Cass Business School
City of London

Cass means business

Faculty of Actuarial Science and Statistics

Lee-Carter Mortality Forecasting Incorporating Bivariate Time Series.

A. Renshaw and S. Haberman

Actuarial Research Paper No. 153

December 2003

ISBN 1-901615-75-8

Cass Business School
106 Bunhill Row
London EC1Y 8TZ
T +44 (0)20 7040 8470
www.cass.city.ac.uk

Lee-Carter mortality forecasting incorporating bivariate time series

Arthur Renshaw and Steven Haberman

Cass Business School, City University, London, UK

Summary.
We investigate the feasibility of projecting age specific mortality rate by taking a multivariate time series approach to the time components in the single valued decomposition of a matrix of suitably transformed empirical log mortality rates.

Keywords: Mortality forecasting; Lee-Carter methodology; Multivariate time series

1. Introduction

The Lee-Carter (LC) approach to mortality forecasting (Lee and Carter (1992), Lee (2000)) makes use of the singular value decomposition (SVD) of a matrix of suitably centred log mortality rates, cross-classified by age and period (calendar year). Then, assuming that the resulting pair of primary singular vectors adequately captures the salient age-period (and cohort) pattern, standard univariate time series methods are applied to the primary period component vector to generate period specific forecasts. Such forecasts are meaningful in the sense that the modelled age specific mortality trends are projected into the future. Thus, Tuljapurkar *et al.* (2000) report on the application of this methodology to mortality rates for each of the G7 countries, in a comparative study of mortality decline between these countries. In addition, Tuljapurkar *et al* justify forecasts based on the primary singular vectors only on the basis of the proportion of the total temporal variance explained by just the first component of the SVD, which is found to be over 94% of the mortality variation, in every G7 country.

However, on applying the LC approach to comparable England & Wales mortality experiences (for which the proportion of the total temporal variances explained by the first SVD component is again over 94%), Renshaw and Haberman (2003a) report on the failure of the first SVD component vectors to capture important aspects of the data, together with the presence of noteworthy residual patterns in the second SVD component vectors. As a consequence, Renshaw and Haberman (2003b) investigate an augmented version of the LC approach by additionally incorporating the second set of SVD vectors, and applying separate univariate ARIMA processes to the first two period component vectors to generate forecasts. Further, Booth *et al.* (2002) in modelling the Australian mortality experience, also resort to the inclusion of the second and even third SVD component vectors in order to improve the quality of fit, but do not attempt forecasting on the basis of this more detailed model.

In this paper, we investigate the potential role of multivariate (bivariate) time series methods for generating forecasts when the LC approach is augmented to include the secondary SVD component vectors. In order to implement these, we make use of the multivariate facilities (and univariate facilities where necessary) of the time series computer package MICROFIT (Pesaran and Pesaran (1997)) and we supplement these facilities, specifically when constructing limits for the (component) time series forecasts. In common with Renshaw and Haberman (2003a), we again focus on the England & Wales mortality experiences for the two genders separately.

2. An augmented LC approach to mortality forecasting

Let m_{xt} denote the central rate of mortality, cross-classified by age x, grouped into k ordered categories, and by period $t = t_1, t_2, ..., t_n = t_1 + h - 1$, with range $h = t_n - t_1 + 1$. Define

$$\alpha_x = \log \prod_{t=t_1}^{t_n} m_{xt}^{1/h}$$

and consider the SVD of the matrix $[z_{xt}] = [\log m_{xt} - \alpha_{xt}]$. Thus we can write

$$\log m_{xt} = \alpha_x + \sum_{i=1}^{\tau} \beta_x^{(i)} \kappa_t^{(i)} + \varepsilon_{xt}\,,\ \tau < \min(h,k) \qquad (1)$$

where $\beta_x^{(i)}, \kappa_t^{(i)}$ denote the respective left and right (ordered) singular vectors, subject to the constraints

$$\sum_{all\,x} \beta_x^{(i)} = 1, \sum_{t=t_1}^{t_n} \kappa_t^{(i)} = 0, \forall i \qquad (2)$$

and where ε_{xt} denotes the residual SVD components. The RHS of (1) comprises a parameterised systematic (non-random) component and an additive error component ε_{xt}. The case $\tau = 1$ (with the redundant prefix i omitted) corresponds to the 'basic' LC approach.

The model (1) is fitted to empirical mortality rates $\hat{m}_{xt}$ by the SVD of $[\hat{z}_{xt}]$, subject to the constraints (2), and, in accordance with LC practice, the resulting $\kappa_t^{(1)}$ estimates are further adjusted to ensure that the actual total deaths are identical with the total expected deaths for each t, as a means of improving the quality of fit. For a discussion of three case studies including the England and Wales male experience, for $\tau = 1, 2$ and using univariate time series methods, together with comparative studies of two other related modelling approaches, see Renshaw and Haberman (2003b).

In this paper, we focus on the case $\tau = 2$ and seek to generate forecasts by modelling $\{\hat{\kappa}_t^{(i)} : i = 1,2\}$, using bivariate time series methods. Denoting the resulting forecasts by $\{\dot{\kappa}_{t_n+s}^{(i)} : s > 0; i = 1,2\}$, projected mortality rates

$$\dot{m}_{x,t_n+s} = \hat{m}_{xt_n} \exp \sum_{i=1}^{2} \hat{\beta}_x^{(i)} (\dot{\kappa}_{t_n+s}^{(i)} - \hat{\kappa}_{t_n}^{(i)}); s > 0$$

are then computed by alignment to the latest available empirical mortality rates $\hat{m}_{xt_n}$. Given the nature of this expression, in keeping with the terminology of UK actuarial mortality studies, we could refer to the expression

$$\log F(x, t_n + s) = \sum_{i=1}^{2} \hat{\beta}_x^{(i)} (\dot{\kappa}_{t_n+s}^{(i)} - \hat{\kappa}_{t_n}^{(i)}); s > 0$$

as a log mortality reduction factor: see Renshaw and Haberman (2003c) for further discussion.

3. Time series modelling for the England & Wales mortality experience

3.1. *Preliminaries*

The data comprise the number of deaths with matching person-years of exposure to the risk of death as supplied by the Office of National Statistics. Cross-classification is by age, categorised in years {<1, 1-4, 5-9, 10-14, … , 80-84, 85+}, and by individual calendar years, from 1950 to 1998 inclusive, for each gender. The age specific mortality trends are plotted in Renshaw and Haberman (2003a). Although the proportion of the temporal variance in the transformed death rates explained by the first SVD component exceeds 94% (Renshaw and Haberman (2003a)), we justify the retention of the second SVD component (for both genders) on the basis of the distinctive patterns in its constituent parts. In particular, the SVD component estimates $\{\hat{\kappa}_t^{(i)} : i = 1,2\}$, (incorporating the adjustment to $\hat{\kappa}_t^{(1)}$), are depicted in each of the respective sets of left and right hand frames in Figs 1 & 2. (A full explanation of Cases I-III is given in Sections 3.2 and 3.3, to follow). We note that Booth *et al.* (2002) report a similar U-shaped pattern for $\{\hat{\kappa}_t^{(2)}\}$ in their SVD analysis of annual age-specific death rates for Australian, 1968-99.

Further details of the multivariate time series methods used below, together with their implementation using MICROFIT, are given in Pesaran and Pesaran (1997). Other relevant details are given in Chapters 11, 19 and 20 of Hamilton (1994). A distillation of specific bivariate time series is given in an appendix to a fuller working paper, available from the authors.

3.2. *Vector autoregressive modelling*

We consider the augmented vector autoregressive model of order p

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t$$

where $\mathbf{y}_t$ is a vector of jointly determined dependent variables, $\boldsymbol{\varepsilon}_t$ is a vector of errors, $\mathbf{a}_0, \mathbf{a}_1$ are vectors of intercept and trend parameters and Φ_i are square matrices of autoregressive parameters. It is assumed that the vector of errors $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{\Omega})$ with symmetric positive define $\mathbf{\Omega}$, and that all roots of

$$|\mathbf{I} - \sum_{i=1}^{p} \mathbf{\Phi}_i \lambda^i | = 0$$

fall outside the unit circle for stability.

We focus on bivariate models for which $\boldsymbol{\kappa}'_t = (\kappa_t^{(1)}, \kappa_t^{(2)})$ and consider Case I: $\mathbf{y}_t = \Delta \boldsymbol{\kappa}_t$ (where Δ denotes the differencing operator) and Case II: $\mathbf{y}_t = \boldsymbol{\kappa}_t$. Given the extensive use of first order integrated I(1) processes in the basic LC approach, and the use of two such separate I(1) processes when $\tau = 2$ (Renshaw and Haberman (2003b)), it is likely that Case II would fail to satisfy the unit root criterion. Further evidence for this, is provided by the Dicky-Fuller (DF) and the augmented Dicky-Fuller (ADF(p)) tests for unit roots, based on the respective (univariate) regression models

$$y_t = \mathrm{a}_0 + \mathrm{a}_1 t + \rho y_{t-1} + \varepsilon_t$$

and

$$y_t = \varsigma_1 \Delta y_{t-1} + \varsigma_2 \Delta y_{t-2} + \ldots + \varsigma_{p-1} \Delta y_{t-p+1} + \mathrm{a}_0 + \mathrm{a}_1 t + \rho y_{t-1} + \varepsilon_t$$

details of which are presented in Table 1. Here the unit root (null) hypothesis $H_0 : \rho = 1$ is not rejected, with the one exception, using the estimated OLS t test statistic, determined by reference to the Akaike information, Schwarz Bayesian and Hannan-Quinn criteria. Details of the test statistics are discussed in Chapter 17 of Hamilton (1994) and in the appendix to our fuller working paper, available from the authors. The exception concerns the unit root tests for the first SVD component for the female experience, reflected below, in our choice of simple regression model for this situation, based on further considerations.

In both cases, we set $p = 1$. This is justified partly on pragmatic grounds, consistent with the statistical significance of model parameters, and partly on the basis of the Akaike information criterion, the Schwarz Bayesian criterion and likelihood ratio test for selecting the order of the model, which we report in Tables 2 and 3. Thus for Case I (Table 2), the likelihood ratio test rejects the null hypothesis $H_0 : p = 0$ but not the null hypothesis $H_0 : p = 1$ for both genders, while one of the two Akaike and Schwarz criteria are also supportive of $p = 1$, for each gender. For Case II (Table 3) and the male experience, the likelihood ratio test rejects $H_0 : p = 0$ but not $H_0 : p = 1$ and both the Akaike and Schwarz criteria are supportive of $p = 1$. On the other hand, for the female experience, these criteria are essentially supportive of $p = 2$. However, on fitting this model, it is readily seen to be heavily over-parameterised.

Define

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \Phi \mathbf{y}_{t-1} + \varepsilon_t \tag{3}$$

with

$$\mathbf{a}_0 = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix}, \mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \Phi_1 = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

Case I : $\mathbf{y}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$

The Wald test for restrictions imposed on parameters in the form of the null hypothesis $H_0 : a_{11} = 0, \phi_{12} = 0, \phi_{21} = 0$ implies that H_0 is not rejected on the basis of the statistic $\chi_3^2 = 3.50\,(p - value = .32)$ for males and $\chi_3^2 = 6.76\,(p - value = .08)$ for females. Thus, to reflect this hypothesis, we are justified in fitting an ARIMA(1,1,0) process to $\{\kappa_t^{(1)}\}$ and a separate ARIMA(1,1,0) process plus deterministic trend term to $\{\kappa_t^{(2)}\}$. Details of the parameter estimates are recorded in Table 2 and the resulting time series projections to the year 2020 depicted in Figs 1(a) & 2(a). In particular we note the quadratic nature of the projections for $\{\kappa_t^{(2)}\}$, which is discussed, together with other implications for mortality rate projections, in Section 4.

Case II : $\mathbf{y}_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$

For males, the block Granger non-causality null hypothesis $H_0 : \phi_{21} = 0$ is not rejected on the basis of the likelihood ratio test statistic $\chi_1^2 = 0.021\,(p - value = .88)$ so that the model, reported in Table 3, is fitted to reflect this hypothesis. Thus $\kappa_{t-1}^{(1)}$ and $\kappa_{t-1}^{(2)}$ impact directly on $\kappa_t^{(1)}$, but only $\kappa_{t-1}^{(2)}$ impacts directly on $\kappa_t^{(2)}$. Projections to the year 2020, based on this model, are depicted in Fig 1(b). Here, the approximate linearity of the projection of $\{\kappa_t^{(2)}\}$ is in marked contrast to the quadratic projection of

Case I. However, the narrowness of the forecast limits, expressly for the first component, is a cause for concern. This is discussed further in Section 4.

