Identifiability in Age/Period/Cohort Mortality Models

Andrew Hunt*
Cass Business School, City University London
Corresponding author: andrew.hunt.1@cass.city.ac.uk†

David Blake
Pensions Institute, Cass Business School, City University London

September 2015

Abstract

The addition of a set of cohort parameters to a mortality model can generate complex identifiability issues due to the collinearity between the dimensions of age, period and cohort. These issues can lead to robustness problems and difficulties making projections of future mortality rates. Since many modern mortality models incorporate cohort parameters, we believe that a comprehensive analysis of the identifiability issues in age/period/cohort mortality models is needed. In this paper, we discuss the origin of identifiability issues in general models before applying these insights to simple but commonly used mortality models. We then discuss how to project mortality models so that our forecasts of the future are independent of any arbitrary choices we make when fitting a model to data in order to identify the historical parameters.

*We are grateful to Bent Nielsen for his advice in improving this paper immeasurably, to Matthias Börger, Andrew Cairns and Pietro Millossovich for their helpful comments on earlier drafts of this paper, and to Andrés Villegas for many useful discussions on this and related topics.

†All comments and criticisms are gratefully received and we hope to be able to respond to them in further revisions of this paper.

JEL Classification: C15, C51, C53
Keywords: Mortality modelling, age/period/cohort models, identification issues, projection

1 Introduction

Many modern models of mortality include parameters to capture the impact of lifelong mortality effects which follow individuals from birth, building on the findings of studies such as Wilmoth (1990) and Willets (1999, 2004). Understanding such "cohort" effects can be of critical importance, especially for those interested in understanding the mortality experience of a specified group of lives, such as members of a pension scheme or policyholders in an annuity book. Examples of models incorporating cohort parameters include those proposed in Renshaw and Haberman (2006), Cairns et al. (2009), Plat (2009), O'Hare and Li (2012), Börger et al. (2013) and Hunt and Blake (2014).

In Hunt and Blake (2015f), we argued that the time has come to undertake a more holistic analysis of the class of age/period/cohort (APC) models and began this analysis by outlining their common structure. In Hunt and Blake (2015d), we focused on the subset of this class without a cohort term, namely on age/period (AP) models, and examined their identifiability issues.

We found that, for AP models, there are a number of "invariant transformations" which change the parameters, but not the fitted mortality rates. The existence of these transformations lead to identifiability issues, meaning that there are certain features of the parameters in a model which are not defined by the data. Instead, they are only determined by the arbitrary identifiability constraints we impose, and therefore have no independent meaning. Consequently, we must be careful to ensure that our results from using mortality models do not depend upon these features of the parameters. These issues with identifiability can lead to models which lack robustness when fitted to data, cause us to draw faulty and erroneous conclusions when analysing the historical data, and bias our projected mortality rates in future. We also found that, unless we choose our projection methods carefully, our projections of mortality can depend upon the arbitrary choice of identifiability

constraint. This should be avoided, so we discussed how to choose projection methods which give "well-identified" projections of mortality rates.

The addition of a set of cohort parameters to a mortality model can generate additional identifiability issues which are fundamentally unlike anything present in otherwise similar AP models. These are caused by the collinearity between age, period and cohort. In the context of the APC mortality models discussed in this study, we find that certain deterministic trends found within the fitted parameters are unidentifiable by the models, and therefore do not possess any meaning other than that imposed by our arbitrary identifiability constraints. This, in turn, means that it is both more important and more difficult to ensure that projections from these models are well-identified, as we must separate these unidentified trends (which depend entirely upon the identifiability constraints) from the variation around the trends, which is meaningful and needs to be projected consistently with what has been observed in the past. Thus, although the present study extends the work of Hunt and Blake (2015d), it is necessary to view the underlying identifiability issues in a fundamentally different way and, consequently, develop a new set of tools to solve them.

In this paper, we study the identifiability issues present in some of the simplest APC models in order to demonstrate the problems in action and their potential resolution. In these simple cases, the identifiability issues can appear trivial, and their impact on our analysis of historical and projected mortality rates relatively minor. However, we believe that it is vital to fully understand these issues in the context of simple models, since they become considerably more important in more complicated models. Indeed, recognising these issues and solving them was vital to the development of the "general procedure" for constructing APC mortality models, described in Hunt and Blake (2014), and appropriately projecting such models, as we discuss in Hunt and Blake (2015b), Hunt and Blake (2015a) Hunt and Blake (2015c) and Hunt and Blake (2015e).

The outline of the paper is as follows. Section 2 reviews the structure of general APC mortality models described in Hunt and Blake (2015f). Section 3 introduces the concept of identifiability in the context of the simplest and most widely used APC model and develops our understanding of how cohort effects create fundamentally new identification issues in this model compared

with the simpler AP model. Section 4 generalises this by examining the issue of identifiability in more general APC models with parametric age functions. Section 5 investigates the consequences of identification for projection, first by looking at the model discussed in Section 3 and then in a more general case. Finally, Section 6 concludes.

2 Structure of age/period/cohort models

An APC mortality model is one which assumes that mortality rates can be modelled as a series of terms involving functions of age, x, period, t, and year of birth, $y = t - x$.[1] This can be written as

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x} \tag{1}$$

where

- $\eta_{x,t}$ is a link function to transform the response variable into a form suitable for modelling and linking it to the proposed predictor structure;
- α_x is a static function of age;[2]
- $\kappa_t^{(i)}$ are period functions governing the evolution of mortality with time;
- $\beta_x^{(i)}$ are age functions modulating the impact of the period function dynamics over the age range; and
- γ_y is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as as it ages.

We also note that the general APC mortality model in Equation 1 can be re-written as

$$\eta_{x,t} = \alpha_x + \boldsymbol{\beta}_x^{\top} \boldsymbol{\kappa}_t + \gamma_{t-x} \tag{2}$$

[1] In this paper we assume that $x \in [1, X]$ and $t \in [1, T]$ and hence that years of birth, y, are in the range $(1 - X)$ to $(T - 1)$. In practice, x and t will be given by the range of the data being used.

[2] We consider models of the form of Equation 1 but without a static age function in Appendix B.

where

$$\boldsymbol{\kappa}_t = \left(\kappa_t^{(1)}, \quad \ldots \quad \kappa_t^{(N)}\right)^\top$$
$$\boldsymbol{\beta}_x = \left(\beta_x^{(1)}, \quad \ldots \quad \beta_x^{(N)}\right)^\top$$

This form is useful when projecting these models, as discussed in Section 5.

The general structure of APC models was discussed in detail in Hunt and Blake (2015f). In particular, we found that APC mortality models have different demographic significance[3] depending on whether the age functions $\beta_x^{(i)}$ are non-parametric[4] or parametric.[5]

In Hunt and Blake (2015d), we used linear algebra to analyse the structure of AP mortality models as mappings from a space of parameters to a model space, and found that in order for these mapping to be unique, the spaces had to have the same dimension. In addition, AP models can be subdivided into those with parametric age functions and those where the age functions are non-parametric. While the two families have similar identifiability issues, these needed to be solved using different methods in order to preserve the demographic significance of the parametric age functions.[6] It is important to note that AP mortality models are nested within the class of APC models, and, therefore, all of the issues raised in Hunt and Blake (2015d) are still applicable for APC mortality models.

APC models have additional identifiability issues which are fundamentally different from anything present in otherwise similar AP models, hence alternative methods are necessary to analyse them. They are caused by the

[3]Demographic significance is defined in Hunt and Blake (2015f) as the interpretation of the components of the mortality model being explainable in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates.

[4]The values of the age functions $\beta_x^{(i)}$ at different ages x are fitted without any a priori structure or functional form. See Hunt and Blake (2015f).

[5]The age functions $\beta_x^{(i)}$ take a specific functional form $\beta_x^{(i)} = f^{(i)}(x;\theta^{(i)})$, defined in advance of fitting the model to data. For simplicity, the dependence of the age functions on $\theta^{(i)}$ is suppressed in the remainder of this paper.

[6]These different methods are not germane to the arguments in this paper. Interested readers should consult Hunt and Blake (2015d).

collinearity between the dimensions of age, period and cohort, because period = year of birth + age. This gives us the freedom to re-write functions of cohort as functions of age and period, or vice versa. The additional identifiability issues generated by the cohort term depend fundamentally on the definition of the age functions within the model, and so are specific to the model in question. We find that APC models with non-parametric age functions do not have any extra identifiability issues beyond those discussed for AP models in Hunt and Blake (2015d), as shown in Appendix A. Models with certain types of parametric age functions require additional identification as discussed in Section 4.

In Hunt and Blake (2015f), we also found that difficulties with estimating and assigning demographic significance to the cohort parameters mean that, in practice, most models use only one cohort term (without any modulating age function) and do not involve any age/cohort interactions for reasons of both simplicity and robustness. We follow the same approach in this paper, and so do not consider models such as that proposed in Renshaw and Haberman (2006) or Model M8 in Cairns et al. (2009).

3 Identifiability in the classic APC model

The simplest APC model (referred to here as the "classic APC model") has a long history and is widely used in the fields of medicine, epidemiology and sociology as well as in demography and actuarial science.[7] It has the following form

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \tag{3}$$

It can be seen that the classic APC model has one age/period term with $f(x) = 1$, which is parametric in the sense defined in Hunt and Blake (2015f).

A model is fully identified when all the parameters in it can be uniquely determined by reference to the available data. In contrast, the classic APC model (as with most APC models) is not fully identified, because there exist

[7]For instance, see Hobcraft et al. (1982), Osmond (1985), O'Brien (2000), Carstensen (2007) and Kuang et al. (2008b).

different sets of parameters which will give the same fitted mortality rates and consequently the same goodness of fit for any data set. This phenomenon is not unique to APC mortality models. However, it is very widespread in such models and has significant implications when we come to make projections using them.

The issue of identifiability in the classic APC model also has a very long history.[8] It is, therefore, a good starting point to determine whether the issues raised in identifying the parameters in Equation 3 can be generalised to the more complex APC models used in mortality modelling. We can see that this model is not fully identified, since if we use the transformations in Equations 4, 5 and 6 to obtain new sets of parameters, we do not change the fitted mortality rates and hence the fit to the data

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - a, \kappa_t + a, \gamma_y\} \tag{4}$$
$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - b, \kappa_t, \gamma_y + b\} \tag{5}$$
$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x + c(x - \bar{x}), \kappa_t - c(t - \bar{t}), \gamma_y + c(y - \bar{y})\} \tag{6}$$

where a bar denotes the arithmetic mean of the variable over the relevant data range.[9] We call such transformations "invariant" for this reason. The existence of invariant transformations means that the model possesses identifiability issues, because no one set of parameters is determined uniquely from the data.

The transformation in Equation 6 is fundamentally unlike any of the transformations present in AP models discussed in Hunt and Blake (2015d), since it involves functions of age, period and year of birth rather than constants. It is a consequence of the collinearity between these dimensions, $y = t - x$, which enables us to decompose a linear function of year of birth into linear functions of age and period, and vice versa. This transformation generalises for many, more complex APC models with parametric age/period terms, as we discuss in Section 4.

[8] For instance, see Glenn (1976), Fienberg and Mason (1979), Rodgers (1982), Holford (1983), Clayton and Schifflers (1987), Wilmoth (1990), Yang et al. (2004), Kuang et al. (2008a) and O'Brien (2011).

[9] e.g., $\bar{x} = \frac{1}{X}\sum_x x = 0.5(X + 1)$.

We say that linear trends in the data are "unidentifiable" by the model, that is, they cannot be uniquely apportioned to either age, period or year of birth (as was discussed in Wilmoth (1990)). The linear trends observed in the parameters of the classic APC model therefore have no independent meaning, as different sets of parameters, with different linear trends will give exactly the same observable quantities such as fitted mortality rates.

The existence of unidentifiable linear trends in the classic APC model is of practical as well as theoretical importance. This is because we often see features of the (transformed) mortality rates which are approximately linear in age and time. For instance, the shape of the age function, α_x, is approximately linear at high ages,[10] whilst κ_t is often approximately linear.[11] The structure of the model means that we are fundamentally unable to separate these linear trends from a linear trend in the cohort parameters.

Because different sets of parameters give the same fit to the data, we cannot use the data to apportion the linear trend to either the age, period or cohort terms. One method of solving this issue is to move to a "maximally invariant" set of parameters, as discussed in Kuang et al. (2008a) and Nielsen and Nielsen (2014), which involves reparameterising the model in an equivalent form with reduced dimensionality, which avoids the identifiability issues. This approach is discussed in Appendix C.

An alternative and much more common approach is to impose additional identifiability constraints on the parameters in order to specify them uniquely.[12] These constraints manually apportion the linear trend between the different terms in the model. Imposing suitable constraints on the model involves the selection of a single set of parameters from the family of equivalent parameter sets, all of which give identical fitted mortality rates. In this sense, the manual apportionment is arbitrary - it does not depend upon

[10] If $\eta_{x,t} = \ln(\mu_{x,t})$, this is the Gompertz model, whilst if $\eta_{x,t} = \text{logit}(q_{x,t})$, this is the Perks model for mortality.

[11] See, for instance, Tuljapurkar et al. (2000), who went so far as to call this the *"universal pattern of mortality decline"*.

[12] We say that the transformations in Equations 4, 5 and 6 cause issues with the *identifiability* of the model.*Identification* of the model is accomplished by imposing a set of identifiability constraints and using the invariant transformations to achieve these constraints.

any observable property of the data, but is a product of the model user's subjective interpretation of the demographic significance of the parameters.

For example, one set of identifiability constraints is $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = 0$.[13] These identifiability constraints allow us to impose our interpretation of the demographic significance of the parameters onto the model. For example, the first two of the constraints above mean that α_x can be interpreted as an "average" level of mortality at age x, over the period, with κ_t and γ_y representing deviations from this average level. The third constraint requires that there are no deterministic linear trends within the fitted cohort parameters, since any linear trend in these parameters will be arbitrarily assigned to the age and period effects by using the transformation in Equation 6. This is in line with the demographic significance we assign to the cohort parameters in Hunt and Blake (2015f).

However, it is important to note that these additional identifiability constraints are arbitrary. For instance, the constraints $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y (y - \bar{y}) = 0$ (used later in Section 5.2) could also be imposed and would give different estimated parameters with exactly the same fit to data and have the same demographic significance. Further, the choice of having no linear trend in the cohort parameters does not have any independent meaning, since it is entirely dependent upon the identifiability constraints chosen. While these constraints might allow us to interpret the demographic significance of the parameters, this interpretation nevertheless depends entirely on the user's judgement rather than on the underlying data. For instance, a different choice of identifiability constraints could be used to impose that the period parameters, κ_t, had no linear trend, which would give the parameters a different demographic significance but leave the fitted mortality rates unchanged. We must, therefore, take care to ensure that our projections of observable quantities such as mortality rates do not depend on our arbitrary identification scheme, as discussed in Section 5.

