Identifiability, Cointegration and the Gravity Model

Andrew Hunt[*]

Cass Business School, City University London

Corresponding author: `andrew.hunt.1@cass.city.ac.uk`

David Blake

Pensions Institute, Cass Business School, City University London

July 2017

Abstract

The gravity model of Dowd et al. (2011) was introduced in order to achieve coherent projections of mortality between two related populations. However, this model as originally formulated is not well-identified since it gives projections which depend on the arbitrary identifiability constraints imposed on the underlying mortality model when fitting it to data. In this paper, we discuss how the gravity model can be modified to give well-identified projections of mortality rates and how this result can be generalised to more complicated mortality models.

JEL Classification: C32

Keywords: Mortality modelling, age/period/cohort models, multi-population modelling, coherent mortality projection

1

1   Introduction

Hunt and Blake (2015b) and Hunt and Blake (2015c) discussed the issue of identifiability in single population age/period/cohort (APC) mortality models, and in particular how to obtain projections of mortality rates which do not depend upon the arbitrary identifiability constraints imposed.

Issues with identifiability in projections also exist if we project mortality for multiple populations rather than just one. Such multi-population projections are vital in order to allow for the correlations and dependencies between related populations that are influenced by similar biological and socio-economic drivers of changing mortality. It is essential that, in such a model, our projections do not depend on the arbitrary identifiability constraints imposed when fitting the model, but only on the underlying drivers of mortality evolution.

Many multi-population mortality models go beyond merely allowing for covariation between the stochastic evolution of mortality in different populations, and instead impose the stronger assumption of "coherence", i.e., that mortality rates in different populations should not diverge with time. Such an imposition is popular and intuitively appealing; however, we find that it usually cannot be imposed on a model in a fashion which does not depend on the arbitrary identifiability constraints. In addition, it can often lead to overriding the evidence from the historical data in order to impose our preconceptions on projected mortality rates in a manner which we consider to be unscientific.

One model designed to achieve coherent projections of mortality between two populations is the "gravity model" of Dowd et al. (2011). This model adopted a cointegration framework to project the period and cohort terms from the classic APC model fitted to each population. However, as originally formulated, the model is not well-identified, since the projections from it depend on the identifiability constraints imposed when fitting the classic APC model. Later work by Zhou et al. (2014) applied the framework of the gravity model to the period terms from the Lee-Carter model (Lee and Carter (1992)) and avoided some of the issues present in the original model of Dowd et al. (2011). However, this new form of the model is still not well-identified, since it gives projections dependent upon the identifiability constraints im-

posed by the user.

In this study, we discuss the issue of identifiability in cointegration models and apply this to the specific context of the gravity model in order to obtain a well-identified model. Section 2 discusses the classic APC model which was used in Dowd et al. (2011) to fit mortality rates in both populations. Section 3 outlines the gravity model introduced in Dowd et al. (2011) and places it in the context of more general cointegration models. Section 4 discusses why the gravity model is not well-identified and how it can be modified to give well-identified projections. Section 5 discusses the model of Zhou et al. (2014), how it differs from the gravity model of Dowd et al. (2011) and the issues with identifiability which are still present. Finally, Section 6 generalises these results to a broader class of mortality models and Section 7 concludes.

2   Identifiability in the classic APC model

The simplest APC model (referred to here as the "classic APC model") has a long history and is widely used in the fields of medicine, epidemiology and sociology as well as in demography and actuarial science. It has the form in Equation 1[1]

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

The parameters in the classic APC model cannot be estimated uniquely by reference to the data alone. A model is fully identified when all the parameters in it can be uniquely determined by reference to the available data. In contrast, the classic APC model is not fully identified because there exist different sets of parameters which will give the same fitted mortality rates and consequently the same goodness of fit for any data set.

We can see that this model is not fully identified, since if we use the transformations in Equations 2, 3 and 4 to obtain new sets of parameters, we do not change our fit to the data (we call such transformations "invariant"

---

[1]In this paper, we assume that ages, x, are in the range $[1, X]$ and periods, t, are in the range $[1, T]$ and therefore that years of birth, y, are in the range $(1 - X)$ to $(T - 1)$.

for this reason)

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - a, \kappa_t + a, \gamma_y\} \tag{2}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - b, \kappa_t, \gamma_y + b\} \tag{3}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x + cx, \kappa_t - ct, \gamma_y + cy\} \tag{4}$$

Because different sets of parameters give the same fit to the data, we cannot use the data to choose between them. Typically, we impose identifiability constraints on the parameters in order to specify them uniquely. For instance, a commonly used set of identifiability constraints is $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = 0$.[2] We refer to these identifiability constraints as "natural", since they allow us to impose our interpretation of the demographic significance[3] of the parameters onto the model. For example, the first two of these constraints mean that α_x can be interpreted as an "average" level of mortality at age x over the period, with κ_t and γ_y representing deviations from this average level. The third constraint requires that there are no deterministic linear trends within the fitted cohort parameters, since any linear trend has been arbitrarily assigned to the age and period effects. This is in line with the demographic significance we assign to the cohort parameters in Hunt and Blake (2015e), namely that the cohort parameters should be centred around zero and should not show any long term trends. This means that cohort effects are interpreted as deviations in the mortality experienced by one cohort relative to that of adjacent years of birth.

However, it is important to note that these additional identifiability constraints, although having a natural interpretation, are arbitrary and ad hoc. While they might allow us to interpret the parameters in terms of their demographic significance, this interpretation nevertheless depends entirely on the user's judgement rather than on the underlying data. Of specific importance in the context of this study, Dowd et al. (2011) used the constraints $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and a third constraint described in terms of minimising a tilting parameter δ, which can be written as $\sum_{x,t}(x - \bar{x})\gamma_{t-x} = 0$. The

---

[2] Here n_y is the number of observations of cohort y in the data and so $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x}$, and a bar denotes the arithmetic mean of the variable over the relevant data range, e.g., $\bar{y} = \frac{1}{X+T-1}\sum_y y = 0.5(X + T)$.

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

impact of using either the "natural" or the Dowd et al. (2011) identifiability constraints when making projections is assessed in Section 4.3.

Since the identifiability constraint we choose to impose are arbitrary and do not affect the historical fitted mortality rates, they should also not affect the future projected mortality rates either. In consequence, we should obtain the same projected mortality rates for any set of identifiability constraints, including but not limited to the two discussed above. We say that models with this property are "well-identified".

