Modelling Mortality for Pension Schemes: Online Appendices

1 Identifiability in the sub-population model

In Hunt and Blake (2015a) and Hunt and Blake (2015b), we discussed the identifiability issues in age/period and age/period/cohort mortality models, respectively. In particular, we find that almost all APC mortality models possess "invariant" transformations, i.e., transformations of the parameters of the model which leave the fitted mortality rates unchanged. In order to find a unique set of parameters, we impose a set of identifiability constraints on them. Typically, these are chosen so that we can assign our desired interpretation of the demographic significance to the parameters in question. However, because this interpretation is subjective, it is important that our choice of identifiability constraints does not have any impact on any observable quantities. For instance, we discuss in Hunt and Blake (2015a,b) how to ensure that projected mortality rates are independent of the choice of identifiability constraints.

The model in Equation 2 of the main paper does not possess any additional identifiability issues in and of itself, once the parameters from the reference population are known. However, due to its structure, transformations of the parameters in the reference population model will have knock-on effects for those in the sub-population model. It is important therefore that invariant transformations of the reference model are also invariant for the sub-population model, so that our choice of identifiability constraints for the reference population does not affect the suitability of the sub-population model. This requirement will determine both the nature of the set of deterministic functions of year of birth, X_y in Equation 2, and the nature of any parametric simplification imposed upon $\alpha_x^{(\Delta)}$, i.e., if $\alpha_x^{(\Delta)}$ is restricted to be

a linear combination of a set of basis functions

$$\alpha_x^{(\Delta)} = \sum_{i=1}^{n} \alpha^{(i)} g^{(i)}(x)$$

then the nature of the basis functions, $g^{(i)}(x)$, will be determined by the identifiability issues present in the model. We, therefore, consider each of the different forms that the invariant transformations of the reference model can take in turn, in order to ensure that they will not affect the sub-population model.

First, the scaling factors in the sub-population model do not depend upon the normalisation scheme of the age/period terms in the reference model. Normalisation schemes are imposed by using a transformation of the form

$$\{\hat{f}^{(R,i)}(x), \hat{\kappa}_t^{(R,i)}\} = \left\{\frac{1}{a^{(i)}} f^{(R,i)}(x), a^{(i)} \kappa_t^{(R,I)}\right\}$$

and so it is obvious that $\Lambda^{(i)} \hat{f}^{(R,i)}(x) \hat{\kappa}_t^{(R,i)} = \Lambda^{(i)} f^{(R,i)}(x) \kappa_t^{(R,i)}$.

Second, we know from Hunt and Blake (2015a) that all APC models are invariant under the transformation

$$\{\hat{\alpha}_x^{(R)}, \hat{f}^{(R,i)}(x), \kappa_t^{(R,i)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - a^{(i)} f^{(R,i)}(x), f^{(i)}(x), \kappa_t^{(R,i)} + a^{(i)}, \gamma_y^{(R)}\} \tag{1}$$

i.e., the model using the transformed parameter set gives exactly the same fitted mortality rates. This allows us to impose the "level" of the period functions, $\kappa_t^{(R,i)}$, via the identifiability constraints, such as imposing $\sum_t \kappa_t^{(R,i)} = 0$ or $\kappa_T^{(R,i)} = 0$. However, such a set of identifiability constraints is arbitrary, and so should not have any consequences for our modelling approach.

Accordingly, we require that our model in Equation 2 is also invariant if the transformed parameters are used for the reference population. In order to ensure this, we require that Equation 2 is invariant under the transformation

$$\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - a^{(i)} \Lambda^{(i)} f^{(R,i)}(x) \tag{2}$$

This transformation can be accommodated without $\alpha_x^{(\Delta)}$ fundamentally changing form if

2

1. $\alpha_x^{(\Delta)}$ is non-parametric, as in the original specification in Equation 2; or

2. if $\alpha_x^{(\Delta)}$ is restricted to be of parametric form, then $\alpha_x^{(\Delta)} = \sum_{i=1}^{N} \alpha^{(i)} f^{(i)}(x) + \sum_{i=N+1}^{n} \alpha^{(i)} g^{(i)}(x)$, i.e., the age functions in the reference model form a subset of the basis functions, $g^{(i)}(x)$.

