*British Actuarial Journal* (2024), Vol. **29**, e19, pp. 1–23
doi:10.1017/S1357321724000175

Institute and Faculty of Actuaries

SESSIONAL PAPER

# Robust mortality forecasting in the presence of outliers

Stephen J. Richards

Longevitas Ltd., Edinburgh EH6 3AJ, UK
Email: stephen@longevitas.co.uk

**Abstract**

Stochastic mortality models are important for a variety of actuarial tasks, from best-estimate forecasting to assessment of risk capital requirements. However, the mortality shock associated with the Covid-19 pandemic of 2020 distorts forecasts by (i) biasing parameter estimates, (ii) biasing starting points, and (iii) inflating variance. Stochastic mortality models therefore require outlier-robust methods for forecasting. Objective methods are required, as outliers are not always obvious on visual inspection. In this paper we look at the robustification of three broad classes of forecast: univariate time indices (such as in the Lee-Carter and APC models); multivariate time indices (such as in the Cairns-Blake-Dowd and newer Tang-Li-Tickle model families); and penalty projections (such as with the 2D P-spline model). In each case we identify outliers using quantitative methods, then co-estimate outlier effects along with other parameters. Doing so removes the bias and distortion to the forecast caused by a mortality shock, while providing a robust starting point for projections. Illustrations are given for various models in common use.

**Keywords:** Mortality shocks; outliers; robust forecasting; ARMA; ARIMA; Covid-19

## 1. Introduction

Covid-19 (The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020) caused mortality shocks in many countries – Figures 1(b) and (d) show the sharp increase in death counts in England and Wales in 2020 compared to 2019. These shocks were also present in insurer data sets; Richards (2022a) demonstrated Covid-related mortality shocks in annuity portfolios in France, the UK and the USA. The other strong feature of Figures 1(b) and (d) is the inflection point of 2011, but this paper is concerned with outliers, rather than trend changes.

Actuaries typically split mortality bases into two components: (i) current levels and (ii) future improvements. For current mortality levels actuaries use a portfolio's own recent experience, often using individual lifetimes and multiple covariates; Richards (2022b) presents a methodology allowing for Covid-19 when conducting such analysis. However, to obtain a long enough time series for forecasting future improvements, actuaries typically use population data. Such data has a very different structure, namely grouped counts at individual ages with no covariate information beyond separate data for males and females (Macdonald *et al.*, 2018, Chapter 10).

Many insurers use stochastic mortality models calibrated to population data for future-improvement modelling, and thus risk management, solvency and reporting. There is a broad selection of available stochastic mortality models, each with different quantitative and qualitative properties (Cairns *et al.*, 2009). However, few of these models are designed to cope with outliers, such as pandemic shock mortality. This is a problem for actuaries, as "the presence of even a few anomalous data points can lead to model misspecification, biased parameter estimation and poor

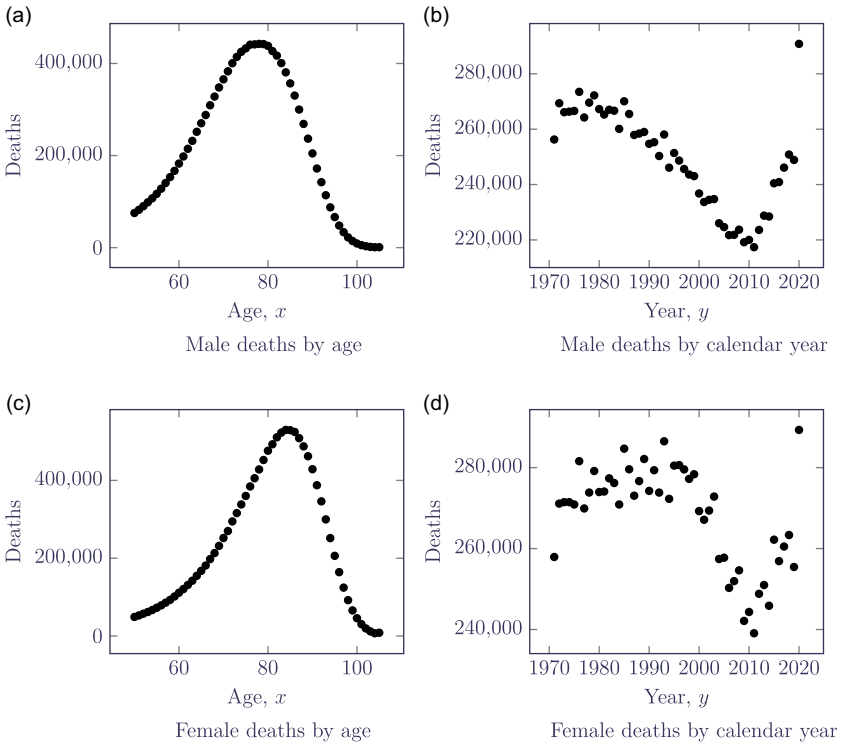


**Figure 1.** Marginal death counts in England and Wales, ages 50–105, 1971–2020
Source: HMD data

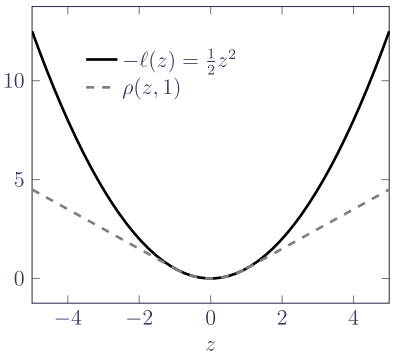


**Figure 2.** Contribution to (negative) log-likelihood for normal variable and robustified contribution using the Huber $\rho$-function of Equation (2) with $k = 1$

forecasts" (Galeano *et al.*, 2006, p. 654). The 2020 mortality experience constitutes just such a distorting anomaly.

One easy approach is just to ignore the affected data. This is only a short-term solution, however, as the 2020 and 2021 experience will over time move from the trailing edge towards the middle of the exposure period. This "deletion approach" is subjective, and two parties may not always agree on what counts as an outlier; Figures 6 and 7 give an example for M9, where it is by no means obvious from visual inspection that 2020 is an outlier. Indeed, outliers can sometimes be masked (Hadi, 1992), and specific examples of this are given in Section 5 and Figure 8. Finally,

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

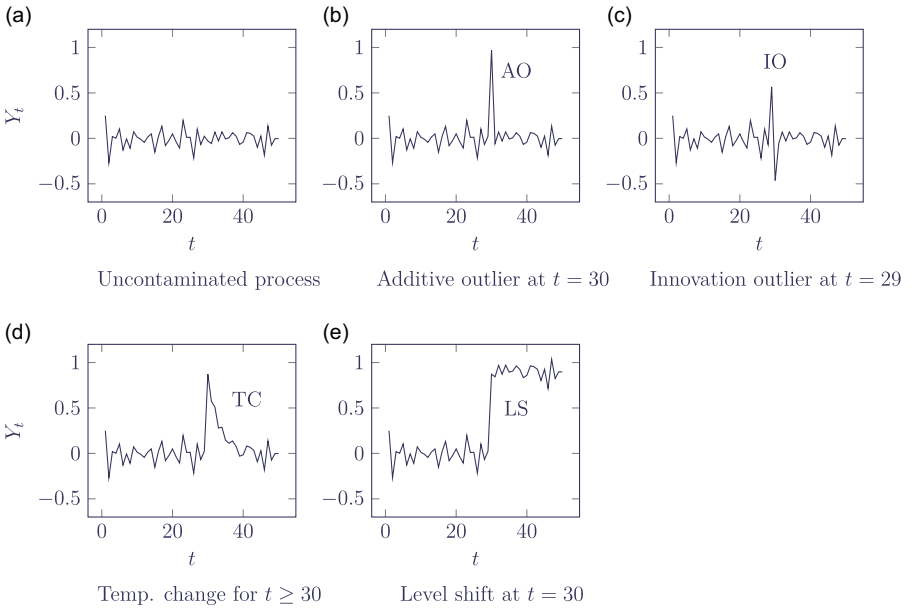


**Figure 3.** Types of outliers in moving-average process $Y_t = \varepsilon_t - 0.8\varepsilon_{t-1}$. Simulation using the R code in Appendix A

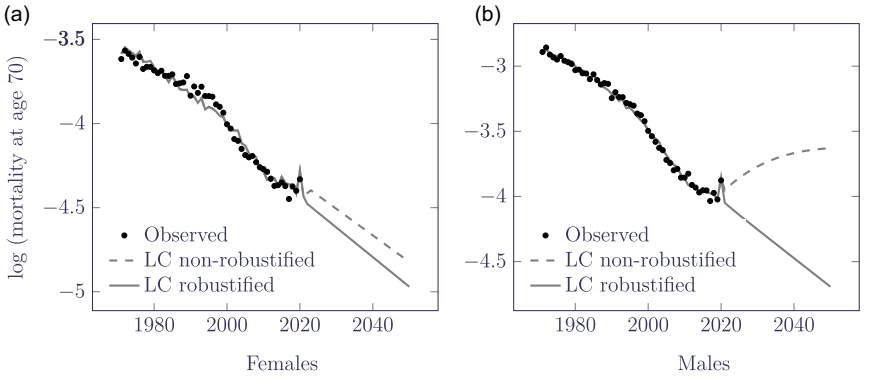


**Figure 4.** Observed mortality rates at age 70 with Lee-Carter forecasts using ordinary regression ARIMA model (Appendix B.2) and robustified regression ARIMA model with critical value 3.5 (Appendix B.3). Data for males and females in England and Wales aged 50–105, over the period 1971–2020

there is an insurance-specific aspect, whereby actuaries repeatedly refit models to measure recalibration risk. An important example is for value-at-risk calculations for longevity trend risk under Solvency II (Richards *et al.*, 2014). For such tasks, an automated outlier-detection procedure is required, for which an objective approach is needed.

The layout of the rest of this paper is as follows: Section 2 describes the data set used; Section 3 outlines some other approaches to dealing with mortality outliers; Section 4 discusses a methodology for robustifying forecasts of univariate time indices, such as the Lee-Carter and APC models; Section 5 describes approaches for multivariate random walks, such as the Cairns-Blake-Dowd and Tang-Lee-Tickle model families; Section 6 presents the results of an approach to the 2D *P*-spline model; Section 7 provides conclusions.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

## 2. Data Description

The data consist of observed deaths in England and Wales, $d_{x,y}$, at age $x$ last birthday in calendar year $y$, together with corresponding mid-year population estimates, $E^c_{x,y}$. The data thus lend themselves to modelling the central mortality rate, $m_{x,y}$, without adjustment. Alternatively, the data can be viewed as suitable for modelling the mortality hazard at age $\left(x + 1/2\right)$ at time $\left(y + 1/2\right)$.

