
A stochastic implementation of the APCI model for mortality projections

S. J. Richards*, I. D. Currie, T. Kleinow and G. P. Ritchie

Abstract

The Age-Period-Cohort-Improvement (APCI) model is a new addition to the canon of mortality forecasting models. It was introduced by Continuous Mortality Investigation as a means of parameterising a deterministic targeting model for forecasting, but this paper shows how it can be implemented as a fully stochastic model. We demonstrate a number of interesting features about the APCI model, including which parameters to smooth and how much better the model fits to the data compared to some other, related models. However, this better fit also sometimes results in higher value-at-risk (VaR)-style capital requirements for insurers, and we explore why this is by looking at the density of the VaR simulations.

Keywords

Mortality Projections; APC; Lee–Carter; ARIMA Models; Solvency II; VaR

1. Introduction

Continuous Mortality Investigation (2016b) introduced a new model for fitting to mortality data: the Age-Period-Cohort-Improvement (APCI) model. This is an extension of the Age-Period-Cohort (APC) model, but it also shares an important feature with the model from Lee and Carter (1992). The APCI model was intended to be used as a means of parameterising a deterministic targeting model for mortality forecasting. However, it is not the purpose of this paper to discuss the Continuous Mortality Investigation's approach to deterministic targeting. Readers interested in a discussion of stochastic versus deterministic projections, in particular the use of targeting and expert judgement, should consult Booth and Tickle (2008). Rather, the purpose of this paper is to present a stochastic implementation of the APCI model for mortality projections, and to compare the performance of this model with various other models sharing similar structural features.

2. Data

The data used for this paper are the number of deaths $d_{x,y}$, aged x last birthday during each calendar year y , split by gender. Corresponding mid-year population estimates, $E_{x,y}^c$, are also given. The data therefore lend themselves to modelling the force of mortality, $\mu_{x+\frac{1}{2},y+\frac{1}{2}}$, without further adjustment. However, for brevity we will drop the $\frac{1}{2}$ and just refer to $\mu_{x,y}$.

*Correspondence to: Longevitas Ltd, 24a Ainslie Place, Edinburgh, EH3 6AJ. E-mail: stephen@longevitas.co.uk

We use data provided by the Office for National Statistics for the population of the United Kingdom. For illustrative purposes we will just use the data for males. As we are primarily interested in annuity and pension liabilities, we will restrict our attention to ages 50–104 over the period 1971–2015. Although data are available for earlier years, there are questions over the reliability of the population estimates prior to 1971. All death counts were based on deaths registered in the United Kingdom in a particular calendar year and the population estimates for 2002–2011 are those revised to take account of the 2011 census results. More detailed discussion of this data set, particularly regarding the current and past limitations of the estimated exposures, can be found in Cairns *et al.* (2015).

One consequence of only having data to age 104 is having to decide how to calculate annuity factors for comparison. One option would be to create an arbitrary extension of the projected mortality rates up to (say) age 120. Another alternative is to simply look at temporary annuities to avoid artefacts arising from the arbitrary extrapolation, as used by Richards *et al.* (2014). We use the latter approach in this paper, and we therefore calculate expectations of time lived and continuously paid temporary annuity factors as follows:

$$\bar{e}_{x,y:\overline{105-x}} = \int_0^{105-x} {}_t p_{x,y} dt \quad (1)$$

$$\bar{a}_{x,y:\overline{105-x}} = \int_0^{105-x} {}_t p_{x,y} v(t) dt \quad (2)$$

where $v(t)$ is a discount function and ${}_t p_{x,y}$ is the probability a life aged x at outset in year y survives for t years:

$${}_t p_{x,y} = \exp\left(-\int_0^t \mu_{x+s,y+s} ds\right) \quad (3)$$

Restricting our calculations to temporary annuities has no meaningful consequences at the main ages of interest, as shown in Richards *et al.* (2014). The methodology for approximating the integrals in equations (1–3) is detailed in Appendix A.

For discounting we will use UK government gilt yields, as shown in Figure 1. The broad shape of the yield curve in Figure 1 is as one would expect, namely with short-term yields lower than longer-term ones. However, there is one oddity, namely that yields decline for terms above 24 years.

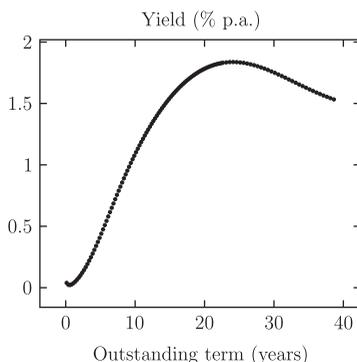


Figure 1. Yields on UK government gilts (coupon strips only, no index-linked gilts) as at 20 April 2017. Source: UK Debt Management Office (DMO, accessed on 21 April 2017)

For $v(t)$ we will follow McCulloch (1971) and McCulloch (1975) and use a spline basis for representing the yields. Note, however, that McCulloch placed his splines with knot points at non-equal distances, whereas we will use equally spaced splines with penalisation as per Eilers and Marx (1996); the plotted points in Figure 1 are sufficiently regular that they look like a smooth curve already, so no distortion is introduced by smoothing. In this paper the P -spline smoothing is applied to the yields directly, rather than to the bond prices as in McCulloch (1971) and McCulloch (1975). The resulting P -spline-smoothed yield curve reproduces all the main features of Figure 1.

3. Model Fitting

We fit models to the data assuming a Poisson distribution for the number of deaths, i.e.

$$D_{x,y} \sim \text{Poisson}(\mu_{x,y} E_{x,y}^c) \quad (4)$$

where $E_{x,y}^c$ denotes the central exposure to risk at age x last birthday in calendar year y . The Poisson assumption means that the variance of the death counts is equal to the mean, which is not true in practice due to over-dispersion. There are several ways of allowing for this over-dispersion: Li *et al.* (2009) proposed a gamma-distributed heterogeneity parameter which varied by age, while Djeundje and Currie (2011) used a single over-dispersion parameter across all ages and years. However, in this paper we will not make an allowance for over-dispersion for simplicity, as we are primarily interested in comparing the models for $\mu_{x,y}$ with each other.

The models we will fit are the following:

$$\text{Age-Period} \quad \log \mu_{x,y} = \alpha_x + \kappa_y \quad (5)$$

$$\text{APC} \quad \log \mu_{x,y} = \alpha_x + \kappa_y + \gamma_{y-x} \quad (6)$$

$$\text{Lee-Carter} \quad \log \mu_{x,y} = \alpha_x + \beta_x \kappa_y \quad (7)$$

$$\text{APCI} \quad \log \mu_{x,y} = \alpha_x + \beta_x (y - \bar{y}) + \kappa_y + \gamma_{y-x} \quad (8)$$

where \bar{y} is the mean over the years 1971–2015. We have selected these models as they are related to each other, but some other models are considered in Appendix F.

Following Brouhns *et al.* (2002) we estimate the parameters using the method of maximum likelihood, rather than the singular-value decomposition of Lee and Carter (1992) or a Bayesian approach. Our focus is on the practical implementation of stochastic models in industry applications, so we estimate $\hat{\kappa}$ and then, for forecasting, fit a variety of models to $\hat{\kappa}$ treating it as an observed time series. This is preferable to re-estimating κ every time we change the model for it, as a fully Bayesian analysis would require. For a discussion of this practical aspect in insurance applications, see Kleinow and Richards (2016).

The Age-Period, APC and APCI models are all linear in the parameters to be estimated, so we will use the algorithm of Currie (2013, pages 87–92) to fit them. The Currie algorithm is a generalisation of the iteratively reweighted least-squares algorithm of Nelder and Wedderburn (1972) used to fit generalised linear models (GLMs), but extended to handle models which have both identifiability

constraints and smoothing via the penalised splines of Eilers and Marx (1996); see Appendix D for an overview. The Lee–Carter model is not linear, but it can be fitted as two alternating linear models as described by Currie (2013, pages 77–80); as with the other three models, constraints and smoothing via penalised splines are applied during the fitting process. Smoothing will be applied to α_x and β_x , but not to κ_y or γ_{y-x} ; smoothing of α_x and β_x reduces the effective number of parameters and improves the quality of the forecasts by reducing the risk of mortality rates crossing over at adjacent ages; see Delwarde *et al.* (2007) and Currie (2013). The fitting algorithm is implemented in R (R Core Team, 2013).

4. Smoothing

An important part of modelling is choosing which parameters to smooth. This is not merely an aesthetic consideration – Delwarde *et al.* (2007) showed how judicious use of smoothing can improve the quality of forecasts, such as by reducing the likelihood of projected mortality rates crossing over at adjacent ages in the future. Figure 2 shows the $\hat{\alpha}_x$ estimates for each of the four models. There is a highly regular, linear-like pattern in each case. We can therefore replace the 55 $\hat{\alpha}_x$ estimates with a smooth curve, or we could even replace them with a straight line for a Gompertz-like version of each model (Gompertz, 1825). This will have the benefit of reducing the effective dimension of the models, i.e., the effective number of parameters. For smoothing we use the penalised splines of Eilers and Marx (1996).

