Bootstrap for mortality projections on dependent data

V. D’Amato¹, S. Haberman², G. Piscopo³, M. Russolillo¹

1 Department of Economics and Statistics, University of Salerno – Italy - e-mail: vdamato@unisa.it, mrussolillo@unisa.it

2 Faculty of Actuarial Science and Insurance, Cass Business School, City University, London – UK - e-mail: s.haberman@city.ac.uk

3 Department of Mathematics for Decisions, University of Florence, Italy - e-mail: gabriella.piscopo@unifi.it
AGENDA

• The motivation
• The Lee Carter Model
• The Lee Carter Sieve Bootstrap
• Numerical Applications
Motivation

Because of the nonlinear nature of the quantities of interest, such as life expectancy, annuity premiums and so on, an analytic approach to the calculation of prediction intervals is intractable, so that it is necessary to resort to a simulation approach.
The presence of *dependence* across time leads to systematic over-estimation or under-estimation of uncertainty in the mortality estimates, caused by whether negative or positive dependence dominates.
The correlation structure between the residuals has to be tackled. Otherwise prediction intervals for projections underestimate the actual longevity risk.

In other words, it is necessary to assess a significant and further source of risk: a sort of *dependency risk*. 
Lee and Carter (1992) suggested a log-bilinear form for the force of mortality:

\[ m_{xt} = \exp(\alpha_x + \beta_x k_t + u_{xt}) \]

\[ \ln(m_{xt}) = \alpha_x + \beta_x k_t + u_{xt} \]

\[ \sum_t k_t = 0 \quad \sum_x \beta_x = 1 \]
The LC Sieve Bootstrap

In the literature, there is more than one bootstrap method for dependent data as for example block, local, wild, Markov bootstrap, sub-sampling and sieve.

Choi and Hall (2000) show that the sieve bootstrap has substantial advantages over blocking methods, to such an extent that block–based methods are not really competitive. In particular, other authors show that the sieve bootstrap outperforms the block bootstrap (Hardle et al. 2003).
The LC Sieve bootstrap

**Notation:**

- $u_{xt}$: error term
- $\varepsilon_{xt}$: innovation term
- $r_{xt}$: estimated innovation or residual
- $\bar{r}_{xt}$: mean value of the residuals
- $r_{xt} - \bar{r}_{xt}$: centred residuals
- $\hat{F}_r$: ecdf of residuals
- $u^*_{xt}$: bootstrap error
- $\varepsilon^*_{xt}$: IID term from $\hat{F}_r$
The LC Sieve Bootstrap

The Scheme:

The error term is approximated by an $AR(\infty)$ representation:

$$u_{xt} = \sum_{j=1}^{\infty} \phi_j u_{xt-j} + \epsilon_{xt} \quad x = 1, 2, \ldots, m$$
The LC Sieve Bootstrap

**The Steps:**

1. **Fit the model and obtain the OLS estimates**:

\[
\hat{\mathbf{u}}_{xt} = \sum_{j=1}^{\hat{p}(n)} \varphi_j \hat{\mathbf{u}}_{xt-j} + \mathbf{e}_{xt} \quad x = 1, 2, \ldots, m
\]
The Lc Sieve Bootstrap

The Steps:

2. Specify the lag length $\hat{p}(n)$ by BIC, AIC, etc
The LC Sieve Bootstrap

The Steps:

3. **Calculate the autoregressive coefficients** by the Ordinary Least Squares or by using the Yule-Walker method

$$\hat{\phi}_j, \quad j = 1, \ldots, \hat{p}(n)$$
The LC Sieve Bootstrap

The Steps:

4. Calculate the residuals (or estimated innovations) associated with \( \hat{\phi}_j \) according the following formula:

\[
    r_{xt} = \hat{u}_{xt} - \sum_{j=1}^{\hat{p}(n)} \hat{\phi}_j \hat{u}_{xt-j} \\
    x = 1, 2, ..., m \\
    t = \hat{p}(n) + 1, ..., n
\]
The LC Sieve Bootstrap

The Steps:

5. Calculate the centred residuals

\[ \tilde{r}_{xt} = r_{xt} - \bar{r}_{xt} \]
The LC Sieve Bootstrap

The Steps:

6. Define the empirical distribution function of the centred residuals

\[ \hat{F}_{xr} (y) = \frac{1}{n - p} \sum_{t=p+1}^{n} 1\{\tilde{r}_{xt} \leq y\} \]
The LC Sieve Bootstrap

The Steps:

7. Draw $\epsilon^*_{xt}$ IID terms from $\hat{F}_{xr}$ with replacement
The LC Sieve Bootstrap

**The Steps:**

8. **Bootstrap** \( u^*_{xt} \) **are simulated** by recursion according to the bootstrap regression model:

\[
u_{xt}^* = \sum_{j=1}^{\hat{p}(n)} \hat{\phi}_j u_{xt-y}^* + \varepsilon_{xt}^* \quad x = 1, 2, \ldots, m
\]
Summary:

In other words, the values of $\varepsilon_{xt}^*$ are obtained by randomly sampling with replacement from $\hat{F}_{xr}$ and consequently the simulated $u_{xt}^*$ are computed and the $m_{xt}^*$ are mapped. Finally the estimates $\hat{\alpha}_{x}^*, \hat{\beta}_{x}^*, \hat{\kappa}_{t}^*$ are obtained by fitting the log-bilinear structure to the $m_{xt}^*$. 
Application scheme:

- Model Fitting
- Analysis of residuals
- Simulation algorithm
- Comparison of the results
Numerical Application

Dataset:

The population data is composed by the Italian male from 1980 up to 2006 from 0 up to 100 years, collected from Human Mortality Database. The death rates above age 100 have been aggregated in an open age group 100+.
Numerical Application

Italy: male death rates (1980-2006)

Figure 1 - log death rates - Italian male population, age from 0 to 100
Figure 2- ax, bx, kt, basic LC model - Italian male population, age from 0 to 100
## Numerical Application

### ERROR MEASURES BASED ON MORTALITY RATES

#### Averages across ages:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>Mean error</td>
<td>-0.00008</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
<td>0.00028</td>
</tr>
<tr>
<td>MPE</td>
<td>Mean Percentage Error</td>
<td>0.01102</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
<td>0.08361</td>
</tr>
</tbody>
</table>

#### Averages across years:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IE</td>
<td>Integrated Error</td>
<td>-0.00455</td>
</tr>
<tr>
<td>ISE</td>
<td>Integrated Squared Error</td>
<td>0.01930</td>
</tr>
<tr>
<td>IPE</td>
<td>Integrated Percentage Error</td>
<td>1.10273</td>
</tr>
<tr>
<td>IAPE</td>
<td>Integrated Absolute Percentage Error</td>
<td>8.19838</td>
</tr>
</tbody>
</table>
## Numerical Application

### ERROR MEASURES BASED ON LOG MORTALITY RATES

<table>
<thead>
<tr>
<th>Averages across ages:</th>
<th>ME (Mean error)</th>
<th>MSE (Mean Squared Error)</th>
<th>MPE (Mean Percentage Error)</th>
<th>MAPE (Mean Absolute Percentage Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00367</td>
<td>0.01487</td>
<td>-0.01132</td>
<td>0.03596</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Averages across years:</th>
<th>IE (Integrated Error)</th>
<th>ISE (Integrated Squared Error)</th>
<th>IPE (Integrated Percentage Error)</th>
<th>IAPE (Integrated Absolute Percentage Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.36583</td>
<td>1.43059</td>
<td>-1.00498</td>
<td>3.40482</td>
</tr>
</tbody>
</table>
Figure 3 - Fitted ax vs residuals
Numerical Application

Figure 4 - Fitted bx vs residuals
Numerical Application

Figure 5 - Fitted $k_t$ vs residuals
Numerical Application

Figure 6 – Residuals years vs age – basic LC on Italian data
Figure 7 – Paths for ax – Sieve Bootstrap
Figure 8 - Simulated paths for bx – Sieve Bootstrap
Figure 9 - Simulated paths for $k_t$ – Sieve Bootstrap
Figure 10 - Forecasted kt – Sieve Bootstrap
Figure 11 – Paths for ax – Residual standard bootstrap
Numerical Application