For females, the choice of model is different. Specifically, the Wald test for restrictions imposed on parameters in the form of the null hypothesis $H_0: \phi_{11} = 0, \phi_{12} = 0, \phi_{21} = 0$ is not rejected on the basis of the Wald test statistic $\chi_3^2 = 2.48\,(p - value = .48)$. Thus, under this hypothesis we are justified in fitting a straight line to $\{\kappa_t^{(1)}\}$ and a completely separate ARIMA(1,1,0) process plus a deterministic trend term to $\{\kappa_t^{(2)}\}$. Details are presented in Table 3 and the associated projections depicted in Fig 2(b). Comparison with Fig 1(b) for the male experience reveals similar patterns and concerns about the narrowness of the first component prediction limits, this time based on a separate univariate analysis, which are discussed further in Section 4.

3.3. *Vector error correction model (co-integration)*

We consider the following simple version of the vector error correction model

$$\Delta\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t - \Pi\mathbf{y}_{t-1} + \sum_{i=1}^{p-1}\Gamma_i\Delta\mathbf{y}_{t-1} + \varepsilon_t$$

where $\mathbf{y}_t$ is a vector of jointly determined I(1) variables, $\mathbf{a}_0, \mathbf{a}_1$ are vectors of intercept and trend parameters, Π is a square long-run multiplier matrix of parameters and Γ_i are square matrices of parameters capturing short term dynamic effects. It is assumed that the vector of errors $\varepsilon_t \sim \text{iidN}(\mathbf{0}, \Omega)$, with symmetric positive definite Ω. There are five options to consider in general, determined by the combination of constraints imposed on the intercept and trend parameters $\mathbf{a}_0, \mathbf{a}_1$. On the basis of trial and error, we conclude that just two of these are appropriate. Again the focus is on bivariate models with $\mathbf{y}_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$ and we set p = 1 for the same reasons as before.

Case III: $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 = \mathbf{0}$ (unrestricted intercepts and no trends)

For this case

$$\Delta\mathbf{y}_t = \mathbf{a}_0 - \Pi\mathbf{y}_{t-1} + \varepsilon_t .$$

Let $\Pi = \alpha\beta'$ where α and β are $2 \times r$ matrices and $r = \text{rank}(\Pi)$. The columns of β define co-integration vectors and the rows of $\beta'\mathbf{y}_t$ define co-integration relationships. Because of the nature of the multiplicative decomposition of Π, it is necessary to place additional constraints on the co-integration vectors prior to estimation. We use Johansen estimation through MICROFIT. The value of r is determined first with the aid of co-integration likelihood ratio tests based on maximum eigenvalue and trace statistics, together with the Akaike information, Schwartz Bayesian and Hannan-Quinn model selection criteria, reported in Table 4. On the basis of complete agreement between these tests and criteria, we set r =1, resulting in the parameter estimates reported in Table 4 and projections depicted in Figs 1 (c) & 2(c).

Case IV: $\mathbf{a}_0 \neq \mathbf{0}$ *and* $\mathbf{a}_1 = \Pi\gamma$ (unrestricted intercepts and restricted trends)

For this case

$$\Delta\mathbf{y}_t = \mathbf{a}_0 - \Pi(\mathbf{y}_{t-1} - \gamma t) + \varepsilon_t$$

where γ is 2×1. On introducing $\Pi=\alpha\beta'$ it follows that the rows of $\beta'(\mathbf{y}_t-\gamma t)$ define the co-integration relationships. This time, on the basis of co-integration likelihood ratio tests and model selection criteria, we set $r = 2$ when fitting the model. The details are reported in Table 5. Since, when $r = 2$, this model is synonymous with a re-parameterisation of model (3) under Case II, we do not pursue this case further.

4. Discussion

The time series models used to construct the log mortality reduction factors

$$\log F(x,t_n+s)=\sum_{i=1}^{c}\hat{\beta}_x^{(i)}(\dot{\kappa}_{t_n+s}^{(i)}-\hat{\kappa}_{t_n}^{(i)});s>0$$

are summarised as follows:

	Male experience	Female experience
Case I	ARIMA(1,1,0) / ARIMA(1,1,0) + trend	ARIMA(1,1,0) / ARIMA(1,1,0) + trend
Case II	VAR(1) + block Granger non-causality	trend only / ARIMA(1,0,0) + trend
Case III	Co-integration model: (rank = 1) unrestricted intercept, no trend	Co-integration model: (rank = 1) unrestricted intercept, no trend
LC2	ARIMA(1,1,0) / ARIMA(1,1,1)	-
LC	ARIMA(1,1,0)	ARIMA(1,1,0)

Thus $c = 2$, with the exception of the basic LC approach, for which $c = 1$. Whereas a bivariate time series approach is adopted in Cases I to III, model selection procedures within this approach, justify the retention of a bivariate model in Case II for males only and in Case III for both genders, otherwise separate univariate time series are fitted in the remaining cases. Under the two component Lee-Carter approach for males only, denoted LC2 (Renshaw and Haberman (2003b)), the comparable modelling equivalent to Case II, separate univariate time series are fitted by design. Under the LC or LC2 approaches, coupled to univariate ARIMA(0,1,0) processes (parameters $\lambda^{(i)}$), the log mortality reduction factor reduces to

$$\log F(x,t_n+s)=\left(\sum_{i=1}^{c}\hat{\beta}_x^{(i)}\hat{\lambda}^{(i)}\right)s,\,s>0\,,$$

which has the same mathematical form as the log mortality reduction factor used in the GLM approach of Renshaw and Haberman (2003a), using either a hinged or straight line predictor.

For males, the age specific period profiles of the resulting log mortality reduction factors for the three cases (I-III) are depicted in Fig. 3(a)-3(c). These bear direct comparison with each other and with the two equivalent profiles: Fig. 2(h) and Fig. 3(f) in Renshaw and Haberman (2003a), generated respectively using the basic LC approach (utilising only the first SVD component vectors) and using a generalised linear model (GLM) regression based approach (utilising a hinged predictor).

In order to extend this comparison, the year 2020 projections of all five log mortality reduction factors are superimposed and depicted in Fig 3(d). For completeness, we additionally include the 2020 projection of the log mortality reduction factor based on the separate, two component, univariate time series Lee-Carter type analysis, denoted LC2, reported in Renshaw and Haberman (2003b): using respective ARIMA(1,1,0) and ARIMA(1,1,1) processes to generate forecasts for the first and second SVD components. In Fig 3(d), we have highlighted the basic LC projection, which, although accounting for over 94% of the total temporal variance, fails to capture the same degree of age variation as the other five approaches. In particular, as noted in Renshaw and Haberman (2003a), the basic LC approach fails to capture and project the reported rise in recent male mortality rates, in the 20 to 34 age bands, which is consistent with positive, or near positive valued log mortality reduction factors in this age range. In contrast, the extreme nature of the projections generated by Case I, which is directly attributable to the quadratic nature of the $\kappa_t^{(2)}$ projections, is clearly implausible. The narrowness of the forecast limits in Case II, especially for the first component, is both unrealistic and a topic for further investigation. We note that no allowance is made for the uncertainty in the parameters used when computing the mean square forecast error. However, this is also true of Case III and we believe this effect to be small. As already noted in Section 3.2, there is evidence for rejecting this model because it fails to satisfy the unit root criterion for stability. Further, the pattern of residuals in the early part of the period (1950s) is also a cause for concern. For a comparison of Case II and Case III (non-standardised) residuals, see Fig. 4. Under the LC2 univariate time series approach, which cannot now be justified from the wider perspective of bivariate time series modelling, there is a failure to capture the more favourable mortality projections, in the 45-49 age band and above, associated with the other approaches, with the exception of the basic LC approach. On the basis of these considerations, there is a case for choosing Case III and comparing the resulting predictions with those generated by the GLM (hinged) approach.

The equivalent age specific profiles and year 2020 projections for females are depicted in Fig. 5. For females, Case I is also implausible because of unrealistic extreme mortality projections at certain ages, again due to the curvature of the $\kappa_t^{(2)}$ time series forecasts. Similarly, narrow forecast limits are also a feature of the first component Case II female experience, albeit this time based on separate univariate time series analysis as opposed to the bivariate analysis for males. In contrast to the male experience, the basic LC approach and GLM (line) approach generate near identical predictions across the whole of the age range, which, in turn, differ to any noteworthy extent from the Case III (and Case II) predictions, only in the 15-19 to 35-39 age bands. Within these age bands, Case III predicts heavier mortality than the LC and GLM approaches, but not to the same extent as in the male experience.

At the practical level, details of the multivariate time series analysis described above have been dictated by the choice of menus available in the computer software package MICROFIT (Pesaran and Pesaran (1997)). In particular, we note that the Lagrange multiplier test provides evidence of residual serial correlation, present mainly in the first component dependent variable (almost exclusively so for males), in both of the bivariate analyses reported under Cases II and III. A possible means of removing this effect is by increasing the order of the respective multivariate models. However, such a solution runs counter to the tests applied initially to determine p and results in substantial over parameterisation. Specifically for males, while this approach is successful in removing residual serial correlation effects in Case III, three of the four constituent components of the additional matrix Γ_1 are not statistically significant when estimated. Further, both versions ($p = 1$ and $p = 2$) in Case III generate (effectively) identical projections in the age bands 40 and above, while the over parameterised version ($p = 2$) generates marginally less optimistic forecasts than its counterpart ($p = 1$) in the age bands below 40.

5. Conclusions

Given the wide interest in the application of the LC approach to generate age-specific mortality forecasts, it is perhaps somewhat surprising that no previous attempt appears to have been made to incorporate finer age specific detail into the forecasting process through the inclusion of secondary SVD components and then undertake a subsequent multivariate time series analysis. As illustrated in this application, by incorporating distinctive features in the secondary SVD components and not restricting the methodology to the first SVD components solely on the basis of a high proportion of the total temporal variation captured by the first components, important age specific features may be captured and projected.

Acknowledgements

The authors wish to thank the CMI Bureau of the Institute & Faculty of Actuaries for providing financial support.

References

Booth, H., Maindonald, J. and Smith, S. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies,* 56, 325-336.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

Lee, R. D. (2000). The Lee-Carter method of forecasting mortality, with various extensions and applications. *North American Actuarial Journal* (with discussion), 4, no 1, 80-93.

Lee, R. D. and Carter, L. (1992). Modelling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* (with discussion), 87, 659-671.

Pesaran, M. H. and Pesaran, B. (1997). *Working with Microfit 4.0. Interactive Econometric Analysis*. Oxford University Press.

Renshaw, A. and Haberman, S. (2003a). Lee-Carter mortality forecasting, a parallel GLM approach, England & Wales mortality projections. *Applied Statistics,* 52, 119-137.

Renshaw, A. and Haberman, S. (2003b). Lee-Carter mortality forecasting with age specific enhancement. *Proceedings, IME, Lisbon, July 2002. Insurance: Mathematics and Economics* (to appear)

Renshaw, A. and Haberman, S. (2003c). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, 32, 379-401.