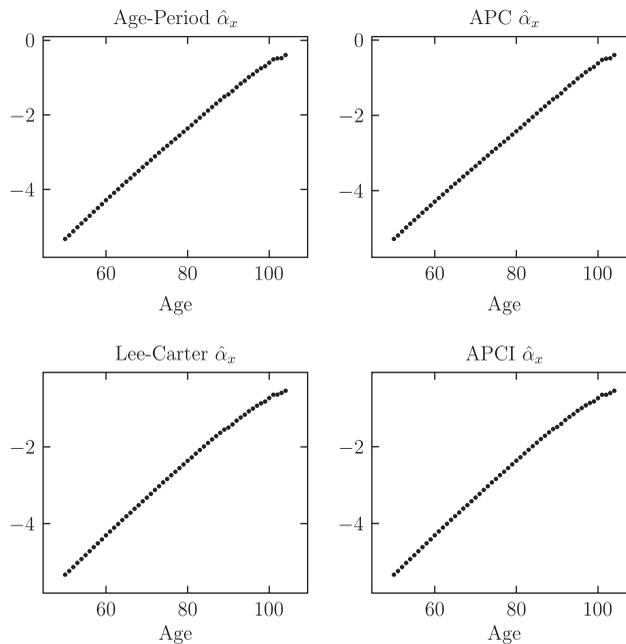


Figure 2. Parameter estimates $\hat{\alpha}_x$ for four unsmoothed models. The α_x parameters play the same role across all four models, i.e., the average log(mortality) value across 1971–2015, when the constraint $\sum_y \kappa_y = 0$ is applied.

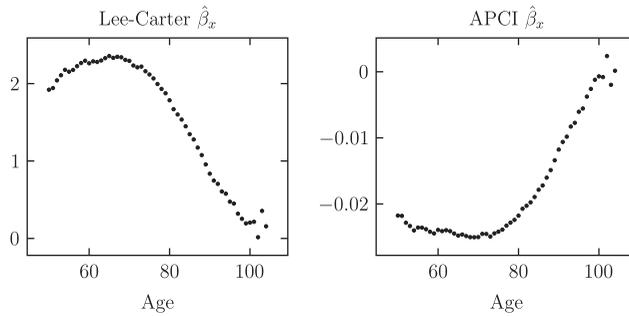


Figure 3. Parameter estimates $\hat{\beta}_x$ for Lee–Carter and APCI models (both unsmoothed). Despite the apparent difference, a switch in sign shows that the β_x parameters play analogous roles in the Lee–Carter and APCI models, namely an age-related modulation of the response in mortality to the time index

Figure 3 shows the $\hat{\beta}_x$ values for the Lee–Carter and APCI models. Although the pattern of the β_x values looks different, this arises from the parameterisation of the models. One view of the APCI model is that it separates the linear component of the time trend in the Lee–Carter model and makes κ the residual, non-linear part of the time trend, i.e.

$$\beta_x \kappa_y = \beta_x (\kappa_y + y - \bar{y} - y + \bar{y}) \tag{9}$$

$$= \beta_x (y - \bar{y}) + \beta_x (\kappa_y - y + \bar{y}) \tag{10}$$

$$\approx \beta_x (y - \bar{y}) + \kappa_y^{APCI} \tag{11}$$

where $\kappa_y^{APCI} \approx \beta_x (\kappa_y - y + \bar{y})$. Now we see why the β_x term has a reverse sign under the APCI model, since in the Lee–Carter model κ_y has a negative slope but in the APCI model $y - \bar{y}$ has a positive slope. It would seem more sensible to have $-\beta_x (y - \bar{y})$ in the APCI model to align the parameterisations, but we will stick with the parameterisation of Continuous Mortality Investigation (2017). Figure 3 shows regularity in the $\hat{\beta}_x$ estimates, albeit not as strong as for the $\hat{\alpha}_x$ estimates in Figure 2. Again, we can replace the fifty-five $\hat{\beta}_x$ estimates with a smooth curve to reduce the effective dimension of the Lee–Carter and APCI models. The greater variability of the $\hat{\beta}_x$ estimates in Figure 3 shows that smoothing here will make an important contribution to reducing the likelihood of mortality rates crossing over in the forecast; smoothing of the $\hat{\beta}_x$ terms for this reason was first proposed by Delwarde *et al.* (2007). Smoothing the $\hat{\beta}_x$ values will therefore both reduce the dimension of the model and improve the forecast quality.

In contrast to Figures 2 and 3, Figures 4 and 5 suggest that smoothing κ and γ is less straightforward. In particular, the $\hat{\kappa}$ estimates for the APCI model in the lower right panel of Figure 4 do not have a clear trend, in which case smoothing $\hat{\kappa}$ under the APCI model would make little sense. In Figure 5 the pattern does not look to be regular and well-behaved enough to warrant smoothing, even though it is technically feasible; the repeated changes in direction suggest that extrapolation would be unwise. In both cases we prefer to leave both $\hat{\kappa}$ and $\hat{\gamma}$ unsmoothed so that we can project them using time-series methods. Note that Richards and Currie (2009) presented a version of the Lee–Carter model with $\hat{\kappa}$ smoothed and thus projected using the penalty function.

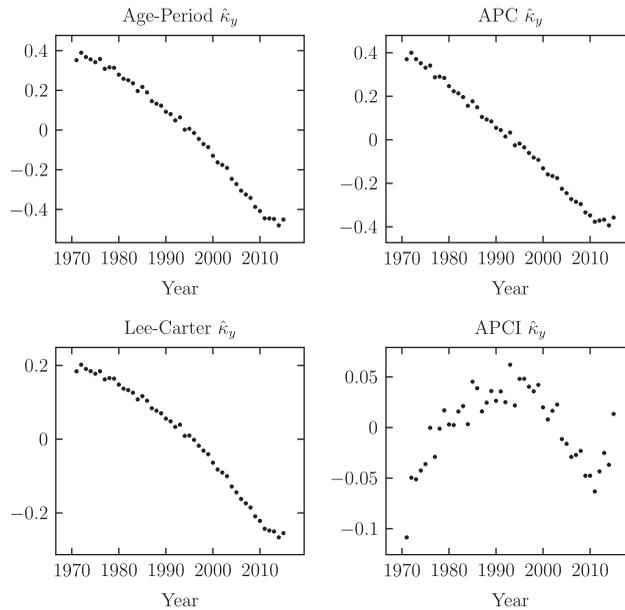


Figure 4. Parameter estimates $\hat{\kappa}_y$ for four unsmoothed models. While κ_y plays a similar role in the Age–Period, Age–Period–Cohort (APC) and Lee–Carter models, it plays a very different role in the APCI model. The APCI $\hat{\kappa}_y$ estimates are an order of magnitude smaller than in the other models, and with no clear trend. In the APCI model κ_y is much more of a residual or left-over term, whose values are therefore strongly influenced by structural decisions made elsewhere in the model

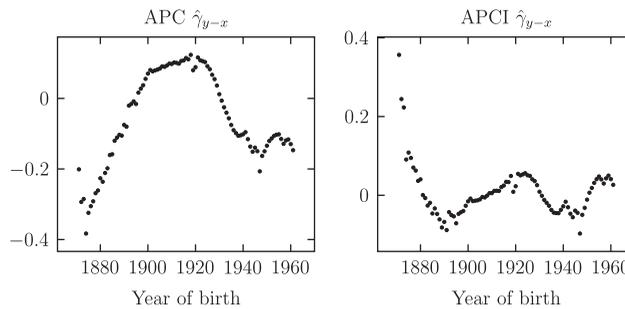


Figure 5. Parameter estimates $\hat{\gamma}_{y-x}$ for Age–Period–Cohort (APC) and APCI models (both unsmoothed). The γ_{y-x} values play analogous roles in the APC and APCI models, yet the values taken and the shapes displayed are very different.

We have omitted plots of the $\hat{\alpha}_x$ and $\hat{\beta}_x$ estimates in the smoothed models as they are just smoothed-curve versions of Figures 2 and 3. Similarly, we have also omitted plots of $\hat{\kappa}_y$ and $\hat{\gamma}_{y-x}$ under the smoothed models, as they are to all practical purposes identical to Figures 4 and 5. This is an important aspect about smoothing – if it is done appropriately, the smoothing of one parameter vector should not make a major impact on any unsmoothed parameter vectors.

