Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

Pitfalls and merits of cointegration-based mortality models

Søren F. Jarner *, Snorre Jallbjørn

Danish Labour Market Supplementary Pension Fund (ATP), Kongens Vænge 8, 3400 Hillerød, Denmark
University of Copenhagen, Department of Mathematical Sciences, Universitetsparken 5, 2100 Copenhagen Ø, Denmark

ARTICLE INFO

ABSTRACT

In recent years, joint modelling of the mortality of related populations has received a surge of attention. Several of these models employ cointegration techniques to link underlying factors with the aim of producing coherent projections, i.e. projections with non-diverging mortality rates. Often, however, the factors being analysed are not fully identifiable and arbitrary identification constraints are (inadvertently) allowed to influence the analysis thereby compromising its validity. Taking the widely used Lee–Carter model as an example, we point out the limitations and pitfalls of cointegration analysis when applied to semi-identifiable factors. On the other hand, when properly applied cointegration theory offers a rigorous framework for identifying and testing long-run relations between populations. Although widely used as a model building block, cointegration as an inferential tool is often overlooked in mortality analysis. Our aim with this paper is to raise awareness of the inferential strength of cointegration and to identify the time series models and hypotheses most suitable for mortality analysis. The concluding application to UK mortality shows by example the insights that can be obtained from a full cointegration analysis.

1. Introduction

Mortality models have many applications in areas such as demography, epidemiology, economics and actuarial sciences. In some applications we are interested in a single life expectancy projection, e.g. a unisex projection for a given country, but in many cases we are more interested in simultaneous projections for groups of related (sub)populations. Examples of the latter include joint modelling of males and females in a population, coherent forecasts for countries in a given region, projecting the life expectancy of smokers and non-smokers, modelling of insured lives relative to a national population, and assessing the effectiveness of a mortality hedge with the presence of basis risk; see e.g. Chen and Millossovich (2018), Kleinow (2015), Bergeron-Boucher et al. (2017), Janssen et al. (2013), Jarner and Kryger (2011), Cairns et al. (2011a), Dowd et al. (2011) and Cairns et al. (2011b).

Models applied independently to separate populations often lead to diverging forecasts. This is the case even for closely related populations. Tuljapurkar et al. (2000) found that applying the model of Lee and Carter (1992) separately to the G7 countries over a 50-year forecast horizon resulted in a life expectancy gap between the countries as large as eight years; despite the countries sharing long-term trends in mortality and convergence

of social and economic factors. The projected divergence is due to small differences in the timing and magnitude of historical improvements being magnified by the separate analyses.

Joint mortality models are based on an assumption of non-divergence, or coherence, of mortality rates of the group of populations under consideration. Coherence is typically achieved by imposing a specific structure of the joint time series model used for forecasting factors driving mortality improvements in each population. Formally, the factors are assumed to cointegrate, i.e. to exhibit stationary relations preventing them from diverging. Cointegration theory offers a rigorous statistical framework for identifying and testing such stationary relations. However, this framework is rarely exploited in full since the structure is often imposed rather than tested, see e.g. Li and Lee (2005), Li and Hardy (2011), Dowd et al. (2011) and Cairns et al. (2011a).

The purpose of the present paper is twofold. First, we wish to advocate that cointegration analysis has more to offer than assuring coherence. Indeed, we will demonstrate the insights and "surprising" models that can arise from a full analysis. Second, we wish to highlight some of the pitfalls and limitations of cointegration analysis when applied to factors that are not fully identifiable, e.g. the mortality index of the popular Lee–Carter model. The overall message is that cointegration-based mortality models have much to offer as an inferential tool, but also that extreme care must be exercised when dealing with semi-identifiable factors often encountered.

---

* Corresponding author.
 E-mail address: sj@atp.dk (S.F. Jarner).

1.1. Cointegration

Cointegration is rooted in econometrics. It was introduced by Engle and Granger (1987) as a methodology for testing for stationary relations between non-stationary time series. The basic idea is that if two variables share a (stochastic) trend, it might be possible to find a linear combination of the variables that cancels the trend resulting in a stationary process. The linear combination is referred to as a cointegrating relation. The Engle–Granger methodology is limited to only a single cointegrating relation, while the more general and comprehensive setup developed by Johansen (1995) allows for an arbitrary number of variables and cointegrating relations, at least in principle. In effect, each (linearly independent) cointegrating relation reduces the dimension of the "driving forces" of the system by one.

While the aim of an econometric analysis is to infer and interpret the (economic) system, the typical focus of a mortality analysis is to produce a plausible forecast with proper quantification of the uncertainty. With the aim of improving gender-specific mortality forecasts, Carter and Lee (1992) suggested cointegration as a possible tool. In recent years, there has been a proliferation of papers using cointegration techniques to obtain coherent forecasts, see e.g. Darkiewicz and Hoedemakers (2004), Lazar and Denuit (2009), Njenga and Sherris (2011), Gaille and Sherris (2011), Yang and Wang (2013), Hyndman et al. (2013), Zhou et al. (2014), Hunt and Blake (2015c), Salhi and Loisel (2017), and Li and Lu (2017). Many of the authors arrive at complex, high-dimensional models which are difficult to interpret, but potentially good at forecasting.

We find that, although cointegration can certainly be used as a tool to impose coherence, the real strength of cointegration lies in inference and interpretation. We believe that this aspect is largely absent in the actuarial literature and that important subject knowledge can be gained from a more statistical approach to mortality modelling, a point also made by Arnold and Sherris (2016). A number of the cited papers do in fact test for cointegration rank as part of their model selection, but formulating and testing hypotheses on parameters is not part of the analysis. The primary aim of this paper is to demonstrate the value of cointegration-based inference in a mortality context.

Cointegration theory is a technically sophisticated field and some preliminary work is needed to establish the type of cointegration models and hypotheses suitable for mortality modelling. Once established, we present a cointegration analysis of male and female UK mortality. We consider both a two-dimensional analysis based on the Lee–Carter model and a four-dimensional analysis based on the logistic two-factor model of Cairns et al. (2006). In principle, any number of factors can be analysed, but to aid interpretation and the formulation of hypotheses it is useful to consider only a moderate number of factors.

1.2. Identifiability

Many mortality models, including the Lee–Carter model and its many variants, are overparametrized and parameters are therefore only identifiable after adding one or more constraints. In the Lee–Carter model, for example, two constraints are needed to ensure identification of the time-varying index and the age-specific parameters. By definition, the fitted mortality rates are unaffected by the identification scheme, but the forecasted mortality rates are not necessarily unaffected. Forecasts are based on time series models for the time-varying parameters and these models might not be invariant to the identification scheme.

Several recent papers have addressed the issue of identifiability and forecasting in mortality models, see in particular Nielsen and Nielsen (2014), Kuang et al. (2008), Hunt and Blake (2015a,b,

2017), and Beutner et al. (2017). In summary, this body of work shows that for forecasts to be unaffected by the identification scheme the time series model should be flexible enough to "preserve" reparametrizations, i.e. forecasting and reparameterizing should be interchangeable operations. As Hunt and Blake (2017) point out, this seemingly innocent requirement can in fact be at odds with model structures imposed to achieve coherence. We will return to this point later in the paper.

Identifiability issues affect the interpretation of parameters and – unless properly addressed – might lead to conclusions resting entirely on arbitrary constraints. Non-trivial issues arise even in the standard setting of an age–period model of the Lee–Carter type, and the complexity of the issues increase rapidly with the number of time-varying indices, see Hunt and Blake (2015b). Further issues arise in joint models with cointegrating parameters, which is the focus of this paper. In this case, the semi-identifiability of the parameters severely limits the choice of meaningful time series models and hypotheses.

To guarantee identification invariant inference, Nielsen and Nielsen (2014) advocate an approach based on maximal invariants of reduced dimensionality in terms of which all estimation and forecasting must be formulated. Although theoretically elegant, researchers might be reluctant to adopt this idea, since the interpretation of the original parametrization is lost. Also, we fear that the approach adds to the impression held by many that identifiability concerns are an esoteric topic which overcomplicate simple problems. In contrast to Nielsen and Nielsen (2014), we do not develop any formal theory in this paper, but rather illustrate by examples some of the pitfalls and problems of semi-identifiability. Hopefully, our exposition will be both accessible and illuminating to a wide audience. We also prefer to retain the original parametrization of the models to make the examples as relevant and familiar as possible. In these respects, our work is similar in spirit to the analysis of the gravity model presented in Hunt and Blake (2017).

1.3. Outline

The rest of the paper is organized as follows. In Section 2 we introduce the mortality models we will use as examples throughout and we discuss identification issues in the familiar setup of the Lee–Carter model. Section 3 covers background information on cointegration theory, while Section 4 specializes the discussion of cointegration to mortality models and illustrates the problems with applying cointegration to semi-identifiable parameters. Section 5 contains a comprehensive analysis of UK mortality data applying cointegration techniques to both the semi-identifiable Lee–Carter model and the fully identified model of Cairns et al. (2006). Finally, Section 6 offers some concluding remarks.

2. Mortality modelling

The object of study for mortality models is the age-specific death rates (ASDR's) in a given population. We assume that data consist of death counts, $D_{t,x}$, and corresponding exposures, $E_{t,x}$, over time–age cells of the form $[t, t+1) \times [x, x+1)$ for a range of calendar years t and (integer) ages x. We also assume that the underlying force of mortality, $\mu_{t,x}$, is constant over each of these cells. It then follows that $\mathbb{E}[D_{t,x}] = \mu_{t,x}\mathbb{E}[E_{t,x}]$, where $\mathbb{E}$ denotes the expectation operator.[1] The observed death rate is defined as

---

[1] More precisely, $\mathbb{E}$ denotes the conditional expectation given $\mu_{t,x}$, since $\mu_{t,x}$ is itself a stochastic quantity. Thus, $\mathbb{E}$ averages over the random times of death given the force of mortality. Note that the exposure is also a stochastic quantity since it depends on the life spans. For modelling purposes, we (implicitly) condition on the exposures by treating them as fixed.

the ratio $m_{t,x} = D_{t,x}/E_{t,x}$. This is commonly used as an (empirical) estimate of the underlying force of mortality. When considering more than one population we add an identifying superscript, e.g., $\mu_{t,x}^i$ denotes the force of mortality at time t and age x in population i.

Most mortality models capture the time evolution of mortality rates (the period effect) by one, or more, time-varying indices (factors), k_t. Below we introduce the one-factor model of Lee and Carter (1992) and the two-factor model of Cairns et al. (2006).

2.1. The Lee–Carter model

The single population model proposed by Lee and Carter (1992) is used in a great number of mortality studies due to its simplicity and ease of interpretation; the ASDR's are modelled by a log-linear relation where an age-dependent a_x describes the general shape of the force of mortality and a single time-varying index, k_t, describes the speed of mortality improvements, governed by an age response b_x. The model is given by

$$\log m_{t,x} = \log \mu_{t,x} + \varepsilon_{t,x} = a_x + b_x k_t + \varepsilon_{t,x}, \qquad (2.1)$$

where $\varepsilon_{t,x}$ are homoscedastic error terms with mean 0 and variance σ_ε^2.

We note that the model is invariant under the parameter transformations

$$\{a_x , \; b_x , \; k_t\} \mapsto \{a_x - b_x c , \; b_x/d , \; d(k_t + c)\}, \qquad (2.2)$$

where $c \in \mathbb{R}$ and $d \in \mathbb{R} \setminus \{0\}$, in the sense that these transformations all yield the same model for $\log m_{t,x}$. In other words, the parameters are not fully identifiable since k_t is only determined up to a linear transformation, b_x up to a multiplicative constant, and a_x up to a shift proportional to b_x. To ensure identification, the parameters are typically subject to the constraints

$$\sum_t k_t = 0, \quad \sum_x b_x = 1. \qquad (2.3)$$

In Lee and Carter (1992) the parameters are estimated by ordinary least squares (OLS), i.e. by minimization of the quantity $\sum_{t,x} (\log m_{t,x} - a_x - b_x k_t)^2$. Under the constraints (2.3), $\hat{a}_x$ equals the time average of $\log m_{t,x}$, and $\hat{b}_x$ and $\hat{k}_t$ can be obtained from the first component of a singular value decomposition of the matrix $\{\log m_{t,x} - \hat{a}_x\}_{t,x}$. The parameter estimates thus obtained equal the maximum likelihood estimates under the additional assumption that the errors are normally distributed.[2]

The assumption of homoscedastic errors is questionable as we would expect observed death rates to fluctuate more when death counts are low. In addition, the use of OLS has the practical problem of how to handle cells with zero death counts frequent in small data sets. In the application to UK data we use instead the Poisson variant of the Lee–Carter model proposed by Brouhns et al. (2002). The two Lee–Carter variants have the same parametric structure and the points made later regarding identification issues in relation to forecasting and cointegration analysis therefore apply to both of them.

