

포아송 로그-이중선형 모형을 이용한 한국 남성 사망률의 베이지안 추정 및 활용

The Bayesian Estimation on Korean Male Mortality Rates Using Poisson Log-Bilinear Model and its Application

황 지 연* · 고 방 원**

Ji-Yeon Hwang · Bangwon Ko

본 연구는 최근 각광받고 있는 베이지안 통계기법을 이용해 한국 남성의 장래사망률을 추정하고, 베이지안 관점에서 기대여명을 비롯한 종신연금과 종신보험의 보험수리적 현가를 분석한 것이다. 시간이 경과함에 따라 사망률이 개선되고 있는 현상을 확률적으로 모델링하기 위해, 전통적인 Lee-Carter 모형을 보완한 포아송 로그-이중선형 모형을 가정했고, Czadob 외 2인(2005)에서 제안된 베이지안 방법론을 적용하여 장래사망률을 추정했다. 추정방법별 추정치들의 차이를 분석하기 위해 특이값 분해 및 최대가능도를 이용한 전통적 추정방법을 병행해 결과를 비교했다. 추정결과에 따르면, 추정방법에 의한 장래사망률의 차이는 크지 않으나, 전통적인 추정방법의 경우 장래사망률의 신뢰구간을 과소평가하는 경향이 있는 것으로 나타났다. 본 연구에서 사용된 베이지안 추정방법은 이러한 문제에 대한 해결책을 제시할 뿐만 아니라 향후 보험부채의 평가에도 유용하게 활용될 수 있을 것으로 기대한다.

국문 색인어: Lee-Carter 모형, 베이지안 추정, 장래사망률, 포아송 로그-이중선형 모형
한국연구재단 분류 연구분야 코드: C030805

* 숭실대학교 정보통계보험수리학과 박사과정(hjyn1203@ssu.ac.kr), 제1저자

** 숭실대학교 정보통계보험수리학과 조교수(bko@ssu.ac.kr), 교신저자

논문 투고일: 2015. 05. 26, 논문 최종 수정일: 2015. 10. 05, 논문 게재 확정일: 2016. 02. 15

I. 서론

통계청 조사에 따르면 우리나라는 2000년대 초반 이미 65세 이상 인구비율이 7%를 넘어서는 고령화 사회에 진입했고, 앞으로도 고령화 현상은 지속되어 2026년에는 65세 이상 인구비율이 20%에 이르는 초고령화 사회에 진입할 것으로 전망된다. 사망률의 개선은 수명연장이라는 긍정적인 측면도 가지고 있지만, 개인의 노후재정이나 보험사의 재무건전성에 심각한 부작용을 초래할 수 있기 때문에 최근 국내외에서 가장 중요한 사회·경제적 이슈로 부각되고 있다. 따라서 이러한 장수위험에 효과적으로 대응하기 위한 활발한 연구가 진행 중이며, 그 중에서도 장래사망률을 모델링하기 위한 연구는 다른 연구에 선행되어야 하는 중요한 선결과제라 할 수 있다.

장래사망률의 연구에서 최근 가장 널리 사용되고 있는 것은 Lee-Carter 모형(1992, 이하 LC모형)이다. LC모형은 모수추정이 용이하고, 단순하지만 사망률변동에 대한 설명력이 우수하기 때문에 많은 각광을 받아왔다. 이후 LC모형의 단점으로 거론된 문제점을 약간 보완한 유사 LC모형들¹⁾도 많이 제안되었는데, 그 중에서는 포아송 로그-이중선형 모형(Poisson log bilinear model)²⁾이 널리 사용되는 모형 중 하나이다. 포아송 로그-이중선형 모형은 포아송 회귀모형(Poisson regression model)과 LC모형을 결합한 것으로 기존 LC모형의 단점으로 지적된 오차의 등분산성 문제를 해결하기 위해 Wilmoth(1993)와 Alho(2000)가 제안하고, Brouhns 외 2인(2002)이 최대가능도 추정방법(MLE; Method of Maximum Likelihood Estimation)을 이용해 추정한 것이다.

최근 국내에서도 이러한 해외연구를 따라 LC모형 또는 유사 LC모형을 적용한

1) 가령, 코호트 효과를 반영한 Renshaw-Haberman 모형(2006), 사회·경제적 환경이 유사한 국가나 그룹의 사망률 패턴이 유사할 것이라는 가정을 반영한 Li-Lee 모형(2005) 등을 들 수 있다. Cairns 외 6인(2009)은 실증분석을 통해 유사 LC모형들을 포함한 8개의 사망률 모형을 비교했다.

2) LC모형과 포아송 로그-이중선형 모형은 차이가 크지 않기 때문에 두 모형을 구분하지 않고 포아송 로그-이중선형 모형을 LC모형으로 언급하는 선행연구도 많이 있으나, 본 연구에서는 모형과 추정방법에 따른 차이를 조사하기 위해 두 모형을 구분하여 기술했다.

연구결과들을 찾아볼 수 있다. 가령, 김세중(2012)은 LC모형과 포아송 로그-이중선형 모형을 이용해 국내의 장래사망률을 추정했고, 백혜연 외 2인(2013)은 전통적 LC모형의 추정치를 바탕으로 생명보험 상품의 보험수리적 현가를 산출했다. 또한, 주효찬·이항석(2014)은 포아송 로그-이중선형 모형을 활용해 사망률의 개선이 생명보험 부채평가에 미치는 영향을 분석했다. 한편, 박유성 외 2인(2013)은 국내 초고령 사망률 자료의 신뢰성 문제를 언급하며 벡터오차수정 LC모형의 사용을 제안했다.

이러한 최근 노력에도 불구하고 국내에서는 아직 베이지안 기법을 활용한 LC모형 또는 유사 LC모형에 대한 추정은 시도되지 않고 있다. 기본적으로 베이지안 통계기법은 모수의 사전확률분포(prior distribution)를 이용하여 불확실성을 반영한 뒤, 자료를 통해 사후확률분포(posterior distribution)를 구하고, 모수에 대한 통계적 추론을 도출해내는 통계적 방법론이다. 베이지안 추정방법은 최대가능도 추정과는 달리 추정 알고리즘의 초기값에 민감하게 반응하지 않으며, 모수를 포함한 복잡한 비선형함수의 추론에 용이하고, 내포되지 않는(nonnested) 모형들 간의 비교를 허용하는 등 많은 장점을 가지고 있다.

따라서 해외의 경우 베이지안 방법론에 바탕을 둔 장래사망률의 추정이나 응용에 관한 연구가 많이 진행되고 있다. 가령, Czado 외 2인(2005)은 계층적 베이지안 모형을 통해 포아송 로그-이중선형 모형의 모수들을 추정했고, Pedroza(2006)는 LC모형에서 모수의 추정오차를 반영하지 않을 경우 장래사망률에 대한 예측구간이 비정상적으로 좁아진다는 것을 베이지안 관점에서 논의하였다. 또한 Kogure 외 2인(2009)은 베이지안 방법을 이용해 일본의 사망률 모형을 비교하고 장수위험을 평가했고, Kogure and Kurachi(2010)는 이를 장수채권(survivor bond), 장수스왑(survivor swap)과 같은 장수 파생상품의 가격산출에 활용했다.

본 연구는 포아송 로그-이중선형 모형하에서 Czado 외 2인(2005)에서 제안된 베이지안 방법론을 적용하여 국내 남성의 장래사망률을 추정한 것이다. 또한, 추정된 장래사망률을 바탕으로 종신보험이나 종신연금의 보험수리적 현가에 대한 사후분포를 구해 장래사망률의 개선현상이 보험부채의 평가에 미치는 영향을 베이

지안 관점에서 분석했다. 즉, 베이지안 관점에서 장래사망률의 불확실성을 고찰하고, 보험계리 분야에서의 활용가능성을 타진해 봄으로써 기존 국내연구와는 차별성을 가진다고 할 수 있다.