Table 1 summarises our approach to smoothing the various parameters across the four models. The impact of the decision to smooth is shown in the contrast between Tables 2 and 3. We can see that

Table 1. Smoothed and Unsmoothed Parameters

Model	Smoothed	Unsmoothed
Age-Period	$\hat{\alpha}_x$	$\hat{\kappa}_y$
APC	$\hat{\alpha}_x$	$\hat{\kappa}_y, \hat{\gamma}_{y-x}$
Lee-Carter	$\hat{\alpha}_x, \hat{\beta}_x$	$\hat{\kappa}_y$
APCI	$\hat{\alpha}_x, \hat{\beta}_x$	$\hat{\kappa}_y, \hat{\gamma}_{y-x}$

Table 2. Expected time lived and annuity factors for unsmoothed models, together with the Bayesian Information Criterion (BIC) and Effective Dimension (ED).

Models	$\bar{e}_{70,2015:35}$	$\bar{a}_{70,2015:35}$	BIC	ED
Age-Period	15.739	13.811	78,667	99.0
APC	15.217	13.510	15,916	188.0
Lee-Carter	15.196	13.531	13,917	153.0
APCI	15.579	13.812	7,140	241.0

The yield curve used to discount future cashflows in the annuity factors is shown in Figure 1.

Table 3. Expected time lived and annuity factors for smoothed (S) models, together with the Bayesian Information Criterion (BIC) and Effective Dimension (ED).

Models	$\bar{e}_{70,2015:35}$	$\bar{a}_{70,2015:35}$	BIC	ED
Age-Period(S)	15.739	13.811	78,527	55.7
APC(S)	16.949	14.701	15,770	144.8
Lee-Carter(S)	15.199	13.534	13,506	63.8
APCI(S)	15.585	13.816	6,724	151.8

The yield curve used to discount future cashflows in the annuity factors is shown in Figure 1.

smoothing has little impact on either the forecast time lived or the annuity factors for the Age-Period, Lee-Carter and APCI models. However, smoothing has led to a major change in the central forecast in the case of the APC model; this is due to a different *autoregressive, integrated moving average* (ARIMA) model being selected as optimal for the κ_y terms: ARIMA(0, 1, 2) for the unsmoothed APC model, but ARIMA(3, 2, 0) for the smoothed version. This large change in forecast is an interesting, if extreme, example of the kind of issues discussed in Kleinow and Richards (2016). An ARIMA(p , 1, q) process models the differences in κ_y , i.e., a model for improvements, whereas an ARIMA(p , 2, q) process models the rate of change in differences in κ_y , i.e., accelerating or decelerating improvements. Smoothing has also improved the fit as measured by the Bayesian information criterion (BIC) – in each case the BIC for a given smoothed model in Table 3 is smaller than the equivalent unsmoothed model in Table 2. This is due to the reduction in the effective number of parameters from the penalisation of the spline coefficients; see equation (21) in Appendix D.

One other interesting aspect of Tables 2 and 3 is the dramatic improvement in overall fit of the APCI model compared to the others. However, it is worth repeating the caution of Currie (2016) that an “oft-overlooked caveat is that it does not follow that an improved fit to data necessarily leads to improved forecasts of mortality”. This was also noted in Kleinow and Richards (2016), where the best-fitting ARIMA process for κ in a Lee–Carter model for UK males led to the greatest parameter uncertainty in the forecast, and thus higher capital requirements under a value-at-risk (VaR) assessment. As we will see in Section 7, although the APCI model fits the data best of the four related models considered, it also produces relatively high capital requirements.

From this point on the models in this paper are smoothed as per Table 1, and the smoothed models will be denoted (S) to distinguish them from the unsmoothed versions.

5. Projections

The κ values will be treated throughout this paper as if they are known quantities, but it is worth noting that this is a simplification. In fact, the $\hat{\kappa}$ estimates have uncertainty over their true underlying value, especially if they are estimated from the mortality experience of a small population; see, for example, Chen *et al.* (2017). The true κ can be regarded as a hidden process, since we cannot observe κ directly and can only infer likely values given the random variation from realised deaths in a finite population. As a result, the estimated variance of the volatility in any ARIMA process fitted to $\hat{\kappa}$ will be an over-estimate, as the estimated $\hat{\kappa}$ values are subject to two sources of variation. There is a parallel here to the concept of a Kalman filter, which models an observable process (the estimated κ) which is itself a realisation of a hidden underlying linear process (the true κ). The Kalman filter therefore allows for two types of noise: measurement error and volatility. In this paper ARIMA models for κ and γ will be estimated using R’s `arima()` function, which uses a Kalman filter to estimate the underlying parameter values, but assuming that there is no measurement error.

As in Li *et al.* (2009) we will adopt a two-stage approach to mortality forecasting: (i) estimation of the time index, κ_y , and (ii) forecasting that time index. The practical benefits of this approach over Bayesian methods, particularly with regards to VaR calculations in life-insurance work, are discussed in Kleinow and Richards (2016). The same approach is used for γ_{y-x} . The details of the ARIMA models used are given in Appendix E.

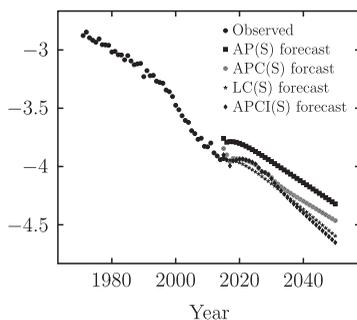


Figure 6. Observed mortality rates at age 70 and projected rates under Age–Period (smoothed) (AP(S)), Age–Period–Cohort (smoothed) (APC(S)), Lee–Carter (smoothed) (LC(S)) and APCI(S) models

Central projections under each of the four models are shown in Figure 6. The discontinuity between observed and forecast rates for the Age-Period(S) model arises from the lack of age-related modulation of the κ_y term – at ages 50–60 there is continuity, at ages 65–75 there is a discontinuity upwards and at ages 85–90 there is a discontinuity downwards. It is for this kind of reason that the Age-Period model is not used in practice for forecasting work.

6. Constraints and Cohort Effects

All four of the models in the main body of this paper require identifiability constraints, and the ones used in this paper are detailed in Appendix C. There is a wide choice of alternative constraint systems. For example, R's `gnm()` function deletes sufficient columns from the model matrix until it is of full rank and the remaining parameters are uniquely estimable and hence identifiable; see Currie (2016). Cairns *et al.* (2009) imposed weighted constraints on the γ_{y-x} parameters that performed a dual purpose: (i) acting as identifiability constraints, and (ii) imposing behaviour on γ_{y-x} to make forecasting assumptions valid. In explaining their choice of identifiability constraints, Cairns *et al.* (2009) stated that their choice ensured that the fitted γ_{y-x} will fluctuate around 0 and will have no discernible linear trend or quadratic curvature.

However, one consequence of the treatment of corner cohorts described in Appendix B is that it reduces the number of constraints required to uniquely identify parameters in the fitting of the APC and APCI models. Following the rationale of Cairns *et al.* (2009) in imposing behaviour on γ_{y-x} , both we and Continuous Mortality Investigation (2016b) use the full set of constraints, meaning that both we and Continuous Mortality Investigation (2016b) are using over-constrained models. For the data set of UK males used in this paper, this policy of over-constraining has a much bigger impact on the shape of the parameter values in APCI model than in the APC model.

The shape of the APC parameters is largely unaffected by over-constraining, as evidenced by Figure 7. However, it is a matter of concern that the values for $\hat{\kappa}_y$ and $\hat{\gamma}_{y-x}$ change so much for the APCI model in Figure 8, at least for this data set and with the choice of cohort constraints. In the case of $\hat{\kappa}_y$, the suspicion is that this term is little more than a residual or left-over in the APCI model. This is a result of the $\beta_x(y-\bar{y})$ term in the APCI model, which picks up the trend that for the other models is present in κ_y .

The changes in κ_y and γ_{y-x} from a minimal constraint system to an over-constrained system depend on the nature of the additional constraints that are applied. In the example in Figure 8 it so happens that the constraints are not consistent with $\gamma=0$ for corner cohorts. A different set of constraints, rather than those outlined in Appendix C, would lead to different estimated parameter values on the left-hand side of Figure 8 and the changes on the right-hand side of Figure 8 might therefore be made greater or smaller.