The data are sourced from the Human Mortality Database (HMD) and we use the subset $x \in \{50, 51, \ldots, 105\}$ and $y \in \{1971, 1972, \ldots, 2020\}$. We therefore have $n_x = 56$ ages and $n_y = 50$ years. The age category $x = 105$ technically contains all ages 105 and over, but as the numbers are very low we will treat all those aged 105 and over as being age 105 for simplicity. We alternate our illustrations between males and and females for variety, but both sexes exhibit similar forecasting problems caused by pandemic-affected mortality. Figure 1 shows the marginal death counts, with panels (b) and (d) showing the unusually large number of deaths in 2020 due to the Covid-19 pandemic.

Models in this paper assume that the number of deaths at age $x$ in year $y$ follow a Poisson distribution. With the exception of the 2D penalised-spline model, no allowance is made for over-dispersion in the Poisson counts. With the exception of the Lee-Carter model, models are fitted as penalised constrained generalised linear models using the algorithm of Currie (2013).

## 3. Some Other Approaches to Outliers

### 3.1. Ignoring Affected Data

The immediate response by most actuaries was to ignore the 2020 data and to continue using models calibrated up to 2019. This was a practical, short-term approach to a highly unusual situation. However, ignoring the mortality experience of 1 or 2 years is not a permanent solution, and more rigorous approaches are required. This is particularly the case once the affected years move towards the middle of the data series, as is the case at the time of writing.

### 3.2. Robust Likelihood Estimation

Maximum-likelihood methods can be very sensitive to the presence of outliers in the data. To illustrate, consider an observation, $z$, supposedly from a N(0,1) distribution. The contribution of this observation to the negative log-likelihood is quadratic:

$$-\ell(z) \propto \frac{1}{2}z^2 \tag{1}$$

which means that an outlier has an outsized – and unbounded – influence on estimation. As a result, much early work on time series robustification focused on robustifying the likelihood through use of $\rho$ functions. A $\rho$ function is designed to replace the usual quadratic contribution in Equation (1) with a function that is non-quadratic for extreme values. An ideal $\rho$ function should behave approximately quadratically for observations that are a modest distance from the mean, but it should limit the contribution of extreme values (unlike Equation (1), where an outlier can make an unlimited contribution, thus causing bias). There are many options for such functions, such as in Hampel (1974) and Maronna *et al.* (2006, Section 2.2.4). One early example is the $\rho$ function from Huber (1964):

$$\rho(z, k) = \begin{cases} \frac{1}{2}z^2, & |z| \le k \\ k|z| - \frac{1}{2}k^2, & |z| > k \end{cases} \tag{2}$$

which is shown in Figure 2 for $k = 1$. The Huber $\rho$ function is identical to the quadratic $-\ell$ function within one standard error of the mean, but extreme observations have a linearly increasing contribution to the robustified log-likelihood instead of an exponentially increasing

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

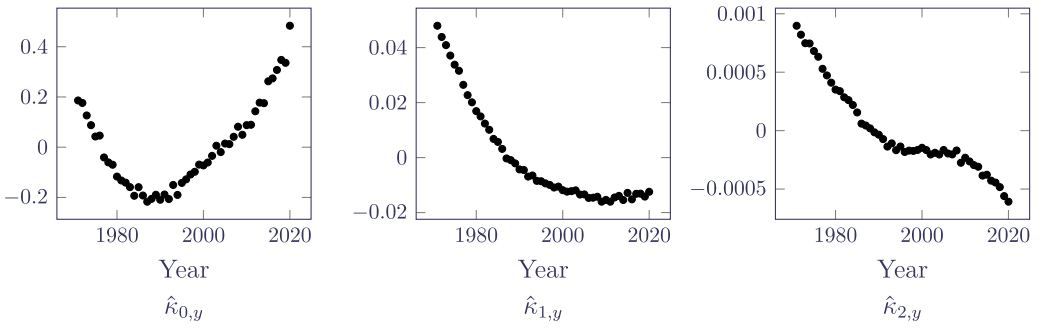


**Figure 5.** Estimates for period effects under the minimally-constrained M9 model. These parameter estimates are not consistent with a multivariate random walk with drift. Contrast with the over-constrained equivalent in Figure 6. The plots are based on HMD data for females aged 50–105, 1971–2020

contribution. The extent to which an outlier contributes to the robustified log-likelihood is dependent on the value of $k$, which requires some judgement.

ARIMA models are a particularly useful representation of stochastic univariate mortality processes, and Martin *et al.* (1983) examined in detail how to robustify the conditional log-likelihood function for such models. One benefit of this approach is the ability to calculate a "clean" version of an affected time series, thus allowing (i) the estimation of an outlier effect, and (ii) the calculation of a sensible forecast start point where the most recent observations contain outliers.

However, robustifying the log-likelihood involves trade-offs. One is reduced estimation efficiency: in Figure 2, observations around 1.5–2 standardised deviations away from the mean make less of a contribution to the log-likelihood than they should in theory, thus making less than full use of all available data. Maronna *et al.* (2006, Section 5.9.1) presented an "optimal" function that balances robustness and efficiency, but it still does not achieve full efficiency. This is a particular issue for mortality-forecasting work, where actuaries often only use data for the past 50 years or so. A particular concern is efficient estimation of the variance of the innovation process, as this plays a large role in determining value-at-risk capital requirements (Kleinow & Richards, 2016, Section 7).

### 3.3. Weighting

Daneel *et al.* (2022) introduced a weighting system for the likelihood, initially with zero weights for pandemic-affected years 2020 and 2021. This was later extended to fractional weights for post-pandemic years in Daneel *et al.* (2023). This approach has at least four problems. Firstly, it requires the analyst to decide which years are outliers, as in Section 3.1. While this can be relatively straightforward for univariate time indices, it is far from simple in multivariate cases; see Section 5, in particular Figures 6 and 7.

Second, weights are manually chosen and therefore arbitrary. This is a potential issue for value-at-risk assessments, which involve repeated simulation and refitting of mortality models. Such repeated model recalibrations require objective criteria for dealing with outliers automatically.

Third, weighting the log-likelihood does not give an estimate of the size of the outlier effect. This means weighting cannot provide a clean starting point for a forecast if recent observations are outliers.

The final problem with weighting the log-likelihood is that it does not fully address the bias problem. Consider weighting the log-likelihood contribution by $w \in (0, 1)$. The weighted contribution to the negative log-likelihood function is then $\rho(z) = \frac{w}{2}z^2$, which is still an exponentially increasing function of the outlier, $z$. Thus, weighting observations is not only arbitrary, but it still leaves the distorting potential of severe outliers.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

If a model cannot handle a feature of real-world data, then a new model is required. In the following three sections we consider three major classes of stochastic model in common actuarial use and how they can be modified to cope with outliers such as those caused by Covid-19.

## 4. Univariate Time Indices

### 4.1. Univariate Mortality Models

Univariate mortality indices are central to several important stochastic projection models, including the model of Lee & Carter (1992):

$$\log m_{x,y} = \alpha_x + \beta_x \kappa_y \tag{3}$$

and the Age-Period-Cohort (APC) model:

$$\log m_{x,y} = \alpha_x + \kappa_y + \gamma_{y-x} \tag{4}$$

Both models require identifiability constraints in order to be fitted, but both fit and forecast for these models are independent of the choice of linear constraints (Currie, 2020). When fitting such models it is useful to smooth $\alpha_x$ and $\beta_x$, as this reduces the dimensionality of the models and improves forecasting performance by reducing the risk of crossover at adjacent ages in the forecast (Delwarde *et al.*, 2007).

To forecast mortality under the Lee-Carter and APC models we need to forecast $\kappa_y$. We can either use a simple random walk with drift, or a full regression ARIMA model; see Appendix A for the structure and operation of both within an ARIMA framework. Note that a random walk with drift is just an ARIMA(0, 1, 0) model, and that a full ARIMA($p$, 1, $q$) model often fits $\kappa_y$ better (Kleinow & Richards, 2016, Table 2). Appendix B.1 considers fitting a random walk with drift in R, while Appendix B.2 considers fitting an ARIMA model around a linear trend.

### 4.2. Outlier Types

For robust estimation of ARIMA models for univariate forecasting we consider the approach of Chen & Liu (1993), which contains two elements. First, Chen & Liu (1993) proposed objective tests to identify where outliers occur. Second, they proposed tests to identify the type of outlier. The ARIMA-robustifying methodology of Chen & Liu (1993) works for two kinds of forecasting model for $\kappa_t$: either a simple random walk with drift, or a regression ARIMA model. Appendix B.3 shows how an outlier effect is co-estimated with the model parameters once an outlier location has been decided.

Chen & Liu (1993) presented tests for four types of outlier: additive outlier (AO), innovation outlier (IO), temporary change (TC) and level shift (LS). Stylised illustrations of each of these are given in Figure 3 for a simple moving-average process, $Y_t$, defined as follows:

$$Y_t = \varepsilon_t - 0.8\varepsilon_{t-1} \tag{5}$$

where $\varepsilon_t \sim \mathrm{N}(0, 0.1)$ and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$. Figure 3(a) shows the first 50 simulations of such a process using the R code in Appendix A.

In Figure 3(b) an additive outlier of 1 has been added to the uncontaminated process at $t = 30$ with all other observations unchanged. The AO represents an external contamination at a specific point without further downstream consequences. In Figure 3(c) an innovation outlier of 0.5 has been added at $t = 29$ – since an IO is an integral part of the process it also affects the series value at $t = 30$. In Figure 3(d) a number of consecutive additive outliers of $0.9{*}0.7^{t-30}$ have been added at $t \geq 30$ to form a temporary change in the series whose impact diminishes. Finally, in Figure 3(e) the series has a permanent shift in level of +0.9 for $t \geq 30$.

