

연구논문

사망률 추계를 위한 오차수정 LC 모형*

VECM-LC Model for Forecasting Mortality in Korea

박유성^{a)} · 장선희^{b)} · 김성용^{c)}

Yousung Park · Sunwha Jang · Seongyong Kim

본 논문에서는 미래 사망률 예측을 위한 모형 적합 시, 자료의 신뢰성이 의심되는 1970년대의 자료가 포함되어야 하는지에 대해 논의하고, 우리나라의 사망률 자료를 이용했을 때의 Lee and Carter(1992) 모형을 평가하도록 한다. 특히 우리나라 통계청에서 이용하고 있는 Li and Lee(2005) 모형의 기본 가정이 우리나라 자료에서는 타당하지 않음을 남녀 사망률의 차이, 공적분 검정, 초고령 연령의 사망률 및 평균수명의 예측치 등을 통해 보이도록 한다. 또한 본 논문에서는 사망률의 횡단면적 상관관계 및 종단면적 상관관계뿐만 아니라, 공적분 관계를 반영하는 새로운 Lee and Carter 모형인 VECM-LC 모형을 제안하고 이를 Li and Lee 모형과 비교하도록 한다.

주제어: Lee and Carter 모형, Li and Lee 모형, 사망률 예측, 오차항 수정 모형

We investigate the 1970s mortality data in Korea, which have been suspected to be unreliable, to determine whether or not inclusion of them can contribute to more accurate forecasts of future mortality rates. Lee and Carter's methods(1994) are carefully assessed their performances for forecasting mortality in Korea. In particular, we show that the model of Li and Lee(2005), which is used in Statistics Korea, is not appropriate for forecasting mortality at least in Korea by

* 이 논문은 2011년 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2010-411-B00028)

a) 고려대학교 통계학과 교수

b) 고려대학교 통계학과 대학원 석사

c) 교신저자(corresponding author): 고려대학교 경제연구소 연구교수 김성용.

E-mail: yaba96@korea.ac.kr

examining the validity of their model assumptions using difference between sex-specific mortality, cointegration test, mortality forecasts of oldest old, and life expectancy at age 0. We propose a new LC model called VECM-LC to accommodate longitudinal and cross-sectional correlations and cointegration relationship in time effects of mortality, and the model performance is compared with the other LC models.

Key words: Lee and Carter model, Li and Lee model, mortality forecasting, vector error correction model

I. 서론

사망률의 추계(mortality projection)는 연금, 보험, 정부정책수립 및 계획에 기초자료를 제공한다. 사망률의 변화는 생활수준, 의료보건수준, 환경적 요인 및 유전적 요인 등에 의존하여 시간, 연령, 코호트에 따라 특수한 패턴을 따르게 된다. 사망률 패턴을 파악하고 이를 이용하여 성별·연령별 사망률의 미래치를 예측하기 위한 사망률 모형은 크게 함수적합(curve fitting)방법과 주성분 분석(principal component analysis)방법으로 나눌 수 있다.

McNown & Roger(1989)는 8개의 모수를 가진 HP함수(Helgman and Pollard 1980)를 시점별·성별·연령별 사망률을 적합하여 8개의 모수추정치에 대한 시계열을 산출한 후 이를 바탕으로 성별·연령별 특성을 규명하였다. 미래의 사망률을 예측하기 위해 8개의 모수를 예측하여야 하지만, 여기에서 두 가지 문제점이 발생한다. 첫 번째로 모수 추정치가 수렴하지 않는 문제이다. 추정해야 할 모수가 8개로 너무 많아 초기치에 따라 모수 추정치가 다르게 나오는 경우가 많다. 두 번째로 모수 추정치가 수렴을 했더라도 변동(variation)이 너무 심해 추정된 모수의 시계열에 ARIMA모형 등을 적합시킬 수 없는(intractable) 경우가 발생한다(Brouhns et al. 2002; Pitacco 2004). 실제로 이러한 문제 때문에 McNown & Rogers(1992)는 8개의 모수 중 6개의 모수는 고정시키고 오직 2개의 모수추정치에 대해 ARIMA모형을 적합시켜 미래의 사망률을 구하였다.

한편, 주성분 분석방법은 Bell & Monsell(1991)과 Lee & Carter(1992)가 연령별 사망률(age-specific mortality)을 예측하기 위해 SVD(Singular Value Decomposition)을 이용한 모수추정 방법을 제안하였다. 특히, Lee and Carter 모형(이하 LC 모형)은 사망률 추정 및 예측방법으로 가장 널리 사용되고 있는 모형이다. Lee & Carter(1992)는 시점별·연령별 로그사망률을 연령 $\times$ 시간 효과로 재 표현하고 이를 제1 주성분을 이용하여 추정하는 방법을 제안하였으며, 시간효과는 절편이 있는 임의보행과정(random walk with drift)으로 예측하였다. Lee & Miller(2001)는 LC 모형에서 사망률의 관측치와 예측치 간의 점프현상을 수정하는 수정 LC 모형을 제시하였으며, Hyndman & Ullah(2006)는 2개 이상의 주성분을 이용하여 기존의 LC 모형을 확장하였다. Li & Lee(2005)는 2개 이상의 그룹이 유사한 사망률 패턴을 보일 경우, 공통 연령 $\times$ 시간 효과 + 개별 연령 $\times$ 시간 효과로 기존의 LC 모형을 확장하였다. 여기에서 그룹은 인접 국가들이나 동일 국가 내의 성별 또는 지역이 될 수 있다. Li and Lee 모형의 특징은 장기적으로 모든 그룹은 공통 연령 $\times$ 시간 효과로 수렴한다는 가정에서 출발하고 있다. Russolillo, Giordano, & Haberman(2010)은 Li and Lee의 관점을 시간 $\times$ 연령 $\times$ 그룹의 승법모형으로 재표현하여 기존의 2-way LC 모형을 3-way LC 모형으로 확장하였으며, 시간 $\times$ 연령 $\times$ 그룹 효과는 three-mode 주성분 분석(Tucker 1966)을 이용하여 추정하였다.

LC류(a type of LC models)의 모형에 대한 약점은 수많은 연구에서 지적되어 왔다. 특히, 연령 $\times$ 시간 효과에서 연령효과가 시간에 의존하지 않는다는 점과 LC 모형의 오차가 등분산을 가정하고 있다는 점이 대표적인 약점이라고 할 수 있다(Renshaw & Haberman 2003; Renshaw & Haberman 2006; Yue et al. 2008; Hatzopoulos & Haberman 2009). 그럼에도 불구하고 LC류 모형이 사망률 변동을 매우 우수하게 설명할 수 있다는 점과 연령별 사망률의 미래치를 예측하는 데에 강건성(robustness)을 가지고 있어 대부분의 국가에서 LC류의 모형을 채택하고 있다.

LC류 모형을 적용하기 전에 선행되어야 할 것은 신뢰성 있는 사망률 자료의 확보와 80세 이상의 초고령 사망률의 연장이다.

첫째로, 신뢰성 있는 사망률 자료의 확보에 대해 고려해 보자. 우리나라 사망률 자료는 1970년부터 존재하지만, Bernett & Horiuchi(1981)와 Hill (1997)은 인구통계학적 분석(demographic analysis)를 통해 1970년대의 우리나라 사망률 자료는 신뢰성에 문제가 있음을 지적하고 있다. 그러나 통계청 인구추계에서는 1970년대의 사망률 자료

를 포함하여 미래의 사망률을 예측하고 있다. 본 연구에서는 1970년대의 사망률 자료를 “동일연령의 사망률은 앞 시점보다 뒤 시점이 낮아야 한다”는 사망률의 일치성(Condran et al. 1991; Wilmoth 1995; Boleslawski & Tabeau 2001; 김성용 외 2011)의 관점에서 미래 사망률 예측을 위한 모형 적합시 1970년대 사망률 자료를 포함해야 하는지를 파악해 보도록 한다.

두 번째로, 초고령 사망률의 연장에 대해 고려해 보자. 고령화(aging)연구의 핵심은 80세 이상 연령의 사망률의 수준, 패턴, 그리고 추이의 파악에 있다(김성용 외 2011). 그러나 우리나라뿐만 아니라 대부분의 나라에서 1990년대까지는 80세 이상 초고령 사망률 자료가 존재하지 않거나 절단된 형태로만 존재한다(Boleslawski & Tabeau 2001; Beutner 2003; 김성용 외 2011). 이는 80세 이상의 초고령 사망률의 예측치가 있어야만 LC류 모형을 적합할 수 있다는 것을 의미한다. 김성용 외(2011)은 초고령 사망률을 예측하기 위해 2005~2009년의 센서스후 인구(postcensal population)을 이용하여 우리나라의 표준 사망률을 제시하였다. 김성용 외(2011)가 제안한 대로 본 연구에서는 2005~2010년의 센서스인구가 작성되었으므로 이 자료를 바탕으로 확정된 표준 사망률을 재산출하여 제시하고자 한다.

이와 같은 사망률 자료의 신뢰성 확보와 초고령 사망률의 연장 후, 본 논문에서는 LC 모형, Hyndman & Ullah(2006)의 2개 이상의 주성분을 이용한 LC 모형(이하 2LC 모형), Li & Lee(2005)의 모형(이하 LL모형)을 비교하고 우리나라 사망률 적용에 대한 타당성을 점검하고자 한다. 특히, LL모형은 현재 통계청에서 사용하고 있는 모형으로, 2개 이상의 그룹(국가 또는 성별)의 사망률이 수렴해야 한다는 전제 조건이 우리나라 자료에서 성립되는지를 살펴볼 것이다.

LL모형의 대안으로 벡터오차수정(Vector Error Correction Model) LC 모형(VECM-LC 모형)을 제안하고자 한다. 이 VECM-LC 모형은 2개 이상 그룹의 사망률이 장기균형관계를 가지고 있으며 횡단면(cross-sectional) 및 종단면(longitudinal) 상관관계를 가지고 있을 때 매우 유용한 모형이다. LL모형의 공통 연령 $\times$ 시간 효과로의 수렴은 장기균형관계의 특수한 경우로 해석할 수 있으므로 LL모형은 VEC-LC 모형의 특수한 경우로 해석할 수 있다.