When we compare the minimal-constraint fits in Figures 7 and 8, we see that for both models γ_{y-x} approximately follows a quadratic function plus some noise. Due to the additional constraint for the APCI model, the quadratic trend is removed in Figure 8, but only the linear trend is removed in Figure 7. It is then unsurprising that the estimated parameter values change more for the APCI model. However, removing the quadratic trend like this has the advantage of making the γ

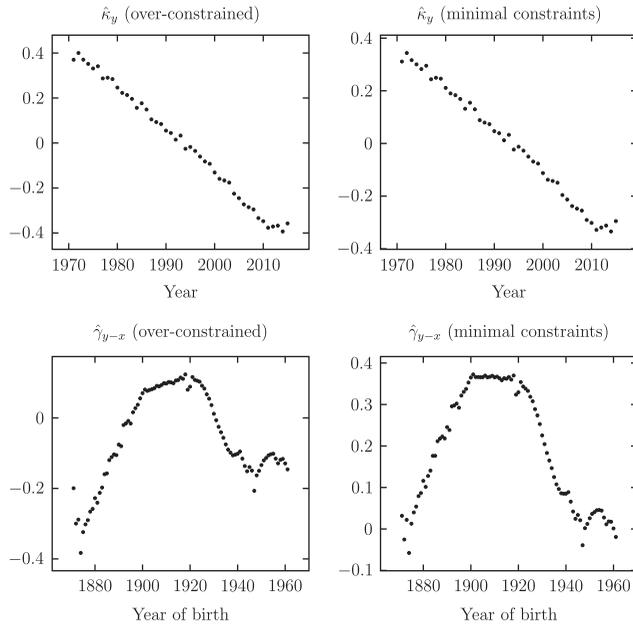


Figure 7. Parameter estimates $\hat{\kappa}_y$ and $\hat{\gamma}_{y-x}$ for the Age-Period-Cohort (smoothed) model: left panels from over-constrained fit, right panels with minimal constraints. The shape of the $\hat{\kappa}_y$ and $\hat{\gamma}_{y-x}$ parameters is largely unaffected by the choice of minimal constraints or over-constraining.

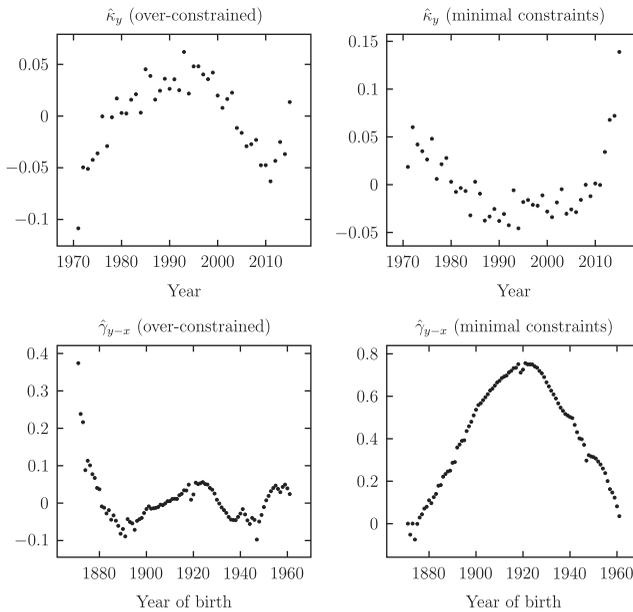


Figure 8. Parameter estimates $\hat{\kappa}_y$ and $\hat{\gamma}_{y-x}$ for APCI(S) model: left panels from over-constrained fit, right panels with minimal constraints. In contrast to Figure 7, the shape of the parameter estimates is heavily affected by the choice to over-constrain the γ_{y-x} parameters.

process in the APCI model look more like a stationary time series, and therefore easier to predict or simulate.

7. VaR Assessment

Insurers in the United Kingdom and European Union are required to use a 1 year, VaR methodology to assess capital requirements for longevity trend risk. Under Solvency II a VaR99.5 value is required, i.e., insurers must hold enough capital to cover 99.5% of possible changes over one year in the financial impact of mortality forecasts. For a set, S , of possible annuity values arising over the coming year, the VaR99.5 capital requirement would be:

$$\frac{q(S, 99.5\%) - q(S, 50\%)}{q(S, 50\%)} \quad (12)$$

where $q(S, \alpha)$ is the α -quantile of the set S , i.e., for any randomly selected value in S , s , we have that $\Pr(s < q(S, \alpha)) = \alpha$. To generate the set S we use the procedure described in Richards *et al.* (2014) to assess long-term longevity trend risk within a one-year framework. The results of this are shown in Table 4.

Table 4 shows VaR99.5 capital requirements at age 70, while Figure 9 shows a wide range of ages. The APCI(S) capital requirements appear less smooth and well-behaved than those for the other models, but the VaR99.5 capital requirements themselves do not appear out of line. We note, however, that the APCI VaR capital requirements exceed the APC(S) and Lee–Carter (S) values at almost every age. How a model's capital requirements vary with age may be an important consideration for life insurers under Solvency II, such as when calculating the risk margin and particularly for closed (and therefore ageing) portfolios.

To understand how the VaR99.5 capital requirements in Table 4 arise, it is instructive to consider the smoothed densities of the annuity factors at age 70 for each model in Figure 10. Here we can see the reason for the higher capital requirement under the APCI model – there is a relatively wider gap between the median and the 99.5% quantile value.

Table 4 shows the impact of model risk in both the median projected annuity factor and the capital requirement. This is a reminder that it is important for practical insurance work to always use a

Table 4. Results of Value-at-Risk Assessment

Models	$\bar{a}_{70,2015:\overline{35} }$		Capital Requirement (%)
	Median	VaR99.5	
AP(S)	13.696	14.263–14.317	4.14–4.54
APC(S)	14.993	15.192–15.316	1.33–2.15
LC(S)	13.447	13.740–13.832	2.18–2.86
APCI(S)	13.692	14.246–14.253	4.05–4.10

The 99.5% quantiles are estimated by applying the estimator from Harrell and Davis (1982) to 5,000 simulations. The ranges given are the 95% confidence intervals computed from the standard error for the Harrell–Davis estimate. The yield curve used to discount future cashflows is shown in Figure 1.

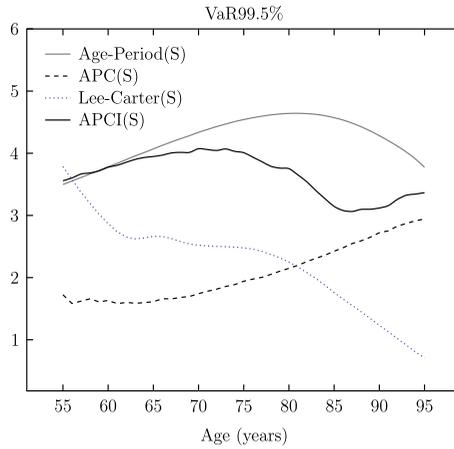


Figure 9. VaR99.5 capital-requirement percentages by age for models in Table 4.

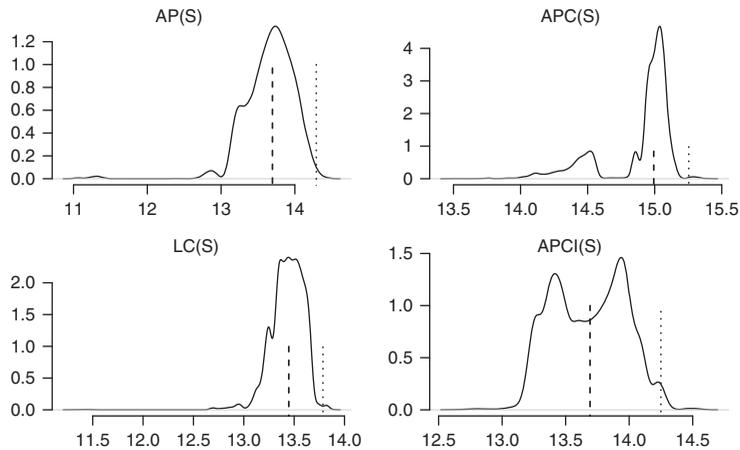


Figure 10. Densities for annuity factors for age 70 from 2015 for 5,000 simulations under the models in Table 4. The dashed vertical lines show the medians and the dotted vertical lines show the Harrell–Davis estimates for the 99.5% quantiles. The shape of the right-hand tail of the APCI (S) model, and the clustering of values far from the median, leads to the higher VaR99.5 capital requirements in Table 4.

variety of models from different families. Indeed, we note that the best estimate under the APC(S) model in Table 4 is higher than the estimated VaR99.5 reserves for the other models, a phenomenon also observed by Richards *et al.* (2014).

8. Conclusions

The APCI model is an interesting addition to the actuarial toolbox. It shares features with the Lee–Carter and APC models and – as with all models – it has its own peculiarities. In the case of the APCI

model, the $\hat{\kappa}_y$ estimates for UK males are heavily dependent on whether the model is over-constrained or not. With minimal constraints the $\hat{\kappa}_y$ estimates in the APCI model for UK male mortality look like a much noisier process than for other models, largely because the linear component of the time trend is accounted for by the $\beta_x(y-\bar{y})$ term. On the face of it this raises questions over whether κ_y should even be kept in the model. However, κ_y captures non-linear curvature in the trend and period deviations – without it the APCI model would have unrealistically low uncertainty over its projections. With minimal constraints it is also trickier to find a forecasting model for the γ process. Either way, with minimal constraints or not, neither κ_y nor γ_{y-x} look like suitable candidates for smoothing.

In the APCI model the $\hat{\gamma}_{y-x}$ estimates change dramatically according to whether the model is over-constrained or not. This is not a feature of the APC model, where the shape of the $\hat{\gamma}_{y-x}$ estimates appear relatively robust to the choice to over-constrain or not, at least for this data set for UK males.