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

Unit root tests for $\kappa_t^{(1)}$. Male experience

	Test Statistic	maximum log-likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn criterion
DF	-2.489	-32.054	-35.054	-37.696	-36.029
ADF(1)	-1.440	-28.205	-32.205	-35.727	-33.504
ADF(2)	-0.718	-23.590	**-28.590**	**-32.590**	**-30.213**
ADF(3)	-0.575	-23.293	-29.293	-34.577	-31.241
ADF(4)	-0.645	-23.220	-30.220	-36.385	-32.493
ADF(5)	-0.673	-23.190	-31.190	-38.234	-33.788

95% critical value for ADF statistic = -3.516 bold = maximum

Unit root tests for $\kappa_t^{(2)}$. Male experience

	Test Statistic	maximum log-likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn criterion
DF	-0.528	152.20	149.20	146.56	148.23
ADF(1)	-0.216	154.56	**150.56**	**147.03**	**149.26**
ADF(2)	-0.151	154.77	149.77	145.37	148.15
ADF(3)	-0.101	155.25	149.25	143.97	147.30
ADF(4)	-0.154	155.33	148.33	142.17	146.06
ADF(5)	-0.152	156.67	148.67	141.63	146.08

95% critical value for ADF statistic = -3.516 bold = maximum

Unit root tests for $\kappa_t^{(1)}$. Female experience

	Test statistic	maximum log-likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn criterion
DF	-5.314*	-38.258	**-41.258**	**-43.900**	**-42.232**
ADF(1)	-3.363	-37.580	-41.580	-45.102	-42.879
ADF(2)	-2.281	-36.275	-41.275	-45.678	-42.899
ADF(3)	-1.914	-36.065	-42.065	-47.349	-44.014
ADF(4)	-2.033	-35.735	-42.735	-48.899	-45.008
ADF(5)	-2.106	-35.507	-43.507	-50.552	-46.105

95% critical value for ADF statistic = -3.516 (*significant); bold = maximum

Unit root tests for $\kappa_t^{(2)}$. Female experience

	Test statistic	maximum log-likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn criterion
DF	-1.923	128.41	125.42	122.78	124.44
ADF(1)	-1.702	133.85	**129.85**	**126.33**	**128.55**
ADF(2)	-1.638	134.70	129.70	125.30	128.07
ADF(3)	-1.626	134.73	128.73	123.44	126.78
ADF(4)	-1.522	134.77	127.77	121.61	125.50
ADF(5)	-1.042	136.07	128.07	121.03	125.47

95% critical value for ADF statistic = -3.516 bold = maximum

Table 1

Case I : $\mathbf{z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$. Male experience

order (p)	log likelihood	Akaike information criterion	Schwarz Bayesian criterion	Likelihood ratio test [p-value]	adjusted LR test [p-value]
3	139.07	123.07	108.62	----	----
2	135.69	123.69	112.85	6.76 [.15]	5.56 [.24]
1	132.77	**124.77**	117.54	12.60 [.13]	10.36 [.24]
0	125.24	121.24	**117.62**	27.67 [.01]*	22.75 [.03]*

bold = maximum * statistically significant

univariate models	parameter estimates [p-value]
$\Delta\kappa_t^{(1)} = a_{10} + \phi_{11}\Delta\kappa_{t-1}^{(1)}$	$\hat{a}_{10} = -.476[.00]; \hat{\phi}_{11} = -.459[.00]$
$\Delta\kappa_t^{(2)} = a_{20} + a_{21}t + \phi_{22}\Delta\kappa_{t-1}^{(2)}$	$\hat{a}_{20} = -.013[.00]; \hat{a}_{21} = .00046[.00]; \hat{\phi}_{22} = -.277[.05]$

Case I : $\mathbf{z}_t = (\Delta\kappa_t^{(1)}, \Delta\kappa_t^{(2)})'$. Female experience

order (p)	log likelihood	Akaike information criterion	Schwarz Bayesian criterion	likelihood ratio test [p-value]	adjusted LR test [p-value]
3	100.34	84.34	69.89	----	----
2	97.21	**85.21**	74.37	6.27 [.18]	5.15 [.27]
1	92.73	84.73	**77.51**	15.22 [.06]	12.51 [.13]
0	78.87	74.87	71.26	42.95 [.00]*	35.31 [.00]*

bold = maximum * statistically significant

univariate models	parameter estimates [p-value]
$\Delta\kappa_t^{(1)} = a_{10} + \phi_{11}\Delta\kappa_{t-1}^{(1)}$	$\hat{a}_{10} = -.551[.00]; \hat{\phi}_{11} = -.493[.00]$
$\Delta\kappa_t^{(2)} = a_{20} + a_{21}t + \phi_{22}\Delta\kappa_{t-1}^{(2)}$	$\hat{a}_{20} = -.018[.00]; \hat{a}_{21} = .00065[.00]; \hat{\phi}_{22} = -.320[.02]$

Table 2

Case II : $\mathbf{y}_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$. Male experience

order (p)	log likelihood	Akaike information criterion	Schwarz Bayesian Criterion	likelihood ratio test [p-value]	adjusted LR test [p-value]
3	147.13	131.13	116.50	----	----
2	144.68	132.68	121.71	4.89 [.30]	4.04 [.40]
1	140.85	**132.85**	**125.53**	12.56 [.13]	10.37 [.24]
0	74.40	70.40	66.74	145.45 [.00]*	120.16 [.00]*

bold = maximum * statistically significant

bivariate model	parameter estimates [p-value]
$\begin{pmatrix}\kappa_t^{(1)}\\ \kappa_t^{(2)}\end{pmatrix} = \begin{pmatrix}a_{10}\\ a_{20}\end{pmatrix} + \begin{pmatrix}a_{11}\\ a_{21}\end{pmatrix} t + \begin{pmatrix}\phi_{11} & \phi_{12}\\ 0 & \phi_{22}\end{pmatrix}\begin{pmatrix}\kappa_{t-1}^{(1)}\\ \kappa_{t-1}^{(2)}\end{pmatrix}$	$\hat{a}_{10} = 4.641[.00]; \hat{a}_{20} = -.010[.00]$ $\hat{a}_{11} = -.190[.00]; \hat{a}_{21} = .00036[.00]$ $\hat{\phi}_{11} = .364[.01]; \hat{\phi}_{12} = -10.407[.00]$ $\hat{\phi}_{22} = .947[.00]$

Case II : $\mathbf{z}_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$. Female experience

order (p)	log likelihood	Akaike information criterion	Schwarz Bayesian criterion	likelihood ratio test [p-value]	adjusted LR test [p-value]
3	108.03	92.30	77.67	----	----
2	104.81	**92.81**	81.83	6.97 [.14]	5.77 [.22]
1	97.16	89.16	**81.84**	22.28 [.00]*	18.41 [.02]*
0	58.97	54.97	51.31	98.65 [.00]*	81.50 [.00]*

bold = maximum * statistically significant

univariate model	Parameter estimates [p-value]
$\kappa_t^{(1)} = a_{10} + a_{11} t$	$\hat{a}_{10} = 8.228[.00]; \hat{a}_{11} = -.329[.00]$
$\kappa_t^{(2)} = a_{20} + a_{21} t + \phi_{22}\kappa_{t-1}^{(2)}$	$\hat{a}_{20} = -.0158[.00]; \hat{a}_{21} = .00054[.00]; \hat{\phi}_{22} = .810[.00]$

Table 3

Case III: $\mathbf{a}_0 \neq \mathbf{0}, \mathbf{a}_1 = \mathbf{0}$ (unrestricted intercepts, no trends). Male experience

eigenvalues: 0.43152 0.1281E-3

null hypothesis	alternative hypothesis	eigenvalue statistic	95% confidence value	trace statistic	95% confidence value
$r=0$	$r=1$	27.11*	14.88	27.12*	17.86
$r \leq 1$	$r=2$	.0062	8.07	.0062	8.07

* statistically significant

rank r	maximised log likelihood	Akaike information criterion	Schwarz Bayesian Criterion	Hannan-Quinn Criterion
0	112.500	110.50	108.63	109.79
1	126.055	**121.06**	**116.38**	**119.29**
2	126.058	120.06	114.44	117.94

bold = maximum

$\hat{\mathbf{a}}_0$ with [p-value]	$\hat{\alpha}$ with [p-value]	cointegration relationship $\beta'\mathbf{y}_t$
$\begin{pmatrix} -.298[.00] \\ -.000565[.58] \end{pmatrix}$	$\begin{pmatrix} .169[.81] \\ -.0388[.00] \end{pmatrix}$	$\left(.0316\kappa_t^{(1)} + 1.881\kappa_t^{(2)}\right)$

Case III: $\mathbf{a}_0 \neq \mathbf{0}, \mathbf{a}_1 = \mathbf{0}$ (unrestricted intercepts, no trends). Female experience

eigenvalues: 0.46031 0.013263

null hypothesis	alternative hypothesis	eigenvalue statistic	95% confidence value	trace statistic	95% confidence value
$r=0$	$r=1$	29.60*	14.88	30.25*	17.86
$r \leq 1$	$r=2$	.64	8.07	.64	8.07

* statistically significant

rank r	maximised log likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn Criterion
0	65.16	63.16	61.29	62.46
1	79.97	**74.97**	**70.29**	**73.20**
2	80.29	74.29	68.67	72.17

bold = maximum

$\hat{\mathbf{a}}_0$ with [p-value]	$\hat{\alpha}$ with [p-value]	co-integration relationships $\beta'\mathbf{z}_t$
$\begin{pmatrix} -.346[.01] \\ -.00168[.38] \end{pmatrix}$	$\begin{pmatrix} -.775[.40] \\ -.0752[.00] \end{pmatrix}$	$\left(.0224\kappa_t^{(1)} + 2.310\kappa_t^{(2)}\right)$

Table 4

Case IV: $\mathbf{a}_0 \neq \mathbf{0}, \mathbf{a}_1 = \Pi\gamma$ (unrestricted intercepts, restricted trends). Males

eigenvalues: 0.44858 0.30103 0.00

null hypothesis	alternative hypothesis	eigenvalue statistic	95% confidence value	trace statistic	95% confidence value
$r=0$	$r=1$	28.57	19.22	45.76	25.77
$r\leq 1$	$r=2$	17.19*	12.39	17.19*	12.39

* statistically significant

rank r	maximised log likelihood	Akaike information criterion	Schwarz Bayesian criterion	Hannan-Quinn criterion
0	112.500	110.50	108.63	109.79
1	126.786	120.79	115.17	118.67
2	135.382	**127.38**	**119.90**	**124.55**

bold = maximum

Case IV: $\mathbf{a}_0 \neq \mathbf{0}, \mathbf{a}_1 = \Pi\gamma$ (unrestricted intercepts, restricted trends). Females

eigenvalues: 0.51666 0.34941 0.00

null hypothesis	alternative hypothesis	eigenvalue statistic	95% confidence value	trace statistic	95% confidence value
$r=0$	$r=1$	34.90	19.22	55.53	25.77
$r\leq 1$	$r=2$	20.63*	12.39	20.63*	12.39

* statistically significant

rank r	maximised log likelihood	Akaike information criterion	Schwarz Bayesian Criterion	Hannan-Quinn Criterion
0	65.164	63.16	61.29	62.46
1	82.613	76.61	71.00	74.49
2	92.930	**84.93**	**77.45**	**82.10**

bold = maximum

Table 5

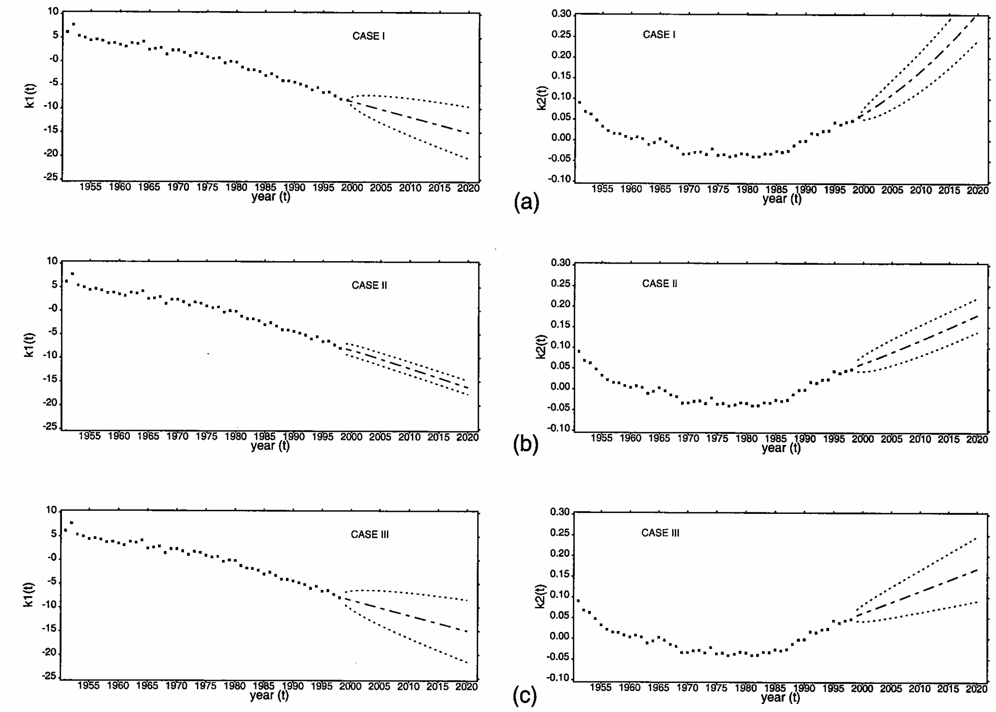

Fig. 1. Male mortality $\kappa_t^{(i)}$ actual and forecast values, with 95% limits (i = 1 LHS, i = 2 RHS): (a) Case I, (b) Case II and (c) Case III.