2.1.1. Forecasting
The time-varying index is typically modelled as a random walk with drift,

$$k_t = k_{t-1} + \theta + \varepsilon_t, \qquad (2.4)$$

---

[2] Lee and Carter (1992) contain a second stage adjustment of the parameters to reproduce the actual number of deaths each year. We do not consider this adjustment here.

where θ is the drift and the ε_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$. The drift and variance are estimated by the sample mean and sample variance, respectively, of the differences $\hat{k}_t - \hat{k}_{t-1}$.

Let T denote the last year of data. Forecasting is based on the conditional distribution of k_{T+h} given $k_T = \hat{k}_T$; we have for $h \geq 1$

$$k_{T+h}|k_T = \hat{k}_T \sim \mathcal{N}\left(\hat{k}_T + h\hat{\theta} , \; h\hat{\sigma}^2\right), \qquad (2.5)$$

from which forecasts and confidence intervals can be derived. In particular, a (median) forecast of future log mortality rates is given by

$$\hat{\mu}_{T+h,x} = \exp\left(\hat{a}_x + \hat{b}_x \left[\hat{k}_T + h\hat{\theta}\right]\right) = \hat{\mu}_{T,x} \exp\left(h\hat{\theta}\hat{b}_x\right). \qquad (2.6)$$

2.1.2. Coherence
The concept of coherent forecasts was introduced by Li and Lee (2005) and formalized by Hyndman et al. (2013). Mortality forecasts for two populations are said to be coherent if the relative mortality rates converge for each age x,

$$\frac{\hat{\mu}_{t,x}^1}{\hat{\mu}_{t,x}^2} \to R_x, \qquad t \to \infty, \qquad (2.7)$$

for positive, age-specific constants R_x. When producing forecasts for populations with historically similar mortality evolutions, the concept of coherence formalizes the intuitively desirable property that the forecasts should reflect these similarities.

On the other hand, coherence is a rather strict requirement which will generally not be satisfied by forecasts obtained by applying e.g. the Lee–Carter model to separate populations. Indeed, it follows from (2.6) that in a Lee–Carter setting a necessary and sufficient condition for coherence is that $\hat{\theta}^1 = \hat{\theta}^2$ and $\hat{b}_x^1 = \hat{b}_x^2$ for all x. In practice, of course, this will never happen. With the aim of obtaining coherent forecasts for a group of populations, Li and Lee (2005) proposed the *augmented common factor* model

$$\log m_{t,x}^i = \log \mu_{t,x}^i + \varepsilon_{t,x}^i = a_x^i + B_x K_t + b_x^i k_t^i + \varepsilon_{t,x}^i. \qquad (2.8)$$

This model produces coherent forecasts when the population specific indices, k_t^i, are modelled as stationary processes, e.g. AR(1)-processes. The common factor, K_t, can be non-stationary, e.g. a random walk with drift as in the original Lee–Carter model.

The notion of coherence has undoubtedly been very influential in setting the standard for joint forecasts. Indeed, many joint mortality models have been devised with the specific aim of achieving coherence, as mentioned in the introduction. The model of Li and Lee (2005) can be seen as an early and very direct way to ensure coherence by equating the driving factors, while the more recent approaches typically combine a specific structure with cointegrating relations, e.g. Dowd et al. (2011). Generally, cointegrating relations do not guarantee coherence, although in the Lee–Carter setting the two concepts are closely linked. We will return to this point in Section 4.3.

It can be argued that coherence is too strict a requirement and that models enforcing coherence risk violating the historic pattern of covariation between populations, see Hunt and Blake (2017). Arguably, it is better to identify cointegrating relations which restrict the joint forecasts in plausible ways, than to insist on coherence. This point will be illustrated in the application section.

2.2. The Cairns–Blake–Dowd model

Originally made to accommodate the British pension market, the model of Cairns et al. (2006) focuses primarily on the post-age
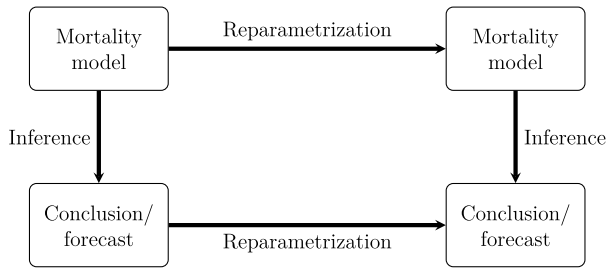
Fig. 1. Illustration of identification invariance whereby the same inferential conclusion is reached for both the original and the reparametrized model. Here, *inference* refers to all aspects of the statistical analysis, e.g., parameter estimation, model selection, hypothesis testing, forecasting, prediction intervals etc.

60 mortality curve and the pricing of immediate life annuities. The model fits the mortality curve by a logistic curve[3]

$$\text{logit}\left(q_{t,x}\right) \triangleq \log\left(\frac{q_{t,x}}{1 - q_{t,x}}\right) = k_{t,1} + k_{t,2}(x - \bar{x}), \qquad (2.9)$$

where $\bar{x} = \frac{1}{N}\sum_i x_i$ is the arithmetic mean of the N ages considered and $q_{t,x} = 1 - \exp(-\mu_{t,x})$ is the probability for an individual aged x at time t to die before $t + 1$. Note that the logit transform is well-defined since q lies between 0 and 1. We also note that $q_{t,x} \approx \mu_{t,x}$ for small $\mu_{t,x}$.

Under the CBD model the logit-transformed curve of death probabilities is linear in age with time-varying parameters. The first index is the *level* of the line, and a decreasing trend in $k_{t,1}$ thus represents an overall improvement in mortality over time. The second index is the *slope* of the line, and an increasing trend in $k_{t,2}$ thus implies a steepening of the mortality curve. The model is fully identified, since there are no invariant parameter transformations.

As is customary, we treat the model as a generalized linear model (GLM) within the binomial family with its canonical logit-link function. In principle, parameters can also be estimated by maximum likelihood assuming Poisson distributed death counts, see Currie (2016) for a comparison of the two approaches. The points made in this paper apply regardless of how parameters are estimated.

2.2.1. Forecasting

Forecasting is performed assuming a bivariate random walk with drift for the two time-varying indices

$$\binom{k_{t,1}}{k_{t,2}} = \binom{k_{t-1,1}}{k_{t-1,2}} + \binom{\theta_1}{\theta_2} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_2\left(\mathbf{0}, \Sigma\right). \qquad (2.10)$$

This projection method entails a dependency structure between the two time-varying indices by allowing for covariation, but omits the possibility of the indices being directly affected by the previous value of one another. As for the Lee–Carter model, the argument for the use of a random walk with drift is that it is often adequate to describe the data. Of course, if deemed necessary, more complicated ARIMA models can be used.

In the application section we shall model female and male mortality by two CBD models with cointegrating parameters. This will give rise to a sort of "logit coherence" rather than the usual (log) coherence.

2.3. Identification invariance

Identification issues arise in many fields of statistics, and mortality modelling is no exception. The problem arises from the fact that many (mortality) models are overparametrized such that there exist different sets of parameters yielding the same fit. From a statistical point of view all these sets are equally good, but to perform the analysis in practice the researcher has to choose one specific set. Therefore, constraints are imposed identifying one of the equivalent parameter sets over the others. The question now arises whether the (arbitrary) choice of constraints influence the forecast, and more generally the statistical inference.

To illustrate the problem, consider the familiar case of the Lee–Carter model as introduced above. Let $(\hat{a}_x, \hat{b}_x, \hat{k}_t)$ denote parameter estimates under the constraints (2.3), and consider the equivalent parameters $(\tilde{a}_x, \tilde{b}_x, \tilde{k}_t) = (\hat{a}_x - \hat{b}_x c, \hat{b}_x/d, d[\hat{k}_t + c])$ for given $c \in \mathbb{R}$ and $d \in \mathbb{R} \setminus \{0\}$. We forecast the time-varying index by the random walk with drift of (2.4). Using the mean difference as estimator for the drift we have $\tilde{\theta} = d\hat{\theta}$, and thereby $\tilde{\theta}\tilde{b}_x = \hat{\theta}\hat{b}_x$ for all x. Since also $\tilde{\mu}_{T,x} = \hat{\mu}_{T,x}$, it follows from (2.6) that for $h \geq 1$

$$\tilde{\mu}_{T+h,x} = \tilde{\mu}_{T,x}\exp\left(h\tilde{\theta}\tilde{b}_x\right) = \hat{\mu}_{T,x}\exp\left(h\hat{\theta}\hat{b}_x\right) = \hat{\mu}_{T+h,x}. \qquad (2.11)$$

This shows that the forecast obtained by the standard Lee–Carter method is in fact invariant to the chosen identification scheme.

However, consider now the case where the time-varying index is modelled as the random walk with drift of (2.4), but with a fixed drift term $\theta = \theta_0$. We then have

$$\tilde{\mu}_{T+h,x} = \tilde{\mu}_{T,x}\exp\left(h\theta_0\tilde{b}_x\right) = \hat{\mu}_{T,x}\exp\left(h\theta_0\hat{b}_x/d\right), \qquad (2.12)$$

which only equals $\hat{\mu}_{T+h,x}$ when $\theta_0\hat{b}_x = 0$. Hence, the forecasts will only be the same if $\theta_0 = 0$ (or $\hat{b}_x = 0$ for all x). Mathematically, the problem is that forecasting and reparametrization are no longer interchangeable operations or, in other words, the restricted forecasting model has different meaning for different parametrizations. One might argue that this is a contrived example as one would never consider this model, but very similar problems arise in cointegrated models where the identification of meaningful hypotheses is much less obvious.

Despite the ease with which specific problems related to lack of identification can be identified, it is surprisingly hard to formulate and justify a general principle that models and inferential procedures must adhere to.[4] Indeed, suggested principles often sound a bit vague: Hunt and Blake (2017) use the term "well-defined" for models giving the same projected mortality rates for any set of identifiability constraints, and Nielsen and Nielsen (2014) talk about "avoiding arbitrariness resulting from the identification process". We propose to use the term "identification invariance" when reparametrization and inference (including forecasting) are interchangeable, cf. Fig. 1. We consider identification invariance a fundamental property of a sound statistical analysis.

Schematically, identification invariance is similar to the classical notion of "parametrization invariance", see e.g. Lindsey (1996). However, where parametrization invariance requires inferential invariance to all one-to-one reparametrizations, identification invariance requires only invariance to parameter transformations induced by different identification constraints. In practice, this amounts to inferential invariance to a specific set of

---

[3] The use of a logistic model as a suitable choice for examining age patterns of human adult mortality is well established, see e.g. Thatcher (1999).

[4] Part of the problem seems to be that some researchers consider the constraints as an intrinsic part of the model, and not merely as (arbitrary) mathematical constraints needed for identification. From this perspective, different constraints imply different models and therefore "naturally" lead to different forecasts (even though the models are statistically identical in terms of describing the observed data).

linear parameter transformations. Note that, since time-varying parameters are typically modelled by linear time series models, the inference is generally not invariant to non-linear parameter transformations, i.e. the inference is generally not fully parametrization invariant.

3. Cointegration theory

In the following we give a brief introduction to cointegrated vector autoregressive (VAR) models, including interpretation and testing of hypotheses; unless explicitly stated otherwise, the exposition relies on Johansen (1995). In the subsequent sections this framework will be applied to mortality modelling.

A p-dimensional VAR(k)-model is a model of the form

$$\mathbf{y}_t = \boldsymbol{\Pi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Pi}_k \mathbf{y}_{t-k} + \boldsymbol{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{3.1}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ are i.i.d. errors and $\mathbf{D}_t$ contains all deterministic terms such as constant, trend and dummy variables. The evolution of each variable in the VAR-model is based on its own lagged values as well as the lagged values of the other variables in the system. This formulation highlights the short-term dynamics, while possible long-run relations between the variables are hard to discern. In order to study long-run relations we introduce the notion of cointegration. First a few preliminary definitions.