3   The gravity model

The "gravity" model was introduced in Dowd et al. (2011) in order to obtain mortality projections for two different populations which do not diverge with time.[4] This model might be appropriate for a small population, such as the lives in an annuity book or pension scheme, which is a subpopulation of a much larger population, such as a national population. The analogy the authors use is of the smaller population being like a planet in orbit around a star (the larger population).

The gravity model requires that the classic APC model of Equation 1 is fitted to two populations[5] and the period functions projected using

$$\kappa_t^{(I)} = \nu^{(I)} + \kappa_{t-1}^{(I)} + \epsilon_t^{(I)}$$
$$\kappa_t^{(II)} = \nu^{(II)} + \kappa_{t-1}^{(II)} + \phi(\kappa_{t-1}^{(I)} - \kappa_{t-1}^{(II)}) + \epsilon_t^{(II)} \tag{5}$$

The parameter $\phi \in [0, 1)$ is designed to ensure that the difference, $\kappa_t^{(I)} - \kappa_t^{(II)}$, is stationary and, therefore, the period functions in the different populations

---

[4]This model is functionally equivalent to the model in Cairns et al. (2011), which differs only in the presentation of the model and the techniques used to fit it to data. Therefore, the comments made in this note for the gravity model are also applicable to the model of Cairns et al. (2011).

[5]In Dowd et al. (2011), these were referred to as populations 1 and 2, with the period and cohort functions numbered accordingly. To avoid confusion with the different period functions $\kappa_t^{(i)}$ for models with more than one age/period term fitted to a single population, we shall refer to the populations as I and II and label the period functions $\kappa_t^{(I)}$ and $\kappa_t^{(II)}$ respectively.

do not diverge.

We can rewrite Equation 5 as

$$\Delta \begin{pmatrix} \kappa_t^{(I)} \\ \kappa_t^{(II)} \end{pmatrix} = \begin{pmatrix} \nu^{(I)} \\ \nu^{(II)} \end{pmatrix} + \begin{pmatrix} 0 \\ \phi \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \kappa_{t-1}^{(I)} \\ \kappa_{t-1}^{(II)} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{(I)} \\ \epsilon_t^{(II)} \end{pmatrix} \tag{6}$$

This model is just a special case of a more general cointegration model, although this interpretation was not commented upon in Dowd et al. (2011). A number of papers have suggested or implemented cointegration as a means of projecting the period parameters of mortality models for different populations. Cointegration was first suggested in the work of Carter and Lee (1992), but was more recently used in the modelling of Li and Hardy (2011) and Yang and Wang (2013).

Cointegration between the period functions requires that we model the vector of time series processes as

$$\Delta \boldsymbol{\kappa}_t = \nu X_t + \sum_{i=1}^{p-1} \Gamma_i \Delta \boldsymbol{\kappa}_{t-i} + \Pi \boldsymbol{\kappa}_{t-p} + \boldsymbol{\epsilon}_t \tag{7}$$

The rank of the matrix Π is then tested in order to identify the number of cointegrating relationships between the period functions in the model. If it is of rank $r < N$ (the number of period functions in $\boldsymbol{\kappa}_t$), then Π can be decomposed as $\Pi = \alpha\beta^\top$, where α and β are $N \times r$ matrices to give the interpretation that the rows of $\beta^\top \boldsymbol{\kappa}_{t-p}$ represent r stationary cointegrating relationships between the different period functions. In order to use cointegration robustly, we need to ensure that any statements we make about the rank of Π are independent of our choice of identifiability constraints.

We can therefore see that the gravity model in Equation 6 has the same form as Equation 7, with $p = 1$, $X_t = \begin{pmatrix} 1 \end{pmatrix}$, $r = 1$, $\alpha = \begin{pmatrix} 0, & \phi \end{pmatrix}^\top$ and $\beta = \begin{pmatrix} 1, & -1 \end{pmatrix}^\top$. The prescribed form for β imposes that there is a stationary cointegrating relationship of the form $\kappa_t^{(I)} - \kappa_t^{(II)} = Z_t$, and so ensures that relative mortality rates will not diverge between the two populations, whilst the prescribed form for α allows the interpretation that population I is dominant and so has no dependence on population II.

A related process was used in Dowd et al. (2011) to project the cohort parameters from the model. This can be written as

$$\Delta \begin{pmatrix} \gamma_y^{(I)} \\ \gamma_y^{(II)} \end{pmatrix} = \begin{pmatrix} \mu^{(I)} \\ \mu^{(II)} \end{pmatrix} + \begin{pmatrix} \alpha^{(I)} & 0 \\ 0 & \alpha^{(II)} \end{pmatrix} \Delta \begin{pmatrix} \gamma_{y-1}^{(I)} \\ \gamma_{y-1}^{(II)} \end{pmatrix} + \begin{pmatrix} 0 \\ \phi \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \gamma_{y-1}^{(I)} \\ \gamma_{y-1}^{(II)} \end{pmatrix} + \begin{pmatrix} \varepsilon_y^{(I)} \\ \varepsilon_y^{(II)} \end{pmatrix}$$

(8)

We can therefore see that this is also similar to the cointegration relationship in Equation 7.[6]

4 Identifiability in the gravity model

4.1 Period functions

The values of $\kappa_t^{(I)}$ and $\kappa_t^{(II)}$ are not uniquely identifiable by the classic APC model, but instead depend upon our choice of identifiability constraints. Equations 2 and 4 give us the freedom to add linear trends in time to either or both time series independently, i.e.

$$\begin{pmatrix} \hat{\kappa}_t^{(II)} \\ \hat{\kappa}_t^{(II)} \end{pmatrix} = \begin{pmatrix} \kappa_t^{(I)} \\ \kappa_t^{(II)} \end{pmatrix} + \begin{pmatrix} a^{(I)} \\ a^{(II)} \end{pmatrix} + \begin{pmatrix} c^{(I)} \\ c^{(II)} \end{pmatrix} t$$

$$\hat{\boldsymbol{\kappa}}_t = \boldsymbol{\kappa}_t + \boldsymbol{a} + \boldsymbol{c}t$$

(9)

However, this transformation, despite leaving the fitted mortality rates unchanged if we make the appropriate offsets to the static age functions and cohort parameters, fundamentally alters the cointegration relationship in Equa-

---

[6] There is a slight difference between Equation 8 and the standard form of the cointegration relationship in Equation 7, in that Equation 8 involves a stationary term in $\boldsymbol{\gamma}_{y-1} = \left(\gamma_{y-1}^{(I)}, \quad \gamma_{y-1}^{(II)} \right)^\top$ rather than $\boldsymbol{\gamma}_{y-2}$. This could be solved by rearranging Equation 8 using $A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\beta^\top\boldsymbol{\gamma}_{y-1} = (A + \alpha\beta^\top)\Delta\boldsymbol{\gamma}_{y-1} - \alpha\beta^\top\boldsymbol{\gamma}_{y-2}$ and redefining the matrix A. However, this solution involves losing the particular structure imposed upon A in Dowd et al. (2011).