[13] Here n_y is the number of observations of cohort y in the data and so $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x}$.

4 Identifiability in APC models with parametric age functions

Many of the more complex APC mortality models being proposed contain cohort parameters in the same form as in the classic APC model (i.e., without an age modulating $\beta_x^{(0)}$ function). Cairns et al. (2009) and Haberman and Renshaw (2011) found that models with a cohort term fit the data better than otherwise similar AP models, especially for the UK population, where a strong cohort effect has been observed by Willets (1999, 2004) and others. It is therefore natural to ask whether the additional issues with identifiability present in the classic APC model are also present in these more complex models.

In Appendix A, we show that APC models with non-parametric age functions do not possess any additional, non-trivial identification issues beyond those found in similar AP models discussed in Hunt and Blake (2015d). We have already seen, however, that in the simplest case of the classic APC model, the additional structure in the model caused by having a parametric age function combined with the collinearity of age, period and cohort can yield new identification issues.

For a general model with parametric age functions

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} + \gamma_{t-x} \tag{7}$$

we can try to generalise Equation 6 to look for invariant transformations of the form

$$\{\hat{\alpha}_x, \hat{f}^{(i)}(x), \hat{\kappa}_t^{(i)}, \hat{\gamma}_y\} = \{\alpha_x - a(x), f^{(i)}(x), \kappa_t^{(i)} - k^{(i)}(t), \gamma_y + g(y)\} \tag{8}$$

where $a(x)$, $k^{(i)}(t)$ and $g(y)$ are smooth functions.[14] Because the formulae used for the age functions define the model being used, in the sense of Hunt and Blake (2015f), we desire that they do not change under the invariant transformations, i.e., $\hat{f}^{(i)}(x) = f^{(i)}(x)$. Transformations which changed the

[14]While, α_x and κ_t are only defined for integer x and t, the parametric age functions $f^{(i)}(x)$ are defined for continuous x and so it make sense to look for transformations which also use continuous functions, as in the classic APC model in Section 3.

age functions in the model would give a fundamentally different model, albeit one which gave the same fit to the data. In Hunt and Blake (2015d), we called different models, with different definitions of the age functions, that gave identical fits to the data "equivalent models".

In order for the transformation in Equation 8 to leave Equation 7 unchanged, we require

$$g(t-x) = a(x) + \sum_{i=1}^{N} f^{(i)}(x)k^{(i)}(t) \tag{9}$$

If this is true, we say that the deterministic trends $k^{(i)}(t)$ and $g(y)$ are "unidentifiable", since the model is unable to apportion them between the age/period and cohort terms, in the same way as with the unidentifiable linear trends in the classic APC model. Instead, we must manually apportion these trends by means of additional identifiability constraints. These deterministic trends in the fitted parameters, therefore, lack any objective meaning, since they are entirely dependent on the choice of identifiability constraints. Nevertheless, they must be allowed for when projecting the APC mortality model, as discussed in Section 5, even if they appear to be comparatively small.

The first thing to note from Equation 8 is the trivial case where Equation 9 holds, i.e., $g(y) = a(x) = b$, a constant, and $k^{(i)}(t) = 0,\ \forall t$. This is simply a transformation of the form in Equation 5. It does not involve any age/period terms and so holds for all APC models, including those with non-parametric age functions.

To find less trivial transformations, we take a Taylor expansion of $g(y)$ around $-x$, assuming that it is an infinitely differentiable function of year of birth

$$g(t-x) = g(-x) + \sum_{j=1}^{\infty} \frac{1}{j!} t^j \left.\frac{d^j g}{dy^j}\right|_{y=-x} \tag{10}$$

Comparing this to Equation 9, we can set $a(x) = g(-x)$ and $k^{(j)}(t) = \frac{1}{j!}t^j$ if $f^{(j)}(x) = \left.\frac{d^j g}{dy^j}\right|_{y=-x}$, i.e., the derivatives of g are a subset of the age functions

of the model. Models of the form in Equation 7 have a finite number, N, of age/period terms and, therefore, we require that $g(y)$ has a finite series of derivatives. There are two cases when g will have a finite sequence of derivatives, either

1. the derivatives terminate after $M \leq N$ terms say, or
2. the form of the derivatives is cyclical so that $\left.\frac{d^{j+M}g}{dy^{j+M}}\right|_{y=-x} = K \left.\frac{d^j g}{dy^j}\right|_{y=-x}$ for some integer $M \leq N$ and constant K.

4.1 Polynomial age functions

For the Taylor series to terminate in a finite number of terms, we require that $\frac{d^j g}{gy^j} = 0$, $\forall j > M$, and therefore that $g(y)$ must be a polynomial in y of order M.

Theorem 1 *APC mortality models of the form in Equation 1 and age functions spanning the polynomials to order $M-1$ possess invariant transformations which add a polynomial of order M to the cohort function.*

Sketch of Proof Take $g(y)$, a general polynomial of order M, and expand as a function of x and t. This can then be regrouped into an equivalent form that corresponds to the age/period terms in the model, in order to see how

$g(y)$ can be absorbed into the age/period structure

$$
\begin{aligned}
g(y) &= \sum_{n=0}^{M} a_n y^n \\
\Rightarrow g(t-x) &= \sum_{n=0}^{M} a_n (t-x)^n \\
&= \sum_{n=0}^{M} a_n \sum_{m=0}^{n} \binom{n}{m} t^m (-x)^{n-m} \\
&= \sum_{n=0}^{M} a_n \left[(-x)^n + \sum_{m=1}^{n} \binom{n}{m} t^m (-x)^{n-m} \right] \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{n=1}^{M} \sum_{l=0}^{n-1} a_n \binom{n}{l} t^{n-l} (-x)^l \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{l=0}^{M-1} (-x)^l \sum_{n=l+1}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{l=0}^{M-1} (-1)^l f^{(l)}(x) \sum_{n=l+1}^{M} a_n \binom{n}{l} t^{n-l} \\
&= a(x) + \sum_{l=0}^{M-1} f^{(l)}(x) k^{(l)}(t)
\end{aligned}
$$

If there are age functions in the model of the form $f^{(j)}(x) = x^j$ of $j = 0, 1, \ldots M-1$, the expression above corresponds to Equation 9 with $a(x) = \sum_{n=0}^{M} a_n (-x)^n$ and $k^{(j)}(t) = (-1)^j \sum_{n=j+1}^{M} a_n \binom{n}{j} t^{n-j}$. More generally, we only require that the age functions span the first $M-1$ polynomials, because these are equivalent to a model with $f^{(j)}(x) = x^j$ such as that in the derivation above. ■

We can think of the transformation as expanding the polynomial $g(y)$ into terms in x and t, grouping these and then combining them with the appropriate age/period terms. A model with age functions spanning the first $M-1$ polynomials therefore has an additional $M+1$ degrees of freedom (represented by the coefficients, a_n, of the general polynomial) which do not affect the fit to the data. This is similar to the analysis in Wilmoth (1990),

which argues that higher order polynomial trends in the cohort parameters will cause identifiability problems in a mortality model if sufficient age/period terms of suitable form exist within the model. These additional degrees of freedom mean that we need to impose an additional $M+1$ identifiability constraints, which assign the $M+1$ unidentifiable polynomial trends between the different age/period and cohort terms in the model.

The simplest example of this is the transformation of the classic APC model described in Section 3. This has a single parametric age function $f(x) = 1$ which spans the polynomials to order 0. The model will then allow first order polynomials (i.e., linear terms) to be added to the cohort parameters with offsets made to the static life function and the period term without changing the fitted mortality rates. These are exactly the invariant transformations described in Equations 5 and 6. Consequently, we impose two additional identifiability constraints for the cohort parameters in the model to identify their level and linear trend.

4.1.1 The Plat models

In Plat (2009), two new APC mortality models were introduced. These can be written[15]

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_{t-x} \tag{11}$$

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{12}$$

The second of these models was introduced as a simplification of the first, with the expectation that it would be more suitable for modelling mortality at high ages. We call the model in Equation 11 the "Plat model" and the model in Equation 12 the "reduced Plat model" for this reason.[16]

The first point to note is that both the Plat and reduced Plat models nest the classic APC model, and therefore the invariant transformations in Equations 4, 5 and 6 are also applicable for both models.

[15] We define $x^+ \equiv max(x, 0)$.

[16] This model can also be thought of as an extension to model M6 in Cairns et al. (2009), with a static age function, or as an extension to the "CBDX" model discussed in Hunt and Blake (2015d) with a cohort term.

The second point to note is that these models also nest simple AP mortality models,[17] and therefore the results of Hunt and Blake (2015d) are still applicable. This means that the "locations" of the period functions are undefined and need to be identified by imposing a constraint on their levels. Usually this is of the form

$$\sum_t \kappa_t^{(i)} = 0$$

These invariant transformations were noted by Plat (2009) and used to impose suitable identifiability constraints.

However, the third point to note is that both of these models have age functions $f^{(1)}(x) = 1$ and $f^{(2)}(x) = (x - \bar{x})$ which span the polynomials to linear order. Using the result of Theorem 1, we should be able to find a transformation of the parameters which adds a quadratic polynomial in y to the cohort parameters, but leaves the fitted mortality rates unchanged. Indeed, we find that the transformation

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{&\alpha_x - d(x-\bar{x})^2, \\ &\kappa_t^{(1)} - d(t-\bar{t})^2, \kappa_t^{(2)} + 2d(t-\bar{t}), \gamma_y + d(y-\bar{y})^2\}\end{aligned} \tag{13}$$

leaves the fitted mortality rates unchanged for both the Plat and reduced Plat models. We say that these models have unidentifiable quadratic trends, which have to be manually allocated between the different parameters via identifiability constraints.

Hence, we require three identifiability constraints on the cohort parameters in the Plat and reduced Plat models, i.e., to apportion the level, linear trend and quadratic trend between the different age/period and cohort terms, plus identifiability constraints on the levels of the period functions. This means that for full identification of the models, we require an additional identifiability constraint to those discussed in Plat (2009).

If the model user fails to allocate the quadratic trend between the different terms via an additional identifiability constraint, then the fitting algorithm

[17] In particular, both models nest the "CBDX" model discussed in Hunt and Blake (2015d).

will make an apportionment in order to achieve convergence. However, this apportionment will not be based on any particular desired demographic significance and will depend on the specific details of fitting algorithm, such as the starting parameter values used. To illustrate, instead of removing quadratic trends from the cohort parameters and apportioning them to the age/period terms, the fitting algorithm may split any quadratic trends between the cohort parameters and the age/period terms, giving values of γ_y with an apparent quadratic trend in y. Not only is this contrary to our desired demographic significance, it can make comparing parameters across datasets difficult due to the presence or absence of quadratic trends which do not have any meaning independent of the data.

In addition, a failure to fully identify the model can lead to inefficient fitting algorithms, which take a long time to converge to a solution, as discussed in Hunt and Villegas (2015). Furthermore, they can also give parameter estimates which are not robust to small changes in the data (e.g., an additional year of data), since such changes can cause the fitting algorithm to abruptly change the allocation of the unidentifiable trends. For these reasons, it is very important to ensure that the APC mortality models we use are fully identified by imposing sufficient identifiability constraints to uniquely estimate all the parameters in the model.

Following the same approach as used for the classic APC model, we might choose to impose the constraints in Section 3 and extend these to impose $\sum_y n_y(y-\bar{y})^2\gamma_y = 0$ to remove quadratic trends in the cohort parameters and allocate them to the age/period terms. However, as with the classic APC model, this choice is arbitrary and a different choice of constraints will make no difference to the fitted mortality rates, only to the interpretation we give to the parameters.

In Section 3, we saw that the lack of identifiability of the linear trends in the model, due to the transformation in Equation 6, was of practical as well as theoretical importance because linear trends were often observed in both the age and period terms. Similarly, the transformation in Equation 13 is of practical importance when fitting the Plat model, because we usually see some curvature in α_x at high ages and also systematic departures from the

linearity of the period functions.[18] These quadratic trends will, therefore, not be distinguishable from a quadratic trend in the cohort parameters in the Plat model. However, because the observed magnitude of such trends is typically smaller than the linear trends observed in the age/period terms, failure to fully identify the quadratic trend in the data will typically have a lower, though still important, impact than a failure to identify the linear trend.

It is worth noting that the transformation in Equation 13 does not treat the different period functions equally, i.e., a term which is quadratic in t is added to $\kappa_t^{(1)}$, a term linear in t is added to $\kappa_t^{(2)}$, whilst $\kappa_t^{(3)}$ is unchanged by the invariant transformation for the Plat model. However, this is true only for the particular definition of the age functions shown. To illustrate, instead of the Plat model in Equation 11, we could instead have chosen an equivalent model of the form

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})^+ \kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_{t-x} \tag{14}$$

Such a model will trivially give the same fitted mortality rates as that in Equation 11 and has the same number of parameters, and so will have the same number of identifiability issues. However, the transformation corresponding to Equation 13 for this model will now add terms linear in t to both $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$. Specifically, for this model, we have the invariant transformation

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{&\alpha_x - d(x - \bar{x})^2, \kappa_t^{(1)} - d(t - \bar{t})^2, \\ &\kappa_t^{(2)} - 2d(t - \bar{t}), \kappa_t^{(3)} + 2d(t - \bar{t}), \gamma + d(y - \bar{y})^2\}\end{aligned} \tag{15}$$

in contrast to the transformation in Equation 13. Specifically, we note that whilst the transformation in Equation 13 did not involve $\kappa_t^{(3)}$, the transformation in Equation 15 does. The invariant transformations of the model are therefore specific to the age functions present, and may be different in different models, even if those models give an equivalent fit to data.

4.2 Exponential and trigonometric age functions

The other case where Equation 10 potentially yields invariant transformations of the parameters occurs when the derivatives of $g(y)$ are cyclical with

[18] For instance, see Booth et al. (2002), who curtailed the use of the data in the Lee and Carter (1992) model based on when a linear assumption for κ_t is no longer appropriate.

period $M \leq N$.

