Evaluation of Statistical Methods for Forecasting Mortality: The Lee-Carter Method and Its Alternative *

Taku Yamamoto[1,4], Hiroaki Chigira[2], and Chisako Yamamoto[3]
[1]Nihon University, Chiyoda-ku, Tokyo, JAPAN
[2]Tohoku University, Sendai, Miyagi-ken, JAPAN
[3]Chukyogakuin University, Mizunami, Gifu-ken, JAPAN
[4]Corresponding author: Taku Yamamoto.
e-mail: yamamoto.taku@nihon-u.ac.jp

Abstract

Forecasting mortality has been a vital issue in demography and actuarial science. It has important implications for the pension plan and long-term economic forecasts of the nation. In the present paper, we evaluate statistical properties of the well-known Lee-Carter method for forecasting mortality. Here, it is assumed that a series of life tables represents a realization of a high dimensional cointegrated process. We also propose an alternative method. Throughout theoretical and experimental analyses, we find that the proposed method is more accurate than the Lee-Carter method for short-term (say, the 1st to the 10th period ahead) and medium-term (say, the 20th to the 30th period ahead) forecasts, although both methods have the similar accuracy for long-term (say, the 50th period and further ahead) forecasts. We further find more favorable results of the proposed method in the analysis of the mortality data of Japanese male for the short-term forecasts.

Keywords: Cointegrated process, life table, Monte Carlo experiment, Mortality forecasting, principal components

1 Review of the Lee-Carter Method and a Mortality Model

(a) Review of the Lee-Carter Method

We first review the Lee-Carter (hereafter, LC) method for forecasting Mortality. Let w_{at} denote the mortality rate of age $a = 1, 2, \cdots, m$, and year $t = 1, 2, \cdots, T$. In the LC method, we actually analyze the log of the original data $y_{at} = \log(w_{at})$. Here, we introduce, for the sake of simple presentation, $(m \times 1)$ vector $y_t = [y_{1t}, \cdots y_{mt}]'$. Next, we define the deviation from mean $\tilde{y}_t = y_t - \bar{y}$, where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ is the sample mean. We proceed to the singular value decomposition: Let f_1 be the first characteristic vector. Next, we obtain the first principal component (PC) as $\{f_1'\tilde{y}_t\}$ and suppose that $f_1'\tilde{y}_t$ is approximated by a random walk with the drift. The drift is estimated as follows:

$$\sum_{t=2}^{T-1} f_1'\Delta\tilde{y}_t == f_1'\overline{\Delta y}.$$

Then, h period ahead forecast of the first PC is given by $f_1'\overline{\Delta y}h + f_1'\tilde{y}_T$. Finally, the LC forecast is obtained by multiplying f_1 to it and adds the sample mean, in order to get forecasts for y_{T+h}:

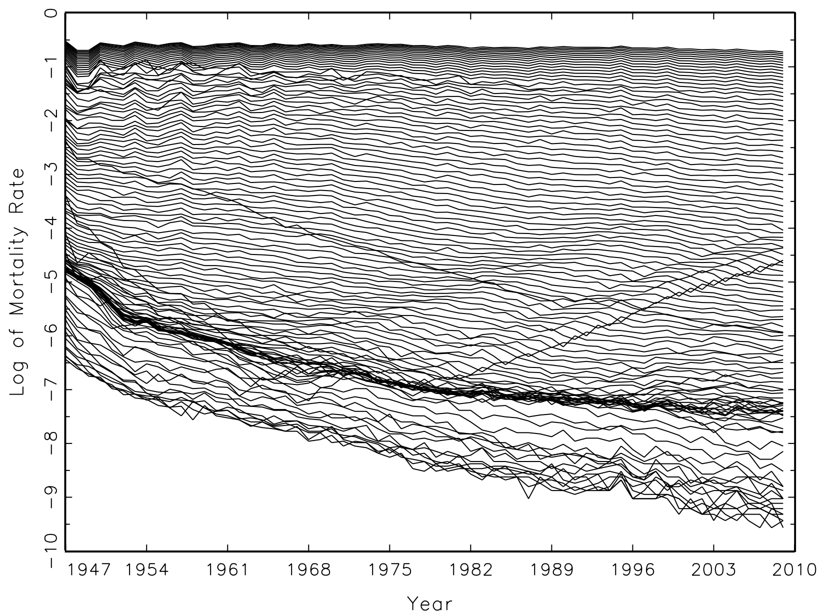
$$\hat{y}_{T+h}^{LC} = f_1(f_1'\overline{\Delta y}h + f_1'\tilde{y}_T) + \bar{y} = f_1 f_1'\overline{\Delta y}h + f_1 f_1'\tilde{y}_T + \bar{y} = f_1 f_1'\overline{\Delta y}h + y_T^{LC}. \tag{1}$$

