The Korean Journal of Applied Statistics (2018) DOI: https://doi.org/10.5351/KJAS.2018.31.5.637
31(5), 637–653

Consideration on assumption and transition of mortality model for Korea

- Discussion on the kinds of Lee-carter -

Jinho Oh[a] · Soon-Young Kim[a,1]

[a]Statistical Research Institute, Statistics Korea

(Received July 23, 2018; Revised September 14, 2018; Accepted September 16, 2018)

Abstract

Rapid aging of the population affects population structure and population aging. Consequently, developed countries have focused on population aging as a major issue in regards to pension sustainability finances as well as health and the elderly welfare system. Mortality projections that result from population structure changes and population aging are increasingly important. This paper compares six mortality models using KOSTAT's life table from 1970 to 2016. The models are rooted in the Lee-Carter (LC) model (Lee and Carter, *Journal of the American Statistical Association*, **87**, 659–671, 1992) and have been modified and improved on the assumptions of the LC model. We examined the improvement process and the check assumption by models in order to find a suitable mortality model for Korea. Korea shows rapid aging and declined mortality rate by age; therefore, it is desirable to estimate and predict mortality from LL&LC-ER models by combining LC-ER, LL, and LC-ER models that reflect the phenomena and modify age-specific mortality patterns without major changes in expected life expectancy.

Keywords: population aging, mortality model, Lee-Carter, LC-ER, LL&LC-ER

1. 서론

저출산과 사망수준의 개선으로 인구고령화가 진행되고 있고 이로 인하여 미래인구와 인구구조에 대한 사회와 정부의 관심이 증대되고 있다. 지속적으로 감소하고 있는 사망률과 기대수명의 증가는 인구의 연령구조 변화에 중요한 영향을 미친다. 따라서 고령층의 증가로 인구규모 및 인구고령화에 미치는 사망률 예측의 중요성은 더욱 조명받고 있다. 또한 사망율 개선이 인구고령화에 미치는 영향은 어떤 연령층에 집중되느냐에 따라 다르다. 젊은층에 집중될 경우 기대수명을 증가시키지만, 노인인구에 집중되면 인구고령화를 심화시킨다.

선진국에서는 고령화로 인하여 우려되는 연금과 건강 및 노인 간호 시스템의 지속 가능성에 집중하고 있다. 이로 인해 정부 정책 입안자들은 인구의 급속한 고령화와 관련된 인구통계학적 문제를 정확하게 진단하고 해결하기 위해 연령별 사망률을 정확히 모델링하고 예측하는 분야에 관심을 갖게 되었다.

연구는 통계청의 공식견해가 아니며 저자의 개인적인 연구결과임을 밝힙니다.
[1]Corresponding author: Statistical Analysis Division Statistical Research Institute, 6F, Statistical Center, 713 Hanbatdaero, Seo-gu, Daejeon 35220. E-mail: sykim0213@korea.kr

미래의 사망률에 대한 예측은 미래에 대한 불확실성을 내포하고 있으므로 장기 예측시 불확실성은 더욱 커지게 된다. 이러한 불확실성을 고려하고 처리하기 위해 확률적 사망률 모형(stochastic mortality model)에 대한 연구가 진행되면서 사망률 예측을 위한 다양한 모형들이 개발되어 왔다. 이는 과거자료의 추세가 미래에도 지속될 것이란 가정 하에 예측자의 주관적 견해를 고려하지 않고 자료에 기반을 두어 미래값을 예측한다.

확률적 사망률 모형에 관한 본격적인 연구는 1992년 미국의 인구통계학자 Lee와 Carter에 의해 시작되었다고 볼 수 있다. Lee와 Carter가 사망률 예측을 위해 개발한 모형은 구조가 단순하고 추정이 간단하다는 장점 때문에 지금까지도 널리 사용되고 있다.

이후 Lee와 Miller (2001), Booth 등 (2002), Li와 Lee (2005), 그리고 De Jong과 Tickle (2006) 등은 Lee-Carter 모형(이하 LC 모형)을 개선하기 위한 연구를 수행하였다. 이밖에도 사망률 예측을 위해 코호트 효과 (Cairns 등, 2006, 2009, 2011; Renshaw와 Haberman, 2006; Hunt와 Villegas, 2015)를 추가하거나 연령구간을 일부구간(고연령)으로 확장 (Kannisto 등, 1994; Thatcher 등, 1998)하는 경우 등 많은 모형들이 연구되어지고 있다.

최근에는 우리나라를 포함한 사망률이 낮은 선진국에서 과거 빠르게 감소했던 젊은층의 사망률은 점진적으로 줄어들고, 고령층 사망률 감소는 상대적으로 빠르게 진행되는 현상이 있다. 이러한 연령별 사망률 개선 추세의 변화를 반영할 수 있는 모형(LC-ER)에 대한 연구도 있다.

이밖에도 이탈리아, 일본, 그리고 덴마크의 생애(lifespan) 평균과 차이(disparity)간의 관계를 분석함으로써 보다 더 심도 있게 사망률 모형의 예측력을 평가한 연구 (Bohk 등, 2017), 사망률 개선에 초점을 맞춰 변수영역 기억모형(random field memory model)에 AR-ARCH 모형을 적용한 연구 (Doukhan 등, 2017), 인접 연령간의 단기 및 장기 사망률의 공간 연관성을 고려하기 위해 vector auto regression (VAR) 프로세스를 활용한 연구 (Li와 Lu, 2017), 그리고 데이터부족을 합리적으로 고려하고 모형의 적합도를 개선하여 예측력을 향상시키기 위해 Elastic-Net (Zou와 Hastie, 2005) 기반의 벌점 VAR (penalized VAR) 모델을 제안한 연구 (Guibert 등, 2017) 등 많은 모형 선행연구들이 있다.

우리나라 사망률 예측에 관한 연구로는 LC 모형에 관한 연구 (Kang 등, 2006), 사망률 모형의 비교에 관한 연구 (Park 등, 2005; Jeong과 Kim, 2011), 그리고 시계열 적용기간에 따른 LC 모형과 LC 코호트효과 확장 모형의 비교 연구 (Jung 등, 2013), 모수와 비모수모형을 활용한 사망률 예측 비교 연구 (Kim과 Oh, 2017), 시계열 이용기간에 따른 사망률 예측 비교 연구 (Kim 등, 2018) 등이 있다.

본 논문은 다음과 같은 점에서 선행연구들과 차별된다.

첫째, 사망률 예측을 위한 대표적인 모형인 LC 모형을 시작으로 이 모형을 개선한 Lee-Miller (LM) 모형 (Lee와 Miller, 2001), Booth-Maindonald-Smith (BMS) 모형 (Booth 등, 2002), Li-Lee (LL) 모형 (Li와 Lee, 2005), 그리고 최근 주목받고 있는 사망률 개선 교대(declined mortality rate with rotation)현상을 반영한 LC-ER 모형(extended of the LC with rotation) (Li 등, 2013)과 LC-ER 모형에 인구 부집단(sub-population)의 정합성 확보를 위해 Li-Lee (2005) 방식을 추가한 LL과 LC-ER의 조합모형(combined LL and LC-ER; LL&LC-ER) (Statistics Canada, 2015)까지의 변천을 살펴보고자 한다. 이중 일부 모델들은 우리나라에서 공식통계를 작성하게 위해 사용한 사망률 예측 모형들이며, 본 연구는 LC 모형 계열에 대한 종합적인 검토를 한 연구로서의 의미가 있다고 할 수 있다.

둘째, 일본, 중국, 한국과 같은 빠른 고령화를 보이는 나라들은 과거의 사망률이 미래에도 지속되고 연령별 사망률이 일정하다는 가정이 더 이상 합리적이지 못한 경우가 많다. 따라서 위에서 소개된 모형에서 사망률을 예측하기 위해 사용되는 각각의 가정들을 검토하고, 빠른 사망률 개선과 고령화로 인한 모형 개선 사항을 소개한다. 특히 일본, 중국, 한국과 같은 빠른 고령화를 보이는 나라들은 과거의 사망률

이 미래에도 지속되고 연령별 사망률이 일정하다는 가정이 더 이상 합리적이지 못한 경우가 많다.

그리고 이러한 차별 연구를 위해 1970–2016년 통계청 KOSIS의 완전생명표(2016) 자료에서 각세별 사망확률 자료를 활용하였음을 밝힌다. 사망확률에서 사망률로의 전환은 사망률 = 사망확률/(1 − 0.5 × 사망확률) 공식을 활용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 LC 계열(LC, LM, BMS, LL, LC-ER, LL&LC-ER)의 사망률 예측모형을 소개하고 이들의 장단점을 살펴본다. 3장에서는 사망률 개선 추세가 변화하고 있는 우리나라 사망률 자료의 특징을 논의하고 LC 계열의 6가지 모형에 적용한 결과를 비교한다. 이 결과들을 통해 빠른 고령화를 보이는 우리나라에 적합한 사망률 모형이 무엇인가를 제안하고자 한다. 마지막으로 4장에서는 연구결과를 정리하였다.