Theorem 2 *APC mortality models of the form in Equation 1 with exponential or trigonometric age functions possess invariant transformations which add similar exponential or trigonometric functions to the cohort parameters.*

Sketch of Proof In order for the derivatives of $g(y)$ to be cyclical with period M, we require

$$\frac{d^M g}{dy^M} = Kg \tag{16}$$

for some non-zero constant K. Substituting this into Equation 10 and comparing with Equation 9 gives

$$\begin{aligned} g(t-x) &= \sum_{j=0}^{M-1} \left.\frac{d^j g}{dy^j}\right|_{y=-x} \sum_{k=1}^{\infty} \frac{1}{(j+kM)!} t^{j+kM} \\ &= \sum_{j=0}^{M-1} f^{(j)}(x) k(t) \end{aligned}$$

This is of the form of Equation 9 if we set $k(t) = \sum_{k=1}^{\infty} \frac{1}{(j+kM)!} t^{j+kM}$ and have M age functions $f^{(j)}(x) = \left.\frac{d^j g}{dy^j}\right|_{y=-x}$ present in the model. It is interesting to note, therefore, that transformations of this form do not involve the static age function, as there is no term in the Taylor expansion of $g(t-x)$ corresponding to $a(x)$.[19]

Equation 16 has solutions of the form

$$g(y) = \sum_{i=1}^{M} \Re[a_i \ \exp(k_i y)]$$

where $\Re[z]$ is the real part of the expression z, and the k_i are the M roots of the equation $k_i^M = K$. In general, these roots will be complex, and,

[19]This means that they are also present in models without a static age function, as discussed in Appendix B.

therefore, $g(y)$ will be exponential, trigonometric or a combination of the two. In addition

$$\begin{aligned} f^{(j)}(x) &= \left.\frac{d^j g}{dy^j}\right|_{y=-x} \\ &= \sum_{i=1}^{M} \Re[a_i k_i^j \ \exp(-k_i x)] \end{aligned}$$

and so the age functions present in the model will also be exponential or trigonometric. ■

Exponential age/period terms can be included in models constructed using the "general procedure" of Hunt and Blake (2014), where they are typically used to explain infant mortality. As an example, consider a model of the form

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + e^{-\lambda x}\kappa_t^{(2)} + \gamma_{t-x} \tag{17}$$

This is an extension of the "exponential" model of Hunt and Blake (2015d), with an additional cohort term. We typically require $\lambda > 0$ to give the age function the demographic significance of governing rates of mortality at low ages. This model will allow the parameters to be transformed using

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)} - a\ e^{\lambda t}, \gamma_y + a\ e^{\lambda y}\} \tag{18}$$

This means that exponential trends in time within the (transformed) data are not uniquely identifiable as either age/period or cohort effects.[20] This transformation gives us an extra degree of freedom in the model which could be used to impose an additional identifiability constraint.

In this case, however, the imposition of an identifiability constraint will be of little practical importance. In Section 3, we said that in order to be practically important, the unidentifiable deterministic trends must be present in both the age and period dimensions of the transformed data. Whilst exponentially increasing trends in the age function are frequently observed in the data (due to low age mortality effects), exponential trends in the period

[20]Note that this transformation has $g(y) = a\exp(\lambda y)$ and therefore $\frac{dg}{dy} = \lambda g$ as per Equation 16.

functions are not.[21] We therefore do not experience problems when fitting the model to data as a result of any failure to be able to assign uniquely such a trend to the either age/period or the cohort terms.

As another example, consider a model with trigonometric age functions of the form

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \cos(\theta x)\kappa_t^{(2)} + \sin(\theta x)\kappa_t^{(3)} + \gamma_{t-x} \tag{19}$$

For this model, we can transform the parameters using

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{&\alpha_x, \kappa_t^{(1)}, \\ &\kappa_t^{(2)} - a\ \cos(\theta t) - b\ \sin(\theta t), \\ &\kappa_t^{(3)} + a\ \sin(\theta t) + b\ \cos(\theta t), \\ &\gamma_y + a\ \cos(\theta y) + b\ \sin(\theta y)\}\end{aligned} \tag{20}$$

This means that periodic patterns are not uniquely identifiable as either age/period or cohort effects.[22]

As with the exponential functions, the presence of unidentifiable trigonometric trends in the model will be of little practical importance. Whilst the (transformed) data often exhibits periodic behaviour in the cohort and period effects, it is rare to see periodic behaviour across ages.[23] Again, we do not have the unidentifiable deterministic trends for the model in both the age and period dimensions and consequently do not experience practical difficulties when fitting the model to data as a result of any failure to be able to assign uniquely such trends to the either age/period or the cohort terms.

[21] An exponential increase or decrease in the period function will typically correspond to super-exponential growth or decline in the observed mortality rates if either $\eta_{x,t} = \ln(\mu_{x,t})$ or $\eta_{x,t} = \text{logit}(q_{x,t})$. Super-exponential growth in mortality rates are not typically observed.

[22] Note that this transformation has $g(y) = a\ \cos(\theta y) + b\ \sin(\theta y)$ and therefore $\frac{d^2 g}{dy^2} = -\theta^2 g$ as per Equation 16.

[23] The lack of periodic structure across ages also explains why trigonometric age functions are not widely used in practice.

4.3 Other age functions

Other parametric age functions do not admit any additional invariant transformations involving the cohort parameters, except in the case where they are actually redefined polynomials, exponentials or trigonometric functions. For instance, the third age/period term in the Plat model did not generate any extra interactions with the cohort parameters, beyond those of the reduced Plat model. This simplifies the identifiability issues of more complex mortality models with different types of age functions, such as those produced by the "general procedure" of Hunt and Blake (2014), compared with what would otherwise be necessary, were, for instance, only polynomial age functions to be used.

4.4 Summary

In summary, issues with the identifiability of APC models relate to functions of year of birth which can be decomposed into purely age/period terms. However, this is only true in models where the age functions take specific parametric forms - namely polynomial, exponential and trigonometric functions. In such models, certain deterministic trends cannot be uniquely allocated between the age/period and cohort terms in the model and so require the imposition of arbitrary identifiability constraints in order to uniquely specify the model.[24] This is summarised in the flow chart in Figure 1.

5 Projection

In the preceding sections, we have seen that APC mortality models are not fully identified and that we can impose arbitrary identifiability constraints on the parameters in order to fit them to the historical data. Two different modellers using the same data and the same model but different arbitrary identification constraints will obtain different sets of parameters, but these will give identical fitted mortality surfaces and, therefore, fits to the data.

[24] As discussed in Appendix B, APC mortality models with non-parametric age functions will not have any additional transformations that leave the fitted mortality rates exactly unchanged. However, such models may have transformations that leave the fitted mortality rates approximately unchanged, as discussed in Hunt and Villegas (2015).

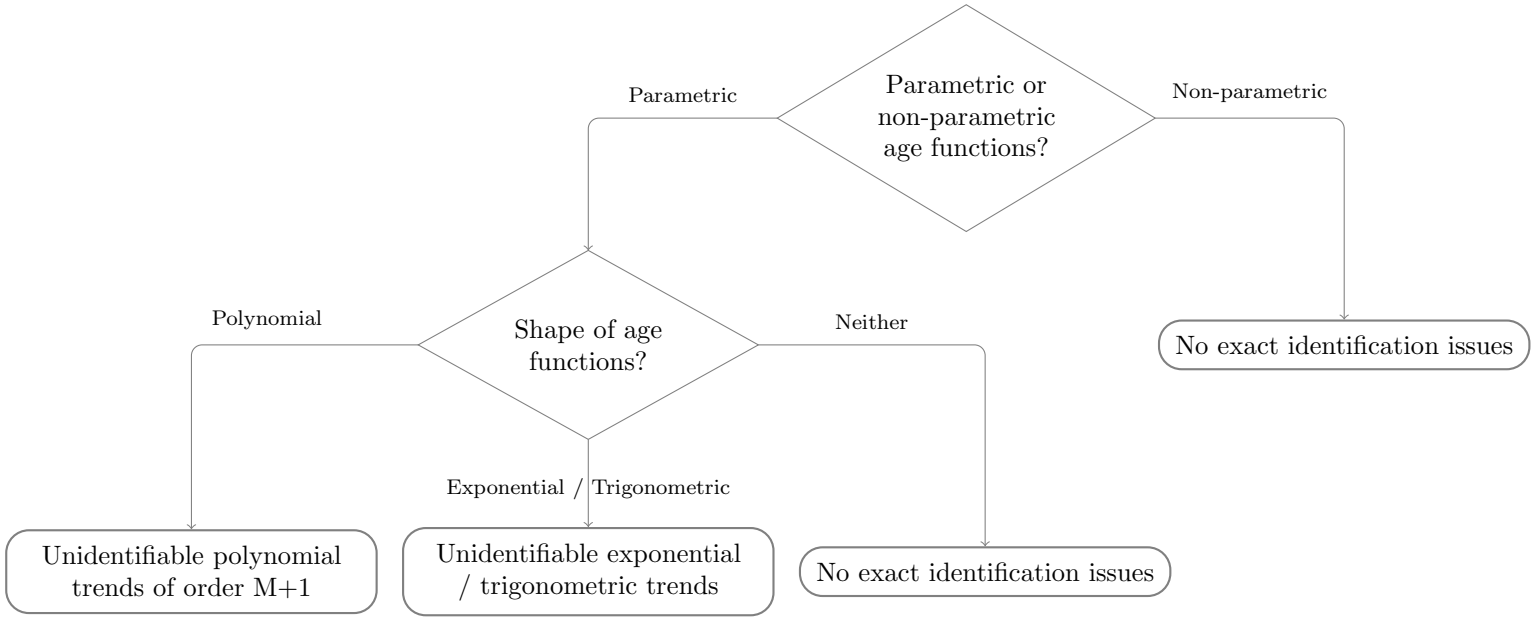


Figure 1: Flow chart of identifiability issues in APC models

For the majority of practical purposes, we not only need to fit a mortality model to historical data but also to use it to project mortality rates into the future. In Hunt and Blake (2015d), we found that we needed to be careful when doing so in AP mortality models in order to ensure that the projected mortality rates will not depend on the arbitrary identifiability constraints imposed when fitting the models to data. The same is true in APC mortality models. However, the addition of a set of cohort parameters and the presence of unidentifiable deterministic trends complicate this analysis significantly.

The most obvious change when moving from an AP to an otherwise similar APC mortality model is the presence of a set of cohort parameters which will also need to be projected into the future. The period and cohort parameters in the APC model are conceptually different and need to be treated separately when making projections. This is because cohort effects have very different demographic significance from the period effects and are treated separately when fitting the model. It is therefore common practice to project the period and cohort parameters independently.

Some authors (e.g., Haberman and Renshaw (2011)) disagree with this approach, arguing that it may only be appropriate to do this when the cohort parameters are estimated using the residuals from the fitted primary age/period structure. This means that the cohort structure fitted by the model is independent of the age/period structure by construction. However, such fitting techniques will not give parameter estimates which maximise the fit to data and can lead to hierarchical issues (because the cohort parameters are only estimated conditional on the previously fitted estimates of the age/period structure). We, therefore, have a clear preference for model fitting techniques where all parameters are estimated together in order to generate the best fit to the historical data.[25]

More generally, it is conceivable that events such as influenza pandemics will cause both an immediate rise in mortality and also lifelong health effects in infants born during the pandemic due to selection effects, leading

[25]For example, in the general procedure of Hunt and Blake (2014), all parameters are re-estimated every time the structure of the model is changed, in order to ensure a close fit to the data.

to correlations between extreme period and cohort effects. However, it is difficult to analyse any dependence structure between the cohort and period parameters as the cohort parameters will be observed over a longer time period, but potentially at a lag of some decades. While it is possible that some extreme mortality events may generate distinctive effects in both the period and cohort parameters, the evidence supporting this conjecture is currently ambiguous (for instance, see Murphy (2009)) and will not generally be relevant for more typical period and cohort effects. An assumption of independence is, therefore, both practical and parsimonious.

In order to make projections of future mortality rates, we typically model the period and cohort parameters as being generated by independent time series processes and use these to project the parameters stochastically into the future. However, the precise form of the time series processes generating the parameters is unknown. Therefore, we analyse the fitted parameters by statistical methods, such as the Box-Jenkins procedure, to determine which processes from the ARIMA family provide the best fit.

Nevertheless, when it comes to projecting mortality rates, we need to recognise that there is a fundamental symmetry between the processes of estimating a model and projecting it: the former takes observations to calibrate the model, whilst the latter uses this calibration to produce projected observations of the future. Due to this symmetry, identification issues which exist when fitting the model may also yield problems when projecting it. When estimating the model, these identifiability issues were solved by imposing arbitrary identifiability constraints on the parameters. However, any time series structure that we find in the parameters needs to be independent of the arbitrary identification scheme used when fitting the model to historical data.

We formalise this by saying that:

> Two sets of model parameters, which give identical fitted mortality rates for the past, should give identical projected mortality rates when projected into the future.

We say that time series processes which satisfy this property are "well-identified".

In particular, the invariant transformations of the parameters of the model which leave the fitted mortality rates unchanged should also leave the projected mortality rates unchanged and, hence, the time series processes used to generate the projected mortality rates unchanged. Consequently, we should use the same time series processes for all sets of parameters from a model which give the same fitted mortality rates. If this is not the case, different processes will be used for different arbitrary identifiability constraints, giving different projected mortality rates. A well-identified time series process should be equally appropriate for all equivalent sets of parameters. To confirm this, we need to check that applying the invariant transformations to the parameters, which leave the fitted mortality rates unchanged, do not also affect the time series processes used to project the parameters.

Hunt and Blake (2015d) discussed how the identification issues in the class of AP models meant that methods for projecting the period parameters from these models into the future needed to be chosen with care in order to ensure they are well-identified. In general, we argued that we should choose to project the model using multivariate methods which are as unstructured as possible, i.e., we should not impose features such as independence, levels of mean reversion or different orders of integration on the time series a priori, but allow these to emerge during the fitting process. However, we also saw that, in models with parametric age functions, the age/period terms were no longer interchangeable once we defined their forms in the model. This allowed us to prioritise biological reasonableness[26] over using the same processes for equivalent models, i.e., models giving the same fitted mortality rates with different definitions of the age functions.

Current practice is to:

1. fit the chosen model to data, imposing any arbitrary identifiability constraints needed in order to specify the parameters uniquely;

2. select time series processes for projecting the parameters based on either using a statistical method (such as the Box-Jenkins procedure to select the preferred processes from the ARIMA class of models) or by

[26]Introduced in Cairns et al. (2006) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

directly choosing the time series processes to ensure biologically reasonable projections by making an appeal to the demographic significance of the parameters.

However, such an approach often leads to projections of mortality rates which are not well-identified. This is because the second step assumes that the parameters found at the first step are known, rather than merely estimated up to an arbitrary identifiability constraint. This means that current practice builds the arbitrary identifiability constraint into the projection process, ensuring that the projected mortality rates are also arbitrary.