(b) A Mortality Model

In the standard prediction theory in time series analysis, we first set the DGP (data generating process), and construct the forecast based upon the assumed DGP. However, Lee and Carter (1992) does not assume any DGP for y_t. In such a situation, we cannot analytically evaluate forecasts. Further, without the DGP for y_t, we do not know what f_1 and $f_1 f_1'\overline{\Delta y}$ in (1) try to estimate. In this paper, we explicitly assume the DGP for y_t.

*Research of Chigira and Yamamoto is partially supported by Grant-in-Aid for Scientific Research (A), No. 23243039.

Figure 1: Log of Mortality Rate of Japanese Male (Age 0 to Age 109)



In order to get insight in the movement of y_t, Figure 1 shows mortality of Japanese male, from age 0 to age 109, for 1947 - 2009, obtained from Human Mortality Database[2]. We can see from this figure that each element of y_{at} has a linear downward trend and shows a similar movement each other. Similar movements among $I(1)$ elements suggest that y_t can be approximated by a cointegrated process. It is in accordance with some previous studies such as Bell (1997) and Darkiewicz and Hoedemakers (2004), which state that cointegration analysis is useful for evaluating and adjusting the Lee-Carter method.

We assume that the DGP for y_t is explicitly expressed as the following model

$$\Delta y_t = C(L)(\varepsilon_t + \Phi D_t), \ \ \varepsilon_t \sim i.i.d.(0, \Omega), \ \ t = 1, \cdots, T, \tag{2}$$

where Δy_t is the first difference of y_t, $C(L) = \sum_{i=0}^{\infty} C_i L^i$ is such that $C(z)$ is convergent for some $|z| < 1 + \delta$ with some $\delta > 0$, D_t is the deterministic term, Φ is its parameter matrix、and Ω is the variance-covariance matrix which is symmetric and positive definite. For the deterministic term ΦD_t, it is modeled as $\Phi D_t = d_1 + d_2 t$,where d_1 and d_2 are $(m \times 1)$ parameter vectors. On the other hand, the level model is written as

$$y_t = \gamma + \mu t + C \sum_{s=1}^{t} \varepsilon_s + C^*(L)\varepsilon_t = \gamma + \mu t + C \sum_{s=1}^{t} \varepsilon_s + \varepsilon_t^*, \tag{3}$$

where $\mu = Cd_1 + C^*(1)d_2 = Cd_1 + d_2^*$, $C = C(1)$, $C^*(L) = \sum_{i=0}^{\infty} C_i^* L^i$, and $C_i^* = -\sum_{j=i+1}^{\infty} C_j$ $(i = 0, 1, \cdots)$. $C^*(z)$ is convergent for $|z| < 1 + \delta$ with some $\delta > 0$. $\gamma = C^*(1)d_1 - \left(\sum_{i=1}^{\infty} iC_i^*\right) d_2$. The representation (2) or (3) is the model for the DGP in this study. We may note that the implication of the deterministic term implies that, multiplying β' to (3) from left, we have

$$\beta' y_t = \beta' \gamma + \beta' d_2^* t + \beta' \varepsilon_t^*. \tag{4}$$

It means that cointegrating relation gives trend-stationarity rather than mean-stationarity.