tion 6 since

$$\Delta \hat{\boldsymbol{\kappa}}_t = \Delta \boldsymbol{\kappa}_t + \boldsymbol{c}$$
$$= \boldsymbol{\nu} + \alpha \beta^\top \boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t + \boldsymbol{c}$$
$$= \boldsymbol{\nu} + \boldsymbol{c} - \alpha \beta^\top (\boldsymbol{a} + \boldsymbol{c}t) + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t$$
$$= \hat{\boldsymbol{\nu}} - \alpha \beta^\top \boldsymbol{c}t + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t$$

The transformed time series has a deterministic linear term, $\alpha \beta^\top \boldsymbol{c}t$, which was not present in the original parameterisation. This means that the time series structure in Equation 6 is not well-identified. In practice, this has the consequence that the gravity model can be difficult to fit to historical time series and may give implausible values.

We might conjecture that a solution to this problem would be to allow for deterministic trends up to linear order in the cointegrating relationship, i.e., using $\boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t$ in place of $\boldsymbol{\nu}$ in Equation 6 to give

$$\Delta \begin{pmatrix} \kappa_t^{(I)} \\ \kappa_t^{(II)} \end{pmatrix} = \begin{pmatrix} \nu_0^{(I)} \\ \nu_0^{(II)} \end{pmatrix} + \begin{pmatrix} \nu_1^{(I)} \\ \nu_1^{(II)} \end{pmatrix} t + \begin{pmatrix} 0 \\ \phi \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \kappa_{t-1}^{(I)} \\ \kappa_{t-1}^{(II)} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{(I)} \\ \epsilon_t^{(II)} \end{pmatrix}$$
$$\Delta \boldsymbol{\kappa}_t = \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \alpha \beta^\top \boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \qquad (10)$$

Such a model is well-identified as it does not change form under the transformation in Equation 9

$$\Delta \hat{\boldsymbol{\kappa}}_t = \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \boldsymbol{c} - \alpha \beta^\top (\boldsymbol{a} + \boldsymbol{c}t) + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t$$
$$= \hat{\boldsymbol{\nu}}_0 + \hat{\boldsymbol{\nu}}_1 t + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t$$
$$\hat{\boldsymbol{\nu}}_0 = \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha \beta^\top \boldsymbol{a}$$
$$\hat{\boldsymbol{\nu}}_1 = \boldsymbol{\nu}_1 - \alpha \beta^\top \boldsymbol{c}$$

However, because we have the first difference of the time series on the left-hand side of Equation 10, when we integrate this equation, we obtain quadratic trends in the levels of the period functions. This is undesirable as we do not generally observe quadratic trends in the fitted parameters and they might change direction when projected into the future with near certainty for no compelling biological reason. Therefore, the model in Equation

8

10 conflicts with our desire for biologically reasonable[7] projections.

There is, however, a way to obtain both biological reasonableness and identifiability under the transformations in Equation 4. This is to restrict the linear deterministic trend in Equation 10 by imposing $\boldsymbol{\nu}_1 = \alpha\beta_1$ where β_1 is an arbitrary constant. This will ensure that the relevant deterministic trend is present in $\boldsymbol{\kappa}_t$, but is constrained within the stationary cointegrating relationships and is not present in the non-stationary part of the relationship.

This means that we need to include constrained deterministic linear trends in the cointegrating relationship, but leave an unconstrained constant term, i.e.

$$
\begin{aligned}
\Delta\boldsymbol{\kappa}_t &= \boldsymbol{\nu}_0 + \alpha\beta_1 t + \alpha\beta^\top\boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{\nu}_0 + \alpha(\beta^\top\boldsymbol{\kappa}_{t-1} + \beta_1 t) + \boldsymbol{\epsilon}_t
\end{aligned}
\tag{11}
$$

To see that this structure is well-identified under the transformations in Equation 4, let us transform the parameters using Equation 9 to obtain

$$
\begin{aligned}
\Delta\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha\beta^\top\boldsymbol{a} + \alpha\left(\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + (\beta_1 - \beta^\top\boldsymbol{c})t\right) + \boldsymbol{\epsilon}_t \\
&= \hat{\boldsymbol{\nu}}_0 + \alpha\left(\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + \hat{\beta}_1 t\right) + \boldsymbol{\epsilon}_t
\end{aligned}
$$

where $\hat{\boldsymbol{\nu}}_0 = \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha\beta^\top\boldsymbol{a}$, as previously, and $\hat{\beta}_1 = \beta_1 - \beta^\top\boldsymbol{c}$. This model also gives biologically reasonable values for ϕ which do not depend upon the identifiability constraints imposed when fitting the models, as demonstrated in Section 4.3.

4.2 Cohort parameters

As with the period parameters, the values of $\gamma_y^{(I)}$ and $\gamma_y^{(II)}$ are not uniquely identifiable in the classic APC model, but instead depend upon our choice of identifiability constraints. Equations 3 and 4 give us the freedom to add

---

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

linear trends in time to either or both time series independently, i.e.

$$\begin{pmatrix} \hat{\gamma}_y^{(II)} \\ \hat{\gamma}_y^{(II)} \end{pmatrix} = \begin{pmatrix} \gamma_y^{(I)} \\ \gamma_y^{(II)} \end{pmatrix} + \begin{pmatrix} b^{(I)} \\ b^{(II)} \end{pmatrix} + \begin{pmatrix} c^{(I)} \\ c^{(II)} \end{pmatrix} y$$

$$\hat{\boldsymbol{\gamma}}_y = \boldsymbol{\gamma}_y + \boldsymbol{b} + \boldsymbol{c}y \tag{12}$$

Rewriting Equation 8 in the form

$$\Delta\boldsymbol{\gamma}_y = \boldsymbol{\mu} + A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\beta^\top\boldsymbol{\gamma}_{y-1} + \boldsymbol{\varepsilon}_y$$

we see that this is also not well-identified as it changes form under the transformation in Equation 12

$$\Delta\hat{\boldsymbol{\gamma}}_y = \boldsymbol{\mu} + \boldsymbol{c} - A\boldsymbol{c} - \alpha\beta^\top (\boldsymbol{b} + \boldsymbol{c}y) + A\Delta\hat{\boldsymbol{\gamma}}_{y-1} + \alpha\beta^\top\hat{\boldsymbol{\gamma}}_{y-1} + \boldsymbol{\varepsilon}_y$$

as the transformed drift term, $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{c} - A\boldsymbol{c} - \alpha\beta^\top (\boldsymbol{b} + \boldsymbol{c}y)$, is now a linear function in year of birth, y.