To avoid this, we propose to work backwards from our desire for projections which are biologically reasonable and well-identified to determine the time series processes we need to use to achieve these aims. Before fitting the model, we need to conduct a thorough analysis of the identifiability issues in the chosen model, using the principles established in Section 4, to determine which features of the parameters are set by the data and which are set by the arbitrary identifiability constraints. Then, suitable time series processes should be selected to model only the former, identifiable features of the parameters, while still allowing for the unidentifiable trends in a way that guarantees that they do not affect the projection of future mortality rates. By following this procedure, we can ensure that the time series processes are well-identified and that the projected mortality rates do not depend on the arbitrary choices we make when fitting the model.

In this section, we will first look at the broad set of criteria needed for well-identified projection methods in general APC mortality models in Section 5.1. Section 5.2 looks in more detail at why current practice can lead to projections which are not well-identified and depend on the arbitrary identifiability constraints chosen in the context of the classic APC model from Section 3. We then revisit the general case of an APC mortality model in Section 5.3, in order to determine general rules for choosing time series processes which are well-identified. These are then applied in the context of the classic APC model again in Section 5.4 and it is demonstrated that projected mortality rates are genuinely independent of the choice of arbitrary identifiability constraint. Section 5.5 then applies the general rules in the context of the Plat model from Plat (2009) and Section 4.1.1 to see how they work in the context of more sophisticated mortality models with more complex

identifiability issues.

5.1 Projecting general APC models

Consider the case of projecting an APC mortality model, which has been fitted using data over the period $[1, T]$ to give mortality rates at time $\tau > T$. From Equation 2, we could write this as

$$\eta_{x,\tau} = \alpha_x + \boldsymbol{\beta}_x^\top \boldsymbol{\kappa}_\tau + \gamma_{\tau-x}$$

If the model has identifiability issues, then the projected mortality rates should be unchanged under exactly the same invariant transformations as the fitted mortality rates were, i.e., if we have an invariant transformation of the form of Equation 8, namely

$$\begin{aligned}
\hat{\alpha}_x &= \alpha_x - a(x) \\
\hat{\boldsymbol{\beta}}_x &= \boldsymbol{\beta}_x \\
\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\kappa}_t - \boldsymbol{k}(t) \\
\hat{\gamma}_y &= \gamma_y + g(y)
\end{aligned}$$

where $a(x)$, $k^{(i)}(t)$ and $g(y)$ satisfy Equation 9, in which case

$$\eta_{x,\tau} = \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top \hat{\boldsymbol{\kappa}}_\tau + \hat{\gamma}_{\tau-x}$$

The projected $\boldsymbol{\kappa}_\tau$ (and potentially the $\gamma_{\tau-x}$) will be random variables, whose distribution is a function of the historical, fitted values, i.e., $\boldsymbol{\kappa}_\tau = P_\kappa(\tau; \{\boldsymbol{\kappa}\})$ and $\gamma_y = P_\gamma(y; \{\gamma\})$. We said previously that we should use the same method of projection for all sets of parameters as a first step to ensure that the projected mortality rates do not depend upon the identifiability constraints. However, for different identifiability constraints, these processes will be estimated from different sets of fitted parameters, e.g., if we use $P_\kappa(\tau; \{\boldsymbol{\kappa}\})$ to project the untransformed period parameters, we must use $P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\})$ to project the transformed period parameters. If we combine this with the invariance of the projected mortality rates, we have

$$\begin{aligned}
\alpha_x + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa}\}) + P_\gamma(\tau - x; \{\gamma\}) &= \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\}) + P_\gamma(\tau - x; \{\hat{\gamma}\}) \\
&= \alpha_x - a(x) + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) + P_\gamma(\tau - x; \{\gamma + g\}) \\
P_\gamma(\tau - x; \{\gamma + g\}) - P_\gamma(\tau - x; \{\gamma\}) &= a(x) + \boldsymbol{\beta}_x^\top \left(P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\})\right)
\end{aligned}$$

Using Equation 9, we can eliminate $a(x)$

$$P_\gamma(\tau - x; \{\gamma + g\}) - P_\gamma(\tau - x; \{\gamma\}) = g(\tau - x) + \boldsymbol{\beta}_x^\top \left(P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) - \boldsymbol{k}(\tau)\right)$$

In order for this to hold for all τ and x requires

$$P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) = P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - \boldsymbol{k}(\tau) \tag{21}$$
$$P_\gamma(y; \{\gamma + g\}) = P_\gamma(y; \{\gamma\}) + g(y) \tag{22}$$

This means that we should obtain the same results if we project the transformed parameters as if we transform the projected parameters, i.e., the processes of projection and transformation are commutative. Consequently, we see that, in order for a projection method to be well-identified under the invariant transformation, it needs to preserve the unidentifiable trends in the model, i.e., P_κ must preserve the trends $\boldsymbol{k}(t)$, and P_γ must preserve the trend $g(y)$. This also means that it does not matter in which order we perform the processes of projection and transformation, the distribution of the transformed parameters projected into the future will be identical to the distribution of the projected parameters which are then transformed.

In addition, since

$$\mathbb{V}ar(\boldsymbol{\kappa}_\tau) = \mathbb{V}ar(\boldsymbol{\kappa}_\tau - \boldsymbol{k}(\tau)) = \mathbb{V}ar(\hat{\boldsymbol{\kappa}}_t)$$
$$\mathbb{V}ar(\gamma_y) = \mathbb{V}ar(\gamma_y + g(y)) = \mathbb{V}ar(\hat{\gamma}_y)$$

we note that the variability of the parameters around the trend is identifiable and so does have a meaning independent of the identifiability constraints imposed. Therefore, we conclude that, while the deterministic trends may be unidentifiable and not meaningful, the variation around the trend is of genuine significance, since it is independent of the identifiability constraints. Therefore, this variation needs to be projected consistent with our demographic significance for the parameters and what has been observed in the historical data.

However, the time series processes selected via current practice often do not preserve the unidentifiable trends in the period and cohort parameters, as we shall now see using the classic APC model.

5.2 Projecting the classic APC model

It has long been known, at least since Osmond (1985), that the lack of identifiability in the classic APC model has important consequences when making projections from the model. Different sets of arbitrary identifiability constraints are based on different allocations of the linear trends in the data between the age, period and cohort parameters. The outcome of current practice can therefore be influenced by the presence or absence of a linear trend in the fitted parameters, despite this being purely dependent upon the identifiability constraints chosen.

To illustrate this, we consider projecting the classic APC model fitted using four different sets of identifiability constraints. The fitted mortality rates given using these four sets of constraints are identical; however, the time series processes found by current practice differ which means that current practice would give different projected mortality rates in the four different cases. Consequently, these time series processes are not well-identified.

We start by fitting the classic APC model to mortality data for the USA from Human Mortality Database (2014) for ages 50 to 100 and year 1950 to 2010. As discussed in Section 3, a number of equally valid identifiability constraints can be imposed on this model, which give identical fitted mortality rates. We consider the following four sets of identifiability constraints:

Case 1: $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x} = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = \sum_{x,t} \gamma_{t-x}((t - \bar{t}) - (x - \bar{x})) = 0$. This was discussed in Section 3 and restricts the cohort parameters to be zero on average and without any linear trends, consistent with our desired demographic significance for the cohort parameters.

Case 2: $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y (y - \bar{y}) = 0$. These constraints impose the same demographic interpretation on the parameters, except that the averages are not weighted by the number of observations of each cohort.

Case 3: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(x - \bar{x}) = 0$. This set of constraints is the same as imposed on the classic APC model in Cairns et al. (2009), where it was written as imposing $\sum_x (\alpha_x - \frac{1}{T}\sum_t \eta_{x,t})(x - \bar{x}) = 0$, i.e., that the static age function, α_x, explains all the linearity across ages in the data.

Case 4: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(t - \bar{t}) = 0$. Similar to Case 3, this set of constraints imposes that the period function, κ_t, accounts for all of the linearity across years in the data.

The first thing to note is that all of these constraints were developed to give the cohort parameters the same demographic significance, i.e., that they should be centred on zero and the other functions in the model should capture any linear trends. Because of this, the fitted parameters in each case are very similar. However, they are not identical, unlike the fitted mortality rates. We therefore see that demographic significance, whilst helpful in selecting an appropriate set of identifiability constraints, does not specify a unique set of constraints to use. Model users with the same interpretation of the parameters can reasonably choose to impose different constraints and obtain different fitted parameters when using the same model with the same data. The fact that demographic significance is subjective and, in practice, different model users adopt a range of interpretations for the different parameters highlights the fact that we must take care to ensure that any conclusions regarding projected mortality rates are independent of the arbitrary choice of constraints made when fitting the model, and underscores the extent to which the identifiability constraints we choose is arbitrary.

Current practice is to take the fitted parameters and then determine which time series processes to use to project them. This may involve performing a Box-Jenkins analysis on the fitted parameters, as was done in Lee and Carter (1992) and Cairns et al. (2011). Alternatively, current practice may appeal to the demographic significance assigned to the parameters, as in Plat (2009). Such an appeal might determine that the period function is non-stationary (as it is primarily responsible for the evolution of mortality) and, based on the discussion in Hunt and Blake (2015f), that the cohort parameters are stationary around zero. It might therefore appear reasonable to choose[27] to use a random walk with drift process for κ_t and an AR(1) process for γ_y

$$\kappa_t = \kappa_{t-1} + \mu + \epsilon_t \tag{23}$$

$$\gamma_y = \rho\gamma_{y-1} + \varepsilon_y \tag{24}$$

[27]Note that we are not saying that these are the most appropriate time series processes to use for this set of parameters. We use them for illustrative purposes as they are relatively simple and not atypical of the processes used in practice. However, it is important to observe that selecting alternative time series processes on a purely statistical basis from the fitted parameters would not solve the issues we have identified.

Table 1 shows the fitted parameters for the four cases above using these time series processes.

	Case 1	Case 2	Case 3	Case 4
κ_{2010}	-0.3526	-0.3439	-0.3550	-0.3478
μ	-0.0110	-0.0107	-0.0111	-0.0109
$\sigma_\kappa = \mathbb{S}t\mathbb{D}ev(\epsilon_t)$	0.0161	0.0161	0.0161	0.0161
γ_{1950}	-0.1459	-0.1125	-0.1422	-0.1530
ρ	0.9513	0.9577	0.9499	0.9542
$\sigma_\gamma = \mathbb{S}t\mathbb{D}ev(\varepsilon_y)$	0.0193	0.0184	0.0193	0.0194

Table 1: Time series parameters for the period and cohort functions in the classic APC model fitted using different identifiability constraints

For $\tau - x > 1950$,[28] we find

$$\mathbb{E}\eta_{x,\tau} = \alpha_x + \kappa_{2010} + (\tau - 2010)\mu + \rho^{\tau-x-1950}\gamma_{1950} \tag{25}$$

We can therefore see that, inserting the fitted time series parameters from Table 1 for the four different cases, we do not find the same expected values for the future mortality rates.[29] This is shown in Figure 2. In addition, the variability of the projected parameters depends on σ_κ, ρ and σ_γ. However, ρ and σ_γ differ between cases, meaning that the variability of projected mortality rates will also be different for the different cases. These differences in the distribution of projected mortality rates might be felt to be relatively small, although they will grow with projection time. However, the most important point is that the differences should not exist at all - the fitted mortality rates for the different cases were *identical* and so should be the distribution of the projected mortality rates. We therefore see that the time series processes used above to project the classic APC model are not well-identified.

[28] That is, for cohort parameters that are projected rather than fitted from historical data, taking into consideration that cohort parameters for the ten most recent years of birth are not fitted from the data due to insufficient observations.

[29] For example, $\mathbb{E}\eta_{60,2020} = -4.5449$ for the Case 1 parameters, -4.5598 for the Case 2 parameters, -4.5459 for the Case 3 parameters and -4.5433 for the Case 4 parameters.

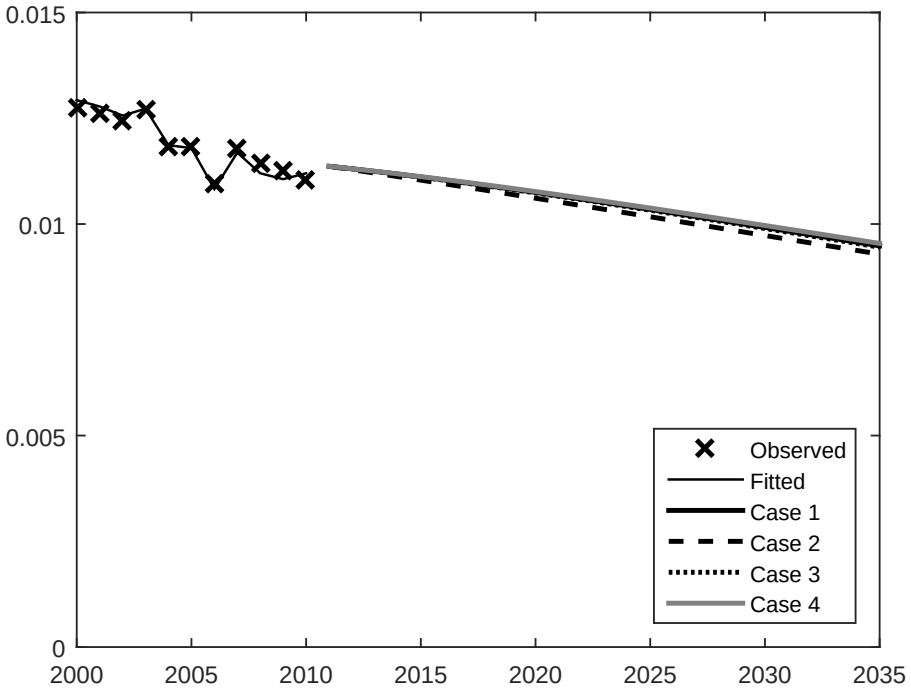


Figure 2: Projected $\mu_{60,t}$ using different sets of identifiability constraints

5.3 Projecting general APC mortality models: Revisited

From Section 5.1 above, we note that we must use the same time series processes to project sets of parameters which give identical fitted mortality rates, i.e., if $P_\gamma(y;\{\gamma\})$ is a suitable process (with time series parameters estimated from the fitted cohort parameters, $\{\gamma_y\}$), then $P_\gamma(y;\{\hat{\gamma}\})$ is a suitable process, albeit with time series parameters estimated from the transformed cohort parameters, $\{\hat{\gamma}_y = \gamma_y + g(y)\}$.

In practice, we usually describe our projection methods in terms of time series processes rather than projection functions. However, the two are equivalent, since the projection function is found by "solving" the difference equation form of the time series. For instance, the AR(1) process has the difference equation form in Equation 24, but has solution

$$P_\gamma(y;\{\gamma\}) = \rho^{y-Y}\gamma_Y + \sum_{s=Y+1}^{y} \rho^{y-s}\varepsilon_s$$

where Y is the last year of birth for which we fitted the cohort parameters.

