

평균수명을 이용한 사망률 예측모형 비교연구

정승환¹ · 김기환²

¹고려대학교 경제통계학과, ²고려대학교 정보통계학과

요약

사망률 예측모형과 생명표 작성방법에 기반을 둔 예측평균수명 작성은 미래의 사망수준을 평가하는 효과적인 방법이 된다. 2006년 통계청에서 장래인구추계 작성 시 예측평균수명을 작성하였으나, 2006년 이후 현재까지 실제평균수명과 적지 않은 차이를 보이고 있어 평균수명의 증가속도를 반영하지 못하고 있다. 이의 원인으로는 전망치에 대한 판단, 사망률 예측모형의 선택과 사용 등이 이유가 될 수 있다. 본 논문에서는 사망률 예측모형의 선택관점에서 이 문제를 살펴보고자 한다. 2011년 장래인구추계 작성을 앞둔 상황에서 오류의 반복을 피하기 위해서는 사망률 예측모형에 대한 특성 및 적용가능성에 대한 충분한 검토가 이루어진 후 적절한 모형을 선택해야 할 것이다. 사망률 예측모형은 주로 사용되고 있는 LC(Lee와 Carter) 모형과 이의 개선모형들, 사망확률 확장모형인 HP8(Heligman과 Pollard 8 parameters) 모형 등 모두 5개의 모형을 비교·분석하였다. 분석결과를 바탕으로 5개의 모형별로 2030년까지의 남녀별 예측평균수명을 작성하여 제시하였고, 이를 통계청에서 제공하는 예측평균수명과 비교하였다. 5개의 모형에 의해 작성된 2030년까지의 새로운 예측평균수명은 통계청의 결과보다 높게 나타나 실제평균수명의 변화를 상대적으로 잘 반영하는 것으로 나타났다.

주요어: 사망률 예측모형, 평균수명, 생명표, Lee-Carter 모형, HP8 모형.

1. 서론

사망자료를 계량적으로 분석하고 예측하는 것은 가까운 장래에 도래할 사회적·경제적 문제점들에 대처하기 위한 정책수립에 중요한 바탕이 되므로 매우 중요한 연구라고 할 수 있다. 우리나라의 경우, 사회의 고령화 속도가 다른 나라들에 비해 매우 빠르게 진행되고 있어 이에 대한 정확한 상황 파악과 예측에 근거한 정책개발이 요구되고 있는 실정이다. 그러나 통계청의 장래인구추계에서 제공된 평균수명 예측값들은 실제 당해 연도 평균수명과 많은 차이를 보이고 있어 고령화, 저출산, 연금, 보험 등에 대한 효과적인 정부정책이 개발되고 있는 것인지에 의문을 품지 않을 수 없다. 표 1.1은 통계청에서 제공하는 장래인구추계 상의 예측평균수명과 실제 당해 연도의 평균수명을 비교·정리한 표이다. 표 1.1에서 볼 수 있듯이 장래인구추계에서 예측한 평균수명은 실제 당해 연도 평균수명보다 낮게 예측되고 있음을 알 수 있다. 1996년 추계에서 예측한 2030년 평균수명은 78.95세로 2005년 실제 평균수명(경험치) 78.63세와 비슷한 수준이었다. 그리고 실제 2008년 평균수명은 80.08세로 '2006년 장래인구추계'의 2015년 예측값인 80.55세에 근접하고 있다. 비록 최근들어 통계청에서 작성된 예측평균수명이 과거에 작성하였던 예측평균수명과는 다르게 실제 평균수명과 차이가 줄어드는 경향을 보이고 있지만, 지속적으로 평균수명 예측값이 실제 경험치보다 낮게 예측되고 있다. 평균수명 예측결과는 급격한 고령화로 예상되는 사회문제들에 대비하기 위한 정부 정책을 결정하는데 기초자료가 되므로 현재 통계청의 평균수명 예측방법이

본 연구는 고려대학교 특별연구비에 의하여 수행되었음.

²교신저자: (339-700) 충청남도 연기군 서창리 208번지 고려대학교 세종캠퍼스, 정보통계학과, 부교수.

E-mail: korpen@korea.ac.kr

표 1.1. 장래인구추계에서의 예측평균수명과 실제평균수명 비교

	2000	2005	2006	2007	2008	2010	2015	2020	2030
경험치	76.02	78.63	79.18	79.56	80.08				
1996년 추계	74.85	76.10				76.99	77.48	78.08	78.95
2001년 추계		77.70				78.78	79.73	80.73	81.53
2005년 특별추계		77.90				79.06	80.08	81.00	81.90
2006년 추계						79.60	80.55	81.45	83.13

앞으로도 계속된다면, 향후의 변화에 적절한 대처를 하지 못할 수도 있다. 통계청은 2011년에 2010년 센서스 결과를 기반으로 작성한 새로운 ‘장래인구추계’를 발표하게 되므로 평균수명 예측과정에 대한 점검은 반드시 필요한 절차라 하겠다.