As an example, consider the case where our model for the reference population is the "classic APC" model of Hobcraft et al. (1982)

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \kappa_t^{(R)} + \gamma_{t-x}^{(R)}$$

$$R_{x,t} = \alpha_x^{(\Delta)} + \Lambda^{(1)} \kappa_t^{(R)} + \Lambda^{(\gamma)} \gamma_{t-x}^{(R)} + \nu X_{t-x}$$

The classic APC model is invariant under the transformation

$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - a, \kappa_t^{(R)} + a, \gamma_y^{(R)}\}$$

Substituting the transformed parameters into the sub-population model gives

$$\hat{R}_{x,t} = \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)} \hat{\kappa}_t^{(R)} + \hat{\Lambda}^{(\gamma)} \hat{\gamma}_{t-x}^{(R)} + \hat{\nu} X_{t-x}$$

$$= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)} (\kappa_t^{(R)} + a) + \hat{\Lambda}^{(\gamma)} \gamma_{t-x}^{(R)} + \hat{\nu} X_{t-x}$$

In order to ensure $\hat{R}_{x,t} = R_{x,t}$, we must have $\hat{\Lambda}^{(1)} = \Lambda^{(1)}$, $\hat{\nu} = \nu$ and $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - a\Lambda^{(1)}$. The requirement that $\hat{\alpha}_x^{(\Delta)}$ is of the same form as $\alpha_x^{(\Delta)}$ implies that any parametric simplification for $\alpha_x^{(\Delta)}$ must be of the form $\alpha_x^{(\Delta)} = \alpha^{(1)} + \sum_{j=2}^{n} \alpha^{(i)} g^{(i)}(x)$, i.e., it has a constant basis function, $g^{(1)}(x) = 1$, in order that the sub-population model does not change if the levels of the period functions are transformed.

Third, the values of $\Lambda^{(i)}$ depend upon the definition of the age functions in the reference model. "Equivalent" models for the reference population, which use different definitions for the age functions but give identical fitted mortality rates, will give different values of $\Lambda^{(i)}$. To see this, consider a

3

reference model of the form[1]

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

$$R_{x,t} = \alpha_x^{(\Delta)} + \Lambda^{(1)}\kappa_t^{(R,1)} + \Lambda^{(2)}(x - \bar{x})\kappa_t^{(R,2)} + \Lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \nu X_{t-x}$$

The model for the reference population is equivalent to a model of the form

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \hat{\kappa}_t^{(R,1)} + x\hat{\kappa}_t^{(R,2)} + \gamma_{t-x}^{(R)}$$

with $\hat{\kappa}_t^{(R,1)} = \kappa_t^{(R,1)} - \bar{x}\kappa_t^{(R,2)}$ and $\hat{\kappa}_t^{(R,2)} = \kappa_t^{(R,2)}$. The corresponding sub-population model in this case is

$$\hat{R}_{x,t} = \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}\hat{\kappa}_t^{(R,1)} + \hat{\Lambda}^{(2)}x\hat{\kappa}_t^{(R,2)} + \Lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \hat{\nu}X_{t-x}$$

However, in this situation, we find that we require $\hat{\Lambda}^{(2)}x = \Lambda^{(2)}(x - \bar{x}) + \Lambda^{(1)}\bar{x}$ to give the same fitted mortality rates for both reference models. In this case, the relationship between the two is a function of age, x, which contradicts the assumption that the scaling factors are constants. Consequently, we find that the values of the scaling factors and the fit provided by the model for the sub-population will depend on the specifics of the age functions in the reference model and will differ between equivalent models.

Finally, identifiability under transformations of the cohort parameters is not as straightforward. From Hunt and Blake (2015b), we found that APC models may have unidentifiable trends which are allocated between the age/period and cohort terms by the identifiability constraints. Invariance of the mortality rates in the sub-population model to a different allocation of these trends in the reference model depends upon the deterministic regressors, X_y, we added to the sub-population model in Equation 2, and the form of any parametric simplification of $\alpha_x^{(\Delta)}$. This is illustrated by the following example.

---

[1]We call this model the "reduced Plat" model, since it was suggested in Plat (2009) as being a reduced form of the model tested in that paper that might be more suitable for high ages. This model can also be thought of as an extension to model M6 in Cairns et al. (2009), with a static age function, or as an extension to the "CBDX" model discussed in Hunt and Blake (2015a) with a cohort term.