2. 사망률 예측모형

본 절에서는 사망률 예측을 위한 대표적인 모형인 LC 모형과 이를 개선한 LM 모형, BMS 모형, LL 모형을 소개한다. 또한 최근 주목받고 있는 사망률 개선 교대현상을 반영한 LC-ER 모형과 LC-ER모형에 인구 부집단의 정합성 확보를 위해 LL (Li와 Lee, 2005) 방식을 조합한 모형을 살펴보겠다.

2.1. Lee-Carter 모형

LC 모형은 1992년 인구통계학자인 Lee와 Cater에 의해 제안된 미래 사망률을 예측하는 대표적인 모형으로 로그 변환된 사망률을 연령효과(age effect)와 시간효과(period effect)의 선형으로 표현한다. 로그 연령별 사망률(age-specific mortality) 예측모델의 구조는 식 (2.1)과 같다.

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \quad t = 1, \ldots, n. \tag{2.1}$$

여기서, $m_{x,t}$는 연령 x와 시간 t에서의 사망률(death rate)을 나타내고, a_x는 연령에 따른 평균적인 로그사망률의 수준, b_x는 각 연령에서 로그사망률의 변화, k_t는 시간 t에 따른 로그사망률 변화 수준을 나타내고, $\epsilon_{x,t}$는 평균 0, 분산 σ_ϵ^2인 관측되지 않은 오차항을 의미한다.

연령에 따른 평균적인 로그사망률의 수준(a_x)과 각 연령에서 로그사망률의 변화(b_x)는 시간에 무관한 상수이고, 시간에 따른 전반적인 사망률 개선정도를 타나내는 사망률 지수(mortality index)인 k_t만이 시간에 의존하므로 추정된 k_t의 예측을 통해 미래 사망률을 예측하게 된다.

모수 a_x는 시간에 대한 $\ln(m_{x,t})$의 평균으로 계산되어지고, b_x와 k_t는 $[\ln(m_{x,t}) - a_x]$에 비정칙분해(singular value decomposition; SVD)방법을 적용하여 얻어진 첫 번째 주성분을 사용하여 추정되어진다. 이때 b_x와 k_t의 유일한 추정치를 보장하기 위해 $\sum_t k_t = 0$, $\sum_x b_x = 1$와 같은 제약조건이 필요하다. 또한, 추정된 $\hat{k}_t$는 추정된 사망자수가 실제 사망자수와 일치하도록 총사망자수 $D_t = \sum_x D_{x,t}$의 기준에서 $\hat{k}_t$를 조정한다. 여기서, $D_{x,t}$는 t연도, x연령의 사망자수이다.

Figure 2.1은 LC 모형에서 각각의 모수들의 연령과 시간(1970–2016)에 따른 추이를 살펴본 것이다. 첫 번째 그림은 연령에 따른 로그사망률 패턴인 a_x, 두 번째는 각 연령에서 로그사망률이 변화하는 정도인 b_x, 세 번째는 시간 t에 따른 로그사망률 수준의 변화를 나타내는 k_t이다. a_x는 J자형 커브를 나타내고, b_x는 절대값이 크면 연령 x의 사망률의 변화가 크다는 것을 의미한다. k_t는 시간에 따라 선형으로 감소하는 패턴으로 이는 시간에 따라 사망률이 개선됨을 의미한다.

세 모수 중 a_x와 b_x는 자료로부터 추정되고, k_t만이 시간에 의존하는 모수이므로 LC 모형에서의 관건은 k_t를 얼마나 정확하게 추정하고 예측하느냐의 문제로 귀결된다. 추정된 $\hat{k}_t$의 예측은 기본적으로 식

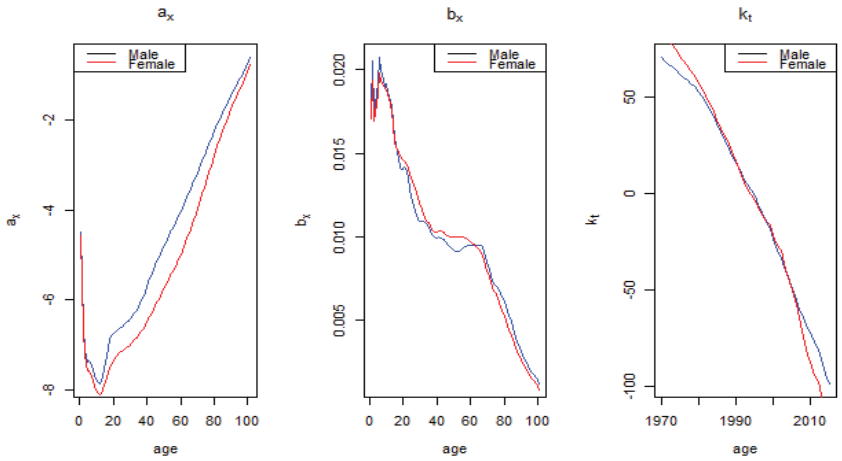


Figure 2.1. Trends of parameters of Lee-Carter model.

(2.2)와 같은 절편을 갖는 임의보행모형(random walk with drift; RWD)을 사용하지만 다른 ARIMA 모형의 적합도 가능하다.

$$k_t = k_{t-1} + d + e_t. \tag{2.2}$$

여기서 d는 임의보행과정의 추세를 나타내는 절편모수로 k_t의 평균 연간 변화를 나타내고, e_t는 평균이 0이고 분산이 σ_ϵ^2인 정규분포를 따르는 확률변수이다.

예측한 $\hat{k}_{n+j}$ $(j = 1, 2, \ldots)$와 추정된 $\hat{a}_x$와 $\hat{b}_x$를 식 (2.1)에 대입하면 미래의 로그사망률을 예측할 수 있다. 즉, $k_{n+h|n}$의 h-단계 예측(h-step-ahead forecasts)을 $\hat{k}_{n+h|n}$로 표기한다면 $\ln(m_{x,n+h})$의 h-단계 예측은 식 (2.3)과 같다.

$$\ln(m_{x,n+h}) = \hat{a}_x + \hat{b}_x \hat{k}_{n+h|n}. \tag{2.3}$$

그리고 첫 번째 주성분(b_x)과 식 (2.1)에서의 오차항 사이의 직교성으로 인해 전반적으로 예측분산은 아래 식 (2.4)와 같이 도출할 수 있다.

$$\mathrm{Var}[\ln(m_{x,t})] \approx \hat{b}_x^2 u_{n+h|n} + v_x. \tag{2.4}$$

여기서, $\hat{b}_x^2$는 첫 번째 주성분의 분산이고, $u_{n+h|n} = \mathrm{Var}(k_{n+h}|k_1, \ldots, k_n)$는 시계열 모형으로부터 얻을 수 있다. 그리고 모형의 잔차 분산인 v_x는 식 (2.1)에서 각각의 x에 대하여 제곱화된 잔차 $\{\epsilon_{x,1}^2, \ldots, \epsilon_{x,n}^2\}$의 평균에 의해 추정된다. 이를 이용하여 미래의 불확실성에 대한 예측값과 신뢰구간을 제시할 수 있다.

LC 모형은 구조가 간단하고 사망률 예측이 용이하며, 예측면에서도 비교적 우수하다고 알려져 있어 사망률 연구에 가장 기본적인 모형으로 현재까지 널리 사용되고 있다. 하지만 실제 적용상에서 여러 가지 문제점들이 발견되었으며 이를 해결하기 위해 다음절에 소개될 여러 개선 모형들이 제안되었다. 이들 모형은 LC 모형의 기본구조를 유지하면서 최소한의 변형을 가한 것으로 주된 초점은 연령별 사망률 b_x 개선과 미래 k_t를 얼마나 잘 산출해 낼 것인가이다.

2.2. Lee-Miller 모형

LM 모형 (Lee와 Miller, 2001)은 LC 모형의 변형으로 다음 3가지 면에서 LC 모형을 수정하였다.

첫째, LC 모형은 시간의 경과와 관계없이 연령별 사망률 변화(b_x)가 일정하다 가정하였다. 그러나 Lee와 Miller (2001)는 미국, 스웨덴, 캐나다, 프랑스 등의 지난 20세기 전반기(1900–1950년)과 후반기(1950–1995년)의 사망률 감소 패턴을 비교한 결과 전염병과 2차 세계대전 등으로 인하여 20세기 사망률 패턴의 구조적 변화가 발생하여 1950년을 기준으로 연령별 사망률의 변화가 일정하지 않음을 발견하였다. LM 모형은 변화하는 연령별 사망률의 패턴(b_x)의 문제를 해결하고 LC 모형에서 사용한 가정의 적합성을 높이기 위해 모형 적합시 1950년 이후의 자료를 사용하였다.

둘째, LC 모형은 추정된 k_t를 총사망자수를 이용하여 조정하였으나, 이는 연도별 사망자수(D_t)와 같은 추가 인구데이터가 필요하다. 이러한 인구데이터의 사용을 피하기 위해 t년도의 기대수명(e_0)을 이용하여 추정된 k_t를 조정하였다.