Chen & Liu (1993) developed statistical tests to identify outliers and classify them according to type. However, while it is possible to identify an outlier anywhere in the series, "it is impossible to

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

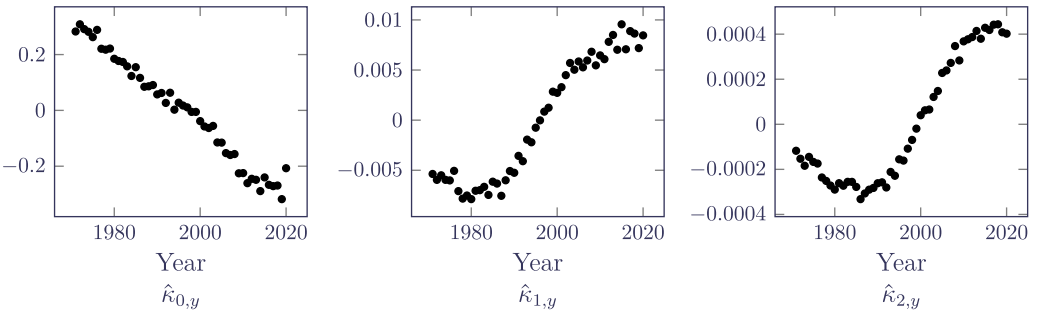


**Figure 6.** Estimates for forecasting parameters under the over-constrained model M9. Contrast with the minimally-constrained equivalent in Figure 5. The plots are based on HMD data for females aged 50–105, 1971–2020

empirically distinguish the type of an outlier occurring at the very end of a series" (Chen & Liu, 1993, p. 286). This has direct relevance to present-day modelling – whilst Covid-19 is a detectable outlier, approaches like that of Chen & Liu (1993) cannot tell us the nature of the outlier until we have observations after the end of the pandemic.

In mortality work we seek to robustify by identifying and measuring AO, TC and LS outliers[1]. In contrast, IO outliers are left in as they are part of the underlying process. The rationale for this is that occasional winters of heavy mortality are a recurring feature, so we do not want to exclude them. Furthermore, "least squares estimates are less affected by innovation outliers than by additive outliers" (Martin *et al.*, 1983, p. 6). Also, excluding IOs would lead to underestimation of the variance of the ARIMA innovation process, $\sigma^2$; this would be undesirable in actuarial work, since this parameter is a major driver of value-at-risk capital requirements for longevity trend risk (Kleinow & Richards, 2016, Section 7).

### 4.3. Choice of Critical Value

Outliers are determined by Chen & Liu (1993) with reference to a critical value. This varies by researcher, but many have opted for a critical threshold of 3 standardised deviations (Maronna *et al.*, 2006, p. 6); with a normally distributed sample, only around 0.3% of observations should be that far away from the mean.

Chen & Liu (1993, Section 3) carried out extensive tests of critical values between 2.25 and 3.5 using simulated series with 100 observations. However, the mortality patterns of the 1920s and 1930s are not generally pertinent to modern actuarial work, so actuaries typically use time series of much shorter lengths, around 50 years. As a result, it is often best to use a critical threshold greater than 3. Consider the example of a Lee-Carter model fitted to the mortality of females in England & Wales over 1971–2020. The best-fitting ARIMA model for $\kappa_y$ using the Chen & Liu (1993) outlier-identification approach leads to the results shown in Table 1.

Using a critical threshold of 3, the methodology of Chen & Liu (1993) finds outliers in 10 out of 50 years, which is excessive. This suggests that for univariate mortality models a higher critical threshold is required. In contrast, using a critical threshold of 3.5 for females in England and Wales identifies the one expected outlier in 2020. This is supported by the test results for simple AR(1) and MA(1) models in Chang *et al.* (1988, Tables 1 and 2), where a critical value of 3.5 also worked well for a series length of 50. Once again, there is also a specific actuarial reason to use a higher critical threshold – Table 1 shows that excluding too many observations leads to an artificially low estimate of $\sigma^2$. Since $\sigma^2$ is a main driver of value-at-risk capital requirements for

[1]It is important to be aware of the distinction between an outlier in a differenced series and an outlier in an undifferenced series. For example, an AO in a differenced series is equivalent to a LS in the undifferenced series (and vice versa).

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

**Table 1.** Selected results from fitting a Chen & Liu (1993) regression ARIMA model to the $\{\hat{\kappa}_y\}$ values in a Lee-Carter model. Mortality data for females in England and Wales, ages 50–105, years 1971–2020

| Critical value | 3 | 3.25 | 3.5 |
|---|---|---|---|
| Years identified as containing AOs | 1973, 1976, 1984, 1991, 1994, 1996, 2003, 2011, 2014 and 2020 | 2003, 2011 and 2020 | 2020 only |
| $\hat{\sigma}^2$ | $2.78 \times 10^{-5}$ | $9.37 \times 10^{-5}$ | $1.29 \times 10^{-4}$ |
| Estimated effect of 2020 outlier | 0.0954 | 0.0572 | 0.0597 |
| $t$-value of 2020 outlier | n/a | 5.23 | 4.96 |

**Table 2.** Selected members from the CBD model family for $\log m_{x,y}$. The layout of the formulae emphasises the commonality and differences between adjacent models. $\bar{x} = \sum_{x=x_{\min}}^{x_{\max}} x/n_x = 77.5$ is the unweighted mean age, while $\hat{\sigma}^2 = \sum_{x=x_{\min}}^{x_{\max}} (x-\bar{x})^2/n_x = 252.25$ is a normalising constant[2] based on the average squared deviation around $\bar{x}$

| Model | Formula | Reference |
|---|---|---|
| M5 | $\kappa_{0,y} + \kappa_{1,y}(x-\bar{x})$ | Cairns *et al.* (2006) |
| M6 | $\kappa_{0,y} + \kappa_{1,y}(x-\bar{x}) + \gamma_{y-x}$ | Cairns *et al.* (2009, Section 4.6) |
| M7 | $\kappa_{0,y} + \kappa_{1,y}(x-\bar{x}) + \gamma_{y-x} + \kappa_{2,y}\left((x-\bar{x})^2 - \hat{\sigma}^2\right)$ | Cairns *et al.* (2009, Section 4.7) |
| M9 | $\kappa_{0,y} + \kappa_{1,y}(x-\bar{x}) + \gamma_{y-x} + \kappa_{2,y}\left((x-\bar{x})^2 - \hat{\sigma}^2\right) + \alpha_x$ | Dowd *et al.* (2020, Section 2) |

longevity trend risk (Kleinow & Richards, 2016, Section 7), too low a critical value would lead to under stated capital requirements.

A critical value of 3.5 performed well in identifying an AO of five standardised deviations in Chang *et al.* (1988, Table 4), and the $t$-values in Table 1 suggest that the Covid-19 mortality in 2020 was also a five-sigma event, i.e. an event five standard deviations away from the mean. However, a true five-sigma event might occur a handful of times every million years, and yet the two mortality spikes due to Covid-19 in 2020 and 2021 were no worse than the two mortality spikes due to influenza in 1918 and 1919 (Richards, 2022b, Figure 2). Two "five-sigma" events in 100 years suggests that the model is wrong (either the Lee-Carter structure or the ARIMA assumption is incorrect) or that $\hat{\sigma}^2$ underestimates the true variance of the innovation process. Either way, it is a reminder that value-at-risk methodologies calibrated to 50 years of data can only set a lower bound for the capital required.

Figure 4 shows the impact of robustifying the regression ARIMA model for forecasting. According to Martin *et al.* (1983, p. 2) "outliers can (lead to) incorrect model identification (. . . ) and seriously impede the construction of forecasts based upon these historical data." We see an example of this in Figure 4(b), where the unrobustified forecast for males is rendered nonsensical due to the bias in the likelihood caused by Covid-19 mortality. However, the robustified forecast has a more sensible starting point and direction.

## 5. Multivariate Time Indices

In this section we look at stochastic mortality models that forecast using a multivariate random walk with drift. There are two broad families: (i) Cairns-Blake-Dowd models, which assume a

[2] $\hat{\sigma}^2$ is the standard notation for this normalising constant, which is used to keep parameter estimates well scaled. This notation unfortunately clashes with the $\hat{\sigma}^2$ used in Table 1 and elsewhere in this paper, which is the standard notation for the variance of the innovation process of a univariate time series.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

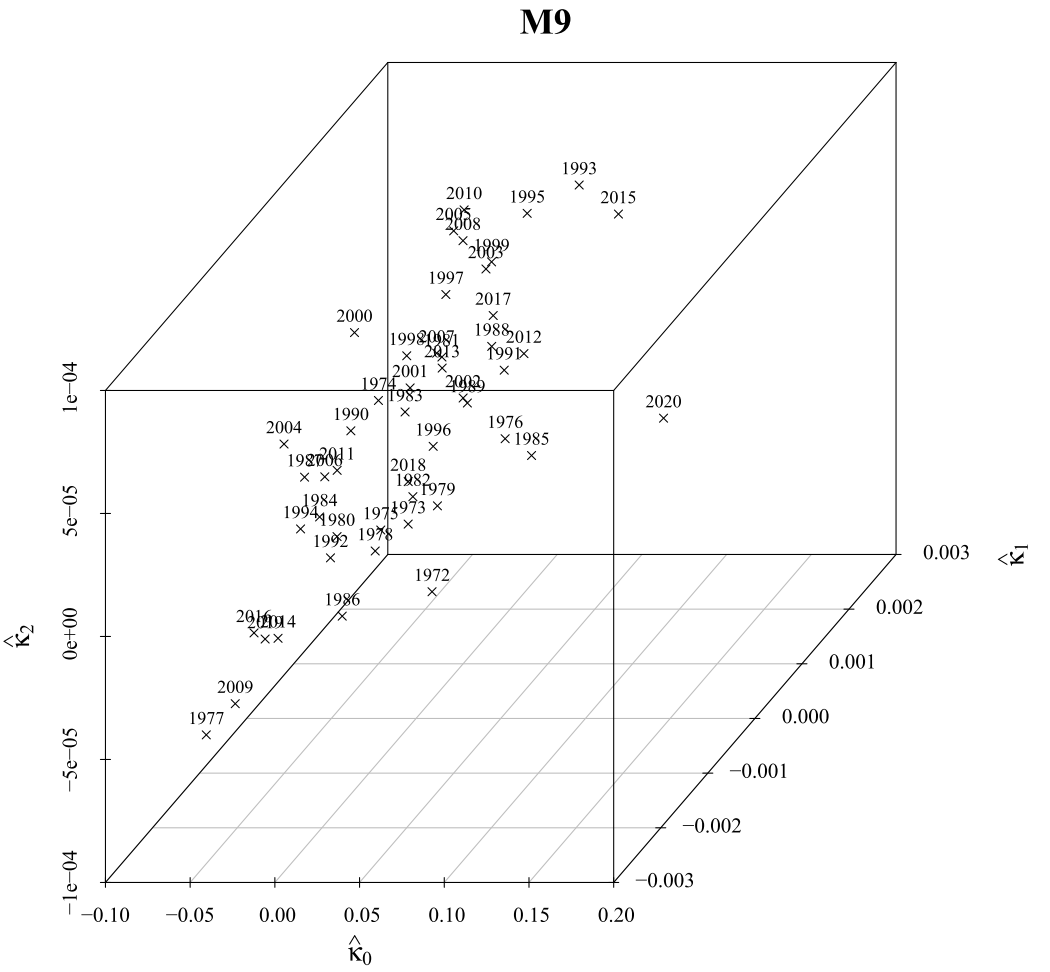


**Figure 7.** Scatterplot of $\Delta\hat{\kappa}$ for M9. The plot is based on HMD data for females aged 50–105, 1971–2020

Gompertz mortality pattern with age and (ii) Tang-Li-Tickle models, which use Hermite splines to allow for slower mortality increases at advanced ages. Both families are well suited to actuarial work because they naturally extrapolate mortality rates to ages higher than the maximum age of the calibrating data set.

