

## A comparison of mortality projection by different time period in time series

Soon-Young Kim<sup>a,1</sup> · Jinho Oh<sup>a</sup> · Kee-Whan Kim<sup>b</sup>

<sup>a</sup>Statistical Research Institute, Statistics Korea;

<sup>b</sup>Department of Applied Statistics, Korea University

(Received September 5, 2017; Revised November 13, 2017; Accepted December 15, 2017)

---

### Abstract

In Korea, as the mortality rate improves in a shorter period of time than in developed countries, it is important to consider the selection of the time series as well as the model selection in the mortality projection. Therefore, this study proposed a method using the multiple regression model in respect to the selection of the time series period. In addition, we investigate the problems that arise when various time series are used based on the Lee-Carter (LC) model, the kinds of LC model along with Lee-Miller (LM) and Booth-Maindonald-Smith (BMS), and the non-parametric model such as functional data model (FDM) and Coherent FDM, and examine differences in the age-specific mortality rate and life expectancy projection. Based on the analysis results, the age-specific mortality rate and predicted life expectancy of men and women are calculated for the year 2030 for each model. We also compare the mortality rate and life expectancy of the next generation provided by Korean Statistical Information Service (KOSIS).

Keywords: mortality projection model, life expectancy, age-specific mortality rates, Lee-Carter model, FDM, Coherent FDM

---

### 1. 서론

미래의 사망률을 예측하기 위해 과거의 추세를 잘 반영하는 사망률 모형의 선택도 중요하지만 서로 다른 시계열 이용기간에 따라 결과가 상이하게 도출되는 경우가 많다 (Cairns 등, 2009).

일반적으로 장기간(long-term)의 동일 시계열 확보가 예측이나 적합에서 중요하지만 현실은 그렇지 못하다. UNPD (2015)에 따르면 219개국 중 89개(40%) 나라만 인구동태통계(vital statistics)의 품질(data quality)과 완비성(completeness)이 우수한 것으로 보고되었으며 39개국의 상세한 사망률 및 인구 데이터를 제공하기 위해 작성된 데이터베이스인 Human Mortality Database (HMD)에 따르면 스웨덴(1751–2014), 덴마크(1835–2014), 핀란드(1878–2015), 프랑스(1816–2015)는 장기간의 기간 생명표(period life table)을 제시하는 국가인 반면 이스라엘(1983–2014), 칠레(1992–2005), 크로아티아(2002–2015)는 상대적으로 매우 짧다.

연구는 통계청의 공식견해가 아니며 저자의 개인적인 연구결과임을 밝힙니다.

<sup>1</sup>Corresponding author: Statistical Analysis Division Statistical Research Institute, 6F, Statistical Center, 713 Hanbatdaero, Seo-gu, Daejeon 35220, Korea. E-mail: [sykim0213@korea.kr](mailto:sykim0213@korea.kr)

유엔 인구처(United Nations Population Division; UNPD)는 선진국의 패턴 및 사전 지식과 국제기구 등의 서베이 자료 등을 확보한 후 베이지안 방법을 통해 동일 시계열을 생성한다. 따라서 국가별 비교를 위해서는 기간의 동일성이 중요한 고려사항이다. 사망률 시계열 적용기간을 어떻게 설정할 것인가에 대한 해외 연구로 Cairns 등 (2009)은 장래 사망률 예측을 위해서는 시계열 적용기간에 따른 모수 추정값의 로버스트 문제 또한 중요한 이슈라고 강조하였으며, 국내 연구로 Jung과 Kim (2012), Jung 등 (2013)은 1983-2005년까지의 시계열을 이용하여 예측한 결과와 1976-1982년까지의 시계열을 추가해서 도출된 결과가 차이가 있음을 확인하였다.

선행연구에 따르면 사망률 예측을 위해서는 시계열 적용기간에 유의할 필요가 있으며, 서로 상이한 기간의 시계열을 적용할 때 모수추정이나 예측값에 차이가 있음을 알 수 있다.

본 연구는 기간의 동일성 측면보다는 사망률 모형으로 가장 널리 사용되는 Lee-Carter (LC)모형, LC계열의 모형(Lee-Miller (LM), Booth-Maindonald-Smith (BMS)) 그리고 비모수 모형(functional data model (FDM), Coherent FDM)에 사망률 시계열 이용기간을 다르게 적용할 경우 연령별 사망률과 기대수명 예측력에 어떠한 차이를 보이는지 검토하고자 한다. 추가적으로 예측력 비교 평가를 위해 사망률 자료를 모형적합기간(fitting period)과 모형검증기간(forecast period)으로 분할한 후 평가검증(validation test)을 수행한다. 본 연구에서 제시하고자 하는 것은 다음과 같다. 모형별 시계열 이용기간에 따른 모형 적합 설명력(자료의 총변동 중 적합된 모형에 의해 설명되는 변동의 비율, the proportion of the variation that explained by the model-specific)과 적합력을 살펴보고, 시계열 이용기간에 따른 로그사망률과 기대수명의 예측력을 평가 검증한다. 시계열 이용기간을 선정하기 위해 자료의 구조적 변화가 발생하는 시점을 탐색하였으며 이를 위해 회귀모형을 이용하는 방법을 제안하고, 미래 예측에 대한 불확실성(uncertainty)을 감안하여 연령별 로그사망률과 기대수명에 대한 구간예측을 검토하고 불확실성을 평가한다.