내용의 전개를 위하여 2절에서는 통계청에서 장래인구추계를 위해 사용하였던 사망률 예측모형에 대하여 살펴보았으며, 3절에서는 통계청에서 사망률 예측에 사용되었던 LC 모형 (Lee와 Carter Model, 1992)의 장단점과 개선 모형들의 특성에 대하여 살펴보고 비교하였다. 4절에서는 LC 모형과는 다른 관점에서 사망확률을 모형화하는 HP8 모형 (Heligman과 Pollard 8 parameter model, 1980)의 특성에 대하여 자세히 살펴보았다. 3절, 4절에서 실제 우리나라 사망률 자료로 모형적합 했던 결과를 바탕으로 5절에서는 모형별로 2030년까지의 우리나라의 남녀별 평균수명을 예측하고 통계청의 2006년 발표결과와 비교하였으며, 6절에서는 각 모형과 예측결과에 대하여 정리·평가하였다.

2. 통계청의 사망률 예측방법

평균수명이란 연령별 사망률(연령별 사망자수/해당인구)을 이용하여 작성한 생명표(life table)의 0세 기대여명을 말하고, 예측평균수명이란 연령별 사망률의 예측값을 이용하여 작성한 생명표의 0세 기대여명을 말한다. 따라서 예측평균수명을 구하기 위한 핵심은 연령별사망률 예측을 위해 어떤 예측모형을 어떻게 적용하였는가가 될 것이다. 본 논문에서는 사망률 예측모형의 비교가 주요 목적이므로 비교의 기준으로 통계청에서 2006년 장래추계인구 작성 시 사용하였던 모형을 살펴보고자 한다. ‘2006년 장래인구추계 보고서’에 기술된 연령별 사망률예측에 관한 내용은 다음과 같다.

- 1) Lee-Carter 모형에 의해 추정된 후 보간법을 이용하여 연도별 사망확률을 추정.
- 2) 0세 및 1-4세 경우는 Lee-Carter 방법 적용 시 너무 급격하게 감소하여 2050년 일본의 사망확률로 수렴할 것으로 가정.
- 3) 75세 이상은 기초자료 부족으로 Brass Logit 방법을 적용.

통계청에서는 위의 설명 이외에는 내용을 공개하지 않고 있어 세부적인 부분까지 어떻게 적용하였는지는 알 수 없으나 세 항목에 대한 개략적인 내용은 문헌연구를 통하여 파악할 수 있었다.

$$\begin{aligned}
 m(x, t) &= \exp(a_x + b_x k_t + \epsilon_{x,t}), \\
 \log(m(x, t)) &= a_x + b_x k_t + \epsilon_{x,t}.
 \end{aligned}
 \tag{2.1}$$

LC 모형은 비정칙치분해(singular value decomposition)를 통해 얻어지는 첫 번째 주성분을 이용하여 모형을 구성하게 되며, 기본 식은 식 (2.1)과 같고, $m(x, t)$ 는 연령 x 와 시간 t 에서의 사망률을, $\log(m(x, t))$ 는 로그사망률을 나타낸다. a_x 는 연령에 따른 일반적인 사망률 패턴을 나타내며, b_x 는 연령에 따른 사망률 변화, k_t 는 시간에 따른 사망률 변화를 나타낸다. LC 모형에 대한 평가는 3절에서 다

표 3.1. Lee-Carter 모형의 장단점

장점	<ul style="list-style-type: none"> - 인구학적(demographic) 모형과 통계적인 시계열 모형이 결합. - 예측값에 대한 신뢰구간을 제공한다. - 주관적인 판단에 의존하지 않고 장기간의 사망률 추세를 이용하여 예측 가능하다. - 비교적 만족스러운 (장기)예측값을 제공한다. - 현재 사망률 예측을 위해 가장 널리 사용되고 있다.
단점	<ul style="list-style-type: none"> - 연령별 사망률 구조가 시간과 관계없이 일정하게 유지되는 것으로 가정. - 새로운 추계를 시작하는 연도의 자료(jump-off year data)에 대한 적합력 부족.

루게 되며, 모형이 세부적인 내용은 박유성 등 (2005)을 참고하기 바란다.

Brass Logit 방법은 Brass (1971)에 의해 제안된 로짓(logit) 모형으로써, 표준사망률과 생명표 상 사망률의 선형관계를 이용하여 작성되는 모형이다. 이들 사망률 사이에는 다음과 같이 선형관계가 형성된다.

$$\text{logit}(m(x, t)) = \alpha_t + \beta_t \text{logit}(m(x, s)).$$

여기서 $m(x, t)$ 는 식 (2.1)과 동일하고, $m(x, s)$ 는 표준인구의 연령별 사망률, 즉 표준사망률을 나타낸다. 표준사망률은 국제연합 신모형 생명표 (UN, 1982)에서 제공하는 사망 유형별 평균 로짓(logit) 사망률을 말하고, 권태환과 김태현 (1990)에 따르면, 우리나라의 경우 남자는 ‘극동유형’, 여자는 ‘일반유형’의 평균 로짓사망률을 따른다고 알려져 있다. Brass Logit 모형은 표준사망률($m(x, s)$)과 연령별 사망률($m(x, t)$)의 비교 모수인 α_t 와 β_t 가 정해지면 $m(x, s)$ 를 이용하여 $m(x, t)$ 를 구성할 수 있게 되는 구조이다. 통계청에서 밝혔듯이 Brass Logit 모형은 부족한 75세 이상의 고령자 자료를 보완하기 위해 사용된 것으로 보인다. 그러나 LC 모형 적용 후 추정사망률에 적용한 보간법에 관한 내용, 75세에서 LC 모형과 Brass Logit 모형을 어떻게 통합하였는지는 관련내용을 확인할 수 없다.

