The Diagnostic Checking of the Lee-Carter Mortality Forecasting Method[1]

Arlt Josef
University of Economics, Prague, Department of Statistics and Probability
nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic
E-mail: arlt@vse.cz

Arltová Markéta
University of Economics, Prague, Department of Statistics and Probability
nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic
E-mail: arltova@vse.cz

Klůfa Jindřich
University of Economics, Prague, Department of Mathematics
nám. W. Churchilla 4, 130 67 Prague 3, Czech Republic
E-mail: klufa@vse.cz

1. The Lee-Carter Method

1.1 The Lee-Carter model (LC model)

Since its introduction (Lee and Carter, 1992) the Lee-Carter method become the "leading statistical model of mortality in the demographic literature" (Deaton and Paxson, 2004). Lee and Carter developed their method for U.S. mortality data in years 1933-1987. This method has been than applied to analyze and mainly to forecast the mortality in many other countries.

The principle of the Lee-Carter method is relatively simple. If m_{gt} denote the log of mortality rate in the age group $g = 1, 2, ..., G$ and time $t = 1, 2, ..., T$, the Lee-Carter model has form

$$m_{gt} = \alpha_g + \beta_g \gamma_t + \varepsilon_{gt},$$

where α_g, β_g and γ_t are parameters to be estimated and ε_{gt} are homoskedastic normally distributed random disturbances with mean 0 and variance σ_ε^2. The identification of the model is ensured by the constraints

$$\sum_{t=1}^{T} \gamma_t = 0 \text{ and } \sum_{g=1}^{G} \beta_g = 1.$$

The parameters of model (1) have the following interpretation: Because of the constraint $\sum_{t=1}^{T} \gamma_t = 0$ the parameter $\exp\alpha_g$ represents the general shape of the mortality schedule. The parameter β_g indicates the dependence of the log mortality at age g on the time trend γ_t, which can be interpreted as the overall mortality index (the general time level of mortality). The shape of the function β_g shows how quickly the mortality rates decline over time in comparison with trend γ_t.

Lee and Carter created the model mainly for forecasting purposes. They assumed $\hat{\alpha}_g$ and $\hat{\beta}_g$ remain constant over time and the overall mortality index $\hat{\gamma}_t$, which is in the fact the univariate time

[1] This paper was written with the support of Grant Agency of the Czech Republic No. 402/09/0369 „Modelling of Demographic Time Series in Czech Republic“

series, will be modeled by the principles of the Box-Jenkins methodology. The aim is to create its forecasts and consequently with the help of estimates $\hat{\alpha}_g$ and $\hat{\beta}_g$ the forecasts of the mortality. After testing of several ARIMA models, they find that the most appropriate model for forecasting of the US data is a random walk with drift.

1.2 The Augmented Common Factor Lee-Carter model (ACFLC model)

Li and Lee (2005) pointed that the Lee-Carter method works well for a single population - for one sex or two sexes combined. But there is problem with the application of this method to forecast mortality for the two sexes in the same country. In the case of the separate forecasting of two sexes, the male and female mortality can diverge increasingly over time even if this effect has not been observed in history. The similar problem comes up in the forecasting of the mortality in different provinces or races in a country, different countries in a region etc.

To avoid this problem Li and Lee suggested to extend the Lee-Carter model in to the so called augmented common factor model. If m_{gti} denotes the log of mortality rate in the age group $g = 1, 2, ..., G$, time $t = 1, 2, ..., T$ and for example sex $i = 1, 2$ in one country, the augmented common factor Lee-Carter model has form

$$m_{gti} = \alpha_{gi} + \mathrm{B}_g \Gamma_t + \beta_{gi}\gamma_{ti} + \varepsilon_{gti}.$$

here α_{gi}, B_g, Γ_t, β_{gi} and γ_{ti} are parameters to be estimated and ε_{gti} are homoskedastic normally distributed random disturbances with mean 0 and variance σ_ε^2.