본 연구는 총 5장으로 구성되어 있다. II장에서는 1970년대의 사망률 자료의 신뢰성을 사망률의 일치성과 Brass-logit 모형의 회귀계수를 통해 점검할 것이다. III장에서는 LC 모형, 2LC 모형, LL모형의 특성과 문제점을 논의한다. 특히, LL모형은 우

리나라 사망률자료에 적용하기에는 많은 문제점이 있다는 논증이 구체적으로 제시될 것이다. IV장에서는 새로운 VECM-LC 모형을 제안하고 모형 적합도와 평균수명의 관점에서 VECM-LC 모형과 기존 모형들을 비교하고자 한다. V장에서는 결론과 향후과제에 대해 논의할 것이다.

II. 사망률 자료

사망률은 연령별 사망자 수를 연령별 인구로 나누어 산출되게 된다. 우리나라 사망자 수 자료는 1970년부터 존재하지만, 통계청은 통계의 일관성 확보와 국제 비교가 가능하도록 1983년부터 사망자 통계를 보정하여 제공하고 있다. 인구는 추계인구, 주민등록인구, 총조사인구가 있다. 이 중 총조사인구가 초고령 연령까지 제공하고 있다. 총조사인구는 1970년부터 2000년까지는 85+세까지, 2005년 이후는 100+세까지 집계되어 있으며 추계인구의 경우 1960년부터 1999년까지는 80+세, 2000년부터 2005년까지는 95+세, 2006년부터는 100+세로 집계되어 있다. 본 연구에서는 초고령 사망률을 최대한 확보하기 위해 총조사인구를 사용할 것이다.

1. 1970년대 사망 자료의 신뢰성

Bernett & Horinchi(1981)와 Hill(1997)은 인구통계학적 분석을 통해 1970년대 우리나라 사망률 자료에 치명적인 오류가 있다고 지적하고 있다. 미래시점의 추계된 사망률의 신뢰성을 확보하기 위해서는 신뢰성 있는 사망률 자료가 전제되어야만 가능하다.

그러나 우리나라의 1970~1979년과 1980~1985년의 연령별 사망률의 평균을 비교하여 본 결과, 남성의 경우 1970~1979년의 38~42세, 73~77세, 그리고 81~84세의 평균 사망률이 1980~1985년보다 낮게 나왔으며, 여성의 경우 70~78세와 80~84세까지의 1970~1979년의 평균사망률이 1980~1985년의 평균사망률보다 낮게 집계되었다. 이는 1970년대 사망 자료에 체계적 오류가 있음을 보여주고 있다. 이와 같이, 70세 이상 대부분의 연령대에서 발생한 사망률의 역전 현상은 미래사망률을 과소추계하게 되므로 1970년대 사망률의 자료는 제외하는 것이 타당하다고 할 수 있다.

2. 초고령 사망률의 확장과 사망률 자료의 신뢰성

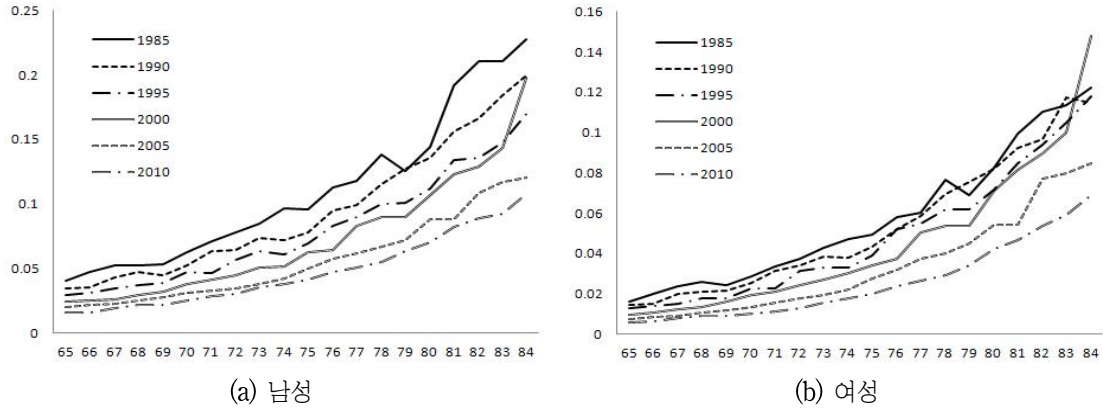
미래의 사망률을 추계하기 위한 절차는 사망률의 변동성을 줄이기 위한 평활화 과정(smoothing)과 85+ 또는 100+ 형태로 절단된 사망률을 110+ 또는 그 이상의 사망률로의 확장이다(김성용 외 2011). 김성용 외(2011)는 2005~2009년의 추계 사망 자료를 이용하여 우리나라의 표준사망률을 제시하고 이 표준사망률을 Brass-logit 모형(Brass 1977)에 적용하여 초고령 사망률을 확장하였다. 김성용 외(2011)가 제안한 대로 센서스 자료가 관찰되었으므로 2005~2010년의 실측자료를 이용하여 우리나라의 표준사망률을 재산출하였다. 재 산출된 표준사망률은 부록에 제시되어 있다. Brass(1977)가 제안한 Brass-logit 모형은

$$\log \frac{m_{x,t}}{1-m_{x,t}} = \beta_{0,t} + \beta_{1,t} \log \frac{m_x^s}{1-m_x^s} + \epsilon_t \quad (1)$$

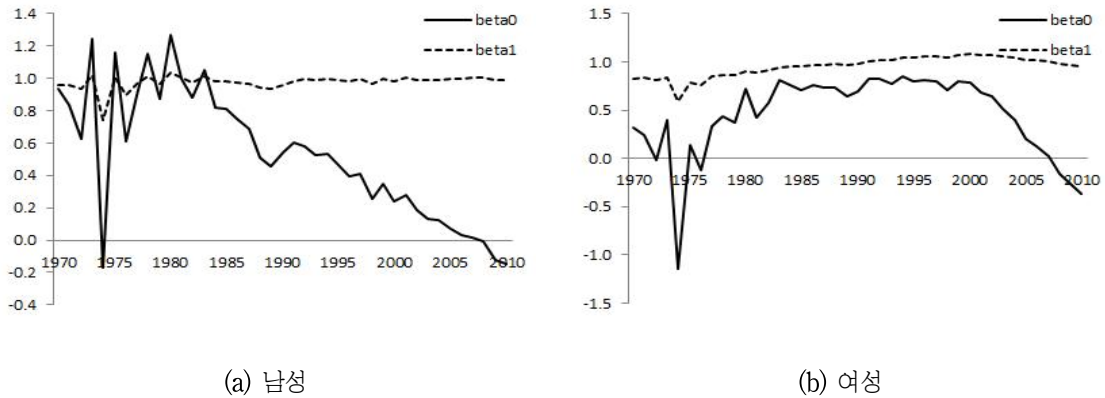
의 단순회귀모형이다. 여기에서 $m_{x,t}$ 는 t 년의 x 세 사망률이며 m_x^s 는 x 세의 표준사망률이다. 표준사망률은 2005~2010년의 사망률 자료로부터 산출되므로 우리나라 사망률의 최근 패턴을 나타내고 있다. 그러므로 $\beta_{1,t}$ 가 1에 가까울수록 시점 t 의 사망률 패턴이 최근의 사망률 패턴과 유사하다는 것을 의미한다. $\beta_{0,t}$ 는 시점 t 에서의 사망률의 평균을 의미하므로 시간이 갈수록 점차 감소하는 패턴을 보이는 것이 자연스러운 현상이다.

식(1)을 적합하기 위해서는 적합연령 구간의 선택이 필요하다. 초고령사망률의 패턴은 유아 사망률이나 청년 사망률의 패턴과 다르므로, 본 논문에서는 김성용 외(2011)에서와 같이 45세부터의 사망률을 적합연령으로 선택하였다. <그림 1>은 65세~84세까지의 사망률의 시계열 추이(1985~2010)를 보여주고 있다.

<그림 1>에서 볼 수 있듯이 75세 이상에서 사망률의 변동이 커지기 시작할 뿐만 아니라, 초고령에서의 기간별 사망률의 순서가 0세에서의 기대수명과 일치해야 한다(Beutler 2003)는 일치성을 위배하고 있다. 이러한 관점에서 Brass-logit 모형의 적합연령을 45~74세로 하여 1970~2010년 자료를 식(1)에 적용하였다. 그 결과, 남성의 경우 R^2 가 최소 0.964, 최대 0.996이며 여성의 경우 최소 0.969, 최대 0.997로 매우 만족스러운 결과를 구할 수 있었다. 단, 1974년 자료의 경우 R^2 가 매우 낮게 (남성의 경우 0.486이며, 여성의 경우 0.472) 나와 자료의 신뢰성이 매우 낮은 것으로 나타났다.



〈그림 1〉 65세~84세까지의 사망률



〈그림 2〉 Brass-logit 모형의 연도별 회귀계수

〈그림 2〉는 1970~2010년까지 성별·연도별 Brass-logit 모형의 절편 $\beta_{0,t}$ 와 기울기 $\beta_{1,t}$ 의 플롯이다. 먼저 $\beta_{1,t}$ 를 살펴보면, 남성의 경우 1980년 이전까지는 $\beta_{1,t}$ 가 불안정하다가 1980년 $\beta_{1,t} = 1.034$ 를 시작으로 2010년까지 1에 매우 근접한 값을 보이고 있다. 그러므로 남성의 경우 1980년부터 최근의 사망률 패턴과 유사하다고 할 수 있다. 한편 여성의 경우, 1984년 $\beta_{1,t} = 0.954$ 으로 이 시점부터 최근의 사망률 패턴과 비슷해지고 있음을 알 수 있다. $\beta_{0,t}$ 은 남녀 모두 1982년까지 매우 불안정하다가, 1983

년부터 남성의 경우는 감소하며 여성의 경우는 안정적인 경향을 보여주고 있다. 즉, 1970년대와 1980년 초반 사망률 자료는 신뢰성에 문제가 있으며 최근의 사망률 패턴을 반영하기 위해 최근의 자료를 사용하는 국제적인 경향(Statistics Canada 2009)의 관점에서 볼 때도 부적절한 것으로 보인다. 그러므로 본 논문에서는 남성은 110+세, 여성은 112+세까지 확장한 1985년부터 2010년까지의 사망률 자료를 이용하여 사망률 추계를 하고자 한다.