한편, Lee and Carter(1992)에서 언급된 바와 같이 LC모형은 모수 추정치의 안정성을 확보하기 위해 충분히 긴 기간(적어도 30년 이상)의 자료를 필요로 한다. 국내의 경우 신뢰할 만한 자료는 1983년 이후부터 통계청에서 제공되고 있기 때문에 모수 추정치의 안정성을 확보하기 위한 경계에 놓인 것으로 보인다. 따라서 기존의 추정방법과는 다른 베이지안 방법을 통해 기존의 추정결과를 다시 확인해 보는 것도 의미 있는 연구라 할 수 있다. 또한, 조만간 도입이 예정된 보험회계기준 개정안(IFRS Phase2)은 보험부채평가에 위험조정액을 반영하도록 요구하는 바, 베이지안 접근을 통해 평가대상의 사후확률분포를 도출한다면 이를 위험조정액 산출 등에 활용할 수 있을 것으로 기대한다.

본 논문은 다음과 같이 구성되었다. 2장은 LC모형과 포아송 로그-이중선행 모형을 소개하고, 베이지안 추정과 관련된 이론적 배경, 계층적 베이지안 모형의 구조, 고령자 사망률을 포함하기 위한 생명표의 확장 등을 다루었다. 3장은 추정방법에 따른 추정치의 차이를 비교하기 위해, 베이지안 추정치와 더불어 특이값 분해(SVD; Singular Value Decomposition) 및 최대가능도를 이용한 추정결과를 제시했고, 이를 바탕으로 예측정확도를 비교했다. 4장은 장래사망률의 개선이 기대여명과 보험수리적 현가를 산출하는 데 미치는 영향을 사후분포와 최고사후밀도구간(HPDI; Highest Posterior Density Interval)을 통해 분석했고, 마지막으로 5장은 향후 연구과제와 함께 본 논문을 마무리했다.

II. 이론적 배경

1. 전통적 LC모형과 포아송 로그-이중선형 모형

Lee and Carter(1992)에서 제안된 전통적인 LC모형은 다음과 같이 연령(x)과 시간(t)효과를 결합하여 장래사망률의 추이를 확률적으로 모델링한 것이다.

$$\ln m_{x,t} = \alpha_x + \beta_x k_t + \varepsilon_{x,t}. \quad (1)$$

여기서 $m_{x,t}$ 는 t 년도 x 세인 사람의 중앙사망률(central death rate)을 나타낸다. 일반적으로 중앙사망률은 $E_{x,t}$ 가 t 년도 x 세 평균 인구 수이고, $D_{x,t}$ 가 t 년도 x 세 사망자 수를 나타낸다고 할 때 $m_{x,t} = D_{x,t} / E_{x,t}$ 로 정의된다. 정수 연령구간에 서 사력($\mu_{x,t}$)이 상수인 경우 $m_{x,t} = \mu_{x,t}$ 가 성립하기 때문에 (1)의 좌변을 중앙사망률 대신 사력으로 나타내기도 한다. 한편 $\varepsilon_{x,t}$ 는 평균이 0이고 분산이 동일한 오차항을 나타내고, α_x 는 연령에 따른 전반적인 사망률 수준을, β_x 는 사망률 개선에 따른 연령별 민감도를, 그리고 κ_t 는 시간경과에 따른 사망률의 개선추세를 의미한다. 일반적으로 모수의 식별가능성(identifiability)을 확보하기 위해 다음의 제약조건이 부과된다.

$$\sum_t \kappa_t = 0, \quad \sum_x \beta_x = 1.$$

Lee and Carter(1992)에서 제안된 추정방법은 특이값 분해를 이용해 모수의 추정치를 구한 뒤, 시계열 분석을 활용해 κ_t 에 대한 모델링과 모수의 추정치를 구하는 것이다. 보다 자세한 추정방법은 박유성 외 3인(2005), 김세중(2012) 등 국내 논문에 소개되어 있으므로 생략한다.

전통적인 LC모형은 각 연령별 로그사망률의 오차항이 등분산성을 만족한다고 가정하고 있는데 이것은 다소 비현실적인 가정이다. 이는 일반적으로 고연령대에서 사망률의 변동성이 커지는 경향을 보이기 때문이다. 따라서 Wilmoth(1993)와

Alho(2000)는 오차항의 등분산성 가정의 문제점을 보완하기 위해 다음과 같은 포아송 로그-이중선형모형을 제안하였다.

$$D_{x,t} \sim \text{Poisson}(E_{x,t} m_{x,t}) \text{ with } \ln m_{x,t} = \alpha_x + \beta_x \kappa_t. \quad (2)$$

식 (2)에서 볼 수 있듯이 포아송 로그-이중선형모형은 LC모형에 포아송 회귀모형을 결합하여 LC모형의 장점을 유지하며, 등분산성의 문제를 해결하도록 고안된 것이다. 이 경우 최대가능도 추정을 위한 가능도함수는 다음과 같다.

$$L(\alpha, \beta, \kappa) \propto \prod_x \prod_t \exp(-E_{x,t} \exp(\alpha_x + \beta_x \kappa_t) + D_{x,t}(\alpha_x + \beta_x \kappa_t)).$$

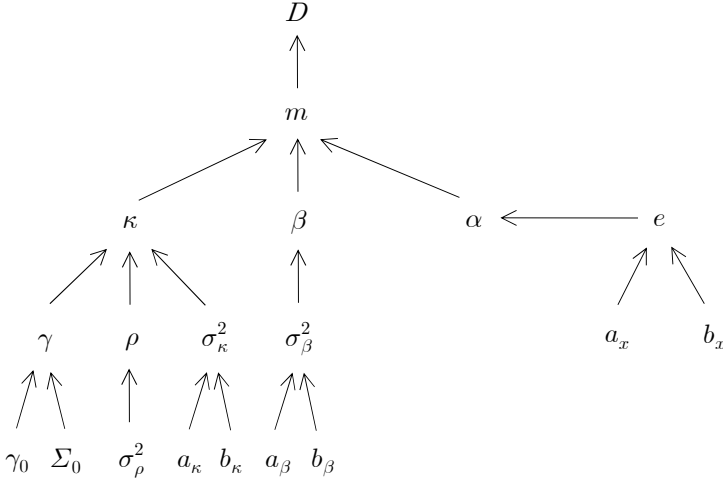
가능도함수가 매우 복잡한 비선형 함수로 표현되기 때문에 일반적인 통계프로그램을 이용해 해를 찾는 것이 쉽지 않다. 따라서 Brouhns 외 2인(2002)은 LEM이라는 범주형 데이터분석 소프트웨어를 이용해 모수를 추정했다. 이 경우 α_x , β_x , κ_t 의 추정은 Newton-Raphson 방법을 통해 순차적으로 이루어지며, 전통적인 LC모형의 특이값 분해를 이용한 추정에서 이루어졌던 κ_t 에 대한 추가적인 추정은 수행되지 않는다.

2. 베이지안 추정방법

가. 계층적 베이지안 모형의 구조

Czado 외 2인(2005)에서 제안된 계층적 베이지안 모형(hierarchical bayesian model)은 대략적으로 <그림 1>과 같은 구조를 가진다. 편의상 첨자는 생략했고, 모수의 사전분포와 사후분포에 대해서는 다음 절에서 설명한다.