따라서 기존의 모형비교 (Kang 등, 2006; Park 등, 2005; Jeong과 Kim, 2011)에 의한 모형별 차이가 아닌 시계열 이용기간에 따른 비교와 평가를 도출하여 선행연구를 확장하고자 한다. 세부적으로 시계열 이용기간에 따른 각 모형별 예측력에 차이가 있는지를 알아보기 위해 민감도 테스트를 실시하였으며, 예측 정확성을 위해 점예측(point forecast)과 구간예측(interval forecast)을 동시에 검토하고 이를 구현하기 위해 R (Hyndman, 2010)의 인구 통계 패키지를 사용하였다.

본 연구의 구성은 다음과 같다. 2장에서는 사망률 자료 설명 및 시계열 이용기간 선정 방법을 제안한다. 3장에서는 모형의 예측 정확도를 평가하는 지표를 소개하고 시계열 이용기간에 따라 사망률 예측 모형을 연령별 사망률과 기대수명 측면에서 모형의 적합도 및 예측의 정확도를 비교한 주요 분석 결과들을 살펴본다. 또한 향후 15년(2016-2030)의 사망률과 예측기대수명을 작성하여 통계청에서 제공하는 장래 연령별 사망률과 기대수명과 비교한다. 끝으로 4장에서는 연구결과를 정리한다.

## 2. 데이터 설명 및 자료이용 기간 선정

### 2.1. 사망률 지표와 추세

사망률은 사망자수와 인구수를 기초로 산출된다.  $E_{x,t}$ 를  $t$ 연도에  $x$ 세인 연앙인구수(그해의 중간인 7월 1일 기준의 인구수),  $D_{x,t}$ 를  $t$ 연도에  $x$ 세인 사망자수라고 하면  $t$ 연도에  $x$ 세의 사망률(crude death rate)은 식 (2.1)과 같이 정의된다.

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}. \quad (2.1)$$

우리나라는 선진국에 비해 사망자료의 시계열이 짧으며 1970년부터 2015년까지 약 45년 정도의 연령별

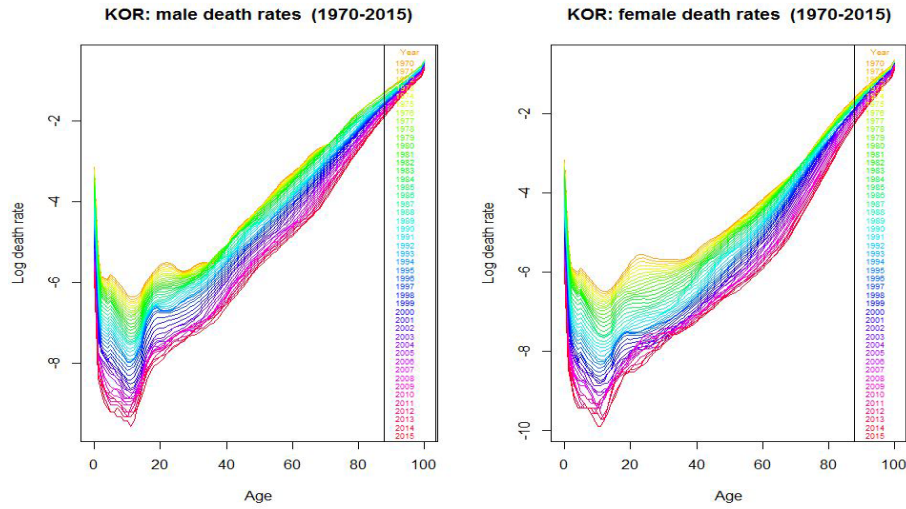


Figure 2.1. Age-specific log mortality by gender (1970-2015).

사망 자료가 있다. 사망은 신고에 의해 수집되는 인구동태자료로 신고누락, 지연 등의 기본적인 문제를 포함하고 있지만 신고율 제고와 함께 품질 개선이 이루어지고 있다. 과거 영아 사망의 경우 약 70%정도 누락되는 것으로 추정되었으나, 1999년 이후부터 영아사망 화장신고 자료를 수집 보완하면서 품질이 크게 향상되었다 (Park, 2011).

