SOCIETY OF ACTUARIES

Article from:

ARCH 2013.1 Proceedings

August 1- 4, 2012

Defang Wu, Xiaoming Liu, Yu Hao

Assessing systematic bias in mortality prediction of the Lee-Carter model

Defang Wu, Xiaoming Liu, Hao Yu

Department of Statistics and Actuarial Science
University of Western Ontario, London
Ontario, Canada

dwu87@uwo.ca

August 3, 2012

Motivation

- It has been found in literature that there exists bias when using the Lee-Carter(LC) model for mortality prediction.

Motivation

- It has been found in literature that there exists bias when using the Lee-Carter(LC) model for mortality prediction.
- An overall under-prediction of mortality decline and life expectancy gain is reported.
  For example, Lee and Miller(2001), Bell(1997), Booth(2002), etc.

Motivation

- It has been found in literature that there exists bias when using the Lee-Carter(LC) model for mortality prediction.
- An overall under-prediction of mortality decline and life expectancy gain is reported.
  For example, Lee and Miller(2001), Bell(1997), Booth(2002), etc.
- Possible reasons are put forward to explain such bias,
  such as: error correlation by age and horizon, changing age-shape of mortality, etc.

Motivation

- It has been found in literature that there exists bias when using the Lee-Carter(LC) model for mortality prediction.
- An overall under-prediction of mortality decline and life expectancy gain is reported.
  For example, Lee and Miller(2001), Bell(1997), Booth(2002), etc.
- Possible reasons are put forward to explain such bias,
  such as: error correlation by age and horizon, changing age-shape of mortality, etc.
- Corresponding modifications are developed to LC model.

Motivation

- In paper of Liu and Yu(2011), they found systematic bias of the LC model in forecast of life expectancy based on simulated data.

Motivation

- In paper of Liu and Yu(2011), they found systematic bias of the LC model in forecast of life expectancy based on simulated data.
- The advantage of using simulated data is to separate the potential model mis-specification issue from the model effectiveness test.

Motivation

- In paper of Liu and Yu(2011), they found systematic bias of the LC model in forecast of life expectancy based on simulated data.
- The advantage of using simulated data is to separate the potential model mis-specification issue from the model effectiveness test.
- Systematic bias in forecast of life expectancy were found even we eliminate potential model mis-specification.

Objectives

The main purpose of our paper is to:

1. measure the magnitude of the bias using the bootstrap method
2. provide suggestions on how to correct the bias
3. illustrate the effectiveness of correction through examining the forecast performance

Brief review of LC model

The LC model

$$\log(m_{xt}) = a_x + b_x k_t + \varepsilon_{xt}$$
$$k_t = k_{t-1} + c + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2)$$

- a_x describes the age pattern of mortality averaged over time
- b_x describes the deviations from the averaged pattern when k_t varies
- k_t describes the variation in the level of mortality over time
- ε_{xt} is the error term

Brief review of LC model

Since the study bases on simulated data, we consider two situations when generating new sample paths of k_t for LC data set:

1. simulate a random sample of ξ_t following normal distribution $N(0, \sigma^2)$
2. simulate a random sample of ξ_t following normal distribution with CDF:
$$F_\xi(x) = 0.95N(0, \sigma^2/2.2) + 0.05N(0, (5\sigma)^2/2.2)$$

Case 1 follows original LC model and case 2 represents the situation where irregular large shocks may happen occasionally.

Systematic bias of LC model

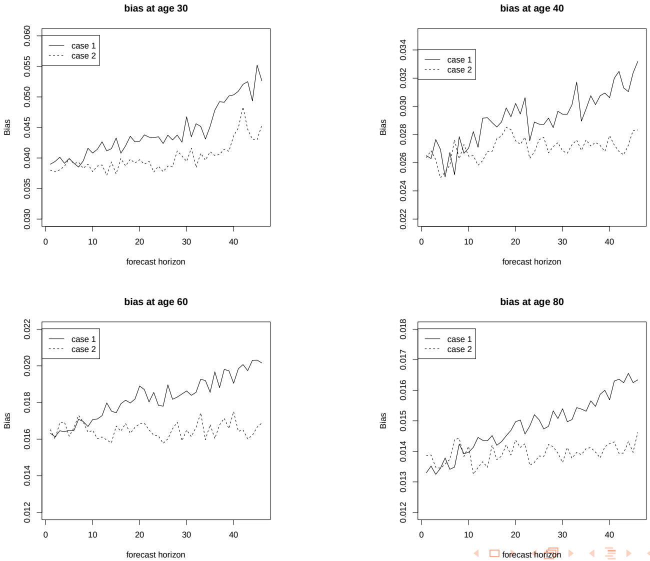
How do we find the systematic bias of LC model by bootstrap method?

Systematic bias of LC model

How do we find the systematic bias of LC model by bootstrap method?

- For each given simulated data set, generate 10,000 sample paths of forecast k_t by bootstrap method.
- Use a_x, b_x and forecast k_t to obtain 10,000 sample of matrix of $\log(m_{xt})$.
- Use median as point forecast of $\log(m_{xt})$ and compare with the "real" mortality.
- Generate 10,000 simulated data set and take average of the difference between predicted value and its corresponding "real" value.

