Introduction to the Lee-Carter Model

November 23, 2017

Definition

- (x): A life aged x; $x \geqslant 0$.
- T_x: Future lifetime of (x), a random variable; so $x + T_x$ denotes age-at-death of (x).
- $F_x(t)$: The distribution function of T_x; $F_x(t) \equiv \mathrm{P}(T_x \leqslant t)$. (Should have been written as $F_{T_x}(t)$; an abbreviation.)
- $S_x(t)$: $S_x(t) \equiv 1 - F_x(t) = \mathrm{P}(T_x > t)$.

Interpretation of $F_x(t), S_x(t)$

- $F_x(t)$: the probability that (x) survives $\leqslant t$ years.
- $S_x(t)$: the probability that (x) survives $> t$ years.

Important Postulate

$$\mathrm{P}(T_x \leqslant t) = \mathrm{P}(T_0 \leqslant x + t \,|\, T_0 > x)$$

Note that from the rules of conditional probabilities,

$$\mathrm{P}(T_0 \leqslant x + t \,|\, T_0 > x) = \frac{\mathrm{P}(x < T_0 \leqslant x + t)}{\mathrm{P}(T_0 > x)}$$

$$F_x(t) \equiv \mathrm{P}(T_x \leqslant t) = \mathrm{P}(T_0 \leqslant x + t \,|\, T_0 > x) = \frac{F_0(x+t) - F_0(x)}{S_0(x)}$$

Use of $S_x(t) = 1 - F_x(t)$,

$$F_x(t) = \frac{F_0(x+t) - F_0(x)}{S_0(x)} \quad \Longrightarrow \quad S_x(t) = \frac{S_0(x+t)}{S_0(x)}$$

Hence

$$S_x(t+u) = \frac{S_0(x+t+u)}{S_0(x)} = \frac{S_0(x+t)}{S_0(x)} \frac{S_0(x+t+u)}{S_0(x+t)} = S_x(t)\, S_{x+t}(u)$$

and note the

Conditions of S_x

- $S_x(0) = 1$.
- $S_x(t) \to 0$ as $t \to \infty$.
- $S_x(t)$ is non-increasing in t.

The Actuarial Notation

- ${}_tp_x \equiv \mathrm{P}(T_x > t) = S_x(t)$, the probability that (x) survives to at least $x+t$ (i.e. (x) survives $> t$ years).
- ${}_tq_x \equiv \mathrm{P}(T_x \leqslant t) = F_x(t)$, the probability that (x) dies before $x+t$ (i.e. (x) survives $\leqslant t$ years).

N.B. ${}_1p_x, {}_1q_x$ are often denoted by p_x, q_x respectively.

Two Most Important Identities

$$ {}_tp_x + {}_tq_x = 1 $$

$$ {}_{t+u}p_x = {}_tp_x \cdot {}_up_{x+t} $$

Given a survival model with ${}_tp_x$, construct the life table from initial age x_0 to maximum age ω, i.e. define the function l_x, $x_0 \leqslant x \leqslant \omega$:

- Fix l_{x_0}, radix of the table.
- For $0 \leqslant t \leqslant \omega - x_0$, define

$$l_{x_0+t} = l_{x_0} \cdot {}_tp_{x_0}$$

- Hence for $x_0 \leqslant x \leqslant x + t \leqslant \omega$

$$\begin{aligned} l_{x+t} &= l_{x_0+(x-x_0+t)} \\ &= l_{x_0} \cdot {}_{x-x_0+t}p_{x_0} \\ &= \underbrace{l_{x_0} \cdot {}_{x-x_0}p_{x_0}}_{=l_x} \cdot {}_tp_x \quad (\text{using } {}_{t+u}p_x = {}_tp_x \cdot {}_up_{x+t}) \\ &= l_x \cdot {}_tp_x \end{aligned}$$

Suppose we have l_x independent lives (x), and each has the probability ${}_tp_x$ survives to $x+t$. Then

$$(\#\text{ of survivors to } x+t) \equiv \mathcal{L}_t \sim \text{binomial}(l_x, {}_tp_x).$$

So

$$(\text{expected } \#\text{ of survivors to } x+t) = \mathrm{E}\{\mathcal{L}_t\} = l_x \cdot {}_tp_x = l_{x+t}.$$

Interpretation of l_{x+t}

Out of $\# l_x$ of independent (x), the expected number of survivors to $x+t$.

Suppose we have l_x independent lives (x), and each has the probability ${}_1q_x$ survives to $x+1$. Then

$$(\#\text{ of deaths to } x+1) \equiv \mathcal{D}_1 \sim \text{binomial}(l_x, {}_1q_x).$$

So

$$(\text{expected } \#\text{ of deaths to } x+1) = \mathrm{E}\{\mathcal{D}_1\} = l_x \cdot {}_1q_x$$

Interpretation of d_x

Out of $\# l_x$ of independent (x), the expected number of death to $x+1$.