Consider the example of the classic APC model for the reference population again. In addition to the transformation above, the classic APC model is also invariant under the following two transformations involving the cohort parameters

$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - b, \kappa_t^{(R)}, \gamma_y^{(R)} + b\}$$

$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} + c(x - \bar{x}), \kappa_t^{(R)} - c(t - \bar{t}), \gamma_y^{(R)} + c(y - \bar{y})\}$$

where $\bar{x}$ and $\bar{t}$ are defined in a similar fashion to $\bar{y}$ in Appendix 2. Invariance of the sub-population model under the first of these transformations requires $\hat{\Lambda}^{(\gamma)} = \Lambda^{(\gamma)}$ and $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - b\Lambda^{(\gamma)}$, and therefore that any parametric restriction placed upon $\alpha_x^{(\Delta)}$ must have a constant basis function, $g^{(1)}(x) = 1$, as discussed above in respect of the level of $\kappa_t^{(R)}$.

However, substituting the transformed parameters from the second transformation in Equation 2, we find

$$\hat{R}_{x,t} = \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}\hat{\kappa}_t^{(R)} + \hat{\Lambda}^{(\gamma)}\hat{\gamma}_{t-x}^{(R)} + \hat{\nu}X_{t-x}$$
$$= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}(\kappa_t^{(R)} - c(t - \bar{t})) + \hat{\Lambda}^{(\gamma)}(\gamma_{t-x}^{(R)} + c((t - \bar{t}) - (x - \bar{x}))) + \hat{\nu}X_{t-x}$$

In order to have $\hat{R}_{x,t} = R_{x,t}$, we require

- $\hat{\Lambda}^{(j)} = \Lambda^{(j)}$, i.e., that our sensitivities do not change from one set of identifiability conditions to any other;

- $\hat{\nu}X_y = \nu X_y - c(\lambda^{(\gamma)} - \lambda^{(1)})(y - \bar{y})$, i.e., we can add terms linear in year of birth to the deterministic term without it fundamentally changing form, and therefore that our deterministic regressors contain a linear trend in year of birth; and

- $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - c\lambda^{(1)}(x - \bar{x})$, i.e., we can add linear functions to any parametric form for $\alpha_x^{(\Delta)}$ without it fundamentally changing form, and therefore that it must be either non-parametric or have a linear function of age, $g^{(2)}(x) = x - \bar{x}$, amongst the basis functions used in any parametric restriction.

In addition to the identifiability issues discussed here, it is also important that any parametric simplification for $\alpha_x^{(\Delta)}$ consists of more than one, constant term. As discussed in Tuljapurkar and Edwards (2009), multiple terms

in $\alpha_x^{(\Delta)}$ allow higher moments of the observable distribution of deaths in the sub-population (such as the variance of age at death) to be captured by the model, as well as the difference in life expectancy between the two populations. These higher moments are important in the allowance for idiosyncratic risk in the sub-population, which is likely to be important in many circumstances, such as those discussed in Hunt and Blake (2016).

We also see from the analysis above that the form of our deterministic regressors, X_y, will depend upon the mortality model being used for the reference population. From Hunt and Blake (2015b), if the model for the reference population contains age functions which span the polynomials to order p, then there will be unidentified polynomial trends in the cohort parameters of order $p + 1$. We must therefore ensure that the deterministic regressors in Equation 2 span the polynomials to order $p + 1$ and that any parametric simplification for the age function, $\alpha_x^{(\Delta)}$, also contains a basis function of the form $g^{(i)}(x) = x^{p+1}$.

For the classic APC model, $p = 0$ and therefore we would require that the deterministic regressors and age function are, at least, of linear order. Similarly, for the reduced Plat model and the model constructed by the general procedure in Section 4.1 and Online Appendix 2, $p = 1$ and we require that the deterministic regressors are at least of quadratic order.