The APCI model fits the data better than the other models considered in this paper, but fit to the data is no guarantee of forecast quality. Interestingly, despite having an improved fit to the data, the APCI model leads to higher capital requirements under a VaR-style assessment of longevity trend risk than most of the other models considered here. These higher requirements vary by age, emphasising that insurers must not only consider multiple models when assessing longevity trend risk, but also the distribution of liabilities by age.

Acknowledgements

The authors thank Alison Yelland, Stuart McDonald, Kevin Armstrong, Yiwei Zhao and an anonymous reviewer for helpful comments. Any errors or omissions remain the responsibility of the authors. All models were fitted using the Projections Toolkit. Graphs were done in R and tikz, while typesetting was done in L^AT_EX. Torsten Kleinow acknowledges financial support from the Actuarial Research Centre (ARC) of the Institute and Faculty of Actuaries through the research programme on Modelling, Measurement and Management of Longevity and Morbidity Risk.

References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317–333.
- Booth, H. & Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science*, **3**(I/II), 3–44.
- Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**(3), 373–393.
- Cairns, A., Blake, D., Dowd, K. & Kessler, A. (2015). Phantoms never die: living with unreliable mortality data. *Journal of the Royal Statistical Society, Series A*, **179**(4), 975–1005.
- Cairns, A.J.G., Blake, D. & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, **73**, 687–718.
- Cairns, A.J.G., 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.
- Chen, L., Cairns, A.J.G. & Kleinow, T. (2017). Small population bias and sampling effects in stochastic mortality modelling. *European Actuarial Journal*, **7**(1), 193–230.

- Continuous Mortality Investigation (2016a). CMI Mortality Projections Model consultation – consultation responses and plans for CMI_2016. Working Paper No. 93.
- Continuous Mortality Investigation (2016b). CMI Mortality Projections Model consultation – technical paper. Working Paper No. 91.
- Continuous Mortality Investigation (2017). CMI Mortality Projections Model: Methods. Working Paper No. 98.
- Currie, I.D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, 13(1), 69–93.
- Currie, I.D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016, 356–383.
- Currie, I.D., Durban, M. & Eilers, P.H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4, 279–298.
- Delwarde, A., Denuit, M. & Eilers, P.H.C. (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized likelihood approach. *Statistical Modelling*, 7, 29–48.
- Diggle, P.J. (1990). *Time Series: A Biostatistical Introduction*. Oxford: Oxford Science Publications.
- Djeundje, V.A.B. & Currie, I.D. (2011). Smoothing dispersed counts with applications to mortality data. *Annals of Actuarial Science*, 5(I), 33–52.
- Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.
- Giroi, F. & King, G. (2008). *Demographic Forecasting*. Princeton, New Jersey: Princeton University Press.
- Gompertz, B. (1825). The nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society*, 115, 513–585.
- Harrell, F.E. & Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika*, 69, 635–640.
- Harvey, A.C. (1981). *Time Series Models*. Oxford: Philip Allan.
- Harvey, A.C. (1996). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–318.
- Kleinow, T. & Richards, S.J. (2016). Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, 2016(10), 1–25.
- Lee, R.D. & Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87, 659–671.
- Li, J.S.-H., Hardy, M.R. & Tan, K.S. (2009). Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. *ASTIN Bulletin*, 39(1), 137–164.
- McCullagh, P. & Nelder, J.A. (1989). Generalized linear models in Editors D. R. Cox, D. V. Hinkley, N. Reid, D. B. Rubin and B. W. Silverman, *Monographs on Statistics and Applied Probability* (2nd ed.), Volume 37. Chapman and Hall, London.
- McCulloch, J.H. (1971). Measuring the term structure of interest rates. *The Journal of Business*, 44, 19–31.
- McCulloch, J.H. (1975). The tax-adjusted yield curve. *The Journal of Finance*, XXX(3), 811–830.
- Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135(Pt 3), 370–384.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (2005). *Numerical Recipes in C++: The Art of Scientific Computing*. New York: Cambridge University Press.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing.

- Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**, 556–570.
- Richards, S.J. (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal*, **14**(II), 257–326 (with discussion).
- Richards, S.J. & Currie, I.D. (2009). Longevity risk and annuity pricing with the Lee-Carter model. *British Actuarial Journal*, **15**(II) 65), 317–365 (with discussion).
- Richards, S.J., Currie, I.D. & Ritchie, G.P. (2014). A value-at-risk framework for longevity trend risk. *British Actuarial Journal*, **19**(1), 116–167.
- Richards, S.J., Kirkby, J.G. & Currie, I.D. (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, **12**(I), 5–61 (with discussion).
- Solo, V. (1984). The order of differencing in ARIMA models. *Journal of the American Statistical Association*, **79**(388), 916–921.

Appendices

A. Integration

We need to evaluate the integrals in equations (1–3). There are several approaches which could be adopted when the function to be integrated can be evaluated at any point, such as adaptive quadrature; see Press *et al.* (2005, page 133) for details of this and other methods. However, since we only have data at integer ages, ${}_t p_{x,y}$ can only be calculated at equally spaced grid points. Since we cannot evaluate the function to be integrated at any point we like, we maximise our accuracy by using the following approximations.

For two points separated by one year we use the Trapezoidal Rule:

$$\int_a^{a+1} f(x) \approx \frac{1}{2} [f(a) + f(a+1)] \quad (13)$$

For three points spaced 1 year apart we use Simpson's Rule:

$$\int_a^{a+2} f(x) \approx \frac{1}{3} [f(a) + 4f(a+1) + f(a+2)] \quad (14)$$

For four points spaced one year apart we use Simpson's 3/8 Rule:

$$\int_a^{a+3} f(x) \approx \frac{3}{8} [f(a) + 3f(a+1) + 3f(a+2) + f(a+3)] \quad (15)$$

For five points spaced 1 year apart we use Boole's Rule:

$$\int_a^{a+4} f(x) \approx \frac{2}{45} [7f(a) + 32f(a+1) + 12f(a+2) + 32f(a+3) + 7f(a+4)] \quad (16)$$

To integrate over n equally spaced grid points we first apply Boole's Rule as many times as possible, then Simpson's 3/8 Rule, then Simpson's Rule and then the Trapezoidal Rule for any remaining points at the highest ages.

B. Corner Cohorts

One issue with the APC and APCI models is that the cohort terms can have widely varying numbers of observations, as illustrated in Figure A1; at the extremes, the oldest and youngest cohorts have just

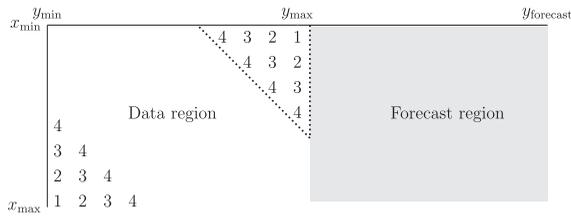


Figure A1. Number of observations for each cohort in the data region

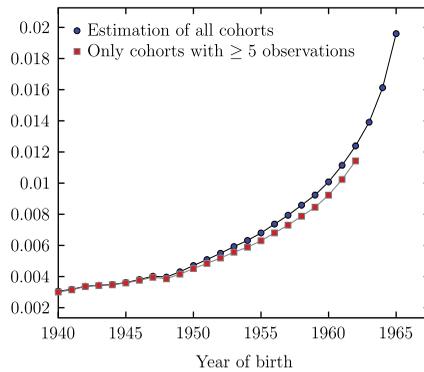


Figure A2. Standard errors of $\hat{\gamma}_{y-x}$ for APCI(S) model with and without estimation of corner cohorts. $\hat{\gamma}_{y-x}$ terms for the cohort cohorts in Figure A1 have very high standard errors; not estimating them has the additional benefit of stabilising the standard errors of those cohort terms we do estimate.

a single observation each. A direct consequence of this limited data is that any estimated γ term for the corner cohorts will have a very high variance, as shown in Figure A2. Cairns *et al.* (2009) dealt with this by simply discarding the data in the triangles in Figure A1, i.e., where a cohort had four or fewer observations. Instead of the oldest cohort having year of birth $y_{\min} - x_{\max}$, for example, it becomes $c_{\min} = y_{\min} - x_{\max} + 4$. Similarly, the youngest cohort has year of birth $c_{\max} = y_{\max} - x_{\min} - 4$ instead of $y_{\max} - x_{\min}$.

There is a drawback to the approach of Cairns *et al.* (2009), namely it makes it harder to compare model fits. We typically use an information criterion to compare models, such as the AIC or BIC. However, this is only valid where the data used are the same. If two models use different data, then their information criteria cannot be compared. This would be a problem for comparing the models in Tables 2, 3 and 12, for example, as the fit for an APC or APCI model could not be compared with the fits for the Age–Period and Lee–Carter models if corner cohorts were only dropped for some models. One approach would be to make the data the same by dropping the corner cohorts for the Age–Period and Lee–Carter fits, even though this is technically unnecessary. This sort of thing is far from ideal, however, as it involves throwing away data and would have to be applied to all sorts of other non-cohort-containing models.