A *linear process* is defined by $\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{C}_i \boldsymbol{\varepsilon}_{t-i}$, where $\boldsymbol{\varepsilon}_t$ are i.i.d. with mean zero and finite variance and $\mathbf{C}(z) = \sum_{i=0}^{\infty} \mathbf{C}_i z^i$ is convergent for $|z| < \delta$ for some $\delta > 1$. An I(0) process is a linear process with the additional requirement that $\sum_{i=0}^{\infty} \mathbf{C}_i \neq 0$, or such a process with a deterministic trend added.[5]

The difference operator, Δ, is defined by $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$. A stochastic process $\mathbf{y}_t$ is called integrated of order 1, or I(1), if $\Delta(\mathbf{y_t} - \mathbb{E}[\mathbf{y}_t])$ is an I(0) process. Loosely speaking, the stochastic component of an I(1) process behaves like a random walk.

Definition 3.1. Let $\mathbf{y}_t$ be integrated of order 1. We say that $\mathbf{y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} \neq 0$ if $\boldsymbol{\beta}' \mathbf{y}_t - \mathbb{E}[\boldsymbol{\beta}' \mathbf{y}_t]$ admits a stationary distribution. The cointegrating rank is the number of linearly independent cointegrating vectors, and the cointegration space is the space spanned by the cointegrating vectors.

3.1. The vector error correction model

We are interested in conditions for the VAR-model to be integrated of order 1. For this purpose we subtract $\mathbf{y}_{t-1}$ from both sides of (3.1) and rearrange terms to obtain the equivalent *vector error correction model* (VECM), where the increment is expressed in terms of differences, lagged differences and the level of the process itself

$$\Delta \mathbf{y}_t = \boldsymbol{\Pi} \mathbf{y}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{3.2a}$$

$$\boldsymbol{\Pi} = (\boldsymbol{\Pi}_1 + \cdots + \boldsymbol{\Pi}_k) - \mathbf{I}, \tag{3.2b}$$

$$\boldsymbol{\Gamma}_i = -(\boldsymbol{\Pi}_{i+1} + \cdots + \boldsymbol{\Pi}_k), \quad i = 1, \ldots, k-1. \tag{3.2c}$$

The behaviour of $\mathbf{y}_t$ is most easily studied in terms of its characteristic polynomial given by $\mathbf{A}(z) = (1-z)\mathbf{I} - \boldsymbol{\Pi}z - (1-z)\sum_{i=1}^{k-1}\boldsymbol{\Gamma}_i z^i$ with determinant $|\mathbf{A}(z)|$. If $\mathbf{A}$ has a unit root then $\boldsymbol{\Pi} = -\mathbf{A}(1)$ is singular and the process is non-stationary. In this

case, $r = \text{rank}(\boldsymbol{\Pi}) < p$ and there exist two $p \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'. \tag{3.3}$$

This is essentially the requirement for I(1). However, to avoid explosive and seasonal roots and to ensure invertibility we also need the following technical condition.

Condition 3.2. *If $|\mathbf{A}(z)| = 0$, then either $|z| > 1$ or $z = 1$. Further, the matrix $\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp$ has full rank, where $\boldsymbol{\Gamma} = \mathbf{I} - \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i$ and $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are $p \times (p-r)$ matrices spanning the orthogonal complement of $\text{span}(\boldsymbol{\alpha})$ and $\text{span}(\boldsymbol{\beta})$, respectively.*

None of the matrices $\boldsymbol{\alpha}, \boldsymbol{\alpha}_\perp, \boldsymbol{\beta}$ or $\boldsymbol{\beta}_\perp$ are uniquely defined, but the conditions and conclusions do not depend on which versions we use. Let $\boldsymbol{\Gamma}(z) = \mathbf{I} - \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i z^i$ whereby $\boldsymbol{\Gamma}(1) = \boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}(\mathbf{L})\mathbf{y}_0 = \mathbf{y}_0 - \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \mathbf{y}_{-i}$, where $\mathbf{L}$ is the lag operator. We can now formulate the celebrated Granger Representation Theorem in a version due to Hansen (2005).

Theorem 3.3. *If $|\mathbf{A}(1)| = 0$ and Condition 3.2 is satisfied, then $\mathbf{y}_t$ can be represented as the sum of a random walk and a stationary process*

$$\mathbf{y}_t = \mathbf{C} \sum_{i=1}^{t} (\boldsymbol{\varepsilon}_i + \boldsymbol{\Phi} \mathbf{D}_i) + \sum_{i=0}^{\infty} \mathbf{C}_i^* (\boldsymbol{\varepsilon}_{t-i} + \boldsymbol{\Phi} \mathbf{D}_{t-i}) + \mathbf{C} \boldsymbol{\Gamma}(\mathbf{L})\mathbf{y}_0, \tag{3.4}$$

where $\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp$, and $\mathbf{C}_i^*$ is defined recursively by

$$\mathbf{C}_i^* = (\mathbf{I} + \boldsymbol{\Pi}) \mathbf{C}_{i-1}^* + \sum_{j=1}^{k-1} \boldsymbol{\Gamma}_j \Delta \mathbf{C}_{i-j}^*, \quad i = 1, 2, \ldots,$$

with $\mathbf{C}_0^* = \mathbf{I} - \mathbf{C}$ and $\mathbf{C}_{-1}^* = \cdots = \mathbf{C}_{-k+1}^* = -\mathbf{C}$. In particular, if $r > 0$ then $\mathbf{y}_t$ is a cointegrated I(1) process with cointegrating vectors $\boldsymbol{\beta}$.

Intuitively, the process evolves as a random walk in $\text{span}(\boldsymbol{\beta}_\perp)$ while at the same time it tries to establish the equilibrium relation for $\boldsymbol{\beta}' \mathbf{y}_t$ with a force that depends on the adjustment coefficients $\boldsymbol{\alpha}$ and the equilibrium error $\boldsymbol{\beta}' \mathbf{y}_t - \mathbb{E}[\boldsymbol{\beta}' \mathbf{y}_t]$.

The factorization (3.3) defines the cointegration space, but the individual cointegrating relations are not unique without further normalization. Johansen (1995) suggests letting the first part of $\boldsymbol{\beta}$ be an r-dimensional identity matrix making for a just-identified normalization and we adopt this approach throughout without further notification.

3.2. Deterministic terms and trends

The deterministic term is an important part of the specification of the model affecting both the trend of $\mathbf{y}_t$ and the test statistics for cointegration rank. Under the conditions of Theorem 3.3, the process $\mathbf{y}_t$ has, in general, a trend of the form $\mathbf{C}\boldsymbol{\Phi} \sum_{i=1}^{t} \mathbf{D}_i + \sum_{i=0}^{\infty} \mathbf{C}_i^* \boldsymbol{\Phi} \mathbf{D}_{t-i}$. Note that we refer to the deterministic part of $\mathbf{y}_t$ as a trend, regardless of its order.

In general, the deterministic terms accumulate to a trend one order higher. More precisely, however, it is only the terms $\mathbf{C}\boldsymbol{\Phi}\mathbf{D}_t$ that accumulate to a higher order. To illustrate the implication of this in more detail, we consider deterministic terms of the form

$$\boldsymbol{\Phi}\mathbf{D}_t = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 t, \tag{3.5}$$

for p-dimensional vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. Following Johansen (1995), we decompose each $\boldsymbol{\theta}_i$ as $\boldsymbol{\theta}_i = \boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_i + \boldsymbol{\alpha} \boldsymbol{\rho}_i$. Since $\mathbf{C}\boldsymbol{\alpha} = 0$, it follows from the Granger representation (3.4) that only $\boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_1 i$ enters into the i'th term of the random walk component. In particular, there is a quadratic trend in the level of the process with coefficient $\frac{1}{2}\mathbf{C}\boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_1$.

---

[5] Sometimes a deterministic trend is allowed in the definition of a linear process. Here, however, we follow the terminology of Johansen (1995) whereby a linear process has mean zero.

Model	Deterministic term	Trend in $\mathbf{y}_t$	$\mathbb{E}\left[\Delta \mathbf{y}_t\right]$	$\mathbb{E}\left[\boldsymbol{\beta}'\mathbf{y}_t\right]$
$H_0(r)$	$\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 t$	Quadratic	Linear	Linear
$H_0^*(r)$	$\boldsymbol{\theta}_0 + \boldsymbol{\alpha}\boldsymbol{\rho}_1 t$	Linear	Constant	Linear
$H_1(r)$	$\boldsymbol{\theta}_0$	Linear	Constant	Constant
$H_1^*(r)$	$\boldsymbol{\alpha}\boldsymbol{\rho}_0$	Constant	Zero	Constant
$H_2(r)$	0	Zero	Zero	Zero

The decomposition of $\boldsymbol{\theta}_i$ gives rise to five nested models defined by restricting the number of non-zero terms. The models and the trending behaviour they entail are summarized in Table 1.

3.3. Cointegration rank and parameter estimation

In some situations, the cointegration rank can be justified on the basis of prior knowledge. Often, however, the cointegration rank needs to be inferred from the data. Let $H(r)$ denote the hypothesis that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ for two $p \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Without further restrictions, this is equivalent to the hypothesis that rank$(\boldsymbol{\Pi}) \leq r$. This creates a set of nested hypotheses

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(p).$$

The cointegration rank can be determined by testing these hypotheses sequentially, starting from $H(0)$ and stopping when the first acceptance is encountered. The cointegration rank is r, say, if $H(0), \ldots, H(r-1)$ are rejected, while $H(r)$ is accepted. Johansen (1995) derives the likelihood-ratio test, known as the trace test, for performing these tests. The distribution of the test statistic is non-standard and it depends on the specification of the deterministic term. Critical values are tabulated in Section 15.3 of Johansen (1995) for the five models considered in Section 3.2. The testing procedure is complicated by the fact that we might need to infer the model for the deterministic term and the cointegration rank simultaneously. We will return to this point in the application section.

Given the (maximal) cointegration rank and specification of the deterministic term, the maximum likelihood estimates of the parameters are obtained by reduced rank regression. For completeness, the estimates and the trace test statistic can be found in the Appendix.

4. Cointegration in mortality models

In this section we first discuss the trend models most relevant in a mortality context. Next, we show how the identification issues of the Lee–Carter model severely limit the set of testable (cointegration) hypotheses. Finally, we comment on alternative approaches to cointegration within the Lee–Carter framework.

4.1. Linear trend models

We restrict our attention to the case of analysing the period effect within a given parametric mortality model. Assume that $\boldsymbol{k}_t$, consisting of the combined time-varying indices of (separate) age–period models, can be shown to form an I(1) process. This is a reasonable assumption, since we expect the period effect to cause at least one of the time-varying indices (for each population) to accumulate annual improvements over time and hence to behave like a random walk with drift. In general, the drift term itself could be time-varying giving rise to trends of all shapes and orders. Indeed, Arnold and Sherris (2016) find quadratic trends when analysing cause-specific mortality. However, for the purpose of this exposition we focus on linear trends only, which

are, arguably, also the only type of trends suitable for robust forecasting.

To characterize the relevant models in more detail, let us consider deterministic terms of the form $\boldsymbol{\Phi}\mathbf{D}_t = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 t$, where $\boldsymbol{\theta}_i = \boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_i + \boldsymbol{\alpha}\boldsymbol{\rho}_i$ for $i = 0, 1$, cf. Section 3.2. The absence of a quadratic trend implies that $\boldsymbol{\gamma}_1 = 0$. Hence, in the current context the largest model of interest is $H_0^*(r)$ of Table 1. Under this model, $\boldsymbol{k}_t$ has the representation

$$\boldsymbol{k}_t = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t + \mathbf{C} \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i + \sum_{i=0}^{\infty} \mathbf{C}_i^* \boldsymbol{\varepsilon}_{t-i} + \mathbf{C}\boldsymbol{\Gamma}(\mathbf{L})\boldsymbol{k}_0; \qquad (4.1)$$

expressions for the intercept, $\boldsymbol{\tau}_0$, and slope, $\boldsymbol{\tau}_1$, can be found in Hansen (2005). It can be shown that the cointegrating relations are *trend stationary*, i.e. they can be decomposed as a linear trend and a stationary process, $\boldsymbol{u}_t$, as $\boldsymbol{\beta}'\boldsymbol{k}_t = \boldsymbol{\beta}'\boldsymbol{\tau}_0 - \boldsymbol{\rho}_1 t + \boldsymbol{u}_t$. Thus, parameters drift further and further apart over time even though they engage in an equilibrium correcting relationship.