In summary, the identifiability issues present in APC mortality models and discussed in Hunt and Blake (2015a,b) have important consequences for the mortality modelling approach used in this study. Most importantly, we require an additional νX_y term in the model and must be careful when specifying any parametric simplification for $\alpha_x^{(\Delta)}$, in order to ensure that our results do not depend on the arbitrary identifiability constraints we impose on the reference model. In the context of the reference model used in this study, described in Section 4.1, this means that we need the term

$$\nu X_y = \nu_1(y - \bar{y}) + \nu_2 \left((y - \bar{y})^2 - \sigma_y\right)$$

in Equation 2, and any parametric simplification of $\alpha_x^{(\Delta)}$ must be of the form

$$\alpha_x^{(\Delta)} = \begin{pmatrix} \alpha^{(1)}, & \alpha^{(2)}, & \alpha^{(3)}, & \alpha^{(4)}, & \alpha^{(5)} \end{pmatrix} \begin{pmatrix} f^{(1)}(x) \\ f^{(2)}(x) \\ f^{(3)}(x) \\ f^{(4)}(x) \\ ((x - \bar{x})^2 - \sigma_x) \end{pmatrix}$$

$$= \sum_{i=1}^{N+1} \alpha^{(i)} \tilde{f}^{(R,i)}(x)$$

where $f^{(i)}(x)$ are the parametric age functions in the reference model, described in Table 1.

2 Models constructed by the "general procedure" for the UK

In Hunt and Blake (2014), a "general procedure" for constructing mortality models tailored to the specific features of individual datasets was proposed. In outline, this

- starts from a simple static mortality model with a non-parametric static age function;

- sequentially adds age/period terms to the model to detect and capture the age/period structure in the data:

  - structure is detected by adding a non-parametric age/period term which will identify the feature explaining the largest proportion of the remaining structure in the data;

  - then this term is simplified into a parametric form which identifies the same feature more parsimoniously and with greater demographic significance;

  - then the statistical significance and robustness of the term is tested;

- finally adds a cohort term once all age/period structure has been captured by the model;

- tests the standardised deviance residuals of the model for any remaining structure, independence, and normality.

This procedure was applied to data from the Human Mortality Database (2014) for men in the UK for ages 50 to 100 and years 1950 to 2011 in order to construct mortality models capable of capturing all the relevant information in the data and therefore allowing it to be projected appropriately.

A brief description of the terms in the models and their demographic significance is given in Table 1 in the main report. A fuller list of the parametric age functions in the "toolkit" developed as part of the general procedure is given in the Appendix of Hunt and Blake (2014).

Models generated by the GP impose the following standard identifiability constraints

$$\sum_{x=50}^{100} |f^{(R,i)}(x)| = 1 \; \forall i\}$$
(3)

on the age functions to ensure that they have a consistent normalisation scheme. This enables us to compare the magnitudes of the period functions both with each other and between populations and gauge their relative importance.[2]

In order to assist the visual comparison between the UK and SAPS data (the latter of which only spans ages 60 to 90 and years 2000 to 2011), we impose the following constraint on the period functions

$$\sum_{t=2000}^{2011} \kappa_t^{(R,i)} = 0 \; \forall i$$
(4)

This means that the period functions represent deviations from an "average" level of mortality in the period covered by the SAPS data, rather than over the whole period of the UK data.

The results of Hunt and Blake (2015b) also indicate that we need to impose constraints on the levels, linear and quadratic trends present in the

---

[2]For both women and men, the second and third age/period terms use age functions which are "self-normalising" in the sense of Hunt and Blake (2015a).

cohort parameters. To identify their levels, we impose the following constraints on the cohort parameters for each of the reference populations

$$\sum_{y=1910}^{1951} n_y^{(S)} \gamma_y^{(R)} = 0$$

where $n_y^{(S)}$ is the number of observations of each cohort in the SAPS data. As with the period functions, this means that the cohort parameters should be centred around zero over the range of the SAPS data, not the full range of the data covered for the UK population. To constrain the linear and quadratic trends in the cohort parameters, we impose

$$\sum_{y=1850}^{1961} n_y^{(R)} \gamma_y^{(R)} (y - \bar{y}) = 0 \tag{5}$$

$$\sum_{y=1850}^{1961} n_y^{(R)} \gamma_y^{(R)} \left((y - \bar{y})^2 - \sigma_y \right) = 0 \tag{6}$$

where $n_y^{(R)}$ is the number of observations of each cohort in the UK national data, $\bar{y} = 0.5(X + T - 1)$ and $\sigma_y = \frac{1}{X+T-1} \sum_y (y - \bar{y})^2$.