3. LC 모형과 개선 모형

LC 모형은 단순명료한 구조로 적용이 간단하며, 사망률 예측이 용이하고 예측력 면에서도 우수하다고 알려져 있어 현재까지 가장 많이 사용되는 예측기법이다 (Lee와 Miller, 2001).

식 (2.1)의 LC 모형을 보면 시간에 무관한 a_x , b_x 와 시간에 의존하는 k_t 로 이루어져 있다. LC 모형은 k_t 의 예측을 통하여 예측사망률을 얻게 된다. 이때 기본적으로 식 (3.1)의 예측모형(random walk with drift)을 사용한다. 그러나 다른 단변량 시계열예측모형의 사용도 가능하다.

$$k_t = k_{t-1} + d + e_t. \quad (3.1)$$

LC 모형은 표 3.1에 정리한 것과 같이 사용상의 용이성과 안정적인 결과를 제공해주는 장점이 있으나, 실제 적용상에서 여러 가지 문제점들이 발견되었으며, 이를 해결하기 위한 개선 모형들이 제안되었다. LC 모형의 개선은 여러 요인에 대해서 이루어지지만 주된 초점은 k_t 를 얼마나 잘 구해낼 것인가에 맞추어진다. 관련 연구가 Hyndman 등 (2006)에 의하여 진행되었으며, 각 모형의 적합에 대한 내용을 R 패키지로 작성하여 공개하고 있다(<http://robjhyndman.com/software/demography>).

3.1. Lee-Miller(LM) 모형

LM 모형은 Lee와 Miller (2001)가 제안한 모형으로 LC 모형보다 시간의 변화를 좀 더 잘 반영하는 k_t 의 추정과 센서스 이후 새로운 추계를 시작하는 연도의 자료에서 발생하는 편의(jump-off bias, 이후

탈선편의)의 제거를 목적으로 하고 있다. 제안 당시의 LC 모형은 1920년대 자료부터 사용하였으며 많은 자료를 사용할수록 안정적인 결과를 얻을 수 있다고 하였으나 Lee와 Miller는 연령별 사망률 패턴의 변화 b_x 가 1950년을 기준으로 달라지는 것을 발견하고 1950년 이후 사망률 자료를 사용할 것을 권하였다(미국, 스웨덴, 프랑스, 캐나다가 이에 해당하며, 전염병, 2차 세계대전이 원인). 또한, LC 모형에서 추정된 k_t 는 총사망자수를 이용하여 조정하게 되어 있으나 LM 모형에서는 평균수명을 이용하여 k_t 를 조정하였다. LC 모형은 연령별 사망구조가 일정하게 유지되는 것을 가정하고 있기 때문에 탈선편의를 반영할 수 없지만, LM 모형에서는 a_x 를 가장 최근 연령별 사망률의 실제 값으로 대체함으로써 이 문제를 해결하였다.

3.2. Booth-Maindonald-Smith(BMS) 모형

Booth 등 (2002)은 k_t 추정 후 조정과정의 개선과 모형적합에 사용되는 자료의 기간을 통계량에 의하여 선택하도록 하는 측면에서 LC 모형을 개선하였다. LC 모형처럼 k_t 의 적합과정에서 총사망자수로 조정하는 경우 연령별 사망자수 분포에서 왜곡이 발생할 수 있음을 지적하고 BMS 모형에서는 k_t 의 적합과정 중 조정을 연령별 사망자수로 하였고 동시에 탈선편의를 회피하는 효과도 얻을 수 있었다. 또한, BMS 모형에서는 모형적합을 위해 사용하는 자료의 기간을 식 (3.2)의 적합도 통계를 이용하여 결정하게 된다.

$$\text{deviance}_t = 2 \sum_x \left(D_{x,t} \ln \frac{D_{x,t}}{D'_{x,t}} - (D_{x,t} - D'_{x,t}) \right). \quad (3.2)$$