An alternative approach is to use all the data, but to simply not fit cohort terms in the corners of Figure 1. This preserves the easy comparability of information criteria between different model fits. To avoid fitting cohort terms where they are too volatile we simply assume a value of $\gamma = 0$ where there are four or fewer observations. This means that the same data are used for models with and

without cohort terms, and thus that model fits can be directly compared via the BIC. Currie (2013) noted that this had the beneficial side effect of stabilising the variance of the cohort terms which are estimated, as shown in Figure A2.

For projections of γ we forecast not only for the unobserved cohorts, but also for the cohorts with too few observations, i.e., the cohorts in the dotted triangle in Figure A1.

C. Identifiability Constraints

The models in equations (5)–(8) all require identifiability constraints. For the Age–Period model we require one constraint, and we will use the following:

$$\sum_y \kappa_y = 0 \tag{17}$$

For the Lee–Carter model we require two constraints. For one of them we will use the same constraint as equation (17), together with the usual constraint on β_x from Lee and Carter (1992):

$$\sum_x \beta_x = 1 \tag{18}$$

There are numerous alternative constraint systems for the Lee–Carter model – see Girosi and King (2008), Renshaw and Haberman (2006) and Richards and Currie (2009) for examples. The choice of constraint system will affect the estimated parameter values, but will not change the fitted values of $\hat{\mu}_{x,y}$.

For the APC model we require three constraints. For the first one we will use the same constraint as equation (17), together with the following two from Cairns *et al.* (2009):

$$\sum_c w_c \gamma_c = 0 \tag{19}$$

$$\sum_c w_c (c - c_{\min} + 1) \gamma_c = 0 \tag{20}$$

where w_c is the number of times cohort c appears in the data. Continuous Mortality Investigation (2016b) uses unweighted cohort constraints, i.e., $w_c = 1, \forall c$, but we prefer to use the constraints of Cairns *et al.* (2009), as they give less weight to years of birth with less data.

For the APCI model we require five constraints. We will use equations (17), (19) and (20), together with the following additional two:

$$\sum_y (y - y_{\min}) \kappa_y = 0 \tag{21}$$

$$\sum_c w_c (c - c_{\min} + 1)^2 \gamma_c = 0 \tag{22}$$

where equation (22) is the continuation of the pattern in equations (19) and (20) established by Cairns *et al.* (2009).

The number of constraints necessary for a linear model can be determined from the rank of the model matrix. Note that the approach of not fitting γ terms for cohorts with four or fewer observations, as

outlined in Appendix B, makes the constraints involving γ unnecessary for identifiability. As in Continuous Mortality Investigation (2016b), this means that the APC and APCI models in this paper are *over-constrained*, and will thus usually produce poorer fits than would be expected if a minimal constraint system were adopted. Over-constraining has a different impact on the two models: for the APC model it leads to relatively little change in κ , as shown in Figure 7. However, for the APCI model κ is little more than a noise process in the minimally constrained model (see Figure 8), while any pattern in κ from the over-constrained model appears likely to have been caused by the constraints on γ_{y-x} .

D. Fitting Penalised Constrained Linear Models

The Age–Period, APC and APCI models in equations (5), (6) and (8) are Generalized Linear Models (GLMs) with identifiability constraints. We smooth the parameters as described in Table 1. We accomplish the parameter estimation, constraint application and smoothing simultaneously using the algorithm presented in Currie (2013). In this section, we outline the three development stages leading up to this algorithm.

Nelder and Wedderburn (1972) defined the concept of a GLM. At its core we have the *linear predictor*, η , defined as follows:

$$\eta = X\theta \quad (23)$$

where X is the *model matrix* or *design matrix* and θ is the vector of parameters in the model. For the model to be identifiable we require that the rank of X equals the length of θ ; the model of Cairns *et al.* (2006) is just such a mortality model (also referred to as M5 in Cairns *et al.*, 2009). Nelder and Wedderburn (1972) presented an algorithm of *iteratively weighted least squares* (IWLS), the details of which vary slightly according to (i) the assumption for the distribution of deaths and (ii) the link function connecting the linear predictor to the mean of that distribution. This algorithm finds the values, $\hat{\theta}$, which jointly maximise the (log-)likelihood.

X can also contain basis splines, which introduces the concept of smoothing and penalisation into the GLM framework; see Eilers and Marx (1996). Currie (2013) extended the IWLS algorithm to find the values, $\hat{\theta}$, which jointly maximise the *penalised* likelihood for some given value of the smoothing parameter, λ . The optimum value of λ is determined outside the likelihood framework by minimising an information criterion, such as the BIC:

$$\text{BIC} = \text{Dev} + \log(n)\text{ED} \quad (24)$$

where n is the number of observations, Dev is the *model deviance* (McCullagh and Nelder, 1989) and ED is the *effective dimension* of the model (Hastie and Tibshirani, 1986). In the single-dimensional case, as λ increases so does the degree of penalisation. The penalised parameters therefore become less free to take values different from their neighbours. The result of increasing λ is therefore to reduce the effective dimension of the model, and so equation (24) balances goodness of fit (measured by the deviance, Dev) against the smoothness of the penalised coefficients (measured via the effective dimension, ED). Currie *et al.* (2004) and Richards *et al.* (2006) used such penalised GLMs to fit smooth, two-dimensional surfaces to mortality grids.

We note that penalisation is applied to parameters which exhibit a smooth and continuous progression, such as the α_x parameters in equations (5)–(8). If a second-order penalty is applied, as $\lambda \rightarrow \infty$ the smooth curve linking the parameters becomes ever more like a simple straight line, i.e., the effective dimension of α_x would tend to ED = 2. Alternatively, the α_x could be replaced with two

parameters for a simple straight-line function of age. In the case of equations (5)–(8) this would simplify the models to variants of the Gompertz model of mortality (Gompertz, 1825).

Many linear mortality models also require identifiability constraints, i.e., the rank of the model matrix is less than the number of parameters to be estimated. The Age–Period, APC and APCI models of the main body of this paper fall into this category: they are all linear, but in each case $\text{rank}(X) < \text{length}(\theta)$. The gap between $\text{rank}(X)$ and $\text{length}(\theta)$ determines the number of identifiability constraints required. To enable simultaneous parameter estimation, smoothing and application of constraints, Currie (2013) extended the concept of the model matrix, X , to the *augmented model matrix*, X_{aug} , defined as follows:

$$X_{\text{aug}} = \begin{pmatrix} X \\ H \end{pmatrix} \tag{25}$$

where H is the *constraint matrix* with the same number of columns as X and where each row of H corresponds to one linear constraint. If $\text{rank}(X_{\text{aug}}) = \text{length}(\theta)$, the model is identifiable. If $\text{rank}(X_{\text{aug}}) > \text{length}(\theta)$, then the model is *over-constrained*; see Appendix C. Note that the use of the augmented model matrix, X_{aug} , here restricts H to containing linear constraints.

In this paper we use a Poisson distribution and a log link for our GLMs; this is the *canonical link function* for the Poisson distribution. This means that the fitted number of deaths is the anti-log of the linear predictor, i.e., $E^c \times e^\eta$. However, Currie (2014) noted that a logit link often provides a better fit to population data. This would make the fitted number of deaths a logistic function of the linear predictor, i.e., $E^c \times e^\eta / (1 + e^\eta)$. If the logistic link is combined with the straight-line assumption for α_x in equations (5)–(8), this would simplify the models to variants of the Perks model of mortality; see Richards (2008). Currie (2016; Appendix 1) provides R code to implement the logit link for the Poisson distribution for the number of deaths in a GLM. From experience we further suggest specifying good initial parameter estimates to R’s `glm()` function when using the logit link, as otherwise there can be problems due to very low exposures at advanced ages. The `start` option in the `glm()` function can be used for this. In Appendix F we use a logit link to make a M5 Perks model as an alternative to the M5 Gompertz variant using the log link. As can be seen in Table A8, the M5 Perks model fits the data markedly better than the other M5 variants.