The second model of interest for mortality modelling is $H_1(r)$ of Table 1. This model has $\boldsymbol{\rho}_1 = \boldsymbol{\gamma}_1 = 0$, i.e. the previous model with the further restriction that the cointegrating relations do not trend. The level of the process still possesses a linear trend. Demographically, the lack of a trend in the cointegrating relations is appealing, but it cannot be assumed in advance. In the application section we will use the statistical setup of Johansen (1995) to test for $\boldsymbol{\rho}_1 = 0$.

Technically, it is also possible to test for further model restrictions, i.e. model $H_1^*(r)$ for absence of a linear trend altogether and model $H_2(r)$ for zero mean. However, neither of these latter models are relevant for modelling mortality data with period effects.

As described in Section 2, applications of the Lee–Carter model and the CBD-model often employ a simple random walk with drift to describe the time-varying indices. In the spirit of preserving as much of the marginal structure as possible, the natural candidate for joint modelling is therefore the VECM with zero lagged differences and no quadratic trend

$$\Delta \boldsymbol{k}_t = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \boldsymbol{\rho}_1 t\right) + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t$$
$$= \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \boldsymbol{\rho}_0 + \boldsymbol{\rho}_1 t\right) + \boldsymbol{\alpha}_\perp \boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}_t. \qquad (4.2)$$

For this model, the linear trend of (4.1) is given by $\boldsymbol{\tau}_1 = \mathbf{C}\boldsymbol{\theta}_0 - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\rho}_1$. Further, if $\boldsymbol{\rho}_1 = 0$ then $\boldsymbol{\tau}_1 = \mathbf{C}\boldsymbol{\theta}_0$ and $\mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right] = -\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{\theta}_0$, and we have

$$\Delta \boldsymbol{k}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} - \mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right]\right) + \boldsymbol{\tau}_1 + \boldsymbol{\varepsilon}_t. \quad (4.3)$$

Models (4.2)–(4.3) will be our workhorse models in the application section.

4.2. Identification invariance of cointegrated Lee–Carter models

The Lee–Carter model is the predominant single population mortality model. At first thought, it therefore seems natural to use the Lee–Carter model as the underlying model for a joint analysis of related populations. However, it turns out that the identifiability issues of the Lee–Carter model severely limit its usefulness for this purpose.

The lack of identifiability of the Lee–Carter model and the consequences for interpretation and forecasting have also been studied by other authors, see in particular Nielsen and Nielsen (2014) and Hunt and Blake (2017). In contrast to these contributions, we here focus on how the identifiability issues restrict the set of testable hypotheses, i.e. the hypotheses that can form part of a statistical analysis. More precisely, we are interested in characterizing the identification invariant cointegration models for the time-varying mortality index of two Lee–Carter models, i.e. the

Table 2
Overview of models for deterministic terms (drift) and hypotheses for parameter constraints. A check mark indicates that the model/hypothesis is identification invariant/testable in the context of a VECM for the mortality indices of two Lee–Carter models, and a minus sign that it is not.

Model	Drift	Invariant	Hypothesis	Testable
$H_0(r)$	$\theta_0 + \theta_1 t$	✓	$\alpha = \mathbf{A}\psi$	÷
$H_0^*(r)$	$\theta_0 + \alpha\rho_1 t$	✓	$\beta = \mathbf{H}\varphi$	÷
$H_1(r)$	θ_0	✓	$\theta_0 = \theta$	÷
$H_1^*(r)$	$\alpha\rho_0$	✓	$\rho_0 = \rho$	÷
$H_2(r)$	0	÷	$\rho_1 = \rho$	✓

cointegration models for which forecasting and reparametrization of the underlying Lee–Carter models are interchangeable, cf. Section 2.3.

Let $\boldsymbol{k}_t = (k_t^1, k_t^2)'$ denote the vector of mortality indices of two Lee–Carter models with given identification schemes. Assume that we choose to model this as a VECM with a linear trend of class $H_0^*(r)$,

$$\Delta\boldsymbol{k}_t = \sum_{i=1}^{l-1}\Gamma_i\Delta\boldsymbol{k}_{t-i} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \rho_1 t\right) + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t, \qquad (4.4)$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, \Sigma)$. Now, consider a reparametrization of the underlying Lee–Carter models or, equivalently, different identification schemes. According to (2.2) the vector of reparametrized mortality indices takes the form $\tilde{\boldsymbol{k}}_t = \mathbf{D}(\boldsymbol{k}_t + \boldsymbol{c})$, where $\mathbf{D} = \mathrm{diag}(d_1, d_2)$ with $d_1, d_2 \in \mathbb{R} \setminus \{0\}$, and $\boldsymbol{c} = (c_1, c_2)'$ with $c_1, c_2 \in \mathbb{R}$. Applying the same transformation to (4.4) yields

$$\Delta\tilde{\boldsymbol{k}}_t = \sum_{i=1}^{l-1}\mathbf{D}\Gamma_i\mathbf{D}^{-1}\Delta\tilde{\boldsymbol{k}}_{t-i} + \mathbf{D}\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\left(\mathbf{D}^{-1}\tilde{\boldsymbol{k}}_{t-1} - \boldsymbol{c}\right)\right.$$
$$\left. + \rho_1 t\right) + \mathbf{D}\boldsymbol{\theta}_0 + \mathbf{D}\boldsymbol{\varepsilon}_t \qquad (4.5)$$

$$= \sum_{i=1}^{l-1}\tilde{\Gamma}_i\Delta\tilde{\boldsymbol{k}}_{t-i} + \tilde{\boldsymbol{\alpha}}\left(\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{k}}_{t-1} + \rho_1 t\right) + \tilde{\boldsymbol{\theta}}_0 + \tilde{\boldsymbol{\varepsilon}}_t, \qquad (4.6)$$

where $\tilde{\Gamma}_i = \mathbf{D}\Gamma_i\mathbf{D}^{-1}$, $\tilde{\boldsymbol{\alpha}} = \mathbf{D}\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\beta}} = \mathbf{D}^{-1}\boldsymbol{\beta}$, $\tilde{\boldsymbol{\theta}}_0 = \mathbf{D}\boldsymbol{\theta}_0 - \mathbf{D}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{c}$, and $\tilde{\boldsymbol{\varepsilon}}_t \sim \mathcal{N}_2(\mathbf{0}, \mathbf{D}\Sigma\mathbf{D})$. Since $\mathbf{D}$ and $\mathbf{D}^{-1}$ have full rank, $\mathrm{rank}\left(\boldsymbol{\alpha}\boldsymbol{\beta}'\right) = \mathrm{rank}\left(\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}'\right)$, and it follows that (4.6) belongs to $H_0^*(r)$ for the same value of r as in (4.4). Thus, in the context of cointegrated Lee–Carter models, $H_0^*(r)$ is identification invariant, cf. Fig. 1.

Similar calculations show that the first four model classes of Table 1 are all identification invariant, while the fifth, $H_2(r)$, is not (unless $r = 0$). In other words, constant and linear trends are preserved by linear transformations of the two indices being modelled—*as long as we allow all other parameters to vary freely*.

While it is generally valid to impose different models for the deterministic terms, e.g. $H_0^*(r)$ or $H_1(r)$, for freely varying parameters, it is generally *not* valid to test or impose further parameter constraints. Indeed, since $\mathbf{D}$ is a diagonal matrix with an arbitrary, non-zero diagonal and since $\tilde{\boldsymbol{\alpha}} = \mathbf{D}\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\beta}} = \mathbf{D}^{-1}\boldsymbol{\beta}$, constraints on either $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ will typically not be satisfied by the transformed model. In particular, the hypothesis of prime interest, $\boldsymbol{\beta} \propto (1, -1)'$, is not testable, i.e. the corresponding model is not identification invariant. Nor can we test hypotheses on the (relative) magnitude of the adjustment coefficients. In fact, apart from the cointegration rank, the only testable hypothesis of some demographic interest is $\rho_1 = \rho$ for given ρ. The ρ_1 parameter is identifiable and can be interpreted as a 'divergence' measure between (related) populations.

In summary, obeying identification invariance we can infer the cointegration rank and distinguish between the two types of linear trend models most relevant for mortality modelling (and also

between other less relevant models). However, within a given model class we can in general not restrict parameters further without violating identification invariance, i.e. essentially all hypotheses of interest are non-testable. The situation is summarized in Table 2. For someone interested purely in joint forecasting of two populations this might not pose a problem. However, if the aim is to analyse the nature of the joint behaviour in more detail, cointegrated Lee–Carter models are of limited use.

4.3. Alternative approaches of the Lee–Carter type

Arguably, the simplest way to avoid the problems due to lack of identifiability of the Lee–Carter model is to use another model. Indeed, if we base our (joint) analysis on a fully identifiable mortality model, e.g. the CBD model, all (joint) hypotheses are well-defined and testable. However, due to the familiarity and widespread use of the Lee–Carter model some researchers might be reluctant to follow this route. For that reason, we consider below two alternative approaches to obtain identification invariant inference within the Lee–Carter framework.

The first approach is to impose further restrictions on the parameters of the underlying Lee–Carter models, thereby implicitly restricting the set of invariant transformations of the mortality indices. As an example of this approach, Zhou et al. (2014) assume equality of the age response parameters of two Lee–Carter models, i.e. $b_x = b_x^1 = b_x^2$ for all x,

$$\log m_{t,x}^i = \log \mu_{t,x}^i + \varepsilon_{t,x}^i = a_x^i + b_x k_t^i + \varepsilon_{t,x}^i \quad \text{for } i = 1, 2. \qquad (4.7)$$

This is similar in spirit to the augmented common factor model of Li and Lee (2005), cf. (2.8). Let $\boldsymbol{k}_t = (k_t^1, k_t^2)'$ denote the vector of mortality indices for given identification scheme. Due to the constraint on the b-parameters, different identification schemes lead to mortality indices of the form $\tilde{\boldsymbol{k}}_t = d(\boldsymbol{k}_t + \boldsymbol{c})$, with $d \in \mathbb{R} \setminus \{0\}$, and $\boldsymbol{c} = (c_1, c_2)'$ with $c_1, c_2 \in \mathbb{R}$. The point to note is that, in contrast to the situation in Section 4.2, the mortality indices are always scaled by the same constant.

If $\boldsymbol{k}_t$ is modelled by the VECM of (4.4) then $\tilde{\boldsymbol{k}}_t$ satisfies (4.6) with $\tilde{\boldsymbol{\alpha}} = d\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\beta}} = d^{-1}\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\theta}}_0 = d\boldsymbol{\theta}_0 - d\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{c}$. We notice that the transformed adjustment coefficients and long-run relations are proportional to their previous values, hence hypotheses on the relative magnitude of these parameters are testable, while hypotheses on the "intercepts" $\boldsymbol{\theta}_0$ and ρ_0 are still not testable. In particular, $\boldsymbol{\beta} \propto (1, -1)'$, imposed – but not tested – by Zhou et al. (2014) in their analysis, is in fact a testable hypothesis.

In some sense the model of Zhou et al. (2014) "solves" the identification issue of separate Lee–Carter models by imposing just enough additional structure to allow the formulation of well-defined joint hypotheses of interest. However, it comes at the price of a very restrictive parameter structure. It is unlikely that identical age response parameters for two separate populations is a reasonable assumption, and in general the model must be expected to fit data rather poorly. Also, it is a somewhat indirect way to address the identification issue.[6]

In the generic mortality modelling setup considered so far one or more time-varying factors are first extracted from data and then forecasted by a time series model. It is tacitly assumed that the number of factors is low; the Lee–Carter model, for instance, uses a single factor to capture the mortality evolution over time.