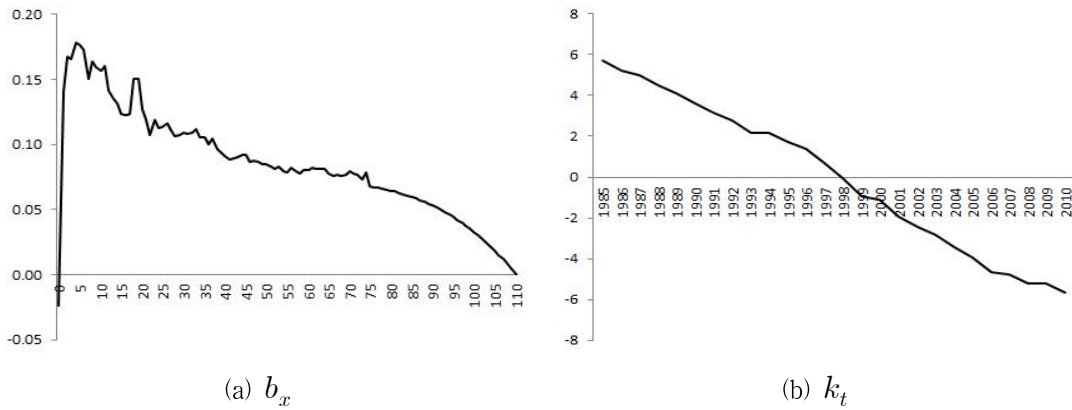
III. LC류 모형의 특성과 문제점

Lee & Carter(1992)가 제안한 LC 모형은

$$\log(m_{x,t}) = a_x + b_x k_t + \epsilon_t \quad (2)$$

으로 정의되며 ϵ_t 는 평균이 0이고 분산이 σ^2 인 *i.i.d*한 오차항이다. 모수 a_x 는 x 세의 평균사망률을 의미하며 $\frac{1}{T} \sum_{t=1}^T m_{x,t}$ 로 추정되며, b_x 와 k_t 는 비정칙치 분해(singular value decomposition)을 통해 구한 첫 번째 주성분(the first principal component)에 의해 추정된다. k_t 는 시점 t 에서의 사망률의 수준(level)을 나타내며 일반적으로 선형적으로 감소한다. 로그사망률 $\log(m_{x,t})$ 는 시점 t 에 대해 감소함수(즉, x 세의 사망률은 시간이 감에 따라 감소) 이므로 b_x 는 모든 x 에 대해 양의 값을 가져야 한다. k_t 는 궁극적으로 음의 값이므로 b_x 가 음의 값을 가지면 시간이 감에 따라 x 세의 사망률이 증가하는 현상이 발생하기 때문이다. a_x 와 b_x 가 시간 t 에 의존하지 않으므로 미래의 사망률 $m_{x,t}$ 는 k_t 의 시계열에 모형을 적합시켜 k_t 의 미래치 $\hat{k}_t$ 를 구한 후 $\hat{m}_{x,t} = \exp(a_x + b_x \hat{k}_t)$ 으로 추정하게 된다.

<그림 3>은 우리나라 남성 사망률의 b_x 와 k_t 이다. k_t 는 시점 t 에 대해 선형적으로 음의 방향으로 감소하여 사망률이 연령에 관계없이 시간이 감에 따라 감소한다는 것을 의미한다. 한편 b_x 는 0세에서만 유일하게 음의 값을 가지고 있으며, x 가 커짐에 따라 일반적으로 감소하는 경향을 보인다. 이는 고연령일수록 시간이 지남에 따라 사망률의 감소속도가 줄어든다는 것을 의미한다.



〈그림 3〉 우리나라 남성 사망률의 b_x 와 k_t



〈그림 4〉 0세에서의 사망률

그런데 〈그림 3〉과 같이 b_x 가 0세에서 음의 값을 가지는 경우 식(2)의 b_0k_t 에 의해 0세에서 시간이 감에 따라 양의 방향으로 커지게 되며 사망률이 증가한다는 모순에 빠지게 된다. 우리나라 사망률 자료에 대한 이러한 현상은 모든 LC류의 모형이 공통적으로 가지고 있는 약점이다.

〈그림 4〉에서 볼 수 있듯이 우리나라 0세 사망률은 남녀 모두 1998년까지 0.002 수준까지 감소하다가 1999년에 0.004 이상의 수준(남성: 0.0047, 여성: 0.0041)으로 급상승하여 신뢰도가 전반적으로 매우 떨어진다. 출생 직후 사망 시의 신고 누락이나 출생 신고의 지연 등이 주요 이유이다. 2000년에 이르러서 통계청은 화장장 등을 조사하여

우리나라 0세 사망률의 신뢰도를 높이고 있어, 최근의 0세 사망률에 가장 높은 가중치를 주는 이중지수평활법(박유성 · 김기환 2002)을 이용하여 미래의 0세 사망률로 대체하였다.

1. LC 모형

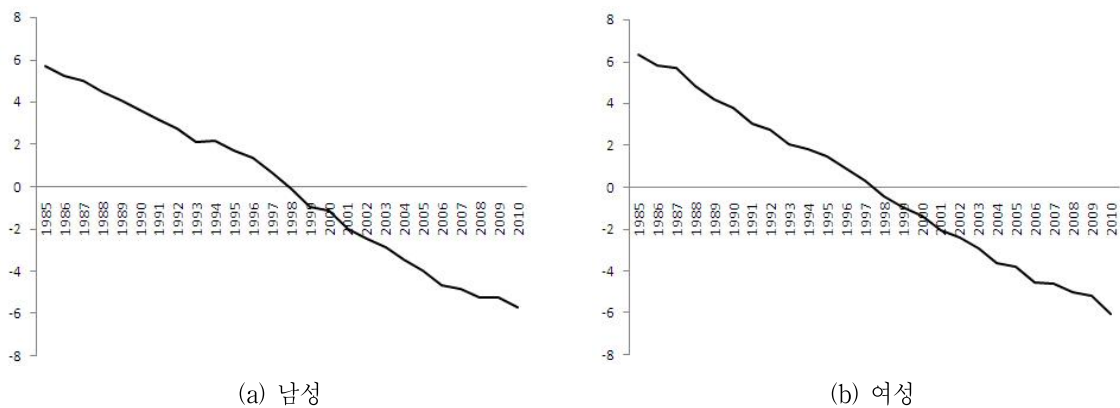
II장의 1세 이상 사망률 자료를 이용하여 식(2)의 k_t 추정치를 구한 결과는 <그림 5>와 같다.

<그림 5>에서 볼 수 있듯이 남녀 모두 k_t 가 선형적으로 감소 추세를 가지고 있으므로, 1차 단위근(the first-order unit root)을 가지는 것으로 보인다. 이를 점검하기 위해 아래의 두 모형, 즉,

$$\text{ADF1} : \nabla k_t = \alpha_0 + \gamma k_{t-1} + \sum_{i=1}^{p-1} \beta_i \nabla k_{t-i} + \epsilon_t$$

$$\text{ADF2} : \nabla k_t = \alpha_0 + \gamma k_{t-1} + \alpha_2 t + \sum_{i=1}^{p-1} \beta_i \nabla k_{t-i} + \epsilon_t$$

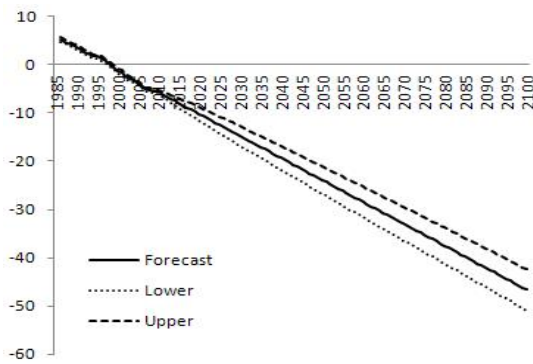
에 대해 증대된 DF-검정(augmented Dickey-Fuller test, Dickey & Fuller 1979)를 실시한 결과, <표 1>에서 볼 수 있듯이 남녀 모두 k_t 가 1차 단위근을 가지고 있는 것으로 나타났다. 단, 잔차검정 결과 ADF1 및 ADF2 모형 모두 $p=2$ 로 설정하였다.



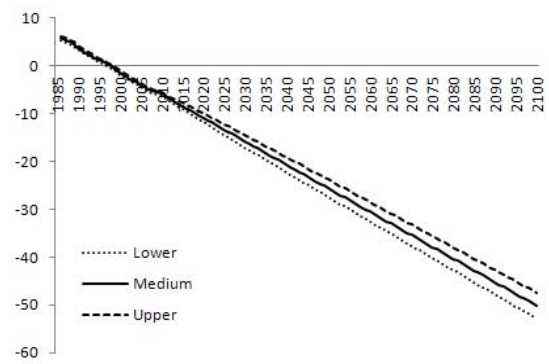
<그림 5> k_t 추정치

〈표 1〉 단위근 검정 결과

모형	남성		여성	
	τ	p-value	τ	p-value
ADF1	-0.73	0.8198	-1.45	0.5395
ADF2	-1.96	0.592	-2.02	0.5574



(a) 남성



(b) 여성

〈그림 6〉 우리나라 성별 LC 모형의 k_t 적합치와 예측치

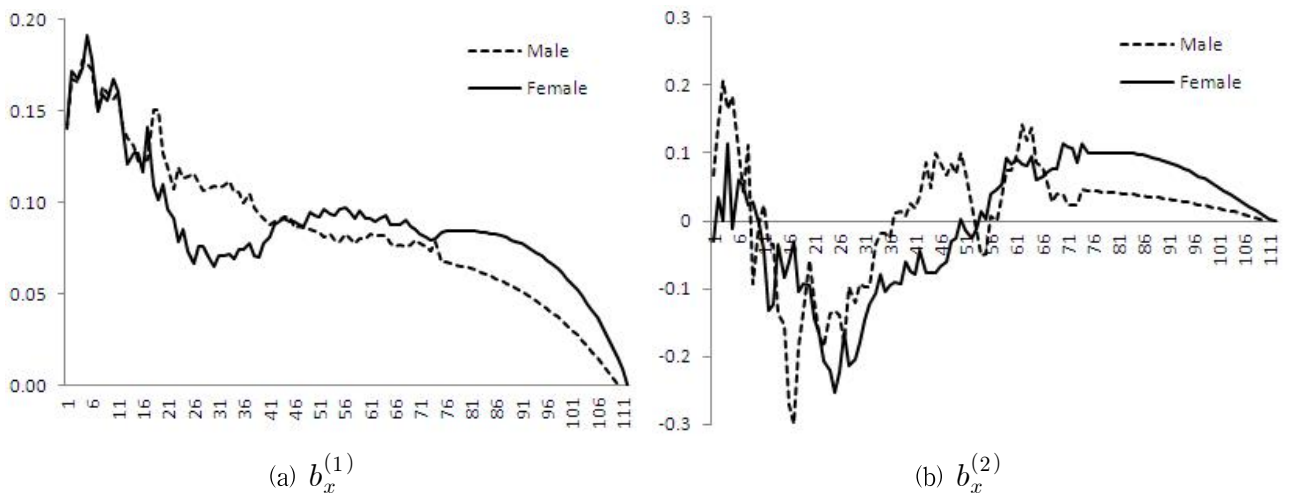
따라서 k_t 에 대해 ARIMA 모형(Autoregressive Integrated Moving Average)을 적합한 결과, 남성 및 여성의 k_t 에 대한 모형은