〈그림 1〉 계층적 베이지안 모형의 구조



나. 사전분포

먼저 κ 의 사전분포로 다음을 가정했다(이하 굵은 글씨는 열벡터나 행렬을 나타낸다).

$$\kappa_t - (\gamma_1 + \gamma_2 t) = \rho (\kappa_{t-1} - (\gamma_1 + \gamma_2 (t-1))) + \epsilon_t, \quad t = t_{\min}, \dots, t_{\max}. \quad (3)$$

여기서 ϵ_t 는 평균이 0이고, 분산이 σ_ϵ^2 인 정규분포를 따르는 오차항이다. 즉, (3)은 시간에 대해 선형적인 추세를 가지도록 κ_t 를 모델링한 것으로, 다변량 정규분포를 이용할 경우 $\kappa \sim N(\mathbf{X}\boldsymbol{\gamma}, \sigma_\epsilon^2 \mathbf{Q}^{-1})$ 와 같이 간결하게 표현될 수 있다. 단, 여기서

$$\mathbf{X} = \begin{pmatrix} 1 & t_{\min} \\ \vdots & \vdots \\ 1 & t_{\max} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & \vdots \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}.$$

또한 초모수(Hyperparameter)에 대해서는 $\gamma \sim N(\gamma_0, \Sigma_0)$, $\sigma_\kappa^{-2} \sim \text{Gamma}(a_\kappa, b_\kappa)^3$, 그리고 ρ 는 구간 $(-1, 1)$ 에서 절단된(truncated) $N(0, \sigma_\rho^2)$ 을 따른다고 가정했다.

모수 β 의 사전 분포로 $\beta \sim N(0, \sigma_\beta^2 I_M)$ 을(단, I_M 은 $M = x_{\max} - x_{\min} + 1$ 차원의 단위행렬), 초모수 σ_β^2 의 사전분포로 역감마(inverse gamma)분포, 즉, $\sigma_\beta^{-2} \sim \text{Gamma}(a_\beta, b_\beta)$ 를 가정했다. 또한 모수 α 에 대해 $e_x = \exp(\alpha_x)$ 의 변환을 취했는데 이에 대한 사전분포로 $e_x \sim \text{Gamma}(a_x, b_x)$ 를 이용했다. 이러한 사전분포의 가정은 사후분포의 계산상 편의를 위해 베이지안 추정에서 흔히 이루어지는 것이다.

다. 사후분포

이번 절에서는 논문의 완결성(completeness)을 위해 사후분포와 모수의 추출과정을 $\kappa_t, \beta_x, \alpha_x$ 와 초모수의 순으로 요약해 제시하였다. 보다 자세한 유도과정은 Czado 외 2인(2005)을 참고하기 바란다.

(1) κ 의 사후분포

먼저 κ 와 D 에서 t 번째 성분을 제거한 열벡터를 다음과 같이 정의한다.

$$\begin{aligned}\kappa_{-t} &= (\kappa_{t_{\min}}, \dots, \kappa_{t-1}, \kappa_{t+1}, \dots, \kappa_{t_{\max}})', \\ D_t &= (D_{x_{\min}, t}, \dots, D_{x_{\max}, t})', D_{-t} = (D_{t_{\min}}, \dots, D_{t-1}, D_{t+1}, \dots, D_{t_{\max}})'. \end{aligned}$$

여기서 $()'$ 은 전치행렬을 나타낸다. 주어진 계층적 구조로부터 κ_t 에 대해 다음의 조건부 확률분포를 얻을 수 있다(유도과정 생략).

3) $\text{Gamma}(a, b)$ 에서 a 는 감마분포의 형태모수(shape parameter)를, b 는 척도모수(scale parameter)의 역수를 나타낸다.

$$f(\kappa_t | \kappa_{-t}, \alpha, \beta, D, \sigma_\kappa^2, \sigma_\beta^2, \gamma, \rho) \propto f(D_{t_{\min}} | \kappa_{t_{\min}}, \alpha, \beta) f(\kappa_{t_{\min}} | \gamma, \sigma_\kappa^2) \\ \times \prod_{s=t_{\min}+1}^{t_{\max}} f(D_s | \kappa_s, \alpha, \beta) f(\kappa_s | \kappa_{s-1}, \gamma, \sigma_\kappa^2, \rho).$$

$t = t_{\min}, t = t_{\max}, t_{\min} < t < t_{\max}$ 으로 경우를 나누어 Metropolis-Hastings 알고리즘을 이용해 κ_t 를 추출할 수 있다.

(2) β 의 사후분포

κ 의 경우와 마찬가지로 Metropolis-Hastings 알고리즘을 이용해 β 의 사후분포를 추출할 수 있다. $\beta_{-x} = (\beta_{x_{\min}}, \dots, \beta_{x-1}, \beta_{x+1}, \dots, \beta_{x_{\max}})'$ 라 할 때

$$f(\beta_x | \beta_{-x}, \alpha, \kappa, D, \sigma_\kappa^2, \sigma_\beta^2, \gamma, \rho) \\ \propto \prod_t \exp(-E_{x,t} \exp(\alpha_x + \beta_x \kappa_t)) \prod_t \exp(\beta_x \kappa_t D_{x,t}) \exp(-\frac{1}{2\sigma_\beta^2} \beta_x^2)$$

가 성립한다(유도과정 생략).

(3) α 의 사후분포

앞에서 언급한 바와 같이 α 의 경우 $e_x = \exp(\alpha_x)$ 의 변환을 거쳐 Gibbs sampling을 통해 추출할 수 있다.

$$\alpha_{-x} = (\alpha_{x_{\min}}, \dots, \alpha_{x-1}, \alpha_{x+1}, \dots, \alpha_{x_{\max}})', \\ c_x = \sum_t E_{x,t} \exp(\beta_x \kappa_t), \quad D_{x \cdot} = \sum_t D_{x,t}$$

라 할 때

$$(e_x | \beta, \kappa, D, \sigma_\kappa^2, \sigma_\beta^2, \gamma, \rho) \sim \text{Gamma}(\alpha_x + D_{x \cdot}, b_x + c_x)$$

가 성립한다(유도과정 생략).

(4) 초모수의 조건부분포

초모수의 조건부분포를 요약하면 다음과 같다.

$$(i) \quad (\rho | \alpha, \beta, \kappa, D, \sigma_{\kappa}^2, \sigma_{\beta}^2, \gamma) \sim N(\mu_{\rho}^*, \sigma_{\rho}^{2*}) \text{ truncated to } (-1, 1).$$

$$(ii) \quad (\sigma_{\kappa}^{-2} | \alpha, \beta, \kappa, D, \sigma_{\beta}^2, \gamma, \rho) \sim$$

$$Gamma\left(a_{\kappa} + \frac{T}{2}, b_{\kappa} + \frac{1}{2} \sum_t (\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right).$$

$$(iii) \quad (\sigma_{\beta}^{-2} | \alpha, \beta, \kappa, D, \sigma_{\beta}^2, \gamma, \rho) \sim Gamma\left(a_{\beta} + \frac{M}{2}, b_{\beta} + \frac{1}{2} \beta^t \beta\right).$$

$$(iv) \quad (\gamma | \kappa, \sigma_{\kappa}^2, \sigma_{\beta}^2, \rho) \sim N(\gamma^*, \sigma_{\kappa}^2 \Sigma^*).$$

단, 여기서

$$\mu_{\rho}^* = b_{\rho} / \left(a_{\rho} + \frac{\sigma_{\kappa}^2}{\sigma_{\rho}^2}\right), \quad \sigma_{\rho}^{2*} = \sigma_{\kappa}^2 / \left(a_{\rho} + \frac{\sigma_{\kappa}^2}{\sigma_{\rho}^2}\right), \quad T = t_{\max} - t_{\min} + 1,$$

$$\Sigma^* = (X' Q X + \sigma_{\kappa}^2 \Sigma_0^{-1})^{-1}, \quad \gamma^* = \Sigma^* (X' Q \kappa + \sigma_{\kappa}^2 \Sigma_0^{-1} \gamma_0)$$

이다.