The general form of ARIMA difference equations for γ_y can be written as[30]

$$(1-L)^d\Phi(L)(\gamma_y - \Gamma(y)) = \Psi(L)\varepsilon_y \tag{26}$$

where L is the lag operator, d is the order of integration of the process, Φ and Ψ are polynomials of order p and q governing the autoregressive and moving average parts of the process, respectively,[31] ε_y are the innovations and $\Gamma(y)$ is a deterministic function of year of birth. Taking unconditional expectations (i.e., with no conditioning on previous lags of the process), we see that

$$\mathbb{E}\left[\gamma_y - \Gamma(y)\right] = 0 \quad \forall y$$

and that the function $\Gamma(y)$ represents the trend around which the cohort parameters vary.

[30] For simplicity, we use the cohort function as an illustrative case. The analysis is identical for $\boldsymbol{\kappa}_t$, however.

[31] In order to be stationary, these polynomials have roots with modulus less than unity.

The invariant transformation of the model in Equation 9 adds a deterministic function - the unidentifiable trend $g(y)$ - to the cohort parameters. However, this deterministic function must not change the error term, ε_y, of a well-identified process and so

$$\begin{aligned}\varepsilon_y &= (1-L)^d \Psi^{-1}(L)\Phi(L)(\gamma_y - \Gamma(y)) \\ &= (1-L)^d \Psi^{-1}(L)\Phi(L)(\hat{\gamma}_y - \hat{\Gamma}(y)) \\ &= (1-L)^d \Psi^{-1}(L)\Phi(L)(\gamma_y + g(y) - \hat{\Gamma}(y))\end{aligned}$$

In order to ensure that the variation around the trend, given by the error term, remains unchanged by the invariant transformation, we require

$$\hat{\Gamma}(y) = \Gamma(y) + g(y)$$

In this case, the deterministic trend, $\Gamma(y)$, has changed under the invariant transformation but not the variation around the trend.

We stated above that the time series processes being used for the parameters should be equally applicable for all sets of parameters which give the same fitted mortality rates. This implies that the form of the deterministic trends should be the same, and, therefore, that $\hat{\Gamma}(y)$ is of the same form as $\Gamma(y)$. This can only be true if $\hat{\Gamma}(y)$, $\Gamma(y)$ and $g(y)$ are all of the same form. For instance, if $g(y)$ is a linear function of year of birth (as in the case of the classic APC model), then $\Gamma(y)$ and $\hat{\Gamma}(y)$ must also be linear functions of year of birth and so will not change form under the invariant transformations of the model.

If we solve Equation 26, we see that

$$\gamma_y = P_\gamma(y; \{\gamma\}) = \frac{\Psi(L)}{(1-L)^d \Phi(L)}\varepsilon_y + \Gamma(y) \tag{27}$$

In this form, it can also be seen that such time series processes preserve unidentified trends in the manner discussed in Section 5.1

$$\begin{aligned}\hat{\gamma}_y &= \gamma_y + g(y) \\ &= \frac{\Psi(L)}{(1-L)^d \Phi(L)}\varepsilon_y + \Gamma(y) + g(y) \\ &= \frac{\Psi(L)}{(1-L)^d \Phi(L)}\varepsilon_y + \hat{\Gamma}(y)\end{aligned}$$

i.e., the projected parameters after applying the invariant transformation will have the same variation, $\frac{\Psi(L)}{(1-L)^d\Phi(L)}\varepsilon_y$, but around a different deterministic trend, $\hat{\Gamma}(y)$, compared with the original parameters projected using the same method. The use of the invariant transformations will not affect our measurement of any coefficients in $\Psi(L)$ or $\Psi(L)$ at the fitting stage. Thus, we also see that the two ways of looking at the projected parameters, namely as time series processes and via projection functions, are equivalent.

As an example, consider the cohort parameters in the classic APC model. From Section 3, we see that, in this model, the cohort parameters have an unidentified constant and linear trend, i.e., $g(y) = b+c(y-\bar{y})$ from Equations 5 and 6. In Section 5.2, we said that current practice might use an AR(1) process for the cohort parameters, which has ARIMA form

$$(1 - \rho L)\gamma_y = \varepsilon_y$$

Comparing this with Equation 26, we see that current practice assumes that $\Gamma(y) = 0$, which is not of the same form as $g(y)$ above. Therefore, the time series process changes form when using an alternative set of parameters $\hat{\gamma}_y = \gamma_y + g(y)$ in place of γ_y,

$$\begin{aligned}(1 - \rho L)\hat{\gamma}_y &= (1 - \rho L)(\gamma_y + b + c(y - \bar{y})) \\ &= (1 - \rho L)\gamma_y + (1 - \rho)(b + c(y - \bar{y})) + \rho c \\ &= \varepsilon_y + (1 - \rho)(b + c(y - \bar{y})) + \rho c \\ &\neq \varepsilon_y\end{aligned}$$

and therefore the process is not well-identified.

When analysed in this form, however, a solution becomes immediately apparent: we need to introduce a linear function, $\Gamma(y) = \beta_0 + \beta_1 y$, into the AR(1) process to ensure that the process is well-identified, i.e.,

$$(1 - \rho L)(\gamma_y - \beta_0 - \beta_1 y) = \varepsilon_y \tag{28}$$

Using the alternative parameters $\hat{\gamma}_y$ would produce

$$\begin{aligned}(1 - \rho L)(\hat{\gamma}_y - \hat{\beta}_0 - \hat{\beta}_1 y) &= (1 - \rho L)(\gamma_y + b + c(y - \bar{y}) - \hat{\beta}_0 - \hat{\beta}_1 y) \\ &= (1 - \rho L)(\gamma_y - \beta_0 - \beta_1 y) \\ &= \varepsilon_y\end{aligned}$$

if $\hat{\beta}_0 = \beta_0 - b - c\bar{y}$ and $\hat{\beta}_1 = \beta_1 - c$. Therefore, the form of Equation 28 does not change under the invariant transformations of the classic APC model, and we conclude that this time series process is well-identified. Again, we also see that the variation around the linear trend, given by ε_y, is unchanged by the invariant transformation, whilst the unidentifiable trend is affected by the invariant transformation.

The time series process in Equation 28 has been suggested previously for the cohort parameters in Cairns et al. (2009) where it was referred to as the "*AR(1) process around a linear drift*". However, in Cairns et al. (2009), it was not used for the classic APC model, nor was it selected for being well-identified, but rather on the grounds of fitting the observed cohort parameters well.

The AR(1) around linear drift process is solved to give

$$P_\gamma(y;\{\gamma\}) = \rho^{y-Y}(\gamma_Y - \beta_0 - \beta_1 Y) + \beta_0 + \beta_1 y + \sum_{s=Y+1}^{y} \rho^{y-s}\varepsilon_s$$

We can also verify, by substituting the forms for $\hat{\gamma}_y$, $\hat{\beta}_0$ and $\hat{\beta}_1$ found above, that this process also satisfies the requirement of Equation 22 in Section 5.1, namely

$$P_\gamma(y;\{\hat{\gamma}\}) = P_\gamma(y;\{\gamma\}) + a + b(y - \bar{y})$$

Hence, projecting the transformed cohort parameters gives us the same results as transforming the projected cohort parameters.

Returning to the form of the time series process in Equation 26, it is common to write this in an alternative, but equivalent form

$$\begin{aligned}(1-L)^d\Phi(L)\gamma_y - (1-L)^d\Phi(L)\Gamma(y) &= \Psi(L)\varepsilon_y \\ (1-L)^d\Phi(L)\gamma_y &= \xi(y) + \Psi(L)\varepsilon_y \qquad (29)\end{aligned}$$

where $\xi(y)$ is a deterministic function of y and $\Gamma(y)$ solves the difference equation

$$(1-L)^d\Phi(L)\Gamma(y) = \xi(y) \qquad (30)$$

In this form, $\xi(y)$ is often referred to as the "drift". Knowing the form that $\Gamma(y)$ must take (i.e., the same form as $g(y)$ from the unidentifiable trends in the model in Equation 8), we can therefore specify the correct form of $\xi(y)$.

As an example of this, consider the classic APC model again, but, this time, consider the period parameters. We know from Section 3 that the period parameters have an unidentified linear trend in much the same way as the cohort parameters, i.e., $k(t) = a - c(t - \bar{t})$ if we re-write Equations 4 and 6 using the notation of Equation 9. Random walk processes are often used for the period parameters, i.e., we assume $d = 1$ and $\Phi(L) = \Psi(L) = 1$. It is then important to specify the correct form for the drift $\xi(t)$. Based on similar arguments to the ones used above for the cohort parameters, we should look for time series processes of the form

$$(1 - L)(\kappa_t - \nu_0 - \nu_1 t) = \epsilon_t$$

which has a linear trend $K(t) = \nu_0 + \nu_1 t$. To obtain a well-identified time series of the form of Equation 29, we need the drift, $\xi(t)$, of the random walk to satisfy

$$\begin{aligned}\xi(t) &= (1 - L)(\nu_0 + \nu_1 t) \\ &= \nu_0 + \nu_1 t - \nu_0 - \nu_1(t - 1) \\ &= \nu_1\end{aligned}$$

i.e., the drift is constant. This shows that the random walk with drift is well-identified for the period parameters in the classic APC model.

We can also verify this directly, since

$$\begin{aligned}\epsilon_t &= \kappa_t - \kappa_{t-1} - \mu \\ &= \hat{\kappa}_t - a + c(t - \bar{t}) - \hat{\kappa}_{t-1} + a - c(t - 1 - \bar{t}) - \mu \\ &= \hat{\kappa}_t - \hat{\kappa}_{t-1} - \hat{\mu}\end{aligned}$$

if $\hat{\mu} = \mu - c$. Thus the transformed period parameters, $\hat{\kappa}_t$, follow a random walk with drift if the original period parameters do. However, the value of the drift, which determines the unidentifiable linear trend, will change under the invariant transformation, although the innovations, ϵ_t, which determine the variability around this drift do not.

In summary, we have the following procedure for selecting a well-identified time series process for any specific APC mortality model:

1. Determine the identifiability issues in the specific APC model by finding the unidentifiable deterministic trends for the parameters which cannot be assigned between the different age/period and cohort terms in the specific model. This will need to be done prior to the fitting stage in order to fit the model robustly to data.

2. Specify a time series process for the variation around these trends. This can either be done by analysing this variation using statistical techniques, or by selecting a process which accords with our demographic significance for the parameters. Doing so will set the form of $\Phi(L)$ and $\Psi(L)$, which determine the stochastic structure of the ARIMA process.

3. Specify the deterministic trends, $\Gamma(y)$, in the time series process in Equation 26, which will need to be of the same form as $g(y)$. Equivalently, this can be achieved by finding a drift function, $\xi(y)$, in the alternative form of the time series process in Equation 29, with the requirement that $(1-L)^d\Phi(L)\Gamma(y) = \xi(y)$.

It is important to recognise that this procedure works backwards from the variation around the trends in the parameters, which is independent of the identifiability constraints and then adds back in the unidentifiable trends which will depend upon the specific set of identifiability constraints we use when fitting the model. In this fashion, we can ensure that the projected parameters are both well-identified and possess our desired demographic significance when specifying a suitable form for the time series process.

5.4 Projecting the classic APC model: Revisited

In Section 5.2, it was demonstrated that the current practice approach to selecting time series processes for the period and cohort parameters in the classic APC model yielded projections of mortality rates which depended upon arbitrary choices made when fitting the model. In Section 5.3, we then showed that the issue in this case was not the use of the random walk with drift for the period parameters, but the selection of an AR(1) process, rather than an AR(1) process around a linear drift for the cohort parameters.

If we use the AR(1) around linear drift process for the cohort parameters for the four cases discussed in Section 5.2, we obtain the time series parameters in Table 2.

	Case 1	Case 2	Case 3	Case 4
γ_{1950}	-0.1459	-0.1125	-0.1422	-0.1530
β_0	0.1388	0.1852	0.1388	0.1388
β_1	-0.0053	-0.0056	-0.0052	-0.0055
ρ	0.9636	0.9636	0.9636	0.9636
$\sigma_\gamma = \mathbb{S}t\mathbb{D}ev(\varepsilon_y)$	0.0184	0.0184	0.0184	0.0184

Table 2: Time series parameters for different identifiability constraints

As previously mentioned in Section 5.2, ρ and σ_γ control the variation of projected cohort parameters. It is, consequently, important to see that these parameters do not change in the four different cases using the well-identified time series processes. The variability of projected mortality rates will be identical in each of the four cases. Using the AR(1) around linear drift process, we also find

$$\begin{aligned}\mathbb{E}\eta_{x,\tau} &= \alpha_x + \kappa_{2010} + (\tau - 2010)\mu \\ &\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) + \beta_0 + \beta_1 \times (\tau - x) \qquad (31)\end{aligned}$$

From the results of Section 5.3, we can see that if we transform the parameters of the classic APC model using the transformation in Equations 4, 5 and 6, and then project them using well-identified time series processes, we obtain

$$\begin{aligned}\hat{\alpha}_x &= \alpha_x - a - b + c(x - \bar{x}) \\ \mathbb{E}\hat{\kappa}_\tau &= \hat{\kappa}_{2010} + \hat{\mu}(\tau - 2010) \\ &= \kappa_{2010} + a - c(2010 - \bar{t}) + (\mu - c)(\tau - 2010) \\ &= \kappa_{2010} + a - c(\tau - \bar{t}) + \mu(\tau - 2010) \\ \mathbb{E}\hat{\gamma}_{\tau-x} &= \rho^{\tau-x-1950}(\hat{\gamma}_{1950} - \hat{\beta}_0 - \hat{\beta}_1 \times 1950) + \hat{\beta}_0 + \hat{\beta}_1 \times (\tau - x) \\ &= \rho^{\tau-x-1950}(\gamma_{1950} + b + c(1950 - x - \bar{y}) - \beta_0 - b - c\bar{y} - (\beta_1 + c) \times 1950) \\ &\quad + \beta_0 + b + c\bar{y} + (\beta_1 + c) \times (\tau - x) \\ &= \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) \\ &\quad + \beta_0 + \beta_1(\tau - x) + c(\tau - x - \bar{y})\end{aligned}$$

Hence, the expectation of $\eta_{x,t}$ in Equation 31, after applying the invariant transformations, becomes

$$\begin{aligned}
\mathbb{E}\hat{\eta}_{x,\tau} &= \hat{\alpha}_x + \hat{\kappa}_{2010} + (\tau - 2010)\hat{\mu} \\
&\quad + \rho^{\tau-x-1950}(\hat{\gamma}_{1950} - \hat{\beta}_0 - \hat{\beta}_1 \times 1950) + \hat{\beta}_0 + \hat{\beta}_1 \times (\tau - x) \\
&= \alpha_x - a - b + c(x - \bar{x}) + \kappa_{2010} + a - c(\tau - \bar{t}) + \mu(\tau - 2010) \\
&\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) \\
&\quad + \beta_0 + \beta_1(\tau - x) + c(\tau - x - \bar{y}) \\
&= \alpha_x + \kappa_{2010} + (\tau - 2010)\mu \\
&\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) + \beta_0 + \beta_1 \times (\tau - x) \\
&= \mathbb{E}\eta_{x,\tau}
\end{aligned}$$

We can therefore see how changes in the linear drift of the period functions between the different cases cancel with the changes in the linear drift in the cohort functions to give exactly the same expected projected mortality rates in all four cases.[32] We, therefore, see in practice what was derived theoretically in Section 5.3, namely that using a random walk with drift process for the period parameters and an AR(1) around linear drift process for the cohort parameters gives well-identified projections for the classic APC model, and so the projected mortality rates which do not depend upon the identifiability constraints imposed.