또한 2000년대 들어 사망신고의 질이 크게 향상됨에 따라 2006년 추계 이후에는 우리나라의 자료를 최대한 활용하는 방법을 적용하고 있다. 통계청은 사망관련 시계열을 보완 정비하여 품질을 크게 향상시키고 있으며 2016년 12월에는 과거 45개년(1970-2014년)의 생명표 작성방법을 일괄 개선하여 국가통계포털(Korean Statistical Information Service; KOSIS)에서 자료를 제공하고 있다.

본 연구에서는 최근 정비된 자료인 2016년에 개선한 생명표의 자료를 이용하였다. 생명표에서는 기대수명을 도출하기 위해 사용되는 사망확률( $q_x$ )이 제공된다. 사망률 모형에서는 사망확률이 아닌 사망률( $m_x$ ) 자료를 이용해야 하므로, 식 (2.2)와 같은 사망률과 사망확률의 관계식을 이용하여 생명표의 사망확률을 사망률로 변환하고 이를 이용하여 사망률 예측 모형을 검토하였다. 0-99세는 완전생명표의 사망확률을 식 (2.2)를 이용하여 사망률로 변환하였으며 100세 이상의 사망률은 생명표의 총생존자수를 총생존년수로 나누어 산출하였다.

$$q_x = \frac{m_x}{1 + \frac{1}{2}m_x}. \quad (2.2)$$

부록 A는 1970-2015년 남자와 여자의 각세별 사망률 시계열 그림을 나타낸 것이다. 1970년대 남녀 사망률 모두 높은 수치로 변동이 크나 1983년 이후로는 비교적 안정적으로 감소하는 패턴을 보이고 있다.

Figure 2.1은 우리나라 1970-2015년 성별에 따른 연령별 로그사망률을 나타낸 것으로 로그사망률(log-mortality)은 식 (2.3)과 같이 정의된다. 로그사망률은 J자 커브를 그리는데 0세의 사망률은 높고 5세 전까지는 점진적으로 감소하다가 10-15세 이후부터 증가하기 시작해 고령층으로 이어진다.

$$\log(m_{x,t}) = \log\left(\frac{D_{x,t}}{E_{x,t}}\right). \quad (2.3)$$

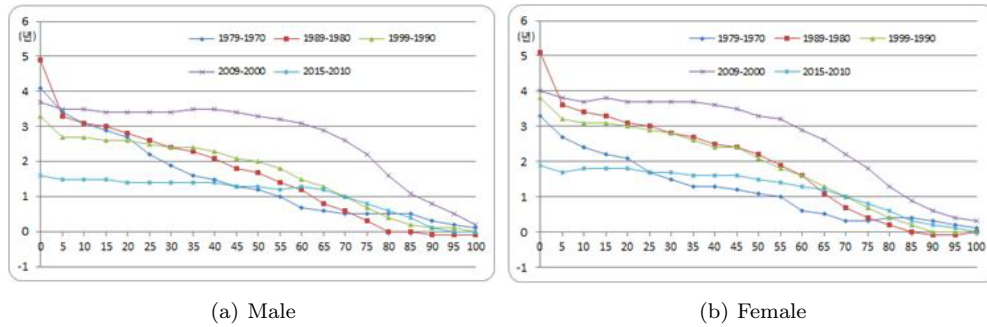


Figure 2.2. Change of Life expectancy within various period.

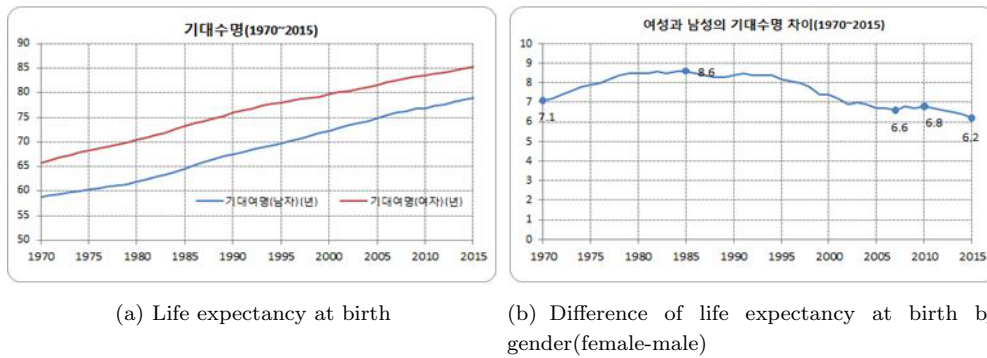


Figure 2.3. Trend of life expectancy at birth.