3. 고령자 사망률로의 확장

본 논문에서 사용된 데이터는 통계청에서 제공하는 1983~2013년 사이의 남성 사망자 수와 추계인구이다. 이 자료는 기준 연도별로 한계연령에 차이를 가지고 있기 때문에 고령자의 사망률을 분석하기 위해 한계연령 이전의 사망률을 한계연령 이후의 사망률로 확장할 필요가 있다. 국내의 경우 김성용 외 2인(2011)과 주효

찬·이항석(2014)이 각각 Brass-logit 모형(Brass, 1977)과 Gompertz 모형을 적용해 기존의 생명표를 가지고 고령자의 사망률로 확장했다. 본 연구에서는 분석의 편의를 위해 주효찬·이항석(2014)을 따라 Gompertz 모형을 적용했다.

Gompertz의 사망법칙(Gompertz law of mortality)은 중년 이후의 사망률에 대한 횡단면적(cross sectional) 모델링을 위해 제안된 고전적인 사망률 모형이다. Gompertz 모형은 x 세의 사력에 대해

$$\mu_x = BC^x \quad (B > 0, C > 1) \quad (4)$$

을 가정하며, 이때 x 세의 사망률에 대해 $q_x = 1 - e^{B(1-C^x)/\ln C}$ 가 성립한다. 이 모형은 단순하지만 중년 이후의 자료에 대한 횡단면적 적합도가 우수하기 때문에 보험계리 분야의 최근 연구에서도 종종 사용되고 있다.

일반적으로 Gompertz 모형의 모수는 최소제곱법을 통해 추정된다. Gompertz 모형은 정수 연령구간에서 사력이 상수(constant force of mortality)임을 가정하고 있지 않지만 앞 절에서의 가정과 일치성을 위해 식 (4)가 정수 연령에서만 성립하고 나머지 연령에서는 사력이 상수임을 가정했다. 이때 $p_x = e^{-\mu_x}$ 가 성립하므로(단, x 는 정수) 다음과 같이 오차의 제곱 합이 최소가 되도록 모수를 추정할 수 있다.

$$\min \left(\sum_x \ln(-\ln p_x) - \ln B - x \ln C \right)^2$$

Gompertz 모형의 특징 및 모수추정에 관한 자세한 설명은 Carriere(1994)를 참고하기 바란다.

III. 추정결과

앞에서 언급한 바와 같이 본 연구는 통계청에서 제공하는 1983~2013년 사이의 남성 사망자 수와 추계인구자료를 이용하였다. 기준 연도별로 한계연령에 차이가 존재하기 때문에 3장에서는 공통적으로 포함되어 있는 79세 이전까지의 사망률 자료를 이용한 추정결과를 제시했다. 또한, 기존의 전통적 추정방법과 비교하기 위해 특이값 분해를 이용한 추정결과(LC모형)와 최대가능도를 이용한 추정결과(포아송 로그-이중선형 모형)를 함께 제시하였다. 이후 Gompertz 모형을 통해 기존 사망률 자료를 79세 이후로 확장하고, 모수를 재추정한 뒤 기대여명, 보험수리적 현가 등의 계산에 활용했는데 그 산출결과를 4장에 제시했다. 모수추정을 비롯한 모든 계산과정은 통계분석 패키지 R을 이용했다.

1. 모수의 추정결과

〈표 1〉~〈표 3〉은 추정방법별 모수의 추정치와 베이지안 신뢰구간에 해당하는 95% HPDI⁴⁾를 제공하고 있다. 먼저 SVD열은 전통적인 LC모형(1)하에서 특이값 분해를 이용한 추정결과이고, MLE열은 포아송 로그-이중선형 모형(2)하에서 최대가능도 방법을 이용한 추정결과이고, Bayesian열은 포아송 로그-이중선형 모형(2)하에서 베이지안 방법을 이용한 추정결과이다. 결과를 살펴보면 추정방법에 따른 모수의 추정치에는 큰 차이가 존재하지 않는 것을 확인할 수 있다. MLE의 경우 모든 추정치가 95% HPDI에 포함되었고, SVD의 경우도 거의 대부분 95% HPDI에 포함되는 것을 확인할 수 있다. 다만 몇 개의 추정치가 95% HPDI에 포함되지 않은 것은 두 추정방법에서 가정하고 있는 사망률 모형이 정확히 일치하지는 않기 때문인 것으로 판단된다. 〈그림 2〉는 몇 개의 선택된 연령과 연도에서 α_x , β_x , κ_t 의 사후확률밀도함수를, 〈그림 3〉과 〈그림 4〉는 각각 α_x 와 β_x 의 추정치(사후평균)와 95% HPDI를 그래프로 나타낸 것이다.

4) Y 가 자료를 나타낼 때 다음 조건을 만족하는 집합 C 를 θ 에 대한 $100(1-\alpha)\%$ HPDI라 한다.

(i) $P(\theta \in C|Y) \geq 1-\alpha$

(ii) 모든 $\theta_1 \in C$, $\theta_2 \notin C$ 에 대해 $P(\theta_1 \in C|Y) \geq P(\theta_2 \in C|Y)$

〈표 1〉 특이값 분해, 최대가능도, 베이지안 방법을 이용한 α_x 의 추정결과

나이	SVD	MLE	Bayesian	95% HPDI	
1	-7.19898	-7.2217	-7.22162	-7.24879	-7.19516
5	-7.76496	-7.77523	-7.77501	-7.8091	-7.73944
10	-8.19767	-8.20133	-8.20042	-8.24167	-8.16242
15	-7.66896	-7.65489	-7.65493	-7.68246	-7.62679
20	-7.04718	-7.04238	-7.042	-7.06252	-7.02202
25	-6.77632	-6.77034	-6.77015	-6.78799	-6.75357
30	-6.60175	-6.5945	-6.59453	-6.60956	-6.57877
35	-6.23146	-6.22841	-6.22837	-6.24106	-6.21533
40	-5.76439	-5.75932	-5.75933	-5.77033	-5.74946
45	-5.29727	-5.29362	-5.29367	-5.30267	-5.28465
50	-4.89974	-4.8955	-4.89552	-4.90359	-4.88787
55	-4.52918	-4.5261	-4.52609	-4.53365	-4.51888
60	-4.13916	-4.13486	-4.13486	-4.14144	-4.12775
65	-3.71494	-3.71083	-3.71093	-3.71743	-3.70446
70	-3.23425	-3.23277	-3.23279	-3.23878	-3.2267
75	-2.76906	-2.76778	-2.76785	-2.77428	-2.76161
79	-2.37841	-2.37867	-2.37868	-2.38536	-2.37163

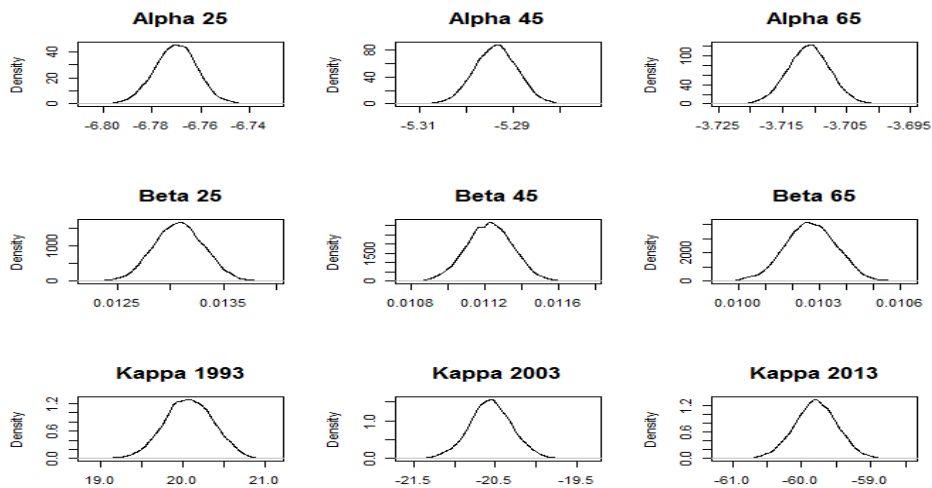
〈표 2〉 특이값 분해, 최대가능도, 베이지안 방법을 이용한 β_x 의 추정결과