$$\text{남성: } k_t = k_{t-1} - 0.455 + \epsilon_t$$

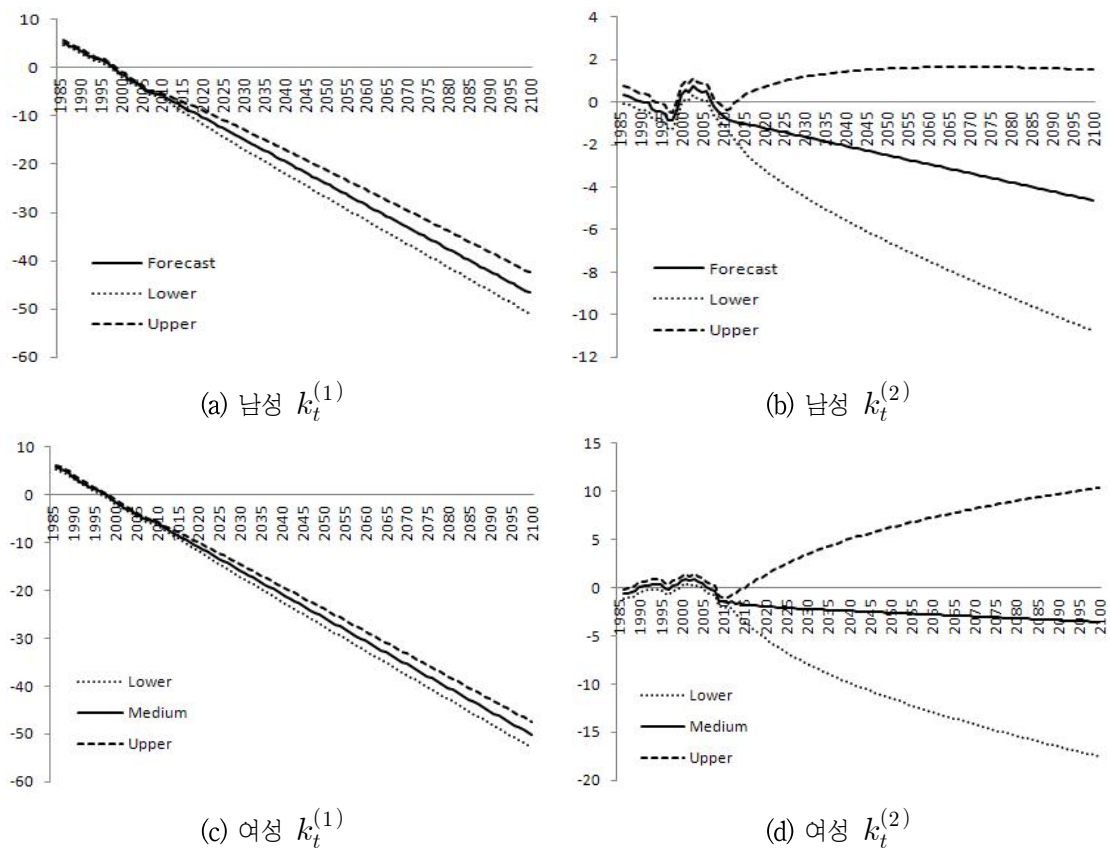
$$\text{여성: } k_t = k_{t-1} - 0.491 - 0.478(k_{t-1} - k_{t-2}) + \epsilon_t \quad (3)$$

으로 추정되었으며, 각 추정치에 대한 p -값은 모두 0.001보다 작았다. 남성의 경우 절편이 있는 임의보행과정이며, 여성의 경우는 절편과 자기회귀과정을 따르는 임의보행과정이다. 남성과 여성의 a_x 와 b_x 는 지면상 생략한다. 〈그림 6〉은 위 모형으로 예측한 남성과 여성 사망률의 적합치 및 95% 신뢰구간이다.

〈그림 6〉에서 알 수 있듯이 LC 모형의 가장 큰 단점 중 하나는 k_t 의 미래치가 절편(drift)에 의해 거의 결정됨으로써 k_t 의 신뢰구간이 너무 좁아, 대응되는 $m_{x,t}$ 의 신뢰구간이 너무 좁게 나오는 것이다.



〈그림 7〉 남성과 여성의 $b_x^{(1)}$ 와 $b_x^{(2)}$



〈그림 8〉 남성 및 여성 $k_t^{(1)}$ 와 $k_t^{(2)}$ 의 예측치와 신뢰구간

다음으로 Hyndman & Ullah(2006)가 제안한 2개 이상의 주성분을 이용한 2LC 모형을 우리나라의 사망률 자료에 적합해 보도록 한다. 2LC 모형은

$$\log(m_{x,t}) = a_x + b_x^{(1)}k_t^{(1)} + b_x^{(2)}k_t^{(2)} + \epsilon_t \quad (4)$$

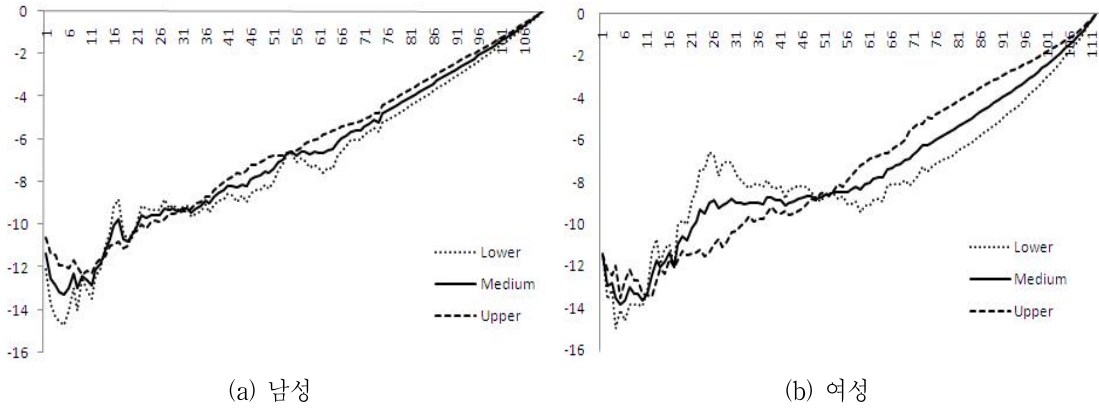
이며, 여기에서 $b_x^{(1)}k_t^{(1)}$ 와 $b_x^{(2)}k_t^{(2)}$ 는 제 1 및 제 2 주성분으로 구한 추정치이다. $\log(m_{x,t})$ 의 분산 중 제 1 주성분의 설명력은 남성(여성)은 95.5%(93.8%)였고, 제 2 주성분의 설명력은 남성(여성)은 1.41%(2.14%)였다. 그러므로 제 1 및 제 2 주성분으로 $\log(m_{x,t})$ 의 변동성은 거의 설명된다고 볼 수 있다.

〈그림 7〉은 남성 및 여성 사망률에 대한 $b_x^{(1)}$ 와 $b_x^{(2)}$ 의 플롯이다. $b_x^{(1)}$ 는 남녀 모두 양의 값을 가지지만, $b_x^{(2)}$ 의 경우 남녀 모두 10~55세까지 음의 값을 가지며, 55세 이상은 양의 값을 가지고 있다. $b_x^{(2)}$ 의 이러한 행태는 〈그림 8〉의 $k_t^{(2)}$ 와 곱해져서, 10~55세의 사망률의 수준을 더 높게, 55세 이상의 사망률을 더 낮게 낮추는 역할을 하게 된다. 미래의 사망률 예측을 위한 $k_t^{(1)}$ 와 $k_t^{(2)}$ 의 모형은

$$\begin{aligned} \text{남성: } k_t^{(1)} &= k_{t-1}^{(1)} - 0.455 + \epsilon_t^{(1)} \\ k_t^{(2)} &= \epsilon_t^{(2)} - 0.624\epsilon_{t-2}^{(2)} \\ \text{여성: } k_t^{(1)} &= k_{t-1}^{(1)} - .491 - .478(k_{t-1}^{(1)} - k_{t-2}^{(1)}) + \epsilon_t^{(1)} \\ k_t^{(2)} &= \epsilon_t^{(2)} + .664\epsilon_{t-2}^{(2)} \end{aligned} \quad (5)$$

이며, p 값은 모두 0.05보다 작다. 모형 (5)로부터 남성 및 여성의 2100년까지의 예측치와 신뢰구간은 〈그림 8〉과 같다.

〈그림 8〉에서 볼 수 있듯이 $k_t^{(1)}$ 은 신뢰구간이 매우 좁지만 $k_t^{(2)}$ 의 신뢰구간은 상대적으로 크게 나와, 주성분 하나만을 사용하는 식(2)의 LC 모형으로 계산된 $m_{x,t}$ 의 신뢰구간보다 식(4)로부터 계산된 $m_{x,t}$ 의 신뢰구간이 더 넓게 나오게 된다. 이는 사망률이 너무 좁게 나온다는 인구통계학적 비판에는 장점으로 작용한다. 그러나 가장 치명적인 약점은 $b_x^{(2)}$ 가 연령에 의존하여 음 또는 양의 값을 가지게 되어 로그사망률의 신뢰하한과 신뢰상한이 교차하는 현상이 벌어지게 된다는 점이다. 〈그림 9〉는 2060년에 대한 식(5)로부터 계산된 로그사망률의 예측치와 95% 신뢰구간이다.