셋째, Lee와 Miller (2001)는 적합된 기간의 마지막 연도에 적합된 값과 실제값의 불일치(jump-bias, 점프 편의)로 인하여 새로운 추계를 시작하는 연도의 자료에서 편의(jump-off bias)가 발생한다. 이로 인하여 남성과 여성의 기대수명에 있어 0.6년의 차이가 발생함을 발견했다 (Lee와 Miller, 2001). LM 모형은 새로운 추계를 시작하는 연도 자료에서 발생하는 편의를 줄이고 그 결과 기대수명(e_0)의 예측에 차이가 발생하는 것을 극복하기 위해 적합된 기간의 마지막 연도에 적합된 값 대신 실제값을 사용하였다.

2.3. Booth-Maindonald-Smith 모형

BMS 모형 (Booth 등, 2002)은 다음과 같은 3가지 면에서 LC 모형을 변형한 모형이다.

첫째, Booth 등 (2002)은 LC 모형처럼 추정된 k_t를 총사망자수(D_t)를 이용하여 조정할 경우 연령별 사망자수 분포에 왜곡이 발생할 수 있음을 지적하고, k_t를 보다 정교하게 조정하기 위해 총사망자수(D_t)나 기대수명(e_0) 대신 연령별 사망자수($D_{x,t}$)를 이용하였다.

$$\ln(D_{x,t}) = \ln(N_{x,t}) + \ln(m'_{x,t}) + \epsilon'_{x,t}. \tag{2.5}$$

이는 연령별 사망자수($D_{x,t}$)에 식 (2.5)와 같은 포아송 모형(Poisson regression model)을 적용하여 사망을 모델화하는 사망자의 연령분포에 맞춘다. 여기서 $\ln(m'_{x,t})$는 $a_x + b_x k'_t$이며, k'_t는 재조정된 k_t를 의미하며 $\epsilon'_{x,t}$는 k_t 조정 후의 잔차이다 (Brillinger, 1986; Booth 등, 2002).

둘째, 시간에 따른 로그사망률의 수준 변화(k_t)가 선형이라는 가정하에서 통계적 적합도(goodness-of-fit)기준을 이용하여 모형적합에 사용하는 자료이용기간을 결정하였다. LC 모형은 시간에 따른 사망률 수준의 변화(k_t)가 선형이고, 각 연령에서의 로그사망률의 변화(b_x)가 일정하다고 가정한다. 호주의 사망률 데이터가 이 두 가정으로부터 유의미한 이탈을 보임에 따라, 선형 k_t의 가정하에서 모형적합을 위해 사용하는 자료의 이용기간을 연령별 사망자수 추정값과 실제값 간의 차이를 최소화하는 기준으로 아래 식 (2.6)의 적합도 통계를 이용하여 결정하도록 하였다 (Booth 등, 2002).

$$\text{deviance}_t = 2\sum_x \left\{ D_{x,t} \times \ln\left(\frac{D_{x,t}}{D'_{x,t}}\right) - (D_{x,t} - D'_{x,t}) \right\}. \tag{2.6}$$

여기서 $D_{x,t}$는 연령별 사망자수, $D'_{x,t}$는 적합된 사망자수이고, $D_{x,t} = N_{x,t} \times \exp(a_x + b_x k'_t)$로 구해진다 ($N_{x,t}$는 인구수, k'_t는 재조정된 k_t).

셋째, 적합된 기간의 마지막 연도에 실제값과 적합값의 불일치로 인하여 새로운 추계를 시작하는 연도의 자료에서 발생하는 편의와 기대수명(e_0) 예측의 차이는 연령별 사망자수 분포에 적합하도록 조정된

k_t'를 이용한 적합된 값을 이용함으로써 해결하였다.

2.4. Li-Lee 모형

LC 모형은 앞절에서 살펴본 바와 같이 많은 선형변형이 이루어졌으며 단일인구내에서 사망률의 적합과 예측에 대표적으로 널리 사용되는 모형들의 근간이 되어 왔다.

그런데 인구집단을 어떤 특성(성(sex) 또는 지역(region))을 기준으로 나누어 몇 개의 그룹(다중인구 또는 인구부집단으로 정의)으로 분석하고자 할 때, 사회 경제적인 조건이 유사한 인구집단의 사망률을 고려한다면 이들 집단간 사망률 패턴에 연관성이 존재한다.

이와 같은 다중인구에 대해 LC 모형을 적용할 경우 다중인구들간의 사망률의 변화(k_t)값들이 서로 달라 이들 집단 간에 존재하는 연관성을 고려할 수 없어, 그 결과 집단간 격차가 발생해 장기 예측에서 합리적인 결과를 도출할 가능성이 높지 않다.

예를 들면 동일 국가 내에서는 성별 격차가 장기적으로 발산할 가능성은 높지 않으나, 남자와 여자의 사망률을 LC 모형을 이용하여 개별적으로 예측하게 되면 성별 사망률 격차가 발산하는 추세가 나타날 수 있다. 따라서 인구학자들은 인구사회학적으로 밀접히 연관된 하위 집단의 사망률 격차가 장기적으로 확대되지 않도록 내적 일관성(coherent)이 충족될 수 있는 사망률 예측모형이 필요하다고 주장하였다.

Li와 Lee (2005)는 사회 경제적인 조건이 유사한 인구집단의 사망률 패턴은 장기적으로 공통 사망률로 수렴할 것이라 가정하에, 기존의 LC 모형을 다중인구로 확장하기 위해 공통사망경향에 대한 시간×연령 효과를 추가하여 식 (2.1)을 식 (2.7)과 같은 사망률 예측모형으로 제시하였다.

$$\ln(m_{x,t,i}) = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + \epsilon_{x,t,i}. \tag{2.7}$$

여기서, $m_{x,t,i}$는 i번째 그룹의 시간 t에서 연령 x의 사망률을 나타내고, LC 모형과 비교할 때, 가장 큰 차이점은 우변의 $B_x K_t$항이다. 이는 모든 그룹의 공통사망경향으로 전체 집단의 연령별 로그사망률 변화정도(B_x)와 시간에 따른 로그사망률 수준의 변화(K_t)를 나타내며 하위집단의 사망률이 장기적으로 발산하지 않도록 하는 역할을 한다. $b_{x,i}k_{t,i}$는 i번째 그룹의 개별사망경향으로 개별 집단 i의 연령별 로그사망률 변화정도($b_{x,i}$)와 시간에 따른 로그사망률 수준 변화($k_{t,i}$)를 나타내며 하위 집단들이 공통추세에서 벗어난 단기적 변동을 설명하는 역할을 한다. 그리고 $a_{x,i}$는 그룹 i의 연령에 따른 평균적인 로그사망률의 수준을 나타내다. 여기서 그룹은 인접국가(지역) 또는 동일국가 내 성별 등이 될 수 있다.

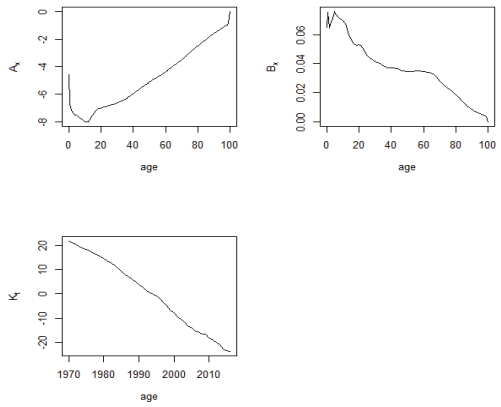
이런 의미에서 LL 모형은 LC 모형의 확장이라 할 수 있고, 공통 사망률 패턴을 고려한 사망률 추계에 활용되고 있다.

LL 모형은 LC 모형의 모수추정 과정을 두 번 적용한다. 먼저 그룹을 구분하지 않고 전체집단을 대상으로 식 (2.8)과 같이 LC 모형을 적용하여 SVD 방법에 의하여 공통사망경향 $B_x K_t$를 추출한다. 여기서, A_x는 공통사망경향에 대한 평균사망률을 의미한다.