However, in the same manner as used for the period parameters above, we can introduce a constrained linear trend into the cointegrating relationship in order to give well-identified projections which are biologically reasonable

$$\Delta\boldsymbol{\gamma}_y = \boldsymbol{\mu} + A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\left(\beta^\top\boldsymbol{\gamma}_{y-1} + \tilde{\beta}_1 y\right) + \boldsymbol{\varepsilon}_y \tag{13}$$

This can be shown to be well-identified by transforming the cohort parameters in a similar fashion.

4.3 Application to England & Wales and CMI Assured Lives data

In order to illustrate how the original gravity model gives projections of mortality which depend upon the identifiability constraints chosen, we apply the gravity model to the same data used in Dowd et al. (2011), i.e., the dominant population is the combined populations of England & Wales and the subordinate population is that of assured lives in the UK as recorded by the Continuous Mortality Investigation, i.e., those people who purchase life assurance policies with UK insurance companies. In both cases, we use data

for ages 50 to 90 and years 1947 to 2006.[8]

We start by fitting the classic APC model to the data.[9] In doing so, we have a choice over the identifiability constraints imposed on the models for England & Wales and the CMI Assured Lives. We investigate four different sets of identifiability constraints, which were used for the classic APC model in Hunt and Blake (2015c), i.e.,

Case 1: $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x} = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = \sum_{x,t} \gamma_{t-x}((t - \bar{t}) - (x - \bar{x})) = 0$.

Case 2: $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y (y - \bar{y}) = 0$.

Case 3: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(x - \bar{x}) = 0$.

Case 4: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(t - \bar{t}) = 0$.

We investigate the constraints shown in Case 1 and Case 3 as they are the "natural" constraints and the constraints used in Dowd et al. (2011), respectively, as discussed in Section 2. The constraints in Case 2 are similar to those in Case 1, except that the summations are taken over each year of birth rather than over all ages and years in the dataset. This has the effect of moving from a weighted average of the cohort parameters being equal to zero (with the weights determined by the number of observations for each cohort) in Case 1 to a simple arithmetical average in Case 2, and similarly for the linear trend. Although not used for the classic APC model, similar constraints were imposed on the cohort term in Model M6 in Cairns et al. (2009) and so have been included for comparison. As discussed in Section 2, the logic underpinning the selection of the Case 3 constraints in Dowd

---

[8]Data for England & Wales is taken from Human Mortality Database (2014) and we are indebted to the Continuous Mortality Investigation for providing providing the CMI Assured Lives dataset.

[9]To do this, we use a two-step procedure to fit the model for simplicity, i.e., we fit the classic APC model to the data first, and then fit the time series process to the fitted parameters using a least squares approach. This is in contrast with the approach used in Dowd et al. (2011), where a one-step method is used. However, Dowd et al. (2011) introduce additional parameters into the one-step method in a Bayesian-type approach whose purpose appears to be to constrain the value of ϕ and prevent it taking values which are not biologically reasonable. However, as discussed later, this issue arises because the gravity model is not well-identified and therefore should not be necessary in a well specified model.

et al. (2011) was that the static age function in the model should explain all the observed linearity across ages. We can apply similar logic to the period function in the classic APC model, i.e., that the period function, κ_t should explain all of the observed linearity across time, to give the constraints in Case 4.

It is important to note that all four sets of constraints were developed to give the same demographic significance to the cohort parameters, i.e., that they should be centred around zero and the other functions in the model should capture any linear trends. Because of this, these four sets of constraints give very similar sets of fitted parameters when these are plotted. These sets of parameters also give identical fitted mortality rates, since they can be transformed into each other using Equations 2, 3 and 4. However, the different sets of parameters are not identical. We therefore see that demographic significance, whilst helpful in selecting an appropriate set of identifiability constraints, does not specify a single, unique set of constraints to use. Model users with the same interpretation of the parameters can reasonably choose to impose different constraints and obtain different fitted parameters when using the same model with the same data. Furthermore, the fact that demographic significance is subjective and, in practice, different model users adopt a range of interpretations for the different parameters highlights the fact that we must take care to ensure that the projected mortality rates are independent of the arbitrary choice of constraints made when fitting the model, and underscores the extent to which the identifiability constraints we choose is arbitrary.

In each case, we apply the same identifiability constraints to both populations. Figure 1 shows the fitted values of $\kappa_t^{(I)} - \kappa_t^{(II)}$ using the Case 1 constraints. Dowd et al. (2011) assumed that these differences are stationary, however, Figure 1 shows that they have a clear linear trend which would bias the estimation of ϕ in the original specification of the model. Since the magnitude and direction of this trend is dependent upon the identifiability constraints imposed, the degree of this bias is dependent upon our choice of identifiability constraints. However, the modified gravity model allows for the potential presence of a linear trend in the cointegrating relationship and therefore any estimates for ϕ will not be biased by such a trend.

To demonstrate this numerically, for each set of fitted period parameters,
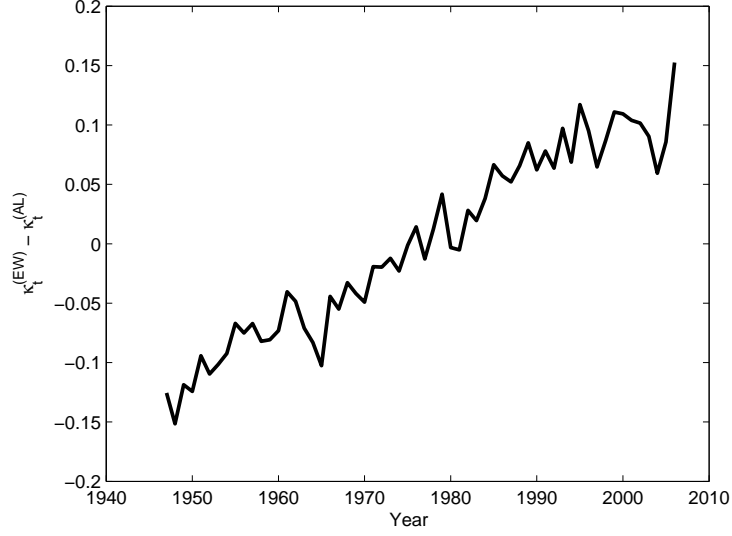
12

Figure 1: Difference between the period functions

we then first fit the original gravity model in Equation 6 and then the modified model in Equation 11. We pay particular attention to the estimated value of ϕ found, as this will determine the rate at which divergence between the two populations mean reverts.