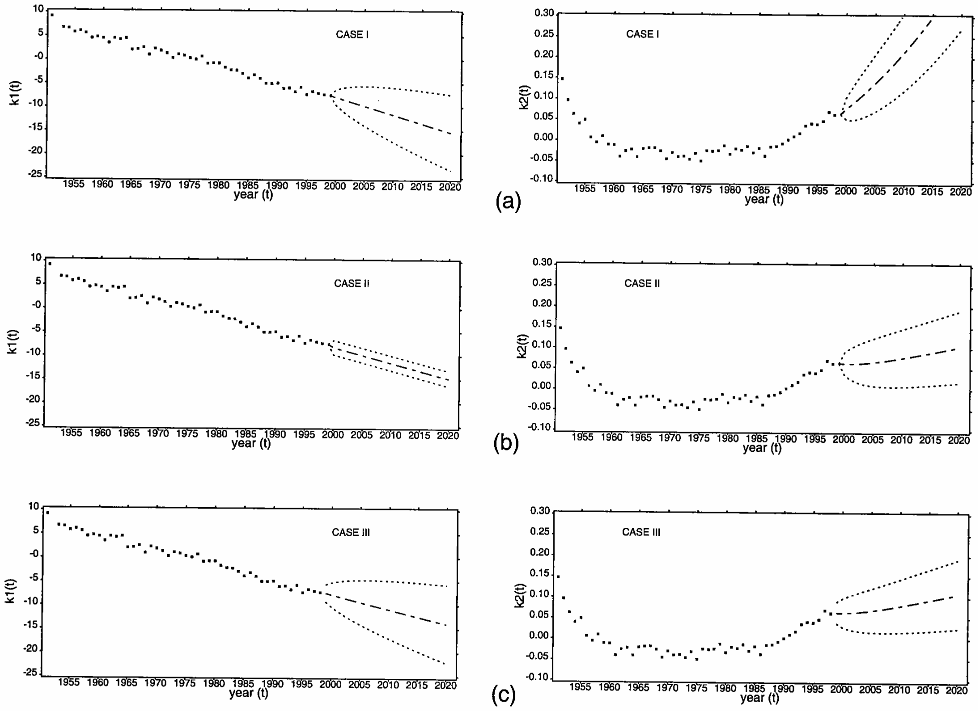

Fig. 2. Female mortality $\kappa_t^{(i)}$ actual and forecast values, with 95% limits ($i = 1$ LHS, $i = 2$ RHS): (a) Case I, (b) Case II and (c) Case III.

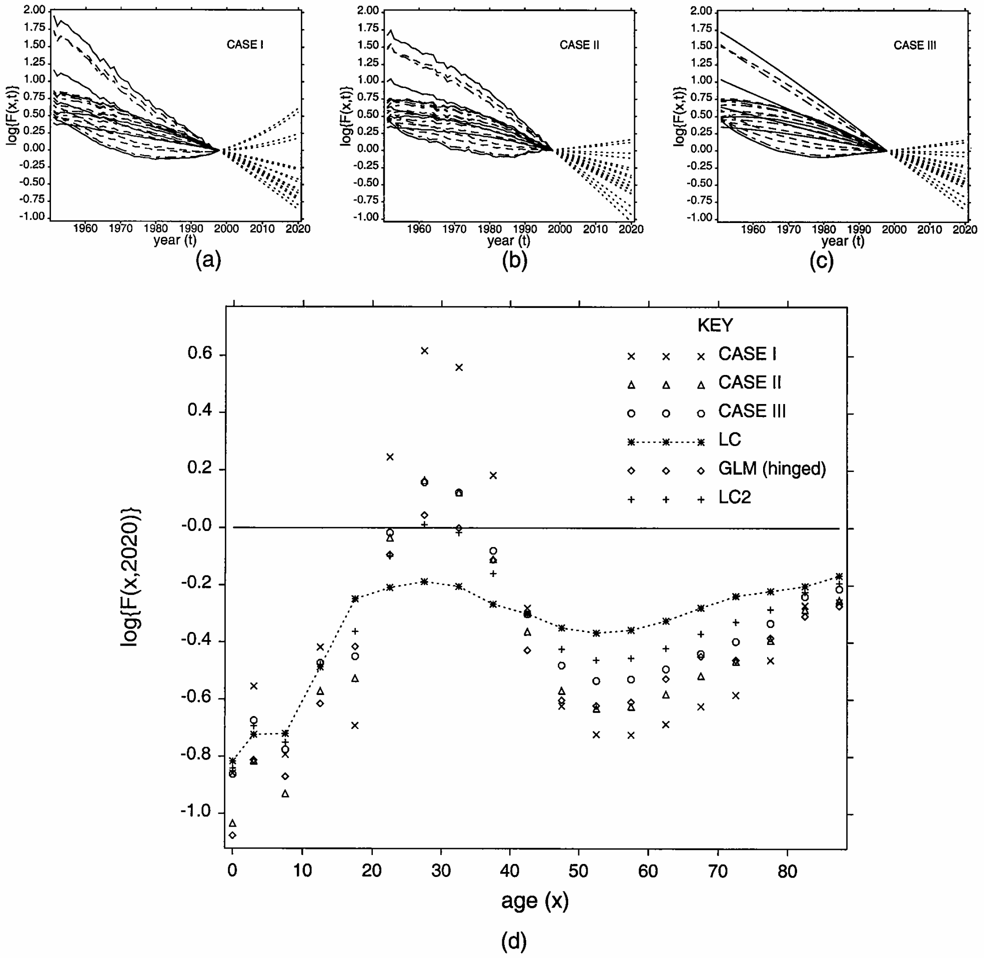

Fig. 3. Male mortality age specific fitted and projected log mortality reduction factors: (a) Case I, (b) Case II, (c) Case III, and (d) year 2020 log mortality reduction factor projections by age for five different modelling approaches, comprising Cases I to III, the basic LC approach and a GLM based approach.

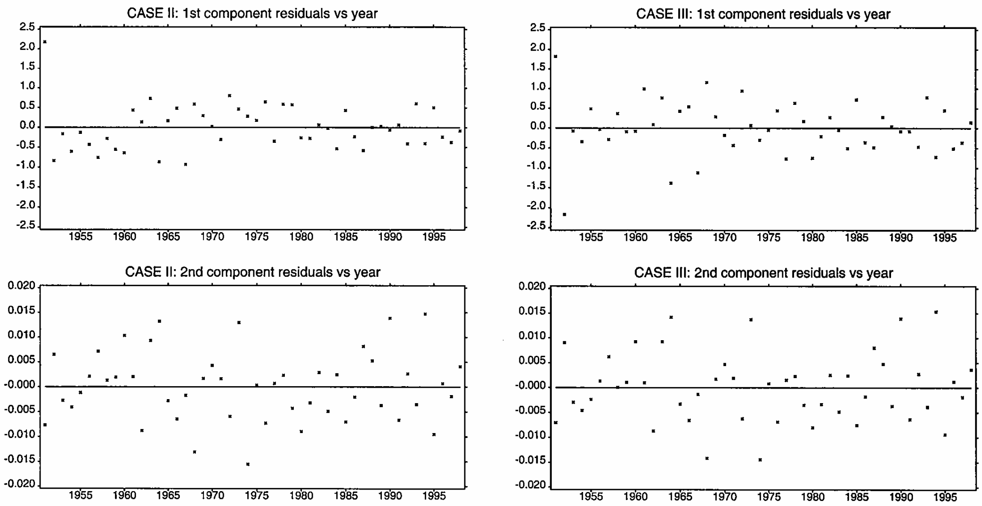

Fig. 4. Male mortality: Case II & Case III residual plots vs. year, by SVD component.

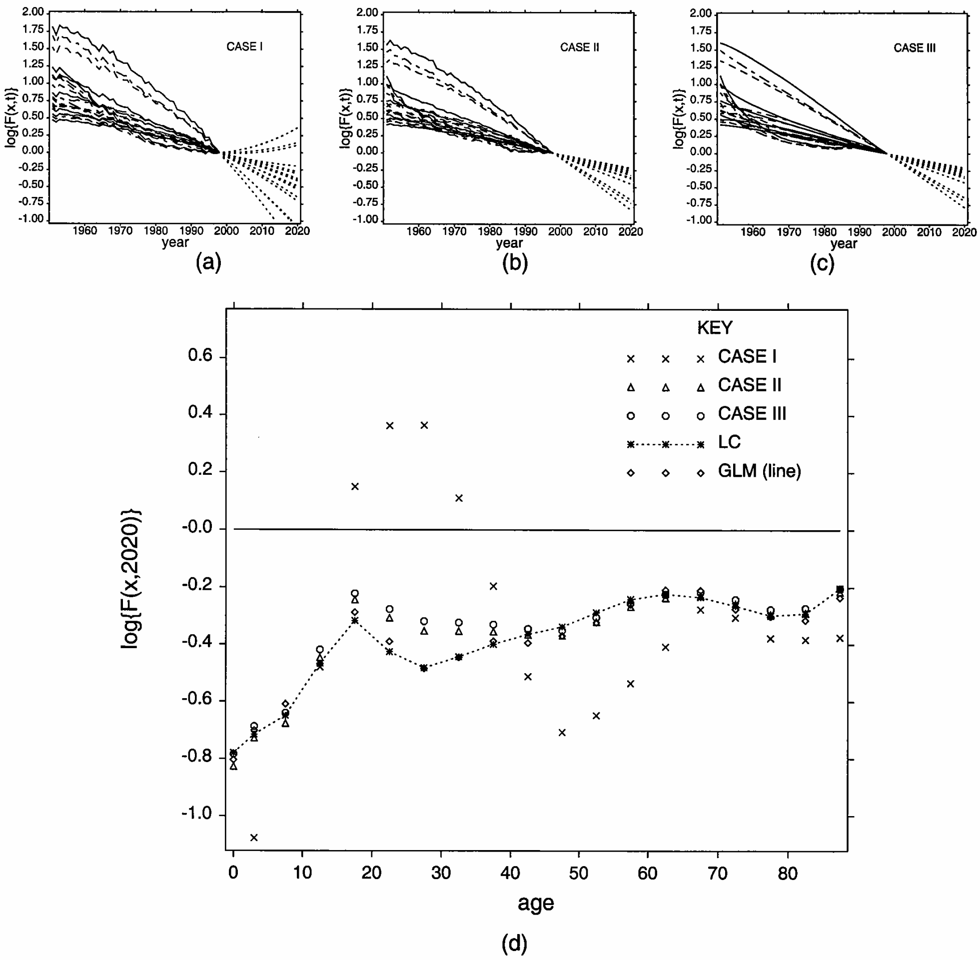

Fig. 5. Female mortality age specific fitted and projected log mortality reduction factors: (a) Case I, (b) Case II, (c) Case III, and (d) year 2020 log mortality reduction factor projections by age for five different modelling approaches, comprising Cases I to III, the basic LC approach and a GLM based approach.

APPENDIX

This appendix contains an outline of mainly first order autoregressive bivariate time series models. Attention is given to model fitting and the subsequent construction of bivariate forecasts and their mean square forecast error (*MSFE*). Whereas the orthogonal decomposition of the *MSFE*, as described below, is readily available in MICROFIT, its value, which is used to construct forecast limits, is not available. Thus all the details described below have been programmed independently of MICROFIT and the results cross-referenced, whenever possible.

Bivariate Vector Autoregressive (VAR) Time Series Modelling

The Model.

