An Extension of the Lee-Carter Model for Mortality Projections

By

Udi Makov,
Shaul Bar-Lev, Yaser Awad,
Extensions 1: Bayesian Trend choice
The Lee-Carter model:

\[ y_{x,t} = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \varepsilon_{x,t} \sim F(\sigma^2) \]

\[ k_t = k_{t-1} + \theta + w_t, \quad w_t \sim N(\sigma_w^2) \]

- \( y_{x,t} \): the logarithm of the central death rate for age \( x \) in year \( t \).
- \( \alpha_x \): general shape across age of the mortality schedule.
- \( \beta_x \): indicates the tendency of the logarithm of the force of mortality at age \( x \) when the general level of mortality (\( k_t \)) changes.
- \( k_t \): is an index that describes the variation in the level of mortality at time \( t \).
Estimation Methodology

The methodology is based on the following steps:

1. Fixing appropriate prior distributions for the unknown parameters

2. Using the Kalman Filter algorithm to estimate the predicted values of the time series process, including the variance of the prediction

3. Using the Gibbs sampler to estimate the posterior distributions of the unknown parameters, including the parameters of the unobserved time series index, whose values were estimated by the Kalman Filter algorithm.
The Lee Carter model as a state-space model

Equations below are describing the state process model and the observation model as follow:

\[
\begin{align*}
\begin{cases}
k_t &= k_{t-1} + \theta + \nu_t, & \text{state equation} \\
y_t &= \alpha + \beta k_t + \nu_t, & \text{measurement equation}
\end{cases}
\end{align*}
\]

We assumed the process noise is \( \nu_t \sim N(0, \sigma_w^2) \) and the measurement noise is \( \nu_t \sim N_p(0, \sigma_{\epsilon}^2 I) \)
General Presentation of time trend model

\[
\begin{align*}
k_t &= \rho k_{t-1} + \theta + \eta_t - \rho \eta_{t-1} + w_t - \varphi w_{t-1} \\
\eta_t &= \gamma_1 + \gamma_2 t
\end{align*}
\]
In vectorial presentation

\[(k - X\gamma) = P(k - X\gamma) + \theta + (I - \Psi)w\]

where

\[k = (k_{\text{min}}, k_{\text{min}+1}, \cdots, k_{\text{max}}) \quad \theta = (\theta, \theta, \cdots, \theta)\]

\[w = (w_{\text{min}}, w_{\text{min}+1}, \cdots, w_{\text{max}})\]

\[P = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \rho & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho & 0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & t_{\text{min}} \\ \vdots & \vdots \\ 1 & t_{\text{max}} \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \varphi & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \varphi & 0 \end{bmatrix}\]
where

\[ k \sim N_T(X\gamma + \theta, \sigma_k^2(UQ^{-1})) \]

\[ Q = (I = P)'(I - P) = \begin{bmatrix}
1 + \rho^2 & -\rho & \cdots & \cdots & 0 \\
-\rho & 1 + \rho^2 & -\rho & \cdots & 0 \\
0 & -\rho & 1 + \rho^2 & -\rho & 0 \\
\vdots & \vdots & -\rho & 1 + \rho^2 & -\rho \\
0 & \cdots & 0 & -\rho & 1
\end{bmatrix} \quad U = \begin{bmatrix}
1 + \varphi^2 & -\varphi & \cdots & \cdots & 0 \\
-\varphi & 1 + \varphi^2 & -\varphi & \cdots & 0 \\
0 & -\varphi & 1 + \varphi^2 & -\varphi & 0 \\
\vdots & \vdots & -\varphi & 1 + \varphi^2 & -\varphi \\
0 & \cdots & 0 & -\varphi & 1
\end{bmatrix} \]
Extensions

Model 1: $\rho = 1, \gamma_1 = \gamma_2 = 0, \varphi = 0 \Rightarrow \text{ARIMA}(0,1,0)$