The parameter $\exp\alpha_{gi}$ represents the general shape of the mortality schedule of sex i. In this model it is assumed that the both sexes in the group have the same β_g and γ_t, which are denoted as B_g and Γ_t. The factor $\beta_{gi}\gamma_{ti}$ is specific for sex i and allows for a short-term or medium-term difference between the rate of change in sex i mortality rates and that rate of change implied by the common factor Γ_t.

The common time factor $\hat{\Gamma}_t$ follows the random walk model with drift or some form of the nonstationary ARIMA model. If the sex specific time factor $\hat{\gamma}_{ti}$ is nonstationary time series, than the forecasts would be divergent. In the case of their stationarity the mortality forecasts would follow the non divergent behavior of the both sexes mortality in the past.

1.3 The Cointegrated Lee-Carter model (CLC model)

The problem with the forecasting of both sexes in the same country, mentioned in the part 1.2, can be solved also by another way. It can be hypothesized that the overall mortality index of men $\hat{\gamma}_{tM}$ and women $\hat{\gamma}_{tW}$ are in some relationship. From the empirical analysis it follows that the both indexes are the nonstationary time series, which can be generally represented by the ARIMA(., 1, .) models. Therefore, it is suitable to apply the cointegration analysis and to identify the presence of the long-run and the short-run relationships. The cointegration of the overall mortality indexes $\hat{\gamma}_{tM}$ and $\hat{\gamma}_{tW}$ will lead to the error correction model which will be used for their forecasting (Arlt, Arltová, Bašta, Langhamrová, 2010). The presence of the long-run relationship will tie up both $\hat{\gamma}_{tM}$ and $\hat{\gamma}_{tW}$ forecasts and thus also the forecasts of male and female mortalities.

2. The Practical Application of the Lee-Carter model

2.1 The Diagnostic Checking

The important feature of the Lee-Carter model is the long term relationship between the log of mortality rate in the age group g and the overall mortality index γ_t. Darkiewicz and Hoedemakers (2004) showed the cointegration relationship between the log-mortality rates in the two different ages. In more detail, consider age x_1 and age x_2, than

$$m_{x_1 t} = \alpha_{x_1} + \beta_{x_1}\gamma_t + \varepsilon_{x_1 t},$$
$$m_{x_2 t} = \alpha_{x_2} + \beta_{x_2}\gamma_t + \varepsilon_{x_2 t}.$$

From the relationship

$$\beta_{x_2} m_{x_1 t} - \beta_{x_1} m_{x_2 t} = (\beta_{x_2}\alpha_{x_1} - \beta_{x_1}\alpha_{x_2}) + (\beta_{x_2}\beta_{x_1} - \beta_{x_1}\beta_{x_2})\gamma_t + (\beta_{x_2}\varepsilon_{x_1 t} - \beta_{x_1}\varepsilon_{x_2 t})$$

it follows that if $\varepsilon_{x_1 t}$ and $\varepsilon_{x_2 t}$ are stationary time series, than the mortality rates in ages x_1 and x_2 are cointegrated.

The same principal can be applied to other models. From the Augmented Common Factor Lee-Carter model it follows for the population i the relationship

$$B_{x_2} m_{x_1 ti} - B_{x_1} m_{x_2 ti}$$
$$= \left(B_{x_2}\alpha_{x_1 i} - B_{x_1}\alpha_{x_2 i}\right) + (B_{x_2}B_{x_1} - B_{x_1}B_{x_2})\Gamma_t + (B_{x_2}\beta_{x_1 i} - B_{x_1}\beta_{x_2 i})\gamma_{ti} + (B_{x_2}\varepsilon_{x_1 ti} B_{x_1}\varepsilon_{x_2 ti}).$$

It is clearly seen that when γ_{ti} is stationary time series the log mortality rates in age x_1 and x_2 in the population i are cointegrated.