E. Projecting κ and γ

A time series is a sequence of elements ordered by the time at which they occur; stationarity is a key concept. Informally, a time series $\{Y(t)\}$ is stationary if $\{Y(t)\}$ looks the same at whatever point in time we begin to observe it – see Diggle (1990, page 13). Usually we make do with the simpler second-order stationarity, which involves the mean and autocovariance of the time series. Let:

$$\mu(t) = E[Y(t)] \tag{26}$$

$$\text{Cov}(t, s) = E[\{Y(t) - \mu(t)\}\{Y(s) - \mu(s)\}] \tag{27}$$

be the mean and autocovariance function of the time series. Then the time series is second-order stationary if:

$$\mu(t) = \mu, \forall t \tag{28}$$

$$\text{Cov}(t, s) = \text{Cov}(|t-s|), \text{ i.e. Cov}(t, s) \text{ depends only on } |t-s| \quad (29)$$

that is, the covariance between $Y(t)$ and $Y(s)$ depends only on their separation in time; see Diggle (1990, page 58). In practice, when we say a time series is stationary we mean the series is second-order stationary. The assumption of stationarity of the first two moments only is variously known as weak-sense stationarity, wide-sense stationarity or covariance stationarity

The lag operator, L , operates on an element of a time series to produce the previous element. Thus, if we define a collection of time-indexed values $\{\kappa_t\}$, then $L\kappa_t = \kappa_{t-1}$. Powers of L mean the operator is repeatedly applied, i.e., $L^i\kappa_t = \kappa_{t-i}$. The lag operator is also known as the backshift operator, while the difference operator, Δ , is $1 - L$.

A time series, κ_t , is said to be *integrated* if the differences of order d are stationary, i.e., $(1 - L)^d\kappa_t$ is stationary.

A time series, κ_t , is said to be *autoregressive* of order p if it involves a linear combination of the previous values, i.e. $(1 - \sum_{i=1}^p ar_i L^i)\kappa_t$, where ar_i denotes an autoregressive parameter to be estimated. An AR process is stationary if the so-called characteristic polynomial of the process has no unit roots; see Harvey (1996). For an AR(1) process this is the case if $-1 < ar_1 < 1$. For empirically observed time series stationary can be tested using unit-root tests.

A time series, κ_t , is said to be a *moving average* of order q if the current value can be expressed as a linear combination of the past q error terms, i.e. $(1 + \sum_{i=1}^q ma_i L^i)\epsilon_t$, where ma_i denotes a moving-average parameter to be estimated and $\{\epsilon_t\}$ is a sequence of independent, identically distributed error terms with zero mean and common variance, σ_ϵ^2 . A moving-average process is always stationary.

A time series, κ_t , can be modelled combining these three elements as an ARIMA model (Harvey, 1981) as follows:

$$\left(1 - \sum_{i=1}^p ar_i L^i\right)(1-L)^d \kappa_t = \left(1 + \sum_{i=1}^q ma_i L^i\right) \epsilon_t \quad (30)$$

An ARIMA model can be structured with or without a mean value. The latter is simply saying the mean value is set at 0. The behaviour and interpretation of this mean value is dependent on the degree of differencing, i.e., the value of d in ARIMA(p, d, q).

For the Age-Period, APC and Lee-Carter models (but not the APCI model), an ARIMA model for κ with $d=1$ is broadly modelling mortality improvements, i.e., $\kappa_{t+1} - \kappa_t$. It will be appropriate where the rate of mortality improvement has been approximately constant over time, i.e., without pronounced acceleration or deceleration. An ARIMA model with $d=1$ but no mean will project gradually decelerating improvements. An ARIMA model with $d=1$ and a fitted mean will project improvements which will gradually tend to that mean value. In most applications the rate at which the long-term mean is achieved is very slow and the curvature in projected values is slight. However, there are two exceptions to this:

- Pure moving-average models, i.e., ARIMA(0, d, q) models. With such models the long-term mean will be achieved quickly, i.e., after $q + d$ years.

- ARIMA models where the autoregressive component is weak. For example, an ARIMA(1, d , q) model where the ar_1 parameter is closer to 0 will also converge to the long-term mean relatively quickly, with the speed of convergence inversely proportional to the absolute value of the ar_1 parameter.

For the Age–Period, APC and Lee–Carter models (but not the APCI model), an ARIMA model for κ with $d=2$ is broadly modelling the rate of change in mortality improvements, not the improvements themselves. Thus, with $d=2$ we are modelling $(\kappa_{t+2} - \kappa_{t+1}) - (\kappa_{t+1} - \kappa_t)$. Such a model will be appropriate where the rate of mortality improvement has been accelerating or decelerating over time. An ARIMA model with $d=2$ and without a mean will project a gradual deceleration of the rate of change in mortality improvements.

To project κ and/or γ in each of the models in the paper, we fit an ARIMA model. We fit ARIMA models with a mean for κ in the Age–Period, APC and Lee–Carter models. We fit ARIMA models without a mean for γ in the APC and APCI models, and also for κ in the APCI model.

The ARIMA parameters, including the mean where required, are estimated using R's `arima()`, which estimates ARIMA parameters assuming that κ_y and γ_{y-x} are known quantities, rather than the estimated quantities that they really are.

While R's `arima()` function returns standard errors, for assessing parameter risk we use the methodology outlined in Kleinow and Richards (2016). The reason for this is that sometimes ARIMA parameter estimates can be borderline unstable, and this can lead to wider confidence intervals for the best-fitting model, as shown in Kleinow and Richards (2016).

To fit an ARIMA model we require to specify the autoregressive order (p), the order of differencing (d) and the order of the moving average (q). For a given level of differencing we fit an ARMA(p , q) model and choose the value of p and q by comparing an information criterion; in this paper we used Akaike's Information Criterion (Akaike, 1987) with a small-sample correction (AICc). Choosing the order of differencing, d , is trickier, as the data used to fit the ARMA(p , q) model are different when $d=1$ and $d=2$: with n observations there are $n-1$ first differences, but only $n-2$ second differences. To decide on the ARIMA(p , d , q) model we select the best ARMA(p , q) model for a given value of d using the AICc, then we pick the ARIMA(p , d , q) model with the smallest root mean squared error as per Solo (1984).

The choice of differencing order is thorny: with $d=1$ we are modelling mortality improvements, but with $d=2$ we are modelling the rate of change of mortality improvements. The latter can produce

Table A1. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(1,1,2) process for κ in smoothed Age–Period model

Parameters	Estimate	Standard Error
ar_1	0.742	0.262
ma_1	-1.366	0.350
ma_2	0.846	0.255
Mean	-0.018	0.004

The ARIMA process order was selected as the best-fitting one from ARIMA(0:3,1:2,0:3). $\hat{\sigma}^2 = 0.000302$.

Table A2. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(1,1,2) process for κ in smoothed Lee–Carter model

Parameters	Estimate	Standard Error
ar ₁	0.821	0.215
ma ₁	-1.306	0.315
ma ₂	0.719	0.264
Mean	-0.010	0.002

The ARIMA process order was selected as the best-fitting one from ARIMA(0:3,1:2,0:3). $\hat{\sigma}^2 = 0.000073$.

Table A3. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(0,1,2) process for κ in smoothed APC model

Parameters	Estimate	Standard Error
ma ₁	-0.682	0.151
ma ₂	0.353	0.225
Mean	-0.017	0.002

The ARIMA process order was selected as best-fitting combination from ARIMA(0:3,1:2,0:3). $\hat{\sigma}^2 = 0.000468$.

Table A4. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(2,1,0) process for γ in smoothed APC model

Parameters	Estimate	Standard Error
ar ₁	-0.037	0.079
ar ₂	0.438	0.098

The ARIMA process order was selected as the best-fitting one from ARIMA(0:3,1:1,0:3). $\hat{\sigma}^2 = 0.000468$.

very different forecasts, as evidenced by comparing the life expectancy for the APC(S) model in Table 3 (with $d=2$) with the life expectancy for the APC model in Table 2 (with $d=1$).

For a VaR assessment of in-force annuities we need to simulate sample paths for κ . If we want mortality rates in the upper right triangle of Figure A1, then we also need to simulate sample paths for γ . We use the formulae given in Kleinow & Richards (2016) for bootstrapping the mean (for κ only) and then use these bootstrapped parameter values for the ARIMA process to include parameter risk in the VaR assessment (Tables 5–10).

F. Other Models

In their presentation of a VaR framework for longevity trend risk, Richards *et al.* (2014) included some other models not considered in the main body of this paper. For interest we present comparison figures for members of the Cairns–Blake–Dowd family of stochastic projection models. We first consider a model sub-family based on Cairns *et al.* (2006) (M5) as follows (Table A6):

Table A5. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(1,1,2) process for κ in smoothed APCI model

Parameters	Estimate	Standard Error
ar ₁	0.754	0.255
ma ₁	-1.375	0.346
ma ₂	0.860	0.278

The ARIMA process order was selected as best-fitting combination from ARIMA(0:3,1:2,0:3). $\hat{\sigma}^2 = 0.000514$.

Table A6. Parameters for Autoregressive, Integrated Moving Average (ARIMA)(2,1,0) process for γ in smoothed APCI model

Parameters	Estimate	Standard Error
ar ₁	0.057	0.056
ar ₂	0.536	0.078

The ARIMA process order was selected as the best-fitting one from ARIMA(0:3,1:1,0:3). $\hat{\sigma}^2 = 0.000514$.