Projections using an AR(1) process around a linear drift might be felt to conflict with our desired demographic significance for the cohort parameters, i.e., that they should exhibit no long-term trends. However, demographic significance is subjective and so should not be used to override a greater concern that the projected mortality rates do not depend upon the arbitrary identifiability constraints. Fortunately, there are methods for obtaining well-identified projections of the cohort parameters which do conform to our desired demographic significance of trendlessness.

In order to lack trends, the drift coefficients of the process, β_0 and β_1, should be zero. Looking again at Table 2, one might think that the values of β_0 and β_1 are quite small, and therefore be tempted to test them statistically with a view to setting them to zero. This, however, would be a mistake.

[32] For example, in all four cases $\mathbb{E}\eta_{60,2020} = -4.6413$.

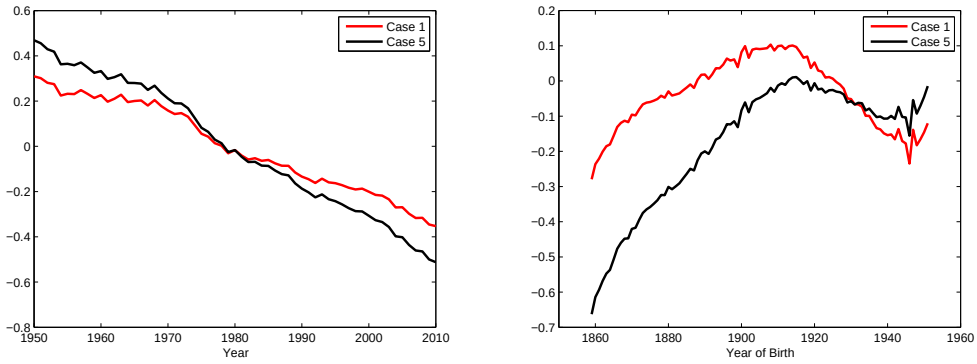
As shown in Section 5.3, the values of β_0 and β_1 change under the invariant transformations of the classic APC model and, therefore, will depend upon the identifiability constraints chosen. Consequently, the results of any statistical analysis of their significance will also depend upon the arbitrary identifiability constraints, which is not desirable.

The reason that β_0 and β_1 are "small" is because we have imposed this via the identifiability constraints. All four sets of identifiability constraints were chosen to set the level of the cohort parameters to be around zero and to have no linear trends over the whole range of the data. Therefore, we would expect to find low values of β_0 and β_1, which control the level and drift to which the process mean-reverts. We could have chosen other, equally reasonable constraints based on alternative subjective interpretations of the demographic significance of the period and cohort parameters which would have resulted in far larger values of β_0 and β_1 and given exactly the same fitted and projected mortality rates. We therefore see that whether or not these parameters are "small", and consequently whether or not they pass a statistical test of their significance, is solely dependent upon the arbitrary identifiability constraints we have chosen.

The four cases in Section 5.2 were motivated by the same desired demographic significance for the cohort parameters - that they should be centred around zero and not have any linear trends. However, the four different cases used four different interpretations of these subjective requirements, and therefore arrived at four different interpretations of what it means to be centred around zero and trendless. These different interpretations resulted in the four different sets of identifiability constraints. Using an AR(1) around linear drift process to project the cohort functions introduces a fifth interpretation for the meaning of being centred around zero and having no linear drift, in this case, that the time series parameters β_0 and β_1 are equal to zero. Therefore, we could use another set of parameters with the identifiability constraints

Case 5: $\sum_t \kappa_t = 0$, $\beta_0 = 0$ and $\beta_1 = 0$

This set of constraints gives identical fitted and projected mortality rates to the other cases, but gives projected cohort parameters which mean-revert around zero, which accords better with our demographic significance. However, the restrictions in Case 5 cannot be known at the time of fitting the



(a) Period parameters κ_t (b) Cohort parameters γ_y

Figure 3: Projecting the parameters of the classic APC model: Cases 1 and 5

model to data, since the appropriate time series process that will be used to project the cohort parameters cannot be known at that stage. To use this set of constraints, we need to do the following:

1. fit the model to data, applying some convenient set of identifiability constraints which can be known in advance of analysing the time series structure of the parameters, e.g., those in Case 1;

2. estimate values for β_0 and β_1 for these historical parameters by fitting the AR(1) around a linear drift process in Equation 28 to them;

3. use these estimated values for β_0 and β_1 in the transformations in Equations 5 and 6 to obtain a new set of (equivalent) age, period and cohort parameters.

The period and cohort parameters for Case 5, compared with those for Case 1, are shown in Figure 3. Using the Case 5 parameters may appear unnatural as the cohort parameters in this case appear to possess a linear trend. However, when we project using the well-identified AR(1) around linear drift process, we find no linear drift in these parameters, merely mean reversion to a level of zero, which fits well with the demographic significance for the cohort parameters discussed in Hunt and Blake (2015f).

5.5 Projecting the Plat model

We will now use this analysis to specify a set of well-identified projection processes for the Plat model discussed in Section 4.1.1. As described in that section, the invariant transformations of the model can be written in the form of Equation 9 with

$$\begin{aligned}
\hat{\alpha}_x &= \alpha_x - a_1 - a_2 - a_3 - b + c(x-\bar{x}) - d(x-\bar{x})^2 && = \alpha_x - a(x) \\
\hat{\kappa}_t^{(1)} &= \kappa_t^{(1)} + a_1 - c(t-\bar{t}) - d(t-\bar{t})^2 && = \kappa_t^{(1)} - k^{(1)}(t) \\
\hat{\kappa}_t^{(2)} &= \kappa_t^{(2)} + a_2 + 2d(t-\bar{t}) && = \kappa_t^{(2)} - k^{(2)}(t) \\
\hat{\kappa}_t^{(3)} &= \kappa_t^{(3)} + a_3 && = \kappa_t^{(3)} - k^{(3)}(t) \\
\hat{\gamma}_y &= \gamma_y + b + c(y-\bar{y}) + d(y-\bar{y})^2 && = \gamma_y + g(y)
\end{aligned}$$

by composing the transformations in Equations 4 (for each period function), 5, 6 and 13.

Starting with the cohort parameters, we may wish to retain the demographic interpretation that they should be stationary and mean reverting and so wish to use an AR(1) structure. However, from the discussion in Section 5.3 and the observation that $g(y)$ is quadratic for the Plat model, we therefore require that $\Gamma(y)$ in Equation 26 is quadratic. In order to give well-identified projections, we would therefore project the cohort parameters using an AR(1) around quadratic drift process, i.e.,

$$(1-\rho L)(\gamma_y - \beta_0 - \beta_1 y - \beta_2 y^2) = \varepsilon_y \tag{32}$$

Simple insertion of $\hat{\gamma}_y = \gamma_y + g(y)$ into this shows that it does not change structure under the invariant transformation and so is well-identified. In principal, we could then decide to switch to an equivalent set of parameters with the constraints $\beta_0 = \beta_1 = \beta_2 = 0$ in the same manner as for the classic APC model. This may be desirable as it gives projected cohort parameters which mean-revert around zero, in line with our demographic significance. In addition, when more complicated methods are used to project the cohort parameters, it might be felt to simplify the process of projection.[33]

For the period parameters, we may wish to use a random walk with drift structure as we did for the classic APC model on the demographic

[33] For an example where this is the case, see Hunt and Blake (2015b).

interpretation that the period functions should be non-stationary. This would be written as

$$(1-L)\boldsymbol{\kappa}_t = \boldsymbol{\xi}(t) + \boldsymbol{\epsilon}_t \tag{33}$$

where $\boldsymbol{\kappa} = \left(\kappa_t^{(1)}, \quad \kappa_t^{(2)}, \quad \kappa_t^{(3)}\right)^\top$ as discussion in Section 2 and similarly for $\boldsymbol{\xi}(t)$ and $\boldsymbol{\epsilon}_t$.

Using this notation, we can group the transformations of the period functions as

$$\begin{aligned}\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\kappa}_t + \begin{pmatrix} a_1 + c\bar{t} - d\bar{t}^2 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} -c + 2d\bar{t} \\ 2d \\ 0 \end{pmatrix} t + \begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix} t^2 \\ &= \boldsymbol{\kappa}_t + \boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2\end{aligned}$$

In Section 5.3, we showed that in order to ensure identifiability, we needed

$$\begin{aligned}\boldsymbol{\xi}(t) &= (1-L)(\boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2) \\ &= \boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2 - \boldsymbol{k}_0 - \boldsymbol{k}_1(t-1) + \boldsymbol{k}_2(t-1)^2 \\ &= \boldsymbol{k}_1 - \boldsymbol{k}_2 + 2\boldsymbol{k}_2 t \\ &= \begin{pmatrix} -c + 2d\bar{t} + d \\ 2d \\ 0 \end{pmatrix} + 2\begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix} t\end{aligned}$$

Therefore, we see that, in order for the Plat model to have well-identified projections, we require a constant drift component for $\kappa_t^{(2)}$ (i.e., $\xi^{(2)}(t) = \mu_0^{(2)}$, a constant) and a linear drift component for $\kappa_t^{(1)}$ (i.e., $\xi^{(1)}(t) = \mu_0^{(1)} + \mu_1^{(1)} t$, a linear function of time). This can be written as

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \mu X_t + \boldsymbol{\epsilon}_t \tag{34}$$

where

$$\mu = \begin{pmatrix} \mu_0^{(1)} & \mu_t^{(1)} \\ \mu_0^{(2)} & 0 \\ 0 & 0 \end{pmatrix}$$

and $X_t = \begin{pmatrix} 1, & t \end{pmatrix}^\top$. We can see that this form of the random walk with drift process extends naturally to allow for other unidentifiable trends by choosing the "trend" matrix, X_t, and corresponding "drift" matrix, μ, appropriately. The need to use a random walk with linear drift is often overlooked, for instance in Plat (2009) and Börger et al. (2013) (who used a model which nests the reduced Plat model) - see also Hunt and Blake (2015b).

We also see that different drifts are required for different period functions in order to give well-identified projections of mortality rates. This runs counter to the desire to treat all the period functions the same, as discussed in Hunt and Blake (2015d). However, using the same drifts for all the period functions can give projections which are not biologically reasonable. For example, allowing for a quadratic trend in $\kappa_t^{(3)}$ can result in apparent changes in trend which are inconsistent with the historical data. In Hunt and Blake (2015d), we also found that we can treat different period functions differently in models with parametric age functions, because there were no invariant transformations of the model which could be used to interchange the age/period terms. It may, therefore, be preferable to allow for different drifts in different period functions in the Plat (2009) model to obtain well-identified projected mortality rates which are also biologically reasonable.[34] We should, therefore, be prepared to override the desire to treat the period functions identically if the alternative is to put biological reasonableness at stake. See Hunt and Blake (2015b) for an example of this issue in practice.

5.6 Summary

APC mortality models which have unidentifiable trends at the fitting stage require extra care when projected to ensure that the projections do not depend on the identifiability constraints chosen. In general, we find that the projection method used must preserve whatever trends were unidentifiable at the fitting stage. For example, the processes which were well-identified for the classic APC model discussed in Section 5.4 preserved linear trends, which were shown to be unidentifiable in Section 3.

Such an approach generalises naturally for more complicated mortality

[34] Using different drifts for the different period functions will mean, however, that time series processes will be required for equivalent models.

models, such as the Plat model discussed in Sections 4.1.1 and 5.5. However, models with higher order polynomial age functions have higher order unidentifiable trends (as shown in Section 4.1), and so require projection processes which allow for these trends. This may cause problems for long term projections.

For example, consider the model

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma_x)\kappa_t^{(3)} + \gamma_{t-x} \qquad (35)$$

which extends model M7 of Cairns et al. (2009) with a static age function (as was done in Haberman and Renshaw (2011)). We can see that a model of this form possesses age functions which span the polynomials to quadratic order. From Section 4.1, we know, without performing any additional analysis, that it has unidentifiable cubic trends in both the cohort parameters and $\kappa_t^{(1)}$ which will need to be allowed for in projection. However small they may be in the historical data, these cubic trends will eventually come to dominate the long term evolution of mortality rates, potentially yielding projected mortality rates which lack biological reasonableness due to apparent changes in trend.

Consequently, it may be prudent to avoid unidentifiable cubic (and higher) order polynomial trends in an APC mortality model. Such trends arise when we use more complicated models with higher-order polynomial age functions. It is therefore useful, when selecting such models, to have a larger "toolkit" of age functions for use in the models than simply extending existing models by using higher-order polynomial terms. Hunt and Blake (2014) proposed such a toolkit, which allows for more complicated mortality models that do not suffer from excessive identifiability issues and can give biologically reasonable, well-identified projections of mortality rates, as shown in Hunt and Blake (2015b,a,e).

6 Conclusions

In Hunt and Blake (2015d), we saw how AP mortality models are not fully identified, and that in order to identify these models, most users impose additional arbitrary identifiability constraints on them when fitting the models to data. Some APC mortality models have extra identifiability constraints,

caused by the collinearity between age, period and cohort, which are unlike anything found in similar AP models. These depend upon the form of the age functions in the model and so are specific to individual models. The identifiability issues involve deterministic trends which cannot be uniquely allocated between the age, period or cohort terms and so an arbitrary allocation must be made via additional arbitrary identifiability constraints. The nature of the unidentifiable trends present in specific models are summarised in Figure 1.

These unidentifiable deterministic trends have important consequences when we come to project the model. We must first determine the identifiability issues in the specific model we are using, in order to find which deterministic trends are unidentifiable. When this is done, we can specify suitable time series processes for the variation around these trends. Only by doing this can we ensure that our projected mortality rates are independent of the arbitrary identifiability constraints imposed when fitting the model.