For the populations i and j than

$$B_{x_2} m_{x_1 ti} - B_{x_1} m_{x_2 tj} =$$
$$= \left(B_{x_2}\alpha_{x_1 i} - B_{x_1}\alpha_{x_2 j}\right) + (B_{x_2}B_{x_1} - B_{x_1}B_{x_2})\Gamma_t + B_{x_2}\beta_{x_1 i}\gamma_{ti} - B_{x_1}\beta_{x_2 i}\gamma_{tj} + (B_{x_2}\varepsilon_{x_1 ti} - B_{x_1}\varepsilon_{x_2 tj}).$$

The log mortality rates in age x_1 and x_2 in the population i and j are cointegrated when γ_{ti} and γ_{tj} are both stationary.

In the case of the Cointegrated Lee-Carter method the long-term relationship of the log mortality rates in the population i and j has form

$$\beta_{x_2} m_{x_1 ti} - \beta_{x_1} m_{x_2 tj} = \left(\beta_{x_2}\alpha_{x_1 i} - \beta_{x_1}\alpha_{x_2 j}\right) + \beta_{x_2}\beta_{x_1}\gamma_{ti} - \beta_{x_1}\beta_{x_2}\gamma_{tj} + (\beta_{x_2}\varepsilon_{x_1 ti} - \beta_{x_1}\varepsilon_{x_2 tj}).$$

The log mortality rates in age x_1 and x_2 and in population i and j are cointegrated when also γ_{ti} and γ_{tj} are cointegrated.

The above mentioned properties of different forms of the Lee-Carter model can be used for their empirical diagnostic checking and practical verification. It follows the simple rule. If the age-specific log-mortality rates are pairwise cointegrated, than the Lee-Carter model is applicable to the log-mortality data. It will be illustrated for the log-mortality rates of the Czech Republic. We have got the mortality data for yearly time period from 1952 to 2008. The source of this data is the Human Mortality Database (2008).

First, the analyzed time series of the log-mortality rates will be tested for the presence of the unit root. Table 1 contains the results of the ADF test. It is seen that except of the female log-mortality rate for age 25-29 the hypotheses of existence of unit root cannot be rejected (5 % sign. level).

Table 1: ADF test of log-mortality rates

	Male		Female	
Age group	p-value	Lag	p-value	Lag
10-14	0.5486	2	0.0762	1
25-29	0.2437	0	0.0494	2
40-44	0.9920	1	0.2156	2
55-59	0.9614	0	0.7312	0
70-74	0.9918	0	0.9629	0

Table 2 shows the results of the cointegration test of the age-specific log-mortality rates given by the ADF test of the residual time series from the static regression. We use the notation introduced by Darkiewicz and Hoedemakers (2004), i. e. the label *x*S*y* means the cointegration test for the sex S, where "M" means males, "F" means females and "MF" means that the first variable represents males and the second represents females, *x* and *y* means ages. It is clearly seen that the most of tested male and female pairs of the log-mortality rates are cointegrated which leads to the conclusion that the classical Lee-Carter model is applicable for the Czech data. On the other side only approximately half of pairs of combined male and female log-mortality rates are cointegrated which does not give much support for the practical application of the Augmented Common Factor Lee-Carter model and Cointegrated Lee-Carter model.

Table 2: Cointegration test

Age group	p-value	Lag	Age group	p-value	Lag
Male			Male & Female		
10-14M25-29	0.0000	0	10-14MF10-14	0.0016	1
10-14M40-44	0.0089	1	10-14MF25-29	0.0054	1
10-14M55-59	0.0094	1	10-14MF40-44	0.0026	1
10-14M70-74	0.0009	1	10-14MF55-59	0.0180	1
25-29M40-44	0.0083	0	10-14MF70-74	0.0135	1
25-29M55-59	0.0063	0	25-29MF10-14	0.0002	0
25-29M70-74	0.0002	0	25-29MF25-29	0.1010	2
40-44M55-59	0.0849	1	25-29MF40-44	0.0100	1
40-44M70-74	0.0561	1	25-29MF55-59	0.0039	0
55-59M70-74	0.0391	0	25-29MF70-74	0.0003	0
Female			40-44MF10-14	0.4953	2
10-14F25-29	0.0000	0	40-44MF25-29	0.7444	1
10-14F40-44	0.0000	0	40-44MF40-44	0.6252	1
10-14F55-59	0.0055	1	40-44MF55-59	0.3912	1
10-14F70-74	0.0022	1	40-44MF70-74	0.3053	1
25-29F40-44	0.0000	0	55-59MF10-14	0.2643	0
25-29F55-59	0.0809	2	55-59MF25-29	0.2540	0
25-29F70-74	0.0521	2	55-59MF40-44	0.1285	0
40-44F55-59	0.0052	1	55-59MF55-59	0.0489	0
40-44F70-74	0.0020	1	55-59MF70-74	0.1257	0
55-59F70-74	0.0000	0	70-74MF10-14	0.5037	1
			70-74MF25-29	0.5362	1
			70-74MF40-44	0.2465	1
			70-74MF55-59	0.0053	0
			70-74MF70-74	0.0232	0