Table A7. Definition of M5 Family Under Equation (31)

Models	$g(\mu_{x,y})$	$w(x)$
M5 Gompertz	log	$x - \bar{x}$
M5 Perks	logit	$x - \bar{x}$
M5 P-spline	log	$\sum_j \theta_j B_j(x)$

\bar{x} Represents the mid-point of the age range, $B_j(x)$ is the j th spline evaluated at x and θ_j the corresponding spline coefficient. Note that the M5 Gompertz and Perks models do not require any identifiability constraints, but the M5 P-spline model needs two.

$$g(\mu_{x,y}) = \kappa_{0,y} + w(x)\kappa_{1,y} \tag{31}$$

for some functions $g()$ and $w()$ where κ_0 and κ_1 form a bivariate random walk with drift. The three members of the M5 family used here are defined in Table A7, with the results in Tables A8 and A9. We also consider two further models from Cairns *et al.* (2009). First, M6:

$$\log \mu_{x,y} = \kappa_{0,y} + (x - \bar{x})\kappa_{1,y} + \gamma_{y-x} \tag{32}$$

Model M6 in equation (32) needs two identifiability constraints and we use equations (19) and (20). As with the M5 family, κ_0 and κ_1 form a bivariate random walk with drift and γ is projected using an ARIMA model (as done for the APC and APCI models). We also consider M7 from Cairns *et al.* (2009):

$$\log \mu_{x,y} = \kappa_{0,y} + (x - \bar{x})\kappa_{1,y} + \left((x - \bar{x})^2 - \hat{\sigma}^2 \right) \kappa_{2,y} + \gamma_{y-x} \tag{33}$$

where $\hat{\sigma}^2 = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$. Model M7 in equation (33) needs three identifiability constraints and

Table A8. Expected time lived and annuity factors for unsmoothed models, together with Bayesian Information Criterion (BIC) and Effective Dimension (ED)

Models	$\bar{e}_{70,2015: > 35}$	$\bar{a}_{70,2015: > 35}$	BIC	ED
M5 Gompertz	15.519	13.734	46,597	90.0
M5 Perks	15.500	13.702	28,950	90.0
M5 P-spline	15.423	13.646	33,428	99.7
M6	16.960	14.730	7,969	179.0
M7	15.514	13.763	7,956	223.0

The yield curve used to discount future cashflows is shown in Figure 1.

Table A9. Results of Value-at-Risk assessment for models in Table A8

Models	$\bar{a}_{70,2015: > 35}$:		Capital Requirement (%)
	Median	VaR99.5	
M5 Gompertz	13.836	14.268–14.328	3.13–3.55
M5 Perks	13.805	14.278–14.345	3.42–3.91
M5 P-spline	13.747	14.192–14.250	3.24–3.66
M6	14.925	15.418–15.481	3.31–3.72
M7	13.867	14.277–14.336	2.95–3.39

The 99.5% quantiles are estimated by applying the estimator from Harrell and Davis (1982) to 5,000 simulations. The ranges given are the 95% confidence intervals computed from the standard error for the Harrell–Davis estimate. The yield curve used to discount future cashflows is shown in Figure 1.

we use equations (19), (20) and (22). κ_0 , κ_1 and κ_2 form a trivariate random walk with drift and γ is projected using an ARIMA model (as done for the APC and APCI models). As with the APC and APCI models, M6 and M7 do not need all these constraints with our treatment of corner cohorts described in Appendix B. Thus, M6 and M7 here are also over-constrained.

Comparing Table A8 with Tables 2 and 3 we can see that the stochastic version of the APCI model produces similar expected time lived and temporary annuity factors to most models, apart from the APC and M6 models. This suggests that the best-estimate forecasts under the APCI model are consistent and not extreme.

Comparing Table A9 with Table 4 we can see that, while the AP(S) and APCI(S) models produce the largest VaR99.5 capital requirements at age 70, these are not extreme outliers.

A comparison of Table 4 with the equivalent figures in Richards *et al.* (2014, Table 4) shows considerable differences in VaR99.5 capital at age 70. There are two changes between Richards *et al.* (2014) and this paper that drive these differences. The first change is that Richards *et al.* (2014) discounted cashflows using a flat 3% per annum, whereas in this paper we discount cashflows using the yield curve in Figure 1. The second change lies in the data: in this paper we use UK-wide data for 1971–2015, whereas Richards *et al.* (2014) used England and Wales data for 1961–2010. There are three important sub-sources of variation buried in this change in the data: the first is that the population estimates for 1961–1970 are not as reliable as the estimates which came after 1970; the

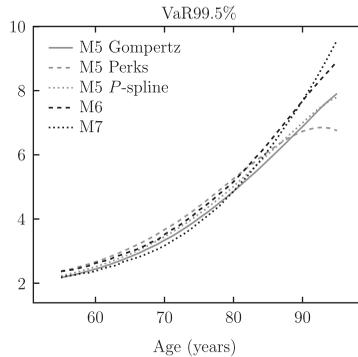


Figure A3. VaR99.5 capital requirements by age for models in Table A8

second is that the data used in this paper include revisions to pre-2011 population estimates following the 2011 census; and the third is that mortality experience after 2010 has been unusual and is not in line with trend. The combined effect of these changes to the discount function and the data has led to the VaR99.5 capital requirements at age 70 for the models in Table A9 being around 0.5% less than for the same models in Richards *et al.* (2014, Table 4). However, a comparison between Figures 9 and A3 shows that these results are strongly dependent on age. As in Richards *et al.* (2014), this means that it is insufficient to consider a few model points for a VaR assessment – insurer capital requirements not only need to be informed by different projection models, but they must take account of the age distribution of liabilities.

G. Differences compared to Continuous Mortality Investigation approach

In this paper we present a stochastic implementation of the APCI model proposed by Continuous Mortality Investigation (2016b). This is the central difference between the APCI model in this paper and its original implementation in Continuous Mortality Investigation (2016a, 2016b). However, there are some other differences of note and they are listed in this section as a convenient overview.

As per Cairns *et al.* (2009) our identifiability constraints for γ_{y-x} weight each parameter according to the number of times it appears in the data, rather than assuming equal weight as in Continuous Mortality Investigation (2016b, page 91). As with Continuous Mortality Investigation (2016b) our APC and APCI models are over-constrained (see Appendix C and section 6).

For cohorts with four or fewer observed values we do not estimate a γ term – see Appendix B. In contrast, Continuous Mortality Investigation (2016a, pages 27–28) adopts a more complex approach to corner cohorts, involving setting the cohort term to the nearest available estimated term.

For smoothing α_x and β_x we have used the penalised splines of Eilers and Marx (1996), rather than the difference penalties in Continuous Mortality Investigation (2016b). Our penalties on α_x and β_x are quadratic, whereas Continuous Mortality Investigation (2016b) uses cubic penalties. Unlike Continuous Mortality Investigation (2016b) we do not smooth κ_y or γ_{y-x} . We also determine the optimal level of smoothing by minimising the BIC, whereas Continuous Mortality Investigation (2016b) smooths by user judgement.

As described in Section 3, for parameter estimation we use the algorithm presented in Currie (2013). This means that constraints and smoothing are an integral part of the estimation, rather than separate steps applied in Continuous Mortality Investigation (2016b, page 15).

Unlike Continuous Mortality Investigation (2016b) we make no attempt to adjust the exposure data.

For projections we use ARIMA models for both κ_y and γ_{y-x} , rather than the deterministic targeting approach of Continuous Mortality Investigation (2016b, pages 31–35). Unlike Continuous Mortality Investigation (2016b) we do not attempt to break down mortality improvements into age, period and cohort components, nor do we have a long-term rate to target and nor do we have any concept of a “direction of travel”(Continuous Mortality Investigation, 2016b, page 14).

H. Suggestions for Further Research

There were many other things which could have been done in this paper, but for which there was not the time available. We list some of them here in case others are interested in doing so:

- Female lives. To illustrate our points, and to provide comparable figures to earlier papers such as Richards *et al.* (2014) and Kleinow and Richards (2016), we used the data for males. However, both insurers and pension schemes have material liabilities linked to female lives, and it would be interesting to explore the application of the APCI model to data on female lives.
- Back-testing. It would be interesting to see how the APCI model performs against other models in back-testing, i.e., fit the models to first part of the data set and see how the resulting forecasts compare to the latter part of the data.
- Sensitivity testing. Some models are sensitive to the range of ages selected or the period covered. It would be interesting to know how sensitive the APCI model is to such changes.
- Canonical correlation. Models with both period and cohort terms, such as the APC and APCI models, usually have these terms projected as if they are independent. However, such terms are usually correlated, making the assumption of independence at best a simplifying assumption for convenience. It would be interesting to compare the correlations of κ and γ for the APC and APCI models. Joint models for κ and γ could be considered.
- Over-dispersion. To fit the models, both we and Continuous Mortality Investigation (2017, page 5) assume that the number of deaths follows a Poisson distribution, i.e., that the variance is equal to the mean. However, in practice death counts are usually over-dispersed, i.e., the variance is greater than the mean.