(c) Criterion of Forecasting Accuracy

In this paper, we here employ trace MSE (mean squared error) as a a measure of forecasting accuracy which has been widely adapted in previous studies. Trace MSE

[2]See reference for its URL.

is defined as follows: Let e^{i}_{T+h} be the forecasting error of the i-th forecasting method, $e^{\mathrm{i}}_{T+h} = y_{T+h} - \hat{y}^{\mathrm{i}}_{T+h}$, where $\hat{y}^{\mathrm{i}}_{T+h}$ is the $(m \times 1)$ forecasting vector of the i-th method for period $T + h$. Trace MSE is defined as expectation of squared e^{i}_{T+h} as follows:

$$\text{tr MSE}(\hat{y}^{\mathrm{i}}_{T+h}) = \text{tr } (E(e^{\mathrm{i}}_{T+h} {e^{\mathrm{i}}}'_{T+h})) = E({e^{\mathrm{i}}}'_{T+h} e^{\mathrm{i}}_{T+h}).$$

2 Theoretical Properties of the Lee-Carter Method

(a) Strong Points of the Lee-Carter Method

The LC forecast is asymptotically given by

$$\hat{y}^{LC}_{T+h} \approx \mu h + y^{LC}_{T}, \tag{5}$$

where "$\approx$" means the asymptotic property when T is large.

From (5), we can see that the LC method correctly captures the drift term. Thus, the LC method is optimal in the long-term forecasting as discussed in Chigira and Yamamoto (2012).

Another strong point is that the LC method is very simple to execute. Namely, it estimates only one parameter, that is, the drift term of the first PC. Conventional forecasting methods for a cointegrated process need to estimate the whole system, and the number of parameters to estimate is $O(m^2)$, and accuracy of estimation deteriorates when the sample size T is small, and then the forecasting accuracy will also deteriorate. The LC method does not suffer the problem. This point is quite important for the analysis of mortality data where T is usually not sufficiently large relative to m.

Further, as Lee and Carter (1992) asserts, it is easy to construct the confidence interval of the forecast.

(b) Weak Points of the Lee-Carter Method

As Girosi and King (2007) states that, since the LC method uses only the first principal component, information contained in the rest of principal components are thrown away. The information loss results in the following two weak points.

(i) The starting point of forecast y^{LC}_{T} in (5) deviates from y_T.

(ii) The drift term is inefficiently estimated.

Further, the LC method applies the principal component analysis without detrending the data y_t, although it apparently has a negative trend. It results in the third weak point, that is, lack of cointegration restriction in its forecasts.

3 Alternative Method: The MTV Method

In order to overcome the weak points of the LC method, we employ the MTV（multivariate time series variance component） method. the MTV method for a cointegrated process was introduced in Chigira and Yamamoto (2009). The MTV method employs the principal component analysis as in the LC method, but they are different in two respects: First, the MTV method applies the principal component analysis after detrending the data y_t. Second, the MTV method uses all principal components in constructing forecasts. Thus, it does not loose information and cointegration restrictions are imposed on its forecasts.

We briefly describe the MTV method below: In order to apply the MTV method,to the model (2) or (3), we regress y_t on the constant term and the trend term in order to detrend the data y_t. The detrended residual vector is given by $\check{y}_t = y_t - \hat{\gamma} - t\hat{\mu}_{trend}$, where $\hat{\gamma}$ and $\hat{\mu}_{trend}$ are the the OLS estimates of the constant and the trend terms, respectively.