2.2 The Mortality Rate Forecasting

The Fig. 1 and 2 show the forecasts of the mortality rates computed on the base of the equation

$$\widehat{m}_{gt} = \hat{\alpha}_g + \hat{\beta}_g \hat{\gamma}_t.$$

The second type of graphs has the color table ordered accordance with the succession of rainbow, so the oldest time series have the red color, the youngest have the violet color. It is clearly seen that the mortality declines with the growing forecast horizon, more in lower ages.

Fig. 1. The mortality rate of the Czech men forecasts by Lee-Carter model for years 2009-2050

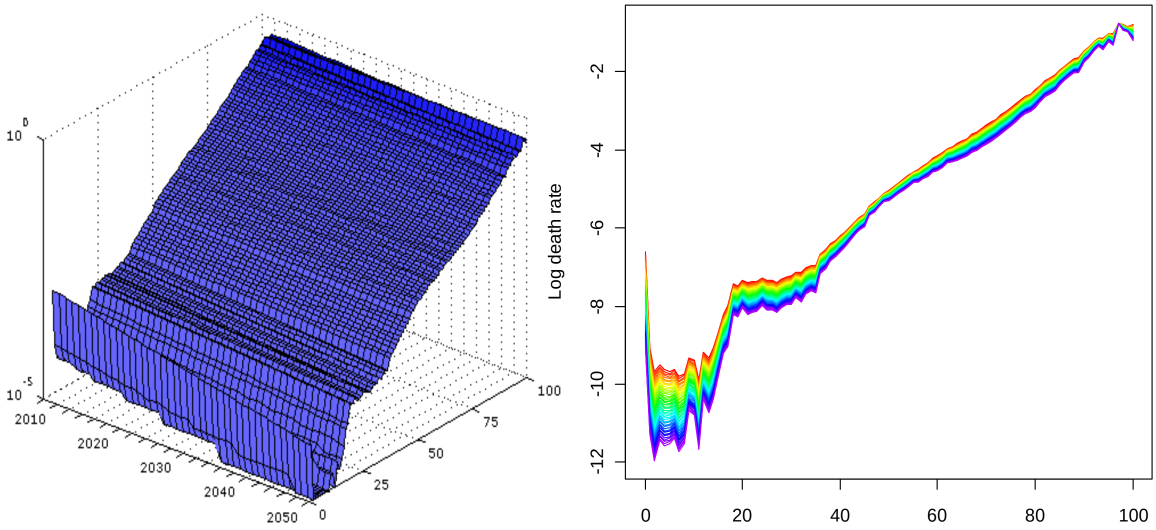
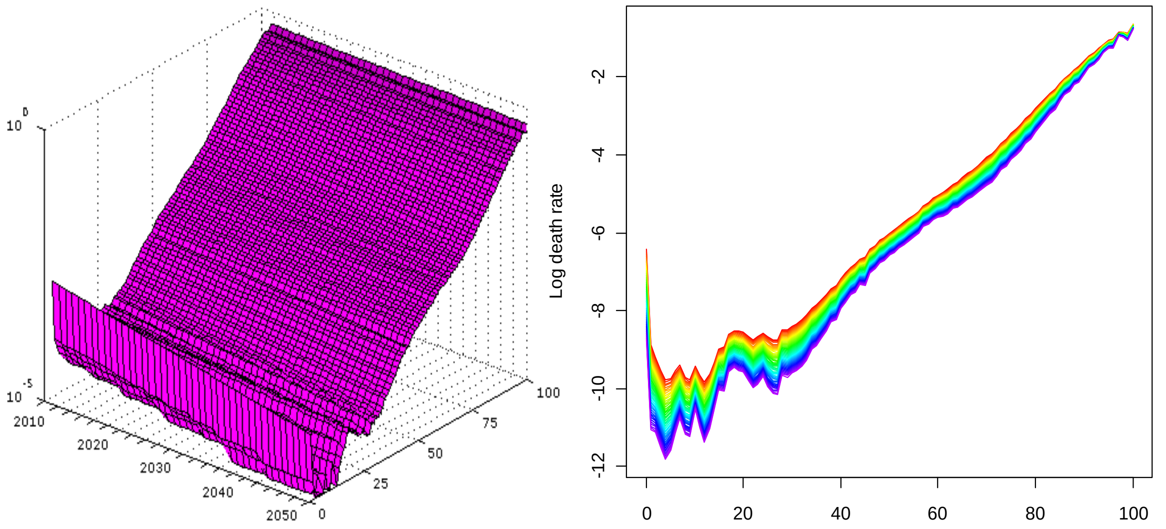