〈그림 9〉 식(5)의 2060년 남성 및 여성 로그사망률 예측치와 95% 신뢰구간

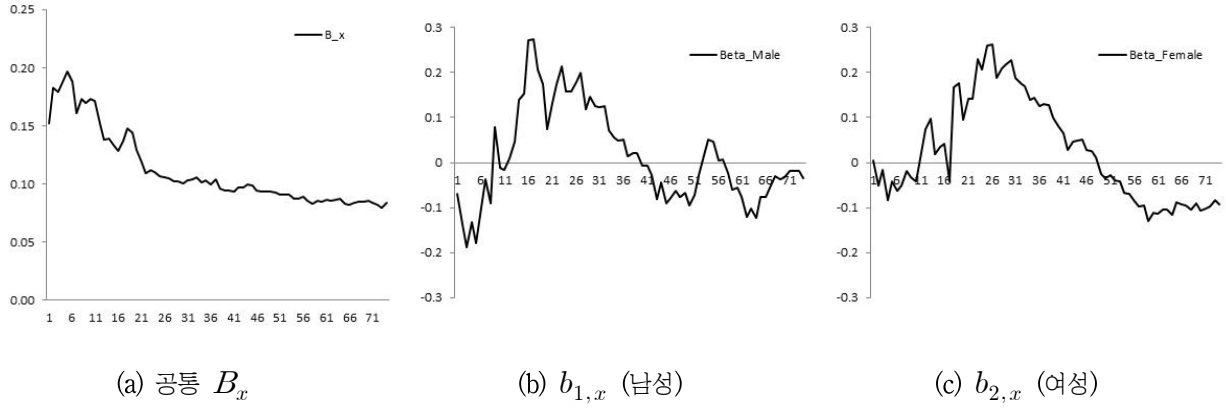
〈그림 9〉에서 “Lower”는 $k_t^{(1)}$ 과 $k_t^{(2)}$ 의 신뢰하한을 이용한 로그 사망률이며 “Upper”는 $k_t^{(1)}$ 과 $k_t^{(2)}$ 의 신뢰상한을 이용한 로그 사망률이다. 위 그림에서 알 수 있듯이 남녀 모두 특정 연령구간에서 “Lower”가 평균보다 크거나 혹은 “Upper”가 평균보다 더 작은 값을 가지고 있으며, 여성의 경우 55세를 기점으로 “Lower”와 “Upper”가 교차하는 현상이 명확하게 나타나고 있다. 이와 같은 교차현상은 〈그림 7〉 및 〈그림 8〉에서 알 수 있듯이 $b_x^{(2)}$ 가 10~55세까지 음의 값이고 2060년의 $k_t^{(2)}$ 의 하한값이 역시 음의 값이므로 $b_x^{(2)} \times (k_t^{(2)})$ 의 하한값이 양의 값을 갖게 되기 때문이다. 이러한 문제가 해결되지 않으면 2개 이상의 주성분을 사용하는 LC 모형은 신중해야 하며 우리나라의 경우 적합하지 않는 것으로 판단된다.

2. Li and Lee 모형

Li & Lee(2005)는 경제, 사회적 환경이 유사한 국가나 그룹은 장기적으로 동일한 사망률 패턴을 가진다는 가정하에서 $m_{i,x,t}$ 를 i 번째 그룹의 시간 t 에서의 x 세의 사망률이라고 할 때,

$$\log(m_{i,x,t}) = a_{i,x} + b_{i,x}k_{i,t} + B_xK_t + \epsilon_{i,x,t} \quad (6)$$

으로 정의한다. 여기에서 $a_{i,x}$ 는 i 번째 그룹의 x 세의 평균사망률, $b_{i,x}k_{i,t}$ 는 i 번째 그



〈그림 10〉 LL모형에서 공통효과 B_x 와 남성 및 여성의 $b_{i,x}$

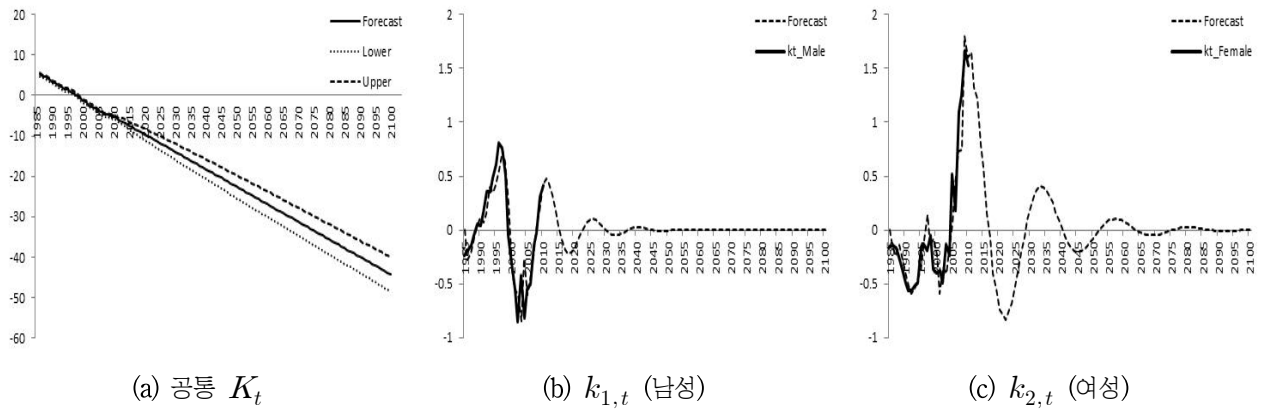
룹의 개별효과이며, $B_x K_t$ 는 모든 그룹의 공통효과이다. 우리나라 통계청에서는 그룹을 남녀로 정의하고 식(6)의 LL모형을 이용하고 있다. 본 논문에서도 그룹으로 남녀를 정의하고, 남자인 경우 $i = 1$, 여자인 경우 $i = 2$ 라고 하자. 우리나라 사망률 자료에 적합한 LL모형의 모수추정치는 〈그림 10〉과 같다.

〈그림 10〉은 공통효과 B_x , $b_{1,x}$, 그리고 $b_{2,x}$ 추정치의 플롯이다. 공통효과 B_x 는 LC 모형과 동일하게 양의 값을 가지고 있지만 남성과 여성의 개별효과 $b_{1,x}$ 와 $b_{2,x}$ 는 연령에 의존하여 음의 값 또는 양의 값을 가지고 있음을 알 수 있다.

공통시간효과 K_t 와 남녀시간효과 $k_{1,t}$ 와 $k_{2,t}$ 의 모형은 다음과 같이

$$\begin{aligned}
 K_t - K_{t-1} &= -0.433 + \epsilon_t \\
 k_{1,t} &= 1.148 k_{1,t-1} - 0.404 k_{1,t-2} + \epsilon_{1,t} \\
 k_{2,t} &= 0.938 k_{2,t-1} + 0.712 k_{2,t-2} - 0.783 k_{2,t-3} + \epsilon_{2,t}
 \end{aligned} \tag{7}$$

으로 추정되었으며, p -값은 모두 0.005보다 작았다. 이 모형에 따른 미래 예측치는 〈그림 11〉과 같다.



〈그림 11〉 공통시간효과 K_t 와 남녀시간효과 $k_{1,t}$ 와 $k_{2,t}$ 의 적합치와 예측치

식(7)과 〈그림 11〉에서 볼 수 있듯이 공통시간효과 K_t 는 단위근(unit root)을 가진 비정상성 과정(nonstationary process)이며 $k_{1,t}$ 와 $k_{2,t}$ 는 0으로 수렴하는 정상성 과정(stationary process)임을 알 수 있다. 이러한 현상은 로그사망률($\log m_{i,x,t}$)이 시간 t 에 대해 감소하는 경우 LL모형에서 항상 성립한다. 그러므로 식(6)은 시점 t 가 충분히 크면

$$\log m_{i,x,t} = a_{i,x} + B_x K_t + \epsilon_{i,x,t} \quad (8)$$

으로 수렴하게 된다. LL 모형을 이용하기 위해서는 충분히 큰 t 에 대해 식(8)에 의해 모든 x 에서

$$\log m_{1,x,t} - a_{1,x} - (\log m_{2,x,t} - a_{2,x}) \approx 0$$

이 되어야 한다. 만약 $\sum_x |\log m_{1,x,t} - a_{1,x} - (\log m_{2,x,t} - a_{2,x})|$ (이를 sum of differences of adjusted mortality(SDAM)이라고 정의한다)이 시점 t 에 대해 감소하면 LL 모형은 좋은 후보모형이 될 것이다. 〈표 2〉는 $x = 0$ 세부터 $x = 84$ 세까지 1985~2010년 동안 계산된 SDAM을 제공하고 있다.

〈표 2〉 조정된 남녀 로그사망률 차(SDAM)

년도	1985	1990	1995	2000	2005	2010
SDAM	6.670	6.199	7.108	7.265	8.903	11.709

〈표 2〉에서 볼 수 있듯이 SDAM이 기대와는 반대로 시간이 지나감에 따라 증가하는 것으로 나타나 우리나라 남녀사망률 자료에 Li and Lee 모형을 적합하는 것은 문제가 있는 것으로 나타났다. 좀 더 구체적으로, 식(6)의 공통효과 $B_x K_t$ 의 존재여부를 위한 통계적인 검정을 살펴보자.

우선, 1절의 LC 모형으로 구한 k_t 와 b_x 를 남자의 경우 k_t^{male} , b_x^{male} 이라 하고 여자의 경우 k_t^{female} , b_x^{female} 이라고 하자. 그러면 LL 모형이 성립하기 위해서는 $b_x^{male} k_t^{male}$ 과 $b_x^{female} k_t^{female}$ 이 모두 장기적으로 $B_x K_t$ 로 수렴해야 한다. 이러한 관계는 아래에서 기술한 것과 같이 k_t^{male} 과 k_t^{female} 간의 공적분(cointegration) 관계의 특수한 경우로 표현할 수 있으며, Johanson 검정(Johanson 1988)을 통해 식(6)의 공통효과와 $B_x K_t$ 의 존재여부를 통계적으로 파악할 수 있다.

LC 모형으로부터 구한 $b_x^{male} k_t^{male}$ 과 $b_x^{female} k_t^{female}$ 이 모두 $B_x K_t$ 로 수렴하고, k_t^{male} 과 k_t^{female} 이 아래의 공적분 관계를 가지고 있다고 가정하자(1절에서 밝혔듯이 k_t^{male} 과 k_t^{female} 은 일차 단위근을 가지고 있다).

$$\alpha + \beta k_t^{male} - k_t^{female} = 0$$

위 식의 양변에 b_x^{female} 을 곱하면 $\alpha b_x^{female} + b_x^{female} \beta k_t^{male} = b_x^{female} k_t^{female}$ 이 된다. 가정에 의해 우변의 $b_x^{female} k_t^{female}$ 가 $B_x K_t$ 로 수렴하므로 $\alpha b_x^{female} + b_x^{female} \beta k_t^{male}$ 도 $B_x K_t$ 로 수렴하게 된다. 한편,

$$\alpha b_x^{female} + b_x^{female} \beta k_t^{male} = \alpha b_x^{female} + \frac{b_x^{female}}{b_x^{male}} \beta b_x^{male} k_t^{male}$$

이고, 가정에 의해 $b_x^{male} k_t^{male} \rightarrow B_x K_t$ 이므로

$$B_x K_t = \alpha b_x^{female} + \frac{b_x^{female}}{b_x^{male}} \beta B_x K_t$$

의 등식이 성립해야 한다. 즉,

$$\alpha b_x^{female} = 0, \frac{b_x^{female}}{b_x^{male}} \beta = 1$$

의 두 등식이 모두 성립되어야 한다.