	Original gravity model	Modified gravity model
Case 1	0.0706	0.3234
Case 2	0.0702	0.3234
Case 3	0.0701	0.3234
Case 4	0.0700	0.3234

Table 1: Values of ϕ for different identifiability constraints

The results shown in Table 1 indicate that the rate of mean reversion (and therefore the distribution of projected mortality rates) is dependent upon the identifiability constraints using the original gravity model, whereas this is not the case for the modified model. The differences between the cases for the original model appear relatively small. However, this is because the four sets of identifiability constraints used were selected on the basis of the

13

same demographic significance for the parameters and therefore the fitted parameters were broadly comparable. This will not necessarily always be the case, as demographic significance is subjective and different model users may have very different understandings as to the interpretation of the parameters.

The most important point is not how small the differences are but that they are different at all. The identifiability constraints made no difference to the the fitted mortality rates for the different cases - they were *identical*. However, the distribution of the projected mortality rates depends upon ϕ, which varies between the four cases in the original specification of the model. Therefore, the projected mortality rates would depend upon the choice of identifiability constraints. This is inconsistent with the fitting stage, where the choice of identifiability constraints made no difference to the fitted mortality rates. By contrast, the modified gravity model avoids this, as shown by the fitted value of ϕ being identical in all four cases in Table 1.

In particular, we note that it is possible that some sets of identifiability constraints for the classic APC model would give values of ϕ in the original gravity model which were greater than unity or less than zero. Therefore, the arbitrary choice of identifiability constraint may lead to diverging projections of mortality in the original gravity model, despite having the same historical fitted mortality rates as the cases shown. This is clearly something which should be avoided by use of the modified gravity model.

It is also interesting to note that the modified gravity model gives values for ϕ which are considerably larger than in the original model. This is because the parameter now captures the genuine reversion between the period functions (i.e., the saw-tooth pattern in Figure 1) without additionally trying to capture the linear trend.

These modifications make a significant difference to the projected parameters when using the gravity model, as shown in Figure 2 using the Case 1 identifiability constraints.

As can be seen in Figure 2a, the original specification of the gravity model adjusts the central trend so that there is a sharp change of trend in the CMI population at the point where the projections begin. In contrast, the modified gravity model in Figure 2b allows the trends observed in either

14

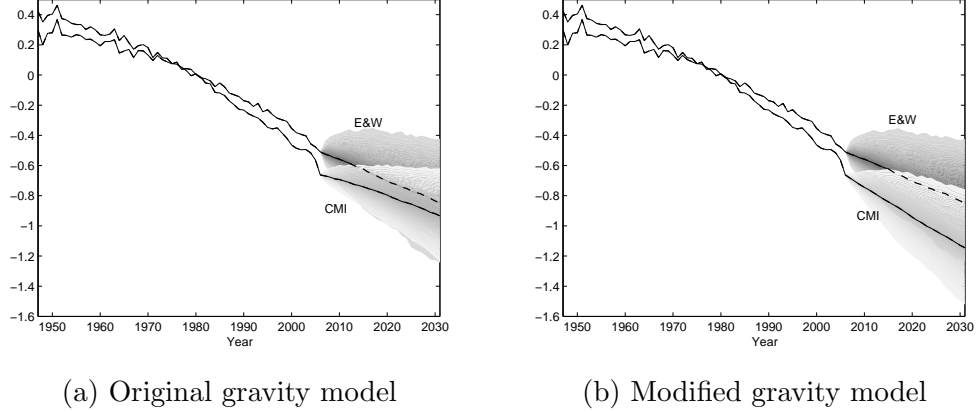(a) Original gravity model   (b) Modified gravity model

Figure 2: Projected period parameters

population to continue in future.

The change in trend exhibited by the period parameters in the original gravity model is explained by the transition between the past, where the linear trends are diverging in the period parameters fitted to the historical data, and the future, where the gravity model is forcing them together. Since the linear trends in the fitted period parameters were unidentifiable and, hence, entirely dependent upon the identifiability constraints imposed upon the model, the magnitude of the trend change also depends solely upon the arbitrary identifiability constraints. Therefore, the existence of such a trend change is not well-identified and leads to projected mortality rates which depend upon the identifiability constraints chosen, unlike the fitted mortality rates.

Furthermore, the existence of such a trend change leads to inconsistencies between the past and the future. This is not compatible with the extrapolative approach to projecting mortality, as discussed in Hunt and Blake (2015a). Although there might be insufficient evidence in the historical data to support the existence of changes in trend in the fitted period parameters, the original gravity model imposes a trend change, precisely at the transition between the historical data and the projected mortality rates. The implication of this is that the data has been collected at a unique point in time that is qualitatively different from the periods before or after it. We do not

15

believe that such an assumption is tenable.

In contrast, the modified gravity model does not predict a change in trend at the transition between past and future. As discussed in Hunt and Blake (2015c), the linear trends in the classic APC model are unidentifiable and depend entirely upon the identifiability constraints, whereas the variation around those trends is identifiable. Therefore, the modified gravity model leaves the linear trend in both populations unchanged, but allows the variation around these trends to be cointegrated. This means that decreases in mortality which are faster than expected in England & Wales are correlated with faster than expected declines in mortality rates in the CMI Assured Lives population. Capturing this correlation is vital in the measuring of basis risk between populations, as in Li and Hardy (2011) and Coughlan et al. (2011), and when modelling liabilities and securities which depend upon mortality in multiple populations, as discussed in Hunt and Blake (2015d).

Not only is the modified gravity model well-identified, we also believe that it gives projections which give greater consistency between the past and the future. The behaviour of the fitted parameters has been analysed and projected into the future, without assuming a priori that this behaviour will change. Such an approach is far more consistent with the extrapolative approach to projecting mortality rates discussed in Section 2 of Hunt and Blake (2015a) than the assumption of a trend change present in the original gravity model.

Furthermore, we believe that an assumption whereby projections maintain the same trends in each population but allow for correlated variation around these trends is more justified in terms of biological reasonableness than assuming that the period parameters converge in future. The factors impacting deviations in mortality rates from trend in one population are likely to be common across populations, leading to correlated variation around the trend in the two populations. In contrast, the differing trend rates of mortality improvement are likely to be generated by more fundamental socio-economic causes, which will remain unchanged for the foreseeable future.