Based upon $\check{y}_t$, we proceed to the singular value decomposition. Let $B_{(m-r)} = \begin{bmatrix} b_1 & \dots & b_{m-r} \end{bmatrix}$ and $B_{(r)} = \begin{bmatrix} b_{m-r+1} & \dots & b_m \end{bmatrix}$, where $b_1, \cdots, b_m$ are the characteristic vectors corresponding to the characteristic values $\pi_1 \geq \cdots \geq \pi_m$. We have the following property:

$$\begin{cases} B'_{(m-r)}\check{y}_t \approx I(1), \\ B'_{(r)}\check{y}_t \approx I(0). \end{cases} \tag{6}$$

Table 1: Distribution of estimated rank of cointegration ($\hat{r}$)

(a) No restriction

$T\backslash\hat{r}$	26	27	28	29	30
50	1.6	24.5	48.7	13.1	12.1
100	2	57.3	35.2	2.7	2.8
200	0.9	89.5	9.1	0.5	0
500	0.8	93.8	5.2	0.2	0

(b) A priori restriction $r \leq 29$

$T\backslash\hat{r}$	26	27	28	29	30
50	2.5	30.3	53.4	13.8	0
100	2.1	59.1	36	2.8	0
200	0.9	89.5	9.1	0.5	0
500	0.8	93.8	5.2	0.2	0

We, then, forecast with an ARIMA(p,1,q) model for each of the first to the $(m-r)$-th principal components $\widehat{B'_{(m-r)}\breve{y}_{T+h}}$, and with an ARIMA($p$,0,$q$) model for each of the $(m-r+1)$-th to the m-th principal components $\widehat{B'_{(r)}\breve{y}_{T+h}}$.

We next multiply the characteristic vector matrix to the above forecasts in order to obtain the forecast of $\breve{y}_t$. Finally, we add the constant and the drift terms back to get the forecast of y_{T+h}. Specifically, we have

$$\hat{y}_{T+h}^{MTV} = (T+h)\hat{\mu}_{trend} + \hat{\gamma} + B_{(m-r)}\widehat{B'_{(m-r)}\breve{y}_{T+h}} + B_{(r)}\widehat{B'_{(r)}\breve{y}_{T+h}}. \tag{7}$$

Thus, the number of parameters to estimate increases as $O(m)$, and substantially less than that of the VEC (vector error correction) model which is the standard one used for the analysis of a cointegrated process. We can also show that the MTV method satisfies the cointegration restrictions.

4 Monte Carlo Experiment

(a) Design of Experiment

The following VEC is used as the DGP:

$$\Delta y_t = \alpha\beta' y_{t-1} + \mu + \varepsilon_t, \;\; \varepsilon_t \sim NID(0, I_m), \tag{8}$$

where α is $(m \times r)$ full-rank matrix.

We consider the case where $m = 30$ and $r = 27$, $T = 50, 200$ and $h = 1, 2, \cdots, 5, 10, 20, 30, 40, 50$. The number of replication is 1000. This case mimics the actual situations where m is relatively large to the sample size T. Specific parameter values of α and β are omitted in lieu of space.

We compare the forecasts of the MTV method (7) and the LC method (1), and the individual ARIMA method. The individual ARIMA method is to fit ARIMA model to each series $\{y_{at}\}$ $(a = 1, 2, \cdots, m)$. The orders of AR and MA parts are determined by SBIC for each series. We denote forecasts of the individual ARIMA method as $\hat{y}_{T+h}^{ARIMA}$.

In practice, the cointegration rank r is unknown and we have to estimate it. For that purpose, as a practical expedient, we sequentially apply the stationary test of Kurozumi and Tanaka (2010) which is the improved version of Kwiatkowski et al. (KPSS) (1992) test to ecah principal component $\{b_i'\breve{y}_t\}$ $(i = 1, 2, \cdots, m)$. The cointegration rank obtained in this way is denoted as $\hat{r}$. Table 1 shows how it works in our experiments.

In order to compare forecasting accuracy, we calculate the following 2 trace MSE ratios:

$$\text{ratio(MTV)} = \frac{\text{tr MSE}(\hat{y}_{T+h}^{MTV})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})}, \quad \text{and} \quad \text{ratio(LC)} = \frac{\text{tr MSE}(\hat{y}_{T+h}^{LC})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})}.$$

If the ratio is less than unity, the forecast in the numerator is more accurate than the individual ARIMA method and vice versa.

