for a given coding of the covariates).

In fact this would also be the case for any inner product induced by a diagonal matrix with entries $w_i = q_d D_i + q_y Y_i$, $q_d \geq 0, q_y \geq 0$. Moreover, only inner products of this form has this property.

Since the projection properties of an inner product induced by a diagonal matrix does not change by multiplication of all elements of the diagonal by a positive constant, all the inner products of interest are induced by diagonal matrices with entries of the form $w_i = D_i + q Y_i$, where $q = 0$ corresponds to using D_i as weights and $q \to \infty$ corresponds to using Y_i as weights. These inner products would be natural choices for definition of orthogonality in extraction of trends in APC modeling of rates.

It should be noted that in cases with sparse event data, some records in data may contain 0 events ($D = 0$), causing the inner product based on D to ignore these records. This would be a reason to always include Y as part of the inner product weights.

5.1 The second order differences

Using factor-coded effects of age, period and cohort, Clayton & Schifflers [4] suggested to use the 2nd order differences of the parameters as these are invariant under reparametrization of the model. These can be obtained using the `contr.2nd` contrasts in the `Epi` package. Which incidentally is just a coding of the factor levels by a linear spline with knots at all points. This is only sensible if age and period classes are all of the same length.

One might think that the subspace spanned by these columns is the subspace orthogonal to the linear trend using an inner product where weights are chosen so that the *sum* of the weights is the same for all levels of the variable. This is however not the case; the 2nd order differences of the period/cohort effects will be the same regardless of the way we extract the linear trend. So the 2nd order differences is not related to the way the linear trend is extracted; they are always the same — provided of course the *model* is the factor model.

The trend extraction by projections leaves estimates of the non-linear part of period / cohort, on the (log-)RR scale; the second order differences are not on this scale, they are *curvature* parameters.

6 Practical implications for APC modeling

It is important to recognize that regardless of how the parametrization is constructed, the model fit is the same. So the discussion above only relates to the definition (*i.e.* extraction) of the trend as a particular *parameter* of the model.

Moreover, even if we restrict attention to inner products that have the invariance property outlined above, the choice is not unique.

In exploring the effect of different inner products over the range of the q defined above, we have to decide when q is large enough to effectively be ∞. We could check the size of the extracted trend for $q = (0, \infty)$, and using $K = \text{mean}(D_i)/\text{mean}(Y_i)$ (median or some other summary could be used as well) using values $q = K \times (0.2, 0.5, 1, 2, 5, 10)$. Note the necessity of a relative scaling of D and Y, since D is an absolute count and Y is scaled measurement (risk-time).

A more modest approach would be to just report the extracted trend using Y and D respectively; both of the latter options are implemented in the `apc.fit` function in the `Epi` package for R.

6.1 A worked example: Lung cancer in Denmark

The `Epi` package for R contains functions to fit APC models and extract trends. The package also contains a number of example data sets. Here we load the lung cancer data, which has male cases and risk time in 5-year Lexis triangles [2] in ages 40–90 for the period 1943–1998, incl. Hence the age and period are coded as the midpoints in the Lexis triangles.

```
> library( Epi )
> data( lungDK )
> lu <- lungDK[,c("Ax","Px","D","Y")]
> names( lu )[1:2] <- c("A","P")
> lu$Y <- lu$Y/10^5
> head( lu, 4 )
```

```
          A         P  D        Y
1 43.33333 1944.667 52 3.362338
2 41.66667 1946.333 28 3.578127
3 43.33333 1949.667 51 3.637837
4 41.66667 1951.333 30 3.909858
```

Different parametrizations of a model

We analyze data by APC models, using the default model with natural splines with 5 knots for each of the three effects, but with three different parametrizations — we see that the model fit is the same:

```
> apc.1 <- apc.fit( lu, parm="AdPC", dr.extr="N", ref.p=1980, pr=F )
> apc.d <- apc.fit( lu, parm="AdPC", dr.extr="D", ref.p=1980, pr=F )
> apc.y <- apc.fit( lu, parm="AdPC", dr.extr="Y", ref.p=1980, pr=F )
> c( apc.1$Model$deviance,
+    apc.d$Model$deviance,
+    apc.y$Model$deviance )
[1] 549.3752 549.3752 549.3752

> round( cbind( ci.exp(apc.1$Model),
+               ci.exp(apc.d$Model),
+               ci.exp(apc.y$Model) ), 3 )
```