(Lee & Carter, 1992; Pedroza, 2006)

$$k_t = k_{t-1} + \theta + w_t, \quad \text{with } w_t \sim N(0, \sigma_w^2)$$

Then the resulting distribution is:

$$k \sim N_T(\theta, \sigma_w^2 Q_{\rho=1}^{-1})$$
Model 2: \( \theta = 0, \varphi = 0 \Rightarrow \text{ARIMA}(1,0,0) \) with linear trend (Czado et al., 2006)

\[
\begin{aligned}
\begin{cases}
 k_t = \rho k_{t-1} + \eta_t - \rho \eta_{t-1} + w_t \\
 \eta_t = \gamma_1 + \gamma_2 t
\end{cases}
\end{aligned}
\]

Then the resulting distribution is:

\[ k \sim N_T (X\gamma, \sigma_k^2 Q^{-1}) \]
Model 3: $\rho = 1, \gamma_1 = \gamma_2 = 0 \Rightarrow \text{ARIMA}(0,1,1)$

Simple exponential smoothing with growth

$$k_t = k_{t-1} + \theta + w_t - \varphi w_{t-1}$$

Then the resulting distribution is:

$$k \sim N_T(\theta, \sigma_w^2 UQ_{\rho=1}^{-1})$$
Model 4: $\gamma_1 = \gamma_2 = 0, \varphi = 0 \Rightarrow \text{ARIMA}(1,0,0)$

As case 2 without linear trend

$$k_t = \rho k_{t-1} + \theta + w_t$$

Then the resulting distribution is:

$$k \sim N_T(\theta, \sigma_w^2 Q^{-1})$$
Model 5: $\gamma_1 = \gamma_2 = 0 \Rightarrow \text{ARIMA}(1,0,1)$

\[ k_t = \rho k_{t-1} + \theta + w_t - \phi w_{t-1} \]

Then the resulting distribution is:

\[ k \sim N_T(\theta, \sigma_w^2 Q^{-1} U) \]
Model Choice

Model 1: $\rho = 1, \gamma_1 = \gamma_2 = 0, \phi = 0 \Rightarrow ARIMA(0,1,0)$

(Lee & Carter, 1992; Pedroza, 2006)

Model 2: $\theta = 0, \phi = 0 \Rightarrow ARIMA(1,0,0)$ with linear trend, (Czado et al., 2005)

Model 3: $\rho = 1, \gamma_1 = \gamma_2 = 0 \Rightarrow ARIMA(0,1,1)$

Simple exponential smoothing with growth

Model 4: $\gamma_1 = \gamma_2 = 0, \phi = 0 \Rightarrow ARIMA(1,0,0)$

As case 2 without linear trend

Model 5: $\gamma_1 = \gamma_2 = 0 \Rightarrow ARIMA(1,0,1)$
Some data consideration
We analyze the data for **U.S** and **Ireland** male. The data obtained from the Human Mortality Database.

The data consists of annual age-specific death rates for years 1959-1999.

The age group for **U.S** are 0, 1-4, 5-9, ..., 105-109, 110+.

The age group for **Ireland** are 0, 1-4, 5-9, ..., 105-109.

Therefore, for **U.S** (P=24, T=6), for **Ireland** (P=23, T=6), that is 1958-1989 grouped by 5 years.
The distribution of the mortality rates by age groups and time
U.S Male

Smrlo'10 an International Symposium on February 8-11, 2010
The distribution of the mortality rates by age groups and time

Ireland Male

Smrlo'10 an International Symposium on February 8-11, 2010
Prior distribution

\[ \alpha \sim N_p(a_{\alpha}, \sigma_{\alpha}^2 I), \quad \sigma_{\alpha}^{-2} \sim Gamma(a_x, b_x) \]

\[ \beta \sim N_p(0, \sigma_{\beta}^2 I), \quad \sigma_{\beta}^{-2} \sim Gamma(a_{\beta}, b_{\beta}) \]

\[ \rho \sim N(0, \sigma_{\rho}^{-2}) \text{ truncated to the interval (-1,1)} \]

\[ \sigma_{w}^{-2} \sim Gamma(a_w, b_w), \quad \sigma_{\varepsilon}^{-2} \sim Gamma(a_{\varepsilon}, b_{\varepsilon}) \]

\[ \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \gamma_{01} \\ \gamma_{02} \end{pmatrix}, \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \right) \]
Bayesian Model Choice