Consider

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t \,;\ t = 1,2,\ldots,n \qquad \text{(A.1)}$$

where

$\mathbf{y}_t, \boldsymbol{\varepsilon}_t$ are $m \times 1$ vectors of joint responses and independent errors

$\mathbf{a}_0, \mathbf{a}_1$ are $m \times 1$ vectors of intercept and trend parameters

$\mathbf{\Phi}_i$ are $m \times m$ matrices of autoregressive parameters

Assumptions:

$\boldsymbol{\varepsilon}_t \sim \text{iidN}(\mathbf{0}, \mathbf{\Omega})$ where $\mathbf{\Omega}$ is symmetric positive definite, and all roots of $|\,\mathbf{I}_m - \sum_{i=1}^{p} \mathbf{\Phi}_i \lambda^i \,| = 0$ fall outside the unit circle.

Consider first order bivariate cases in detail, for which $m = 2, p = 1,\ \mathbf{\Phi}_1 = \mathbf{\Phi}$.

Notes:

1. Higher order processes ($p > 1$), while of potential interest, induce over-parameterisation in our context.
2. Adopt the convention that $\{\mathbf{y}_t : t = 0,1,2,\ldots,n\}$ is observed. Thus
   n = number of observations – 1.
3. The implementation of such models is described in Chapters 7 and 19 of Pesaran, W.H. and B. Pesaran (1997). Working with Micofit 4.0 *Oxford University Press.*
4. The case $\mathbf{a}_1 = \mathbf{0}$ is described in detail in Chapter 11 of Hamilton, J.D. (1994). Time Series Analysis, *Princeton University Press, Princeton, New Jersey.*
5. The specific choice of Granger causality model described below is consistent with the ordering (by SVD) of the time series components.
6. When the deterministic trend term is pre-set to zero $(\mathbf{a}_1 = \mathbf{0})$, the appropriate rows and columns are omitted from the detailed matrices below.
7. Repeated reference to the multivariate Gaussian model is a common theme throughout this appendix.

Model Fitting.

Refer to the multivariate Gaussian model

$$\mathbf{Y} = \mathbf{GA} + \boldsymbol{\varepsilon}$$

for which

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ & \\ y_{1n} & y_{2n} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & 1 & y_{10} & y_{20} \\ 1 & 2 & y_{11} & y_{21} \\ & & & \\ 1 & n & y_{1,n-1} & y_{2,n-1} \end{pmatrix}, \ \mathbf{A}' = \begin{pmatrix} \mathrm{a}_{01} & \mathrm{a}_{11} & \phi_{11} & \phi_{12} \\ \mathrm{a}_{02} & \mathrm{a}_{12} & \phi_{21} & \phi_{22} \end{pmatrix}$$

and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$, so that

$$\mathbf{G'G} = \begin{pmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} y_{1,t-1} & \sum_{t=1}^{n} y_{2,t-1} \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} ty_{1,t-1} & \sum_{t=1}^{n} ty_{2,t-1} \\ \sum_{t=1}^{n} y_{1,t-1} & \sum_{t=1}^{n} ty_{1,t-1} & \sum_{t=1}^{n} y_{1,t-1}^2 & \sum_{t=1}^{n} y_{1,t-1}y_{2,t-1} \\ \sum_{t=1}^{n} y_{2,t-1} & \sum_{t=1}^{n} ty_{2,t-1} & \sum_{t=1}^{n} y_{1,t-1}y_{2,t-1} & \sum_{t=1}^{n} y_{2,t-1}^2 \end{pmatrix}, \mathbf{G'Y} = \begin{pmatrix} \sum_{t=1}^{n} y_1 & \sum_{t=1}^{n} y_2 \\ \sum_{t=1}^{n} ty_{1t} & \sum_{t=1}^{n} ty_{2t} \\ \sum_{t=1}^{n} y_{1,t-1}y_{1t} & \sum_{t=1}^{n} y_{1,t-1}y_{2t} \\ \sum_{t=1}^{n} y_{2,t-1}y_{1t} & \sum_{t=1}^{n} y_{2,t-1}y_{1t} \end{pmatrix}$$

Then the ordinary least squares estimates (OLS) are

$$\hat{\mathbf{A}} = (\mathbf{G'G})^{-1}(\mathbf{G'Y}), \ \hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} / \nu$$

with $\nu = n - k$ degrees of freedom, where $\hat{\mathbf{A}}$ is $k \times 2$.

The residual vector is

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}}.$$

Forecasting.

Successive substitution gives

$$\mathbf{y}_{t+s} = \{\mathbf{I} + \boldsymbol{\Psi}_1 + \ldots + \boldsymbol{\Psi}_{s-1}\}\mathbf{a}_0 + \{(t+s)\mathbf{I} + (t+s-1)\boldsymbol{\Psi}_1 + \ldots(t+1)\boldsymbol{\Psi}_{s-1}\}\mathbf{a}_1 + \boldsymbol{\Psi}_s\mathbf{y}_t$$
$$+ \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t+s-1} + \ldots + \boldsymbol{\Psi}_{s-1}\boldsymbol{\varepsilon}_{t+1}$$

where $\boldsymbol{\Psi}_i = \boldsymbol{\Phi}^i$.

Thus the *s* step ahead forecast, from time *t* (= *n*), can be expressed either as

$$\hat{\mathbf{y}}_{t+s|t} = \{\mathbf{I} + \boldsymbol{\Psi}_1 + \ldots + \boldsymbol{\Psi}_{s-1}\}\mathbf{a}_0 + \{(t+s)\mathbf{I} + (t+s-1)\boldsymbol{\Psi}_1 + \ldots(t+1)\boldsymbol{\Psi}_{s-1}\}\mathbf{a}_1 + \boldsymbol{\Psi}_s\mathbf{y}_t$$

or as

$$\hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\mu} - \left\{\sum_{i=0}^{s-1} i\boldsymbol{\Phi}^i\right\} + \boldsymbol{\Phi}^s(\mathbf{y}_t - \boldsymbol{\mu}) \text{ where } \boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1}\{\mathbf{a}_0 + (t+s)\mathbf{a}_1\},$$

on replacing $\boldsymbol{\varepsilon}_{t+s} : s > 0$, by their expected values **0**.

The mean square forecast error (*MSFE*) and its orthogonal decomposition by component are obtained in an identical manner to co-integration modelling, described below.

Bivariate Granger Causality Modelling.

Consider the first order bivariate VAR model under the null hypothesis $H_0 : \phi_{21} = 0$ so that

$$y_{1t} = a_{10} + a_{12}t + \varphi_{11}y_{1,t-1} + \varphi_{12}y_{2,t-1} + \varepsilon_{1t}$$
$$y_{2t} = a_{20} + a_{22}t \qquad\qquad + \varphi_{22}y_{2,t-1} + \varepsilon_{2t}$$

with $\varepsilon \sim N(\mathbf{0}, \Omega)$ and symmetric variance-covariance matrix $\mathbf{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

Model Fitting.

Method I:

Stage 1: Implement the (restricted) regression

$$y_{2t} = a_{20} + a_{21}t + \varphi_{22}y_{2,t-1} + \varepsilon_{2t}$$

for which

$$\mathbf{Y} = \mathbf{GA} + \varepsilon$$

with

$$\mathbf{Y} = \begin{pmatrix} y_{21} \\ y_{22} \\ \\ y_{2n} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & 1 & y_{20} \\ 1 & 2 & y_{21} \\ & & \\ 1 & n & y_{2,n-1} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} a_{20} \\ a_{22} \\ \varphi_{22} \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \\ \varepsilon_{2n} \end{pmatrix}$$

and $\varepsilon_{2t} \sim iidN(0, \sigma_{22})$, where $\hat{\mathbf{A}}$ is $k \times 1$, so that

$$\hat{\mathbf{A}} = (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{G}'\mathbf{Y}),\ \hat{\varepsilon} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}} \text{ and } \hat{\sigma}_{22} = \hat{\varepsilon}'\hat{\varepsilon}/(n - k).$$

Stage 2: Implement the (intermediate) regression

$$y_{1t} = a_0 + a_1 t + d_0 y_{2t} + d_1 y_{1,t-1} + d_2 y_{2,t-1} + \varepsilon_t$$

for which

$$\mathbf{Y} = \mathbf{GA} + \varepsilon$$

with

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \\ y_{1n} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & 1 & y_{21} & y_{10} & y_{20} \\ 1 & 2 & y_{22} & y_{11} & y_{21} \\ & & & & \\ 1 & n & y_{2n} & y_{1,n-1} & y_{2,n-1} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ d_0 \\ d_1 \\ d_2 \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_n \end{pmatrix}$$

and $\varepsilon_t \sim iidN(0, h)$, so that

$$\hat{\mathbf{A}} = (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{G}'\mathbf{Y}),\ \hat{\varepsilon} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}} \text{ and } \hat{h} = \hat{\varepsilon}'\hat{\varepsilon}/(n - k + 1).$$

Stage 3: Use the following sequence of transformations to compute the outstanding parameters

$$\sigma_{12} = \sigma_{21} = d_0\sigma_{22}$$

$$a_{10} = a_0 + \frac{\sigma_{12}}{\sigma_{22}}a_{20}, a_{11} = a_1 + \frac{\sigma_{12}}{\sigma_{22}}a_{21}$$

$$\varphi_{11} = d_1, \varphi_{12} = d_2 + \frac{\sigma_{12}}{\sigma_{22}}\varphi_{22};\ \sigma_{11} = h + \frac{\sigma_{12}\sigma_{21}}{\sigma_{22}}.$$

Method II:
Refer to the partitioned multivariate Gaussian model

$$\mathbf{Y} = \mathbf{G}\mathbf{A} + \boldsymbol{\varepsilon}$$

or

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix}$$

with

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \\ y_{1n} \\ y_{21} \\ y_{22} \\ \\ y_{2n} \end{pmatrix}, \ \mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & y_{10} & y_{20} & 0 & 0 & 0 \\ 1 & 2 & y_{11} & y_{21} & 0 & 0 & 0 \\ & & & & & & \\ 1 & n & y_{1,n-1} & y_{2,n-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & y_{20} \\ 0 & 0 & 0 & 0 & 1 & 2 & y_{21} \\ & & & & & & \\ 0 & 0 & 0 & 0 & 1 & n & y_{2,n-1} \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \mathrm{a}_{01} \\ \mathrm{a}_{11} \\ \phi_{11} \\ \phi_{12} \\ \mathrm{a}_{02} \\ \mathrm{a}_{12} \\ \phi_{22} \end{pmatrix}$$

and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$. Also

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n, \ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

so that

$$\mathbf{G}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{G} = \begin{pmatrix} \sigma_{11}^{-}\mathbf{G}_{11} & \sigma_{12}^{-}\mathbf{G}_{12} \\ \sigma_{21}^{-}\mathbf{G}_{21} & \sigma_{22}^{-}\mathbf{G}_{22} \end{pmatrix}, \ \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma_{11}^{-} & \sigma_{12}^{-} \\ \sigma_{21}^{-} & \sigma_{22}^{-} \end{pmatrix}$$

where

$$\mathbf{G}_{11} = \begin{pmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} y_{1,t-1} & \sum_{t=1}^{n} y_{2,t-1} \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t y_{1,t-1} & \sum_{t=1}^{n} t y_{2,t-1} \\ \sum_{t=1}^{n} y_{1,t-1} & \sum_{t=1}^{n} t y_{1,t-1} & \sum_{t=1}^{n} y_{1,t-1}^2 & \sum_{t=1}^{n} y_{1,t-1} y_{2,t-1} \\ \sum_{t=1}^{n} y_{2,t-1} & \sum_{t=1}^{n} t y_{2,t-1} & \sum_{t=1}^{n} y_{1,t-1} y_{2,t-1} & \sum_{t=1}^{n} y_{2,t-1}^2 \end{pmatrix}, \mathbf{G}_{12} = \begin{pmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} y_{2,t-1} \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t y_{2,t-1} \\ \sum_{t=1}^{n} y_{1,t-1} & \sum_{t=1}^{n} t y_{1,t-1} & \sum_{t=1}^{n} y_{1,t-1} y_{2,t-1} \\ \sum_{t=1}^{n} y_{2,t-1} & \sum_{t=1}^{n} t y_{2,t-1} & \sum_{t=1}^{n} y_{2,t-1}^2 \end{pmatrix}$$