	exp(Est.)	2.5%	97.5%	exp(Est.)	2.5%	97.5%	exp(Est.)	2.5%	97.5%
MA	101.293	99.206	103.425	103.481	101.415	105.588	89.855	87.918	91.835
MA1	2.818	2.728	2.910	3.172	3.072	3.274	3.273	3.172	3.378
MA2	2.907	2.834	2.981	3.368	3.288	3.451	3.503	3.420	3.589
MA3	7.783	7.532	8.043	10.067	9.764	10.379	10.782	10.468	11.106
MA4	1.952	1.909	1.997	2.352	2.306	2.400	2.472	2.419	2.527
I(P - p0)	1.032	1.031	1.033	1.020	1.019	1.020	1.020	1.019	1.021
MPr1	1.329	1.289	1.370	1.329	1.289	1.370	1.329	1.289	1.370
MPr2	1.252	1.210	1.296	1.252	1.210	1.296	1.252	1.210	1.296
MPr3	1.463	1.399	1.530	1.463	1.399	1.530	1.463	1.399	1.530
MCr1	1.234	1.197	1.272	1.234	1.197	1.272	1.234	1.197	1.272
MCr2	1.297	1.269	1.325	1.297	1.269	1.325	1.297	1.269	1.325
MCr3	2.085	2.023	2.149	2.085	2.023	2.149	2.085	2.023	2.149

We extracted the trend using different inner products to define orthogonality, and now show the trend parameters as annual percent changes (incidentally, these are in the `Drift` component of the `apc` object) — here we show the average annual trend:

```
> dr.1 <- apc.1$Drift
> dr.y <- apc.y$Drift
> dr.d <- apc.d$Drift
> round( org <- (rbind( dr.1[1,,drop=F],
+                       dr.d[1,,drop=F],
+                       dr.y )-1)*100, 2 )
                exp(Est.) 2.5% 97.5%
APC (1-weights)      3.23 3.15  3.30
APC (D-weights)      1.96 1.90  2.02
APC (Y-weights)      2.03 1.95  2.11
A-d                  2.35 2.30  2.40
```

We see that the standard inner product produces a substantially different trend estimate from the ones based on D or Y, and that they are on either side of the estimate of the drift from the age-drift model [3, 4].

A plot of the estimated effects (figure 1) shows the three different sets of effects, that however all multiply to the same fitted rates:

```
> par(mar=c(3,4,0.1,4),las=1,mgp=c(3,1,0)/1.6,lend="butt")
> lt <- c("solid","solid","11")
>  plot( apc.d, "Male lung cancer incidence rate per 100,000 PY",
+               col="red"  , lty=lt )
cp.offset    RR.fac
     1765       100

> lines( apc.1, col="black", lty=lt )
> lines( apc.y, col="blue" , lty=lt )
> text( 135, 400*c(0.85,1,0.85^2),
+       paste( c("D: ","1: ","Y: "),
+              formatC((c(dr.d[1,1],dr.1[1,1],dr.y[1,1])-1)*100,format="f",dig=2),
+              " %/year", sep=""),
+       col=c("red","black","blue"), adj=1 )
```

Uninformative data expansion

Now we duplicate some of the records so we have duplicate occurrences of some combinations of age and period (and hence cohort), but we redistribute events and risk time between the duplicate records:

```
> set.seed( 1952 )
> wh <- (1:nrow(lu))[lu$P>1980]
> ld <- lu[wh,]             # records for the latest periods
> wd <- runif( nrow(ld) ) # random fractions for events
> wy <- runif( nrow(ld) ) #                     and person-years
> l1 <- transform( ld, D=round(wd*D), Y=wy*Y ) # data frame with these fractions of D and Y
> l2 <- merge( l1, ld, by=c("A","P") )         # merge with the original data
>                      # merge renames non-by variables with .x and .y suffixes
> l2 <- transform( l2, D=D.y-D.x, Y=Y.y-Y.x )[,c("A","P","D","Y")] # complementary fraction
> # check that D and Y is the same for all combinations of A and P:
> range( xtabs( cbind( D,Y ) ~ A + P, data = ld ) -
+        xtabs( cbind( D,Y ) ~ A + P, data = l1 ) -
+        xtabs( cbind( D,Y ) ~ A + P, data = l2 ) )
[1] 0 0

> # so l1 + l2 together have the same events and PY as do ld
> lx <- rbind( lu[-wh,], l1, l2 ) # data frame with uninformatively duplicated records
```