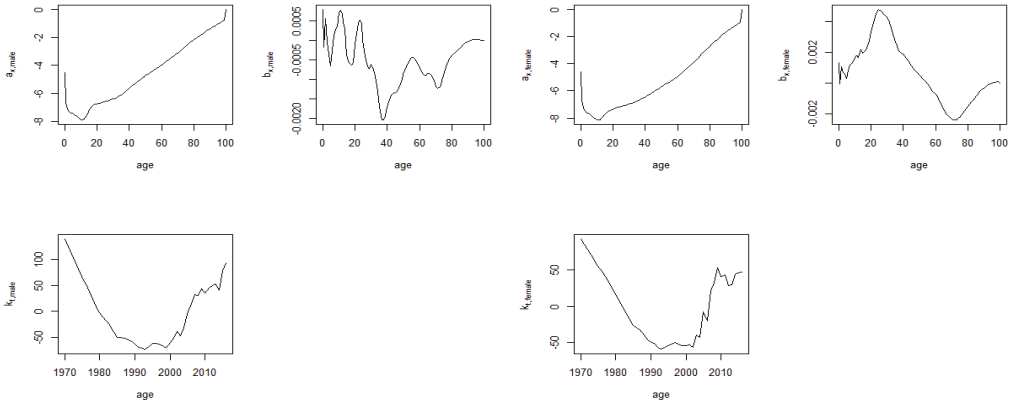
$$\ln(m_{x,t}) = A_x + B_x K_t + \epsilon_{x,t}. \tag{2.8}$$

다음으로 앞단계에서 추출한 공통사망경향 추정치 $\hat{B}_x\hat{K}_t$에 기초하여 각각의 인구부집단으로부터 도출된 잔차 $[\ln(m_{x,t,i}) - a_{x,i} - \hat{B}_x\hat{K}_t]$에 다시 SVD 방법을 적용하여 개별 그룹의 사망경향 $b_{x,i}k_{t,i}$를 추출한다. 또한 추정된 k_t와 K_t의 예측은 LC 모형과 같이 기본적으로 RWD을 사용하지만 다른 ARIMA 모형적합도 가능하다.

Figure 2.2는 인구집단을 성별그룹으로 구분하여 LL 모형을 적용했을 때 각 모수값들의 연령과 시간에 따른 추이를 살펴본 것으로 시간 t에 의존하는 모수는 K_t와 $k_{t,i}$이다. 공통사망경향에 대한 평균사



(a) Coherent mortality trend



(b) Mortality trend by male (c) Mortality trend by female

Figure 2.2. Trends of parameters of Li-Lee model. 자료 : KOSIS 생명표(1970–2016년 사망률 자료).

망률(A_x)과 B_x, K_t의 추이를 연령과 시간에 대해서 나타낸 것이다. 개별사망경향에 대한 남자의 평균적인 로그사망률($a_{x,\text{male}}$)과 $b_{x,\text{male}}$, $k_{t,\text{male}}$를 연령과 시간에 대해서 나타내고, 여자의 대한 평균적인 로그사망률($a_{x,\text{female}}$)과 $b_{x,\text{female}}$, $k_{t,\text{female}}$를 연령과 시간에 대해서 나타낸 것이다. LC 모형의 Figure 2.1과 매우 유사함을 확인할 수 있다.

2.5. LC-ER 모형

최근 우리나라를 포함한 사망률이 낮은 선진국에서 유년층의 사망률 개선은 점점 감소하고 노년층의 사망률 개선은 점점 증가하는 현상이 발생하고 있다 (Horiuchi와 Wilmoth, 1995; Kannisto 등, 1994; Li와 Gerland, 2011; Li 등, 2013; Kim과 Oh, 2017; Kim과 Han, 2017).

이것은 젊은층과 노년층의 사망률 감소 패턴이 교대(순환, rotation)되는 현상으로 기대수명이 증가함에 따라 발생하고 있으며 LC 모형에서 연령별 사망률 변화를 나타내는 b_x가 시간의 경과에 따라 변하고 있음을 의미한다.

Li와 Gerland (2011)은 이와 같은 현상을 사망률 개선 교대이라 명명하고 시간의 변화에 따라 연령별 사망률 변화 패턴이 변하는 사망률 개선 교대 현상을 반영하기 위해 LC 모형의 b_x에 강건한 순환(robust rotation)을 도입하여 주관적이고 강건한 수정을 할 수 있는 Lee-Carter method with robust rotation (LC-RR) 모형을 제안하였다.

그리고 2년 뒤에 Li 등 (2013)은 LC-RR 모형의 회전모델을 타당하고 유연하게 변화를 줄 수 있도록 아래 식 (2.9)와 같은 LC-ER 모형을 제안하였다. 즉, 이는 LC 모형에서 시간의 변화에 관계없이 일정한 b_x에 시간 변수를 고려한 $B_{x,t}$형태로 변형한 모형이다. 또한 식 (2.9)를 성별에 따라 확장하면 식 (2.10)과 같다.

$$\ln(m_{x,t}) = a_x + B_{x,t}K_t + \epsilon_{x,t}. \tag{2.9}$$

$$\begin{aligned} &\text{남자}: \ \ln(m_{x,t,m}) = a_{x,m} + B_{x,t,m}K_{t,m} + \epsilon_{x,t,m}, \\ &\text{여자}: \ \ln(m_{x,t,f}) = a_{x,f} + B_{x,t,f}K_{t,f} + \epsilon_{x,t,f}. \end{aligned} \tag{2.10}$$

여기서, $B_{x,t}$는 t시점에서 연령 x세의 사망률 변화(패턴)으로 사망률 교대 현상이 반영되도록 아래 식 (2.11)과 같이 정의하였다. 이는 LC 모형 본래의 b_x와 사망률 개선속도가 더 이상 변하지 않고 일정하게 유지되는 최종 시점(u)의 b_x^u의 가중선형 결합으로 표현된다.

가중치 $w_x(t)$는 시간에 대한 함수로 시간이 지남에 따라 1에 가까워지고 그 결과 전체적인 $B_{x,t}$도 시간이 지남에 따라 점점 b_x^u에 가까워진다. e_0^u는 사망률 개선속도가 더 이상 변하지 않는 시점(즉, 교대현상이 끝나는 시점)의 기대수명을 나타낸다.

$$B_{x,t} = \begin{cases} b_x, & e_0^t < 80, \\ (1 - w_s(t))b_x + w_s(t)b_x^u, & 80 \leqq e_0^t < e_0^u, \\ b_x^u, & e_0^t > e_0^u, \end{cases} \tag{2.11}$$

$$b_x^u = \begin{cases} \bar{b}_{15-64}, & 0 \leqq x \leqq 64, \\ b_x \times \dfrac{b_{u,60-64}}{b_{65-70}}, & 65 \leqq x, \end{cases} \qquad w(t) = \frac{e_0^t - 80}{e_0^u - 80}, \quad w_s(t) = \left[0.5 \times \left\{1 + \sin\left(\frac{\pi}{2}(2w(t) - 1)\right)\right\}\right]^p.$$

Li 등 (2013)은 타당한 회전 모델을 제안하기 위해 $B_{x,t}$를 산출할때 선진 20개국의 사망률 자료 분석에 근거한 세 가지 가정을 활용하였다.

첫째, 연령별 사망률 개선 패턴은 기대수명이 80세에 도달했을 때 교대하기 시작해서 102세에 도달할 때까지 지속되며, 그 이후에는 일정한 것으로 가정한다. 따라서 기대수명이 80세보다 작은 경우는 LC 모형과 동일하고, 80세 이상이면 사망률 개선은 빨라져 $B_{x,t}$는 연령별 사망률이 일정하다고 가정한 LC 모형의 b_x에서 점차 벗어나게 될 것이다.

둘째, 연령별 사망률 개선 속도(패턴)가 더 이상 변하지 않고 일정하게 유지되는 최종시점(u로 표기)의 b_x^u는 사망률 개선 패턴이 65세 미만의 모든 연령에서 동일($\bar{b}_{15-64}$)하고, 65세 이상 연령에 대해서는 증가할수록 감소한다고 가정한다. 이는 선진 20개국 국가의 과거 사망률 자료의 경험적 분석 결과로 65세까지는 모든 연령대에서 사망률 감소 속도가 각 국가마다 정도는 다르지만 궁극적으로 하나의 일반적인 수준으로 수렴한다는 것을 의미한다.

셋째, 식 (2.11)의 가중함수 $w_s(t)$의 80이다. 가중함수 $w_s(t)$의 계산은 두 단계를 거친다. 먼저 t시점에서 기대수명 e_0^t에 근거한 선형 가중치 함수(w_t)는 선진 20개국 인구의 기대수명의 대략적인 중앙값인 80세를 사용하여 $w(t) = (e_0^t - 80)/(e_0^u - 80)$와 같이 제안된다. 여기서 e_0^u는 앞에서 언급하였듯이 사망률 개선 속도가 더 이상 변하지 않는 시점의 최종 기대수명이므로 교대는 $e_0^t = 80$일 때 $w_t = 0$에

서 시작하며 $e_0^t = e_0^u$일 때 1까지 변한다. 다음으로 이렇게 정의된 가중함수를 보다 매끄럽게 하기 위해 w_s^t를 삼각함수로 정의한다. 여기서 p는 교대의 속도를 통제하는 모수$(0 < p < 1)$이고 디폴트로 0.5를 사용한다.