나이	SVD	MLE	Bayesian	95% HPDI	
1	0.016917	0.018568	0.018565	0.017904	0.01925
5	0.021254	0.022106	0.022105	0.021223	0.022959
10	0.018546	0.019122	0.019085	0.018054	0.02006
15	0.01486	0.014222	0.014223	0.013522	0.014936
20	0.014347	0.014026	0.014017	0.013486	0.014556
25	0.013147	0.013081	0.013081	0.012626	0.013555
30	0.012938	0.012806	0.01281	0.012386	0.01325
35	0.012712	0.012669	0.012669	0.01227	0.013015
40	0.011628	0.011508	0.011513	0.011194	0.011799
45	0.011204	0.01122	0.011226	0.010974	0.011467
50	0.010239	0.010212	0.010215	0.010008	0.010432
55	0.010259	0.010174	0.010178	0.009968	0.010376
60	0.010211	0.01012	0.010126	0.009919	0.010322
65	0.010379	0.010266	0.010269	0.010083	0.010453
70	0.009228	0.009101	0.009104	0.00893	0.009275
75	0.008882	0.008779	0.008785	0.008618	0.00896
79	0.008454	0.008534	0.008538	0.008355	0.008733

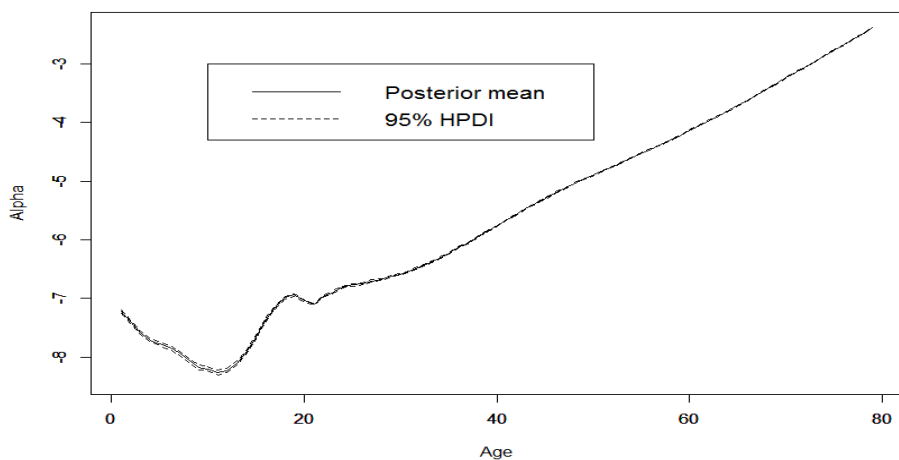
〈표 3〉 특이값 분해, 최대가능도, 베이지안 방법을 이용한 κ_t 의 추정결과

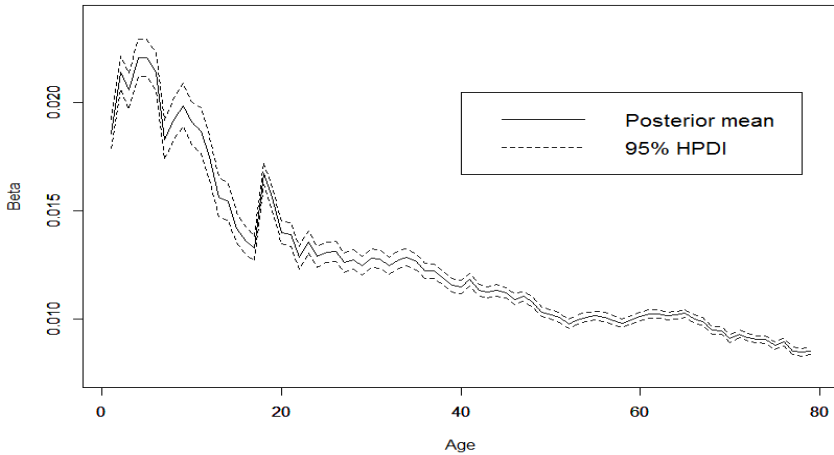
연도	SVD	MLE	Bayesian	95% HPDI	
1983	60.34562	59.95822	59.82294	59.21861	60.37556
1984	51.32354	51.44543	51.62812	51.06111	52.15294
1985	50.09923	49.96973	49.89449	49.39127	50.36573
1986	46.9106	46.61656	46.63919	46.10974	47.13658
1987	45.1395	44.89785	44.77286	44.27569	45.26122
1988	39.73693	39.80725	39.84634	39.33769	40.32109
1989	36.72456	36.76677	36.73082	36.25237	37.21619
1990	33.33951	33.32539	33.304	32.80934	33.77454
1991	29.89335	29.83457	29.75552	29.2575	30.25443
1992	24.21278	24.39479	24.41646	23.90808	24.92969
1993	19.99477	19.96716	20.06768	19.49816	20.64808
1994	19.15861	19.24221	19.12761	18.57833	19.68626
1995	15.39774	15.34507	15.32248	14.81651	15.803
1996	11.20649	11.40524	11.36463	10.88971	11.86315
1997	6.712393	6.637769	6.654651	6.136031	7.143168
1998	3.079118	2.687913	2.632478	2.133724	3.106152
1999	-1.90052	-2.55829	-2.52431	-3.02162	-2.03906
2000	-5.94087	-6.51102	-6.58008	-7.09269	-6.03574
2001	-11.7269	-12.4396	-12.3744	-12.9018	-11.8453
2002	-16.0819	-16.7811	-16.7481	-17.248	-16.2391
2003	-20.1093	-20.5646	-20.5561	-21.0935	-20.0293
2004	-23.8867	-24.5844	-24.6117	-25.154	-24.0691
2005	-29.0791	-29.7901	-29.7837	-30.3588	-29.2089
2006	-34.4569	-35.3524	-35.2593	-35.8501	-34.6845
2007	-38.6211	-39.3328	-39.2847	-39.8368	-38.7542
2008	-42.0799	-42.7176	-42.6876	-43.2167	-42.1468
2009	-45.7079	-46.046	-45.9859	-46.5756	-45.4026
2010	-47.6475	-48.1012	-48.1568	-48.7164	-47.5837
2011	-51.9286	-52.4645	-52.3801	-52.9343	-51.7704
2012	-54.4785	-55.2105	-55.2521	-55.8145	-54.6745
2013	-59.147	-59.8479	-59.7954	-60.4113	-59.1804

〈그림 2〉 $\alpha_x, \beta_x, \kappa_t$ 의 사후확률밀도함수:
 ($x = 25, 45, 65, t = 1993, 2003, 2013$)