For models comparison we estimated the Deviance information criterion – DIC for normal likelihood. smaller values for “better” models. (Spiegelhalter, et al. 2002).

$$DIC = D(\Theta) + 2pD$$

where:

$$D(\Theta) = -2 \log(f(y/\Theta))$$

$$p = \text{the effective number of parameters}$$

$$\Theta = \text{the sample average of the simulated values of } \Theta$$

We used the data of 1958-1989 to estimate the models parameters and the DIC index.
Bayesian Model Choice

The DIC indices for all five models using U.S male data are as follows:

DIC[ model 1:ARIMA(0,1,0)] = -149.25
DIC[ model 2:ARIMA(1,0,0)] = -154.26
DIC[ model 3:ARIMA(0,1,1)] = -146.55
DIC[ model 4:ARIMA(1,0,0)] = -108.96
DIC[ model 5:ARIMA(1,0,1)] = -104.96

The DIC indices for all five models using Irish male data are as follows:

DIC[ model 1:ARIMA(0,1,0)] = -130.07
DIC[ model 2: ARIMA(1,0,0)] = -141.77
DIC[ model 3: ARIMA(0,1,1)] = -123.91
DIC[ model 4: ARIMA(1,0,0)] = -88.78
DIC[ model 5: ARIMA(1,0,1)] = -87.35
We used the data of 1958-1989 to estimate the models parameters.

Then we used the data from 1990-1999 to assess the predictive quality of the models and we compared between them using also graphical presentation.
Comparison between model 1 and model 2

U.S Male

AGE group 90-94

<table>
<thead>
<tr>
<th>Forecast Year</th>
<th>Observed</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0.20</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>1991</td>
<td>0.21</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1992</td>
<td>0.22</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1993</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1994</td>
<td>0.24</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1995</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1996</td>
<td>0.26</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1997</td>
<td>0.27</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1998</td>
<td>0.28</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>1999</td>
<td>0.29</td>
<td>0.24</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Comparison between model 1 and model 2

Ireland Male

AGE group 90-94

Forecast Year

observed  model 1  model 2

Death rate
Extensions 2:

Mortality Projections Simultaneous for Multiple populations
Exchangeable sequence of random variables

Formally, an exchangeable sequence of random variables is a finite or infinite sequence $X_1, X_2, X_3, \ldots$ of random variables such that for any finite permutation $\sigma$ of the indices 1, 2, 3, ..., i.e. any permutation $\sigma$ that leaves all but finitely many indices fixed, the joint probability distribution of the permuted sequence

$$X_{\sigma(1)}, X_{\sigma(2)}, X_{\sigma(3)}, \ldots$$

is the same as the joint probability distribution of the original sequence.
Exchangeability between $\theta$'s parameters:

Suppose we have mortality projection for $m$ countries and we believe the time trends parameters belong to a joint distribution

Suppose $\theta = (\theta_1, \theta_2, \cdots, \theta_m)$ are i.i.d variables, with

$$\pi(\theta_j / b, \eta) \rightarrow N(b, \eta), \forall j, \eta > 0$$

The joint distribution of $\theta$'s is:

$$\pi(\_ / b, \eta) = \prod_{j=1}^{m} \pi(\theta_j / b, \eta)$$

Suppose $b \sim N(\mu_b, \sigma_b^2)$ and $\eta \sim IG(a_\eta, c_\eta)$, where $\mu_b, \sigma_b^2, a_\eta, c_\eta$ assumed known (maybe chosen to reflect vague prior information)
REPLACE: Sampling from $\theta$

$$(\theta / K, k_0, \sigma_w^2) \rightarrow N \left( \frac{k_n - k_0}{T}, \frac{\sigma_w^2}{T} \right)$$
By the following posterior distributions, for m=2:

$$(\theta_1 / \Psi^*, b, \eta) \rightarrow \mathcal{N}\left(\frac{k_{T,1} - k_{0,1}}{T} \eta + b \frac{\sigma_{w,1}^2}{T}, \left[\frac{\sigma_{w,1}^2}{T} + \eta\right]; \left[\frac{\sigma_{w,1}}{T} \eta\right] / \left[\frac{\sigma_{w,1}}{T} + \eta\right]\right)$$

$$(\theta_2 / \Psi^*, b, \eta) \rightarrow \mathcal{N}\left(\frac{k_{T,2} - k_{0,2}}{T} \eta + b \frac{\sigma_{w,2}^2}{T}, \left[\frac{\sigma_{w,2}^2}{T} + \eta\right]; \left[\frac{\sigma_{w,2}^2}{T} \eta\right] / \left[\frac{\sigma_{w,2}}{T} + \eta\right]\right)$$