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

3.3. Hyndman-Ullah(HU) 모형

Hyndman과 Ullah (2007)는 LC 모형의 개선을 LM 모형, BMS 모형과는 다른 관점에서 시도하였다. HU 모형은 함수적 자료분석(functional data analysis), 비모수 평활기법(nonparametric smoothing)을 이용한 원자료의 평활을 포함하고 있다. 원자료에 적용되는 평활식은 식 (3.3)과 같으며

$$\log(m_t(x)) = f_t(x) + \sigma_t(x)\epsilon_t(x) \quad (3.3)$$

평활함수 $f_t(x)$ 에 의하여 적합 되는 부분과 확률변수 $\epsilon_t(x)$, 연령별 잡음 $\sigma_t(x)$ 에 의한 적합 되지 않는 부분으로 구성된다. 평활함수 $f_t(x)$ 는 식 (3.4)와 같이 정의된다. 식 (3.4)의 특징은 LC 모형의 비정칙치분해로부터 얻어지는 주성분 중 가장 큰 것 하나만을 사용하는 반면, 이 식에서는 두 개 이상의 주성분을 사용하여 첫 번째 주성분이 설명하지 못한 부분들로부터 추가적인 정보를 획득하는 것이다. HU 모형에서는 k_t 에 대한 추가적인 보정은 시행하지 않는다.

$$f_t(x) = a(x) + \sum_{j=1}^J k_{t,j} b_j(x) + e_t(x). \quad (3.4)$$

여기서 $a(x)$ 는 나이에 따른 일반적인 로그사망률 패턴을 나타내고, $k_{t,j}$ 와 $b_j(x)$ 는 해당 주성분 j 에서 시간에 따른 사망수준의 변화와 각 연령의 로그사망률 변화 정도를 나타낸다.

표 3.2. 2000~2008년 남녀별 실제평균수명과 예측평균수명 비교

	2000	2001	2002	2003	2004	2005	2006	2007	2008	MAD
실제평균수명(남)	72.25	72.82	73.40	73.86	74.51	75.14	75.74	76.13	76.54	
LC-남	72.92	73.46	73.99	74.53	75.08	75.64	76.20	76.78	77.36	0.62
LM-남	72.78	73.26	73.75	74.24	74.74	75.24	75.75	76.26	76.78	0.27
BMS-남	73.00	73.53	74.07	74.62	75.17	75.73	76.30	76.88	77.47	0.71
HU-남	73.11	73.96	74.39	75.02	75.58	75.95	76.56	77.23	77.83	1.03
실제평균수명(여)	79.60	80.04	80.45	80.81	81.35	81.89	82.36	82.73	83.29	
LC-여	80.26	80.64	81.02	81.40	81.77	82.13	82.49	82.85	83.21	0.38
LM-여	80.14	80.50	80.85	81.20	81.55	81.89	82.24	82.57	82.91	0.30
BMS-여	80.40	80.78	81.16	81.54	81.91	82.28	82.64	83.00	83.36	0.51
HU-여	80.29	80.59	80.96	81.19	81.65	82.05	82.41	82.82	83.23	0.31

3.4. 실제평균수명과 예측평균수명의 비교

3절에서 소개했던 4개의 사망률예측모형의 예측력을 확인하기 위하여 1980~1999년 우리나라 사망률자료를 이용하여 2000~2008년 남녀별 평균수명을 예측하고 예측결과들은 국가통계포털(KOSIS, <http://www.kosis.kr>)에서 제공하는 2000~2008년 실제 남녀별 평균수명과 비교하였다. 각 모형을 이용한 예측에는 Hyndman 등의 R 패키지를 사용하였다. 표 3.2는 이를 정리한 결과이며, 그림 3.1은 모형별로 예측된 로그사망률을 그림으로 표시한 것이다. 그림 3.1의 (a)~(d)는 모형별 예측 여성 로그사망률을 나타낸 것으로 일반적인 연령별 로그사망률의 패턴(나이키 패턴)을 따르고 있는 것을 공통적으로 볼 수 있다. 그림의 크기 때문에 관찰에 어려움이 있으나, 네 가지 모형 모두 2000~2008년 기간 로그사망률의 지속적인 감소(나이키 패턴의 하강)로 예측한 것을 확인할 수 있으며, 3절에서 설명한 네 가지 모형의 차이점, 즉 LC 모형의 개선을 위해 반영된 부분들의 효과가 확연히 드러나고 있는 것을 볼 수 있다. 이러한 모습들은 연령별 남성 예측 로그사망률에서도 동일하게 관찰된다.

표 3.2에 정리된 결과를 보면 모형별, 남녀별 예측평균수명이 실제평균수명과 매우 흡사함을 알 수 있고 이 두 값 간의 절대평균오차(Mean Absolute Deviation; MAD)도 낮은 수준을 보이고 있다. 통계청이 발표한 예측평균수명이 모형의 적용을 통해 어렵지 않게 구한 예측평균수명보다 예측력이 떨어지는 이유는 예측모형의 선택과 엄격한 적용 등에서 찾을 수도 있겠으나, Lee와 Miller가 지적했던 전문가 의견의 비관적 편향(pessimistic bias of expert opinion) 즉, 정부기관이 발표하는 장기예측값이 담당 실무 경험자 또는 해당 분야 전문가들의 의견에 따른 조정으로 인해 실제보다 낮게 예측되는 경향이 주된 원인일 것으로 예상된다.

