Invariance and the forecasting of mortality

Iain Currie¹

¹ Heriot-Watt University, UK

E-mail for correspondence: I.D.Currie@hw.ac.uk

Abstract: The forecasting of human mortality is an import topic for providers of pensions and care of the elderly. Many models of mortality are not identifiable so parameter constraints are used to obtain parameter estimates that can be used for forecasting. We show that when an ARIMA model is used to forecast parameter estimates the resulting forecasts of mortality are invariant with respect to the choice of constraints. These results remain true when some model terms are smoothed. We illustrate our results with Portuguese data.

Keywords: Forecasting; Identifiability; Invariance; Mortality.

1 Introduction

The mortality of an individual depends on their current age, the current year and their year of birth (among other risk factors). These determinants are generally known as the age effect, the period effect and the cohort effect. A problem of major interest to the providers of pensions and care of the elderly is the forecasting of mortality. In the financial world the usual approach is to construct a model in terms of the age, period and cohort effects. The model is fitted to appropriate data, the period and cohort effects are forecast, and a forecast of mortality is obtained.

However, most such models are not identifiable and it is not clear what exactly is being forecast. For example, Clayton and Schifflers (1987) in a carefully argued paper “doubt the wisdom” of such forecasting. Nevertheless the method does give plausible answers that are consistent across a range of models. Our purpose here is to try to explain why this is so.

We use Portuguese mortality data downloaded from the Human Mortality Database on December 18, 2018. We have the number of deaths \(d_{x,y}\) and the corresponding central exposed to risk \(e_{x,y}\) for ages 50 to 90 and years 1970 to 2015. For simplicity we will index the ages by \(x_n = (1, \ldots, n)^2\), the
years by $\mathbf{x}_y = (1, \ldots, n_y)^T$ and the years of birth by $\mathbf{x}_\text{c} = (1, \ldots, n_{\text{c}})^T$ where $n_{\text{c}} = n_a + n_y - 1$ is the number of distinct cohorts. With this convention age $x$ has index $i = x - \min(x) + 1$ and year $y$ has index $j = y - \min(y) + 1$. The oldest cohort in the first year is indexed one; thus, the cohort index for age $x$ in year $y$ is $n_{\text{a}} - i + j$. We suppose that the number of deaths at age $x$ in year $y$ follows a Poisson distribution $\mathcal{P}(e^{x,y} \lambda_{x,y})$ where $\lambda_{x,y}$ is the force of mortality or hazard of death at age $x$ in year $y$. With our data we have $n_a = 41$, $n_y = 46$ and $n_{\text{c}} = 86$.

The plan of the paper is: section 2 describes our method, section 3 gives an example and section 4 contains some concluding remarks.

## 2 Method

We consider a generalized linear model or GLM with model matrix $\mathbf{X}$, $n \times p$, $n > p$ and rank $p - q$ where $q \geq 1$. We denote the vector of parameters by $\mathbf{\theta}$. Since $\mathbf{X}$ is not of full rank $\mathbf{\theta}$ is not identifiable. However, there exists a matrix $\mathbf{H}$, $q \times p$, with rank $q$ such that $\mathbf{H} \mathbf{\theta} = \mathbf{0}$. Now, subject to the condition that $\mathbf{H} \mathbf{\theta} = \mathbf{0}$, we do have a unique estimate of $\mathbf{\theta}$. We refer to $\mathbf{H}$ as a constraints matrix and we note that $\mathbf{H}$ is not unique.

We will use the $P$-spline system of smoothing when we wish to smooth certain parameters in our models; see Eilers and Marx (1996) for a general introduction and Currie et al. (2004) for details in our present application. We denote the penalty matrix of the $P$-spline system by $\mathbf{P}$.

Currie (2013) generalized the Nelder and Wedderburn (1972) algorithm for estimation of $\mathbf{\theta}$ in a GLM. The following is a scoring algorithm for estimation of $\mathbf{\theta}$ subject to (a) the constraint $\mathbf{H} \mathbf{\theta} = \mathbf{0}$ and (b) smoothing via the penalty matrix $\mathbf{P}$

\[
\left( \begin{array}{c} X^T \tilde{W} X + P \\ H \end{array} \right) \left( \begin{array}{c} \hat{\mathbf{\theta}} \\ \hat{\omega} \end{array} \right) = \left( \begin{array}{c} X^T \tilde{W} \tilde{z} \\ 0 \end{array} \right); \tag{1}
\]

here $\tilde{W}$ is the diagonal matrix of weights, $\tilde{z}$ is the so-called working variable and $\hat{\omega}$ is an auxiliary variable. If a canonical link is used then (1) is a Newton-Raphson scheme.