### 5.1. Cairns-Blake-Dowd Models

Table 2 gives an overview of four CBD models. The missing model, M8, is excluded as it tends to produce unstable forecasts (Cairns *et al.*, 2011, Section 6), although as a three-dimensional random walk with drift it could be robustified in the same way as M7 and M9.

The parameter naming convention in Table 2 is kept consistent with Cairns *et al.* (2006) and Cairns *et al.* (2009), where the various $\kappa$ terms form the bivariate and trivariate random walks with drift for forecasting. However, it is worth emphasising that these are dependent parameters, and that their values and role depend on the other parameters in the model. This is particularly the case for M9, where the age term, $\alpha_x$, changes the shape, scale and nature of the $\kappa$ terms compared to the other three models.

The models in Table 2 are all fitted as Generalised Linear Models (GLMs) with a Poisson assumption for the number of deaths at each combination of age and year (Currie, 2016). We

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

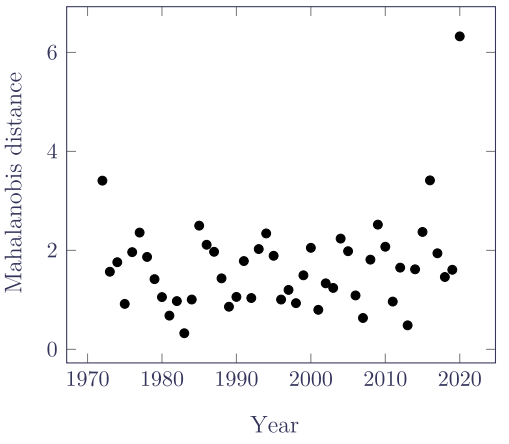


**Figure 8.** Mahalanobis distance for $\Delta\hat{\kappa}$ for M9. The plot is based on HMD data females aged 50–105, 1971–2020

ignore the over-dispersed nature of the death counts, as over-dispersion does not bias estimated means. Djeundje & Currie (2011) discuss dispersed mortality counts in more detail.

Many stochastic mortality models require identifiability constraints (Currie, 2020), but M5 does not require any. M6, M7 and M9 contain a cohort term, $\gamma_{y-x}$, which ordinarily requires two or more identifiability constraints. However, following Richards *et al.* (2019, Appendix B) we do not estimate cohort terms with fewer than four observations, which means no identifiability constraints are required for M6 or M7. This contrasts with the implementation in Cairns *et al.* (2009), where two identifiability constraints were used for M6 and three for M7. If we drop cohort terms with fewer than four observations, the only CBD model requiring identifiability constraints is M9, which needs three linear constraints. Here we use the following for M9:

$$\sum_{y=y_{\min}}^{y_{\max}} \kappa_{0,y} = \sum_{y=y_{\min}}^{y_{\max}} \kappa_{1,y} = \sum_{y=y_{\min}}^{y_{\max}} \kappa_{2,y} = 0 \tag{6}$$

However, there are nevertheless circumstances when we might want to impose more constraints than are mathematically necessary for identifiability. One such reason is to obtain parameter estimates that are forecastable. Figure 5 shows some parameter estimates under minimal identifiability constraints, and these patterns are not consistent with a multivariate random walk with drift. In this paper we therefore over-constrain our M9 models – Figure 6 shows that additional constraints on $\gamma_{y-x}$, although mathematically unnecessary for identifiability, do nevertheless produce parameter estimates more consistent with the forecasting assumption of a multivariate random walk with drift. Currie (2020) provides an extensive treatise on identifiability constraints in linear models, including tests for determining the minimum number of identifiability constraints required.

For forecasting we define $\kappa$ as an $n_y \times p$ matrix composed of the relevant parameters in Table 2. For example, for the M9 model using data for 1971–2020 we would have the following $50 \times 3$ matrix:

$$\kappa = \begin{pmatrix} \kappa_{0,1971} & \kappa_{1,1971} & \kappa_{2,1971} \\ \kappa_{0,1972} & \kappa_{1,1972} & \kappa_{2,1972} \\ \vdots & \vdots & \vdots \\ \kappa_{0,2020} & \kappa_{1,2020} & \kappa_{2,2020} \end{pmatrix} \tag{7}$$

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

**Table 3.** TLT model family for $\log m_{x,y}$

| Model | Formula | Reference |
|---|---|---|
| HS1 | $\alpha_y h_{00}(x) + \omega_y h_{01}(x)$ | Tang *et al.* (2022) |
| HS2 | $\alpha_y h_{00}(x) + \omega_y h_{01}(x) + s_{0,y} h_{10}(x)$ | Tang *et al.* (2022) |
| HS3 | $\alpha_y h_{00}(x) + \omega_y h_{01}(x) + s_{1,y} h_{11}(x)$ | Tang *et al.* (2022) |
| HS4 | $\alpha_y h_{00}(x) + \omega_y h_{01}(x) + s_{0,y} h_{10}(x) + s_{1,y} h_{11}(x)$ | Tang *et al.* (2022) |
| HS5 | $\alpha_y h_{00}(x) + \omega h_{01}(x) + s_{0,y} h_{10}(x)$ | Richards (2023) |

We will confine ourselves to robustifying the multivariate random walk element of models M5, M6, M7 and M9, which allows the forecasting of existing cohorts. The projection of $\gamma_{y-x}$ under M6, M7 and M9 does not require robustification. We note that such outlier cohorts as have been identified in raw data (Cairns *et al.*, 2015) are largely dealt with by the method protocols used by the HMD. See Boumezoued (2021) for details on how outlier cohorts arise from sudden shifts in birth distribution, and how fertility data can be used to improve exposure estimates.

### 5.2. Tang-Li-Tickle Models

A new class of stochastic mortality models was introduced by Tang *et al.* (2022). Instead of the linear Gompertz assumption in CBD models, TLT models use Hermite splines:

$$h_{00}(x) = \big(1 + 2f(x)\big)(1 - f(x))^2 \tag{8}$$

$$h_{10}(x) = f(x)(1 - f(x))^2 \tag{9}$$

$$h_{01}(x) = f^2(x)\big(3 - 2f(x)\big) \tag{10}$$

$$h_{11}(x) = f^2(x)\big(f(x) - 1\big) \tag{11}$$

$$f(x) = \frac{x - x_0}{x_1 - x_0} \tag{12}$$

where $x_0$ is the minimum modelled age and $x_1$ is the maximum modelled age. Often these are set to the minimum and maximum age of the calibrating data set, but they can lie outside this range. For example, although the maximum age in the England and Wales data set in this paper is 105, we can set $x_1 = 120$ to extrapolate mortality rates to this age. With the Hermite basis splines we can define the Tang-Li-Tickle family of models in Table 3.

The models in Table 3 are all fitted as GLMs and do not require identifiability constraints. The TLT family is new, and so has not yet been extended to include cohort terms. Like CBD models, TLT models are useful actuarially because they can extrapolate fitted mortality rates beyond the highest age of the calibrating data set (Richards, 2023, Figure 1).

As with the CBD models in Table 2, the models in Table 3 are also forecast using multivariate random walks with drift: HS1 and HS5 are bivariate, HS2 and HS3 are trivariate and HS4 is tetravariate. In practice there is often such a weak or non-existent trend in some of these variables that it makes sense to impose a common value across time. This is done for HS5, which is a version of HS2 with $\omega_y = \omega$, i.e. turning a trivariate forecast into a bivariate one. HS5 is essentially the Hermite-spline analogue of M5 in Table 2.

For forecasting we define $\kappa$ as an $n_y \times p$ matrix composed of the relevant parameters in Table 3. For example, for the HS2 model using data for 1971–2020 we would have the following $50 \times 3$ matrix:

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

$$\kappa = \begin{pmatrix} \alpha_{1971} & \omega_{1971} & s_{0,1971} \\ \alpha_{1972} & \omega_{1972} & s_{0,1972} \\ \vdots & \vdots & \vdots \\ \alpha_{2020} & \omega_{2020} & s_{0,2020} \end{pmatrix} \tag{13}$$

### 5.3. Outlier Detection for Multivariate Random Walks

Forecasting for the CBD and TLT families is performed as a $p$-dimensional random walk with drift. This means that the row differences in $\hat{\kappa}$ between any two consecutive years are assumed to have a multivariate normal distribution with mean vector $\mu$ and covariance matrix $\mathbf{\Sigma}$, neither of which vary based on the years in question, i.e. $\Delta\hat{\kappa} \sim \text{MVN}(\mu, \mathbf{\Sigma})$.

The detection of outliers in multivariate data can be tricky – "it is quite possible for data to be outliers in multivariate space, but not outliers in any of the original univariate dimensions" (Hadi *et al.*, 2009, p. 57). A possible example of this is shown for M9 in Figure 6 – there is no trace of an outlier in $\hat{\kappa}_{1,2020}$ or $\hat{\kappa}_{2,2020}$, with only a weak suggestion of a possible outlier in $\hat{\kappa}_{0,2020}$. This is an interesting contrast to the univariate example in Figure 17, where the outlier in 2020 is quite clear.

Since model parameters are dependent on each other, the real question is whether the triplet $\left(\hat{\kappa}_{0,2020}, \hat{\kappa}_{1,2020}, \hat{\kappa}_{2,2020}\right)$ is an outlier? Figure 7 shows a scatterplot of $\Delta\hat{\kappa}$, which further illustrates the difficulty of using visual inspection – is 2020 the outlier, or is 1977? Or are there any outliers at all?