4. Heligman과 Pollard 8 Parameter(HP8) 모형

사망률을 모형화하고 예측하는 LC 모형 및 개선 모형들과는 달리 HP8 모형은 5세 단위의 사망확률을 1세 단위의 사망확률로 확장해주는 방법이다. 사망확률은 사망률을 기반으로 조정된 값으로 생명표 작성과정을 따르면 쉽게 구할 수 있다. HP8 모형은 식 (4.1)과 같이 총 8개의 모수로 구성되며, 이 8개의 모수를 이용하여 로그사망확률을 적합하게 된다. HP8 모형에 관한 연구는 Heligman과 Pollard (1980), Kostaki (1991), Kostaki와 Panousis (2001) 등에 의해 이루어졌고 이를 우리나라 자료에 적용한 연구로는 김기환 등 (2006)이 있다. 이 연구에서 우리나라 1971~2003년, 5세 단위 사망확률을 1세

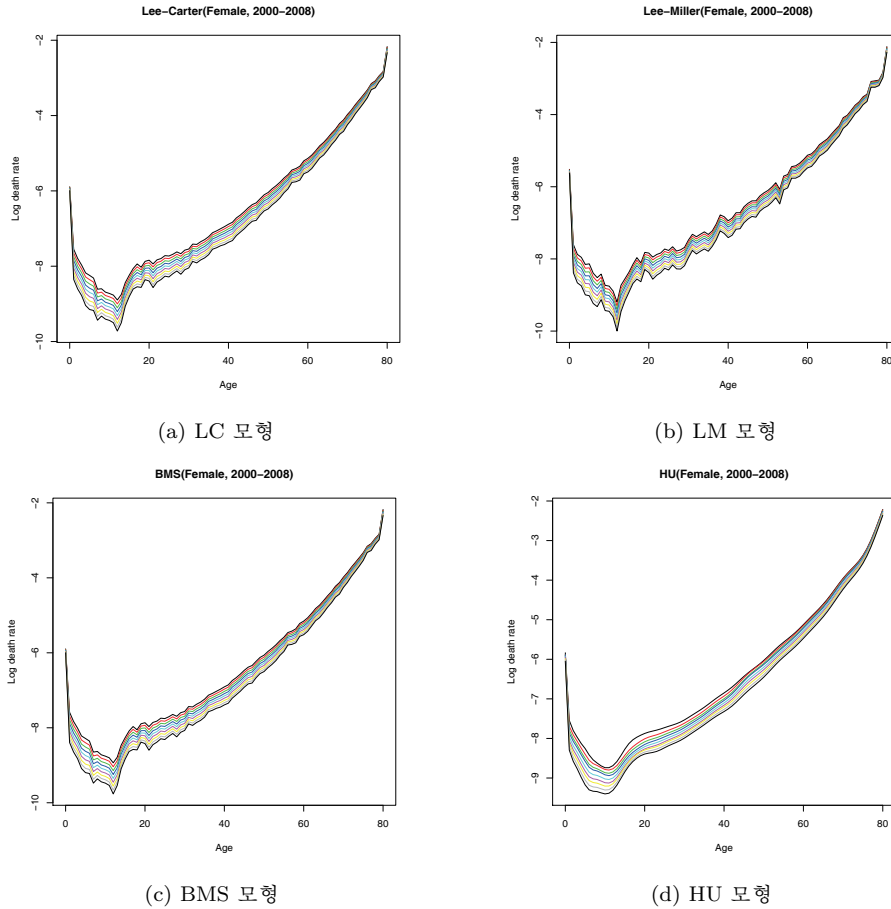


그림 3.1. 적합 모형별 2000-2008년 여성 로그사망률 예측도표

단위 사망확률로 확장하는 과정과 결과를 제시하였다.

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D \exp\left(-E \left(\ln\left(\frac{x}{F}\right)\right)^2\right) + GH^x. \quad (4.1)$$

식 (4.1)의 q_x 는 x -세의 사망확률을 나타내며 $p_x = 1 - q_x$ 이고, A, B, C, D, E, F, G, H 는 q_x/p_x 의 변화를 표현하는 8개의 모수이다. 8개의 모수는 크게 세 부분으로 나누어 볼 수 있다. A, B, C 모수로 구성된 첫 번째 부분은 유아연령층 사망률의 급격한 지수적 감소를 나타내고, D, E, F 모수로 구성된 두 번째 부분은 청년층의 사고로 인한 사망률 증가에 영향을 주는 부분을 나타낸다. G, H 모수로 구성된 세 번째 부분은 노년층 사망률의 급격한 증가를 나타내는 부분이다. 표 4.1은 HP8 모형에서 8개 모수의 개별적 역할을 정리한 것이다.

통계청에서 제공하고 있는 간이생명표의 5세 단위 사망확률을 HP8 모형에 연도별로 적용하면, 추정된 모수로 이루어진 8개의 시계열을 얻을 수 있다. 이 8개의 시계열에 적절한 시계열모형을 적합하게 되면, 예측값을 얻을 수 있다. 예측된 8개의 추정 모수 시계열을 HP8 모형에 재적용하면, 예측사망확률을 얻을 수 있다.