Figure 12 - Simulated paths for bx – Residual standard bootstrap
Figure 13 - Forecasted kt – Residual standard bootstrap
Numerical Application

Figure 14 - Forecasted kt – Residual standard bootstrap
## Numerical Application

<table>
<thead>
<tr>
<th>h</th>
<th>Residual Bootstrap</th>
<th></th>
<th>Sieve Bootstrap</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>95%</td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>1</td>
<td>-38.10</td>
<td>-34.81</td>
<td>-41.75</td>
<td>-28.37</td>
</tr>
<tr>
<td>2</td>
<td>-41.45</td>
<td>-37.53</td>
<td>-45.66</td>
<td>-32.52</td>
</tr>
<tr>
<td>3</td>
<td>-43.66</td>
<td>-39.89</td>
<td>-47.82</td>
<td>-34.01</td>
</tr>
<tr>
<td>4</td>
<td>-46.30</td>
<td>-42.34</td>
<td>-49.82</td>
<td>-37.49</td>
</tr>
<tr>
<td>5</td>
<td>-48.82</td>
<td>-44.78</td>
<td>-52.95</td>
<td>-40.37</td>
</tr>
<tr>
<td>6</td>
<td>-51.40</td>
<td>-47.22</td>
<td>-55.07</td>
<td>-41.89</td>
</tr>
<tr>
<td>7</td>
<td>-53.95</td>
<td>-49.67</td>
<td>-57.69</td>
<td>-43.44</td>
</tr>
<tr>
<td>8</td>
<td>-56.52</td>
<td>-52.13</td>
<td>-61.43</td>
<td>-47.03</td>
</tr>
<tr>
<td>9</td>
<td>-59.08</td>
<td>-54.60</td>
<td>-63.52</td>
<td>-50.13</td>
</tr>
<tr>
<td>10</td>
<td>-61.66</td>
<td>-57.07</td>
<td>-66.63</td>
<td>-51.33</td>
</tr>
<tr>
<td>11</td>
<td>-59.55</td>
<td>-64.23</td>
<td>-68.97</td>
<td>-53.24</td>
</tr>
<tr>
<td>12</td>
<td>-62.02</td>
<td>-66.80</td>
<td>-72.17</td>
<td>-55.72</td>
</tr>
<tr>
<td>13</td>
<td>-69.37</td>
<td>-64.47</td>
<td>-74.42</td>
<td>-57.84</td>
</tr>
<tr>
<td>14</td>
<td>-71.94</td>
<td>-66.91</td>
<td>-77.69</td>
<td>-60.33</td>
</tr>
<tr>
<td>15</td>
<td>-74.50</td>
<td>-69.36</td>
<td>-80.27</td>
<td>-62.98</td>
</tr>
</tbody>
</table>

**Table 3-** Non parametric standard bootstrap and Sieve bootstrap 5% and 95% Confidence Intervals for $k_{t+h}$
Our research proposes a particular bootstrap methodology, the LC Sieve Bootstrap, for capturing the dependence in deriving prediction intervals, thus avoiding a systematic over-estimation or under-estimation of the amount of uncertainty in the parameter estimates, respectively if negative or positive dependence dominates.
Concluding Remarks

• The *standard residual bootstrap* procedure does not preserve the correlation structure in the data.

• The *sieve bootstrap*, on the other hand, captures the dependency structure, leading to more reliable uncertainty measurement.
Main References


Main References

Main References


Main References

• Kim, M.C., A.F. Shapiro, G. Hognas, 2006, Evaluating and Extending the Lee–
  Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. 
  Insurance: Mathematics and Economics 26: 1-20

• Kunsch, H.R., 1989, The jackknife and the bootstrap for general stationary 
  observations, Ann. Statist. 17 1217-1241.

• Lahiri, S.N., 2003, Resampling Methods for Dependent Data, Springer

• Lee, R.D., L. R. Carter, 1992, Modelling and Forecasting U.S. Mortality, Journal of 
  the American Statistical Association, 87, 659-671.