$$\mathbf{G}_{21} = \mathbf{G}'_{12}, \mathbf{G}_{22} = \begin{pmatrix} n & \sum_{t=1}^{n} t & \sum_{t=1}^{n} y_{2,t-1} \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 & \sum_{t=1}^{n} t y_{2,t-1} \\ \sum_{t=1}^{n} y_{2,t-1} & \sum_{t=1}^{n} t y_{2,t-1} & \sum_{t=1}^{n} y_{2,t-1}^2 \end{pmatrix}$$

and

$$\mathbf{G}'(\Sigma^{-1}\otimes\mathbf{I}_n)\mathbf{Y}=\begin{pmatrix}\sigma_{11}^{-}\sum_{t=1}^{n}y_{1t}+\sigma_{12}^{-}\sum_{t=1}^{n}y_{2t}\\ \sigma_{11}^{-}\sum_{t=1}^{n}ty_{1t}+\sigma_{12}^{-}\sum_{t=1}^{n}ty_{2t}\\ \sigma_{11}^{-}\sum_{t=1}^{n}y_{1,t-1}y_{1t}+\sigma_{12}^{-}\sum_{t=1}^{n}y_{1,t-1}y_{2t}\\ \sigma_{11}^{-}\sum_{t=1}^{n}y_{2,t-1}y_{1t}+\sigma_{12}^{-}\sum_{t=1}^{n}y_{2,t-1}y_{2t}\\ \sigma_{21}^{-}\sum_{t=1}^{n}y_{1t}+\sigma_{22}^{-}\sum_{t=1}^{n}y_{2t}\\ \sigma_{21}^{-}\sum_{t=1}^{n}ty_{1t}+\sigma_{22}^{-}\sum_{t=1}^{n}ty_{2t}\\ \sigma_{11}^{-}\sum_{t=1}^{n}y_{2,t-1}y_{1t}+\sigma_{12}^{-}\sum_{t=1}^{n}y_{2,t-1}y_{2t}\end{pmatrix}.$$

Thus the parameter estimates satisfy

$$\hat{\mathbf{A}}=\mathbf{G}'(\Sigma^{-1}\otimes\mathbf{I}_n)\mathbf{G}\mathbf{G}'(\Sigma^{-1}\otimes\mathbf{I}_n)\mathbf{Y} \tag{A.2}$$

Set up the following iterative fitting process:

Step 1: Determine starting values for $\hat{\mathbf{A}}=\begin{pmatrix}\hat{\mathbf{A}}_1\\ \hat{\mathbf{A}}_2\end{pmatrix}$, using

$$\hat{\mathbf{A}}_i=(\mathbf{G}_i\mathbf{G}_i)^{-1}(\mathbf{G}_i'\mathbf{Y}_i), i=1,2$$

where

$$\mathbf{G}_i'\mathbf{G}_i=\mathbf{G}_{ii}, i=1,2; \mathbf{G}_1'\mathbf{Y}_1=\begin{pmatrix}\sum_{t=0}^{n}y_{1t}\\ \sum_{t=0}^{n}ty_{1t}\\ \sum_{t=0}^{n}y_{1,t-1}y_{1t}\\ \sum_{t=0}^{n}y_{2,t-1}y_{1t}\end{pmatrix}, \mathbf{G}_2'\mathbf{Y}_2=\begin{pmatrix}\sum_{t=1}^{n}y_{2t}\\ \sum_{t=1}^{n}ty_{2t}\\ \sum_{t=1}^{n}y_{2,t-1}y_{2t}\end{pmatrix}$$

Step 2: Compute estimates for Σ, using

$$\hat{\sigma}_{ij}=\frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{\sqrt{(n-k_i)(n-k_j)}}, \text{ where } \hat{\varepsilon}_i=\mathbf{Y}_i-\mathbf{G}_i\hat{\mathbf{A}}_i, i=1,2.$$

Step 3: Use (A.2) to update the estimate for $\hat{\mathbf{A}}$.

Step 4: Iterate between Step 2 and Step 3. Convergence is established by monitoring the difference between consecutive estimates of $\hat{\mathbf{A}}$.

Forecasting.

As above, with $\Phi=\begin{pmatrix}\varphi_{11} & \varphi_{12}\\ 0 & \varphi_{22}\end{pmatrix}$.

Univariate Dickey-Fuller (DF) and augmented Dickey-Fuller (ADF) unit root tests.

Consider $n+p$ observations $(y_{-p+1}, y_{-p+2}, \ldots, y_n)$. The DF regression model

$$y_t = \mathrm{a}_0 + \mathrm{a}_1 t + \rho y_{t-1} + \varepsilon_t;\ t = 1,2,\ldots,n$$

and the ADF(p) regression models

$$y_t = \varsigma_1 \Delta y_{t-1} + \varsigma_2 \Delta y_{t-2} + \ldots + \varsigma_{p-1} \Delta y_{t-p+1} + \mathrm{a}_0 + \mathrm{a}_1 t + \rho y_{t-1} + \varepsilon_t;\ t = 1,2,\ldots,n$$

are central to the tests.

Thus writing $u_t = \Delta y_t$, it follows in general, that

$$\mathbf{Y} = \mathbf{GA} + \boldsymbol{\varepsilon}$$

with

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \\ y_n \end{pmatrix}, \mathbf{G} = \begin{pmatrix} u_0 & & u_{-p+1} & 1 & 1 & y_0 \\ u_1 & & u_{-p+2} & 1 & 2 & y_1 \\ \\ u_{n-1} & & u_{n-p+1} & 1 & n & y_{n-1} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \varsigma_1 \\ \\ \varsigma_{p-1} \\ \mathrm{a}_0 \\ \mathrm{a}_1 \\ \rho \end{pmatrix}, \ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_n \end{pmatrix}$$

and $\varepsilon_t \sim iidN(0, \sigma^2)$, so that

$$\hat{\mathbf{A}} = (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{G}'\mathbf{Y}),\ \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}} \text{ and } \hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n-p-3).$$

Then the test statistic for the *OLS t* test for the unit root null hypothesis $\mathrm{H}_0 : \rho = 1$ is

$$\frac{\hat{\rho} - 1}{\sqrt{\hat{\sigma}^2 \mathbf{e}'(\mathbf{GG}')^{-1}\mathbf{e}}} = \frac{\hat{\rho} - 1}{s.e.(\hat{\rho})}$$

where $\mathbf{e}$ is a $(p + 3)$ vector with unity in the last position and zeros elsewhere.

Notes.

(i) The tests may be applied without the linear trend by setting $\mathrm{a}_1 = 0$ and making and consequential adjustments.

(ii) The vector **e** is designed to capture the standard error of $\hat{\rho}$.

(iii) Chapter 17 of Hamilton, J.D. (1994) discusses the tests in great detail.

(iv) In MICROFIT, results are reported by pre-selecting a value p_{max} for p and computing the test statistic for the DF and ADF(p), p = 1, 2, …, p_{max}-1 models, with conditioning on the first p_{max} observations. Akaike information, Schwarz Bayesian and Hannan-Quinn criteria are quoted as a guide to model selection.

Multivariate Vector Autoregressive (VAR) Times Series Modelling

Consider the general VAR(p) process, based on m components, given by (A.1).

Model Fitting.
This proceeds by conditioning on the first p components of the multivariate data set

$$\{\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \dots, \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n\}.$$

Thus

$$\mathbf{Y} = \mathbf{GA} + \boldsymbol{\varepsilon}$$

for which

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{21} & & y_{m1} \\ y_{12} & y_{22} & & y_{m2} \\ & & & \\ y_{1n} & y_{2n} & & y_{mn} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & 1 & y_{10} & & y_{m0} & & y_{1,-p+1} & & y_{m,-p+1} \\ 1 & 2 & y_{11} & & y_{m1} & & y_{1,-p+2} & & y_{m,-p+2} \\ & & & & & & & & \\ 1 & n & y_{1,n-1} & & y_{m,n-1} & & y_{1,n-p+1} & & y_{m,n-p+1} \end{pmatrix}$$

$$\mathbf{A}' = \begin{pmatrix} a_{01} & a_{11} & \phi_{11}^{(1)} & & \phi_{1m}^{(1)} & & \phi_{11}^{(p)} & & \phi_{1m}^{(p)} \\ a_{02} & a_{12} & \phi_{21}^{(1)} & & \phi_{2m}^{(1)} & & \phi_{21}^{(p)} & & \phi_{2m}^{(p)} \\ & & & & & & & & \\ a_{0m} & a_{1m} & \phi_{m1}^{(1)} & & \phi_{mm}^{(1)} & & \phi_{m1}^{(p)} & & \phi_{mm}^{(p)} \end{pmatrix}.$$

Thus $\mathbf{Y}$ is $n \times m$, $\mathbf{G}$ is $n \times (2 + pm)$, $\mathbf{A}$ is $(2 + pm) \times m$, and the estimates

$$\hat{\mathbf{A}} = (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{G}'\mathbf{Y}), \hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} / \nu$$

with $\nu = n - 2 - pm$ degrees of freedom, are computed. The 2nd columns of $\mathbf{G}$ and $\mathbf{A}'$ are deleted if the linear trend term is omitted and ν increased by 1 accordingly. The residual vector is

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}}.$$

Forecasting.
The trick is to expand equation (A.1) into the form of a VAR(1) process

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \\ \mathbf{y}_{t-p+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & t\mathbf{I} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \end{pmatrix} + \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p \\ \mathbf{I} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & & \mathbf{0} & \mathbf{0} \\ \\ \mathbf{0} & \mathbf{0} & & \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \mathbf{y}_{t-3} \\ \\ \mathbf{y}_{t-p} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \mathbf{0} \\ \\ \mathbf{0} \end{pmatrix},$$

which is written as

$$\boldsymbol{\xi}_t = \mathbf{h}_t \varsigma + \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t,$$

where

$\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \mathbf{v}_t$ are $mp \times 1$, $\mathbf{h}_t$ is $mp \times 2m$, ς is $2m \times 1$ and $\mathbf{F}$ is $mp \times mp$.
Successive substitution gives

$$\boldsymbol{\xi}_{t+s} = (\mathbf{h}_{t+s} + \mathbf{F}\mathbf{h}_{t+s-1} + \mathbf{F}^2\mathbf{h}_{t+s-2} + \ldots + \mathbf{F}^{s-1}\mathbf{h}_{t+1})\varsigma + \mathbf{F}^s\boldsymbol{\xi}_t + \mathbf{v}_{t+s} + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{F}^2\mathbf{v}_{t+s-2} + \ldots + \mathbf{F}^{s-1}\mathbf{v}_{t+1}$$

which is written as

$$\boldsymbol{\xi}_{t+s} = \mathbf{H}_{t+s}\varsigma + \mathbf{F}^s\boldsymbol{\xi}_t + \sum_{j=0}^{s-1}\mathbf{F}^j\mathbf{v}_{t+s-j}.$$

Then, extracting the 1st m rows, gives

$$\mathbf{y}_{t+s} = \mathbf{H}_{1,t+s}\varsigma + \mathbf{F}_1^{(s)}\boldsymbol{\xi}_t + \sum_{j=0}^{s-1}\mathbf{\Psi}_j\boldsymbol{\varepsilon}_{t+s-j}$$

where

$\mathbf{H}_{1,t+s}, \mathbf{F}_1^{(s)}$ are the 1st m rows of the respective matrices $\mathbf{H}_{t+s}, \mathbf{F}^s$

and

$\mathbf{\Psi}_j$ is the blocked matrix comprising the 1st m rows and 1st m columns of $\mathbf{F}^j$.
Thus the s step ahead forecast, from time t (=n) is given by

$$\hat{\mathbf{y}}_{t+s|t} = \mathbf{H}_{1,t+s}\varsigma + \mathbf{F}_1^{(s)}\boldsymbol{\xi}_t$$

on replacing $\boldsymbol{\varepsilon}_{t+s} : s > 0$, by their expected values $\mathbf{0}$. For computational purposes

$$\mathbf{H}_{t+s} = \mathbf{h}_{t+s} + \mathbf{F}\mathbf{H}_{t+s-1}, s = 1, 2, 3, \ldots;\ \mathbf{H}_t = \mathbf{0}.$$

The *MSFE* and its orthogonal decomposition by component (generalised to m and not 2 components) are obtained in identical manner to co-integration modelling, described below.