One approach to multivariate random walks is to use the Mahalanobis distance, $D_j$, for a $p$-dimensional observation, $z_j = \Delta\kappa_j$:

$$D_j = \sqrt{(z_j - \mu)^T \mathbf{\Sigma}^{-1} (z_j - \mu)} \tag{14}$$

Figure 8 plots the distance measures for $\Delta\hat{\kappa}$. If there are no outliers, and $z_j$ has a normal distribution, then $D_j \sim \chi^2_p$, where $p$ is the number of columns in $\hat{\kappa}$. The upper 5% quantile of $\chi^2_3$ is 7.815, so at face value it looks like 2020 is not an outlier for females in England and Wales under M9.

However, Hadi *et al.* (2009, p. 60) noted that the "Mahalanobis distance is not robust, as it is affected by masking and swamping." Masking is the phenomenon whereby an outlier is hidden because it inflates the estimate of variance used to detect outliers. Swamping is the phenomenon whereby non-outliers have a large Mahalanobis distance because an outlier has distorted the mean of the process. We therefore have an example of masking in Figure 8 because the estimate $\widehat{\mathbf{\Sigma}}$ used in Equation (14) has been distorted by the presence of the outlier we suspect is there. This is an issue for insurers in particular, since a distorted $\widehat{\Sigma}$ will likely lead to excessive capital requirements. Hadi (1992) presented an approach to identifying multivariate outliers, later updated in Hadi (1994), which used robust measures for the mean and covariance matrix to minimise the risk of masking and swamping.

More recently, Galeano *et al.* (2006) presented a methodology for outlier detection for multivariate data using projection pursuit; see also (Huber, 1985) for an early introduction to this topic. Like the Mahalanobis distance, projection pursuit reduces a multidimensional problem to a univariate one. Galeano *et al.* (2006) presented a methodology that sought a mapping from a vector ARMA (VARMA) model to a univariate ARMA one, and furthermore sought mappings that maximised and minimised the kurtosis coefficient. This is done because the kurtosis coefficient, being the fourth moment, is even more sensitive to outliers than the quadratic function of Equation (1). Illustrations of applying Galeano *et al.* (2006) to some M9 models are given in Figure 9.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

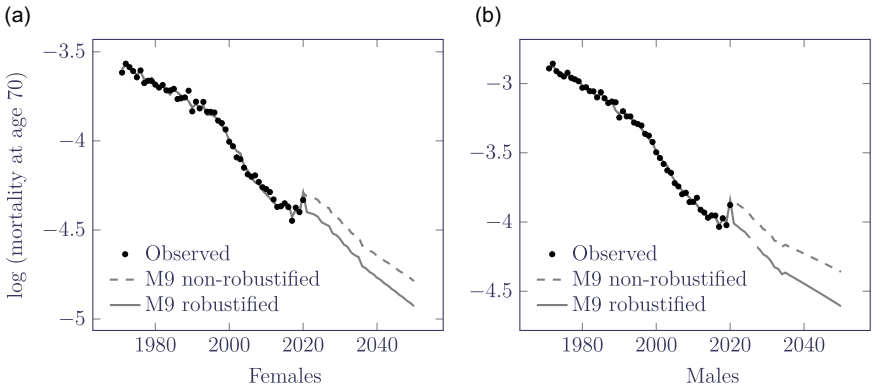


**Figure 9.** Observed mortality rates at age 70 with forecasts using the ordinary M9 model (Dowd *et al*., 2020, Section 2) and robustified M9 model. The data are for males and females in England and Wales aged 50–105, over the period 1971–2020

Galeano *et al*. (2006) considered far more complicated VARMA and VARIMA models with more dimensions and longer time series than actuaries face with CBD and TLT models; Appendix C examines some of the limitations of this methodology when applied to shorter time series with fewer dimensions, which are more typical of stochastic mortality work.

## 6. 2D Penalty Projections

The last of our three classes of projection is the penalty method of the 2D age-period (2DAP) penalised-spline model of Currie *et al*. (2004). This 2DAP model was extended to include period shocks by Kirkby & Currie (2010). In the models fitted in this section we further extend Kirkby & Currie (2010) to allow for over-dispersion.

Kirkby & Currie (2010, Equation 2.9) defined the 2DAP model as a generalised linear array model (GLAM):

$$\log \boldsymbol{M} = \log \boldsymbol{E}^c + \boldsymbol{B}_a \boldsymbol{\Theta} \boldsymbol{B}_y' \tag{15}$$

where $\boldsymbol{M}$ is the matrix of mortality rates indexed by age in the rows and by calendar year in the columns, $\boldsymbol{E}^c$ is the corresponding matrix of population exposures, $\boldsymbol{B}_a$ is a $B$-spline basis matrix for age, $\boldsymbol{B}_y$ is a $B$-spline basis matrix for time, and $\boldsymbol{\Theta}$ is a matrix of regression coefficients.

Fitting the GLAM is done by expressing Equation (15) in vector form:

$$\log \operatorname{vec}(\boldsymbol{M}) = \log \operatorname{vec}(\boldsymbol{E}^c) + \left(\boldsymbol{B}_y \otimes \boldsymbol{B}_a\right)\boldsymbol{\theta} \tag{16}$$

where the vec() function stacks the columns of a matrix into a single vector, $\otimes$ is the Kronecker product and $\boldsymbol{\theta} = \operatorname{vec}(\boldsymbol{\Theta})$. In the fitting procedure, the coefficients in $\boldsymbol{\Theta}$ are simultaneously smoothed in the age and time dimensions by suitable penalty functions. The problem with a mortality shock is that smoothing in the time dimension is no longer sensible. Kirkby & Currie (2010) therefore extended the model in Equation (16) to include period shocks as follows:

$$\log \operatorname{vec}(\boldsymbol{M}) = \log \operatorname{vec}(\boldsymbol{E}^c) + \left(\boldsymbol{B}_y \otimes \boldsymbol{B}_a\right)\boldsymbol{\theta} + \left(\boldsymbol{I}_{n_y} \otimes \breve{\boldsymbol{B}}_s\right)\breve{\boldsymbol{\theta}} \tag{17}$$

where $\boldsymbol{I}_{n_y}$ is the identity matrix for $n_y$ years, $\breve{\boldsymbol{B}}_s$ is a $B$-spline basis in age and $\breve{\boldsymbol{\theta}}$ is a vector of shock coefficients. $\breve{\boldsymbol{B}}_s$ and $\boldsymbol{B}_a$ are both basis matrices in age, but $\breve{\boldsymbol{B}}_s$ typically has fewer knots than $\boldsymbol{B}_a$. When applied to the mortality of females in England and Wales, Equation (17) reveals the period shocks shown in Figure 10.

Figure 10 shows that the model of Kirkby & Currie (2010) not only picks up the Covid-19 mortality shock of 2020, but also various minor period shocks since 1971. However, for

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

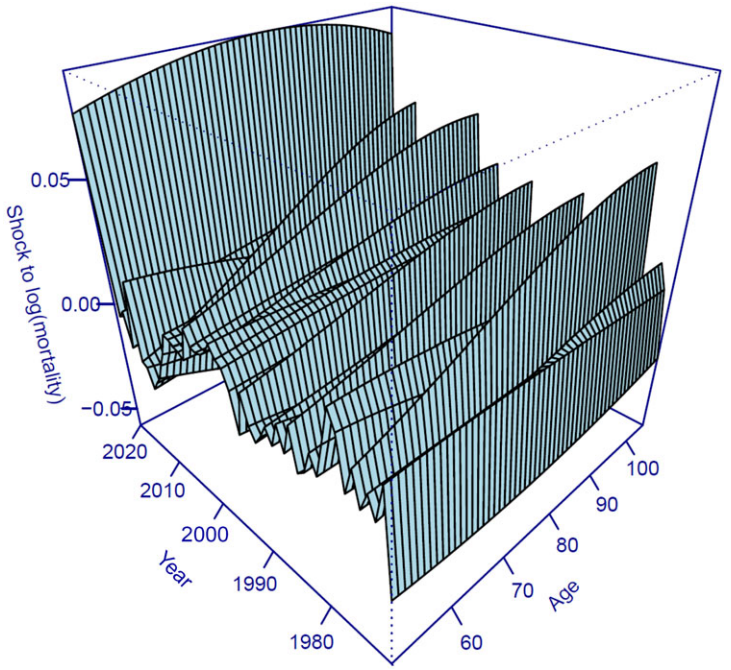


**Figure 10.** Unscaled period shocks under the model in Equation (17). The data are for females in England and Wales aged 50–105, over the period 1971–2020

robustification we are arguably less interested in the minor shocks and would prefer to smooth the most trivial ones towards zero. Kirkby & Currie (2010) therefore proposed an extension whereby the amount of smoothing varies by a scaling factor that is defined by a given year's shock size relative to the largest shock. Thus, if the basic smoothing parameter is $\lambda_s$, the scaled smoothing parameter for year $i$, $\lambda_i$, is then:

$$\lambda_i = \lambda_s \left( \frac{\text{largest shock size}}{\text{shock size for year } i} \right)^{\alpha}, \quad i = 1, \ldots, n_y \tag{18}$$

Figure 11 shows the inverse relative scaling factors ($\lambda_s/\lambda_i$) for simple scaling with $\alpha = 3.24$. The factor $\lambda_s/\lambda_i$ is small for all years relative to 2020, showing that all years bar 2020 have heavy smoothing applied.