표 4.1. HP8 모형의 모수별 역할

	모수	의미
첫 번째 부분	A	1세 사망확률(q_1)과 가장 근접한 값으로 1세 사망률의 수준을 나타냄.
	B	1세 사망률과 0세 사망률의 차이를 나타내는 부분이고 B 값이 클수록 q_0 는 q_1 에 근접함.
	C	1세 이후 사망률이 감소하는 비율을 나타내는 부분으로써, C 값이 증가할수록 사망률은 더 급속하게 하락.
두 번째 부분	D	사고로 인해 발생한 사망률이 얼마나 증가했는지를 나타내는 부분, 즉 곡선의 높이를 나타냄.
	E	사망률이 밀집되어 있는 정도를 나타내는 부분, 즉 곡선의 폭을 나타냄.
	F	사고로 인한 사망률의 증가가 어느 연령대에서 발생하는지를 나타냄.
세 번째 부분	G	노년층 사망률의 기본 수준을 나타냄.
	H	노년층에서 사망률이 증가하는 비율(slope)을 나타냄. G와 H의 패턴은 일반적으로 반대의 패턴을 보인다.

표 5.1. 2004~2008년 HP8 모형 적합에 의한 남녀별 모수값

		A	B	C	D	E	F	G	H
남자	2004	0.001388	0.054949	0.081287	0.001261	8.451932	19.6139	0.000240	1.103677
	2005	0.001291	0.042474	0.075637	0.001293	7.128892	20.5367	0.000209	1.104991
	2006	0.001098	0.026407	0.068554	0.001071	6.422552	19.4557	0.000191	1.105555
	2007	0.001128	0.045119	0.072771	0.001484	5.895348	20.0944	0.000172	1.106582
	2008	0.001082	0.080107	0.083935	0.001655	5.960378	20.9427	0.000166	1.106494
여자	2004	0.001106	0.000280	0.029370	0.001416	5.012144	27.4227	0.000030	1.124175
	2005	0.001027	0.000107	0.026848	0.001568	5.155666	27.0841	0.000029	1.123787
	2006	0.000924	0.000081	0.025758	0.001328	4.824866	27.7637	0.000024	1.126038
	2007	0.000829	0.000007	0.020521	0.002014	5.348784	27.3241	0.000022	1.126362
	2008	0.000715	0.001721	0.009833	0.002026	5.100699	28.2964	0.000020	1.126768

5. 모형별 평균수명 예측과 비교

이 절에서는 LC, LM, BMS, HU, HP8 총 5개 모형을 이용하여 2030년까지 남녀별 평균수명을 예측하고 비교하였다. LC 모형과 개선 모형의 경우 표 3.2 작성을 위해 사용하였던 자료를 확장하여 2030년까지 예측기간을 연장하였다. 즉 1980~2008년까지의 사망률 자료를 이용하여 모형들을 적합하고 2030년까지 남녀, 연령별 사망률을 예측하고 이를 이용하여 연도별 남녀 평균수명을 계산하였다. HP8 모형의 경우 김기환과 정승환 (2006)의 연구결과(1971~2003년은 기존의 연구결과 사용)에 2004~2008년 자료에 대한 적합결과를 추가하여 1971~2008년까지 총 38 시점에 대한 8개의 모수 시계열을 작성하였고, 2030년까지 예측하였다. 표 5.1은 2004~2008년 추가된 남녀별 8개의 모수 값을 정리한 것이다. 남녀별 8개의 추정 모수 시계열은 사망확률의 변화를 반영하기 때문에 매우 완만한 변화를 보이게 되고, 2030년까지 장기예측을 시행하게 되므로 식 (5.1)의 간단한 시계열회귀모형으로 남녀별 추정 모수 시계열을 적합하였다.

$$P_{i,t} = \beta_{0,i} + \beta_{1,i}t + \epsilon_{i,t}, \quad i = A, B, \dots, H, \quad t = 1, 2, 3, \dots, 38 \quad (5.1)$$

여기서 $P_{i,t}$ 는 모수별로 구성된 시계열을 의미한다. 식 (5.1)의 추정결과는 표 5.2에 정리하였다. 모수별로 추정된 시계열 회귀식에서 t 값을 40~60까지 변화시키면 2010~2030년까지 예측 모수 시계열을 얻을 수 있고 식 (4.1)에 대입하면 예측사망확률을 구할 수 있다.

앞서 언급한 5개의 모형을 이용하여 작성한 2010~2030년 남녀별 예측평균수명을 표 5.3에 정리하였고 비교를 위하여 통계청의 장래인구추계 (2006)에 발표된 남녀별 평균수명을 표 5.3에 함께 정리하였다.