Where \( \{K^1, \ldots, K^m, k_{0,1}, \ldots, k_{0,m}, \sigma_{w,1}^2, \ldots, \sigma_{w,m}^2\} \equiv \Psi^* \)

The rest of the posterior distributions should stay the same as before for each country separately.

Similarly; we assumed Exchangeability on other parameters of the model: Alpha and Beta
Validation

• We used the data from U.S and Ireland male of 1958-1989 to estimate the models parameters

• Then we used the data from 1990-1999 to assess the predictive quality of the models and we compared between them using also graphical presentation
Exchangeability w.r.t $\Theta$

Figure 1: Posterior dis. of $\theta$ : USA and Ireland
Exchangeability w.r.t $\alpha$

Figure 3: Posterior dis. of alpha, age-group 105-109: Usa and Ireland
Exchangeability w.r.t $\beta$

Figure 5: Posterior dis. of beta, age-group 105-109: Usa and Ireland
## Impact on $\alpha$ & $\beta$

<table>
<thead>
<tr>
<th>beta_Usa_Ex</th>
<th>beta_Usa</th>
<th>alpha_Usa_Ex</th>
<th>alpha_Usa</th>
<th>beta_ir_Ex</th>
<th>beta_ir</th>
<th>alpha_Ir_Ex</th>
<th>alpha_Ir</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.151</td>
<td>0.145</td>
<td>-3.995</td>
<td>-3.991</td>
<td>0.314</td>
<td>0.313</td>
<td>-4.034</td>
<td>-4.032</td>
<td>0</td>
</tr>
<tr>
<td>0.103</td>
<td>0.102</td>
<td>-7.124</td>
<td>-7.121</td>
<td>0.199</td>
<td>0.200</td>
<td>-7.140</td>
<td>-7.141</td>
<td>1-4</td>
</tr>
<tr>
<td>0.109</td>
<td>0.108</td>
<td>-7.801</td>
<td>-7.807</td>
<td>0.152</td>
<td>0.152</td>
<td>-7.816</td>
<td>-7.822</td>
<td>5-9</td>
</tr>
<tr>
<td>0.078</td>
<td>0.076</td>
<td>-7.736</td>
<td>-7.730</td>
<td>0.079</td>
<td>0.079</td>
<td>-7.961</td>
<td>-7.969</td>
<td>10-14</td>
</tr>
<tr>
<td>0.045</td>
<td>0.047</td>
<td>-2.911</td>
<td>-2.917</td>
<td>0.007</td>
<td>0.011</td>
<td>-2.787</td>
<td>-2.778</td>
<td>70-74</td>
</tr>
<tr>
<td>0.039</td>
<td>0.043</td>
<td>-2.533</td>
<td>-2.544</td>
<td>0.005</td>
<td>0.007</td>
<td>-2.357</td>
<td>-2.352</td>
<td>75-79</td>
</tr>
<tr>
<td>0.031</td>
<td>0.032</td>
<td>-2.121</td>
<td>-2.128</td>
<td>0.012</td>
<td>0.009</td>
<td>-1.930</td>
<td>-1.922</td>
<td>80-84</td>
</tr>
<tr>
<td>0.028</td>
<td>0.028</td>
<td>-1.721</td>
<td>-1.730</td>
<td>0.022</td>
<td>0.022</td>
<td>-1.507</td>
<td>-1.492</td>
<td>85-89</td>
</tr>
<tr>
<td>0.023</td>
<td>0.021</td>
<td>-1.361</td>
<td>-1.371</td>
<td>0.014</td>
<td>0.010</td>
<td>-1.163</td>
<td>-1.156</td>
<td>90-94</td>
</tr>
<tr>
<td>0.007</td>
<td>0.008</td>
<td>-1.041</td>
<td>-1.063</td>
<td>0.006</td>
<td>0.008</td>
<td>-0.748</td>
<td>-0.735</td>
<td>95-99</td>
</tr>
<tr>
<td>-0.030</td>
<td>-0.029</td>
<td>-0.931</td>
<td>-0.937</td>
<td>-0.062</td>
<td>-0.063</td>
<td>-0.661</td>
<td>-0.644</td>
<td>100-104</td>
</tr>
<tr>
<td>-0.064</td>
<td>-0.063</td>
<td><strong>-1.019</strong></td>
<td><strong>-1.032</strong></td>
<td><strong>-0.173</strong></td>
<td><strong>-0.168</strong></td>
<td><strong>-0.749</strong></td>
<td><strong>-0.740</strong></td>
<td>105-109</td>
</tr>
<tr>
<td>-0.062</td>
<td>-0.060</td>
<td>-1.231</td>
<td>-1.258</td>
<td>-0.114</td>
<td>-0.112</td>
<td>-0.704</td>
<td>-0.683</td>
<td>110+</td>
</tr>
</tbody>
</table>
Impact on $\alpha$ & $\beta$