The justification for these constraints is that they allow us to remove linear and quadratic trends in the cohort parameters. This makes them conform better to the demographic significance for cohort parameters described in Hunt and Blake (2015c), namely that the cohort parameters should not have any long-term systematic trends. We impose this over the whole range of the UK data, which is considerably longer than the range covered by the SAPS data, since there appear to be short-term trends (lasting for a few decades) which are then reversed out over a longer time horizon. However, this means that over the shorter range of years of birth covered by the SAPS data, the cohort parameters appear to have strong trends.

It is important to note, however, that our demographic significance for the parameters is highly subjective and our choice of constraints is arbitrary. We have therefore taken appropriate steps in Online Appendix 1 to ensure that our choice of identifiability constraints does not affect either the mortality rates fitted by the model or our overall conclusions.

When fitting the final models, we obtain the parameters shown in Figure 1. This model has a BIC of -1.96×10^4, with 407 free parameters for both populations.[3] We also test the standardised deviance residuals from fitting the model as part of the general procedure. The moments of the residuals and a Jarque-Bera test of their normality is given in Table 1.

	Residual mean	Standard deviation	Residual skewness	Residual kurtosis	Jarque-Bera statistic
Men	0.00	0.93	-0.05	3.15	4.84

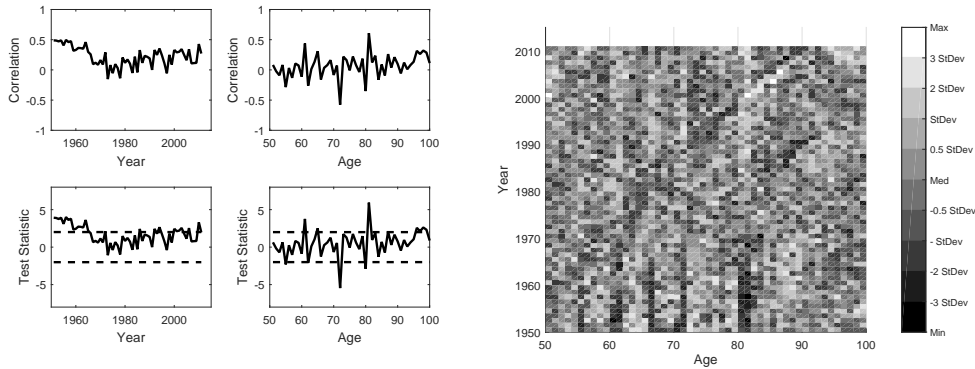Table 1: Moments of the residuals of the reference model

We can see that the residuals are close to normal and therefore pass the relevant Jarque-Bera test for normality at the 5% level (p-value of 8.7%), although they are slightly leptokurtic. We also see from Figure 1a that there appears to be relatively little correlation structure over consecutive ages, although the residuals show significant autocorrelations during the early part of the data range, which diminishes towards the end of the period of the data. The heat maps for the residuals shown in Figure 1b indicate that the residuals have very little remaining structure in them. There is possibly some remaining structure around age 80, although this appears to be specific to only a few neighbouring years and therefore it is difficult to add an age/period term to capture this without overfitting the model.

References

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

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

---

[3]For comparison, the Lee and Carter (1992) model fitted to the same data obtain a BIC of -2.71×10^4 with 161 free parameters.

(a) Correlations for sequential years and ages of the residuals

(b) Heat maps of the residuals

Figure 1: Tests of the residuals from the reference model

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

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

Hunt, A., Blake, D., 2015a. Identifiability in age/period mortality models. Tech. rep., Pensions Institute PI 15-08, Cass Business School, City University London.

Hunt, A., Blake, D., 2015b. Identifiability in age/period/cohort mortality models. Tech. rep., Pensions Institute PI 15-09, Cass Business School, City University London.

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

Hunt, A., Blake, D., 2016. Transferring mortality and longevity risks in pension schemes via bespoke longevity swaps. Work in Progress.

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

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

Tuljapurkar, S., Edwards, R., 2009. Variance in death and its implications for modeling and forecasting mortality. NBER Working Papers.