표 5.2. HP8 모형 추정 모수 시계열 남녀별 적합결과(1971~2030년)

	남자		여자	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
A	0.0022	-0.00003	0.0020	-0.00003
B	-0.0179	0.00123	-0.0294	0.00082
C	0.0806	-0.00034	0.0421	-0.00045
D	0.0022	-0.00003	0.0025	-0.00003
E	12.1462	-0.15900	8.9860	-0.11560
F	18.3248	0.03140	25.4305	0.06480
G	0.000419	-0.000006346	0.0000435	-0.000000547
H	1.0910	0.00041	1.1169	0.00025

표 5.3. 예측평균수명과 장래인구추계 평균수명 비교

	남자						여자					
	추계	LC	LM	BMS	HU	HP8	추계	LC	LM	BMS	HU	HP8
2010	76.15	78.25	78.16	77.97	78.18	77.14	82.88	85.59	85.68	85.21	85.56	84.70
2011		78.77	78.66	78.43	78.70	77.52		86.02	86.14	85.61	85.87	85.24
2012		79.30	79.16	78.89	79.20	77.91		86.44	86.59	86.02	86.20	85.58
2013		79.82	79.66	79.36	79.74	78.30		86.87	87.05	86.42	86.52	85.95
2014		80.36	80.17	79.82	80.26	78.71		87.29	87.51	86.82	86.84	86.29
2015	77.11	80.90	80.69	80.28	80.78	79.11	83.80	87.72	87.96	87.23	87.18	86.60
2016		81.45	81.21	80.74	81.30	79.54		88.15	88.42	87.64	87.50	87.01
2017		82.00	81.74	81.21	81.83	79.96		88.58	88.88	88.04	87.85	87.37
2018		82.56	82.27	81.67	82.40	80.39		89.02	89.35	88.45	88.19	87.70
2019		83.13	82.81	82.14	83.00	80.83		89.46	89.81	88.87	88.54	88.05
2020	78.04	83.71	83.36	82.61	83.65	81.27	84.68	89.90	90.28	89.28	88.89	88.43
2021		84.30	83.92	83.08	84.32	81.71		90.35	90.76	89.70	89.25	88.76
2022		84.90	84.49	83.55	85.00	82.15		90.80	91.23	90.12	89.61	89.12
2023		85.51	85.06	84.03	85.64	82.59		91.25	91.71	90.54	89.96	89.50
2024		86.13	85.65	84.51	86.22	83.02		91.71	92.20	90.97	90.32	89.82
2025	78.93	86.76	86.24	84.99	86.73	83.42	85.50	92.18	92.69	91.40	90.68	90.05
2026		87.41	86.85	85.48	87.20	83.82		92.65	93.19	91.83	91.03	90.41
2027		88.07	87.47	85.98	87.67	84.17		93.12	93.69	92.27	91.39	90.68
2028		88.74	88.10	86.47	88.18	84.48		93.61	94.19	92.72	91.76	90.83
2029		89.43	88.75	86.98	88.77	84.76		94.10	94.71	93.17	92.12	90.99
2030	79.79	90.13	89.41	87.48	89.45	84.99	86.27	94.59	95.23	93.62	92.50	91.17

장래인구추계의 예측평균수명은 5년 단위로만 제공되기 때문에 1년 단위로 제시할 수 없었다. 표 5.3의 결과를 보면 2030년 예측 평균수명은 남자의 경우 LC 모형으로 예측한 평균수명이 90.13세로 가장 높게 나타났고, 여자의 경우 LM 모형으로 예측한 평균수명이 95.23세로 가장 높게 나타났다. 남녀 모두 장래인구추계 예측결과와 비교하였을 때, LC 모형과 개선 모형들을 이용하여 예측한 평균수명과 크게는 9~10세 정도, 작게는 6~8세 정도 차이가 발생하였다. 이를 정리해보면 남녀 모두 통계청의 결과가 평균수명을 가장 낮게 예측하고 있으며, LC 모형과 이의 개선모형들이 통계청 결과에 비하여 높게 예측하고 있다. HP8 모형에 의해 예측된 평균수명은 통계청의 예측수명과 LC 및 개선된 모형들에 의한 예측평균수명들의 사이에 위치한다. 통계청의 예측평균수명은 이미 2008년부터 실제평균수명과 어긋나기 시작했으므로 과소추정된 것이 확실하다고 할 수 있다. LC 및 개선 모형들에 의한 결과 역시 현재의 평

평균수명 증가속도가 2030년까지 그대로 반영된 것이므로 과대추정을 의심할 수 있다. 그 이유는 평균수명의 증가는 노인인구의 증가로 연결되며, 이는 일정시점 이후 사망자의 증가로 나타나고, 이로 인해 평균수명의 증가속도가 향후 둔화할 것임을 예상할 수 있기 때문이다. 그러므로 표 5.3의 결과로만 판단할 때는 HP8 모형에 의한 예측평균수명이 상대적으로 가장 높은 타당성을 갖는 것으로 볼 수 있다.