사망확률에서 중요한 K_t의 예측은 LC 모형을 사용하여 초기 기대수명을 예측한 후, 식 (2.11)의 $B_{x,t}$를 사용하여 기대수명에 맞게 조정된 K_t를 찾는 방법을 사용한다. 즉, LC-ER의 $B_{x,t} \times K_t$와 LC의 $b_x \times k_t$가 동일해지도록 K_t를 수정하되 기대수명이 같아지도록 시뮬레이션을 통해 반복적인 계산을 통해 도출된다. 따라서 기대수명의 예측치가 변경되는 것이 아니라 예측치를 생성하는 사망률의 연령 패턴을 변경하게 되는 것이다. 이는 b_x를 $B_{x,t}$로 변환함에 따라 기대수명 등의 사망률 특성이 왜곡되지 않고 연령별 미묘한 사망률 변화 패턴의 변화를 반영하기 위함이다. 이런 과정을 거쳐 b_x^u, e_0^u 그리고 $B_{x,t}$를 도출한 후 K_t값으로 e_0^t를 예측한다.

2.6. LL과 LC-ER 조합모형(LL&LC-ER 모형)

앞절에서 살펴본 바와 같이 LC-ER 모형은 시간 흐름에 따라 연령별 사망률 변화 패턴이 바뀌는 사망률 개선 교대 현상을 반영한 것이다.

하지만 LC-ER 모형은 연령별 사망률 개선과 단일인구의 사망률 적합 및 예측을 위한 LC 모형의 확장으로 사회 경제적 조건이 유사한 인구 부집단의 사망률을 고려할 때는 집단간에 존재하는 연관성을 고려할 수가 없는 제한점이 있기 때문에 집단간 격차가 발생할 가능성이 높다. 또한 LL 모형은 다중인구 그룹의 집단간 내적 정합성을 유지하여 일괄성 있는 사망률을 예측할 수는 있으나 시간 흐름에 따른 사망률 개선효과를 반영하지 못하므로 장기적으로 예측할 경우에는 한계가 있다.

따라서 다중인구의 경우 내적 정합성을 확보하면서 동시에 사망률 개선 교대 현상을 반영할 수 있는 모형을 고려할 수 있는데 본 연구에서는 LC-ER의 개념에 LL 모형을 적용할 수 있다는 의미에서 LL과 LC-ER의 조합(LL&LC-ER)모형으로 명명한다.

이는 캐나다 통계청 (Statistics Canada, 2015)과 우리나라 장래인구추계 (KOSIS, 2016)에서 적용한 방법으로 아래 식 (2.12)와 같으며 식 (2.7)에서 연령별 사망률 공통변화패턴 모수 B_x를 시간에 따라 변동하는 $B_{x,t}$로 수정한 것이다. 여기서 $m_{x,t,i}$는 i번째 그룹의 시간 t에서 연령 x의 사망률을 나타내고, $B_{x,t}$는 공통인구의 t시점에서 x연령의 사망률 변화 정도(패턴)을 나타내며 하위집단의 사망률이 장기적으로 발산하지 않도록하는 역할을 한다.

$$\ln(m_{x,t,i}) = a_{x,i} + B_{x,t}K_t^* + b_{x,i}k_{t,i} + \epsilon_{x,t,i}. \tag{2.12}$$

이는 공통인구에 LC-ER을 적용하여 $B_{x,t}K_t^*$를 산출한 후 각 하위 인구집단별로 LL 모형을 적용하여 $b_{x,i}k_{t,i}$을 산출하게 된다. 이때, $B_{x,t}K_t^*$이 LL 모형의 B_xK_t와 동일해지도록 K_t^*를 찾는다.

지금까지 논의한 LC 류의 사망률 예측 모형의 관계와 변천을 정리하면 Figure 2.3과 같고 이들의 특징을 간명하게 비교하면 Table 2.1과 같다.

3. 사망률 모형 가정 검토

본 절에서는 사망률 개선 속도가 변화하고 있는 우리나라 데이터의 특징을 살펴보고 LC 류 모형에서 사용하는 가정을 검토한다. 특히 이번 절에서는 LC-ER과 LL과 LC-ER의 조합모형에 초점을 맞춰 살펴보기로한다. 그리고 우리나라 사망률 개선 추세 변화 패턴을 탐색하며 LC 계열의 대표적인 6가지 모형(LC, LM, BMS, LL, LC-ER, LL&LC-ER)을 활용하여 우리나라 기대수명을 산출하고 결과를 비교한다.

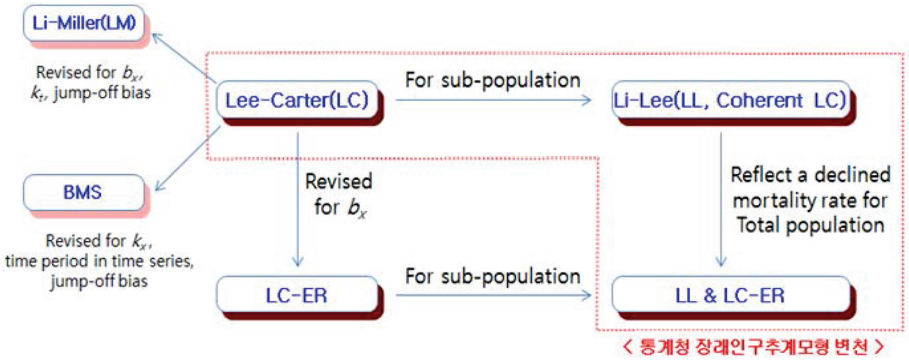


Figure 2.3. Relation and transition of mortality models on the kinds of Lee-Carter model. BMS = Booth-Maindonald-Smith; LC-ER = extended of the LC with rotation; LL&LC-ER = combined LL and LC-ER.

Table 2.1. Comparison of mortality models on the kinds of LC

Method	Data period	Adjustment to match	Jump-off bias	Rotation	Sub population	Reference
LC	All	D_t	fitted	without	individual	Lee and Carter (1992)
LM	1950	$e(0)$	observed	without	individual	Lee and Miller (2001)
BMS	Linear	$D_{x,t}$	fitted	without	individual	Booth *et al.* (2002)
LL	All	D_t	fitted	without	coherent	Li and Lee (2005)
LC-ER	All	D_t	observed	with	individual	Li *et al.* (2013)
LL & LC-ER	All	D_t	observed	with	coherent	Canada Statistics (2015)

주: data period는 각 논문에서 고려한 데이터의 시계열을 의미.
LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; LL = coherent LC; LC-ER = extended of the LC with rotation; LL&LC-ER = combined LL and LC-ER.

3.1. LC-ER 및 LL&LC-ER 모형의 가정 검토

Li 등 (2013)은 선진 20개국의 과거 사망률의 경험적 분석에 근거하여 아래와 같은 3가지(80세, 102세, 65세 미만)의 실험인 결과값을 가정하였다. 이는 식 (2.11)의 $B_{x,t}$와 같이 연령별 사망률 개선 패턴은 기대수명이 80세에 도달했을 때 교대하기 시작해서 102세에 근접할 때까지 지속되며, 그 이후에는 교대현상이 중단되어 연령별 사망률 개선 패턴이 변하지 않고 일정(b_x^u)하다는 것이다. 이때, 교대현상이 멈춰 연령별 사망률 개선 속도가 더 이상 변하지 않고 일정하게 유지되는 최종시점(u)의 b_x^u는 65세 미만 모든 연령에서 사망률 개선패턴은 동일($\bar{b}_{15-64}$)하며 그 이후에 연령이 증가할수록 작아짐을 가정하였다.

Li 등 (2013)은 예전부터 높은 사망률을 보이는 0세의 사망률($m(0)$)과 사망률이 가장 낮은 계층의 사망률($m(15–19)$)에 초점을 맞추어 Figure 3.1과 같이 선진 20개국 $m(0)/m(15–19)$의 비율이 과거로부터 계속 감소했으며 앞으로도 감소할 것으로 예상하였다. Figure 3.1에 의하면 $m(0)/m(15–19)$의 비율은 기대수명이 80세 전후부터 수렴하기 시작함을 알 수 있다. 즉, 이 80세는 선진 20개국가의 기대수명의 대략적인 중앙값이다. 또한 보편적인 가치로 교대현상 멈추는 시점을 기대수명이 102세가 되는 시점으로 선택하였다.

또한 선진 20개국 인구의 과거 사망률의 경험적 분석에 근거하여 65세까지 모든 연령대에서 사망률 감소 속도가 각 인구마다 정도는 다르지만 궁극적으로 하나의 일반적 수준으로 수렴함을 발견하였다.