〈그림 3〉 α_x 의 사후평균과 95% HPDI



〈그림 4〉 β_x 의 사후평균과 95% HPDI

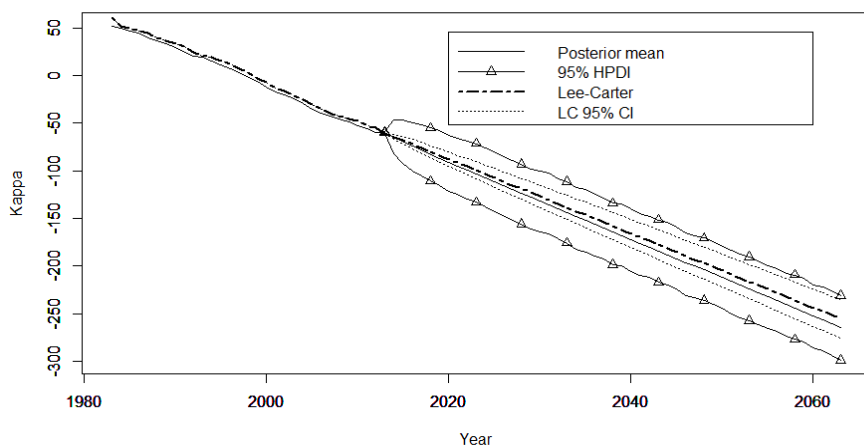
2. 장래사망률의 추정결과

전통적인 추정방법과 달리 베이지안 방법은 κ_t 뿐만 아니라 α_x 와 β_x 의 불확실성을 모두 반영해 장래사망률의 사후분포를 도출하기 때문에 이를 이용하면 장래사망률(q_x)의 사후평균과 HPDI를 쉽게 얻을 수 있다. 〈그림 5〉는 κ_t 의 추정치와 95% 신뢰구간(또는 95% HPDI)을 각각 특이값 분해와 베이지안 방법으로 구하여 그래프로 나타낸 것이다. 〈그림 5〉를 통해 추정방법에 따른 κ_t 의 추정치(Posterior mean과 Lee-Carter로 표시됨)에는 큰 차이가 없지만, 95% 신뢰구간(HPDI와 CI로 표시됨)에는 상당한 차이가 존재하는 것을 확인할 수 있다. 이러한 현상은 LC모형의 단점 가운데 하나로 지적된 것으로 κ_t 의 미래치가 절편(drift)에 의해 거의 결정됨으로써 κ_t 의 신뢰구간이 너무 좁게 나온 것에 기인한다⁵⁾. 최대가능도를 이용하는 경우는 특이값 분해를 이용하는 경우와 큰 차이를 보이지 않았기 때문에 그래프에 표시하지 않았다. 〈그림 6〉은 특이값 분해와 베이지안 방법으로 추정된 κ_t , α_x , β_x 를 바탕으로 현재 45세인 남성의 장래사망률을 구하여 그래프로 나타내었다. 베이지안 방법을 이용할 경우 κ_t 의 신뢰구간에 대한 차이와 더불어 모수 α_x

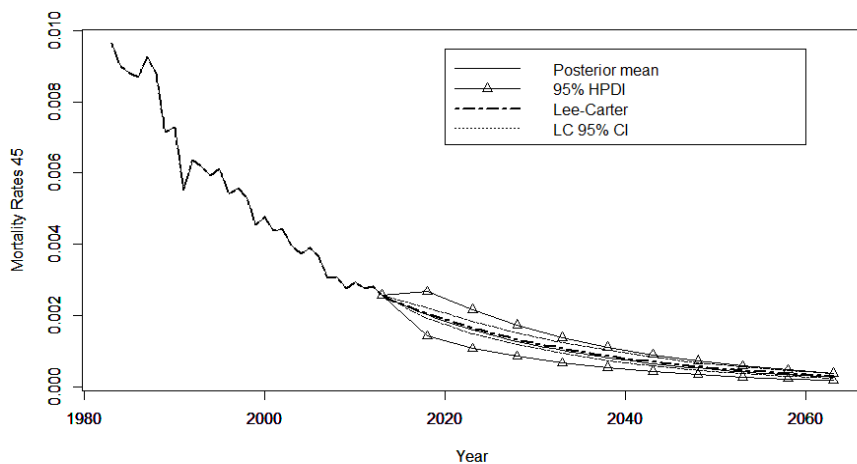
5) 박유성 외 2인(2013)을 참고하기 바란다.

와 β_x 에 대한 불확실성이 추가되기 때문에 장래사망률의 신뢰구간이 훨씬 넓어지는 것을 확인할 수 있다. <그림 7>은 45세 남성의 장래사망률에 대한 사후밀도 함수의 그래프를 20년 간격(2014년, 2034년, 2054년)으로 나타내었다.

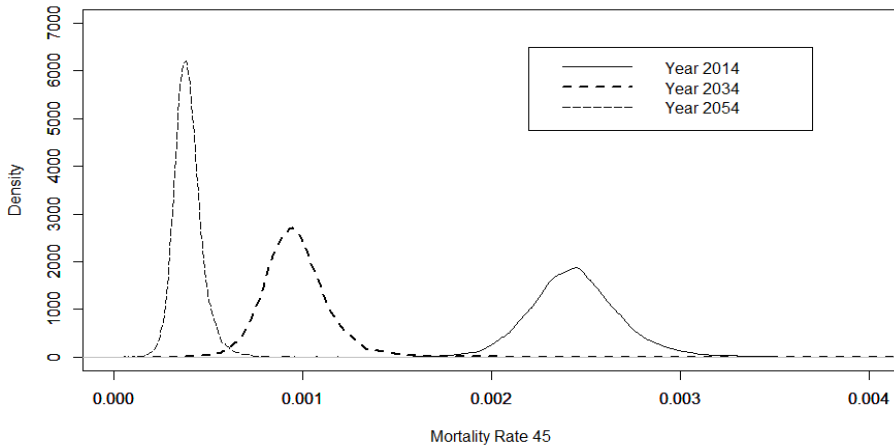
<그림 5> κ_t 의 사후평균과 95% 신뢰구간



<그림 6> 45세 남성 장래사망률의 추정치와 95% 신뢰구간



〈그림 7〉 45세 남성의 장래사망률에 대한 사후밀도함수



3. 추정방법에 따른 예측정확도의 비교

이번 절에서는 1983~2013년의 통계청 자료를 1983~2003년과 2004~2013년의 두 부분으로 나누어 사후검증(backtesting)한 결과를 제시한다. 추정방법별 예측의 정확도(prediction accuracy)를 비교하기 위해 먼저 1983~2003년 사망률 자료를 가지고 장래사망률을 추정한 뒤 2004~2013년 실제 자료에 대한 MAPE(Mean Absolute Percentage Error)⁶⁾ 값을 계산하였다. 이를 〈표 4〉에 정리했는데, 최대가능도 방법, 베이지안 방법, 특이값 분해의 순서로 예측의 정확도가 우수한 것으로 평가되었다. 그러나 이 차이가 통계적으로 의미가 있어 보이지는 않는다. 이보다 〈그림 8〉을 통해 앞에서 언급한 신뢰구간의 과소평가 문제를 재확인해 볼 수 있다. MAPE를 계산할 때와 마찬가지로 1983~2003년 자료를 가지고 각각 특이값 분해와 베이지안 방법으로 장래사망률의 95% 신뢰구간과 95% HPDI를 구한 뒤, 이를 2004~2013

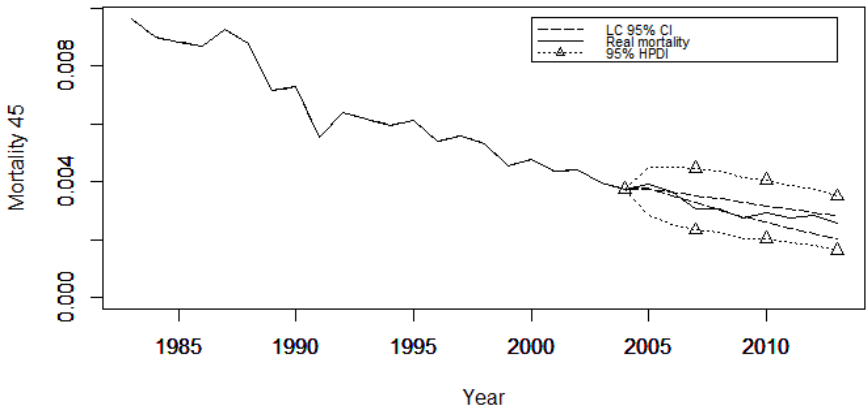
6) $MAPE = \frac{1}{n} \sum_{x,t} \left| \frac{q_{x,t} - \hat{q}_{x,t}}{q_{x,t}} \right|$ 으로 정의된다. 여기서 $q_{x,t}$ 는 실제 사망률을, $\hat{q}_{x,t}$ 는 추정된 미래사망률을, n 은 자료의 개수를 나타낸다. MAPE는 단순하지만 직관적으로 설득력을 가지고 있기 때문에 통계모형이나 추정방법에 따른 예측정확도를 평가하기 위해 많이 사용되는 측도 가운데 하나이다.