Equation (18) gives rise to four smoothing options: (i) no scaling ($\alpha = 0$), as used in the model behind Figure 10; (ii) simple scaling ($\alpha = 1$); (iii) fixed scaling, where $\alpha$ is set to a given value; and (iv) optimised scaling, where $\alpha$ is set by minimising an information criterion such as the BIC. Figure 12 shows the period shocks under optimised scaling, which shows that minor period effects are smoothed towards zero, leaving a focus on the most significant (and otherwise most distorting) shocks. One result is that the magnitude of the 2020 shock is larger in Figure 12 than in Figure 10. Another feature of Figure 12 is that the 2020 shock reduces in size with increasing age. This has parallels with the age-dependent shocks shown in Kirkby & Currie (2010, Figure 5), where the excess mortality of the 1919 influenza pandemic was dramatically higher at ages 20–40 than at post-retirement ages. The reducing shock mortality with increasing age in 2020 contrasts with the pattern of increasing excess mortality with age for the minor shocks in Figure 12. The minor shocks are therefore not just smaller in magnitude, but also qualitatively different.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

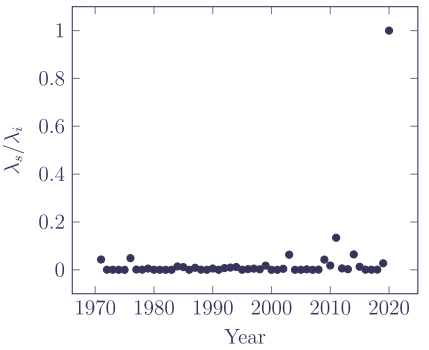


**Figure 11.** Inverse relative scaling factors ($\lambda_s/\lambda_i$) for 2D period-shock model of Kirkby & Currie (2010) with $\alpha = 3.24$. The data are for females in England and Wales aged 50–105, over the period 1971–2020

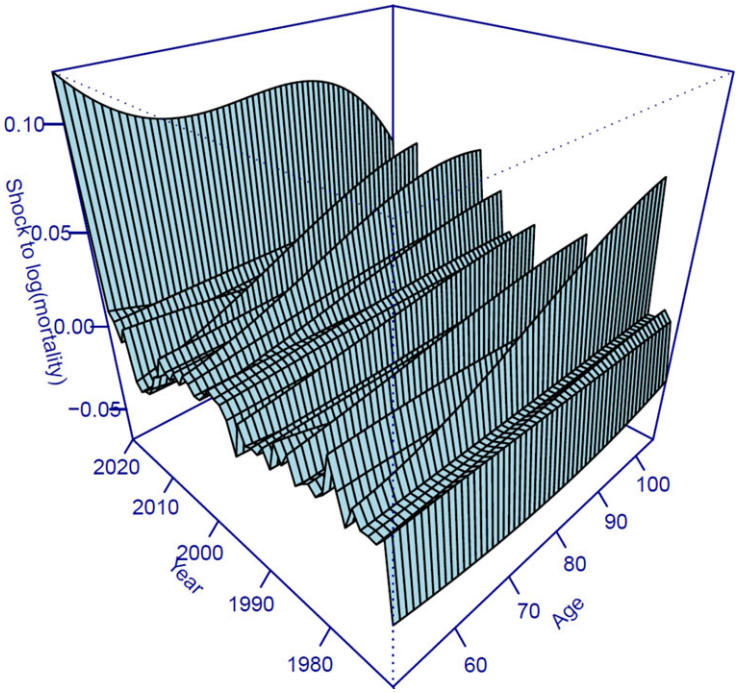


**Figure 12.** Period shocks under optimised scaling using Equation (18). The data are for females in England and Wales aged 50–105, over the period 1971–2020

The use of the 2D period-shock model removes the distorting influence of the Covid-19 pandemic. This produces more sensible forecasts without the undue influence of outliers, as shown in Figure 13.

## 7. Conclusions

The Covid-19 mortality shock of 2020 created outliers in the population mortality data of many countries. These outliers bias parameter estimates in mortality projection models, thus affecting central forecasts and value-at-risk assessments of insurer capital requirements for longevity risk. The short-term solution of ignoring the 2020 and 2021 experience works only as long as those

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

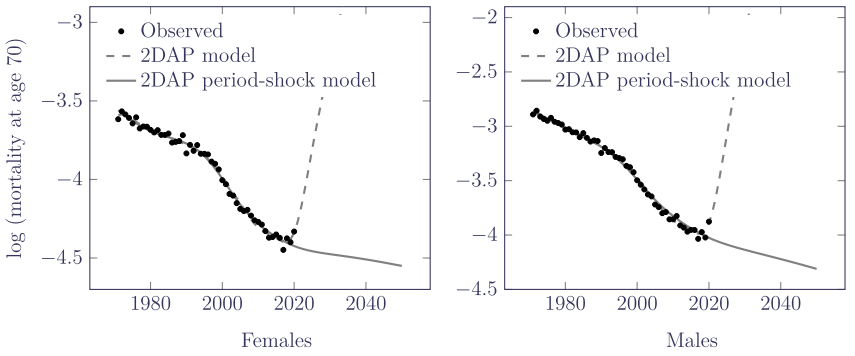


**Figure 13.** Observed mortality rates at age 70 with forecasts using ordinary 2DAP model (Currie *et al*., 2004) and 2DAP period-shock model (Kirkby & Currie, 2010). Data for males and females in England and Wales aged 50–105, over the period 1971–2020

years are at the trailing edge of the data set. As the Covid-affected years move towards the middle of the time series, other approaches are needed, especially objective procedures that can be automated for value-at-risk assessments.

Determination of outliers by visual inspection is unreliable, as how an outlier affects parameter estimation is model-dependent. Simplistic solutions like weighting the log-likelihood are similarly subjective, while also failing to insulate parameter estimates from the bias caused by extreme values. We therefore require objective procedures to (i) identify outliers and – where possible – classify them, (ii) estimate outlier effects alongside forecasting parameters to reduce bias, and (iii) calculate a robust starting point for forecasts where recent observations contain outliers. We outline such objective methods for three classes of stochastic mortality-forecasting models: (i) the approach of Chen & Liu (1993) for univariate mortality indices, (ii) the approach of Galeano *et al.* (2006) for multivariate mortality indices, and (iii) the approach of Kirkby & Currie (2010) for 2D smoothed models. In each case we illustrate how these methods co-estimate parameters and outlier effects, resulting in robust forecasts.

**Acknowledgements.** The author thanks Gavin Ritchie and Torsten Kleinow for helpful comments on earlier drafts of this paper. Any errors or omissions remain the responsibility of the author. Calculations were performed using the Projections Toolkit (Longevitas Development Team, 2024) and bespoke programs written in R (R Core Team, 2021). Graphs were produced in tikz (Tantau, 2024), pgfplots (Feuersänger, 2015) and R, while typesetting was done in LaTeX.

**Disclosure.** The author is a director of Longevitas Ltd, and has a financial interest in the Projections Toolkit.

## References

**Boumezoued, A.** (2021). Improving HMD mortality estimates with HFD fertility data. *North American Actuarial Journal*, **25**(suppl. 1), S255–S279. https://doi.org/10.1080/10920277.2019.1672567.

**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. https://doi.org/10.1111/j.1539-6975.2006.00195.x.

**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. https://doi.org/10.1080/10920277.2009.10597538.

**Cairns, A. J. G., 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**, 355–367.

**Cairns, A. J. G., Blake, D., Dowd, K., & Kessler, A.** (2015). Phantoms never die: Living with unreliable mortality data. *Journal of the Royal Statistical Society, Series A*, **179**, 975–1005. https://doi.org/10.1111/rssa.12159.

**Chang, I., Tiao, G. C., & Chen, C.** (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, **30**(2), 193–204.

**Chen, C., & Liu, L.-M.** (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, **88**(421), 284–297.

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

**Currie, I. D.** (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**(1), 69–93. https://doi.org/10.1177/1471082X12471373.
**Currie, I. D.** (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, **2016**, 356–383.
**Currie, I. D.** (2020). Constraints, the identifiability problem and the forecasting of mortality. *Annals of Actuarial Science*, **14**(2), 537–566. https://doi.org/10.1017/S1748499520000020.
**Currie, I. D., Durban, M., & Eilers, P. H. C.** (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**, 279–298. https://doi.org/10.1191/1471082X04st080oa.
**Daneel, C., Bale, S., Cocevar, P., Hanlon, S., Rimmer, S., Robjohns, N., & Sewell, B.** (2022). *Working paper 160*. Continuous Mortality Investigation Ltd.
**Daneel, C., Bale, S., Cocevar, P., Hanlon, S., Rimmer, S., Robjohns, N., & Sewell, B.** (2023). *Working paper 173*. Continuous Mortality Investigation Ltd.
**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. https://doi.org/10.1177/1471082X0600700103.
**Djeundje, V. A. B., & Currie, I. D.** (2011). Smoothing dispersed counts with applications to mortality data. *Annals of Actuarial Science*, **5**(I), 33–52.
**Dowd, K., Cairns, A. J. G., & Blake, D.** (2020). CBDX: A workhorse mortality model from the Cairns-Blake-Dowd family. *Annals of Actuarial Science*, **14**, 445–460.
**Feuersänger, C.** (2015). *Manual for Package PGFPLOTS, Version 1.12.1*. http://sourceforge.net/projects/pgfplots.
**Galeano, P., Peña, D., & Tsay, R. S.** (2006). Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, **101**(474), 654–669. https://doi.org/10.1198/0162145000001131.
**Hadi, A. S.** (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society, Series B (Methodological)*, **54**(3), 761–771. http://www.jstor.org/stable/2345856.
**Hadi, A. S.** (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society, Series B* (Methodological), **56**(2), 393–396. http://www.jstor.org/stable/2345910.
**Hadi, A. S., Rahmatullah Imon, A. H. M., & Werner, M.** (2009). Detection of outliers. *WIREs Computational Statistics*, **1**(1), 57–70. https://doi.org/10.1002/wics.6.
**Hampel, F. R.** (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**(346), 383–393.
**Harvey, A. C.** (1981). *Time Series Models*. Philip Allan.
**Huber, P.** (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**(1), 73–101.
**Huber, P.** (1985). Projection pursuit. *The Annals of Statistics*, **13**(2), 435–475. https://doi.org/10.1214/aos/1176349519.
**Kirkby, J. G., & Currie, I. D.** (2010). Smooth models of mortality with period shocks. *Statistical Modelling*, **10**(2), 177–196. https://doi.org/10.1017/1471082X0801000204.
**Kleinow, T., & Richards, S. J.** (2016). Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, **2016**(10), 1–25. https://doi.org/10.1080/03461238.2016.1255655.
**Lee, R. D. and Carter, L.** (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659–671. http://www.jstor.org/stable/2290201.
**Longevitas Development Team** (2024). *Projections Toolkit v2.8.7 User Guide*. Longevitas Ltd, Edinburgh, United Kingdom.
**Macdonald, A. S., Richards, S. J., & Currie, I. D.** (2018). *Modelling Mortality with Actuarial Applications*. Cambridge: Cambridge University Press.
**Maronna, R. A., Martin, R. D., & Yohai, V. J.** (2006). *Robust Statistics: Theory and Methods*. John Wiley and Sons Ltd.
**Martin, R. D., Samarov, A., & Vandaele, V. J.** (1983). Robust methods for ARIMA models. In E. Zellner (ed.), *Applied Time Series Analysis of Economic Data*, pp. 153–177. Washington Bureau of the Census.
**Petris, G., Petrone, S., & Campagnoli, P.** (2009). Dynamic Linear Models with R. *Springer*. https://doi.org/10.1007/b135794.
**R Core Team** (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
**Richards, S. J.** (2022a). Real-time measurement of portfolio mortality levels in the presence of shocks and reporting delays. *Annals of Actuarial Science*, **16**(3), 430–452. https://doi.org/10.1017/S1748499522000021.
**Richards, S. J.** (2022b). Allowing for shocks in portfolio mortality models. *British Actuarial Journal*, **27**, 1–22 (with discussion). https://doi.org/10.1017/S1357321721000180.
**Richards, S. J.** (2023). Some comments on "a hermite-spline approach for modelling population mortality" by Tang, Li & Tickle (2022). *Annals of Actuarial Science*, **17**(3), 643–646. https://doi.org/10.1017/S174849952300012X.
**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. https://doi.org/10.1017/S1357321712000451. https://www.longevitas.co.uk/var.
**Richards, S. J., Currie, I. D., Kleinow, T., & Ritchie, G. P.** (2019). A stochastic implementation of the APCI model for mortality projections. *British Actuarial Journal*, **24**, e13. https://doi.org/10.1017/S1357321718000260. https://www.longevitas.co.uk/apci.