6. 결과 정리

본 논문에서는 예측평균수명 작성 시 사망률예측모형의 영향력을 평가하기 위하여 LC 모형을 비롯한 총 5개의 모형을 비교하였고, 우리나라 사망률자료를 이용하여 모형의 적절성을 평가하고 2010~2030년 남녀별 평균수명을 예측하였다. 이 과정에서 통계청의 2006년 ‘장래인구추계’에 포함된 예측평균수명이 현재의 평균수명 증가속도를 반영하지 못하고 있음을 확인하였다. 통계청의 예측평균수명에 비하여 5개 사망률예측모형에 의한 2030년까지의 남녀별 예측평균수명은 모두 높게 나타나고 있어 상대적으로 현재의 평균수명 증가속도를 잘 반영하고 있는 것으로 나타났다. 그러면 5개 모형에 의한 예측평균수명은 얼마나 타당한 값일까? 5절에서도 언급한 것처럼 앞으로 평균수명의 증가로 늘어난 고령인구들의 사망률 역시 상대적으로 증가할 것이 예상되므로 최근 평균수명의 증가속도가 그대로 유지되지는 못할 것으로 예상된다. 따라서 HP8 모형에 의한 평균수명이 다소 낮게 예측된 것을 제외하고는 나머지 4개 모형(LC, LM, BMS, HU)에 의한 예측평균수명은 과대추정된 것으로 보인다. 그러나 5개 모형은 식 (3.1), (5.1)의 개선을 통해 새로운 예측평균수명을 작성할 수 있으므로 과대추정문제에 대한 조정은 충분히 가능할 것으로 판단된다. 지금까지 통계청에서 작성한 대부분의 예측평균수명은 실제 평균수명보다 과소추정되었다. 따라서 이 연구결과가 2011년 새로운 장래인구추계를 앞두고 있는 통계청이 오류의 반복을 피하는 데 도움이 될 것으로 기대한다.

참고문헌

- 국가통계포털. <http://www.kosis.kr>
- 권태환, 김태현 (1990). <한국인의 생명표>, 서울대학교 출판부.
- 김기환, 이동희, 정승환 (2006). 간이생명표 확장기법을 통한 사망확률 추정, <Journal of the Korean Data Analysis Society>, **8**, 959-969.
- 김기환, 정승환 (2006). 우리나라 생명표의 연령구간 확장 및 기대여명 예측, <Journal of the Korean Data Analysis Society>, **8**, 1723-1733.
- 박유성, 김기환, 이동희, 이연경 (2005). 사망률 예측을 위한 모형 비교, <응용통계연구>, **18**, 639-654.
- 통계청 (2006). 장래인구추계.
- Booth, H., Hyndman, R. J., Tickle, L. and De Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions, *Demographic Research*, **15**, 289-310.
- Booth, H., Mandonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline, *Population Studies*, **56**, 325-336.
- Booth, H., Tickle, L. and Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison, *New Zealand Population Review*, **31**, 13-34.
- Brass, W. (1971). *On the Scale of Mortality*, Biological Aspects of Demography, Taylor and Francis, London.
- Heligman, L. and Pollard, J. H. (1980). The Age pattern of mortality, *Journal of the Institute of Actuaries*, **107**, 49-77.
- Hyndman, R. J., Booth, H., Tickle, L. and Mandonald, J. (2008). *Demography: Forecasting Mortality and Fertility Data*, R package version 0.995.
- Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rate: A functional data approach, *Computational Statistics & Data Analysis*, **51**, 4942-4956.

- Kostaki, A. (1991). The Heligman- Pollard Formula as a tool for expanding an abridged life table, *Journal of Official Statistics*, **7**, 311-323.
- Kostaki, A. and Panousis, V. (2001). Expanding an abridged life table, *Demographic Research*, **5**, 1-22.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting the time series of 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.
- United Nations (1982). *Model Life Tables for Developing Countries*, United Nations publication.

A Comparison Study for Mortality Forecasting Models by Average Life Expectancy

Seunghwan Jeong¹ · Kee-Whan Kim²

¹Department of Economics and Statistics, Korea University

²Department of Informational Statistics, Korea University

Abstract

By use of a mortality forecasting model and a life table, forecasting the average life expectancy is an effective way to evaluate the future mortality level. There are differences between the actual values of average life expectancy at present and the forecasted values of average life expectancy in population projection 2006 from Statistics Korea. The reason is that the average life expectancy forecasts did not reflect the increasing speed of the actual ones. The main causes of the problem may be errors from judgment for projection, from choice, or use of a mortality forecasting model. In this paper, we focus on the choice of the mortality forecasting model to inspect this problem. Statistics Korea should take a mortality forecasting model with considerable investigation to proceed population projection 2011 without the errors observed in population projection 2006. We compare the five mortality forecasting models that are the LC(Lee and Carter) model used widely and its variants, and the HP8(Heligman and Pollard 8 parameter) model for handling death probability. We make average life expectancy forecasts by sex using modeling results from 2010 to 2030 and compare with that of the population projection 2006 during the same period. The average life expectancy from all five models are forecasted higher than that of the population projection 2006. Therefore, we show that the new average life expectancy forecasts are relatively suitable to the future mortality level.

Keywords: Mortality forecasting model, average life expectancy, life table, Lee-Carter model, HP8 model.

This work was supported by the Korea University Grant.

²Corresponding author: Professor, Department of Informational Statistics, Korea University, 208, Seochang-Ri, Jochiwon-Eup, Yeonki-Gun, Chung-nam 339-700, Korea. E-mail: korpen@korea.ac.kr