Systematic bias of LC model



Systematic bias of LC model

It is worth noticing that

- Overall values in the figure above for four different ages are positive. Positive bias indicates of under-prediction of decline of $\log(m_{xt})$ by LC model.
- This is consistent with the fact in Liu and Yu(2011) that bias of $e_0(t)$ is always negative and life expectancy gain is under-predicted.
- The systematic bias found in Liu and Yu(2011) is not the result of functional change of forecast variable from $\log(m_{xt})$ to $e_0(t)$ but effectiveness of LC model.

Systematic bias of LC model

Further, we calculate the percentage of bias.

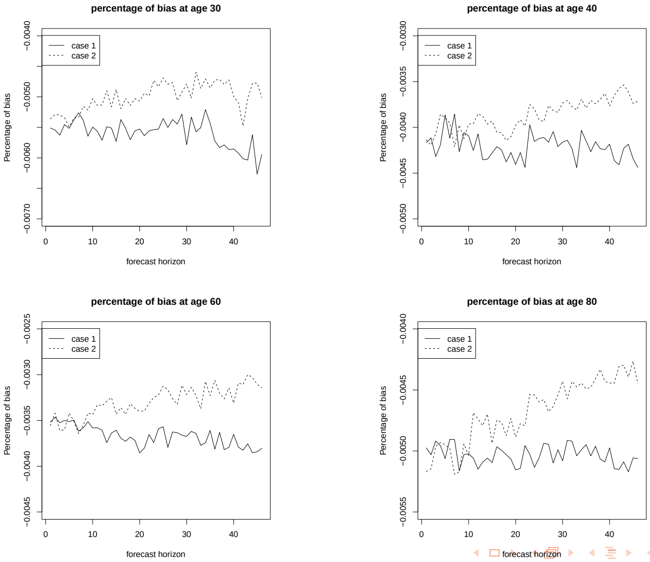
- Percentage of bias = bias/corresponding "real" value of $\log(m_{xt})$.

Systematic bias of LC model

Further, we calculate the percentage of bias.

- Percentage of bias = bias/corresponding "real" value of $\log(m_{xt})$.
- Overall values of percentage of bias are negative, which seems "conflict" to the sign of bias.

Systematic bias of LC model



Bias correction

- Though percentage of bias in forecast of $\log(m_{xt})$ are relatively small, the bias in forecast of $e_0(t)$ could be significant because exponential functions are applied to $\log(m_{xt})$.

Bias correction

- Though percentage of bias in forecast of $\log(m_{xt})$ are relatively small, the bias in forecast of $e_0(t)$ could be significant because exponential functions are applied to $\log(m_{xt})$.
- By applying bias correction to forecast variable, we want forecast performance to be more accurate in forecast of $e_0(t)$.
- The main idea of bias correction is:

Bias correction

- Though percentage of bias in forecast of $\log(m_{xt})$ are relatively small, the bias in forecast of $e_0(t)$ could be significant because exponential functions are applied to $\log(m_{xt})$.
- By applying bias correction to forecast variable, we want forecast performance to be more accurate in forecast of $e_0(t)$.
- The main idea of bias correction is:

Bias Correction

new predicted value = predicted value - estimated bias, where
estimated bias = E[predicted value - value from model predict line]

Bias correction

Remarks:

We fit LC model to mortality data set to obtain parameters of a_x, b_x and k_t, where k_t is modeled by c and σ_ξ^2.

- Predicted value: for forecast purpose, we generate sample path of k_t at time t_{0+i} given the data available up to time t_0 by:

$$k_{t_{0+i}} = k_{t_0} + i \cdot c + \sum_{j=1}^{i} \xi_j.$$

- Value from model predict line: generate sample path of k_t at time t_{0+i} given the data available up to time t_0 by:

$$k_{t_{0+i}} = k_{t_0} + i \cdot c.$$

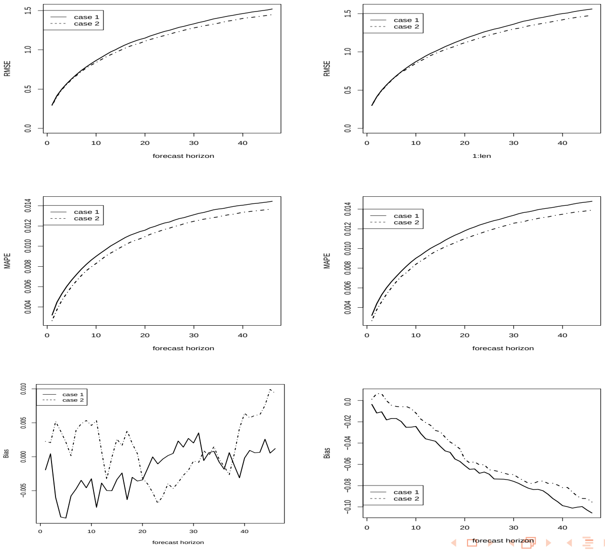
Bias correction

- Apply bias correction to two forecast variables:
  1. applying bias correct to final forecast value of $e_0(t)$
  2. applying bias correct to $\log(m_{xt})$ and then calculating $e_0(t)$ with corrected $\log(m_{xt})$.