**Tang, S., Li, J., & Tickle, L.** (2022). A Hermite spline approach for modelling population mortality. *Annals of Actuarial Science*, **17**, 1–42. https://doi.org/10.1017/S1748499522000173.

**Tantau, T.** (2024). *TikZ and PGF manual for Version 3.1.10*. https://pgf-tikz.github.io/pgf/pgfmanual.pdf.

**The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team** (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) – China, 2020. *China CDC Weekly*, **2**(8), 113–122. https://doi.org/10.46234/ccdcw2020.032. http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51.

## Appendices

### *Appendix A. R Code to Simulate Outlier Types*

```
################################################################################
#
# R script to generate pseudo-process Yt in Equation (5) to illustrate the
# types of univariate outlier in Section 4 and Figure 3.
#
n = 51
set.seed(-1)
epsilon = rnorm(n, sd = 0.1)
kappa = epsilon[2:n] - 0.8*epsilon[1:(n-1)]
t = 1:(n-1)

plotter = function(kappa)
{
    plot(kappa, ylim = range(-1, 1), type = "n")
    lines(kappa, lty = 1, lwd = 2)
}

plotter(kappa)

#
Additive outlier
kappaAO = kappa
kappaAO[30] = kappaAO[30]+1
plotter(kappaAO)

# Innovation outlier
epsilonIO = epsilon
epsilonIO[30] = 0.5
kappaIO = epsilonIO[2:n] - 0.8*epsilonIO[1:(n-1)]
plotter(kappaIO)

# Temporary change
kappaTC = kappa
index = 0:(n-30-1)
kappaTC[30+index] = kappaTC[30+index]+0.9*0.7^index
plotter(kappaTC)

# Level shift
kappaLS = kappa
index = 30:(n-1)
kappaLS[index] = kappaLS[index]+0.9
plotter(kappaLS)
```

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

## Appendix B. AR(I)MA Models for Univariate Mortality Indices

### B.1. ARMA Models for Differences

From an outset year, $y$, we denote by $\kappa_{y+t}$ the value of a univariate mortality index in year $y+t, t \geq 0$. Let $\{X_t\}$ be the stochastic process of the first differences of $\{\kappa_{y+t}\}$, i.e. $X_t = \Delta\kappa_{y+t}$.

We define an autoregressive moving-average (ARMA) model for $X_t$ as follows:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \tag{B1}$$

with autoregressive parameters $\{\phi_i, i = 1, \ldots, p\}$ and moving-average parameters $\{\theta_i, i = 1, \ldots, q\}$. $\{\varepsilon_t\}$ is a white-noise process where each $\varepsilon_t$ is i.i.d. $N(0, \sigma^2)$. Equation (B1) is for an ARMA($p$, $q$) model without a mean. An ARMA($p$, $q$) model for a process $\{X_t\}$ with a mean $\mu$ is defined as follows:

$$(X_t - \mu) = \phi_1 (X_{t-1} - \mu) + \cdots + \phi_p \left(X_{t-p} - \mu\right) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{B2}$$

$\mu$ in Equation (B2) is a constant drift term and represents the long-term tendency for $\kappa_{y+t}$ to change roughly linearly, albeit with potentially long meanders around the linear trend caused by the autoregressive and moving-average parameters.

We further define the backshift operator, $B$, such that $B^i \varepsilon_t = \varepsilon_{t-i}$ for $i \geq 1$. ($B$ is sometimes described as the lag operator and denoted $L$ (Harvey, 1981, p. 26).) For conciseness we can define polynomials in $B$ as $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$. Different authors use different signing conventions – see for example Martin *et al.* (1983, p.5) – but the signing used here is the same as used in R's arima() function. We can then rewrite Equation (B2) more compactly:

$$\phi(B)(X_t - \mu) = \theta(B)\epsilon_t \tag{B3}$$

We will concern ourselves only with stationary processes, which impose bounds on the permissible values taken by parameters (see Harvey (1981, pp.28–35)). We illustrate the estimation of the parameters in Equation (B3) using the univariate mortality index in Table B.1 and illustrated in Figure B.2. The R command in Figure B.1 fits an ARMA(1, 2) model with a mean to the first differences of a variable kappa by maximum-likelihood, where the diff() function calculates the first differences and the arima() function fits the model. The output stage in Figure B.1 shows $\hat{\phi}_1$ as ar1, $\hat{\theta}_1$ as ma1, $\hat{\theta}_2$ as ma2 and $\hat{\sigma}^2$ as sigma2. Somewhat confusingly, the mean, $\hat{\mu}$, is labelled intercept in R's output, a topic we return to in Section B.2.

```
arima(diff(kappa), order=c(1,0,2), method="ML", include.mean=TRUE)

Coefficients:
         ar1      ma1     ma2  intercept
      0.7675  -1.1845  0.6189    -0.0083
s.e.  0.1688   0.1720  0.1322     0.0020

sigma^2 estimated as 5.453e-05:  log likelihood = 166.97,  aic = -323.94
```

**Figure B.1.** R command and output for fitting ARMA(1, 2) model with a mean using the data in Table B.1

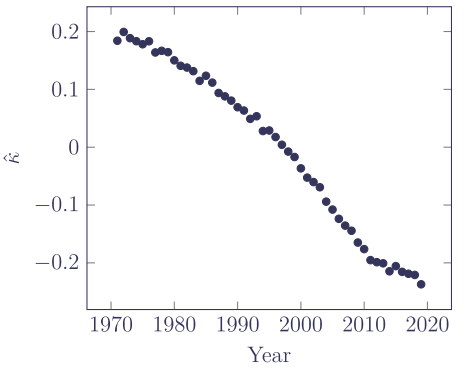


**Figure B.2.** $\hat{\kappa}_{1971+t}$ values from Table B.1

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

**Table B.1.** $\hat{\kappa}_{1971+t}, t = 0, 1, \ldots, 48$ from Lee-Carter model with smoothed $\hat{\alpha}_x$ and $\hat{\beta}_x$ parameters. ONS data for males aged 50–105 in England and Wales, 1971–2019

| | | | |
|---|---|---|---|
| 0.18412 | 0.11476 | 0.00424 | −0.17621 |
| 0.19929 | 0.12355 | −0.00759 | −0.19528 |
| 0.18853 | 0.11168 | −0.01694 | −0.19882 |
| 0.18332 | 0.09380 | −0.03658 | −0.20079 |
| 0.17802 | 0.08792 | −0.05260 | −0.21459 |
| 0.18316 | 0.08054 | −0.06030 | −0.20565 |
| 0.16383 | 0.06914 | −0.06925 | −0.21550 |
| 0.16668 | 0.06339 | −0.09412 | −0.21880 |
| 0.16434 | 0.04904 | −0.10786 | −0.22087 |
| 0.15019 | 0.05352 | −0.12381 | −0.23705 |
| 0.14077 | 0.02797 | −0.13560 | |
| 0.13771 | 0.02890 | −0.14444 | |
| 0.13148 | 0.01764 | −0.16489 | |

## B.2 Regression ARIMA Models

Equation (B3) is a model for $\{X_t\}$, the first differences of $\{\kappa_{y+t}\}$. This is equivalent to a regression ARIMA model for $\kappa_{y+t}$. Such models are termed REGARIMA models by Maronna *et al.* (2006, p. 300). The regression ARIMA model can be fitted to the $\kappa$ variable directly by specifying an external regressor vector containing $1, 2, \ldots, n_y$ as shown in Figure B.3.

A comparison of Figures B.1 and B.3 shows that an ARMA($p$, $q$) model with a mean for the first differences of $\{\kappa_{y+t}\}$ is equivalent to a linear regression model for $\{\kappa_{y+t}\}$ with ARIMA($p$, 1, $q$) errors. The drift estimate, $\hat{\mu}$, is labelled intercept in Figure B.1, but in Figure B.3 it is the coefficient of the external regressor vector 1:ny, i.e. the slope of the assumed linear trend in $\kappa_{y+t}$. It is somewhat less than helpful that the same drift term, $\hat{\mu}$, variously goes by the label "mean," "intercept" and "slope," but the labels stem from the model context. For example, we can re-arrange Equation (B3) as follows:

$$X_t = \mu + \frac{\theta(B)}{\phi(B)}\epsilon_t \tag{B4}$$

If we start at $\kappa_y$, the $t$-steps-ahead forecast, $\kappa_{y+t}$, is then $\kappa_y$ plus the cumulative sum of the next $t$ differences:

$$\begin{aligned} \kappa_{y+t} &= \kappa_y + \sum_{i=1}^{t} X_i \\ &= \kappa_y + t\mu + \sum_{i=1}^{t} \frac{\theta(B)}{\phi(B)}\varepsilon_i \end{aligned} \tag{B5}$$

from which we can see that $\kappa_{y+t}$ is composed of a linear trend, $\kappa_y + t\mu$, plus the addition of a complex ARMA error process.

```
ny = length(kappa)
arima(kappa, order=c(1,1,2), method="ML", xreg=1:ny)

Coefficients:
         ar1      ma1     ma2     1:ny
      0.7675  -1.1845  0.6189  -0.0083
s.e.  0.1688   0.1720  0.1322   0.0020

sigma^2 estimated as 5.453e-05:  log likelihood = 166.97,  aic = -323.94
```

**Figure B.3.** R command and output for fitting a linear regression with ARIMA(1, 1, 2) model for trend deviations using data in Table B.1

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

Thus, the mean (intercept) of the ARMA process for $\{X_t\}$ becomes the slope of the linear trend for $\{\kappa_{y+t}\}$ in the regression ARIMA model. For a regression ARIMA($p$, 1, $q$) model, Equation (B3) can therefore be written directly in terms of $\kappa_{y+t}$ as:

$$\phi(B)(1-B)\big(\kappa_{y+t}-\mu_t\big)=\theta(B)\epsilon_t \tag{B6}$$

where $\mu_t = t\mu$. Note that it is common to fit ARIMA models by recasting them in state-space form and applying a Kalman filter; see Martin *et al.* (1983, Section 5.1) or Harvey (1981, Chapter 3). Indeed, once the model is in state-space form it would be possible to allow $\mu_t$ to vary more flexibly in time as a full dynamic linear model (Petris *et al.*, 2009), but this is beyond the scope of this paper.