년의 실제 사망률(실선으로 표시됨)과 함께 <그림 8>에 표시하였다. 그림을 통해 특이값 분해를 이용한 신뢰구간의 경우 신뢰구간을 벗어나는 실제 사망률의 개수가 기대치보다 많아 신뢰구간이 과소평가되고 있음을 확인할 수 있었다⁷⁾.

<표 4> 특이값 분해, 최대가능도, 베이지안 추정방법에 대한 MAPE 비교

	SVD	MLE	Bayesian
MAPE	0.1028916	0.1023678	0.1023942

<그림 8> 45세 남성의 실제 사망률과 신뢰구간의 비교



IV. 추정결과의 활용방안

이번 장에서는 추정된 장래사망률이 기대여명을 비롯하여 종신연금과 종신보험의 보험수리적 현가 산출에 미치는 영향을 분석하여 보험계리 분야에서의 향후 활용가능성을 살펴보았다. 먼저 Gompertz 모델을 통해 기존의 생명표(1983~2013년,

7) 특이값 분해의 경우 대략 실제 사망률 가운데 대략 70%가 신뢰구간에 포함되었고, 베이지안 방법의 경우 대략 98%가 HPDI에 포함되었다.

79세까지)를 초고령자를 포함한 생명표(119세까지)로 확장한 뒤, 이를 가지고 포아송 로그-이중선형 모형의 모수를 재추정하고, 장래사망률의 예측치를 구해 분석에 활용하였다. 특히 분석과정에서 도출한 사후분포를 함께 제시하여 베이지안 추정이 가지는 차별성을 나타냈다.

1. 기대여명

기대여명은 현재 x 세인 생존자의 잔여수명에 대한 기댓값이다. 이번 절에서는 기대여명의 값으로 다음과 같이 정의된 단축기대여명(또는 개산기대여명, curtate life expectancy)을 고려했다.

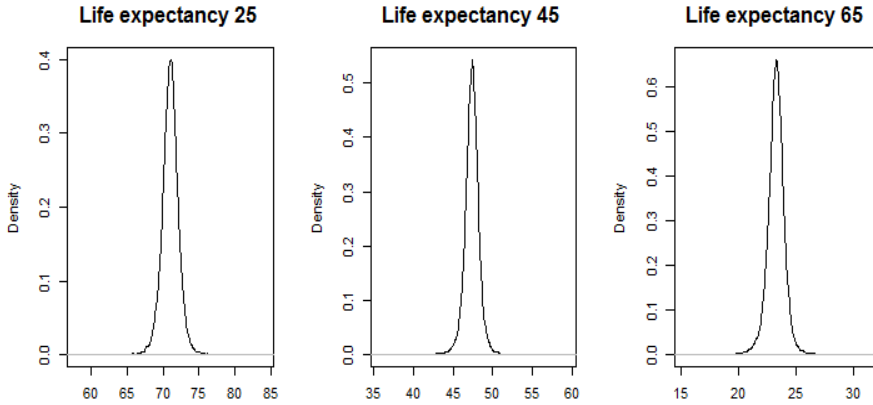
$$e_x = \sum_{k=1}^{\infty} {}_k p_x$$

여기서 ${}_k p_x$ 는 x 세인 생존자가 k 년 시점에 생존해 있을 확률이다. 장래사망률의 개선효과를 반영한 경우와 그렇지 않은 경우로 나누어 25세, 45세, 65세의 기대여명의 추정치(사후평균)를 각각 계산하고, 이를 <표 5>에 정리하였다. 가령, 현재 통계청 자료에 따르면 45세의 기대여명은 대략 35년⁸⁾인 반면, 시간에 따른 장래사망률의 개선효과를 반영하면 기대여명이 47년으로 증가한다. 베이지안 방법으로 기대여명의 사후확률밀도함수를 구하여 이를 <그림 9>에 나타냈다.

<표 5> 단축기대여명(e_x)의 사후평균과 95% HPDI

연령(x)	현행	개선 반영	95% HPDI	
25	54,11082	71,0064	68,61508	73,25366
45	35,26142	47,34397	45,47682	49,14476
65	18,17641	23,29988	21,75295	24,75617

8) 본 논문의 경우 Gompertz 모형으로 생명표를 확장하고, 단축기대여명을 사용했기 때문에 통계청에서 발표되는 완전기대여명(complete life expectancy)과는 약간의 차이가 있다.

〈그림 9〉 단축기대여명(e_x)의 사후밀도함수

2. 생명보험과 생명연금

이번 절은 장래사망률의 개선효과가 생명연금과 생명보험의 보험수리적 현가(actuarial present value)에 미치는 영향을 조사하기 위해 다음과 같은 단순한 계약을 분석했다.

종신연금: 가입 이후 매년 초에 연금 1을 지급

종신보험: 사망이 발생한 연도 말에 보험금 1을 지급

앞에서와 마찬가지로 현행 사망률이 유지될 경우와 장래사망률의 개선효과를 반영할 경우로 나누어 보험수리적 현가를 산출했는데 이자율(effective rate of interest)은 3%의 고정금리를 이용했다. 또한 비용과 해약은 발생하지 않는다는 단순한 가정하에서 보험수리적 현가를 산출했다. <표 6>과 <표 7>에 종신연금과 종신보험의 보험수리적 현가에 대한 추정치(사후평균)와 95% HPDI 계산결과를 제시하였다. 종신연금의 경우 장래사망률의 개선효과를 반영할 때 나이가 증가함에 따라 보험수리적 현가의 증가가 상대적으로 커지는 것을 확인할 수 있다. 즉, 연금수령인의 나이가 25세인 경우 종신연금 보험수리적 현가의 상대적인 증가율이 6.7%이나 65세인 경우 16.3%로 증가한다. 따라서 종신연금의 경우 고령에서 장래

사망률의 개선이 큰 영향을 미친다고 할 수 있다. <표 7>을 통해 종신보험의 경우에는 사망률의 개선효과가 반대로 나타나는 것을 확인할 수 있다. 앞서와 마찬가지로 <그림 10>과 <그림 11>을 통해 종신연금과 종신보험의 보험수리적 현가에 대한 사후확률밀도함수를 그래프로 나타내었다.

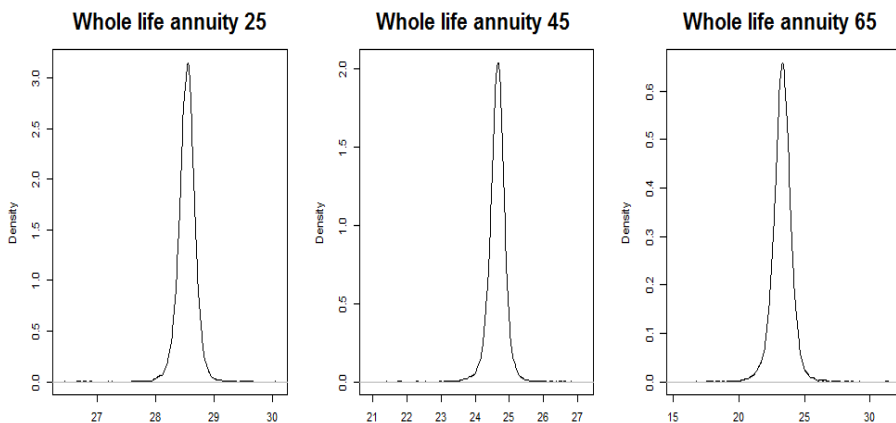
<표 6> 종신연금의 보험수리적 현가($\ddot{a}_x$)에 대한 사후평균과 95% HPDI