Sample Mortality Table: $q_x(t)$

Age Group	1901-11	1911-21	1921-31	1931-41	1941-51	1951-61	1961-71	1971-81	1981-91	1991-01	2001-11
0	0.3731	0.3789	0.2991	0.2549	0.2154	0.1664	0.1556	0.1449	0.0899	0.0703	0.0510
1-4	0.0574	0.0583	0.0463	0.0394	0.0331	0.0250	0.0232	0.0213	0.0096	0.0066	0.0039
5-9	0.0143	0.0139	0.0129	0.0114	0.0098	0.0078	0.0060	0.0044	0.0034	0.0023	0.0012
10-14	0.0114	0.0111	0.0103	0.0091	0.0078	0.0062	0.0048	0.0035	0.0017	0.0014	0.0009
15-19	0.0176	0.0171	0.0159	0.0141	0.0120	0.0096	0.0073	0.0054	0.0029	0.0024	0.0014
20-24	0.0227	0.0220	0.0204	0.0181	0.0155	0.0123	0.0095	0.0069	0.0037	0.0031	0.0019
25-29	0.0233	0.0227	0.0210	0.0186	0.0160	0.0127	0.0097	0.0071	0.0034	0.0032	0.0018
30-34	0.0243	0.0236	0.0219	0.0194	0.0166	0.0132	0.0101	0.0074	0.0033	0.0031	0.0020
35-39	0.0268	0.0260	0.0241	0.0214	0.0184	0.0146	0.0112	0.0082	0.0039	0.0033	0.0025
40-44	0.0308	0.0299	0.0277	0.0246	0.0211	0.0168	0.0129	0.0094	0.0048	0.0041	0.0031
45-49	0.0372	0.0361	0.0335	0.0297	0.0255	0.0203	0.0157	0.0115	0.0068	0.0058	0.0042
50-54	0.0463	0.0449	0.0418	0.0371	0.0319	0.0255	0.0197	0.0144	0.0104	0.0092	0.0065
55-59	0.0594	0.0576	0.0536	0.0478	0.0412	0.0342	0.0256	0.0188	0.0162	0.0152	0.0109
60-64	0.0776	0.0753	0.0702	0.0628	0.0544	0.0439	0.0342	0.0253	0.0269	0.0220	0.0201
65-69	0.1056	0.0995	0.0931	0.0838	0.0731	0.0596	0.0467	0.0348	0.0436	0.0381	0.0312
70-74	0.1296	0.1318	0.1240	0.1125	0.0992	0.0819	0.0651	0.0490	0.0649	0.0556	0.0540
75-79	0.1776	0.1734	0.1644	0.1511	0.1352	0.1138	0.0921	0.0707	0.0843	0.0823	0.0746
80+	0.3145	0.3120	0.3074	0.3004	0.2916	0.2793	0.2660	0.2522	0.1578	0.1302	0.1429

The Lee-Carter (1992) Model

$$\log q_x(t) = a_x + b_x \cdot k_t + \varepsilon_x(t)$$

- x: age group $(x = 0, 1-4, 5-9, \ldots)$
- t: time of life table $(t = 1901-11, 1911-21, 1921-31, \ldots)$
- $q_x(t)$: rate of mortality for age group x at time t
- a_x: average age specific pattern of mortality
- k_t: time trend of the mortality
- b_x: sensitivity of $q_x(t)$ w.r.t k_t
- $\varepsilon_x(t)$: the error associated with x and t

Note that in Lee-Carter model, the parameterization (a_x, b_x, k_t) is invariant under the transformation

$$(a_x, b_x, k_t) \mapsto \left(a_x + c\, b_x, \frac{b_x}{d}, d\,(k_t - c)\right), \quad c, d \in \mathbb{R}.$$

Hence constraints should be used to get result; Lee-Carter adopted

$$\sum_x b_x = 1, \quad \sum_t k_t = 0,$$

which implies a_x is the average of $\log q_x(t)$ over time. However, to fit Lee-Carter model needs other ideas: Principle Component Analysis (PCA).