By understanding these identifiability issues, however, we can build more complex mortality models, for instance, via the "general procedure" of Hunt and Blake (2014), and be confident that they are founded on a secure knowledge of the underlying mathematical structure of APC mortality models. We are also able to use more sophisticated time series projection methods, as in Hunt and Blake (2015b), Hunt and Blake (2015a) and Hunt and Blake (2015e), knowing that our projections are free from dependence on the arbitrary choices we made when fitting the model to data.

A Identifiability in APC models with non-parametric age functions

In discussing whether a model with non-parametric age functions has any additional issues with identifiability when a cohort term is added, it is useful to begin with a recap of some of the notation used and results from Hunt and Blake (2015d).

A.1 Identifiability in AP models

In Hunt and Blake (2015d), we found that it was helpful to write Equation 1 in matrix form as

$$H = \alpha \mathbf{1}_T^\top + \beta\kappa \tag{36}$$

where

- H is the $(X \times T)$ matrix of transformed data (i.e. $H = \{\eta_{x,t}\}$),
- α is a $(X \times 1)$ matrix of the static age function,
- $\mathbf{1}_T$ is a $(T \times 1)$ matrix of ones, and
- β and κ are the $(X \times N)$ and $(N \times T)$ matrices of age and period functions constructed above, respectively.

When expressed in this form, AP models can be analysed through the prism of matrix algebra and linear mathematics. We can then see that there is a lack of identifiability in the model which allowed us to perform certain transformations on the parameters given in Equations 37 and 38 without affecting the fitted mortality rates

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} = \{\alpha, \beta A^{-1}, A\kappa\} \tag{37}$$

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} = \{\alpha - \beta B, \beta, \kappa + B\mathbf{1}_T^\top\} \tag{38}$$

These invariant transformations can be used to impose additional arbitrary identifiability constraints to set the "level" and "normalisation" of the age/period terms and potentially to orthogonalise them.[35] These freedoms allowed us to impose our desired demographic significance on the parameters, but meant that care had to be taken to ensure that projections from the model were identifiable, i.e., were independent of our arbitrary identifiability constraints. In Hunt and Blake (2015d), we also found that our treatment of the identification issues was subtly different depending on whether the model had parametric or non-parametric age functions, as by defining the age functions a priori, we were unable to use the transformations in Equation 37 without altering the age functions and therefore fundamentally changing the model.

[35] In the sense of ensuring that $\sum_x \beta_x^{(i)}\beta_x^{(j)} = 0$ and $\sum_t \kappa_t^{(i)}\kappa_t^{(j)} = 0$ for $i \neq j$.

A.2 Identifiability in APC models

Equation 36 can be extended to allow for cohort effects

$$H = \alpha \mathtt{1}_T^\top + \beta\kappa + \gamma \tag{39}$$

where γ is an $(X \times T)$ Toeplitz matrix, i.e., a matrix where the diagonal elements are constant. It is clear that the transformations in Equations 37 and 38 are still invariant transformations of Equation 39 and therefore the conclusions of Hunt and Blake (2015d) are still applicable in the wider context of APC mortality models. Indeed, the transformation in Equation 4 of the classic APC model is simply the transformation in Equation 38 applied to this specific model.

Generalising Equation 5 in this context, we can see that the transformation

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{\gamma}\} = \{\alpha - c\mathtt{1}_X, \beta, \kappa, \gamma + c\mathtt{1}_X \mathtt{1}_T^\top\} \tag{40}$$

is common to all APC models of the form in Equation 39 (where $\mathtt{1}_X$ has a similar definition as $\mathtt{1}_T$ above). This transformation was also discussed (using alternative notation) in Section 4. This allows us to set the level of the cohort parameters - typically to be around zero to impose the demographic significance discussed in Hunt and Blake (2015f).

To generalise the transformation in Equation 6 for more complicated invariant transformations, if we can find a Toeplitz matrix Γ such that[36]

$$\Gamma = a\mathtt{1}_T^\top + \beta k \tag{41}$$

(with a an $(X \times 1)$ matrix and k an $(N \times T)$ matrix), we then have the transformation

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{\gamma}\} = \{\alpha - a, \beta, \kappa - k, \gamma + \Gamma\} \tag{42}$$

In the case of the classic APC model, we have $\beta = \mathtt{1}_X$ and so can find a Toeplitz matrix $\Gamma = c(\mathtt{1}_X \boldsymbol{T}^\top - \boldsymbol{X}\mathtt{1}_T^\top)$ where $\boldsymbol{X}$ is the $(X \times 1)$ column vector $X_i = \{i - \bar{x}\}$ where i runs from 1 to X (and similarly for $\boldsymbol{T}$).

[36] We actually require the more general statement that $\Gamma = a\mathtt{1}_T^\top + bk$ with b a $(X \times N)$ matrix such that $\beta = bA$, i.e., the columns of b lie within the span of the columns of β. However, without loss of generality, we define $\tilde{k} = Ak$ to obtain the result shown.

Theorem 3 *There are no invariant transformations of general APC mortality models with non-parametric age functions, i.e., no such A, k and Γ exist unless a specific shape for β is assumed in the model.*

Sketch of Proof Consider the general term $a\mathbf{1}_T^\top + \beta k$, which is analogous to the predictor structure of an AP mortality model. As we argue in Hunt and Blake (2015d), this has dimension $X + N(X+T) - N(N+1)$, i.e., the X parameters in a, the NX parameters in β, and the NT in k reduced by the $N(N+1)$ degrees of freedom in the transformations in Equations 37 and 38.

In contrast, in the general case, Γ has dimension $X+T-1$, i.e., one degree of freedom for each diagonal. For Equation 41 to be true, these matrices must have the same dimension and therefore

$$\begin{aligned} X + N(X+T) - N(N+1) &= X + T - 1 \\ N^2 + N(1-X-T) + T - 1 &= 0 \end{aligned} \tag{43}$$

However, N, X and T are integers, set by the structure of the model and the range of the data, and therefore Equation 43 will not generally be true. Hence Equation 42 will not be an invariant transformation of a general APC mortality model with non-parametric age functions.

The argument used in this proof relies on $a\mathbf{1}_T^\top + \beta k$ being of full rank and therefore breaks down if β is of lower dimension than the maximum possible. However, this is equivalent to imposing a parametric form on the age functions and accordingly, the line of reasoning above is not possible in the general case.

Therefore, general non-parametric APC mortality models do not possess any other invariant transformations apart from the ones in Equations 37, 38 and 40. They require only identifiability constraints which set the normalisation scheme of the age functions, impose orthogonality between the age and period functions (both using the transformation in 37), set the levels of the period functions $\kappa_t^{(i)}$ using Equation 38, and the level of the cohort parameters γ_{t-x} using Equation 40.

■

For instance, we see that for the H1 model of Haberman and Renshaw (2009) and Hunt and Villegas (2015),

$$\eta_{x,t} = \alpha_x + \beta_x \kappa_t + \gamma_{t-x} \tag{44}$$

we cannot find an invariant transformation of the parameters similar to that in Equation 6. This is because of the lack of shape in either age or period in the $\beta_x \kappa_t$ term which can be used to decompose the cohort term. However, this model does possess an "approximate" identifiability constraint, which leaves the fitted mortality rates almost unchanged in the majority of cases. This is caused by κ_t often having a form that is close being parametric, which is discussed in detail in Hunt and Villegas (2015).

Some, especially demographers, have argued that all cohort effects are simply mis-specified age/period effects and are best modelled as such.[37] Although this may be true in a strictly mathematical sense, a large number of age/period terms are required to replicate any general cohort term in the model. It is therefore more parsimonious to include a set of cohort parameters rather than multiple age/period terms. This, again, is similar to the argument in Wilmoth (1990), which states that it is plausible and parsimonious to include a single set of cohort parameters rather than an excessive number of age/period terms which achieve the same effect.

Some datasets may show little or no structure across years of birth, in which case the decision to include a cohort term becomes one decided on the basis of the demographic and statistical significance of the parameters for that dataset. Such a decision can be made only after all significant age/period terms have been identified. We therefore recommend a procedure, such as the "general procedure" in Hunt and Blake (2014), which only adds such a term when justified by the data.

B Models without a static age function

As we discuss in Hunt and Blake (2015f), a number of APC mortality models have been proposed which do not have an explicit static age function, α_x, the

[37] For instance, Cairns et al. (2011) raised *"the possibility that cohort effects might be partially or completely replaced by well-chosen age and period effects"* and also see Murphy (2010)

most prominent of which being the extensions of the CBD model in Cairns et al. (2009). If the model does not have an explicit static age function, the age functions in the model must be parametric and therefore known in advance of fitting the model to data. The structure of the APC model in this case is therefore

$$\eta_{x,t} = \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} + \gamma_{t-x}$$

The identifiability issues in such models can be considered in the same fashion as in Section 4. In particular, we noted in Section 4.2 that the invariant transformations of models with exponential or trigonometric age functions did not involve the static age function, and therefore are also applicable in models without one.

The invariant transformations of models with polynomial age functions, in contrast, did involve the static age function explicitly. The proof of Theorem 1 involves expanding a polynomial function of year of birth, $g(y)$, into polynomial terms in x and t and then combining these in the appropriate age/period terms. In particular, the term in this expansion with no t dependence was combined into the static age function. This is seen most clearly in the transformation in Equation 6, but also in the transformation in Equation 13 for the Plat model.

However, we can see that the lack of a static age function to absorb this term in the expansion of $g(y)$ is not an insurmountable problem as long as there is an age/period term with the appropriate age function. This means that if $g(y)$ is a polynomial of order M, we must have age functions in the model up to order M as well. This contrasts with models with a static age function, which only require age functions up to order $M - 1$.

Theorem 4 *APC mortality models with no static age function and age functions spanning the polynomials to order M possess invariant transformations which adds a polynomial of order M to the cohort function.*

Sketch of Proof The proof is similar to that of Theorem 1. Take $g(y)$, a general polynomial of order M, and expand as a function of x and t. This can then be regrouped into an equivalent form that corresponds to the

age/period terms in the model, in order to see how $g(y)$ can be absorbed into the age/period structure

$$\begin{aligned}
g(y) &= \sum_{n=0}^{M} a_n y^n \\
\Rightarrow g(t-x) &= \sum_{n=0}^{M} a_n (t-x)^n \\
&= \sum_{n=0}^{M} a_n \sum_{m=0}^{n} \binom{n}{m} t^m (-x)^{n-m} \\
&= \sum_{n=0}^{M} \sum_{l=0}^{n} a_n \binom{n}{l} t^{n-l} (-x)^l \\
&= \sum_{l=0}^{M} (-x)^l \sum_{n=l}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{l=0}^{M} (-1)^l f^{(l)}(x) \sum_{n=l}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{l=0}^{M} f^{(l)}(x) k^{(l)}(t)
\end{aligned}$$

which is of the form of Equation 9 if the age functions in the model are of the form $f^{(j)}(x) = x^j$ of $j = 0, 1, \ldots M$. ■

To see this in practice, consider model M6 of Cairns et al. (2009)

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{45}$$

and compare it with the reduced Plat model of Equation 12 in Section 4.1.1. For the reduced Plat model, we saw that the transformation in Equation 13 was invariant, and involved adding a quadratic function of year of birth to the cohort parameters, with adjustments to $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and the static age function α_x. For model M6, this transformation is not permitted, as there is no static age function to adjust in this model. Instead, the model only has the simpler linear invariant transformation

$$\{\hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\kappa_t^{(1)} - c(t - \bar{t}), \kappa_t^{(2)} - c, \gamma_y - c(y - \bar{y})\} \tag{46}$$

We can also see this using the analysis of Hunt and Blake (2015f), where it was shown that models without a static age function can be written as though they do have one of a specific, parametric form that has been combined with the other age/period terms in the model. In the case of model M6, we see that this implies a static age function which is a linear function of age, which then could not be used to absorb a quadratic age term coming from the addition of a quadratic function of year of birth to the cohort parameters. Consequently there is a trade-off: models without a static age function have simpler identifiability issues than (otherwise similar) models possessing one, but are unable to provide a good fit to mortality data across the full age range, as discussed in Hunt and Blake (2015f).

C Maximal invariants

An alternative approach to using an arbitrary identification scheme was suggested by Kuang et al. (2008b,a) and Nielsen and Nielsen (2014) for the classic APC model. This is to change the parameterisation of the model to an equivalent form with reduced dimensionality which does not suffer from identifiability issues. The new parameters are known as "maximal invariant" parameters, since they are the set with the largest number of parameters (i.e., are "maximal"), and are injective[38] and give the same fitted mortality rates as the original model in Equation 1 (i.e., the reparameterisation is "invariant") . We can think of this as finding a parameterisation of the model which gives the same fit to data, but where every possible degree of freedom in the model is fully utilised in fitting the data.

Kuang et al. (2008b) and Nielsen and Nielsen (2014) proposed an approach to generating a maximally invariant parameterisation for the classic APC model based on finding the second differences of the age, period and cohort terms. These second differences do not change under the invariant transformations of the model and so have a meaning independent of the identifiability constraints. In this Appendix, we review this approach and discuss how it can be extended to deal with the identifiability issues in some of the more complex APC mortality models. However, we also find that it suffers from a number of limitations which make it unsuitable for many

[38] A transformation is injective if different points in the domain get mapped to different points in the image of the transformation.

APC models and which can cause projections to be biologically unreasonable.