위의 3가지 가정은 우리와 사회 경제적 환경이 다른 선진 20개국의 경험적 사망률 분석 결과에 의한 것

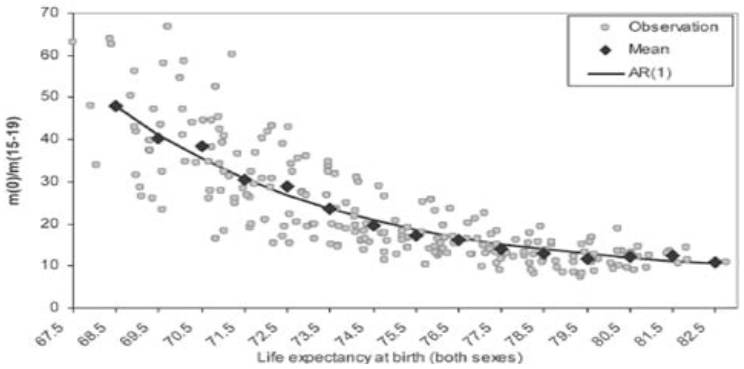
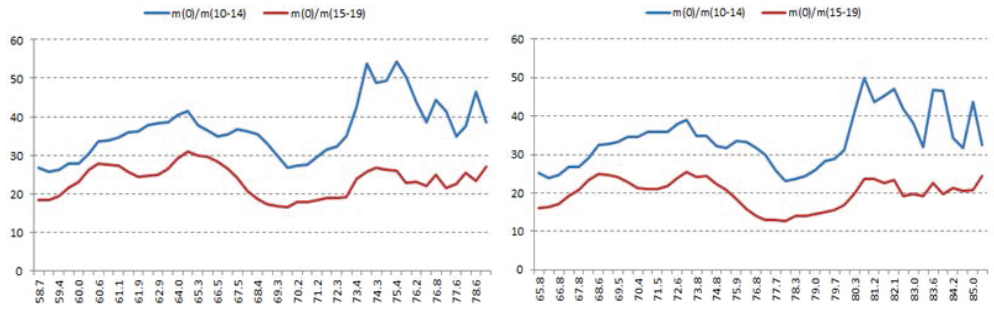


Figure 3.1. Trend of proportion on $m(0)/m(15–19)$ of the 20 developed countries. Li *et al.* (2013).



(a) Male (b) Female

주: 우리나라에서 가장 작은 사망률을 보이는 계층은 15–19세가 아니고 10–14세 계층이다.

Figure 3.2. Trend of proportion on $m(0)/m(15–19)$ and $m(0)/m(10–14)$ for Korea.

이다. 우리나라의 경우 사망률 개선이 다른 나라에 비해 유래없이 빠른 속도로 개선되어 왔으며 사망률 개선 패턴 또한 상이하다. 따라서 이들 가정에 대한 검토가 필요하다.

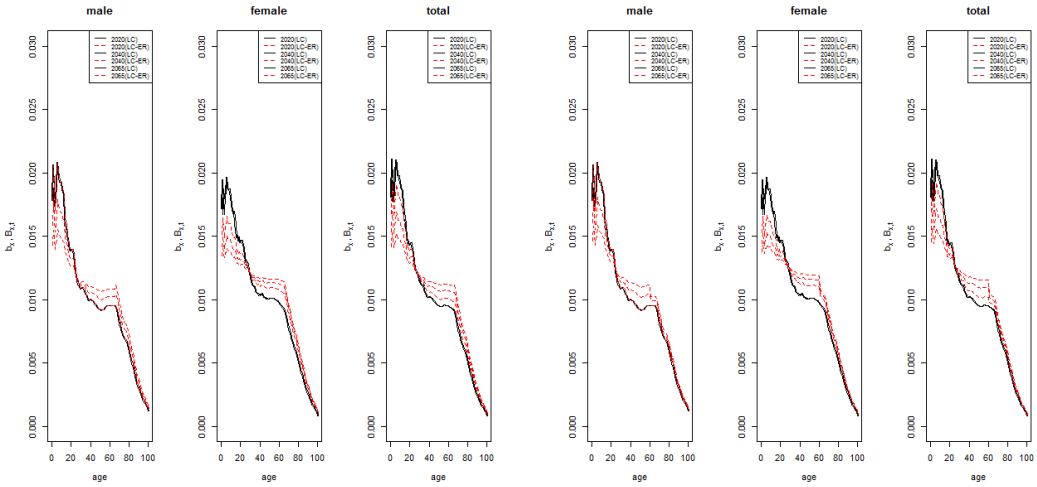
먼저 우리나라의 $m(0)/m(15–19)$ 비율을 살펴보면 Figure 3.2와 같다. 우리나라의 경우 기대수명 80세 이후 $m(0)/m(15–19)$의 비율은 Figure 3.1과 상이한 패턴이다. 우리나라 사망률 교대현상이 나타나는 시점(e_0^t의 t)을 Li 등의 가정과 동일하게 80세로 간주하면 남자는 2019년, 여자는 2001년, 전체는 2009년이다.

Figure 3.3의 첫 번째 그림은 Li 등이 가정한 것처럼 우리나라의 사망률 교대현상이 나타나는 시점을 80세로 간주하고 b_x^u의 경우 65세 미만 모든 연령에서 사망률 개선패턴이 동일하다 가정했을 때, LC 모형의 b_x와 LC-ER 모형의 $B_{x,t}$패턴을 비교한 것이다.

$B_{x,t}$는 시간이 흐름에 따라 b_x와 다르게 연령별 사망률 개선 패턴이 교대현상이 시작되는 시점(남자 2019년, 여자 2001년, 전체 2009년) 이후 매년 다르게 나타남을 확인할 수 있다.

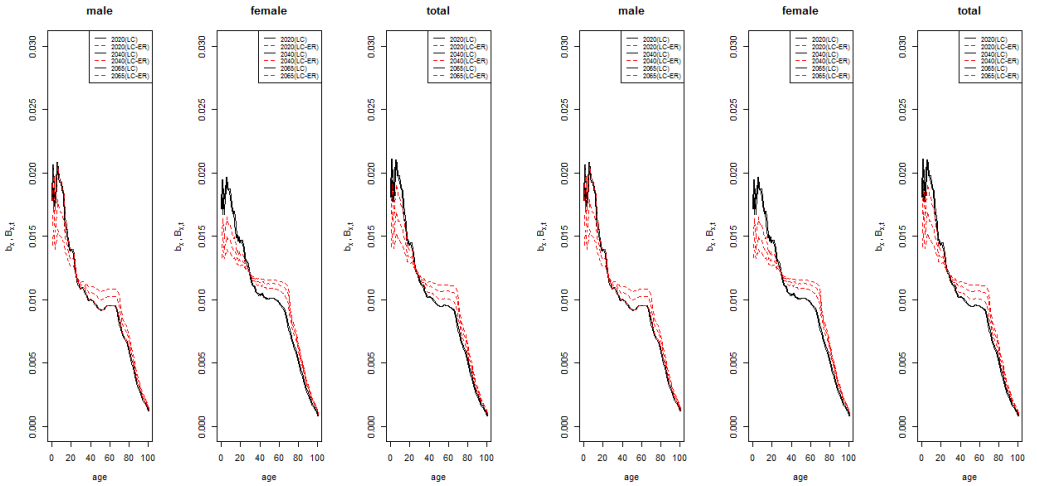
또한 남자의 경우 65세 전후로 사망률 개선 패턴의 점프현상이 발생하고 여자의 경우 사망률 개선 패턴의 단절현상이 발생함을 확인할 수 있었다. 이는 연령별 사망률 개선 속도가 더 이상 변하지 않고 일정하게 유지되는 최종시점에서 '65세 미만'가정이 적합하지 않다는 것을 의미한다.

Figure 3.3의 (b)–(d)는 우리나라의 사망률 교대현상이 나타나는 시점을 80세로 간주하고 b_x^u를 '60세



(a) Assumptions: 80세, 65세 미만

(b) Assumptions: 80세, 60세 미만



(c) Assumptions: 80세, 70세 미만

(d) Assumptions: 80세, 75세 미만

Figure 3.3. b_x of LC and $B_{x,t}$ of LC-ER over assumption change.

미만', '70세 미만', '75세 미만'에서 사망률 개선패턴이 동일하다 각각 가정하여 LC 모형의 b_x와 LC-ER 모형의 $B_{x,t}$패턴을 비교한 것이다.

비교 결과 '60세 미만'으로 가정한 경우 성별에 관계없이 사망률 개선 패턴의 단절 현상이 발생하였다. 그러나 '70세 미만'과 '75세 미만'으로 가정을 설정한 경우 다른 가정에 비해 점프현상 및 단절현상이 발생하지 않았다.

따라서 우리나라의 경우 '65세 미만'의 가정보다 '70세 미만' 또는 '75세 미만'의 가정이 타당할 것으로 보인다.