$b_x^{female} \neq 0$ 이므로(만약 0이라면 모든 연령의 사망률이 시간에 관계없이 상수가 되기 때문) 설령 $(b_x^{female}/b_x^{male})\beta = 1$ 이 성립되더라도 $\alpha \neq 0$ 이면 $b_x^{male} k_t^{male}$ 과 $b_x^{female} k_t^{female}$ 이 모두 $B_x K_t$ 로 수렴할 수 없다는 것을 의미한다. 그러므로 LL모형을 가정하기 위한 전제조건은 LC 모형으로부터 구한 k_t^{male} 및 k_t^{female} 이 $\alpha + \beta k_t^{male} - k_t^{female} = 0$ 의 관계(즉, 공적분 관계)가 있는지를 점검한 후, 관계가 있다면 최소한 $\alpha = 0$ 이 되어야 한다는 것이다.

본 논문에서는 $k_{1,t}$ 와 $k_{2,t}$ 사이에 공적분 관계가 존재하는지 Johansen 검정(Johansen 1991)을 실시하도록 하며, 다음의 공적분 모형을 고려한다.

$$\Delta \mathbf{k}_t = \alpha(\beta', \beta_0)(\mathbf{k}_{t-1}', 1)' + \sum_{i=1}^{p-1} \Phi_i^* \Delta \mathbf{k}_{t-i} + \epsilon_t$$

여기서 $\mathbf{k}_t = (k_t^{male} \ k_t^{female})'$ 이며, 잔차검정 결과 $p = 2$ 로 설정하였다. 위 모형에서 $\alpha(\beta', \beta_0) = \Pi$ 라고 할 때, 가설은

$$H_0 : Rank(\Pi) = 0 \quad vs \quad H_1 : Rank(\Pi) > 0$$

이며, 검정통계량으로는 트레이스 통계량을 이용한다. <표 3>은 공적분 검정 결과이다.

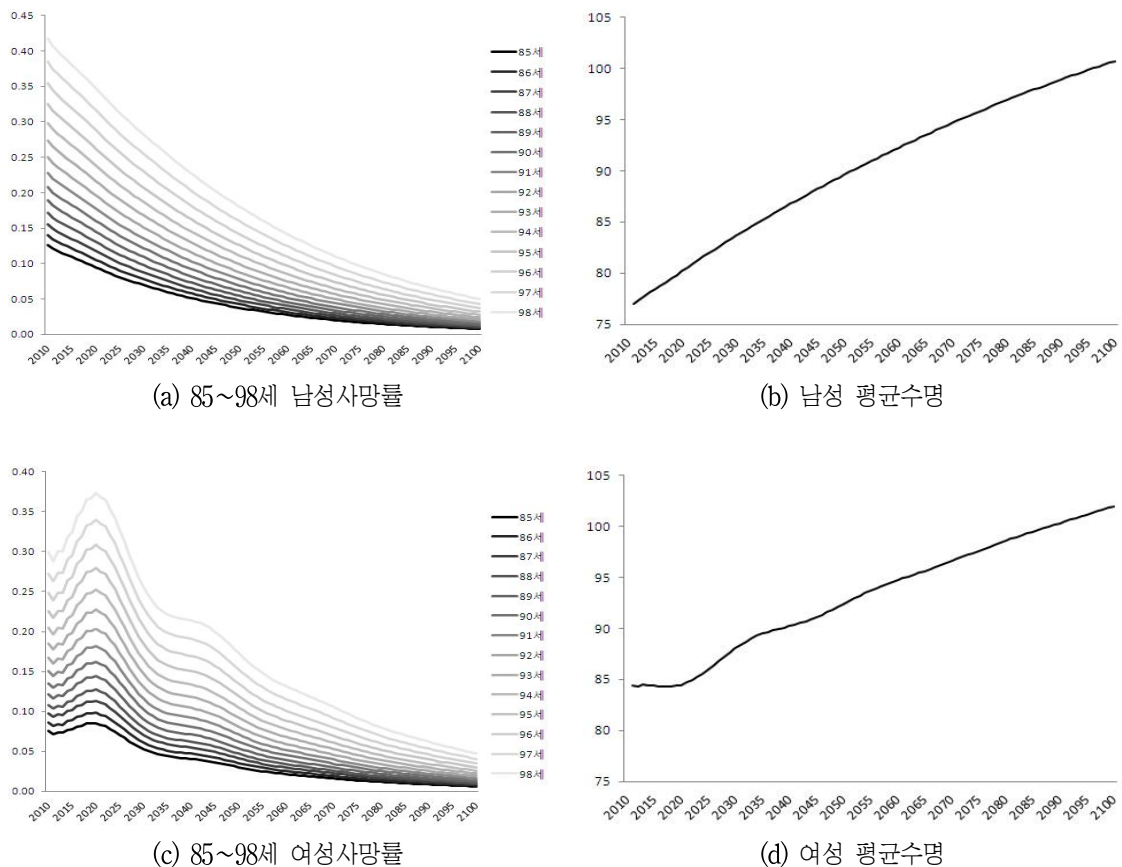
<표 3> Johansen 공적분 검정 결과

H0: Rank = r	H1: Rank > r	Eigenvalue	Trace	5% Critical Value
0	0	0.6972	34.4215	19.99
1	1	0.2131	5.7525	9.13

〈표 3〉에서 알 수 있듯이 하나의 공적분 관계가 존재하며 공적분 관계는 $1.087k_t^{male} + 3.446 - k_t^{female} = 0$ 으로 추정되었다. 또한 α 에 대한 추정치 3.446을 이용하여 $H_0: \alpha = 0$ 을 검정한 결과, 매우 유의하게(p -값이 0.0001 이하) H_0 를 기각하였다.

이상에서 살펴본 단위근 검정과 공적분 검정 등을 통해, 우리나라의 k_t^{male} 과 k_t^{female} 은 $3.446 + 1.087k_t^{male} - k_t^{female} = 0$ 의 장기균형관계를 가지고 있으므로 우리나라 사망률 자료에 LL모형의 적용은 문제가 있는 것으로 나타났다.

LL모형의 또 하나의 치명적인 결점은 〈그림 11〉에서 볼 수 있듯이 $k_{1,t}$ 와 $k_{2,t}$ 모두 크게 증가했다가 감소하는 큰 파장을 가지고 있으며, 여성의 경우 2011~2015년 사이의 파장이 매우 커 동 기간의 사망률이 〈그림 12〉-(c)와 같이 시간이 지남에 따라 증가하는 역전 현상이 발생하게 된다. 동일한 현상이 통계청의 사망률 추계에서도 발생하고 있다.



〈그림 12〉 LL 모형에 의해 예측된 85~98세 남성 및 여성사망률과 평균수명

또한 이러한 역전현상은 <그림 12>-(d)와 같이 우리나라 여성의 평균수명의 예측치가 출렁거리는 현상으로 귀결되고 있다. 그러므로 개별시간 효과 $k_{1,t}$ 와 $k_{2,t}$ 의 변동성이 우리나라와 같이 크게 나타나면 LL모형의 적용에 신중해야 한다.

IV. 벡터오차수정 LC 모형과 모형비교

1. VECM-LC 모형

우리나라 남성과 여성의 로그사망률의 시간효과는 1차 단위근을 가지며 공적분 관계를 가지고 있음을 논의하였다. 또한 이들 시간효과는 서로 간에 상관관계를 가지고 있을 가능성이 높다. 이처럼 경제사회적으로 유사한 그룹의 시간효과 간에 공적분 관계를 가지고 있으며 서로 간에 상관관계를 가지고 있을 때 사용할 수 있는 계량경제학적 모형은 벡터오차수정모형(Vector Error Correction Model : VECM)이다. 이를 이용한 새로운 LC 모형을 VECM-LC 모형으로 정의하고 구체적인 모형은 다음과 같다.

$$\begin{aligned} \log m_{i,x,t} &= a_{i,x} + b_{i,x} k_{i,t} + \epsilon_{i,t}, \quad i = 1, \dots, M \\ \Delta \mathbf{k}_t &= \alpha \beta^T \mathbf{k}_t + \sum_{j=1}^{p-1} \Phi_j \Delta \mathbf{k}_{t-j} + \epsilon_t \end{aligned} \quad (9)$$

여기에서

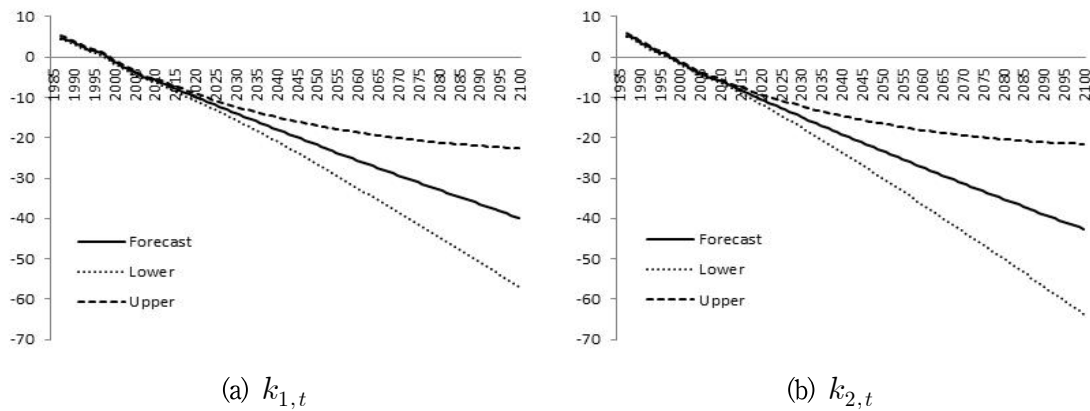
$\Delta \mathbf{k}_{t-j} = (k_{1,t-j} - k_{1,t-j-1}, k_{2,t-j} - k_{2,t-j-1}, \dots, k_{M,t-j} - k_{M,t-j-1})^T$ 이며, $\mathbf{k}_t = (k_{1,t}, k_{2,t}, \dots, k_{M,t})^T$ 이다. α 와 β 는 $M \times r$ 행렬이며 $r \leq M$ 이다. $\beta^T \mathbf{k}_t$ 는 공적분 관계를 나타내고 α 는 $\mathbf{k}_t$ 가 공적분 관계를 벗어났을 때 다시 돌아오는 속도로 해석될 수 있는 계수를 나타낸다(Enders 2004).