Figure 2.1에 의하면 최근에 로그사망률이 하강하는 추세를 보인다. 이는 사망률 개선이 진행되고 있음을 나타내며, 우리나라는 1970년 이후 연령별 사망률 개선이 급격히 이루어지고 있고 최근 2000년 이후 안정화 추세를 보인다(부록 A 참조).

사망수준이 높았을 때인 1990년대 이전에는 영유아의 사망수준 개선이 수명증가에 중요한 역할을 하였지만 시간이 흘러 2000년대 이후로는 중장년 및 노년층의 사망수준 개선이 주요한 요인이 되고 있다. 이와 같은 현상은 Figure 2.2의 연령에 따른 성별 기대여명의 증가 추이를 통해서도 확인가능하다. 특히 60-79세 남자의 사망수준이 빠르게 감소하여 남녀의 기대수명 차이를 좁히고 있다.

Figure 2.3(a)에 의하면 1970년의 기대수명은 남자가 58.7세, 여자가 65.8세였으나 2015년 남자 79.0세, 여자 85.2세로 증가하여 46년 동안 기대수명이 20년 정도 개선되었고, Figure 2.3(b)의 남녀 기대수명의 차이는 1970년 7.1세에서 점차 증가하여 1985년 8.6세로 정점에 도달한 후 점차 감소하여 2015년 6.2세이다.

## 2.2. 사망률 추정과 예측모형

1992년 인구통계학자인 Lee와 Cater에 의해 제안된 LC 모형은 미래 사망률을 예측하는 대표적인 모형이다. 그들이 제안한 로그 연령별 사망률(age-specific mortality) 예측모델의 구조는 식 (2.4)와 같다.

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \quad t = 1, \dots, n, \quad (2.4)$$

여기서  $m_{x,t}$ 는 연령  $x$ 와 시간  $t$ 에서의 사망률(death rate),  $a_x$ 는 연령에 따른 평균적인 로그사망률의 수준,  $b_x$ 는 각 연령에서 로그사망률의 변화,  $k_t$ 는 시간  $t$ 에 따른 로그사망률 변화 수준,  $\epsilon_{x,t}$ 는 평균 0, 분산  $\sigma_\epsilon^2$ 인 관측되지 않은 오차항을 의미한다.

따라서 LC 모형은 로그 변환된 사망률을 연령효과(age effect)와 시간효과(period effect)의 선형으로 표현하는 모형이다.

특히  $a_x$ 와  $b_x$ 는 시간에 무관한 상수이지만 시간에 따른 전반적인 사망률 개선정도를 나타내는 사망률 지수(mortality index)인  $k_t$ 만이 시간에 의존하므로 추정된  $k_t$ 의 예측을 통해 미래 사망률을 예측하게 된다.

Lee와 Carter (1992)는  $b_x$ 와  $k_t$ 의 유일한 추정치를 보장하기 위해  $\sum_t k_t = 0$ ,  $\sum_x b_x = 1$ 과 같은 제약 조건을 부과하였으며 각 모수를 추정하는 방법은 다음과 같다.

모수  $a_x$ 는 시간에 대한  $\ln(m_{x,t})$ 의 평균,  $b_x$ 와  $k_t$ 는  $[\ln(m_{x,t}) - a_x]$ 에 비정칙분해(singular value decomposition; SVD)방법을 적용하여 얻어진 첫 번째 주성분을 사용하여 추정된다. 또한, 추정된  $\hat{k}_t$ 는 사망률이 아닌 로그사망률에 가까워지도록  $k_t$ 를 추정한 것이므로 실제 데이터인 총사망자수  $D_t = \sum_x D_{x,t}$ 의 기준에서  $\hat{k}_t$ 를 조정한다.

식 (2.4)에서 제시한 바와 같이 LC 모형은 간단(simple)하고 단기간 시계열의 사망률 예측이 용이하다. 또한 예측 면에서도 비교적 우수하다고 알려져 있어 사망률 연구에 가장 기본적인 모형으로 현재까지 널리 사용되고 있다. 특히, 실무적용 면에서 용이성과 안정적인 결과를 제공해주는 장점이 있으나, 장기 간보다는 단기간(short-term) 예측에 우월, 새로운 추계를 시작하는 시점에서 발생하는 편의(jump-off bias), 지역별(sub-region) 사망률 예측시 사망률 교차(cross-over) 등 실제 적용상에서 여러 가지 문제점들이 발견되었으며 이를 해결하기 위한 아래와 같은 개선 모형들이 제안되었다.