---

[6] In Zhou et al. (2014), the stated reason for assuming identical b-parameters is to obtain non-divergent mortality rates. It is unclear whether the authors realize that this assumption also ensures identification invariance. Indeed, Hunt and Blake (2017) in their otherwise careful paper seem to overlook this subtlety in their critique of the model by failing to acknowledge the restricted set of invariant transformations (the A-matrix of equation (19) of Section 5 of Hunt and Blake (2017) ought to have identical, rather than freely varying, diagonal terms).
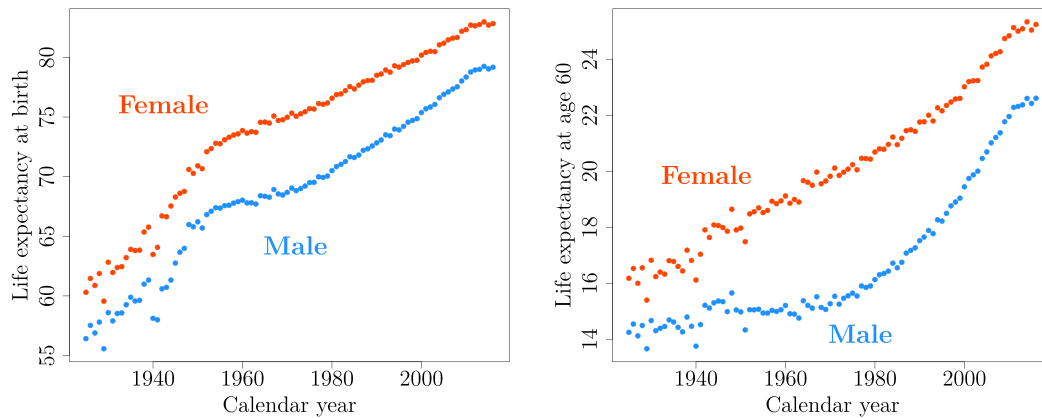
Fig. 2. Empirical period remaining life expectancy for UK males and females at birth (left panel) and at age 60 (right panel) for the period 1930–2016.

An alternative approach is to consider the vector of log mortality rates as a multivariate time series (of high dimension) and model this series directly, i.e. to model directly the N-dimensional series $\mathbf{y}_t = (\log m_{t,x_1}, \ldots, \log m_{t,x_N})'$. By construction, there are no factors and hence no identification issues related to factor identification, but due to the high dimension the time series models are more complex and harder to interpret.

Lazar and Denuit (2009) use the cointegration methodology of Section 2 to model $\mathbf{y}_t$. In this framework, the Lee–Carter model is a special case with cointegration rank $N - 1$ corresponding to one common stochastic trend. Lazar and Denuit (2009) focus on single-population modelling, but the approach extends readily to multi-population modelling by stacking the $\mathbf{y}$-vectors. The VAR/VECM approach to modelling $\mathbf{y}_t$ is also explored in the recent papers by Salhi and Loisel (2017) and Li and Lu (2017).

As mentioned in Section 2.1.2, coherence has received much attention as a desirable property of mortality forecasts. In the Lee–Carter setting of Section 4.2 coherence corresponds to stationarity of $b_x^1 k_t^1 - b_x^2 k_t^2$ for all x. Note that, cointegrated mortality indices, i.e. stationarity of $\boldsymbol{\beta}' \boldsymbol{k}_t$ for *some* $\boldsymbol{\beta}$, do not in itself guarantee coherence. Indeed, for (strict) coherence we must have $\boldsymbol{b}_x \propto \boldsymbol{\beta}$ for all x, where $\boldsymbol{b}_x = (b_x^1, -b_x^2)'$. In practice, this will never be (strictly) satisfied, unless enforced by design as in Zhou et al. (2014). In contrast, when modelling $\mathbf{y}_t$ directly non-diverging rates for different populations (coherence) or for different ages within the same population can more easily be obtained while preserving model flexibility, see e.g. the model of Li and Lu (2017).

The VAR/VECM approach to modelling $\mathbf{y}_t$ directly certainly has its merits as a flexible method for forecasting capable of capturing the dependency structure across ages. The approach also provides a useful framework for determining the number of common stochastic trends, i.e. the dimension of the driving dynamics. However, due to the high dimension the resulting models are often very complex and hard to interpret. For the purpose of gaining demographic insights by formulating and testing hypotheses we find that more parsimonious models with a limited set of interpretable factors are better suited.

5. Applications to UK mortality data

In this section we present two applications of cointegration-based mortality models. The applications focus on the inferential procedure, in particular hypothesis testing and interpretation of the models, rather than on the resulting forecasts. We use UK mortality data for males and females retrieved from the Human Mortality Database (2019).[7]

─────────

[7] Specifically, we use data for United Kingdom with HMD country code GBR_NP.

The first application is based on the Lee–Carter model and illustrates the care with which results must be interpreted due to semi-identifiability. The second application is based on the fully identified CBD-model for which a more detailed analysis is possible. In both cases, the analysis starts with a visual inspection of the mortality indices being modelled and tests for non-stationarity.

5.1. Lee–Carter application

The period remaining life expectancy is a useful summary measure for the mortality conditions of a population at a given point in time. Fig. 2 shows the period remaining life expectancy at birth and at age 60 for UK males and females, calculated using the observed death rates $m_{t,x} = D_{t,x}/E_{t,x}$. The most striking feature is the remarkable increase in life expectancy over the period; an increase also seen in most other developed countries. Another prominent feature is the degree of similarity between the life expectancy evolution of the two genders. This prompts the scientific question of whether the joint behaviour is the result of two random walk-type processes with similar drift and correlated innovations, or whether the two processes do in fact engage in a cointegrating relation? As a first attempt at answering this question, we consider a cointegration analysis of the mortality indices of two Lee–Carter models.

From Fig. 2 we can identify a number of different periods in the life expectancy evolution. After the steep and erratic initial part, the life expectancy evolution changes character during the early 1950s and improvements are hereafter smoother. Around 1970 improvements in male life expectancy pick up speed and the gender gap begins to narrow. Finally, in the last part of the series, around 2010, improvements slow down and life expectancy flattens for both sexes.

The choice of data period for the analysis is a compromise between including as much information as possible versus using only data adequately described by the models. Balancing these concerns we choose to use the period 1960–2016; a period so long that it enables us to capture potential equilibrium relations. As previously advertised we use the Poisson regression version of the Lee–Carter model with the identification constraints of (2.3). The Lee–Carter model is estimated separately for males and females over the period 1960–2016 and ages 0–100. Fig. 3 shows the estimated mortality indices (k-index) and the age response parameters (b-parameters). The mortality indices evolve quite similarly over time, while the age response parameters are rather different between ages 20–80.
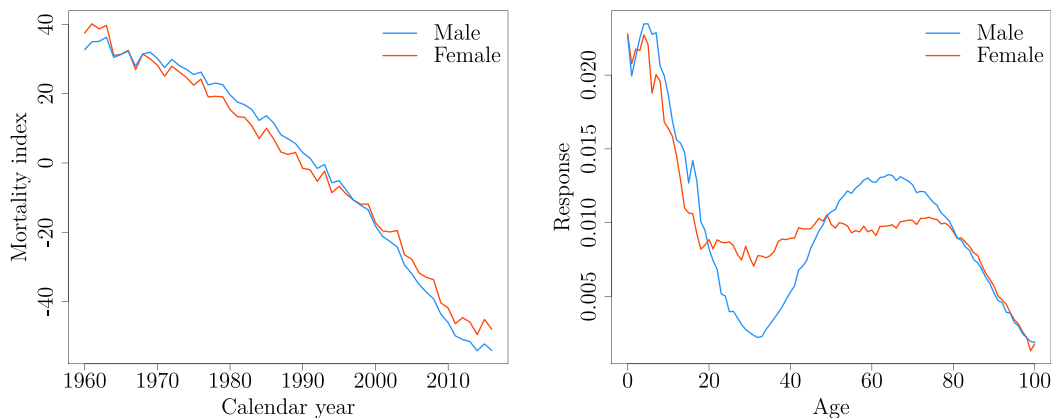
Fig. 3. Parameters for Lee–Carter models fitted to UK male and female mortality over the period 1960–2016 and ages 0–100. Left panel shows the mortality index, $\hat{k}_t$, and right panel shows the age response parameters, $\hat{b}_x$.

5.1.1. Cointegration rank and deterministic structure

From unit root tests we conclude that $\hat{k}_t^{\circlearrowleft}$ and $\hat{k}_t^{\circlearrowleft}$ are indeed I(1)-processes (test results not shown).[8] The next objective of the analysis is to determine whether or not the two indices engage in an equilibrium correcting relationship, i.e. do $\hat{k}_t^{\circlearrowleft}$ and $\hat{k}_t^{\circlearrowleft}$ cointegrate?

To answer this question we employ the VECM of (4.2) and test for cointegration rank and deterministic structure in this model. For given deterministic structure, the test for cointegration rank is based on Johansen's trace statistic, cf. (A.15) of Appendix. The results are shown in Table 3. Recall that the null hypothesis of at most r cointegrated relations, rank($\boldsymbol{\Pi}$) $\leq r$, is tested against an unrestricted $\boldsymbol{\Pi}$. Since the test for cointegration rank is very dependent on the assumed deterministic structure, the two must be determined jointly.

We first note that the two models $H_1(0)$ and $H_0^*(0)$ are identical, since rank($\boldsymbol{\Pi}$) $= 0$ implies that $\boldsymbol{\alpha}$ is zero. Table 3 presents two tests for this model against, respectively, $H_1(2)$ and $H_0^*(2)$. Both of these are rejected at the 5%-significance level. Consequently, we conclude that the system cointegrates and we move on to test the two cointegration models $H_1(1)$ and $H_0^*(1)$. Both of these models are accepted and we therefore take the simpler of these, $H_1(1)$, as describing our data

$$\Delta \boldsymbol{k}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} - \mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right]\right) + \boldsymbol{\tau}_1 + \boldsymbol{\varepsilon}_t, \quad (5.1)$$

where $\mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right] = -\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{\theta}_0$ and $\boldsymbol{\tau}_1 = \boldsymbol{\beta}_{\perp}\left(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\beta}_{\perp}\right)^{-1}\boldsymbol{\alpha}'_{\perp}\boldsymbol{\theta}_0$ are, respectively, the stationary mean of the cointegrating relation and the trend in the level of the process, cf. (4.3).

In an ordinary cointegration analysis model (5.1) would serve as a starting point for formulating and testing hypotheses of interest on the parameters. However, as described in Section 4.2, in the present context there are no further testable hypotheses and all there is left is to interpret the model.

5.1.2. Interpretation of the cointegration model

We obtain the following model, where we have used the just-identified normalization of Section 3.1 by which the first element of $\boldsymbol{\beta}$ is 1,

$$\begin{pmatrix} \Delta k_t^{\circlearrowleft} \\ \Delta k_t^{\circlearrowleft} \end{pmatrix} = \begin{pmatrix} -0.084 \\ 0.018 \end{pmatrix} \begin{pmatrix} 1 & -1.235 \end{pmatrix} \begin{pmatrix} k_{t-1}^{\circlearrowleft} \\ k_{t-1}^{\circlearrowleft} \end{pmatrix} + \begin{pmatrix} -1.560 \\ -1.526 \end{pmatrix} + \boldsymbol{\varepsilon}_t$$

$$= \begin{pmatrix} -0.084 \\ 0.018 \end{pmatrix} (s_{t-1} - 3.056) + \begin{pmatrix} -1.817 \\ -1.471 \end{pmatrix} + \boldsymbol{\varepsilon}_t,$$

Table 3

The trace test for cointegration rank for deterministic terms of class H_1 and H_0^*. The table shows the likelihood ratio test statistic for test of $H(r)$ in $H(2)$ for $r = 0, 1$. Critical values are the 95%-quantiles of the limiting distribution given in Section 15.3 of Johansen (1995).

Model	Cointegrating relations	Deterministic term	Trace statistic	Critical value
$H_1(0)$	0	$\boldsymbol{\theta}_0$	33.95	15.34
$H_1(1)$	1	$\boldsymbol{\theta}_0$	1.18	3.84
$H_0^*(0)$	0	$\boldsymbol{\theta}_0$	34.05	25.47
$H_0^*(1)$	1	$\boldsymbol{\theta}_0 + \boldsymbol{\alpha}\rho_1 t$	1.28	12.39

where $s_t = \hat{\boldsymbol{\beta}}'\boldsymbol{k}_t = k_t^{\circlearrowleft} - 1.235 k_t^{\circlearrowleft}$ and the drift term $\boldsymbol{\theta}_0$ is decomposed as in (5.1). Parameter estimates are obtained by reduced rank regression as described in Appendix.