Let $\hat{\mathbf{\theta}}_i$ be the maximum likelihood estimate of $\mathbf{\theta}$ under the constraint $\mathbf{H}_i \mathbf{\theta} = \mathbf{0}$, $i = 1, 2$. The fitted values are invariant with respect to the choice of constraints so $\mathbf{X} \hat{\mathbf{\theta}}_1 = \mathbf{X} \hat{\mathbf{\theta}}_2$. We are interested in the relationship between $\hat{\mathbf{\theta}}_1$ and $\hat{\mathbf{\theta}}_2$. This is characterized by the null space of $\mathbf{X}$ which we define as

\[
\mathcal{N}(\mathbf{X}) = \{ \mathbf{v} : \mathbf{X} \mathbf{v} = \mathbf{0} \}. \tag{2}
\]

With this notation, $\hat{\mathbf{\theta}}_1 - \hat{\mathbf{\theta}}_2 \in \mathcal{N}(\mathbf{X})$.

The idea is to use two constraint systems: the first will be a system used in the literature, a standard system, and the second will be a random system. We will then use the null space of $\mathbf{X}$ to show that forecasts of mortality
under these two systems are also invariant when an ARIMA model is used to forecast.

3 Example

One popular model for forecasting mortality in the financial world is the age-period-cohort model. We have

\[ \log \lambda_{i,j} = \alpha_i + \kappa_j + \gamma_{n_a-i+j}, \quad i = 1, \ldots, n_a, \quad j = 1, \ldots, n_y, \tag{3} \]

where \( i \) is the age of death, \( j \) is the year of death and \( n_a-i+j \) is the year of birth. Let \( \alpha = (\alpha_1, \ldots, \alpha_{n_a})^T \), \( \kappa = (\kappa_1, \ldots, \kappa_{n_y})^T \) and \( \gamma = (\gamma_1, \ldots, \gamma_{n_c})^T \); let \( \theta = (\alpha^T, \kappa^T, \gamma^T)^T \). Model (3) has \( n_a + n_y + n_c \) parameters but the rank of its model matrix is \( n_a + n_y + n_c - 3 \) so three constraints are required to give a unique estimate of \( \theta \). One common or standard set of constraints is

\[ \sum_1^{n_y} \kappa_j = \sum_1^{n_c} \gamma_c = 0; \tag{4} \]

see Cairns et al. (2009) for example. We also consider a set of random constraints

\[ \sum u_{i,j} \theta_j, \quad i = 1, 2, 3, \quad j = 1, \ldots, n_a + n_y + n_c \tag{5} \]

where the \( u_{i,j} \) are independent uniform variables, \( U(0,1) \). We denote the estimates under the two constraint systems by \( \hat{\theta}_s \) and \( \hat{\theta}_r \) respectively. Figure 1 shows the estimates of \( \alpha, \kappa \) and \( \gamma \) under the two systems for our Portuguese data and one set of random constraints. The estimates are strikingly different in both shape and scale, yet the invariance of the fitted values guarantees that the fitted values \( \log \hat{\lambda} \) are equal, as in the bottom right panel.

A basis for the null space, \( \mathcal{N}(X) \), of the model matrix, \( X \), is

\[ \left\{ \begin{pmatrix} 1_{n_a} \\ -1_{n_y} \\ 0_{n_c} \end{pmatrix}, \quad \begin{pmatrix} 1_{n_a} \\ 0_{n_y} \\ -1_{n_c} \end{pmatrix}, \quad \begin{pmatrix} x_a \\ -x_y \\ x_c-n_a1_{n_a} \end{pmatrix} \right\} \tag{6} \]