Now we can check that we get the same fitted parameters in the sense that the likelihood ratio tests are the same for the two different lay-outs of data, and that the drift from the Age-drift model is the same:

```
> apc.fit( lu, dr.extr="N", ref.p=1990, ref.c=1930, pr=T )$Drift
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
              Model Mod. df.  Mod. dev. Test df. Test dev.     Pr(>Chi) Test dev/df
1               Age      215 15568.2387       NA        NA           NA          NA
2         Age-drift      214  6965.0731        1 8603.1656 0.000000e+00   8603.1656
3        Age-Cohort      211  1152.5297        3 5812.5433 0.000000e+00   1937.5144
4 Age-Period-Cohort      208   549.3752        3  603.1545 2.087058e-130   201.0515
5        Age-Period      211  3211.4282        3 2662.0530 0.000000e+00    887.3510
6         Age-drift      214  6965.0731        3 3753.6448 0.000000e+00   1251.2149
     H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
              exp(Est.)     2.5%    97.5%
APC (1-weights) 1.032266 1.031493 1.033039
A-d             1.023478 1.022963 1.023994
```

```
> apc.fit( lx, dr.extr="N", ref.p=1990, ref.c=1930, pr=T )$Drift

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
                Model Mod. df. Mod. dev. Test df. Test dev.      Pr(>Chi) Test dev/df
1                 Age      285  46148.16       NA        NA            NA          NA
2           Age-drift      284  37544.99        1 8603.1656  0.000000e+00   8603.1656
3          Age-Cohort      281  31732.45        3 5812.5433  0.000000e+00   1937.5144
4  Age-Period-Cohort      278  31129.30        3  603.1545 2.087058e-130    201.0515
5          Age-Period      281  33791.35        3 2662.0530  0.000000e+00    887.3510
6           Age-drift      284  37544.99        3 3753.6448  0.000000e+00   1251.2149
       H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
                 exp(Est.)     2.5%     97.5%
APC (1-weights)   1.027211 1.026466 1.027956
A-d               1.023478 1.022963 1.023994
```

We note that the estimates from the age-drift model (labeled `A-d`) and the likelihood-ratio tests are the same, no matter the data lay-out; they test the exact same hypothesis based on the same amount of information. The residual deviances should not be the same; they are each comparing the fit of the model (which is the same in the two cases) with the model that has one parameter per observational record (which is *not* the same in the two cases).

Then we extract the temporal trend using the standard inner product, as well as D and Y as weights in the inner product:

```
> dx.1 <- apc.fit( lx, dr.extr="N", ref.p=1980, ref.c=1910, pr=F )$Drift
> dx.d <- apc.fit( lx, dr.extr="D", ref.p=1980, ref.c=1910, pr=F )$Drift
> dx.y <- apc.fit( lx, dr.extr="Y", ref.p=1980, ref.c=1910, pr=F )$Drift
> xxx <- (rbind( dx.1[1,,drop=F], dx.d[1,,drop=F], dx.y )-1)*100
> round( cbind( org, xxx ), 4 )
                 exp(Est.)   2.5%  97.5% exp(Est.)   2.5%  97.5%
APC (1-weights)     3.2266 3.1493 3.3039    2.7211 2.6466 2.7956
APC (D-weights)     1.9640 1.9040 2.0240    1.9640 1.9040 2.0240
APC (Y-weights)     2.0291 1.9472 2.1110    2.0291 1.9473 2.1110
A-d                 2.3478 2.2963 2.3994    2.3478 2.2963 2.3994
```

The three leftmost columns here are the estimates using the original dataset; the three rightmost columns are from the uninformatively expanded dataset.

We see that the trend using the Ds or the Ys as weights for the inner product to define orthogonality is invariant under uninformative data-change, whereas using the standard inner product is not. In this lung cancer example, using the standard inner product leads to quite a substantial dependence on the data layout, even if the fitted model is precisely the same for the two different ways of entering data.

7 Discussion

The firm belief in the uniqueness of the temporal trend from an APC model comes from the formally correct use of the word "identifiable", which in statistical terminology refers to a particular *parametrization*, or more specifically a particular set of columns in the model matrix spanning the model space. Identifiability is *not* a property of the *model* (the space spanned by the model matrix), but of the particular parametrization. In this sense Holford's proposal with orthogonal projection w.r.t. the standard inner product is based on

identifiability in a *different parametrization* from the one(s) I propose by using another inner product to define orthogonality. In my proposal here, the trend is also identifiable in the usual statistical sense (by the exact same argument that Holford uses), but the parametrization I propose is different from the one proposed by Holford, and moreover offers invariance under uninformative changes in data lay-out.

8   Conclusion

It is not possible to specify a unique temporal trend in an APC model, as it depends on the inner product chosen to define orthogonality. It is however advisable to use either the number of events and/or the risk time (exposure, risk time) as weights of the inner product when defining orthogonality due to the invariance property of these. An argument can be made for either; the number of events is the information about the log-rate, whereas the amount of follow-up is the information about the square root of the rate — see the technical sections (A2).

9   Acknowledgement

References

[1] TR Holford. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39:311–324, 1983.

[2] B Carstensen. Age-Period-Cohort models for the Lexis diagram. *Statistics in Medicine*, 26(15):3018–3045, July 2007.

[3] D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Statistics in Medicine*, 6:449–467, 1987.