Table 3.1. Comparison of Life expectancy at birth on the kinds of LC

Year	Male						Female						KOSIS[1)]	
	LC	LM	BMS	LL	LC-ER	LL& LC-ER	LC	LM	BMS	LL	LC-ER	LL& LC-ER	Male	Female
1970	58.8	58.8	58.8	58.8	58.8	58.8	65.8	65.8	65.8	65.8	65.8	65.8	58.7	65.8
1975	60.3	60.3	60.3	60.3	60.3	60.3	68.2	68.2	68.2	68.2	68.2	68.2	60.3	68.2
1980	61.9	61.9	61.9	61.9	61.9	61.9	70.4	70.4	70.4	70.4	70.4	70.4	61.9	70.4
1985	64.6	64.6	64.6	64.6	64.6	64.6	73.2	73.2	73.2	73.2	73.2	73.2	64.6	73.2
1990	67.5	67.5	67.5	67.5	67.5	67.5	75.9	75.9	75.9	75.9	75.9	75.9	67.5	75.9
1995	69.7	69.7	69.7	69.7	69.7	69.7	77.9	77.9	77.9	77.9	77.9	77.9	69.7	77.9
2000	72.3	72.3	72.3	72.3	72.3	72.3	79.7	79.7	79.7	79.7	79.7	79.7	72.3	79.7
2005	74.9	74.9	74.9	74.9	74.9	74.9	81.6	81.6	81.6	81.6	81.6	81.6	74.9	81.6
2010	76.8	76.8	76.8	76.8	76.8	77.2	83.6	83.6	83.6	83.6	83.6	83.7	76.8	83.6
2015	78.9	79.0	79.0	78.9	78.9	79.5	85.1	85.1	85.1	85.1	85.1	85.2	79.0	85.2
2020	80.1	80.4	80.4	80.1	80.1	80.1	86.0	86.2	86.2	85.9	86.0	85.7	80.3	86.2
2025	81.4	81.7	81.8	81.3	81.4	81.4	86.9	87.2	87.2	86.7	86.9	86.5	81.6	87.0
2030	82.6	83.0	83.0	82.5	82.6	82.5	87.7	88.1	88.1	87.5	87.7	87.3	82.7	87.8
2035	83.7	84.1	84.1	83.5	83.7	83.6	88.5	88.9	88.9	88.2	88.5	88.0	83.7	88.5
2040	84.7	85.1	85.1	84.4	84.7	84.5	89.1	89.6	89.6	88.8	89.1	88.6	84.7	89.1
2045	85.6	86.0	86.1	85.3	85.6	85.3	89.8	90.2	90.2	89.4	89.8	89.2	85.5	89.7
2050	86.5	86.9	86.9	86.1	86.5	86.2	90.3	90.8	90.8	89.9	90.3	89.8	86.3	90.2
2055	87.2	87.7	87.7	86.8	87.2	86.9	90.8	91.3	91.3	90.4	90.8	90.3	87.1	90.7
2060	88.0	88.4	88.5	87.5	88.0	87.5	91.3	91.8	91.8	90.9	91.3	90.7	87.8	91.2
2065	88.7	89.1	89.2	88.2	88.7	88.2	91.8	92.3	92.3	91.3	91.8	91.2	88.4	91.7
2070	89.3	89.7	89.8	88.7	89.3	88.8	92.2	92.7	92.7	91.7	92.2	91.6	-	-
2075	89.9	90.3	90.4	89.3	89.9	89.3	92.6	93.1	93.1	92.1	92.6	91.9	-	-
2080	90.4	90.8	90.9	89.8	90.4	89.8	92.9	93.4	93.4	92.4	92.9	92.3	-	-
2085	90.9	91.4	91.5	90.3	90.9	90.3	93.2	93.8	93.8	92.7	93.2	92.6	-	-
2090	91.4	91.8	91.9	90.7	91.4	90.7	93.6	94.1	94.1	93.0	93.6	92.9	-	-
2095	91.9	92.3	92.4	91.1	91.9	91.1	93.9	94.4	94.4	93.3	93.9	93.2	-	-
2100	92.3	92.7	92.8	91.5	92.3	91.5	94.2	94.7	94.7	93.6	94.2	93.5	-	-

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; LL = coherent LC; LC-ER = extended of the LC with rotation; LL&LC-ER = combined LL and LC-ER.

1) 1970–2015년 통계청 자료는 생명표이며, 그 이후는「장래인구추계 (2016) 2016–2065」자료임.
모든 결과는 소수둘째자리에서 반올림 원칙을 적용하고, 미래 예측치 k_t, K_t 등은 랜덤워크를 가정함.

3.2. LC 계열 모형 적용 결과 비교

이번 절에서는 지금까지 소개하고 논의한 6가지 LC 계열 모형들을 우리나라 1970–2016년 사망률 자료에 적용해 결과를 비교하고자 한다. LC, LL, LC-ER, 그리고 LL&LC-ER 모형은 R 프로그램을 이용하여 구현하였으며, LM과 BMS 모형은 R에서 오픈소스로 제공하는 demography 패키지를 활용하였다. 이때 LC-ER과 LL&LC-ER 모형은 Li 등 (2003)이 제안한 것과 동일한 가정을 설정하였다.

Table 3.1은 본 논문에서 고려한 6가지 사망률 모형에 의한 결과와 통계청(KOSIS)에서 제공하고 있는 기대수명이다. 6가지 모형에 의한 결과에서 2020부터 2100년까지는 각 모형에 의한 예측치이며, 1970년부터 2015년까지는 각 모형의 적합성을 보여주기 위해 추정치를 기재하였다.

먼저 LC, LM, BMS, LL, LC-ER, 그리고 LL&LC-ER 모형의 기대수명 결과와 통계청 공식통계와 비

Table 3.2. MAE, MAFE for the kinds of LC

		LC	LM	BMS	LL	LC-ER	LC&LC-ER
MAE	Male	0.0022	0.0015	0.0015	0.0019	0.0202	0.0002
	Female	0.0026	0.0022	0.0023	0.0022	0.0199	0.0001
MAFE	Male	0.0032	0.0026	0.0026	0.0034	0.0031	0.0013
	Female	0.0018	0.0015	0.0012	0.0018	0.0025	0.0011

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; LL = Coherent LC; LC-ER = extended of the LC with rotation; LL&LC-ER = combined LL and LC-ER; MAE = mean absolute error; MAFE = mean absolute forecasting error.

교를 해보면, 1970부터 2015년 추정치는 통계청 각세별 완전생명표(1970∼2015년)와 거의 동일하거나 0.1세 정도 차이를 보인다. 그리고 2020년부터 2065년까지 예측치는 통계청 장래인구추계와 남자는 약 0.2–0.4세, 여자는 약 0.2–0.5세 차이를 보인다.

다음으로 2020년 이후 6개 모형의 예측결과는 LL 모형과 LL&LC-ER 모형의 기대수명이 상대적으로 낮은 값을 나타내고 LM과 BMS는 상대적으로 높게 도출되었다. 그리고 LC와 LC-ER은 거의 동일한 수준을 보였다. 따라서 기대수명 예측치 순위는 LM, BMS, LC, LC-ER, 통계청, LL, LL&LC-ER 모형 순이다.

그리고 LC-ER과 LL&LC-ER 모형은 LC과 LL모형에서 연령별 사망률 개선과 인구부집단의 사망 공통성을 고려한 확장 모형이지만 기대수명 측면에서는 큰 차이를 보이지 않는 것으로 판명되었다. 이러한 결과는 LC-ER와 LL&LC-ER 모형 기대수명의 예측치가 변경되는 것이 아니라 예측치를 생성하는 연령별 사망률 패턴을 변경하기 때문이다. 따라서 빠른 고령화와 연령별 사망률 개선을 반영하는 LC-ER과 LL&LC-ER 모형이 우리나라 사망률 예측에 합리적인 모형이라 할 수 있다.

끝으로 적합력과 예측오차 측면에서 모형을 살펴보았다. 본 연구는 6개 모형에 대해 적합성(MAE)과 예측력(MAFE)을 평가하기 위해 사망력 자료(1970–2016년)를 적합기간(1970–2011년, in sample)과 검증기간(2012–2016년, out sample)으로 분할한 후 평균절대오차(mean absolute error; MAE = $\sum_{t=1970}^{2011} |m_t - \hat{m}_t|/n$), 평균절대예측오차(mean absolute forecasting error; MAFE = $\sum_{t=2012}^{2016} |m_t - \hat{m}_t|/n$)를 Table 3.2와 같이 산출하였다.

결과에서도 알 수 있듯이 LL&LC-ER 모형이 타 모형보다 적합력이 좋고 예측오차가 작다는 것을 알 수 있다. 이는 LL 모형의 사망 공통성과 LC-ER 모형의 연령별 사망률 개선을 동시에 고려하였기 때문이다. 더불어 각 모형별로 MAE를 성별에 따라 살펴보면 LC, LL, LC-ER이 유사하고 LM, BMS가 유사한 수준을 보인다.

4. 결론 및 제언

미래의 사망률은 정부의 사회복지 및 의료지출과 관련이 있어 사망률의 예측 정확도가 향상되면 현재와 미래의 자원 배분에 관한 정책 결정에 도움이 된다. 이처럼 사망률 예측은 중요하다.

이런 관점에서 본 논문은 우리나라에서 기대수명 예측에 자주 사용되는 LC류 사망률 모형 변천을 살펴보고, 모형 변화 과정과 가정들의 개선사항들을 정리해 보았다. 또한 이들 모형들에 대해서 우리나라 사망률 자료에 적용해 결과들을 비교하였다.