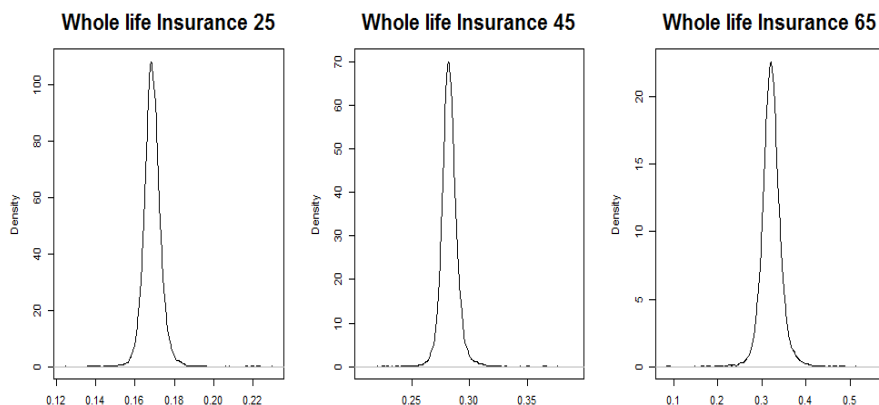
연령(x)	현행	개선 반영	95% HPDI	
25	26,73616	28,53697	28,22352	28,82286
45	21,32947	24,64108	24,13971	25,11306
65	13,69465	15,93218	15,19214	16,63992

<표 7> 종신보험의 보험수리적 현가(A_x)에 대한 사후평균과 95% HPDI

연령(x)	현행	개선 반영	95% HPDI	
25	0,221277	0,1688262	0,1604991	0,1779558
45	0,3787533	0,2822988	0,2685516	0,2969018
65	0,6011267	0,5359561	0,5153421	0,5575105

<그림 10> 종신연금의 보험수리적 현가($\ddot{a}_x$)에 대한 사후확률밀도함수



〈그림 11〉 종신보험의 보험수리적 현가(A_x)에 대한 사후확률밀도함수

V. 결론 및 향후 연구과제

본 연구에서는 최근 각광받고 있는 베이지안 통계기법을 이용해 시간이 경과함에 따라 개선되고 있는 장래사망률을 추정하고, 베이지안 관점에서 이러한 영향을 평가해 보았다. 국내의 연구 가운데 처음으로 베이지안 기법을 통해 장래사망률을 추정했다는 측면에서 의의를 가지며, 베이지안 추정의 장점을 최대한 활용하여 추정치에 대한 사후확률밀도함수와 최고사후밀도구간을 함께 제시하였다. 분석결과에 따르면 추정방법별 추정치의 차이는 크지 않으나, 전통적인 추정방법의 경우 신뢰구간을 과소평가하는 경향을 확인할 수 있었다.

한편 본 연구는 베이지안 관점에서 기본적인 LC모형만을 고려했으나, 향후 보다 다양한 모형을 가지고 분석해 보는 것도 의미 있는 후속 연구가 될 것이다. 또한, 현실적인 보험계리 가정을 추가하여 보험계약을 분석한다면 실무자를 위한 가이드라인이 될 것으로 기대한다. 마지막으로, 본 연구는 베이지안 추정에 주안점을 두기 위해 편의상 Gompertz 모형을 이용해 고연령의 사망률로 확장했지만, 향후 관련연구에서는 이에 관해 보다 세심한 주의가 필요하다고 판단된다.

참고문헌

- 김성용 · 김기환 · 박유성, “우리나라의 초고령 사망률의 확장에 대한 연구”, **조사연구**, 제12권 제2호, 2011, pp. 1-26.
- 김세중, “Lee-Carter 모형을 이용한 사망률 예측에 관한 연구”, **계리학연구**, 제4권 제2호, 2012, pp. 48-66.
- 박유성 · 김기환 · 이연경 · 이동희, “사망률 예측을 위한 모형 비교”, **응용통계연구**, 제18권 3호, 2005, pp. 639-654.
- 박유성 · 장선화 · 김성용, “사망률 추계를 위한 오차수정 LC 모형”, **조사연구**, 제14권 2호, 2013, pp. 19-47.
- 백혜연 · 노주희 · 이항석, “Lee-Carter 모형에서 사망률 추정과 보험수리적 현가분석”, **한국자료분석학회지**, 제15권 제3호, 2013, pp. 1553-1572.
- 주호찬 · 이항석, “사망률의 개선이 책임준비금에 미치는 영향”, **리스크관리연구**, 제25권 제2호, 2014, pp. 3-33.
- Alho, J. M., “Discussion of the Lee-Carter Method for Forecasting Mortality with Various Extensions and Applications”, *North American Actuarial Journal* 4(1), 2000, pp. 91-93.
- Brass, W., “On the Scale of Mortality”, *Biological Aspects of Demography*, 1977.
- Brouhns, N., Denuit, M. and Vermunt, J.K., “A Poisson Log-Bilinear Approach to the Construction of Projected Lifetables”, *Insurance: Mathematics and Economics* 31, 2002, pp. 373-393.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. and Balevich, I., “A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States”, *North American Actuarial Journal* 13(1), 2009, pp. 1-35.
- Carriere, J. F., “An investigation of the Gompertz Law of Mortality”, *Actuarial*

- Research Clearing House* 2, 1994, pp. 161-177.
- Chib, S., "Markov Chain Monte Carlo methods: Computation and Inference", *Handbook of Econometrics* 5, 2001, pp. 3569-3649.
- Czado, C., Delwarde, A. and Denuit, M., "Bayesian Poisson Log-Bilinear Mortality Projections", *Insurance: Mathematics and Economics* 36, 2005, pp. 260-284.
- Huang, H.-C., Yue, J. C., and Yang, S. S., "An Empirical Study of Monthly Models in Taiwan", *Asia Pacific Journal of Risk and Insurance* 3(1), 2008, pp. 150-164.
- Kogure, A. Kitsukawa, K. and Kurachi, Y., "A Bayesian Comparison of Models for Changing Mortalities toward Evaluating Longevity Risk in Japan", *Asia-Pacific Journal of Risk and Insurance* 3, 2009, pp. 1-22.
- Kogure, A. and Kurachi, Y., "Bayesian Approach to Pricing Longevity Risk Based on Risk-Neutral Predictive Distributions", *Insurance: Mathematics and Economics* 46(1), 2010, pp. 162-172.
- Lee, R. D., "The Lee-Carter Method for Forecasting Mortality with Various Extensions and Applications", *North American Actuarial Journal* 4(1), 2000, pp. 80-91.
- Lee, R. D., and Carter, L., "Modelling and Forecasting the Time Series of US Mortality", *Journal of American Statistical Association* 87, 1992, pp. 659-671.
- Li, N. and Lee, R., "Coherent Mortality Forecasts for a Goup of Populations: An Extension of the Lee-Carter Method", *Demography* 42(3), pp. 575-594.
- Pedroza, C., "A Bayeisan Forecasting Model: Predicting U.S. Male Mortality", *Biostatistics* 7, 2006, pp. 530-550.
- Renshaw, A. E., and Haberman, S., "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors", *Insurance: Mathematics and Economics* 38(3), pp. 556-570.
- Wilmoth, J. R., "Computational Methods for Fitting and Extrapolating the

Lee-Carter Model of Mortality Change”, Technical Report, Department of Demography, University of California, Berkeley, 1993.

Abstract

In this paper, we intend to project the future Korean mortality rates, and analyze the effect of mortality improvement on the life expectancies and the actuarial present values of life insurances and life annuities. Under the Poisson log-bilinear model, we estimate the parameters for the Korean males by using the hierarchical Bayesian methodology proposed by Czado et al.(2005). For comparison purpose, we provide the estimation results based on the singular value decomposition and the maximum likelihood estimation. Our results show that the parameter estimates based on three different estimation methods do not exhibit a large difference between each other, but the traditional estimation methods underestimate the prediction interval. We expect that the Bayesian estimation method not only can provide a solution to this underestimation problem, but could be very useful to evaluate insurers' liabilities in the future.

※ **Key words:** Lee-Carter Model, Bayesian Estimation, Mortality Projection, Poisson Log-bilinear Model