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

[5] W Lexis. *Einleitung in die Theorie der Bevölkerungsstatistik*. Karl J Trübner, Strassburg, 1875.
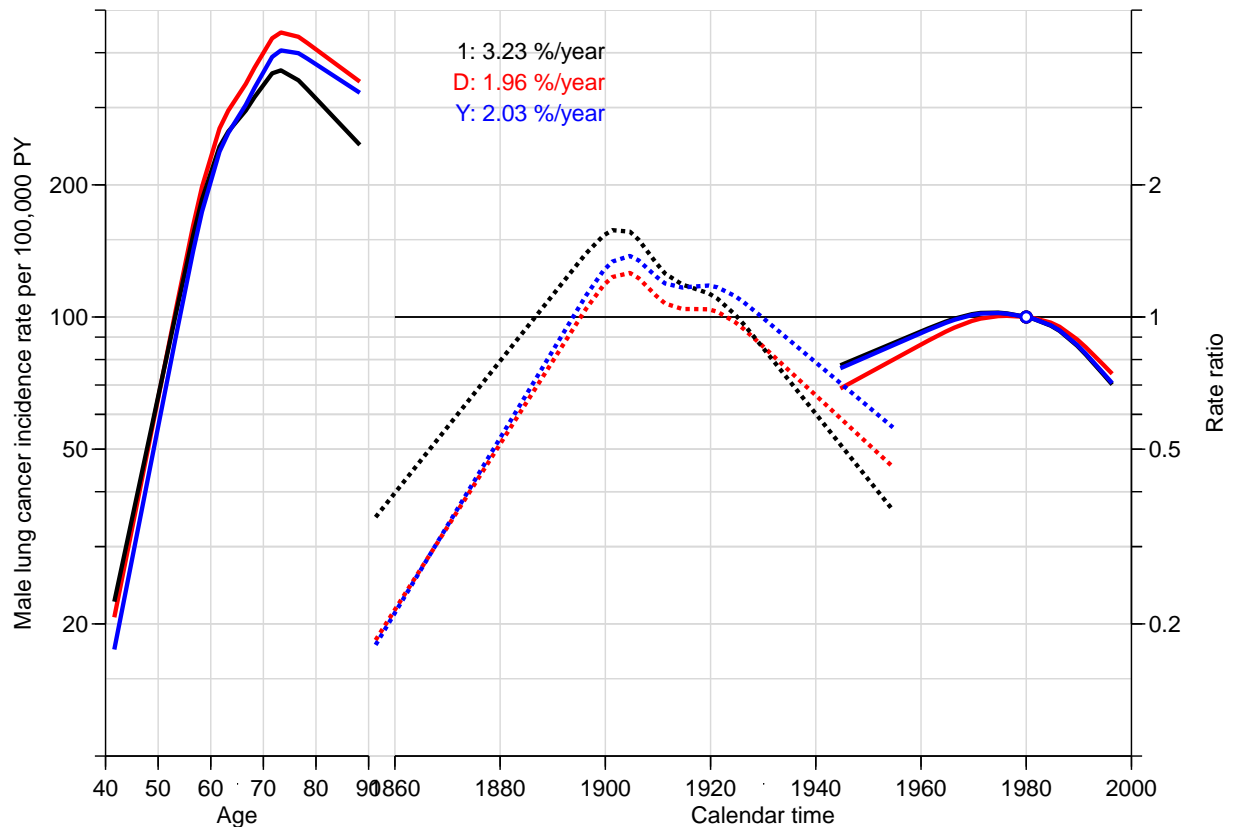
Figure 1: *Three different parametrizations of an APC-model for male lung cancer incidence in Denmark. Black is the traditional projection, red is using no. of deaths as weights and blue is the one using the person-years as weights in the inner product defining orthogonality. All three sets of curves multiply together to the same fitted values. The cohort curves (dotted lines) are 0 on average and has 0 slope on average, but with respect to different inner products — they differ only by affine terms. The period curves have average slopes equal to the extracted drifts but with respect to different inner products; they are constrained to be 1 at 1980 and so differ by terms proportional to $(p - 1980)$.*

A1 Examples

In the `Epi` package are functions for projections using different inner products (`projection.ip`, `detrend`) as well as an `apc.fit` function that fits an APC model (and all the submodels) and returns (among other things) the extracted trend, using a specified inner product. So far the only possible choices of relevance here are `dr.extr="y"` or `dr.extr="d"` and the traditional `dr.extr="1"`, with obvious meaning. Any inner product induced by a diagnonal matrix can be accommodated by supplying the diagonal entries as the argument `dr.extr`.

The models fitted by default by `apc.fit` are using natural splines (restricted cubic splines) to model the effects of age, period and cohort, placing the knots (5 per default) so that the number of events is the same between each pair of knots, and half of this number outside of each of the boundary knots.