LC, LM, BMS, LC-ER, 그리고 LL&LC-ER 모형은 추정치나 예측치 모두 유사한 결과가 산출되었으나, LL과 LL&LC-ER 모형은 상대적으로 약간 낮은 기대수명이 도출되었다. 그리고 이들 모두 KOSIS

생명표와 장래인구추계 기대수명과는 0.3세 이내의 차이를 보일뿐 거의 유사한 것으로 나타났다.

빠른 고령화 그리고 연령별 사망률의 개선 효과를 보이는 우리나라의 경우 기대수명에 큰 변화를 주지 않고 이들 현상을 반영하여 연령별 사망률 패턴을 수정하는 LC-ER과 LL&LC-ER 모형이 바람직하다고 볼 수 있다. 특히 LL&LC-ER 모형은 타 모형들보다 MAE, MAFE 기준으로 판단할때 적합력과 예측력의 우수성을 보였다. 또한, 17개 시도 장래 사망률 예측과 같이 인구부집단이 여러 개일 경우 시도 집단간의 관계를 고려하고 연령별 사망률 개선 효과를 반영할 수 있는 LL&LC-ER 모형의 적용이 적합하다 볼 수 있다.

본 연구를 진행하면서 향후 연구방향과 한계점을 도출할 수 있었다.

먼저, 우리나라 사망률에 대한 코호트효과를 살펴볼 필요가 있다. 본 연구는 기간효과에 대한 LC류 모형을 살펴보았으나 코호트 효과는 고려하지 않았다. Cairns 등 (2006, 2007, 2008)과 Kim (2012)의 선행연구에서도 밝혔듯이 우리나라 사망률은 코호트 효과가 존재하는 것으로 판명되었다. 따라서 기간효과와 코호트효과를 동시에 고려할 수 있는 모형 연구가 필요하다.

다음으로 기대수명 상한연령(open-age)을 100세 이상으로만 간주하여 분석하였다. 하지만 Kannisto 등 (1994)과 Sevcikova 등 (2015)이 제시한 것처럼 상한연령을 110세 또는 120세로 간주하여 고연령 사망률을 도출할 필요가 있다. 특히 우리나라는 고령층에 점진적인 사망률 개선을 보이고 있으므로 이들의 사망률 변화를 살펴보는 것은 의미 있는 연구이다.

끝으로 젊은층의 사망률 개선율은 점진적으로 낮아지고 고령층의 사망률 개선은 높아짐에 따라 사망률 곡선의 변화는 각 연령층마다 다른 기울기를 보일 수 있다. 이처럼 서로 다른 연령별 사망률을 적합할 수 있는 연령별 변화(shift-age)모형도 고안할 필요가 있다고 본다.

이상으로 본 연구 결과가 우리나라 사망률 모형 개선과 예측 정확도에 일조하기를 바란다.

References

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan disparity as an additional indicator for evaluating mortality forecasts, *Demography*, **54**, 1559–1577.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline, *Population Studies*, **56**, 325–336.

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics, *Biometrics*, **42**, 693–734.

Cairns, A., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality: theory and calibration, *Journal of Risk and Insurance*, **73**, 687–718.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States, *North American Actuarial Journal*, **13**, 1–35.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: an analysis of six stochastic mortality models, *Insurance Mathematics and Economics*, **43**, 355–367.

De Jong, P. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting, *Mathematical Population Studies*, **13**, 1–18.

Doukhan, P., Pommeret, D., Rynkiewicz, J., and Salhi, Y. (2017). A class of random field memory models for mortality forecasting, *Insurance : Mathematics and Economics*, **77**, 97–110.

Guibert, Q., Lopez, O., and Piette P. (2017). Forecasting mortality rate improvements with a high-dimensional VAR, *HAL-01613050*, doi:hal.archives-ouvertes.fr/hal-01613050

Horiuchi, S. and Wilmoth, J. (1995). Annual meeting of the population association of America. San Francisco, CA: Population Association of America.

Hunt, A. and Villegas A. M. (2015). Robustness and convergence in the Lee-Carter model with cohort effects, *Insurance: Mathematics and Economics*, **64**, 186–202.

Jeong, S. and Kim K. W. (2011). A Comparison study for mortality forecasting model by average life expectancy, *The Korean Journal of Applied Statistics*, **24**, 115–125.

Jung, K., Back, J., and Kim, D. (2013). Comparison of mortality estimate and prediction by the period of time series data used, *The Korean Journal of Applied Statistics*, **26**, 1019–1032.

Kannisto, V., Lauritsen, J., Thatcher, A. R., and Vaupel, J. W. (1994). Reductions in mortality at advanced ages: several decades of evidence from 27 countries, *Population and Development Review*, **20**, 793–810.

Kang, J. C., Lee, J. C., and Sung, J. H. (2006). A study on methods for forecasting mortality considering longevity risk, *The Journal of Risk Management*, **17**, 153–178.

Kim, S. J. (2012). A comparison study on the stochastic mortality models for measuring longevity risk, *Korean Insurance Journal*, **93**, 213–235.

Kim, S. Y. and Han, M. J. (2017). Mortality forecasting for 2016 Korean Population Projection, *Korean Journal of Population Studies*, **40**, 1–25.

Kim, S. Y. and Oh, J. H. (2017). A study comparison of mortality projection using parametric and non-parametric model, *The Korean Journal of Applied Statistics*, **30**, 701–717.

Kim, S. Y., Oh, J. H., and Kim, K. W. (2018). A comparison of mortality projection by different time period in time series, *The Korean Journal of Applied Statistics*, **31**, 41–65.

KOSIS (2016). Population Projections (2015–2065).

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

Lee, R. D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography*, **38**, 537–549.

Li, H. and Lu, Y. (2017). Coherent forecasting of mortality rates: a Sparse Vector-Autoregression approach, *ASTIN Bulletin: The Journal of the IAA*, **47**, 563–600.

Li, N. and Gerland, P. (2011). Modifying the Lee-Carter Method to Project Mortality Changes up to 2100, the Population Association of America 2011 Annual meeting-Washington, DC, Session 125, formal Demography I: Mathematical Models and Methods.

Li, N. and Lee R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method, *Demography*, **42**, 575–594.

Li, N., Lee, R., and Gerland, P. (2013). Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections, *Demography*, **50**, 2037–2051.

Park, Y. S., Kim, K. W., Lee, D. H., and Lee, Y. K. (2005). A comparison of two model for forecasting mortality in South Korea, *The Korean Journal of Applied Statistics*, **18**, 639–654.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors, *Insurance: Mathematics and Economics*, **38**, 556–570.

Sevcikova, H., Li, N., Kantorova, V., Gerland, P., and Raftery, A. E. (2015). Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections. Working paper no. 150, Center for Statistics and the Social Sciences University of Washington.

Statistics Canada (2015). Population Projections for Canada(2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions.

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The force of mortality at ages 80 to 120, *Odense Monographs on Population Aging 5*, Odense University Press.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, *Royal Statistical Society*, **67**, 301–320.

우리나라 사망력 모형의 변천과 가정 고찰
- Lee-Carter 류를 중심으로 -

오진호[a] · 김순영[a,1]

[a]통계청 통계개발원

(2018년 7월 23일 접수, 2018년 9월 14일 수정, 2018년 9월 16일 채택)

요 약

빠른 고령화로 고령층의 증가는 인구구조 변화와 인구고령화에 영향을 미친다. 예전부터 선진국은 인구고령화를 주요현안으로 간주하여 고령화로 인한 연금 재정건전성, 건강 및 노인 복지 시스템의 지속 가능성에 집중하고 있다. 이처럼 고령층의 증가로 인구구조 변화와 인구고령화에 미치는 사망률 예측은 어느 때보다도 중요하다. 본 논문은 통계청 1970–2016년 각세별 생명표 자료를 활용하여 사망률 모형 6가지를 비교하였다. 이들 모형은 Lee-Carter (LC) 모형 (Lee and Carter, *Journal of the American Statistical Association*, **87**, 659–671, 1992)에 근원을 두고 있으며, LC 의 가정을 수정하고 개선한 것이다. 이들 개선과정과 가정검토를 모형별로 살펴보고 우리나라에 적합한 사망률 모형을 모색했다. 분석결과 빠른 고령화와 연령별 사망률의 개선 효과를 보이는 우리나라의 경우 기대수명에 큰 변화를 주지 않고 이들 현상을 반영하고 연령별 사망률 패턴을 수정하는 LC-ER 모형 (Li 등, *Demography*, **50**, 2037–2051, 2013)과 Li-Lee 모형과 LC-ER모형을 조합한 LL&LC-ER 모형으로 사망률을 예측하는 것이 바람직하다.

주요용어: 인구고령화, 사망률 모형, Lee-Carter 모형, LC-ER, LL&LC-ER

연구는 통계청의 공식견해가 아니며 저자의 개인적인 연구결과임을 밝힙니다.

[1]교신저자: (35220) 대전광역시 서구 한밭대로 713 통계센터 6층, 통계개발원 통계분석실.
E-mail: sykim0213@korea.kr