### B.3 Regression ARIMA models with outliers

We now consider the fitting of a regression ARIMA model for $\{\kappa_{y+t}\}$ where there is contamination with one or more outliers. Table B.2 is equivalent to Table B.1 but with the addition of an extra year's data. Figure B.4 shows the outlier in 2020 caused by the first Covid-19 shock in April and May 2020.

The outlier in 2020 has distorted the model fit in Figure B.5 considerably, as all parameter estimates have taken very different values compared to Figure B.3. To fit a regression ARIMA model for $\{\kappa_{y+t}\}$ with allowance for an outlier in 2020 we define a matrix of external regressors, XREG. The first column of XREG is the linear trend $1:n_y$ as before, whereas the second column is an indicator variable taking the value 1 in 2020 and 0 in all other years. Further outliers for other years can be added similarly as additional columns for XREG. The R commands to fit this outlier-aware model are shown in Figure B.6, along with the output.

In Figure B.6 the estimated drift term, $\hat{\mu}$, now appears as XREG1, while the estimate of the outlier effect of the 2020 Covid-19 mortality appears as XREG2. Note that the estimates $\hat{\phi}_1$, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\mu}$. and $\hat{\sigma}^2$ are little changed between Figures B.3 and B.6, showing that co-estimation of the outlier effect along with the ARIMA parameters eliminates the distortion demonstrated in Figure B.5.

Finally, for forecasting we need to robustify the starting point. In the case of the data in Table B.2, the cleaned value is $\hat{\kappa}_{2020}$ minus XREG2, i.e. $-0.1669 - 0.0631 = -0.2300$.

**Table B.2.** $\hat{\kappa}_{1971+t}, t = 0, 1, \ldots, 49$ from the Lee-Carter model with smoothed $\hat{\alpha}_x$ and $\hat{\beta}_x$ parameters. ONS data for males aged 50–105 in England and Wales, 1971–2020

| | | | |
|---|---|---|---|
| 0.18483 | 0.11651 | 0.00768 | −0.17030 |
| 0.19979 | 0.12518 | −0.00398 | −0.18911 |
| 0.18918 | 0.11349 | −0.01321 | −0.19261 |
| 0.18405 | 0.09588 | −0.03258 | −0.19457 |
| 0.17883 | 0.09010 | −0.04837 | −0.20816 |
| 0.18389 | 0.08283 | −0.05598 | −0.19938 |
| 0.16484 | 0.07160 | −0.06482 | −0.20906 |
| 0.16765 | 0.06593 | −0.08932 | −0.21232 |
| 0.16534 | 0.05181 | −0.10288 | −0.21436 |
| 0.15140 | 0.05622 | −0.11862 | −0.23029 |
| 0.14211 | 0.03106 | −0.13025 | −0.16691 |
| 0.13912 | 0.03197 | −0.13898 | |
| 0.13298 | 0.02088 | −0.15912 | |

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

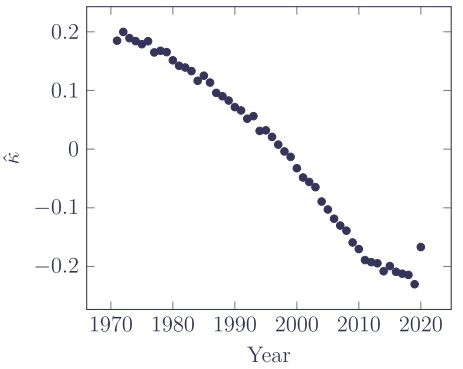


**Figure B.4.** $\hat{\kappa}_{1971+t}$ values from Table B.2, showing the outlier in 2020

```
arima(x = kappa, order = c(1, 1, 2), xreg = 1:ny, method = "ML")

Coefficients:
         ar1      ma1     ma2     1:ny
      0.9533  -1.6968  0.9427  -0.0024
s.e.  0.0523   0.2522  0.2176   0.0065

sigma^2 estimated as 0.0001046:  log likelihood = 152.4,  aic = -294.79
```

**Figure B.5.** R command and output for fitting linear regression with ARIMA(1, 1, 2) model for trend deviations using data in Table B.2

```
ny = length(kappa)
XREG = cbind(1:ny, c(rep(0, ny-1), 1))
arima(kappa, order=c(1,1,2), method="ML", xreg=XREG)

Coefficients:
         ar1      ma1     ma2    XREG1   XREG2
      0.7685  -1.1850  0.6193  -0.0081  0.0631
s.e.  0.1667   0.1699  0.1309   0.0019  0.0081

sigma^2 estimated as 5.184e-05:  log likelihood = 171.7,  aic = -331.39
```

**Figure B.6.** R commands and output for fitting outlier-robustified linear regression with ARIMA(1, 1, 2) model for trend deviations using data in Table B.2

## *Appendix C. Multivariate Outlier Detection in R*

R offers an implementation of Galeano *et al.* (2006) in the outliers.hdts() function in the SLBDD package. This is intended for multivariate data generated by a variety of vector ARMA (VARMA) and vector ARIMA (VARIMA) processes, potentially with a large number of dimensions. In this appendix we look at the performance of the outliers.hdts() function when applied to the multivariate random walks in the CBD and TLT families of stochastic mortality models.

Galeano *et al.* (2006, Table 2) provide the empirical critical thresholds for eight specimen models for time series of various lengths. One issue for actuaries is that the shortest time series tested has 50 observations, whereas typically this would be close to the maximum relevant length for actuarial work. A second issue for actuaries is that none of the eight models tested corresponds to a multivariate random walk with drift. Finally, the outliers.hdts() function uses a hard-coded critical value derived from the square root of a power of the upper 5% value of a $\chi^2_1$ distribution. We are therefore interested in the

https://doi.org/10.1017/S1357321724000175 Published online by Cambridge University Press

**Table C.1.** False-positive rates for outliers.hdts() function with hard-coded p-value of 5% in SLBDD package with 10,000 simulated random walks with drift. For mortality work with CBD and TLT models it is therefore better to robustify the differenced multivariate series. Results derived from simulating random walks with drift specified in Equations (25) and (26)

| Series | Differenced | | Undifferenced | |
|---|---|---|---|---|
| Length | M5 | M7 | M5 | M7 |
| 25 | 4.80% | 4.12% | 15.44% | 10.26% |
| 35 | 3.97% | 3.91% | 15.31% | 9.24% |
| 35 | 3.76% | 3.38% | 14.86% | 8.46% |
| 40 | 3.09% | 2.94% | 14.55% | 7.77% |
| 45 | 3.06% | 2.50% | 13.98% | 7.35% |
| 50 | 2.97% | 2.24% | 13.93% | 7.24% |

performance of the outliers.hdts() function with the shorter series of fewer dimensions that are likely to be encountered in actuarial work.

We fit M5 and M7 models to the mortality experience of males aged 50–105 in England & Wales over the period 1971–2019. The estimated parameters of the bivariate random walk with drift for $\hat{\kappa}_{0,y}$ and $\hat{\kappa}_{1,y}$ in the M5 model are:

$$\begin{aligned} \widehat{\mu} &= \begin{pmatrix} -0.01750290 \\ 0.000331263 \end{pmatrix} \\ \widehat{\mathbf{\Sigma}} &= \begin{pmatrix} 5.081656 \times 10^{-4} & 1.251913 \times 10^{-5} \\ 1.251913 \times 10^{-5} & 7.224136 \times 10^{-7} \end{pmatrix} \end{aligned} \tag{C1}$$

and the estimated parameters of the trivariate random walk with drift for $\hat{\kappa}_{0,y}$, $\hat{\kappa}_{1,y}$ and $\hat{\kappa}_{2,y}$ in the M7 model are:

$$\begin{aligned} \widehat{\mu} &= \begin{pmatrix} -7.283770 \times 10^{-3} \\ -4.100712 \times 10^{-4} \\ -5.730169 \times 10^{-7} \end{pmatrix} \\ \widehat{\mathbf{\Sigma}} &= \begin{pmatrix} 6.811852 \times 10^{-4} & 2.514202 \times 10^{-5} & 5.142949 \times 10^{-7} \\ 2.514202 \times 10^{-5} & 1.321879 \times 10^{-6} & 2.860279 \times 10^{-8} \\ 5.142949 \times 10^{-7} & 2.860279 \times 10^{-8} & 1.239491 \times 10^{09} \end{pmatrix} \end{aligned} \tag{C2}$$

We simulate the random walks with drift for a term of 50 years, first by simulating the differenced series (the VARMA process) then calculating the cumulative sum (the VARIMA process). We then call outliers.hdts() for both the differenced and undifferenced series using the full 50-observation series, but also for the first 25, 30, 35, 40 and 45 values to check the function performance with shorter series. The proportions of simulated series erroneously identified as containing outliers (the false positives) are shown in Table C.1.

Table C.1 shows that the outliers.hdts() function has a radically different performance based on whether the differenced or undifferenced series is used. The false-positive rate for the differenced series is typically smaller than the hard-coded 5% rate, and is in fact smaller than this for mortality series of length 40–50 years. On the basis of this limited assessment, the unmodified outliers.hdts() function looks appropriate for application to the multivariate random walks for CBD and TLT stochastic mortality models, but only when passed the differenced series. In practice, one can edit the source code for outliers.hdts() and related functions to vary the p-value from the hard-coded 5%. However, Table C.1 shows that the false-positive rates of outliers.hdts() vary strongly, with undifferenced bivariate random walks particularly at risk of greater false positives than the programmed p-value implies.

Despite its restricted functional interface, the outliers.hdts() function is useful for robustifying multivariate mortality indices. The function returns a cleaned version of the indices, which can be used to calculate robust estimates of the forecasting parameters while also providing robustified starting points for the forecast. Estimates of the multivariate outliers can be obtained by simply deducting the cleaned series from the original series.

**Cite this article:** Richards SJ. (2024). Robust mortality forecasting in the presence of outliers. *British Actuarial Journal*. https://doi.org/10.1017/S1357321724000175