A1.1 Lung cancer in Denmark: factor model

Here we use the same data, but initially aggregate to 5×5 age by period sets (so-called A-sets in the Lexis diagram):

```
> library(Epi)
> data( lungDK )
> names( lu <- aggregate( lungDK[,c("D","Y")],
+                         lungDK[,c("A5","P5")],
+                         FUN=sum ) )
[1] "A5" "P5" "D"  "Y"

> names( lu )[1:2] <- c("A","P")
> with( lu, table(A,P) )
    P
A    1943 1948 1953 1958 1963 1968 1973 1978 1983 1988 1993
  40    1    1    1    1    1    1    1    1    1    1    1
  45    1    1    1    1    1    1    1    1    1    1    1
  50    1    1    1    1    1    1    1    1    1    1    1
  55    1    1    1    1    1    1    1    1    1    1    1
  60    1    1    1    1    1    1    1    1    1    1    1
  65    1    1    1    1    1    1    1    1    1    1    1
  70    1    1    1    1    1    1    1    1    1    1    1
  75    1    1    1    1    1    1    1    1    1    1    1
  80    1    1    1    1    1    1    1    1    1    1    1
  85    1    1    1    1    1    1    1    1    1    1    1
```

We fit the model (using the exchangeable factor parametrization for the three effects) and extract the trend using different inner products, and display this as annual percent changes:

```
> dr.1 <- apc.fit( lu, dr.extr="1", model="factor", pr=F )$Drift
No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> dr.d <- apc.fit( lu, dr.extr="d", model="factor", pr=F )$Drift
No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> dr.y <- apc.fit( lu, dr.extr="y", model="factor", pr=F )$Drift
No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> round( org <- (rbind( dr.1[1,,drop=F], dr.d[1,,drop=F], dr.y )-1)*100, 2 )
```

```
              exp(Est.) 2.5% 97.5%
APC (1-weights)    3.28 3.15  3.40
APC (D-weights)    1.99 1.93  2.05
APC (Y-weights)    2.13 2.04  2.23
A-d                2.36 2.31  2.41
```

Again we see that the standard inner product produces a substantially different trend estimate from the ones based on D or Y. And that these are on either side of the estimate of the drift from the age-drift model [3, 4].

Now we duplicate some of the records so we have duplicate occurrences of some combinations of age and period (and hence cohort), but we redistribute events and risk time between the duplicates:

```
> set.seed( 1952 )
> wh <- (1:nrow(lu))[lu$P>1980]
> ld <- lu[wh,]
> wd <- runif( nrow(ld) )
> wy <- runif( nrow(ld) )
> l1 <- transform( ld, D=round(wd*D), Y=wy*Y )
> l2 <- merge( l1, ld, by=c("A","P") )
> l2 <- transform( l2, D=D.y-D.x, Y=Y.y-Y.x )[,c("A","P","D","Y")]
> # check that D and Y is the same for all combinations of A and P:
> range( xtabs( cbind( D,Y ) ~ A + P, data = ld ) -
+        xtabs( cbind( D,Y ) ~ A + P, data = l1 ) -
+        xtabs( cbind( D,Y ) ~ A + P, data = l2 ) )
[1] 0 0

> lx <- rbind( lu[-wh,], l1, l2 )
```

Now we can check that we get the same fitted parameters in the sense that the likelihood ratio tests are the same for the two different lay-outs of data, and that the drift parameter from the Age-drift model is the same:

```
> apc.fit( lu, dr.extr="1", model="factor", pr=T )$Drift
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
              Model Mod. df.  Mod. dev. Test df. Test dev.      Pr(>Chi) Test dev/df
1               Age        100 15103.0013       NA        NA            NA          NA
2         Age-drift         99  6417.3811        1 8685.6202  0.000000e+00  8685.62024
3        Age-Cohort         81   829.6293       18 5587.7518  0.000000e+00   310.43065
4 Age-Period-Cohort         72   208.5476        9  621.0817 6.244637e-128    69.00907
5        Age-Period         90  2723.4660       18 2514.9184  0.000000e+00   139.71769
6         Age-drift         99  6417.3811        9 3693.9151  0.000000e+00   410.43501
      H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .
              exp(Est.)     2.5%     97.5%
APC (1-weights)  1.032769 1.031537 1.034003
A-d              1.023580 1.023065 1.024096

> apc.fit( lx, dr.extr="1", model="factor", pr=T )$Drift

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
              Model Mod. df. Mod. dev. Test df. Test dev.      Pr(>Chi) Test dev/df
1               Age        130 51089.37       NA        NA            NA          NA
2         Age-drift         129 42403.75        1 8685.6202  0.000000e+00  8685.62024
3        Age-Cohort         111 36816.00       18 5587.7518  0.000000e+00   310.43065
4 Age-Period-Cohort         102 36194.92        9  621.0817 6.244637e-128    69.00907
```

```
5         Age-Period    120  38709.84      18 2514.9184  0.000000e+00  139.71769
6          Age-drift    129  42403.75       9 3693.9151  0.000000e+00  410.43501
     H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .
                exp(Est.)    2.5%    97.5%
APC (1-weights)  1.028019 1.026911 1.029128
A-d              1.023580 1.023065 1.024096
```