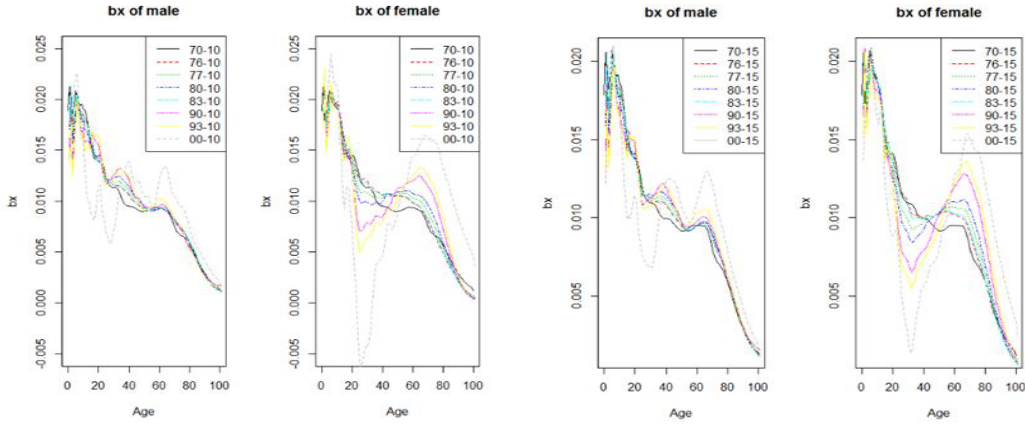
모형 개선을 모수와 비모수모형으로 구분해보면 모수 모형으로 LM 모형 (Lee와 Miller, 2001)은  $b_x$ ,  $k_t$ , 그리고 점프 편의(jump-bias)문제를 수정한 모형이며, BMS 모형 (Booth 등, 2002)은 총사망자( $D_t$ )나 기대수명 대신 연령별 사망자수( $D_{x,t}$ )를 사용하고 포아송 회귀모형을 적용하여 사망률 모델링한다. 비모수 모형으로는 관측치에 존재하는 측정오차와 질병이나 전쟁 등으로 인구동태 자료에서 나타나는 불규칙적인 패턴 등을 교정(비모수 평활기법 사용)한 함수적 자료분석(functional data analysis)모형이 있다. FDM (Hyndman과 Ullah, 2007)과 Coherent FDM (Hyndman 등, 2013)이 여기에 속한다. 이들에 대한 자세한 내용은 참고문헌을 참조하기 바란다.

### 2.3. 회귀모형을 이용한 시계열 이용기간 선정

사망률 모형에 의한 미래의 사망률 예측은 일반적으로 과거 자료의 추세가 미래에도 계속 지속될 것이란 가정하에 이루어진다. 앞 절에서 살펴본 바와 같이 우리나라는 사용 가능한 시계열 자료가 1970-2015년이며 길지 않은 기간 동안 사망률 개선이 급속히 이루어졌다. 또한 1982년 이전의 자료는 고연령층의 자료 부족으로 완비성(completeness)이 떨어져, 1983년 이후 자료를 정비하였다는 점 등을 고려할 때 사망률 예측을 위한 모형 선정과 더불어 예측에 사용할 시계열 이용기간의 적정성 검토도 필요하다.

따라서 사망률 모형 적합에 사용되는 시계열 이용기간에 따라 모형의 예측력에 어떤 변화가 있는지를 살펴보기 위해 가장 전형적인(typical) 사망률 모형인 LC 모형을 사용하여 시계열 이용기간을 선정한다.

Figure 2.4와 같이 LC 모형을 이용하여 자료 적합을 위한 시계열 이용 시작시점과 종료시점을 변화시켜가며 시계열 이용기간에 따른 연령별 사망률 변화를 나타내는  $b_x$ 를 살펴보았다.  $b_x$ 의 절대값이 크면 연령  $x$ 에서 사망률의 변화가 크다는 것을 의미한다. (a)와 (b)에 따르면 남, 여 모두 1990년 이후 시점부



(a) Ending point of fitting period: 2010

(b) Ending point of fitting period: 2015

**Figure 2.4.** Change of log mortality by various fitting period.

더 큰 변화를 나타내는 것을 알 수 있다. 특히 남자보다는 여자가 상대적으로 큰 변화를 보이는 것이 특징이다.

본 연구는 자료이용 시작시점을 선택하기 위한 방법으로 LC 모형에서 연령별 사망률의 변화를 나타내는  $b_x$ 가 변화하는 시점을 찾기 위해 회귀분석을 이용한다. 즉, 자료적합 종료시점을 2015년으로 고정하고 자료 적합 시작시점을 1970년부터 2005년까지 1년씩 증가시켜가며 LC 모형을 적용하여 연령별 사망률 변화  $b_x$ 를 구해보았다. 즉, 시계열 이용 기간에 따라 총 36개의 자료구간을 이용하여 LC 모형의 연령별 사망률  $b_x$ 를 추출하였다.