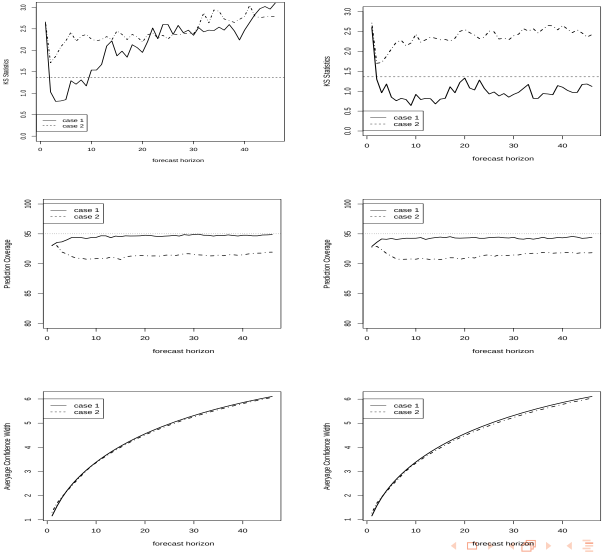
Bias correction

- Apply bias correction to two forecast variables:
  1. applying bias correct to final forecast value of $e_0(t)$
  2. applying bias correct to $\log(m_{xt})$ and then calculating $e_0(t)$ with corrected $\log(m_{xt})$.
- We illustrate the effectiveness of correction under these two methods through examining the forecast performance.
- Forecast performance is evaluated in RMSE, MAPE, bias, Kolmogorov-Smirnov(KS) statistics, coverage and average confidence interval width.

Evaluation


case 1
case 2
RMSE
forecast horizon
1:len
MAPE
Bias
0 10 20 30 40
0.0 0.5 1.0 1.5
0.004 0.006 0.008 0.010 0.012 0.014
−0.005 0.000 0.005 0.010
−0.10 −0.08 −0.06 −0.04 −0.02 0.0

Evaluation



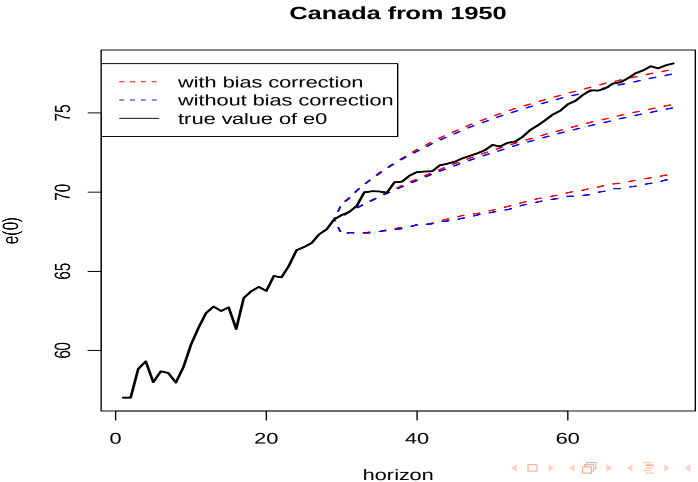
KS Statistics
forecast horizon
case 1
case 2
Prediction Coverage
Averyage Confidence Width

Evaluation

- It is remarkable to notice that:
  - RMSE, MAPE with bias correct applied to $e_0(t)$ is slightly smaller than that with bias correct applied to $\log(m_{xt})$
  - Applying correction to $e_0(t)$ makes bias randomly distribute around zero for both case 1 and case 2.
  - KS statistics of case 1 with bias correct applied to $e_0(t)$ are larger than critical value while that with bias correct applied to $\log(m_{xt})$ is smaller.
- Applying bias correct to $e_0(t)$ is more effective in terms of first three evaluation measurements for forecast performance.

Evaluation

Real mortality data are chosen from Canada(1922-1950) and forecast to year 1995.



Conclusion and Limitation

- It's reported that bias in forecast $\log(m_{xt})$ for LC model are positive in general.
- Positive bias indicates under-prediction of mortality decline but the deviations are less 1% in general.
- In order to obtain more accurate forecast performance of $e_0(t)$, two kinds of bias correction methods are suggested.
- Applying correction to $e_0(t)$ is more effective to obtain better forecast performance based on simulated data.
- Due to dramatically increasing $e_0(t)$ in reality, forecast by LC model is still under-predicted even we try to correct the systematic bias of the model.

Reference

Lee, R. D. and Carter, L. R. 1992. Modeling and Forecasting U.S. Mortality. Journal of the American Stotistical Association 87: 659-675.

Lee, R. D. and Miller, T. 2001. Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. Demography 38: 537-549.

Liu, X. and Yu, H. Assessing and extending the lee-carter model for long-term mortality prediction. Orlando, Fla., January 2011. Living to 100 Symposium.

Thank you for your listening!