We note that the estimates from the age-drift model (labeled `A-d`) and likelihood-ratio tests are the same, no matter the data lay-out; they test the exact same hypothesis based on the same amount of information. The residual deviances should not be the same; they are each comparing the fit of the model (which is the same in the two cases) with the model that has one parameter per observational record (which is *not* the same in the two cases).

Then we extract the temporal trend using the standard inner product, as well as D and Y as weights in the inner product:

```
> dx.1 <- apc.fit( lx, dr.extr="1", model="factor", pr=F )$Drift

No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> dx.d <- apc.fit( lx, dr.extr="d", model="factor", pr=F )$Drift

No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> dx.y <- apc.fit( lx, dr.extr="y", model="factor", pr=F )$Drift

No reference cohort given;  reference cohort for age-effects is chosen as
 the median date of birth for persons  with event:  1913 .

> xxx <- (rbind( dx.1[1,,drop=F], dx.d[1,,drop=F], dx.y )-1)*100
> round( cbind( org, xxx ), 4 )

                exp(Est.)    2.5%  97.5% exp(Est.)    2.5%  97.5%
APC (1-weights)    3.2769  3.1537 3.4003    2.8019  2.6911 2.9128
APC (D-weights)    1.9870  1.9272 2.0468    1.9870  1.9272 2.0468
APC (Y-weights)    2.1348  2.0444 2.2253    2.1348  2.0444 2.2253
A-d                2.3580  2.3065 2.4096    2.3580  2.3065 2.4096
```

So we see that using the Ds or the Ys as weights for the inner product to define orthogonality is invariant under data-change, whereas using the standard inner product is not. In this lung cancer example, using the standard inner product leads to quite a substantial dependence on the data layout, even if the fitted model is precisely the same for the two different ways of entering data.

A1.2 Testis cancer in Denmark

In this example we load the Danish testis cancer data which is counts and risk time in 1-year age and period classes (so-called A-sets of the Lexis diagram), restrict the age-range and analyze it by APC models using different ways of extracting the trend.

```
> library( Epi )
> data( testisDK )
> # restrict to the clinically relevant age-range
> tr <- subset( testisDK, A>14 & A<65 )
> # code by midpoint of intervals
> tr$A <- tr$A + 0.5
> tr$P <- tr$P + 0.5
> head( tr )
      A      P D       Y
16 15.5 1943.5 0 31188.00
17 16.5 1943.5 0 31654.33
18 17.5 1943.5 0 32084.33
19 18.5 1943.5 0 32681.67
20 19.5 1943.5 0 33005.67
21 20.5 1943.5 0 33302.67
```

We fit the model and extract the trend using different inner products, and display this as annual percent changes:

```
> dr.1 <- apc.fit( tr, dr.extr="1", ref.p=1990, ref.c=1950, pr=F )$Drift
> dr.d <- apc.fit( tr, dr.extr="d", ref.p=1990, ref.c=1950, pr=F )$Drift
> dr.y <- apc.fit( tr, dr.extr="y", ref.p=1990, ref.c=1950, pr=F )$Drift
> round( org <- (rbind( dr.1[1,,drop=F], dr.d[1,,drop=F], dr.y )-1)*100, 2 )
                 exp(Est.) 2.5% 97.5%
APC (1-weights)       2.67 2.48  2.85
APC (D-weights)       2.61 2.45  2.77
APC (Y-weights)       2.67 2.50  2.84
A-d                   2.55 2.40  2.71
```

Here we see much less difference between the extracted trends using the different definitions of orthogonality.

Now we duplicate some of the records in the last years, but redistribute events and risk time so the sum remains the same:

```
> set.seed( 1952 )
> wh <- (1:nrow(tr))[tr$P>1970] # which rows to duplicate
> td <- tr[wh,]                 # a copy of these
> wd <- runif( nrow(td) )       # fraction of events
> wy <- runif( nrow(td) )       # fraction of risk time
> t1 <- transform( td, D=round(wd*D), Y=wy*Y ) # events and risk time
> t2 <- merge( t1, td, by=c("A","P") )          # match with original
> t2 <- transform( t2, D=D.y-D.x, Y=Y.y-Y.x )[,c("A","P","D","Y")]
>                             # the complementary events and risk time
> # check we have the same no in td as in t1 and t2 together
> range( xtabs( cbind( D,Y ) ~ A + P, data = td )
+       -xtabs( cbind( D,Y ) ~ A + P, data = t1 )
+       -xtabs( cbind( D,Y ) ~ A + P, data = t2 ) )
[1] 0 0

> # replace with the duplcated records
> tx <- rbind( tr[-wh,], t1, t2 )
```