$b_x$ 의 패턴을 통해 적절한 시계열 이용 시작시점을 탐색하기 위해, LC 모형으로 산출한  $b_x$ 에 대하여 식 (2.5)와 같은 연령에 대한 3차식 선형모형인 중회귀분석(multiple regression)을 실시하여 도출된 회귀계수의 변화를 통해 연령별 사망률 변화 패턴이 변하는 시점을 찾아보았다.

$$b_x = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \beta_3 \text{age}^3, \quad (2.5)$$

$$b_{x,\text{male}} = \beta_{0,\text{male}} + \beta_{1,\text{male}} \text{age} + \beta_{2,\text{male}} \text{age}^2 + \beta_{3,\text{male}} \text{age}^3,$$

$$b_{x,\text{female}} = \beta_{0,\text{female}} + \beta_{1,\text{female}} \text{age} + \beta_{2,\text{female}} \text{age}^2 + \beta_{3,\text{female}} \text{age}^3,$$

여기서  $b_x$ 를 연령에 대한 3차 회귀분석을 실시한 결과 산출된 회귀계수( $\beta_0, \beta_1, \beta_2, \beta_3$ ) 검정 결과 모두 유의수준 1%에서 통계적으로 유의하게 나타났다. 본 연구는 지면관계상 36개 시계열 회귀계수 도출결과를 부록 B에 제시하였다.

Figure 2.5는  $b_x$ 를 연령에 대해 3차 회귀분석을 적합한 결과로 실선은 적합된 회귀식이고 점선은 실제 값이다. Figure 2.6은  $b_x$ 를 연령에 대해 남녀 3차 회귀모형을 적합했을 때 결정계수의 변화이다. 남자의 경우 1996년 이전에는 회귀식의 설명력이 완만히 감소하는 패턴이고 90% 이상으로 상당히 높으나, 그 이후부터 설명력이 급격히 떨어져 2002년엔 약 72% 정도 수준이다. 여자의 경우 1970년 이후 설명력이 점진적으로 하강하고 1985년을 전후로 하여 설명력이 90% 이하로 떨어지다가 2002년을 기점으로 설명력 급감하였다. 즉, 남자의 경우 1996년 이후의 자료, 여자의 경우 1985년 이후의 자료만을 사용할 경우 설명력이 90% 아래로 하회하는 수준이 될 것이며, 특히 여자는 2002년 이후의 자료만을 이용할 경우 설명력이 급감함을 보인다.

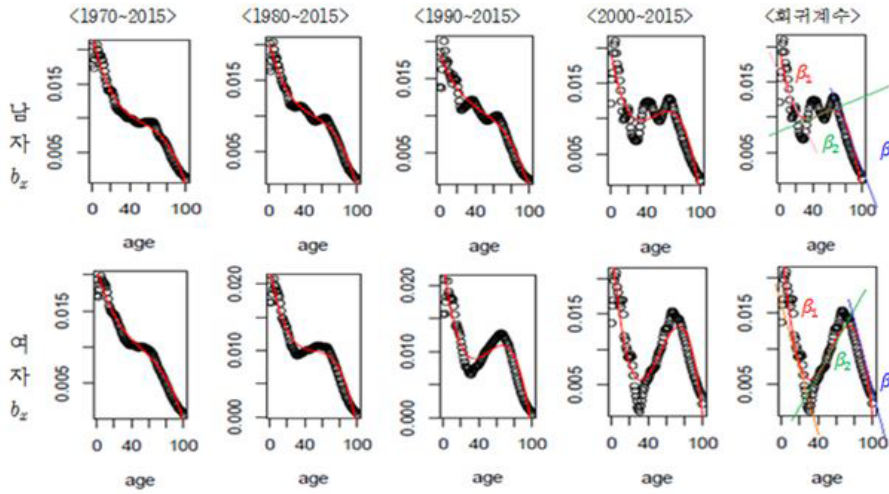


Figure 2.5. Plot of fitted  $b_x$  regression model (line: fitted value, dot: actual value).

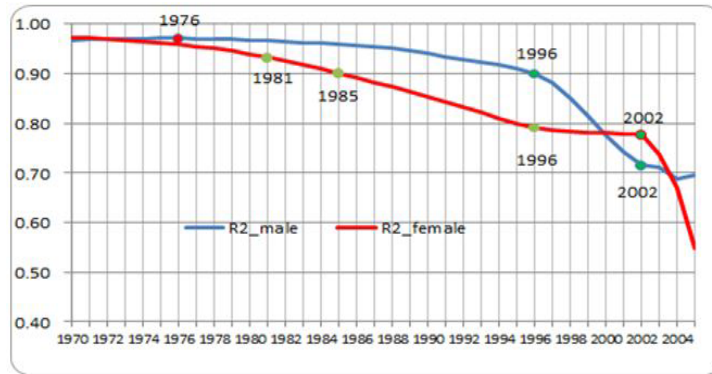


Figure 2.6. Change of R-square in  $b_x$  regression model.