우리나라 남녀의 로그사망률의 시간효과 $k_{1,t}$ 와 $k_{2,t}$ 에 적용하여 구한 VECM 모형은

$$\begin{aligned} \begin{pmatrix} k_{1,t} - k_{1,t-1} \\ k_{2,t} - k_{2,t-1} \end{pmatrix} &= \begin{pmatrix} -0.18 \\ -0.24 \end{pmatrix} (1.087k_{1,t} - k_{2,t} + 3.446) \\ &\quad + \begin{pmatrix} -0.1 & -0.16 \\ 0.1 & -0.67 \end{pmatrix} \begin{pmatrix} k_{1,t-1} - k_{1,t-2} \\ k_{2,t-1} - k_{2,t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \end{aligned}$$

이며 $\epsilon_{1,t}$ 와 $\epsilon_{2,t}$ 의 상관계수는 0.6으로 추정되었다. 그러므로 우리나라 남녀의 로그사망률의 시간효과는 1차 단위근을 가지고 있으며, $1.087k_{1,t} - k_{2,t} + 3.446 = 0$ 의 공적분 관계(또는 장기균형관계)를 가지고 있다. 또한 이들 시간효과는 1차 자기상관관계를 가지고 있으며 0.6의 횡단면적(cross-sectional) 상관계수를 가지는 동적구조를 가지고 있다.

VECM-LC 모형에 의해 추정된 $a_{i,x}$ 와 $b_{i,x}$ 는 LC 모형과 동일하지만 $k_{i,t}$ 의 값은 다르게 추정된다. <그림 13>은 VECM-LC 모형에 의해 추정된 우리나라 남성($k_{1,t}$)와 여성($k_{2,t}$)의 적합값과 예측치의 플롯이다.



<그림 13> VECM-LC 모형에 의한 우리나라 남성과 여성의 시간효과 $k_{1,t}$ 와 $k_{2,t}$ 의 예측치 및 신뢰구간

<그림 13>에서 볼 수 있듯이 VECM-LC 모형의 $k_{1,t}$ 와 $k_{2,t}$ 의 신뢰구간이 <그림 11>에 제시된 LL모형의 공통 K_t 와 <그림 6>에 제시되어 있는 개별 LC 모형의 k_t^{male} 과 k_t^{female} 보다 훨씬 넓다는 것을 알 수 있다. 이와 같이 VECM-LC 모형의 $k_{i,t}$ 의 신뢰구간이 넓은 이유는 LC 모형이나 LL모형과 다르게 절편이 공적분 관계에서만 존재하기 때문이다. LC 모형이나 LL모형과 같이 지나치게 좁은 신뢰구간은 사망률의 상한과 하한이 지나치게 근접하게 예측되어 인구통계학적 해석상 문제점으로 지적되어 왔다. VECM-LC 모형은 이러한 문제점을 어느 정도 해결하는 것으로 판단할 수 있다.

2. 모형비교

우리나라 남녀 사망률 예측을 위한 모형으로 남성과 여성을 독립적으로 적용하는 LC 모형과 2개의 주성분을 이용한 LC 모형, 남성과 여성을 동시에 적용하는 LL 모형과 새로운 VECM-LC 모형을 비교하고자 한다. 모형 적합도는 1985~2010년의 1~84세까지의 실제 관측된 사망률을 이용하여 측정할 것이다. 모형 적합도를 측정하기 위한 통계량은 평균오차(Mean Error : ME)로 모형의 편의(bias)를 측정하고, 평균절대오차(Mean Absolute Error: MAE)와 평균절대퍼센트오차(Mean Absolute Percentage Error : MAPE)로 모형의 정밀도를 측정할 것이다. 이 모든 통계량의 값이 0에 가까울수록 좋은 모형이라고 할 수 있다.

독립적인 LC 모형은 남성의 경우, 110개의 a_x , 110개의 b_x 그리고 1개의 k_t 의 모수를 측정해야 하고 여성의 경우, 112개의 a_x , 112개의 b_x 그리고 2개의 k_t 의 모수를 추정해야 하므로 총 447개의 모수를 이용하여 로그사망률을 추정하게 된다. 2개의 주성분을 사용하는 LC 모형(2LC)은 총 895개의 모수를 추정해야 하며 LL모형은 668개의 모수를, 그리고 VECM-LC 모형은 452개의 모수가 추정되어야 한다.

이러한 모수의 차이를 상쇄시키기 위해 ME, MAE, 그리고 MAPE를 계산할 때 예를들어, $MAE = \frac{1}{nT} \sum_{x=1}^n \sum_{t=1}^T |m_{x,t} - \hat{m}_{x,t}|$ 대신에 $\frac{1}{df} \sum_{x=1}^n \sum_{t=1}^T |m_{x,t} - \hat{m}_{x,t}|$ 를 사용하기로 한다. 여기에서 $\hat{m}_{x,t}$ 는 $m_{x,t}$ 의 추정치이며 $df=nT-$ 모수의 총 수로 정의하며 일반적으로 자유도라고 한다.

〈표 4〉는 우리나라 성별·연령별 사망자수와 기대여명에 대한 모형별 적합도를 보여주고 있다. 남성의 경우 VECM 모형이 가장 우수하고, 여성의 경우 사망률은 VECM 모형이, 기대여명은 2LC 모형이 가장 우수한 것으로 나타났다. 기대여명의 경우, 관측 기대여명과 추정치의 차이는 MAPE에서 볼 수 있듯이 남성은 0.9% 이하, 여성은 1.5% 이하로 적합도가 모형에 관계없이 우수하게 나타났다. 사망률의 경우, MAPE 관점에서 남성은 2LC 모형을 제외하고 9%이하이고 여성은 11%~13.3%로 여성사망률의 정밀도를 더 높이기 위한 모형개발이 필요하다고 할 수 있다.

〈표 4〉 성별·연령별 사망률 및 기대여명에 대한 적합도

	모형 적합도	남				여			
		LC	2LC	LL	VECM	LC	2LC	LL	VECM
사망률	ME	0.000	0.000	0.000	0.000	-0.001	-0.002	-0.002	-0.001
	MAE	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	MAPE	8.527	10.321	8.871	8.493	11.351	13.309	12.274	11.318
기대여명	ME	-.016	-.017	-.065	-.019	-.013	.001	.209	-.012
	MAE	.185	.159	.196	.161	.333	.178	.337	.369
	MAPE	.760	.720	.916	.675	1.358	.678	1.409	1.485

〈표 5〉 모형별 평균수명

연도	남성				여성			
	LC	2LC	LL	VEC	LC	2LC	LL	VEC
2011	76.87	77.17	77.04	76.81	83.81	84.68	84.38	83.73
2020	80.06	80.50	80.19	79.76	87.12	88.29	84.45	86.79
2030	83.26	83.83	83.64	82.67	90.56	91.87	88.04	89.88
2040	86.13	86.81	86.76	85.25	93.75	95.09	90.22	92.72
2050	88.72	89.47	89.62	87.55	96.63	97.92	92.38	95.28
2060	91.04	91.84	92.26	89.61	99.17	100.37	94.71	97.55
2070	93.11	93.93	94.69	91.44	101.35	102.42	96.57	99.53
2080	94.95	95.78	96.90	93.07	103.20	104.13	98.58	101.25
2090	96.59	97.40	98.90	94.52	104.73	105.53	100.31	102.71
2100	98.03	98.82	100.71	95.82	106.00	106.68	101.97	103.96

〈표 5〉는 모형별 평균수명을 정리한 표이다. 남성의 평균수명은 VECM 모형이 가장 낮게, LL 모형이 가장 높게, 그리고 여성의 평균수명은 LL모형이 가장 낮게, LC와 2LC 모형이 가장 높게 예측하고 있다.

VECM 모형의 평균수명이 LC 또는 2LC 모형보다 낮게 나타난 주된 이유는 VECM 모형에 반영된 공적분 관계와 상관관계 때문이다. LL 모형의 특성상 현재 평균수명이 낮은 남성은 빠른 속도로 기대수명을 증가시키고, 현재 평균 수명이 높은 여성은 느린 속도로 기대수명을 증가시켜 궁극적으로 같은 기대수명을 갖도록 하기 때문에 남성의 평균수명을 가장 높게, 여성의 평균수명을 가장 낮게 예측하는 현상이 나타난다.

V. 결론

정밀도 높은 미래의 사망률 예측은 인구 고령화, 재정 추계, 자산 배분, 복지 정책 연구 등에 필수적 요소이다. 정밀도 높은 사망률 예측을 위해서는 양질의 사망률 자료가 필요하며, 이를 위해 그동안 신뢰성에 의구심이 있었던 70년대 사망률 자료의 재검점의 필요성이 요구되었다. 1974년 74세 코호트의 사망률이 1976년까지 예외적으로 낮게 나왔으며 30대 후반 연령과 70~84세의 연령에서 1970년대 평균사망률이 1980년대 평균사망률보다 높게 나와 문제점을 드러내고 있음을 보였다. 또한 초고령 연령으로의 확장을 위해 사용하는 Brass-logit 모형의 모수를 이용하여 1970년대 자료는 신뢰성에 심각한 문제가 있음을 보여주었다.

사망률 추계를 위한 LC 모형, 2개의 주성분을 이용한 LC 모형(2LC), 유사한 사회·경제적 배경을 가진 그룹은 장기적으로 동일한 사망률 패턴을 가진다는 Li and Lee 모형(LL모형), 2개 이상의 그룹의 사망률의 시간 패턴이 공적분 관계(장기균형관계)를 가지고 있으며 상관관계를 가지고 있을 때 새로운 VECM-LC 모형을 소개하고 비교하였다. 우리나라 남성과 여성의 사망률에 2LC 모형과 LL 모형의 적용은 문제가 있음을 보여주었다. 특히, 우리나라 남성과 여성의 사망률의 패턴은 오히려 시간이 갈수록 상이한 현상을 보이고 있다는 점, 공적분 검정을 통해 남성과 여성의 시간효과는 동일한 값으로 수렴하지 않는다는 것을 보여준 점, 남성과 여성의 개별 시간효과가 파동이 크게 나와 사망률의 역전현상과 여성의 평균수명이 출렁거리는 형태로 예측된다는 점 등을 보여줌으로써 우리나라 통계청에서 사용하고 있는 LL 모형은 우리나라 남녀사망률자료에 적합하지 않다는 것을 보여주었다.