(b) Experimental Results

Trace MSE ratios are given in Table 2. When the sample size is small, $T = 50$, the forecasting accuracy of the MTV method is almost the same as the individual ARIMA

Table 2: trace MSE ratios

$T = 50$

h	1	2	3	4	5	10	20	30	40	50
ratio(MTV)	1.04	0.99	0.96	0.96	0.96	0.98	1.00	1.02	1.04	1.05
ratio(LC)	1.82	1.32	1.16	1.11	1.07	1.03	1.01	1.01	1.01	1.01

$T = 200$

h	1	2	3	4	5	10	20	30	40	50
ratio(MTV)	0.92	0.88	0.86	0.86	0.87	0.92	0.96	0.99	1.00	1.01
ratio(LC)	4.15	2.59	2.08	1.86	1.70	1.41	1.24	1.17	1.14	1.12

Table 3: trace MSE ratio（Japanese Male, age 30 ～ age 59）

$m = 30$、$\hat{r} = 27$

h	1	2	3	4	5
ratio(MTV)	1.04	0.74	0.73	0.50	0.25
ratio(LC)	3.39	4.08	4.83	3.27	3.00

method, whereas that of the LC method is worse than the individual ARIMA for the short- and the medium-term forecasts. As T is increased to 200, the MTV method becomes more accurate than the individual ARIMA for the short- and the medium-term forecasts, whereas the forecasting accuracy of the LC method deteriorates for the short- and the medium-term forecasts. For the long-term forecasts, both trace MSE ratios converge to unity as discussed in Chigira and Yamamoto (2012).

5 Applications to Japanese Mortality Data

We apply the three methods to the mortality data of Japanese male from age 30 to 59. The data from 1947 to 2004 is used for estimation and from 2005 to 2009 for forecasts. The data was obtained from Human Mortality Database. Table 3 shows trace MSE ratios of forecasts. The MTV method is substantially better than the ARIMA method except the first period ahead. The LC method extremely worse than the individual ARIMA method, ratio(LC) being greater than 3. Thus, in the short-term forecast, superiority of the MTV method over the LC method is more evident than the experimental results in the previous section suggest.

6 Concluding Remarks

In the present paper, we evaluated the LC method in the perspective of time series analysis, more specifically, in the framework of a cointegrated process. We also proposed an alternative, the MTV method. We have found the MTV method is superior to the LC method in the short- and the medium-term forecasts both in experiments and in an empirical application. It may be further noted that the LC method is inferior even to the individual ARIMA method in the short- and the medium-term forecasts.

References

Bell, W.R. (1997): “Comparing and Assessing Time Series Methods for Forecasting Age Specific Demographic Rates,” *Journal of Official Statistics*, 13, 279–303.

Chigira, H. and T. Yamamoto (2009): “Forecasting in Large Cointegrated Processes,” *Journal of Forecasting*, 28, 631–650.

Chigira, H. and T. Yamamoto (2012): “The Effect of Estimating Parameters on Long-Term Forecasts for Cointegrated Systems,” *Journal of Forecasting*, 31, 344–360.

Darkiewicz, G. and T. Hoedemakers (2004): " How the Co-integration Analysis can Help in Mortality Forecasting," Unpublished Manuscript, Actuarial Science Research Group, Catholic University of Leuven.

Girosi, F. and G. King (2007): " Understanding the Lee-Carter Mortality Forecasting Method," Unpublished Manuscript.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www. mortality.org or www.humanmortality.de (data downloaded on 22/08/2012).

Keilman, N. (2003): "Types of Models for Projecting Mortality," in Bengtsson, T. and N. Keilman eds. *Perspectives on Mortality Forecasting,* 19–27, Swedish National Social Insurance Board.

Kurozumi, E., and S. Tanaka (2010): "Reducing the Size Distortion of the KPSS Test," *Journal of Time Series Analysis*, 31, 415–426.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics,* 54, 159–178.

Lee, R.D. and L.R. Carter (1992): "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association*, 87, 659–671.