In summary, we find that the modified gravity model gives projected mortality rates for England & Wales and the CMI Assured Lives popula-

tions which are well-identified and have variation which is correlated in a biologically reasonable fashion. However, the modified gravity model does not induce the trends present in either population to change sharply at the transition point between past and future, which is a feature of the original gravity model and which was imposed to ensure that mortality rates in the two populations are "coherent".

4.4 Coherence

The term "coherence" was introduced in Li and Lee (2005), and was defined formally in Hyndman et al. (2013) in terms of the relative mortality rates between populations, i.e.,

$$\mathbb{E}\left[\frac{\mu_{x,t}^{(p_1)}}{\mu_{x,t}^{(p_2)}}\right] \rightarrow R_x \tag{14}$$

a function of age only. This means that relative mortality rates are stationary, and so the mortality rates projected in the two populations do not diverge with time. Coherence is a stronger requirement for a multi-population mortality model than simply allowing the covariation observed in the past to continue into the future, as discussed in Section 4.3 above.[10]

The original gravity model was introduced in part to ensure that mortality rates in the England & Wales and CMI Assured Lives populations are coherent. The original gravity model has coherence built into it, since

$$\ln\left[\frac{\mu_{x,t}^{(I)}}{\mu_{x,t}^{(II)}}\right] = \left(\alpha_x^{(I)} - \alpha_x^{(II)}\right) + \left(\kappa_t^{(I)} - \kappa_t^{(II)}\right) + \left(\gamma_{t-x}^{(I)} - \gamma_{t-x}^{(II)}\right)$$

$$= \left(\alpha_x^{(I)} - \alpha_x^{(II)}\right) + \beta^{\top}\left(\boldsymbol{\kappa}_t + \boldsymbol{\gamma}_{t-x}\right)$$

which is stationary in time by construction.[11]

---

[10]Coherence is a potential feature of the projected mortality rates and can result from a number of different techniques for projecting mortality, rather than it being a technique in itself. For instance, the original and modified gravity models both involve the technique of cointegration, but one gives coherent projected mortality rates, whilst the other does not. Conversely, the original gravity model and the SAINT model of Jarner and Kryger (2011) both give coherent mortality rates, but use different techniques to achieve this.

[11]However, the long-run distribution of $\frac{\mu_{x,t}^{(I)}}{\mu_{x,t}^{(II)}}$, and specifically R_x, will depend upon the arbitrary identifiability constraints imposed when fitting the model.

However, when the gravity model is modified to ensure projections are well-identified, coherence no longer necessarily holds, since we can have different linear trends in both populations (i.e., $\beta^\top \left(\boldsymbol{\kappa}_t + \beta_1 t + \boldsymbol{\gamma}_{t-x} + \tilde{\beta}_1 (t - x) \right)$ is stationary, whilst $\beta^\top \left(\boldsymbol{\kappa}_t + \boldsymbol{\gamma}_{t-x} \right)$ is not). The level of divergence will be set by the observed divergence between the populations in the historical dataset, i.e., we will project mortality rates that will continue to diverge if they have been observed to do so in the past. Such an approach gives greater consistency between the historical data and projected mortality rates.

Therefore, we see that there is the potential for conflict between the desire for coherent projections and the need for projections of the model to be well-identified. In general, we believe that obtaining projected mortality rates that do not depend on arbitrary choices made when fitting the model to data is more important than a desire to prevent divergence between populations, for the reasons discussed below. However, we note that identifiability issues in mortality models are features of the *parameters* in mortality models, whereas coherence is a property of the projected mortality *rates*, which should be independent of these issues. If coherence is desired, we therefore believe that methods of imposing it should focus on constraining the projected mortality rates themselves, rather than specific features of the model parameters, which will depend on the identifiability constraints imposed.

However, we would often go further and question the desire to impose coherence a priori on projected mortality rates. Much of the work discussing coherent projections of mortality rates has been based on the idea that mortality rates should not diverge indefinitely in future between related populations. For instance, Li and Lee (2005) stated that *"Obviously, mortality differences between [closely related] populations should not increase over time indefinitely if the similar socio-economic conditions and close connections were to continue."* We believe that there are two problems with this conjecture.

First, whilst it might be true that projecting divergences indefinitely into the future may be unrealistic, we would point out that extrapolating any model indefinitely into the future is likely to give nonsensical results sooner or later. For example, the Lee-Carter model will tend to give mortality rates

arbitrarily close to zero at all ages if projected far enough into the future. However, such a phenomenon is more the fault of a modeller misusing the model to make inappropriate forecasts than it is the fault of the model itself. A general rule of thumb is that a model should not be projected for a longer period than the data used to estimate it. Given this, the question becomes why we should believe that mortality differences cannot diverge for another 50 years (say) if we have observed mortality differences diverging for the previous 50 years. Assuming that the evolution of mortality rates in the future will be qualitatively different from the past is inconsistent with the extrapolative approach.

Second, we believe that it is simply untrue that differences in mortality rates cannot persist for prolonged periods between ostensibly related populations. For example, life expectancy at age 65 varies considerably between areas in the same city[12] in a pattern which has been stable for decades, let alone between different socio-economic groups within the same country (see Harper et al. (2007) and Villegas and Haberman (2014)) or between countries. Whilst coherence does not impose the requirement that these long-established differences decrease, it does assume that they are not expected to grow beyond their current level, which we do not believe is supported by the evidence. It also raises the question as to what is so special about the currently observed differences in mortality that they should act as a barrier beyond which further divergence is not possible.

Therefore, we do not believe that coherence is a desirable property to impose upon an extrapolative multi-population mortality model. As scientific investigators, we should allow the data to speak for itself rather than impose any prior views onto the models that we use. This is consistent with the extrapolative approach discussed in Section 2 of Hunt and Blake (2015a), where analysis of historical data, rather than subjective opinions and biases, is used to project mortality rates. If the data supports our beliefs, that is encouraging. If the data does not, then we need to examine either our preconceptions to determine whether they need to be revised or re-examine the model we are using to analyse and project the data.

Ultimately, many of the preconceptions which lead to a desire for co-

---

[12]Source: `http://data.london.gov.uk/dataset/life-expectancy-birth-and-age-65-ward`

19

herence between different populations have a basis in our knowledge of the specific populations under consideration and the specific factors causing the divergence in these populations. For example, the observed divergence between mortality rates in the England & Wales and CMI Assured Lives populations could be attributed to the selective nature of the CMI Assured Lives dataset, which consists of individuals who are likely to be wealthier than the average citizen of England & Wales. In addition, this selective population may adopt different lifestyles, with less smoking and a better diet than the wider population, for example, leading to a differing pattern of mortality. We might reasonably feel that such differences will get less important with time and the wider population adopts the same lifestyle as the sub-population, and therefore that mortality rates in the two population should stop diverging in future.