Now we can check that we got the same fit to data by comparing the old fit with results based on a data set where a subset of records have been split in two:

```
> apc.fit( tr, dr.extr="1", ref.p=1990, ref.c=1950, pr=T )$Drift
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
              Model Mod. df. Mod. dev. Test df.  Test dev.      Pr(>Chi) Test dev/df
1               Age     2695  4235.078       NA         NA            NA          NA
2         Age-drift     2694  3048.316        1 1186.76204 4.595058e-260 1186.762037
3        Age-Cohort     2691  3022.611        3   25.70531  1.099303e-05    8.568438
4 Age-Period-Cohort     2688  2964.336        3   58.27421  1.373608e-12   19.424738
```

```
5         Age-Period    2691  3025.982         3   61.64591  2.615369e-13    20.548638
6          Age-drift    2694  3048.316         3   22.33361  5.559277e-05     7.444538
     H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
              exp(Est.)     2.5%    97.5%
APC (1-weights)  1.026677 1.024809 1.028548
A-d              1.025537 1.024020 1.027056

> apc.fit( tx, dr.extr="1", ref.p=1990, ref.c=1950, pr=T )$Drift

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
              Model Mod. df. Mod. dev. Test df.   Test dev.      Pr(>Chi) Test dev/df
1               Age    4045 10889.653       NA          NA            NA          NA
2         Age-drift    4044  9702.891        1 1186.76204 4.595058e-260 1186.762037
3       Age-Cohort    4041  9677.186        3   25.70531  1.099303e-05    8.568438
4 Age-Period-Cohort    4038  9618.911        3   58.27421  1.373608e-12   19.424738
5        Age-Period    4041  9680.557        3   61.64591  2.615369e-13   20.548638
6         Age-drift    4044  9702.891        3   22.33361  5.559277e-05    7.444538
     H0
1
2 zero drift
3 Coh eff|dr.
4 Per eff|Coh
5 Coh eff|Per
6 Per eff|dr.
              exp(Est.)     2.5%    97.5%
APC (1-weights)  1.026321 1.024618 1.028026
A-d              1.025537 1.024020 1.027056
```

The printed results shows that the analyses based on the two different data sets give the same models and tests, but not the same deviances and not the same extracted trend. Now we fit the model again and extract the trend in different ways:

```
> dx.1 <- apc.fit( tx, dr.extr="1", ref.p=1990, ref.c=1950, pr=F )$Drift
> dx.d <- apc.fit( tx, dr.extr="d", ref.p=1990, ref.c=1950, pr=F )$Drift
> dx.y <- apc.fit( tx, dr.extr="y", ref.p=1990, ref.c=1950, pr=F )$Drift
> xxx <- (rbind( dx.1[1,,drop=F], dx.d[1,,drop=F], dx.y )-1)*100
> round( cbind( org, xxx ), 4 )
                exp(Est.)   2.5%   97.5% exp(Est.)   2.5%   97.5%
APC (1-weights)    2.6677 2.4809 2.8548    2.6321 2.4618 2.8026
APC (D-weights)    2.6088 2.4523 2.7655    2.6088 2.4523 2.7655
APC (Y-weights)    2.6700 2.4970 2.8433    2.6700 2.4970 2.8433
A-d                2.5537 2.4020 2.7056    2.5537 2.4020 2.7056
```

So we see that using the Ds or the Ys as weights for the inner product to define orthogonality is invariant under uninformative change in data layout, whereas using the standard inner product is not; however in this example with much smaller effect than in the lung cancer example.

A2 Information from the log-likelihood

The log-likelihood for a rate parameter λ in a constant rate model, based on observation of D events during Y risk time is:

$$\ell(\lambda) = D \log(\lambda) - \lambda Y$$

and thus $\hat{\lambda} = D/Y$.

The observed information is minus the 2$^{\text{nd}}$ derivative of ℓ evaluated at $\hat{\lambda}$. By that token the information about any transformation of λ has a dimension which is the dimension of the transformed parameter to the power of -2. So if we are looking for transformation of λ with observed information D (dimensionless) we would look at $\log(\lambda)$, and if we are looking for one with observed information Y (dimension: time) we would look at $\sqrt{\lambda}$ (dimension: time$^{-1/2}$).