Bivariate Cointegration Time Series Modelling

The Model.

We consider

$$\Delta \mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t - \Pi \mathbf{y}_{t-1} + \varepsilon_t,\ t = 1,2,\ldots,n$$

where

$\mathbf{y}_t$	is a $m \times 1$ vector of jointly determined (endogenous) $I(1)$ variables
$\mathbf{a}_0$	is a $m \times 1$ vector of intercept parameters
$\mathbf{a}_1$	is a $m \times 1$ vector of trend parameters
Π	is a $m \times m$ long-run multiplier matrix

Assumption:

$\varepsilon_t \sim \text{iidN}(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ is symmetric positive definite

The following cases are of potential interest:

Case I: $\mathbf{a}_0 = \mathbf{a}_1 = \mathbf{0}$ (no intercepts and no trends)

Case II: $\mathbf{a}_0 = \Pi\mu$ and $\mathbf{a}_1 = \mathbf{0}$ (restricted intercepts and no trends)

Case III: $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 = \mathbf{0}$ (unrestricted intercepts and no trends)

Case IV: $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 = \Pi\gamma$ (unrestricted intercepts and restricted trends)

Case V: $\mathbf{a}_0 \neq \mathbf{0}$ and $\mathbf{a}_1 \neq \mathbf{0}$ (unrestricted intercepts and unrestricted trends)

Thus

Case II: $\Delta \mathbf{y}_t = -\Pi(\mathbf{y}_{t-1} - \mu) + \varepsilon_t,\ t = 1,2,\ldots,n$

Case III: $\Delta \mathbf{y}_t = \mathbf{a}_0 - \Pi \mathbf{y}_{t-1} + \varepsilon_t,\ t = 1,2,\ldots,n$

Case IV: $\Delta \mathbf{y}_t = \mathbf{a}_0 - \Pi(\mathbf{y}_{t-1} - \gamma t) + \varepsilon_t,\ t = 1,2,\ldots,n$

where

γ, μ are $m \times 1$ vectors of parameters.

Consider the bivariate cases in detail, for which $m = 2$.

Notes:

1. Case I acts as an outer marker for the set of model structures and is of little practical value.
2. Case V has been found to generate unrealistic forecasts in our context.
3. Higher order processes, incorporating the terms $\sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i}$ ($p > 1$) on the RHS are of potential interest, but are found to induce over-parameterisation in our context.
4. When $t = 1$, the RHS of the modelling equations requires the value of $\mathbf{y}_0$. We adopt the convention that $\{\mathbf{y}_t : t = 0,1,2,\ldots,n\}$ is observed. Thus n = number of observations – 1.
5. The model is a special case of the more general vector error correction model (VECM), described in Chapters 7 and 19 of Pesaran, W.H. and B. Pesaran (1997). Working with Microfit 4.0 *Oxford University Press.*
6. Cases II & III are discussed in considerable detail in Chapters 11, 19 & 20 of Hamilton, J. D (1994), Time Series Analysis, *Princeton University Press.*
7. Full rank (r = 2) versions of Cases III & IV are re-parameterised versions of the bivariate VAR(1) model discussed above, with and without $\mathbf{a}_1$ pre-set to zero, respectively.

Cointegration relationships.

These require the decomposition

$$\Pi = \alpha\beta'$$

where

α, β are both $2 \times r$ and $r = \text{rank}(\Pi)$.

For Case II $\quad \Pi(\mathbf{y}_{t-1} - \mu) = \alpha\beta'(\mathbf{y}_{t-1} - \mu)$

so that $\beta'(\mathbf{y}_{t-1} - \mu)$ defines the co-integration relationships

$$g_{it} = \beta_{i1} y_{1,t-1} + \beta_{i2} y_{2,t-1} + \beta_{i3},\ i = 1,\ldots,r.$$

For Case III $\quad \Pi\mathbf{y}_{t-1} = \alpha\beta'\mathbf{y}_{t-1}$

while $\beta'\mathbf{y}_{t-1}$ defines the co-integration relationships

$$g_{it} = \beta_{i1} y_{1,t-1} + \beta_{i2} y_{2,t-1},\ i = 1,\ldots,r.$$

For Case IV $\quad \Pi(\mathbf{y}_{t-1} - \gamma t) = \alpha\beta'(\mathbf{y}_{t-1} - \gamma t)$

and $\beta'(\mathbf{y}_{t-1} - \gamma t)$ defines the co-integration relationships

$$g_{it} = \beta_{i1} y_{1,t-1} + \beta_{i2} y_{2,t-1} + \beta_{i3} t,\ i = 1,\ldots,r.$$

Model Fitting.

1. Refer to the multivariate Gaussian model

$$\mathbf{Y} = \mathbf{GA} + \varepsilon$$

where $\varepsilon \sim \text{N}(\mathbf{0}, \Omega)$, so that ordinary least squares (OLS) gives

$$\hat{\mathbf{A}} = (\mathbf{G}'\mathbf{G})^{-1}(\mathbf{G}'\mathbf{Y}),\ \hat{\varepsilon} = \mathbf{Y} - \mathbf{G}\hat{\mathbf{A}},\ \hat{\Omega} = \hat{\varepsilon}'\hat{\varepsilon} / \nu$$

on ν degrees of freedom.

2. Estimating the cointegration relationships:

Step 1: (Case II) Compute $\hat{\varepsilon}_{0t}$, the tth component of the residual vector, from the OLS regression of $\Delta\mathbf{y}_t$ *on* $\mathbf{0}$. Thus trivially

$$\hat{\varepsilon}_{0t} = \begin{pmatrix}\Delta y_{1t} & \Delta y_{2t}\end{pmatrix}.$$

Step1: (Case III and Case IV) Compute $\hat{\varepsilon}_{0t}$, the tth component of the residual vector, from the OLS regression of $\Delta\mathbf{y}_t$ *on* $\mathbf{1}$. Thus

$$\Delta\mathbf{y}_t = \mathbf{a}_0 + \varepsilon_t \text{ for } t = 1,2,\ldots,n$$

gives

$$\mathbf{Y} = \mathbf{GA} + \varepsilon$$

where

$$\mathbf{Y} = \begin{pmatrix}\Delta y_{11} & \Delta y_{21} \\ \Delta y_{12} & \Delta y_{22} \\ & \\ \Delta y_{1n} & \Delta y_{2n}\end{pmatrix},\ \mathbf{G} = \begin{pmatrix}1 \\ 1 \\ \\ 1\end{pmatrix},\ \mathbf{A} = \begin{pmatrix}\mathrm{a}_{01} & \mathrm{a}_{02}\end{pmatrix}.$$

Then

$$\hat{\mathbf{A}} = \frac{1}{\mathrm{n}}\begin{pmatrix}y_{1n} - y_{10} & y_{2n} - y_{20}\end{pmatrix} = \begin{pmatrix}\Delta\bar{y}_{1.} & \Delta\bar{y}_{2.}\end{pmatrix}$$

and

$$\hat{\varepsilon}_{0t} = \begin{pmatrix}\Delta y_{1t} - \Delta\bar{y}_{1.} & \Delta y_{2t} - \Delta\bar{y}_{2.}\end{pmatrix}.$$

Step 2: (Case II) Compute $\hat{\boldsymbol{\varepsilon}}_{1t}$, the tth component of the residual vector, from the OLS regression of $\begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{1} \end{pmatrix}$ on $\mathbf{0}$. Thus trivially

$$\hat{\boldsymbol{\varepsilon}}_{1t} = \begin{pmatrix} y_{1,t\text{-}1} & y_{2,t\text{-}1} & 1 \end{pmatrix}.$$

Step 2: (Case III) Compute $\hat{\boldsymbol{\varepsilon}}_{1t}$, the tth component of the residual vector, from the OLS regression of $\mathbf{y}_{t-1}$ *on* $\mathbf{1}$. Thus

$$\mathbf{y}_{t-1} = \mathbf{a}_0 + \boldsymbol{\varepsilon}_{t-1} \text{ for } t = 1,2,\ldots,n$$

gives

$$\mathbf{Y} = \mathbf{GA} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_{10} & y_{20} \\ y_{11} & y_{21} \\ & \\ y_{1,n-1} & y_{2,n-1} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 \\ 1 \\ \\ 1 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \mathrm{a}_{01} & \mathrm{a}_{02} \end{pmatrix}.$$

Then

$$\hat{\mathbf{A}} = \frac{1}{\mathrm{n}}\left(\sum_{t=1}^{n} y_{1,t-1} \quad \sum_{t=1}^{n} y_{2,t-1} \right) = \begin{pmatrix} \bar{y}_{1.} & \bar{y}_{2.} \end{pmatrix}$$

and

$$\hat{\boldsymbol{\varepsilon}}_{1t} = \begin{pmatrix} y_{1,t\text{-}1} - \bar{y}_{1.} & y_{2,t\text{-}1} - \bar{y}_{2.} \end{pmatrix}.$$

Step2: (Case IV) Compute $\hat{\boldsymbol{\varepsilon}}_{1t}$, the tth component of the residual vector, from the OLS regression of $\begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{t} \end{pmatrix}$ on $\mathbf{1}$. Thus

$$\begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{t} \end{pmatrix} = \mathbf{a}_0 + \boldsymbol{\varepsilon}_{t-1} \text{ for } t = 1,2,\ldots,n$$

gives

$$\mathbf{Y} = \mathbf{GA} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_{10} & y_{20} & 1 \\ y_{11} & y_{21} & 2 \\ & & \\ y_{1,n-1} & y_{2,n-1} & n \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 \\ 1 \\ \\ 1 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \mathrm{a}_{01} & \mathrm{a}_{02} & \mathrm{a}_{03} \end{pmatrix}.$$

Then

$$\hat{\mathbf{A}} = \frac{1}{\mathrm{n}}\left(\sum_{t=1}^{n} y_{1,t-1} \quad \sum_{t=1}^{n} y_{2,t-1} \quad \frac{n(n+1)}{2} \right) = \left(\bar{y}_{1.} \quad \bar{y}_{2.} \quad \frac{n+1}{2} \right)$$

and

$$\hat{\boldsymbol{\varepsilon}}_{1t} = \left(y_{1,t\text{-}1} - \bar{y}_{1.} \quad y_{2,t\text{-}1} - \bar{y}_{2.} \quad t - \frac{n+1}{2} \right).$$

Step3: Compute the matrices

$$\mathbf{S}_{ij} = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}'_{it}\hat{\varepsilon}_{jt}, \quad i, j = 0,1$$

and the eigenvalues λ_i and eigenvectors β_i of the equation

$$\left|\lambda\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\right| = 0.$$

The first r eigenvectors, in order of magnitude of the eigenvalues, form the cointegration vector $\beta = (\beta_1, \beta_2, ..., \beta_r)$. Johansen ((1991) Econometrica, 59, 6, pp 1551-1580) normalisation is by setting

$$\beta'\mathbf{S}_{11}\beta = \mathbf{I},$$

for which

$$\beta'_i(\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01})\beta_j = 0, \forall i \neq j.$$

3. Estimating the parameters:

Recall that for

Case II: $\Delta\mathbf{y}_t = -\alpha\beta'(\mathbf{y}_{t-1} - \mu) + \varepsilon_t$, $t = 1,2,\ldots,n$

Case III: $\Delta\mathbf{y}_t = \mathbf{a}_0 - \alpha\beta'\mathbf{y}_{t-1} + \varepsilon_t$, $t = 1,2,\ldots,n$

Case IV: $\Delta\mathbf{y}_t = \mathbf{a}_0 - \alpha\beta'(\mathbf{y}_{t-1} - \gamma t) + \varepsilon_t$, $t = 1,2,\ldots,n$.