However, this kind of argument for imposing coherence on a model makes use of additional information regarding the causes of any divergence, information that was not used when fitting the model. We therefore believe that, rather than imposing coherence on a model to obtain the results we want, it would be better to incorporate into our model the additional information that justifies our desire for coherence in the first place. Such information may include economic and lifestyle variables, for instance, as in Reichmuth and Sarferaz (2008), Wang and Preston (2009) and French (2014). This may help explain any observed divergence in the past and potentially allow for coherent projections which are still well founded in a rigorous analysis and extrapolation of the data.

5   Identifiability in the cointegrated Lee-Carter model

Zhou et al. (2014) applied a similar cointegration framework as developed for the gravity model to the period parameters of the Lee-Carter model

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \tag{15}$$

for multiple populations. The period parameters are projected using a time

series process of the form

$$\Delta \boldsymbol{\kappa}_t = \boldsymbol{\nu} + \Gamma \Delta \boldsymbol{\kappa}_{t-1} + \alpha \beta^\top \boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \tag{16}$$

which is a cointegrated relationship of the form in Equation 7.[13] As in Dowd et al. (2011), β was constrained so that $\beta = \begin{pmatrix} 1, & -1 \end{pmatrix}^\top$ in order that relative mortality rates do not diverge in the two populations. However, no assumption is made regarding the dominance of one population over the other, and therefore no constraint is made on α, unlike the gravity model where $\alpha = \begin{pmatrix} 0, & \phi \end{pmatrix}^\top$ was used to impose the condition that population I dominates population II.

As discussed in Lee and Carter (1992) and Hunt and Blake (2015b), the Lee-Carter model is also not well-identified and possesses the invariant transformations

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \{\alpha_x, \frac{1}{a}\beta_x, a\kappa_t\} \tag{17}$$

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \{\alpha_x - b\beta_x, \beta_x, \kappa_t + b\} \tag{18}$$

which are used to impose identifiability constraints in a similar fashion to the classic APC model. These invariant transformations can be applied independently to the two populations without affecting the fitted mortality rates, and so we can write

$$\hat{\boldsymbol{\kappa}}_t = A\left(\boldsymbol{\kappa}_t + \boldsymbol{b}\right) \tag{19}$$

with $A = \begin{pmatrix} a^{(I)} & 0 \\ 0 & a^{(II)} \end{pmatrix}$.

If we apply this transformation to the time series process in Equation 16 we obtain

$$\Delta \hat{\boldsymbol{\kappa}}_t = A\boldsymbol{\nu} - A\alpha\beta^\top \boldsymbol{b} + A\Gamma A^{-1}\Delta\hat{\boldsymbol{\kappa}}_{t-1} + A\alpha\beta^\top A^{-1}\hat{\boldsymbol{\kappa}}_{t-1} + A\boldsymbol{\epsilon}_t$$

$$= \hat{\boldsymbol{\nu}} + \hat{\Gamma}\Delta\hat{\boldsymbol{\kappa}}_{t-1} + \hat{\alpha}\hat{\beta}^\top \hat{\boldsymbol{\kappa}}_{t-1} + \hat{\boldsymbol{\epsilon}}_t$$

---

[13]Again, the form of Equation 16 differs from the form of Equation 7 due to the stationary cointegrating term, $\alpha\beta^\top \boldsymbol{\kappa}_{t-1}$, as opposed to $\alpha\beta^\top \boldsymbol{\kappa}_{t-2}$ required by Equation 7. However, this can be resolved in the manner outlined in footnote 6.

which is of the same form as Equation 16 if we redefine the terms appropriately. In particular, this involves setting

$$
\begin{aligned}
\hat{\beta} &= A^{-1}\beta \\
&= \begin{pmatrix} \frac{1}{a^{(I)}} & 0 \\ 0 & \frac{1}{a^{(II)}} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \left(\frac{1}{a^{(I)}}, \quad -\frac{1}{a^{(II)}} \right)^{\top}
\end{aligned}
$$

i.e., if the time series process is well-identified, β cannot be restricted to have any particular form, since these restrictions will only apply for one set of identifiability constraints. We also see that we are free to set $\hat{\alpha} = A\alpha$, since α is not constrained to any particular form initially. Therefore, in order for the model of Zhou et al. (2014) to be well-identified, the restriction on β as well as the restriction on α must also be relaxed. This was commented upon in Nielsen and Nielsen (2014).

The reason for the difference between the models of Zhou et al. (2014) and Dowd et al. (2011) arises because of the differences in the underlying APC mortality models used in either study. In the Lee-Carter model used in Zhou et al. (2014), the "scale" of the period functions is defined by an identifiability constraint on β_x. This scale is arbitrary, and we can change it without affecting the fitted mortality rates from the model. Therefore the projected mortality rates from the model of Zhou et al. (2014) also need to also be invariant to changes in this scale. In contrast, the scale of the period functions is defined by the parametric age function in the classic APC function, and not by an identifiability constraint. Therefore, it cannot be changed in the model, and so we do not have to ensure that the projected mortality rates are invariant to changes in its scale.

Conversely, the Lee-Carter model does not have unidentifiable linear trends, unlike the classic APC model. Therefore the model of Zhou et al. (2014) does not require a linear drift term, i.e., $\beta_1 t$, in the cointegrating relationships. The classic APC model does contain unidentifiable linear trends in the parameters, which can be varied in the historic parameters without affecting the fitted mortality rates. It is, therefore, essential that the projected mortality rates from the model of Dowd et al. (2011) are invariant to changes in the linear trend in the period an cohort functions, which is ensured

by the presence of the linear drift term, $\beta_1 t$, in the cointegrating relationships.

We see, therefore, that the form the gravity model needs to take in order to be well-identified depends on the underlying APC mortality model being used and the identifiability issues within that particular model. It is therefore essential that these identifiability issues are fully analysed and understood, as discussed in Hunt and Blake (2015b,c). In general, we see that it is best to avoid making any impositions on the structure of α and β, and so use the most general form of cointegrating relationship, in order to avoid any potential identifiability issues and avoid constraining the form of the model unnecessarily and, potentially, inappropriately.

6   Extending the cointegration model

For the general cointegration model in Equation 7, our approach generalises naturally to models where there are unidentifiable higher-order polynomial deterministic trends in the parameters. If the period functions of a model have unidentified deterministic trends which are polynomial of order M, then in order to be well-identified under the corresponding invariant transformations, we will need to allow for unconstrained deterministic trends up to polynomial order $M - 1$ and constrained deterministic trends of order M.