Specifically, the information about $\theta = \log(\lambda)$ is as follows:

$$\ell(\theta) = D\theta - \exp(\theta)Y$$
$$\ell'(\theta) = D - \exp(\theta)Y$$
$$I(\theta) = -\ell''(\theta)\Big|_{\theta=\hat{\theta}} = \exp(\theta)Y\Big|_{\theta=\hat{\theta}} = (D/Y)Y = D$$

... and the information about $\gamma = \sqrt{\lambda}$ is as follows:

$$\ell(\gamma) = D\log(\gamma^2) - \gamma^2 Y$$
$$\ell'(\gamma) = 2D\gamma/\gamma^2 - 2\gamma Y = 2D/\gamma - 2\gamma Y$$
$$I(\gamma) = -\ell''(\gamma)\Big|_{\gamma=\hat{\gamma}} = D/\gamma^2 + 2Y\Big|_{\gamma=\hat{\gamma}} = D/(D/Y) + 2Y = 3Y$$

Above we showed that in order to have invariance of trend extraction under uninformative change in data layout, we need an inner product in the observation space which is proportional to a linear combination of D and Y. Thus the weights that we can combine to give inner products that render a trend extraction with this property are combinations of the information about the log rate and the square root of the rate.

A3   Orthogonality and inner product

Two vectors x and y are said to be orthogonal if their *inner product* is 0:

$$x \perp y \quad \Leftrightarrow \quad \langle x|y \rangle = 0$$

The standard inner product of two vectors in a finite-dimensional linear space is defined as the bivariate scalar-valued function:

$$\langle x|y \rangle = \sum_i x_i y_i$$

A general inner product is induced by a positive semi-definite matrix W, as:

$$\langle x|y \rangle_W = x'Wy \geq 0, \quad \forall x, y$$

where the inequality is the definition of semi-definite.

The most common choice of W is as a diagonal matrix with non-negative entries; if $W = \text{diag}(w_1, \ldots, w_n)$, then:

$$\langle x|y \rangle = \sum_i x_i w_i y_i$$

Orthogonality properties of an inner product are invariant under multiplication of the diagonal elements by a positive constant.

The standard inner product is the one induced by the identity matrix, $w_i = 1$, $\forall i$.

Formulae for projection w.r.t. a general inner product can be found in any treatise on linear algebra, but are also briefly mentioned in [2] in connection with practical implementation details of `projection.ip` in the `Epi` package.

A3.1 Table 1 of Holford's paper[1]

Here is an R-representation of table 1:

```
> A <- rep(1:3,each=3)
> P <- rep(1:3,3)
> C <- P-A+3
> Ac <- rep(c(1,-2,1),each=3)
> Pc <- rep(c(1,-2,1),3)
> Cc <- cbind( Cc1=c(-2,-1, 2,-1,-2,-1, 2,-1,-2),
+              Cc2=c( 0,-2, 1, 2, 0,-2,-1, 2, 0),
+              Cc3=c( 6,-4, 1,-4, 6,-4, 1,-4, 6) )
> Al <- A-2
> Pl <- P-2
> Cl <- Pl-Al
> cbind( A,P,C,Ac,Pc,Cc,Al,Pl,Cl)
      A P C Ac Pc Cc1 Cc2 Cc3 Al Pl Cl
 [1,] 1 1 3  1  1  -2   0   6 -1 -1  0
 [2,] 1 2 4  1 -2  -1  -2  -4 -1  0  1
 [3,] 1 3 5  1  1   2   1   1 -1  1  2
 [4,] 2 1 2 -2  1  -1   2  -4  0 -1 -1
 [5,] 2 2 3 -2 -2  -2   0   6  0  0  0
 [6,] 2 3 4 -2  1  -1  -2  -4  0  1  1
 [7,] 3 1 1  1  1   2  -1   1  1 -1 -2
 [8,] 3 2 2  1 -2  -1   2  -4  1  0 -1
 [9,] 3 3 3  1  1  -2   0   6  1  1  0
```

This is the table 1 in the paper, but there is a typo in `Cc` — the 2nd column is not orthogonal to `Cl` by the standard inner product:

```
> t(Cl) %*% Cc
     Cc1 Cc2 Cc3
[1,]   0  -4   0
```

... the 8th element in `Cc2` is 2, but it should be -2:

```
> Cc[8,2] <- -2
> t(Cl) %*% Cc
     Cc1 Cc2 Cc3
[1,]   0   0   0
```

If we want the columns to be orthogonal to the intercept too we just subtract the means:

```
> cnt <- matrix(apply(Cc,2,mean),9,3,byrow=TRUE)
> matrix(1,1,9) %*% (Cc-cnt)
               Cc1 Cc2          Cc3
[1,] -4.440892e-16   0 -3.552714e-15

> t(Cl) %*% (Cc-cnt)
     Cc1 Cc2 Cc3
[1,]   0   0   0
```