Note that the adjustment coefficients have opposite signs for males and females such that the two indices are either pushed together or pushed apart in response to disequilibrium errors. However, since the female coefficient is much smaller than the male coefficient, the adjustments are primarily taken by the male index, while the female index evolves essentially like a random walk.

It is tempting to interpret the fact that the $\boldsymbol{\beta}$-coefficients are of the same magnitude but opposite signs as suggesting that male and female mortality "follow each other closely", i.e. as approximate coherence. However, this is not a valid conclusion since the (relative) magnitude of the $\boldsymbol{\beta}$-coefficient has no significance by itself, but is merely a result of the chosen identification schemes, cf. Section 4.2. The proper conclusion is that forecasted mortalities will be approximately coherent for ages with $\hat{b}_x^{\circlearrowleft} \approx 1.2\hat{b}_x^{\circlearrowleft}$, and for these ages only. From the right panel of Fig. 3 we see that this relation is satisfied for only a small group of ages around age 25 and age 40.

Regarding the cointegrating relation, it follows from (5.1) that

$$s_t = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})\left(s_{t-1} - \mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right]\right) + \mathbb{E}\left[\boldsymbol{\beta}'\boldsymbol{k}\right] + u_t, \quad (5.2)$$

where $u_t = \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$. Thus the cointegrating relation follows an autoregression with AR-coefficient of $1 + \boldsymbol{\beta}'\boldsymbol{\alpha}$.[9] We know from Section 4.2 that alternative identification schemes in the underlying Lee–Carter models lead to a new set of adjustment and long-run coefficients of the form $\tilde{\boldsymbol{\alpha}} = \mathbf{D}\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\beta}} = \mathbf{D}^{-1}\boldsymbol{\beta}$. Observing that $\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\alpha}} = \boldsymbol{\beta}'\boldsymbol{\alpha}$, it follows that the AR-coefficient is in fact invariant to the identification scheme(s). We can therefore

---

[8] Specifically, we use the *Augmented Dickey–Fuller* test in conjunction with the *Phillips–Perron*, see e.g. Said and Dickey (1984) and Phillips and Perron (1988).

[9] Note that, since $|\boldsymbol{A}(z)| = |1 - z||1 - z(1 + \boldsymbol{\beta}'\boldsymbol{\alpha})|$ the first part of Condition 3.2 reduces to $|1 + \boldsymbol{\beta}'\boldsymbol{\alpha}| < 1$, i.e. stationarity of (5.2).
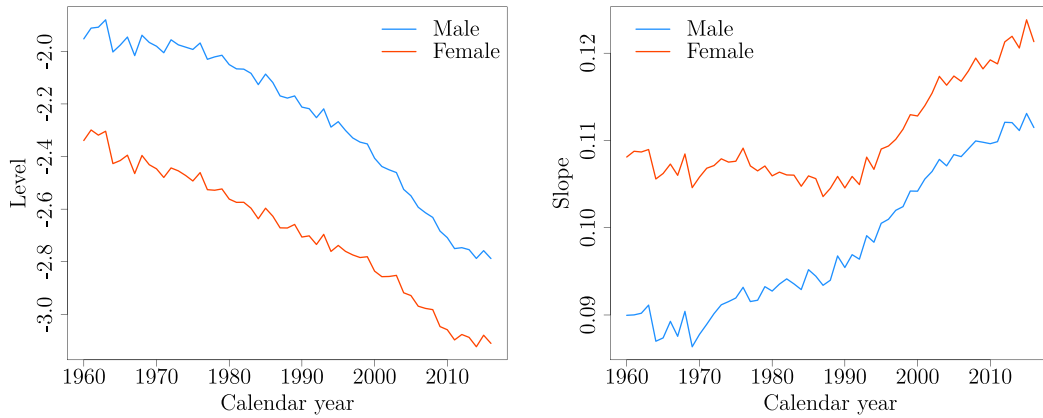
Fig. 4. CBD model fitted to UK male and female mortality. Left panel shows the mortality level, $\hat{k}_{t,1}$, and right panel shows the mortality slope, $\hat{k}_{t,2}$.

conclude that the cointegrating relation is always mean-reverting with an estimated AR-coefficient of $1 + \hat{\beta}'\hat{\alpha} = 0.89$, implying a strong degree of mean reversion. Due to non-identifiability, no further inference can be drawn about the nature of the joint behaviour under the cointegrated Lee–Carter model.

5.2. Cairns–Blake–Dowd application

In contrast to the Lee–Carter model, the CBD model uses two, fully identified factors to describe the mortality evolution of the population under study. In the following we use the CBD model of Section 2.2 as the basis for a four-dimensional cointegration analyses of UK male and female mortality. The higher dimension opens for a richer, yet still interpretable, set of relations while factor identifiability enables the formulation of testable hypotheses.

The CBD model is intended for modelling of pensioners' mortality only and, consequently, we apply it to ages 60–100, rather than the full age span. Fig. 4 shows the two mortality factors of the CBD model estimated separately for UK males and females over the period 1960–2016. The first factor represents the level of mortality and the second factor represents the slope of the mortality curve. Both factors show a clear trend over time. Not surprisingly, the level is generally declining reflecting an overall decrease in mortality over the past decades for both genders. We note that the profile of the level factor bears some resemblance to that of the mortality index of the Lee–Carter model displayed in Fig. 3. However, in contrast to the mortality index of the Lee–Carter model the level factor of the CBD model represents the *absolute* level of mortality and it is therefore markedly higher for males than for females. The slope parameter, shown in the right panel of Fig. 4, is generally upward trending indicating greater mortality improvements at younger ages than at older ages. While we observe a somewhat stable difference between male and female levels of mortality, there appears to be no obvious relation between the slope parameters for the two genders.

5.2.1. Cointegration rank and deterministic structure

Proceeding as in Section 5.1.1, we first verify that the components of $\hat{\boldsymbol{k}}_t = \left(\hat{k}_{t,1}^{\male}, \hat{k}_{t,1}^{\female}, \hat{k}_{t,2}^{\male}, \hat{k}_{t,2}^{\female}\right)'$ are I(1)-processes (test results not shown).

The next step of the analysis is to determine a suitable VECM for the composite index in which we can subsequently formulate and test specific hypotheses. We employ again the VECM of (4.2) and test jointly for cointegration rank and deterministic structure in this model. Table 4 shows the results of Johansen's trace test for rank($\boldsymbol{\Pi}$) $\leq r$ against an unrestricted $\boldsymbol{\Pi}$ for the two deterministic

$$
\begin{array}{ccccccccc}
H_1(0) & \subset & H_1(1) & \subset & H_1(2) & \subset & H_1(3) & \subset & H_1(4) \\
\| & & \cap & & \cap & & \cap & & \cap \\
H_0^*(0) & \subset & H_0^*(1) & \subset & H_0^*(2) & \subset & H_0^*(3) & \subset & H_0^*(4)
\end{array}
$$

Fig. 5. Relation between the I(1)-models considered in Table 4.

Table 4
The trace test for cointegration rank for deterministic terms of class H_1 and H_0^*. The table shows the likelihood ratio test statistic for test of $H(r)$ in $H(4)$ for $r = 0, 1, 2, 3$. The critical value is the 95%-quantile of the limiting distribution given in Section 15.3 of Johansen (1995).

Model	Cointegrating relations	Deterministic term	Trace statistic	Critical value
$H_1(0)$	0	$\boldsymbol{\theta}_0$	80.29	47.21
$H_1(1)$	1	$\boldsymbol{\theta}_0$	38.87	29.38
$H_1(2)$	2	$\boldsymbol{\theta}_0$	12.95	15.34
$H_1(3)$	3	$\boldsymbol{\theta}_0$	0.51	3.84
$H_0^*(0)$	0	$\boldsymbol{\theta}_0$	80.31	62.61
$H_0^*(1)$	1	$\boldsymbol{\theta}_0 + \boldsymbol{\alpha}\rho_1 t$	38.89	42.20
$H_0^*(2)$	2	$\boldsymbol{\theta}_0 + \boldsymbol{\alpha}\rho_1 t$	12.96	25.47
$H_0^*(3)$	3	$\boldsymbol{\theta}_0 + \boldsymbol{\alpha}\rho_1 t$	0.52	12.39

structures of relevance, H_1 and H_0^*. Fig. 5 shows the relation between the various models; note that $H_1(0) = H_0^*(0)$ corresponds to a (multivariate) random walk with drift customarily used for forecasting in the CBD model, cf. Section 2.2.1.

It can be seen from Table 4 that the random walk hypothesis (rank($\boldsymbol{\Pi}$) $= 0$) is rejected in both $H_1(4)$ and $H_0^*(4)$, and we conclude a cointegration rank of at least one. The smallest model of rank 1, $H_1(1)$, is also rejected, while $H_1(2)$ and $H_0^*(1)$ are both accepted. These two models represent the two smallest acceptable models, cf. Fig. 5. Neither of the models is contained in the other, but $H_0^*(1)$ is arguably the "simpler" model introducing only a single additional trend term while $H_1(2)$ introduces an additional cointegrating relation. Consequently, we adopt $H_0^*(1)$ as our starting model, i.e. the VECM

$$\Delta \boldsymbol{k}_t = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{k}_{t-1} + \rho_1 t\right) + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t, \tag{5.3}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are four-dimensional vectors.

5.2.2. Hypothesis testing and interpretation

Having established that the system has a cointegration rank of one, we can now investigate hypotheses on the nature of the equilibrium correcting relation. We are primarily interested in investigating whether the mortality levels of the two genders enter the cointegrating relation with coefficients of the same magnitude and opposite signs, i.e. whether the distance between

the level parameters is the quantity entering the stable relation. In addition, we are also interested in investigating the degree of dependence between the level and slope parameters which can be formulated as restrictions on the adjustment coefficients. Other hypotheses of interest could be formulated, but we restrict ourselves to these two.

We first formulate the hypothesis of main interest, namely that the first two components of $\boldsymbol{\beta}$ are of the same magnitude and opposite signs. The last two components of $\boldsymbol{\beta}$ are left unrestricted. At the same time we would like to test for the absence of a linear drift in the cointegrating relation ($\rho_1 = 0$), since this term muddles the interpretation of the stationary relation. Hence, we consider the composite hypothesis[10]

$$\mathscr{H}_0 \; : \; (\boldsymbol{\beta}', \rho_1) = (1, -1, \varphi_1, \varphi_2, 0), \tag{5.4}$$

where $\varphi_1, \varphi_2 \in \mathbb{R}$ are unrestricted parameters. The test statistic for this hypothesis is $\chi^2(2)$-distributed with a value of 0.005 and a critical value of 5.99 at a 5%-significance level. Hence, we accept hypothesis (5.4) and obtain the model

$$\begin{pmatrix} \Delta k_{t,1}^{\sigma} \\ \Delta k_{t,1}^{\varphi} \\ \Delta k_{t,2}^{\sigma} \\ \Delta k_{t,2}^{\varphi} \end{pmatrix} = \begin{pmatrix} -0.105 \\ -0.042 \\ 0.002 \\ 0.002 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 21.66 \\ -12.44 \end{pmatrix}' \begin{pmatrix} k_{t-1,1}^{\sigma} \\ k_{t-1,1}^{\varphi} \\ k_{t-1,2}^{\sigma} \\ k_{t-1,2}^{\varphi} \end{pmatrix}$$
$$+ \begin{pmatrix} 0.1103 \\ 0.0368 \\ -0.0015 \\ -0.0024 \end{pmatrix} + \boldsymbol{\varepsilon}_t. \tag{5.5}$$

The second question of interest is whether the joint behaviour of the level and slope parameters can be simplified. In particular, whether the adjustment coefficients for the slope parameters, $k_{t,2}^{\sigma}$ and $k_{t,2}^{\varphi}$, as well as the female level parameter, $k_{t,1}^{\varphi}$, are in fact zero, i.e. whether the slope parameters and the female level parameter are weakly exogenous for the long-run coefficients $\boldsymbol{\beta}$. This would imply that while the slope parameters and the female level parameter influence the long-run relation, the long-run relation has no influence on them. To test the hypothesis of weak exogeneity while retaining the established model, we formulate a simultaneous linear restriction on both the adjustment coefficients $\boldsymbol{\alpha}$ and the long-run coefficients $\boldsymbol{\beta}$[11]

$$\mathscr{H}_0 \; : \; (\boldsymbol{\alpha}', \boldsymbol{\beta}', \rho_1) = (\psi_1, 0, 0, 0, 1, -1, \varphi_1, \varphi_2, 0), \tag{5.6}$$

where $\psi_1, \varphi_1, \varphi_2 \in \mathbb{R}$ are unrestricted parameters. The test statistic for this hypothesis is $\chi^2(5)$-distributed with a value of 9.54 and a critical value of 11.07 at a 5%-significance level. We conclude that both slope parameters are weakly exogenous and are therefore not impacted by the long-run relation. We obtain the following final model

$$\begin{pmatrix} \Delta k_{t,1}^{\sigma} \\ \Delta k_{t,1}^{\varphi} \\ \Delta k_{t,2}^{\sigma} \\ \Delta k_{t,2}^{\varphi} \end{pmatrix} = \begin{pmatrix} -0.089 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 11.78 \\ -2.12 \end{pmatrix}' \begin{pmatrix} k_{t-1,1}^{\sigma} \\ k_{t-1,1}^{\varphi} \\ k_{t-1,2}^{\sigma} \\ k_{t-1,2}^{\varphi} \end{pmatrix} \begin{pmatrix} 0.1061 \\ -0.0138 \\ 0.0004 \\ 0.0002 \end{pmatrix} + \boldsymbol{\varepsilon}_t$$
$$= \begin{pmatrix} -0.089 \\ 0 \\ 0 \\ 0 \end{pmatrix} (s_{t-1} - 1.395) + \begin{pmatrix} -0.0181 \\ -0.0138 \\ 0.0004 \\ 0.0002 \end{pmatrix} + \boldsymbol{\varepsilon}_t, \tag{5.7}$$

---

[10] For an in-depth description of this type of test, we refer to Johansen and Juselius (1992).