For instance, the model of Plat (2009)

$$\ln(\mu_{x,t}^{(p)}) = \alpha_x^{(p)} + \kappa_t^{(1,p)} + (x - \bar{x})\kappa_t^{(2,p)} + (x - \bar{x})^+ \kappa_t^{(3,p)} + \gamma_{t-x}^{(p)} \qquad (20)$$

has unidentifiable quadratic trends, as discussed in Hunt and Blake (2015c). If the Plat (2009) model were fitted to two populations, we would have six period functions in total - $\kappa_t^{(1,I)}$ and $\kappa_t^{(1,II)}$ with unidentified quadratic trends, $\kappa_t^{(2,I)}$ and $\kappa_t^{(2,II)}$ with unidentified linear trends and $\kappa_t^{(3,I)}$ and $\kappa_t^{(3,II)}$ with unidentified constants.

We could look for a cointegration model involving all six period functions. Cointegration, by its nature, involves interactions between the different period functions. We therefore are unable to allow for deterministic trends of different order in different period functions. Allowing for cointegration

between all six time series would therefore mean allowing for constrained quadratic trends and unconstrained linear trends in all six period functions, which may lead to projections which are not biologically reasonable for each population.

It is more biologically reasonable to consider each pair of period functions separately based on their shared demographic significance. This would mean looking for cointegrating relationships with constrained quadratic (and unconstrained constant and linear) trends for the two $\kappa_t^{(1,p)}$ functions, relationships with constrained linear trends for the $\kappa_t^{(2,p)}$ functions, and so on. That is, we use

$$\Delta\boldsymbol{\kappa}_t^{(1)} = \boldsymbol{\nu}_0^{(1)} + \boldsymbol{\nu}_1^{(1)}t + \alpha^{(1)}\left(\beta^{(1)\top}\boldsymbol{\kappa}_{t-1}^{(1)} + \beta_2^{(1)}t^2\right) + \boldsymbol{\epsilon}_t^{(1)} \qquad (21)$$

$$\Delta\boldsymbol{\kappa}_t^{(2)} = \boldsymbol{\nu}_0^{(2)} + \alpha^{(2)}\left(\beta^{(2)\top}\boldsymbol{\kappa}_{t-1}^{(2)} + \beta_1^{(2)}t\right) + \boldsymbol{\epsilon}_t^{(2)} \qquad (22)$$

$$\Delta\boldsymbol{\kappa}_t^{(3)} = \alpha^{(3)}\left(\beta^{(3)\top}\boldsymbol{\kappa}_{t-1}^{(3)} + \beta_0^{(3)}\right) + \boldsymbol{\epsilon}_t^{(3)} \qquad (23)$$

to project the period functions. This approach is used in Hunt and Blake (2015d), albeit in a model with unidentified cubic (as opposed to merely quadratic) trends.

7   Conclusions

Cointegration can be a powerful tool for projecting mortality rates in related populations. However, it is a tool which must be used with care to ensure that we have identifiability under any invariant transformations which allocate unidentifiable polynomial trends between the parameters. In the case of the gravity model of Dowd et al. (2011) and the model of Zhou et al. (2014), we have shown how to adapt the process used to project the period functions so that it gives well-identified projections that do not depend on the arbitrary identifiability constraints imposed. We have also shown how this can be generalised to more complicated APC mortality models.

Further, we have shown that we cannot also impose the condition that mortality rates are coherent and do not diverge in future. Not only does imposing coherence mean that the projected mortality rates will depend upon the arbitrary identifiability constraints selected, it is also incompatible with

an extrapolative approach to modelling mortality. An extrapolative approach must, first and foremost, take its lead from the evidence of the historical data. While, in many circumstances, a belief in coherence is quite natural, we believe we should test for its existence in the historical data statistically using well-identified models, rather than assume its existence beforehand as an article of faith. If we do not find any evidence for coherence in the historical data, this should be considered a puzzle to explain using more data and better models, and not just an error to be corrected by an ad hoc fix which overrides the evidence of the data to obtain the results we anticipated in advance. Such an approach is not only more rigorous and more scientific, but can also give new insights into the factors which govern the evolution of mortality rates and enhance our understanding of longevity risk.

References

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

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

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Khalaf-Allah, M., 2011. Bayesian stochastic mortality modelling for two populations. ASTIN Bulletin 41 (1), 29–59.

Carter, R., Lee, D., 1992. Modeling and forecasting US sex differentials. International Journal of Forecasting 8, 393–411.

Coughlan, G. D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A. J., Blake, D., Dowd, K., 2011. Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. North American Actuarial Journal 15 (2), 150–176.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., 2011. A gravity model of mortality rates for two related populations. North American Actuarial Journal 15 (2), 334–356.

French, D., 2014. International mortality modelling - An economic perspective. Economics Letters 122 (2), 182–186.

Harper, S., Lynch, J., Burris, S., Davey Smith, G., 2007. Trends in the black-white life expectancy gap in the United States, 1983-2003. The Journal of the American Medical Association 297 (11), 1224–32.

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

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

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

Hunt, A., Blake, D., 2016. Consistent mortality projections allowing for trend changes. Work in Progress.

Hyndman, R. J., Booth, H., Yasmeen, F., 2013. Coherent mortality forecasting: The product-ratio method with functional time series models. Demography 50, 261–283.

Jarner, S. F., Kryger, E. M., 2011. Modelling mortality in small populations: The SAINT model. ASTIN Bulletin 41 (2), 377–418.

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

Li, J. S.-H., Hardy, M. R., 2011. Measuring basis risk in longevity hedges. North American Actuarial Journal 15 (2), 177–200.

Li, N., Lee, R. D., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography 42 (3), 575–594.

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

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

Reichmuth, W., Sarferaz, S., 2008. Bayesian demographic modeling and forecasting: An application to U.S. mortality. Tech. rep., Humbolt University, Berlin.

Villegas, A. M., Haberman, S., 2014. On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. North American Actuarial Journal 18 (1), 168–193.

Wang, H., Preston, S. H., 2009. Forecasting United States mortality using cohort smoking histories. Proceedings of the National Academy of Sciences of the United States of America 106 (2), 393–8.

Yang, S. S., Wang, C.-W., 2013. Pricing and securitization of multi-country longevity risk with mortality dependence. Insurance: Mathematics and Economics 52 (2), 157–169.

Zhou, R., Wang, Y., Kaufhold, K., Li, J. S.-H., Tan, K. S., 2014. Modeling period effects in multi-population mortality models: Applications to Solvency II. North American Actuarial Journal 18 (1), 150–167.