모형의 적합도를 통해 새롭게 제안한 VECM-LC 모형이 가장 우수한 것으로 나타났다. VECM-LC 모형의 특징은 사망률의 하한과 상한값이 다른 모형에 비해 넓게 나오고, 기존 LC 모형보다 평균수명의 증가속도가 느리게 나와 선진 고령화 국가에서 경험한 인구통계학적 관점에서 좀 더 현실적이라고 할 수 있다.

그러나 새롭게 제시한 VECM-LC 모형을 포함한 LC류의 모형에서 연령효과가 시간에 의존하지 않는다는 지나친 단순화로 인해, 예측오류가 8.5%~13.3%로 비교적 큰 값을 보이고 있어 좀 더 정밀한 모형 연구가 요구된다고 할 수 있다.

참고문헌

- 김성용 · 김기환 · 박유성. 2011. “우리나라의 초고령 사망률의 확장에 대한 연구.” 《조사연구》 12(2): 1-26.
- 박유성 · 김기환. 2011. 《SAS/ETS를 이용한 시계열자료분석 I》. 서울: 자유아카데미.
- 박유성 · 김기환 · 이동희 · 이연경. 2005. “사망률 예측을 위한 모형 비교.” 《응용통계연구》 18(3): 639-654.
- 박유성 · 김성용. 2011. “우리나라 자료에 적합한 생명표 작성방법에 대한 연구.” 《응용통계연구》 24(5): 769-789.
- 정승환. 2011. 《우리나라 생명표 예측에 관한 연구: 사망확률 검토, 사망률 예측과 생명표 분석》. 고려대학교 박사학위 논문.
- Bell, W.R. and B.C. Monsell. 1991. “Using Principal Components in Time Series Modeling and Forecasting of Age-Specific Mortality Rates.” *Proceedings of the Social Statistics Section, American Statistical Association* 154-159.
- Bennett, N.G. and S. Horiuchi. 1981. “Estimating the Completeness of Death Registration in a Closed Population.” *Population Index* 47(2): 207-221.
- Beuttner, T. 2003. “Approaches and Experiences in Projecting Mortality Patterns for the Oldest-Old.” *North American Actuarial Journal* 6(3): 14-29.
- Boleslawski, L. and E. Tabeau. 2001. “Comparing Theoretical Age Patterns of Mortality beyond the Age of 80.” in E. Tabeau, A. Van Den Berg Jeths, and C. Heathcote (eds.), *Forecasting Mortality in Developed Countries Insights from a Statistical, Demographic and Epidemiological Perspective*. Dordrecht: Kluwer Academic Publishers.

- Brass, W. 1977. "On the Scale of Mortality." in W. Brass(ed.), *Biological Aspects of Demography*. London: Taylor and Francis.
- Brouhns, N., M. Denuit, and J. Vermunt. 2002. "A Poisson Log-Bilinear Regression Approach to the Construction of Projected Lifetables." *Insurance: Mathematics and Economics* 31(3): 373-393.
- Coale, A.J. and E.E. Kisker. 1990. "Defects in Data on Old-Age Mortality in the United States: New Procedures for Calculating Schedules and Life Tables at the Highest Ages." *Asian and Pacific Population Forum* 4(1): 1-31.
- Condran, G., C.L. Himes, and S.H. Preston. 1991. "Old Age Mortality Patterns in Low-Mortality Countries: An Evaluation of Population and Death Data at Advanced Ages, 1950 to Present." *Population Bulletin of the United Nations* 30: 23-61.
- Dickey, D. and W.A. Fuller. 1979. "Distribution of the Estimates for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427-431.
- Enders, W. 2004. *Applied Econometric Time Series*. New York: Wiley.
- Engle, R.F. and C.W.J. Granger. 1987. "Co-Integration and Error-Correction: Representation, Estimation and Testing." *Econometrica* 55: 251-76.
- Gompertz, B. 1825. "On the Nature of the Function Expressive of the Law of Mortality." *Philosophical Transactions* 27: 513-585.
- Hatzopoulos, P. and S. Haberman. 2009. "A Parameterized Approach to Modelling and Forecasting Mortality". *Insurance: Mathematics and Economics* 44(1): 103-123.
- Heligman, L. and J.H. Pollard. 1980. "The Age Pattern of Mortality." *Journal of the Institute of Actuaries* 107: 49-80.
- Hill, K. 1997. "The Measurement of Adult Mortality: An Assessment of Data Availability, Data Quality and Estimation Methods: in Health and Mortality Issues of Global Concern." *Proceedings of the Symposium on Health and Mortality*. Brussels: Population and Family Study Centre.
- Hyndman, R.J. and M.S. Ullah. 2007. "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics & Data Analysis* 51(10): 4942-4956.
- Johansen, S. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59(6): 1551-1580.
- Lee, R. and T. Miller. 2001. "Evaluating the Performance of the Lee Carter Mortality Forecasts." *Demography* 38(4): 537-549.

- Lee, R.D. and L.R. Carter. 1992. "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association* 87: 659–671.
- Li, N. and R. Lee. 2005. "Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee–Carter Method." *Demography* 42(3): 575–594.
- McNown, R. and A. Rogers. 1989. "Forecasting Mortality: A Parametrized Time Series Approach." *Demography* 26: 645–660.
- McNown, R. and A. Rogers. 1992. "Forecasting Cause–Specific Mortality Using Time Series Methods." *International Journal of Forecasting* 8: 413–32.
- Pitacco, E. 2004. "Survival Models in a Dnamic Context: A Survey." *Insurance Mathematics & Economics* 35: 279–298.
- Renshaw, A.E. and S. Aberman. 2003. "Lee–Carter Mortality Forecasting with Age–Specific Enhancement." *Insurance: Mathematics and Economics* 33: 255–72.
- Renshaw, A.E., and S. Haberman. 2006. "A Cohort–Based Extension to the Lee–Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics* 38: 556–70.
- Russolillo, M., G. Giordano, and S. Haberman. 2011. "Extending the Lee–Carter Model: A Three–way Decomposition." *Scandinavian Actuarial Journal* 2011(2): 96–117.
- Statistics Canada. 2009. *Population Projections for Canada, Provinces and Territories 2009–2036*. Ottawa: Statistics Canada.
- Tucker, L.R. 1966. "Some Mathematical Notes on Three–mode Factor Analysis." *Psychometrika* 31(3): 279–311.
- Wilmoth, J.R. 1995. "Are Mortality Rates Falling at Extremely High Ages? An Investigation Based on a Model Proposed by Coale and Kisker." *Population Studies* 49(2): 281–295.
- Yue, J.C., S.S. Ynag, and H.C. Hauang. 2008. "A Study of the Lee–Carter Model with Age–Shifts." *Living to 100 and Beyond Symposium, Society of Actuaries* 1–20.

부록. 표준 사망률

연령	남자	여자	연령	남자	여자
0	0.00388	0.00338	57	0.00862	0.00293
1	0.00042	0.00039	58	0.0092	0.00319
2	0.00029	0.00023	59	0.00989	0.00348
3	0.00025	0.0002	60	0.01074	0.00385
4	0.0002	0.00016	61	0.01174	0.00427
5	0.00019	0.00015	62	0.01299	0.00477
6	0.00019	0.00014	63	0.0144	0.00535
7	0.00019	0.00013	64	0.01593	0.00597
8	0.00015	0.00011	65	0.01762	0.00668
9	0.00015	0.0001	66	0.01938	0.00746
10	0.00014	0.00009	67	0.02124	0.00834
11	0.00012	0.00009	68	0.02327	0.00936
12	0.00016	0.00011	69	0.02539	0.01055
13	0.00017	0.00014	70	0.02772	0.01195
14	0.00021	0.00014	71	0.03032	0.01357
15	0.00027	0.00016	72	0.03315	0.01541
16	0.00035	0.0002	73	0.0365	0.01761
17	0.00045	0.0002	74	0.04034	0.02021
18	0.00047	0.00023	75	0.04481	0.02327
19	0.00047	0.00027	76	0.05012	0.02691
20	0.00047	0.00028	77	0.05596	0.03101
21	0.00047	0.00034	78	0.06254	0.03556
22	0.00057	0.00039	79	0.06979	0.04067
23	0.0006	0.00043	80	0.07747	0.04631
24	0.00066	0.00042	81	0.0864	0.05288
25	0.00074	0.00049	82	0.09675	0.06049
26	0.00072	0.00049	83	0.10764	0.06919
27	0.00082	0.00046	84	0.11957	0.07943

28	0.00083	0.00047	85	0.13262	0.09065
29	0.00083	0.00051	86	0.14682	0.10322
30	0.00087	0.00055	87	0.16225	0.11766
31	0.00092	0.00056	88	0.17895	0.13097
32	0.00096	0.00059	89	0.19698	0.14661
33	0.00098	0.0006	90	0.2164	0.1636
34	0.00108	0.00061	91	0.23726	0.18199
35	0.00118	0.00064	92	0.25963	0.20183
36	0.00135	0.0007	93	0.28355	0.2232
37	0.00143	0.00073	94	0.30909	0.24614
38	0.00162	0.00082	95	0.3363	0.27072
39	0.00178	0.00086	96	0.36523	0.29699
40	0.00201	0.00089	97	0.39595	0.32501
41	0.00223	0.00092	98	0.4285	0.35485
42	0.00247	0.00096	99	0.46295	0.38655
43	0.00282	0.00107	100	0.49936	0.42019
44	0.00304	0.00116	101	0.53777	0.45581
45	0.0034	0.00127	102	0.57824	0.49348
46	0.00374	0.00137	103	0.62084	0.53325
47	0.00408	0.00146	104	0.66561	0.57518
48	0.00443	0.00156	105	0.71262	0.61934
49	0.0048	0.00166	106	0.76192	0.66578
50	0.00518	0.00178	107	0.81356	0.71456
51	0.0056	0.00191	108	0.86761	0.76574
52	0.00605	0.00205	109	0.92411	0.81937
53	0.00651	0.0022	110	0.996	0.87552
54	0.00701	0.00235	111		0.93424
55	0.00753	0.00252	112		0.9956
56	0.00804	0.00271			