[11] Likelihood ratio tests for composite hypotheses on cointegrating relations are covered in Johansen (1991).

where $s_t = \hat{\boldsymbol{\beta}}' \boldsymbol{k}_t = k_{t,1}^{\sigma} - k_{t,1}^{\varphi} + 11.78 k_{t,2}^{\sigma} - 2.12 k_{t,2}^{\varphi}$ and the drift term is decomposed in the stationary mean of the cointegrating relation and the trend in the level of the process, cf. (4.3). The cointegrating relation follows an AR-regression of form (5.2) with an AR-coefficient of $1 + \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\alpha}} = 0.91$.

The model can be interpreted as a "mixture" model where the original system is partitioned into a conditional and a marginal system. The marginal system, consisting of the slope parameters and the female level parameter, evolves as a trivariate random walk with drift. Conditioned on these parameters, the male level parameter evolves as the sum of a random walk with drift (and innovations conditioned on the "marginal" innovations) and an error correction term. The error correction term seeks to maintain the long-run relation between the level and slope parameters, but it does so by affecting only the male level parameter.

5.2.3. Implications for forecasting

Mortality modelling is often performed with the aim of forecasting and cointegration models are often enforced to ensure coherence. In this paper we wish to promote the broader use of cointegration as an inferential tool to obtain insights about mortality factor dynamics. Also, we wish to demonstrate that although cointegration does not necessarily imply coherence (in the strict sense of Section 2.1.2) it might still lead to strongly coupled, joint forecasts. Forecasts which might indeed be more plausible being inferred from data, rather than imposed. In this section we illustrate these points for the model (5.7) obtained above.

For forecasting purposes, it is useful to rewrite the model on VAR-form as

$$\boldsymbol{k}_t = \boldsymbol{\Pi}_1 \boldsymbol{k}_{t-1} + \boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_4 \left(\mathbf{0}, \boldsymbol{\Sigma} \right), \tag{5.8}$$

where $\boldsymbol{\Pi}_1 = \mathbf{I} + \boldsymbol{\alpha}\boldsymbol{\beta}'$, cf. (3.1). From this we can readily generate stochastic forecasts, and we also immediately have the forecasting distribution from which the mean forecast and confidence intervals can be derived,

$$\boldsymbol{k}_{T+h}|\boldsymbol{k}_T \sim \mathcal{N} \left(\boldsymbol{\Pi}_1^h \boldsymbol{k}_T + \sum_{i=0}^{h-1} \boldsymbol{\Pi}_1^i \boldsymbol{\theta}_0, \; \sum_{i=0}^{h-1} \boldsymbol{\Pi}_1^i \boldsymbol{\Sigma} (\boldsymbol{\Pi}_1^i)' \right), \tag{5.9}$$

where $h \geq 1$ is the horizon and $\boldsymbol{k}_T = \hat{\boldsymbol{k}}_T$ is the last value of the estimated indices.

Eq. (5.9) is a general result valid for all VAR(1)-models with constant drift. This is useful for numerical computations, but the mean and variance structures are not easily discerned. Using the cointegrating relations it is possible to obtain more revealing expressions for the model at hand.

First, let $\mathbf{B} = \boldsymbol{\alpha} \left(\boldsymbol{\beta}'\boldsymbol{\alpha} \right)^{-1} \boldsymbol{\beta}'$ and $\mathbf{C} = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp \right)^{-1} \boldsymbol{\alpha}_\perp'$ and notice that $\mathbf{I} = \mathbf{C} + \mathbf{B}$, cf. Chapter 3 of Johansen (1995). Next, observe that $\boldsymbol{\Pi}_1$ has eigenvalues 1 and $\lambda = 1 + \boldsymbol{\beta}'\boldsymbol{\alpha}$ with corresponding eigenspaces span($\boldsymbol{\beta}_\perp$) and span($\boldsymbol{\alpha}$), respectively. Hence, for $\boldsymbol{v} \in \mathbb{R}^4$ and $i \geq 0$,

$$\boldsymbol{\Pi}_1^i \boldsymbol{v} = \boldsymbol{\Pi}_1^i \left(\mathbf{C} + \mathbf{B} \right) \boldsymbol{v} = \boldsymbol{\Pi}_1^i \left(\mathbf{C}\boldsymbol{v} + \mathbf{B}\boldsymbol{v} \right) = \mathbf{C}\boldsymbol{v} + \lambda^i \mathbf{B}\boldsymbol{v}, \tag{5.10}$$

since $\mathbf{C}\boldsymbol{v} \in$ span($\boldsymbol{\beta}_\perp$) and $\mathbf{B}\boldsymbol{v} \in$ span($\boldsymbol{\alpha}$). In words, $\boldsymbol{\Pi}_1$ acts on $\boldsymbol{v}$ by leaving intact its $\boldsymbol{\beta}_\perp$-component, while shrinking its $\boldsymbol{\alpha}$-component by a factor of λ (< 1).

Using (5.10) and the formula for the h first terms of a geometric series $\sum_{i=0}^{h-1} \lambda^i = (1 - \lambda^h)/(1 - \lambda)$, we find the following
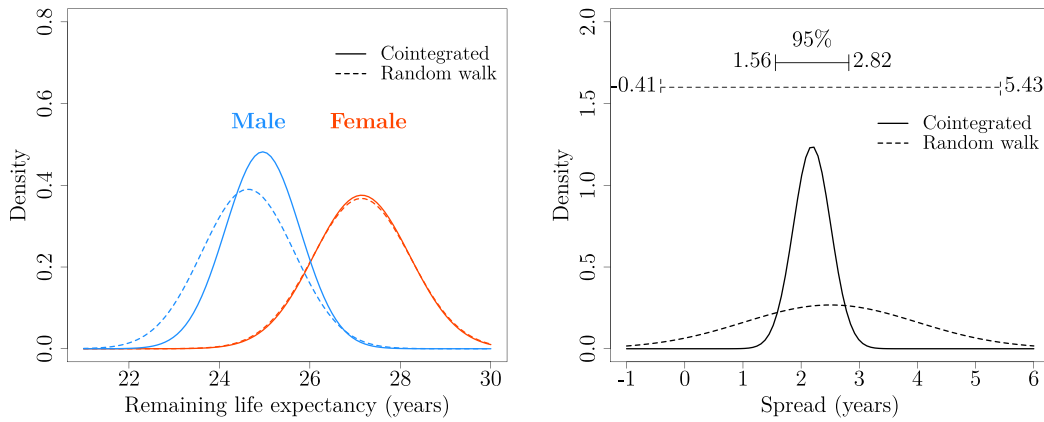
Fig. 6. Forecasting distributions of the cohort remaining life expectancy for the random walk and the cointegration model based on 100,000 simulations. Left panel shows the gender-specific life expectancy, $e_{60}^c(2017)$, and the right panel shows the life expectancy difference (spread) between females and males.

expression for the mean of the forecasting distribution

$$\Pi_1^h \mathbf{k}_T + \sum_{i=0}^{h-1} \Pi_1^i \boldsymbol{\theta}_0 = \mathbf{C}\mathbf{k}_T + \lambda^h \mathbf{B}\mathbf{k}_T + \mathbf{C}\boldsymbol{\theta}_0 h$$
$$+ (1-\lambda)^{-1}(1-\lambda^h)\mathbf{B}\boldsymbol{\theta}_0$$
$$= \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 h + \lambda^h \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\left[\boldsymbol{\beta}'\mathbf{k}_T - \mathbb{E}(\boldsymbol{\beta}'\mathbf{k})\right],$$
(5.11)

where $\boldsymbol{\tau}_0 = \mathbf{C}\mathbf{k}_T + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\mathbb{E}(\boldsymbol{\beta}'\mathbf{k})$, $\boldsymbol{\tau}_1 = \mathbf{C}\boldsymbol{\theta}_0$ and $\mathbb{E}(\boldsymbol{\beta}'\mathbf{k}) = -(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\theta}_0$. We see that asymptotically the mean of the process behaves like a random walk with a drift term that preserves the cointegrating relation. The last term of (5.11) shows how the initial disequilibrium error decays exponentially to zero. This term is not present in the Granger representation of Theorem 3.3, since there we assume that $\boldsymbol{\beta}'\mathbf{k}_T$ is distributed according to its stationary distribution, while in (5.11) we condition on the entire vector $\mathbf{k}_T$.[12]

For the variance of the forecasting distribution, it can be shown that

$$\lim_{h\to\infty} \frac{1}{h} \sum_{i=0}^{h-1} \Pi_1^i \boldsymbol{\Sigma} (\Pi_1^i)' = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}',$$
(5.12)

see pp. 68–69 of Johansen (1995). Thus, asymptotically the variance of the process accumulates linearly, with the variance from the random walk component on $\text{span}(\boldsymbol{\beta}_\perp)$ dominating the total variability.

To give a more intuitive understanding of the role of cointegration, we conclude with a numerical example where we compare the (joint) forecast from the cointegration model (5.7) with separate, gender-specific forecasts based on the bivariate random walk model of Section 2.2.1. We compare the forecasted cohort remaining life expectancy as it depends on the projected mortality surface over a long horizon and it is therefore well suited to capture differences in dependency structures over time.

The two models describe and project the estimated CBD-parameters of Fig. 4 differently. However, the models have similar deterministic structures and we therefore expect similar mean forecasts. This is confirmed by Fig. 6 which shows the forecasting distributions of the cohort remaining life expectancy for a 60-year-old Briton based on 100,000 simulations. The female

distributions align perfectly since the cointegration model (5.7) in fact results in a random walk forecast as well. The male distribution is shifted slightly towards its female counterpart in the case of the cointegrated model. This is the result of the male projection reacting to the perceived "disequilibrium".

As the theoretical analysis showed, the cointegration model generally behaves asymptotically like (coupled) random walks. Therefore, the marginal model for each sex is (at least asymptotically) close to the corresponding random walk model, both in terms of deterministic and stochastic behaviour. The dependency structure of the two models is, however, very different. By construction, the random walk model yields independent forecasts for men and women, while the cointegration model yields highly dependent forecasts. This implies a much more narrow distribution for the difference in life expectancy between men and women under the cointegration model than under the random walk model, as illustrated in the right panel of Fig. 6.

In summary, the cointegration model yields forecasts that are similar to those obtained from the simpler random walk model for each sex, but with a more plausible dependency structure. The cross-gender dependency is achieved by a single cointegrating relation derived from data. The resulting forecasts are well-behaved and empirically justified, but they are not coherent (in the strict mathematical sense). This indicates that the current definition of coherence might be too strict and too specific a requirement, and that other types of "coherence" might be equally good – or even better – when judging the quality of joint forecasts.