Given $\{x_1, x_2, \ldots, x_n\}$, each $x_i \in \mathbb{R}^d$. Goal: Find a vector x_0 that

$$\text{minimize } J(x_0) \equiv \sum_{i=1}^{n} \|x_0 - x_i\|^2$$

Solution of this problem is $x_0 = m$, where

$$m = \frac{1}{n} \sum_{i=1}^{n} x_i$$

This can be shown as

$$\begin{aligned} J(x_0) &= \sum_{i=1}^{n} \|x_0 - x_i\|^2 = \sum_{i=1}^{n} \|(x_0 - m) - (x_i - m)\|^2 \\ &= \sum_{i=1}^{n} \|x_0 - m\|^2 - \underbrace{2\sum_{i=1}^{n}(x_0 - m)'(x_i - m)}_{=2(x_0-m)'\sum_{i=1}^{n}(x_i-m)=0} + \underbrace{\sum_{i=1}^{n} \|x_i - m\|^2}_{\text{independent of } x_0} \end{aligned}$$

- m, the sample mean of $\{x_1, x_2, \ldots, x_n\}$, is the zero-dimension representation of the data set, but does not reveal the variability.
- We will seek the one-dimension representation of the data set by projecting the data onto a line running through the sample mean.
- Let $e \in \mathbb{R}^d$ be a unit vector in the direction of this line; the equation of this line is

$$x = m + a\,e$$

where the scalar a corresponds to the distance of x from m.

Now we represent the data set $\{x_1, x_2, \ldots, x_n\}$ by $\{m + a_1 e, m + a_2 e, \ldots, m + a_n e\}$ with scalars $\{a_1, a_2, \ldots, a_n\}$ and unit vector e. The optimal set of $\{a_1, a_2, \ldots, a_n\}$ is determined via

$$\text{minimize } J_1(\{a_1, a_2, \ldots, a_n\}, e) = \sum_{i=1}^{n} \|(m + a_i e) - x_i\|^2$$

$$= \sum_{i=1}^{n} \|a_i e - (x_i - m)\|^2$$

$$= \sum_{i=1}^{n} a_i^2 \underbrace{\|e\|^2}_{=1} - 2 \sum_{i=1}^{n} a_i e' (x_i - m) + \underbrace{\sum_{i=1}^{n} \|x_i - m\|^2}_{\text{independent of } a_i}$$

Partially differentiate w.r.t a_i and set the derivatives to zero,

$$a_i = e'(x_i - m), \quad 1 \leqslant i \leqslant n.$$

Now substitute a_i back to J_1, we have

$$
\begin{aligned}
J_1(e) &= \sum_{i=1}^{n}[e'(x_i - m)]^2 - 2\sum_{i=1}^{n}[e'(x_i - m)]^2 + \sum_{i=1}^{n}\|x_i - m\|^2 \\
&= -\sum_{i=1}^{n}[e'(x_i - m)]^2 + \sum_{i=1}^{n}\|x_i - m\|^2 \\
&= -\sum_{i=1}^{n} e'(x_i - m)(x_i - m)'e + \sum_{i=1}^{n}\|x_i - m\|^2 \\
&= -e'\,S\,e + \sum_{i=1}^{n}\|x_i - m\|^2, \qquad \text{where } S = \sum_{i=1}^{n}(x_i - m)(x_i - m)'.
\end{aligned}
$$

Minimize $J_1(e) \iff$ Maximize $e'\,S\,e$!

To maximize $e' S e$ under the constraint $||e|| = 1$, we use the method of Lagrange multipliers. Let λ be the undermined multiplier, we differentiate

$$u = e' S e - \lambda(e' e - 1)$$

with respect to e and set to zero to obtain

$$\frac{\partial u}{\partial e} = 2S e - 2\lambda e = 0 \implies S e = \lambda e;$$

e is the eigenvector of S, with corresponding eigenvalue λ, and $e' S e = e' \lambda e = \lambda$. Hence to maximize $e' S e$, we will seek the eigenvector e of S with the greatest eigenvalue.

Singular Value Decomposition Theorem

Given a matrix $M \in \mathbb{R}^{n \times m}$, then M has the singular value decomposition (SVD) as

$$M = U \Sigma V^*$$

where $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$, are unitary matrices and $\Sigma \in \mathbb{R}^{n \times m}$ with its diagonal entries the singular values of M. The columns of U, V are called the left, right-singular vectors of M respectively. The left-singular vectors of M are a set of orthonormal eigenvectors of MM^*; The right-singular vectors of M are a set of orthonormal eigenvectors of $M^* M$.

Inspired by PCA and SVD, given the $q_x(t)$ table, to fit the Lee-Carter model consists of the following 3 steps:

- Take logarithm of $q_x(t)$; the result is a matrix $\widetilde{M}$. Set

$$a_x = \frac{1}{(\# \text{ of t})} \sum_t \log q_x(t)$$

  and $M = \widetilde{M} - a$, where a is the matrix formed from a_x.
- From the SVD of M, $M = U \Sigma V$. Then

$$k_t = U(:,1), \quad b_x = V(1,:)$$

- (Optional) Forecast k_t using ARIMA; update the model.