First, the age, period and cohort functions in the classic APC model are expanded as telescopic sums in terms of their second differences, i.e.,

$$\begin{aligned}
\alpha_x &= \alpha_X - \sum_{i=x+1}^{X} \Delta\alpha_i \\
&= \alpha_X - \sum_{i=x+1}^{X} \left(\Delta\alpha_X - \sum_{j=i+1}^{X} \Delta^2\alpha_j \right) \\
&= \alpha_X - (X-x)\Delta\alpha_X + \sum_{i=x+1}^{X} \sum_{j=i+1}^{X} \Delta^2\alpha_j \\
\kappa_t &= \kappa_1 + (t-1)\Delta\kappa_2 + \sum_{i=2}^{t} \sum_{j=3}^{t} \Delta^2\kappa_j \\
\gamma_y &= \gamma_{1-X} + (y-1+X)\Delta\gamma_{2-X} + \sum_{i=2-X}^{y} \sum_{j=3-X}^{y} \Delta^2\gamma_j
\end{aligned}$$

In the case of the age function, α_x, we work backwards from α_X due to the negative dependence of cohort on age. However, it is important to note that this expansion has not changed the number of parameters in the model, merely written them in a new form. This, of itself, will not solve the identifiability issues. However, Kuang et al. (2008b) and Nielsen and Nielsen (2014) then substituted the second difference expansions of the parameters into the classic APC model and group the deterministic terms together

$$\eta_{x,t} = a_0 + (X-x)a_1 + (t-1)b_1 + \sum_{i=x+1}^{X} \sum_{j=i+1}^{X} \Delta^2\alpha_j + \sum_{i=2}^{t} \sum_{j=3}^{i} \Delta^2\kappa_j + \sum_{i=2-X}^{t-x} \sum_{j=3-X}^{i} \Delta^2\gamma_j \tag{47}$$

where

$$\begin{aligned}
a_0 &= \alpha_X + \kappa_1 + \gamma_{1-X} \\
a_1 &= \Delta\gamma_{2-X} - \Delta\alpha_X \\
b_1 &= \Delta\kappa_2 + \Delta\gamma_{2-X}
\end{aligned}$$

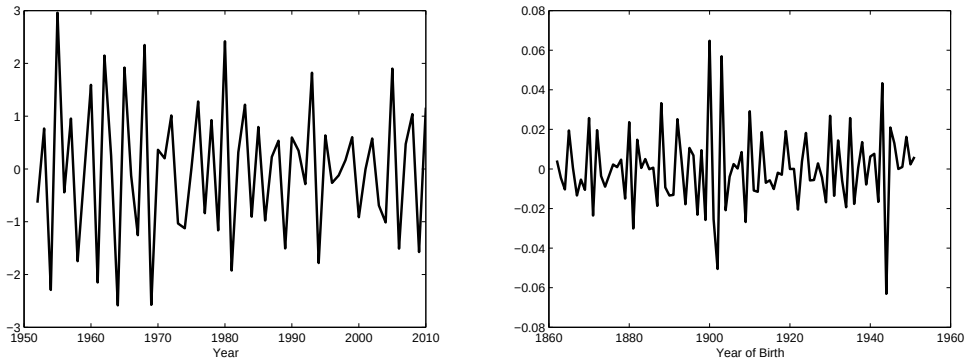
In Kuang et al. (2008b) and Nielsen and Nielsen (2014), these new parameters were introduced by considering three points of the fitted mortality surface. The most important point about the procedure is that it replaces six parameters in the original parameterisation with only three in the maximally invariant parameterisation. The maximally invariant parameterisation therefore contains $3 + (X - 2) + (T - 2) + (T + X - 3) = 2X + 2T - 4$ free parameters. This compares with $2X + 2T - 1$ parameters and the three additional identifiability constraints required by the three invariant transformations - Equations 4, 5 and 6 - for the original parameterisation of the classic APC model. Hence the maximally invariant parameterisation gives the same fitted mortality rates with the same number of effective parameters but without the over-parameterisation and consequent need for identifiability constraints in the original formulation of the model.

However, by doing this, we have lost much of the demographic significance associated with the original parameters in the classic APC model. For example, whilst α_x in the original parameterisation of the classic APC model relates to an age effect specific to age x, $\Delta^2\alpha_x$ relates to the curvature of the mortality curve in the age dimension at age x and will impact mortality rates at all ages below x. It is therefore harder to explain its demographic significance to other model users or develop an intuition about what values are reasonable in order to check the validity of the model. Although demographic significance is subjective, it is still not desirable to lose it if it can be avoided. This may restrict the usefulness of the maximally invariant approach.

In order to project the model into the future, we need to analyse the $\Delta^2\kappa_t$ and $\Delta^2\gamma_y$ parameters as time series. These are shown in Figure 4 for the same dataset as used in Section 5.2. As can be seen,[39] these parameters appear to be stationary and so it is natural to project them using an ARMA process.

If we were to "integrate up" the double differences to recover our original κ_t and γ_y parameters, these would both be $I(2)$ processes. This conflicts with the demographic significance for the cohort parameters discussed in

[39]We have removed the large outlier cohort effects for years of birth 1918/19 using indicator variables, as they are believed to be data artefacts resulting from the surge of births due to the demobilisation of soldiers after the First World War, based on similar reasons as those presented in Richards (2008) and Cairns et al. (2014).



(a) $\Delta^2\kappa_t$ (b) $\Delta^2\gamma_y$

Figure 4: Second differences from the classic APC model

Hunt and Blake (2015f). $I(2)$ processes are also not likely to be biologically reasonable, as the uncertainty in projected mortality rates would grow very quickly. This would have important ramifications if the model is projected.

The maximal invariant approach also works with some other APC mortality models. For instance, consider the reduced Plat model of Equation 12. This model has $X + 2T + (X + T - 1) = 2X + 3T - 1$ parameters and, as discussed in Section 4.1.1, we know that it requires five identifiability constraints to fully identify (two for the level of the period functions and one each for the level, linear trend and quadratic trend in the cohort parameters).

In order to find a maximally invariant parameterisation, we follow the same logic as in Kuang et al. (2008b) and consider the telescopic sums of the parameters. However, as α_x, $\kappa_t^{(1)}$ and γ_y all possess unidentifiable quadratic trends, we need to consider the third differences of these parameters, but only consider the second differences of $\kappa_t^{(2)}$, since it only has unidentifiable

linear trends

$$\alpha_x = \alpha_X - (X-x)\Delta\alpha_X + \frac{1}{2}(X-x)(X-1-x)\Delta^2\alpha_x - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k$$

$$\kappa_t^{(1)} = \kappa_1^{(1)} + (t-1)\Delta\kappa_2^{(1)} + \frac{1}{2}(t-1)(t-2)\Delta^2\kappa_3^{(1)} + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)}$$

$$\kappa_t^{(2)} = \kappa_1^{(2)} + (t-1)\Delta\kappa_2^{(2)} + \sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)}$$

$$\gamma_y = \gamma_{1-X} + (y-1+X)\Delta\gamma_{2-X} + \frac{1}{2}(y-1+X)(y-2+X)\Delta^2\gamma_{3-X} + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k$$

Combining these in Equation 12 and grouping the deterministic terms of the same type reduces the dimension of the parameter set in the same manner as for the classic APC model. Therefore, we find the maximally invariant form of the reduced Plat model

$$\begin{aligned}\eta_{x,t} = {} & a_0 + (x-\bar{x})a_1 + (x-\bar{x})^2a_2 + (t-\bar{t})b_1 + (t-\bar{t})^2b_2 + (x-\bar{x})(t-\bar{t})c_1 \\ & - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)} + (x-\bar{x})\sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)} \\ & + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k \end{aligned} \tag{48}$$

The final step to prove that this is a maximally invariant parameterisation would be to check that each of the parameters can be estimated uniquely from the data. Alternatively and more easily, we can see that it is maximally invariant from a dimensional argument, since the parameterisation has $6 + (X-3) + (T-3) + (T-2) + (X+T-4) = 2X + 3T - 6$ free parameters, which is the same as the number of parameters in the original reduced Plat model less the number of identifiability constraints imposed. Therefore, the freely varying parameter space has the same dimension as the model space and gives the same fitted mortality rates as the original model, and so the parameters represent maximal invariants. Because of this, the

revised model does not possess any identification issues.

As in the case of the classic APC model, moving to a maximally invariant form for the model means losing the demographic significance of the parameters. The maximally invariant form of the reduced Plat model is highly unintuitive compared with the original parameterisation, and it would be difficult to communicate the impact of the various parameters to anyone not intimately familiar with the maximally invariant approach. As discussed in Hunt and Blake (2015f), since demographic significance is a major reason for choosing a model with parametric, as opposed to non-parametric age functions, this is highly undesirable. Also, and again similar to the classic APC model, the use of third differences for $\kappa_t^{(1)}$ and γ_y leads naturally to using $I(3)$ processes when we project the model, which are unlikely to give biologically reasonable projections.

Further, the maximal invariant approach does not work with all APC mortality models. If we follow the same logic to try to find the maximally invariant parameterisation for the full Plat model in Equation 11 we obtain

$$\begin{aligned}\eta_{x,t} &= a_0 + (x-\bar{x})a_1 + (x-\bar{x})^2 a_2 + (t-\bar{t})b_1 + (t-\bar{t})^2 b_2 + (x-\bar{x})(t-\bar{t})c_1 \\ &\quad - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)} + (x-\bar{x})\sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)} \\ &\quad + (x-\bar{x})^+\kappa_1^{(3)} + (x-\bar{x})^+\sum_{i=2}^{t}\Delta\kappa_i^{(3)} + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k \qquad (49)\end{aligned}$$

We know, from Section 4.1.1, that the Plat model has $X+3T+(X+T-1) = 2X+4T-1$ parameters and requires six identifiability constraints (three on the levels of the period functions and one each for the level, linear trend and quadratic trend in the cohort parameters). However, the maximally invariant parameterisation in Equation 49 has $7+(X-3)+(T-3)+(T-2)+(T-1)+(X+T-4) = 2X+4T-6$ free parameters, i.e., one too many. This is because the $(x-\bar{x})^+\kappa_1^{(3)}$ term cannot be combined with the expanded form of α_x, since it is not a polynomial. Consequently, there is no dimensional reduction with respect to this age/period term.

Because of this, we will still require an additional identifiability constraint to fit the model in Equation 49 to data. However, it is no longer clear what

this should be or what the underlying invariant transformation of the parameters is. The maximally invariant approach has therefore not solved the identifiability issues for this model, but has made making an arbitrary identification considerably more difficult.

This will be true for any age/period term which does not have a polynomial age function. As discussed in Section 4.3, such terms do not generate any additional identifiability issues beyond the unidentifiable level of the period function, as discussed in Hunt and Blake (2015d). It therefore may be possible to deal with this using an approach similar to that proposed for the model of Lee and Carter (1992) in Nielsen and Nielsen (2014) and discussed in the Appendix of Hunt and Blake (2015d). However, as these two techniques for obtaining maximally invariant parameterisations are fundamentally different, it is unclear how to combine them in models which mix polynomial and non-polynomial age functions, such as the Plat model.

In summary, the maximally invariant approach proposed in Kuang et al. (2008b) and Nielsen and Nielsen (2014) for the classic APC model can be generalised, but only to models with purely polynomial age functions. For models with other forms for the age functions (or which mix polynomial and non-polynomial age functions), the maximally invariant approach, at best, offers a partial solution. However, in using such an approach, we lose our desired demographic significance regarding the parameters in the model and are likely to obtain projected mortality rates which are not biologically reasonable, so this approach is not, in general, recommended.

References

Booth, H., Maindonald, J., Smith, L., 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56 (3), 325–36.

Börger, M., Fleischer, D., Kuksin, N., 2013. Modeling the mortality trend under modern solvency regimes. ASTIN Bulletin 44 (1), 1–38.

Cairns, A. J., Blake, D., Dowd, K., 2006. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin 36 (1), 79–120.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2011. Mortality density forecasts: An analysis of six stochastic mortality models. Insurance: Mathematics and Economics 48 (3), 355–367.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal 13 (1), 1–35.

Cairns, A. J., Blake, D., Dowd, K., Kessler, A., 2014. Phantoms never die: Living with unreliable mortality data. Tech. rep., Herriot Watt University, Edinburgh.

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

Clayton, D., Schifflers, E., 1987. Models for temporal variation in cancer rates. II: Age-period-cohort models. Statistics in Medicine 6 (4), 469–81.

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.

Glenn, N., 1976. Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. American Sociological Review 41 (5), 900–904.

Haberman, S., Renshaw, A., 2009. On age-period-cohort parametric mortality rate projections. Insurance: Mathematics and Economics 45 (2), 255–270.

Haberman, S., Renshaw, A., 2011. A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics 48 (1), 35–55.

Hobcraft, J., Menken, J., Preston, S. H., 1982. Age, period and cohort effects in demography: A review. Population Index 48 (1), 4–43.

Holford, T. R., 1983. The estimation of age, period and cohort effects for vital rates. Biometrics 39 (2), 311–24.

Human Mortality Database, 2014. Human Mortality Database. Tech. rep., University of California, Berkeley and Max Planck Institute for Demographic Research.
URL www.mortality.org

Hunt, A., Blake, D., 2014. A general procedure for constructing mortality models. North American Actuarial Journal 18 (1), 116–138.

Hunt, A., Blake, D., 2015a. A Bayesian approach to modelling and projecting cohort effects. Work in Progress.

Hunt, A., Blake, D., 2015b. Consistent mortality projections allowing for trend changes. Work in Progress.

Hunt, A., Blake, D., 2015c. Identifiability, cointegration and the gravity model. Tech. rep., Pensions Institute PI 15-05, Cass Business School, City University London.

Hunt, A., Blake, D., 2015d. Identifiability in age/period mortality models. Work in Progress.

Hunt, A., Blake, D., 2015e. Modelling longevity bonds: Analysing the Swiss Re Kortis Bond. Insurance: Mathematics and Economics 63, 12–29.

Hunt, A., Blake, D., 2015f. On the structure and classification of mortality models. Tech. rep., Pensions Institute PI 15-06, Cass Business School, City University London.

Hunt, A., Villegas, A. M., 2015. Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics 64, 186–202.

Kuang, D., Nielsen, B., Nielsen, J. P., 2008a. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 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 (4), 979–986.

Lee, R. D., Carter, L. R., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87 (419), 659– 671.

Murphy, M., 2009. The "golden generations" in historical context. British Actuarial Journal 15 (S1), 151–184.

Murphy, M., 2010. Re-examining the dominance of birth cohort effects on mortality. Population and Development Review 36 (2), 365–90.

Nielsen, B., Nielsen, J. P., 2014. Identification and forecasting in mortality models. The Scientific World Journal, Article ID 347043.

O'Brien, R. M., 2000. Age period cohort characteristic models. Social Science Research 29 (1), 123–139.

O'Brien, R. M., 2011. Constrained estimators and age-period-cohort models. Sociological Methods & Research 40 (3), 419–452.

O'Hare, C., Li, Y., 2012. Explaining young mortality. Insurance: Mathematics and Economics 50 (1), 12–25.

Osmond, C., 1985. Using age, period and cohort models to estimate future mortality rates. International Journal of Epidemiology 14 (1), 124–9.

Plat, R., 2009. On stochastic mortality modeling. Insurance: Mathematics and Economics 45 (3), 393–404.

Renshaw, A., Haberman, S., 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38 (3), 556–570.

Richards, S. J., 2008. Detecting year-of-birth mortality patterns with limited data. Journal of the Royal Statistical Society: Series A (Statistics in Society) 171 (1), 279–298.

Rodgers, W., 1982. Estimable functions of age, period, and cohort effects. American Sociological Review 47 (6), 774–787.

Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405 (6788), 789–792.

Willets, R., 1999. Mortality in the next millennium. Staple Inn Actuarial Society.

Willets, R., 2004. The cohort effect: Insights and explanations. British Actuarial Journal 10 (4), 833–877.

Wilmoth, J. R., 1990. Variation in vital rates by age, period and cohort. Sociological Methodology 20, 295–335.

Yang, Y., Fu, W. J., Land, K. C., 2004. A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. Sociological Methodology 34, 75–110.