For all three cases

$$\mathbf{Y} = \mathbf{GA} + \varepsilon$$

where

$$\mathbf{Y} = \begin{pmatrix} \Delta y_{11} & \Delta y_{21} \\ \Delta y_{12} & \Delta y_{22} \\ & \\ \Delta y_{1n} & \Delta y_{2n} \end{pmatrix}.$$

For Cases III and IV and rank $r = 1$

$$\mathbf{G} = \begin{pmatrix} 1 & -g_{11} \\ 1 & -g_{12} \\ & \\ 1 & -g_{1n} \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} a_{01} & a_{02} \\ \alpha_{11} & \alpha_{12} \end{pmatrix}$$

so that $\mathbf{G}'\mathbf{G} = \begin{pmatrix} n & -\sum_{t=1}^{n} g_{1t} \\ -\sum_{t=1}^{n} g_{1t} & \sum_{t=1}^{n} g_{1t}^2 \end{pmatrix}$ and $\mathbf{G}'\mathbf{Y} = \begin{pmatrix} y_{1n} - y_{10} & y_{2n} - y_{20} \\ -\sum_{t=1}^{n} g_{1t}\Delta y_{1t} & -\sum_{t=1}^{n} g_{1t}\Delta y_{2t} \end{pmatrix}$,

For Cases III and IV and rank $r = 2$

$$\mathbf{G} = \begin{pmatrix} 1 & -g_{11} & -g_{21} \\ 1 & -g_{12} & -g_{22} \\ & & \\ 1 & -g_{1n} & -g_{2n} \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} a_{01} & a_{02} \\ \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$

so that

$$\mathbf{G'G} = \begin{pmatrix} n & -\sum_{t=1}^{n} g_{1t} & -\sum_{t=1}^{n} g_{2t} \\ -\sum_{t=1}^{n} g_{1t} & \sum_{t=1}^{n} g_{1t}^2 & \sum_{t=1}^{n} g_{1t}g_{2t} \\ -\sum_{t=1}^{n} g_{2t} & \sum_{t=1}^{n} g_{1t}g_{2t} & \sum_{t=1}^{n} g_{2t}^2 \end{pmatrix}, \mathbf{G'Y} = \begin{pmatrix} y_{1n} - y_{10} & y_{2n} - y_{20} \\ -\sum_{t=1}^{n} g_{1t}\Delta y_{1t} & -\sum_{t=1}^{n} g_{1t}\Delta y_{2t} \\ -\sum_{t=1}^{n} g_{2t}\Delta y_{1t} & -\sum_{t=1}^{n} g_{2t}\Delta y_{2t} \end{pmatrix}.$$

For Case II delete the 1st columns of $\mathbf{G}$ and $\mathbf{G'G}$ together with the 1st rows of $\mathbf{A}$, $\mathbf{G'G}$ and $\mathbf{G'Y}$.
Then the estimates

$$\hat{\mathbf{A}} = (\mathbf{G'G})^{-1}(\mathbf{G'Y})$$

$$\hat{\Omega} = (\mathbf{Y} - \mathbf{G}\hat{\mathbf{A}})'(\mathbf{Y} - \mathbf{G}\hat{\mathbf{A}})/\nu,$$

are computed according to the relevant cointegration relationships g_{it}.

Tests to Determine the Rank of Π.

Let

$$\mathrm{H}_r : \mathrm{rank}(\Pi) = r, \quad r = 0, 1, 2$$

and let

$$\lambda_1 > \lambda_2 > \ldots > \lambda_r$$

denote the ordered eigenvalues of the cointegration relationships.

1. The likelihood ratio test based on the eigenvalue statistic is

$$\mathrm{LR}(\mathrm{H}_r \mid \mathrm{H}_{r+1}) = -n\log(1 - \hat{\lambda}_{r+1}), \quad r = 0, 1.$$

2. The likelihood ratio test based on the trace statistic is

$$\mathrm{LR}(\mathrm{H}_r \mid \mathrm{H}_2) = -n\sum_{i=r+1}^{2}\log(1 - \hat{\lambda}_i), \quad r = 0, 1.$$

Critical values for these tests are generated by stochastic simulation techniques.

3. Compare the Akaike information criterion

$$\mathrm{AIC} = \ell_r - \varsigma_r,$$

the Schwarz Bayesian criterion

$$\mathrm{SBC} = \ell_r - \frac{\varsigma_r}{2}\log(n)$$

and the Hannan-Quinn criterion

$$\mathrm{HQC} = \ell_r - \varsigma_r \log\{\log(n)\}$$

for r = 0, 1, 2, where

$$\ell_r = -n\left\{1 + \log\left(2\pi\sqrt{\frac{\nu|\hat{\Omega}|}{n}}\right)\right\}$$

is the (constrained) maximum of the appropriate log likelihood, and for

Case II $\quad \varsigma_r = 5r - r^2,$

Case III $\quad \varsigma_r = 2 + 4r - r^2,$

Case IV $\quad \varsigma_r = 2 + 5r - r^2.$

For r = 0, $\Pi = \mathbf{0}$ and the appropriate log likelihood is that associated with Section 2, Step 1 of the model fitting section above.

Forecasting.

1. Forecasts.

Consider Case IV for which

$$\Delta\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t - \Pi\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,\ t = 1,2,\ldots,n$$

so that

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{a}_1 t + (\mathrm{I} - \Pi)\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t .$$

Successive substitution implies

$$\mathbf{y}_{t+s} = \{\mathbf{I} + \Psi_1 + \ldots + \Psi_{s-1}\}\mathbf{a}_0 + \{(t+s)\mathbf{I} + (t+s-1)\Psi_1 + \ldots + (t+1)\Psi_{s-1}\}\mathbf{a}_1 + \Psi_s\mathbf{y}_t$$
$$+\ \boldsymbol{\varepsilon}_{t+s} + \Psi_1\boldsymbol{\varepsilon}_{t+s-1} + \ldots + \Psi_{s-1}\boldsymbol{\varepsilon}_{t+1}$$

where $\Psi_i = (\mathbf{I} - \Pi)^i$.

Thus the s step ahead forecast, from time t (= n), is

$$\hat{\mathbf{y}}_{t+s|t} = \{\mathbf{I} + \Psi_1 + \ldots + \Psi_{s-1}\}\mathbf{a}_0 + \{(t+s)\mathbf{I} + (t+s-1)\Psi_1 + \ldots + (t+1)\Psi_{s-1}\}\mathbf{a}_1 + \Psi_s\mathbf{y}_t$$

on replacing $\boldsymbol{\varepsilon}_{t+s}$ $s > 0$, by its expected value $\mathbf{0}$.

For Case II: $\mathbf{a}_0 = \alpha\,[\beta_{i3}]$, $\mathbf{a}_1 = \mathbf{0}$, $\Pi = \alpha\begin{bmatrix}\beta_{i1} & \beta_{i2}\end{bmatrix}$

For Case III: $\mathbf{a}_1 = \mathbf{0}$, $\Pi = \alpha\beta'$

For Case IV: $\mathbf{a}_1 = \alpha\,[\beta_{i3}]$, $\Pi = \alpha\begin{bmatrix}\beta_{i1} & \beta_{i2}\end{bmatrix}$

where $[\beta_{i3}]$ is $r \times 1$, $\begin{bmatrix}\beta_{i1} & \beta_{i2}\end{bmatrix}$ is $r \times 2$ and α is formed by extracting the relevant elements from **A,** (part 3 of the model fitting section above), prior to transposing.

2. Forecast errors.

Since

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\varepsilon}_{t+s} + \Psi_1\boldsymbol{\varepsilon}_{t+s-1} + \ldots + \Psi_{s-1}\boldsymbol{\varepsilon}_{t+1}$$

the s step ahead mean square forecast error is

$$MSFE(\hat{\mathbf{y}}_{t+s|t}) = \mathrm{E}\{(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'\} = \Omega + \Psi_1\Omega\Psi_1' + \ldots + \Psi_{s-1}\Omega\Psi_{s-1}' .$$

The orthogonal decomposition of the *MSFE* by component, involving the transformation

$$\mathbf{u}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$$

where

$$\Omega = \mathbf{ADA}' = \mathbf{PP}'$$

and

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}, \mathbf{P} = \mathbf{AD}^{1/2}$$

is given by

$$MSFE(\hat{\mathbf{y}}_{t+s|t}) = \sum_{j=1}^{2}\{\mathbf{p}_j\mathbf{p}_j' + \Psi_1\mathbf{p}_j\mathbf{p}_j'\Psi_1' + \ldots + \Psi_{s-1}\mathbf{p}_j\mathbf{p}_j'\Psi_{s-1}'\}$$

where $\mathbf{p}_j$ is the jth column of the Cholesky factor **P**. This decomposition is conditional on the ordering of the component variables in the bivariate time series, in parallel with the ordering of the time series components in their determination by SVD.

Footnotes:

1. For notational convenience, the origin $t = 0$, is set in the first calendar year for which data are available (pth. year in general). To generate the same intercept parameter estimates as MICROFIT, it is necessary to set this origin in the calendar year immediately prior to the first year for which data are available.
2. To generate the same Johansen normalised parameter estimates produced by MICROFIT, it was found necessary to scale the co-integration vector β, defined under Section 2, Step 3, by a further factor of $1/\sqrt{n}$.

FACULTY OF ACTUARIAL SCIENCE AND STATISTICS

Actuarial Research Papers since 2001

135. Renshaw A. E. and Haberman S. On the Forecasting of Mortality Reduction Factors. February 2001.
ISBN 1 901615 56 1

136. Haberman S., Butt Z. & Rickayzen B. D. Multiple State Models, Simulation and Insurer Insolvency. February 2001. 27 pages.
ISBN 1 901615 57 X

137. Khorasanee M.Z. A Cash-Flow Approach to Pension Funding. September 2001. 34 pages.
ISBN 1 901615 58 8

138. England P.D. Addendum to "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving". November 2001. 17 pages.
ISBN 1 901615 59 6

139. Verrall R.J. A Bayesian Generalised Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. November 2001. 10 pages.
ISBN 1 901615 62 6

140. Renshaw A.E. and Haberman. S. Lee-Carter Mortality Forecasting, a Parallel GLM Approach, England and Wales Mortality Projections. January 2002. 38 pages.
ISBN 1 901615 63 4

141. Ballotta L. and Haberman S. Valuation of Guaranteed Annuity Conversion Options. January 2002. 25 pages.
ISBN 1 901615 64 2

142. Butt Z. and Haberman S. Application of Frailty-Based Mortality Models to Insurance Data. April 2002. 65 pages.
ISBN 1 901615 65 0

143. Gerrard R.J. and Glass C.A. Optimal Premium Pricing in Motor Insurance: A Discrete Approximation. (Will be available 2003).

144. Mayhew, L. The Neighbourhood Health Economy. A systematic approach to the examination of health and social risks at neighbourhood level. December 2002. 43 pages.
ISBN 1 901615 66 9

145. Ballotta L. and Haberman S. The Fair Valuation Problem of Guaranteed Annuity Options: The Stochastic Mortality Environment Case. January 2003. 25 pages.
ISBN 1 901615 67 7

146. Haberman S., Ballotta L. and Wang N. Modelling and Valuation of Guarantees in With-Profit and Unitised With-Profit Life Insurance Contracts. February 2003. 26 pages.
ISBN 1 901615 68 5

147. Ignatov Z.G., Kaishev V.K and Krachunov R.S. Optimal Retention Levels, Given the Joint Survival of Cedent and Reinsurer. March 2003. 36 pages.
ISBN 1 901615 69 3

148. Owadally M.I. Efficient Asset Valuation Methods for Pension Plans. March 2003. 20 pages.

ISBN 1 901615 70 7

149. Owadally M.I. Pension Funding and the Actuarial Assumption Concerning Investment Returns. March 2003. 32 pages.

ISBN 1 901615 71 5

150. Dimitrova D, Ignatov Z. and Kaishev V. Finite time Ruin Probabilities for Continuous Claims Severities. August 2003.

151. Iyer S. Application of Stochastic Methods in the Valuation of Social Security Pension Schemes. August 2003. 40 pages.

ISBN 1 901615 72 3

152. Ballotta L., Haberman S. and Wang N. Guarantees in with-profit and Unitized with profit Life Insurance Contracts; Fair Valuation Problem in Presence of the Default Option[1]. October 2003. 28 pages.

ISBN 1-901615-73-1

153. Renshaw A. and Haberman. S. Lee-Carter Mortality Forecasting Incorporating Bivariate Time Series. December 2003. 33 pages.

ISBN 1-901615-75-8

Faculty of Actuarial Science and Statistics

Actuarial Research Club

The support of the corporate members

CGNU Assurance
Computer Sciences Corporation
English Matthews Brockman
Government Actuary's Department
Swiss Reinsurance
Watson Wyatt Partners

is gratefully acknowledged.

ISBN 1-901615-75-8