Fig. 2. The mortality rate of the Czech women forecasts by Lee-Carter model for years 2009-2050



REFERENCES:

Arlt, J., Arltová, M.: Ekonomické časové řady. Professional Publishing. 2009.

Arlt, J., Arltová, M., Bašta, M., Langhamrová, J.: Cointegrated Lee-Carter Mortality Forecasting Method. Paris 22.-27.08.2010. In: *COMPSTAT 2010.* Paris : CNAM and INRIA, s. 713–720, 2010.

Darkiewicz, G., Hoedemakers, T.: How the Co-integration Analysis can Help in Mortality Forecasting. Actuarial Science Research Group, Catholic Univ. of Leuven. 2004.

Deaton, A., Paxson, Ch.: Mortality, Income, and Income Inequality Over Time in the Britain and the United States. Technical Report 8534 National Bureau of Economic Research Cambridge, MA:. <http://www.nber.org/papers/w8534>

Engle, R., Granger, C. W. J.: Cointegration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 1987, 251-276.

Girosi, F., King, G.: Understanding the Lee-Carter Mortality Forecasting Method. 2007. <http://gking.harvard.edu/files/lc.pdf>

Human Mortality Database. 2008 University of California, Berkeley (USA) a Max Planck Institute for Demographic Research (Německo). Dostupné z http://www.mortality.org

Hyndman, R. J.: Demography: Forecasting mortality, fertility, migration and population data, R package v. 1.09, 2011. <http://robjhyndman.com/software/demography>.

Hyndman, R. J., Shang, H. L.: Rainbow plots, bagplots and boxplots for functional data, *Journal of Computational & Graphical Statistics* 19(1), 29-45, 2010. <http://robjhyndman.com/papers/Rainbow5.pdf>.

Lee, R.: The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications, North American Actuarial Journal, Vol. 4, No. 1, 2000.

Lee, R. D., Carter, L.: Modeling and Forecasting the Time Series of U. S. Mortality. Journal of the American Statistical Association 1992; 87, 659-671

Li, N., Lee, R.: Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method, Demography, 2005; 42(3): 575-594

ABSTRACT

The model diagnostic checking is very important part of the model construction. From the definition of the classical Lee-Carter model it follows the long-term relationship between the log-mortality rates and the overall mortality index. Thus if the model is correctly specified the log-mortality rates for ages x_1 and x_2 should be cointegrated. This diagnostic principle is generalized and applied also for both the Augmented Common Factor Lee-Carter Method (ACFLC) model and the Cointegrated Lee-Carter (CLC) model. The practical verification is made on the bases of Czech mortality data.