일반적으로 결정계수는 모형 전체의 설명력 정도만 의미하고 회귀모형 계수값들의 유의성과 변화는 나타내지 못한다. 따라서 본 연구는 회귀계수의 변화도를 Figure 2.7과 Figure 2.8에 제시한다. 이 두 그림은  $b_x$ 에 관한 남녀 회귀모형의 계수의 변화를 나타낸다. Figure 2.7의 남자의 경우 적합된 회귀계수  $\beta_1$ 은 완만히 상승하다가 1991년을 기점으로 하락세로 돌아서며, 1994년 이후 급격히 감소하는 패턴을 보이고 있다. 계수  $\beta_2$ 는 완만히 감소하다가 1990년을 기점으로 상승세로 돌아서며, 1994년 이후 급격히 상승하는 패턴을 보인다. 계수  $\beta_3$ 는 완만히 상승하다가 1989년을 기점으로 하락세로 돌아서 1994년 이후 급격히 감소하는 패턴을 보인다.

Figure 2.8의 여자의 경우 적합된 회귀계수  $\beta_1$ 은 점진적인 하락세를 보이다가 2002년 이후 급격한 상승세를 나타내며,  $\beta_2$ 는 점진적인 상승세를 보이다가 2002년 이후 급격한 하락으로 전환되었다.  $\beta_3$ 도  $\beta_1$ 과 유사하게 점진적으로 하락하다 2002년 이후 급격한 상승세로 돌아섰다.

위의 결과에 따르면 남자의 연령별 사망률 변화  $b_x$ 패턴이 변하는 시점이 1994년이며 1996년부터는 급격히 변화하고 있으며 여자는 2002년이다. 여자의 경우 2002년 이후의 자료를 사용기에는 단기간이므



Figure 2.7. Change of regression coefficient in  $b_x$  male regression model with various fitting period.

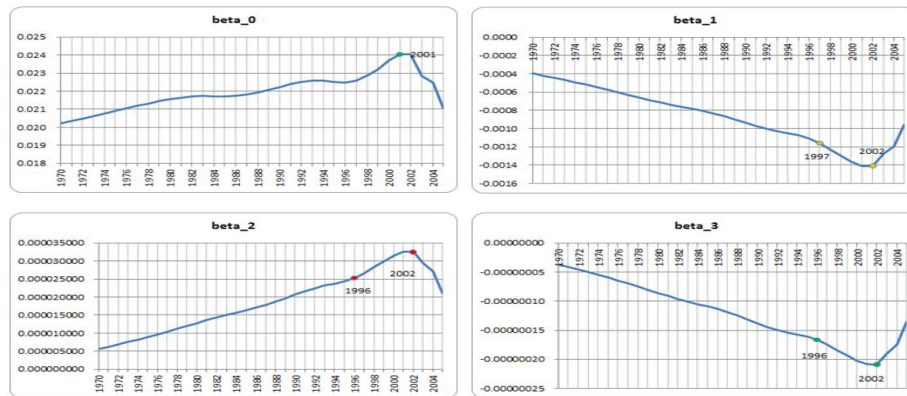


Figure 2.8. Change of regression coefficient in  $b_x$  female regression model with various fitting period.

로 2002년 이전의 연령별 사망률 변화 패턴에 작은 변화가 있는 1996년을 사용하기로 한다. 따라서 회귀모형을 이용하여 연령별 사망률 변화를 나타내는  $b_x$ 가 변하는 시점을 남, 녀 모두 1996년으로 선정하였다.

본 연구는 회귀분석의 결과와 다양한 시계열 자료 이용기간을 종합적으로 반영하기 위해 아래와 같이 5가지로 선정하여 비교한다.