where \( 1 \) and \( 0 \) are vectors of 1s and 0s respectively of the indicated lengths. The estimates \( \hat{\theta}_s \) and \( \hat{\theta}_r \) are intimately related; their difference \( \hat{\theta}_s - \hat{\theta}_r \) lies in \( \mathcal{N}(X) \). We define \( \Delta \hat{\alpha} = \hat{\alpha}_s - \hat{\alpha}_r \), \( \Delta \hat{\kappa} = \hat{\kappa}_s - \hat{\kappa}_r \) and \( \Delta \hat{\gamma} = \hat{\gamma}_s - \hat{\gamma}_r \). Then, equating coefficients in (6), we find

\[ \Delta \hat{\alpha} = (A + B)1_{n_a} + Cx_a \tag{7} \]
\[ \Delta \hat{\kappa} = -A1_{n_y} - Cx_y \tag{8} \]
\[ \Delta \hat{\gamma} = -(B + n_aC)1_{n_c} + Cx_c \tag{9} \]
for some constants $A$, $B$ and $C$; in our example, we found $A = -4.07$, $B = 3.90$ and $C = -0.0818$. Clayton and Schifflers (1987) among others have observed that the age, period and cohort parameters are only estimable up to a linear function. Equations (7), (8) and (9) make this precise; the left panel of Figure 2 illustrates these relations.

We turn now to forecasting. We forecast the period and cohort terms with an ARIMA model and then, with $\alpha$ fixed at its estimated value, use (3) to forecast the log $\hat{\lambda}$. Despite the difference between $\hat{\kappa}_s$ and $\hat{\kappa}_r$, and between $\hat{\gamma}_s$ and $\hat{\gamma}_r$, the forecast values of log $\lambda$ are invariant with respect to the choice of constraints, just like the fitted values. This invariance is illustrated in the bottom right panel of Figure 1; it can also be proved with (7), (8) and (9).

We turn briefly to the effect of smoothing. Regular forecasts of log $\lambda$ are desirable for the pricing and reserving of many financial products. In the age-period-cohort model this regularity can be achieved by smoothing the age term $\alpha$. We use the $P$-spline system with cubic B-splines, a second
order penalty and a knot spacing of $\delta_a = 5$. Let $B_a$, $n_a \times c_a$, be the resulting regression matrix along age and set $\alpha = B_a a$. We refer to this model as the smooth age-period-cohort model. The regression coefficients are $\theta = (a^T, \kappa^T, \gamma^T)^T$ with length $c_a + n_y + n_c$. We use the same approach as in the unsmoothed model. We use the constraints (4) and the random constraints (5) but with $j = 1, \ldots, c_a + n_y + n_c$. A basis for the null space of the model is

$$\left\{ \begin{pmatrix} 1_{c_a} \\ -1_{n_y} \\ 0_{n_c} \end{pmatrix}, \begin{pmatrix} 1_{c_a} \\ 0_{n_y} \\ -1_{n_c} \end{pmatrix}, \begin{pmatrix} \delta_a x_{c_a} \\ -x_y \\ x_c - \omega_c 1_{n_c} \end{pmatrix} \right\}; \quad (10)$$

here $x_{c_a} = (1, 2, \ldots, c_a)^T$ and $\omega_c = n_a + 2\delta_a - 1$. The right panel of Figure 2 illustrates the relations; we note that $0.4876 = \delta_a 0.09751$, in agreement with the basis in (10). Again we see that the age, period and cohort effects are estimable only up to a linear function. It follows in a similar fashion to the unsmoothed case that not only are fitted values invariant with respect to the choice of constraints but so also are their forecast values.

4 Conclusions

The rates of mortality in models of mortality are uniquely estimable, despite the fact that the component terms, ie, the age, period and cohort terms, are not uniquely estimable. This follows from the invariance of fitted values in a GLM. This gives rise to the following paradox: the period and cohort terms are not estimable and so neither are their forecast values. Why is it then that in practice the forecast rates of mortality seem plausible across a range
of mortality models? In this short paper we provide a partial resolution of this paradox, namely that the forecast values of $\log \lambda$ are also invariant with respect to the choice of constraints used to estimate the age, period and cohort effects. For a fuller discussion of these ideas and further examples see Currie (in preparation).

Acknowledgments: I am very grateful to Jim Howie for a tutorial on null spaces and to Longevitas Ltd for financial support.

References