6. Concluding remarks

In this paper we have discussed the interlinked concepts of identifiability, coherence and cointegration in the context of multi-population mortality modelling. We have made the point that cointegration has an important role to play as an inferential tool to obtain insights into the joint dynamics of mortality factors. This role goes beyond the typical usage of cointegration as merely a tool to obtain coherent forecasts—defined as forecasts for which the relative age-specific mortality rates converge over time.[13]

Since its introduction in Li and Lee (2005), the concept of coherence has served as the gold standard for joint forecasts of related populations, and many models have been designed with the explicit goal of achieving coherence. At first sight, coherence seems like a reasonable property, but on further inspection it

---

[12] When $\mathbf{k}_t$ has the form (5.8) Theorem 3.3 reads (in terms of expected value) $\mathbb{E}[\mathbf{k}_{T+h}] = \mathbf{C}\mathbf{k}_T + \mathbf{C}\sum_{i=1}^{h}\boldsymbol{\theta}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\sum_{i=0}^{\infty}(\mathbf{I}+\boldsymbol{\alpha}\boldsymbol{\beta}')^i\boldsymbol{\beta}'\boldsymbol{\theta}_0 = \mathbf{C}\mathbf{k}_T + \mathbf{C}\boldsymbol{\theta}_0 h - \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\theta}_0 = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 h$, i.e. (5.11) without the exponentially decaying error-correction term.

[13] Whether the goal of (statistical) models is to gain insights or to forecast is the object of a long-standing, and still highly relevant, debate, see e.g. Breiman (2001) and Shmueli (2010).

appears somewhat arbitrary and specifically tailored to models of the Lee–Carter type with log-linear modelling of mortality rates. Joint forecasts based on other model types, e.g. logistic, can produce equally plausible dependency structures and thus be equally "coherent", even if they lack relative convergence. In our view, insisting on coherence is not a suitable starting point, nor a reasonable restriction, for a joint analysis and might in fact produce forecasts that are at odds with historic data.

In this paper we have focused on two types of linear trend models suitable for analysing the period effect of age–period mortality models, and we have shown how to interpret these models by use of the Granger decomposition. Cointegration is a technical field, and the analysis and interpretation of models and hypotheses are not straightforward. In many mortality models, e.g. Lee–Carter type models, the factor(s) are only semi-identifiable in which case additional difficulties (and pitfalls) arise. From a statistical point of view, it is in general meaningless to test, or impose restrictions, on the long-run coefficients in the context of semi-identifiable models. In contrast, fully identified models give access to the full inferential power of cointegration and in our opinion this is a strong argument in favour of these models if the aim is to gain subject matter insights.

Declaration of competing interest

Acknowledgements

Appendix. Maximum likelihood estimation of the VECM

Consider the VAR(k) model in VECM form with $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ for $p \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$,

$$\Delta \mathbf{y}_t = \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\Phi}\mathbf{D}_t + \boldsymbol{\varepsilon}_t, \tag{A.1}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(0, \boldsymbol{\Sigma})$ and independent. Based on data for $t = 1, \ldots, T$, maximum likelihood estimates of the freely varying parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \ldots \boldsymbol{\Gamma}_{k-1}, \boldsymbol{\Phi}, \boldsymbol{\Sigma})$ can be obtained by reduced rank regression. Following Juselius (2006), we introduce the shorthand notation

$$\mathbf{Z}_{0t} = \Delta \mathbf{y}_t, \tag{A.2}$$

$$\mathbf{Z}_{1t} = \mathbf{y}_{t-1}, \tag{A.3}$$

$$\mathbf{Z}_{2t} = \left[\Delta \mathbf{y}'_{t-1}, \ldots, \Delta \mathbf{y}'_{t-k+1}, \mathbf{D}'_t \right], \tag{A.4}$$

and write (A.1) on the form

$$\mathbf{Z}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Z}_{1t} + \boldsymbol{\Psi}\mathbf{Z}_{2t} + \boldsymbol{\varepsilon}_t, \tag{A.5}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{k-1}, \boldsymbol{\Phi}]$. Define the product moment matrices

$$\mathbf{M}_{ij} = T^{-1} \sum_{t=1}^{T} \mathbf{Z}_{it} \mathbf{Z}'_{jt}, \quad i, j = 0, 1, 2, \tag{A.6}$$

and the sample-covariance matrices

$$\mathbf{S}_{ij} = \mathbf{M}_{ij} - \mathbf{M}_{i2} \mathbf{M}_{22}^{-1} \mathbf{M}_{2j}, \quad i, j = 0, 1. \tag{A.7}$$

The maximum likelihood estimator of $\boldsymbol{\beta}$ is found by solving the eigenvalue problem

$$|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10} \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{01}| = 0, \tag{A.8}$$

for eigenvalues $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_p > 0$ with associated eigenvectors $\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_p$ that satisfy

$$\hat{\lambda}_i \boldsymbol{S}_{11} \hat{\boldsymbol{v}}_i = \boldsymbol{S}_{10} \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{01}, \quad i = 1, \ldots, p, \tag{A.9}$$

$$\hat{\boldsymbol{V}} \boldsymbol{S}_{11} \hat{\boldsymbol{V}} = \boldsymbol{I}, \tag{A.10}$$

where $\hat{\boldsymbol{V}} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_p)$. The cointegrating relations $\hat{\boldsymbol{\beta}}$ are given by the first r eigenvectors

$$\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_r). \tag{A.11}$$

The remaining parameters are estimated as

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \boldsymbol{S}_{01} \boldsymbol{\beta}(\boldsymbol{\beta}' \boldsymbol{S}_{11} \boldsymbol{\beta})^{-1}, \tag{A.12}$$

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{S}_{00} - \boldsymbol{\alpha}(\boldsymbol{\beta}' \boldsymbol{S}_{11} \boldsymbol{\beta})\boldsymbol{\alpha}', \tag{A.13}$$

$$\hat{\boldsymbol{\Psi}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{M}_{02} \boldsymbol{M}_{22}^{-1} - \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{M}_{12} \boldsymbol{M}_{22}^{-1}, \tag{A.14}$$

with $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}})$.

Trace statistic for test of cointegration rank

Recall that $H(r)$ denotes the hypothesis that rank($\boldsymbol{\Pi}$) $\leq r$, or equivalently that $\boldsymbol{\Pi}$ has at most r non-zero eigenvalues. The likelihood ratio test statistic for $H(r)$ in $H(p)$, known as the *trace statistic*, is given by

$$LR_{trace}(r) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i), \tag{A.15}$$

where $\hat{\lambda}_i$ are the eigenvalues found by solving (A.8). The asymptotic distribution of the trace statistic depends on the deterministic terms. Critical values are tabulated in Section 15.3 of Johansen (1995) for the linear models in Section 3.2. Intuitively, if rank($\boldsymbol{\Pi}$) $= r$ then $LR_{trace}(r)$ will be small, since $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ will all be close to zero. Conversely, the test statistic will be large if rank($\boldsymbol{\Pi}$) $> r$, since at least $\hat{\lambda}_{r+1}$ will deviate from zero.

References

Arnold, S., Sherris, M., 2016. International cause-specific mortality rates: New insights from a cointegration analysis. Astin Bull. 46 (1), 9–38.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., Vaupel, J.W., 2017. Coherent forecasts of mortality with compositional data analysis. Demogr. Res. 37, 527–566.

Beutner, E., Reese, S., Urbain, J.-P., 2017. Identifiability issues of age–period and age–period–cohort models of the Lee–Carter type. Insurance Math. Econom. 75, 117–125.

Breiman, L., 2001. Statistical modeling: The two cultures. Statist. Sci. 16 (3), 199–215.

Brouhns, N., Denuit, M., Vermunt, J., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance Math. Econom. 31 (3), 373–393.

Cairns, A., Blake, D., Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. J. Risk Insurance 73 (4), 687–718.

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

Cairns, A., Dowd, K., Blake, D., Coughlan, G.D., 2011b. Longevity hedge effectiveness: a decomposition. Quant. Finance 14 (2), 217–235.

Carter, L.R., Lee, R.D., 1992. Modeling and forecasting US sex differentials in mortality. Int. J. Forecast. 8 (3), 393–411.

Chen, R.Y., Millossovich, P., 2018. Sex-specific mortality forecasting for UK countries: a coherent approach. Eur. Actuar. J. 8 (1), 69–95.

Currie, I., 2016. On fitting generalized linear and non-linear models of mortality. Scand. Actuar. J. 2016 (4), 356–383.

Darkiewicz, G., Hoedemakers, T., 2004. How the Co-integration Analysis Can Help in Mortality Forecasting. Katholieke Universiteit Leuven, Department of Applied Economics.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Khalaf-Allah, M., 2011. A gravity model of mortality rates for two related populations. N. Am. Actuar. J. 15 (2), 334–356.

Engle, R., Granger, C., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica 55 (2), 251–276.

Gaille, S., Sherris, M., 2011. Modelling mortality with common stochastic long-run trends. Geneva Pap. Risk Insur.- Issues Pract. 36 (4), 595–621.

Hansen, P.R., 2005. Granger's representation theorem: A closed-form expression for I(1) processes. Econometrics J. 8 (1), 23–38.

Human Mortality Database, 2019. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded April 2019. http://www.mortality.org.

Hunt, A., Blake, D., 2015a. Identifiability in Age/Period Mortality Models (WP). The Pensions Institute, Cass Business School, City University London, PI-1508.

Hunt, A., Blake, D., 2015b. Identifiability in Age/period/cohort Mortality Models (WP). The Pensions Institute, Cass Business School, City University London, PI-1509.

Hunt, A., Blake, D., 2015c. Modelling longevity bonds: Analysing the Swiss Re Kortis bond. Insurance Math. Econom. 63, 12–29.

Hunt, A., Blake, D., 2017. Identifiability, cointegration and the gravity model. Insurance Math. Econom. 78, 360–368.

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

Janssen, F., van Wissen, L.J., Kunst, A.E., 2013. Including the smoking epidemic in internationally coherent mortality projections. Demography 50 (4), 1341–1362.

Jarner, S., Kryger, E., 2011. Modelling adult mortality in small populations: The saint model. Astin Bull. 41 (2), 377–418.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.

Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.

Johansen, S., Juselius, K., 1992. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. J. Econometrics 53 (1), 211–244.

Juselius, K., 2006. The Cointegrated VAR Model: Methodology and Applications. Oxford University Press.

Kleinow, T., 2015. A common age effect model for the mortality of multiple populations. Insurance Math. Econom. 63, 147–152.

Kuang, D., Nielsen, B., Nielsen, J., 2008. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 987–991.

Lazar, D., Denuit, M., 2009. A multivariate time series approach to projected life tables. Appl. Stoch. Models Bus. Ind. 25 (6), 806–823.

Lee, R., Carter, L., 1992. Modeling and forecasting U.S. mortality. J. Amer. Statist. Assoc. 87 (419), 659–675.

Li, J., Hardy, M., 2011. Measuring basis risk in longevity hedges. N. Am. Actuar. J. 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.

Li, H., Lu, Y., 2017. Coherent forecasting of mortality rates: A sparse vector-autoregression approach. Astin Bull. 47 (2), 563–600.

Lindsey, J.K., 1996. Parametric Statistical Inference. Oxford University Press.

Nielsen, B., Nielsen, J., 2014. Identification and forecasting in mortality models. Sci. World J. 2014.

Njenga, C.N., Sherris, M., 2011. Longevity risk and the econometric analysis of mortality trends and volatility. Asia-Pac. J. Risk Insur. 5 (2), Article 2.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.

Said, S., Dickey, D., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71 (3), 599–607.

Salhi, Y., Loisel, S., 2017. Basis risk modelling: A cointegration-based approach. Statistics 51 (1), 205–221.

Shmueli, G., 2010. To explain or to predict? Statist. Sci. 25 (3), 289–310.

Thatcher, A., 1999. The long-term pattern of adult mortality and the highest attained age. J. R. Stat. Soc. Ser. A 162 (1), 5–43.

Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405 (6788), 789–792.

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

Zhou, R., Wang, Y., Kaufhold, K., Johnny, L., Tan, K., 2014. Modeling period effects in multi-population mortality models: Applications to solvency II. N. Am. Actuar. J. 18 (1), 150–167.