<table>
<thead>
<tr>
<th>alpha_Usa_Ex</th>
<th>alpha_Usa</th>
<th>beta_ir_Ex</th>
<th>beta_ir</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.019</td>
<td>-1.032</td>
<td>-0.173</td>
<td>-0.168</td>
<td>105-109</td>
</tr>
</tbody>
</table>
## Impact on $k(t)$

<table>
<thead>
<tr>
<th></th>
<th>$k(t)$-Ir-Ex</th>
<th>$k(t)$-Ir</th>
<th>$k(t)$-Usa-Ex</th>
<th>$k(t)$-Usa</th>
<th>year=t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.277</td>
<td>1.259</td>
<td>0.683</td>
<td>0.698</td>
<td>60-64</td>
<td></td>
</tr>
<tr>
<td>0.973</td>
<td>0.979</td>
<td>1.303</td>
<td>1.315</td>
<td>65-69</td>
<td></td>
</tr>
<tr>
<td>0.425</td>
<td>0.432</td>
<td>1.161</td>
<td>1.173</td>
<td>70-74</td>
<td></td>
</tr>
<tr>
<td>-0.074</td>
<td>-0.067</td>
<td>0.116</td>
<td>0.119</td>
<td>75-79</td>
<td></td>
</tr>
<tr>
<td>-0.833</td>
<td>-0.835</td>
<td>-1.134</td>
<td>-1.145</td>
<td>80-84</td>
<td></td>
</tr>
<tr>
<td>-1.768</td>
<td>-1.768</td>
<td>-2.130</td>
<td>-2.160</td>
<td>85-89</td>
<td></td>
</tr>
<tr>
<td>-0.194</td>
<td>-0.189</td>
<td>-0.187</td>
<td>-0.204</td>
<td>theta</td>
<td></td>
</tr>
</tbody>
</table>
Impact on $\Theta$

<table>
<thead>
<tr>
<th>k(t)-Ir-Ex</th>
<th>k(t)-Ir</th>
<th>k(t)-Usa-Ex</th>
<th>k(t)-Usa</th>
<th>theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.194</td>
<td>-0.189</td>
<td>-0.187</td>
<td>-0.204</td>
<td></td>
</tr>
</tbody>
</table>
## Impact on RMSE

<table>
<thead>
<tr>
<th>Mortality at age 105</th>
<th>Improvement</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2755</td>
<td>USA</td>
<td>1.8%</td>
</tr>
<tr>
<td>0.2705</td>
<td>USA Ex</td>
<td></td>
</tr>
<tr>
<td>1.2750</td>
<td>Ireland</td>
<td>18.4%</td>
</tr>
<tr>
<td>1.0401</td>
<td>Ireland Ex</td>
<td></td>
</tr>
</tbody>
</table>
Thank you