- 첫째, 1970–2015년 모든 사망률 자료
- 둘째, 부록 A를 통해 확인하였듯이 1982년 이전의 자료가 고연령층 부분의 부족으로 인해 불완전하므로 사망관련 자료를 정비한 1983–2015년 자료
- 셋째, 1970년 이후 비교적 안정적 추세를 보이는 1976–1982년을 1983년 후 연속 자료에 추가한 자료
- 넷째, 연령별 사망률 변화를 나타내는  $b_x$ 의 패턴이 변화하는 시점인 1996년 이후의 자료
- 다섯째, 부록 A의 결과에 따라 2000년 이후 사망률 개선이 둔화되고 안정화 추세를 보이는 자료 이용

2000년 이후의 자료를 이용하는 경우 시계열이 다소 짧지만, 1970–1999년 자료는 80세 이상의 보완, 95세 이상의 보완과 정제를 거친 자료이나 2000년 이후는 100세까지 자료가 확보되어 보완 및 정제를



**Table 3.1.** Fitting and forecast period of modeling

모형적합기간	모형검증기간(예측기간)
1970-2010(41년)	
1976-2010(35년)	
1983-2010(28년)	2011-2015(5년)
1996-2010(15년)	
2000-2010(11년)	

거치지 않은 자료이므로 2000년 이후의 자료를 이용하는 것도 의미가 있다 고려되어 분석에 포함한다.

### 3. 시계열 이용기간에 따른 사망률 예측 비교

#### 3.1. 예측력 평가 방법

시계열 이용기간에 따른 사망률 모형의 평가는 과거 데이터에 대한 적합도와 미래 사망률에 대한 예측력으로 구분될 수 있다. 즉, 적합도와 예측력 평가를 위해 본 연구는 LC 모형 외 4가지 모형과 5가지 시계열 이용기간을 고려한다.

예측력 비교 평가를 위해 사망률 자료를 모형적합기간과 모형검증기간으로 분할한 후 평가검증을 수행하였다. 본 연구에서는 모형 적합에 이용되는 시계열 시작 시점을 2장에서 서술하였듯이 5가지를 사용한다 (Table 3.1).

예측의 정확성은 실제 연령별 로그사망률 및 기대수명과 예측된 연령별 로그사망률  $\ln(\hat{m}_{x,t})$  및 기대수명  $(\hat{e}_0)$ 의 차이인 예측오차(forecasting error)에 기반한 식 (3.1)과 같은 평균오차(mean error; ME)와 평균절대오차(mean absolute error; MAE)를 활용하였다. 이때 예측 기대수명 산출은 각 모형을 통해 예측된 연령별 사망률에 기초하여 기대수명을 예측하는 방식을 사용하였다.

$$ME = \frac{\sum (y_t - \hat{y}_t)}{n}, \quad MAE = \frac{\sum |y_t - \hat{y}_t|}{n}. \quad (3.1)$$

본 연구에서는 평가검증을 수행하므로 자료이용기간에 따른 모형적합기간에 대한 적합력 비교는 식 (3.1)의 MAE를, 모형검증기간(2011-2015)에 대한 예측력 비교 및 예측값과 국가통계와의 비교(2016-2030)는 평균예측오차(mean forecasting error; MFE)와 평균절대예측오차(mean absolute forecasting error; MAFE)를 활용한다. 예측력 및 국가통계와의 비교를 위한 MFE와 MAFE는 식 (3.2)와 같이 정의한다.

$$MFE = \frac{\sum_{t=2011}^{2015} (y_t - \hat{y}_t)}{n}, \quad MAFE = \frac{\sum_{t=2011}^{2015} |y_t - \hat{y}_t|}{n}. \quad (3.2)$$

#### 3.2. 시계열 이용기간별 비교

본 연구는 시계열 이용기간별 비교를 위해 다음 3가지를 고려한다. 첫째, 모형별 시계열 이용기간에 따른 모형 적합 설명력과 적합력(MAE)을 살펴본다. 둘째, 시계열 이용기간에 따른 로그사망률과 기대수명의 예측력(MFE, MAFE)을 평가검증한다. 셋째, 미래 예측에 대한 불확실성을 감안하여 연령별 로그사망률과 기대수명에 대한 구간예측을 검토하고 불확실성을 평가한다. 미래 예측은 항상 불확실성이 내포되어 있으므로 점예측치와 함께 예측치의 불확실성 정도를 나타낼 수 있는 구간예측(신뢰구간)을 제시하는 것도 중요하다. 따라서 구간예측은 점예측과 관련하여 불확실성을 평가하는데 유용한 도구이다.

**Table 3.2.** The proportion of the variation that explained with the model-specific by fitting period

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.986	0.987	0.983	0.944	0.912	0.977	0.975	0.968	0.895	0.873
	LM	0.986	0.987	0.983	0.944	0.912	0.977	0.975	0.968	0.895	0.873
	BMS	0.984	0.984	0.983	0.935	0.764	0.974	0.974	0.966	0.854	0.854
FDM	1	0.987	0.987	0.984	0.945	0.912	0.977	0.975	0.968	0.895	0.872
	2	0.993	0.993	0.992	0.983	0.958	0.995	0.994	0.990	0.973	0.941
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	0.988	0.988	0.983	0.933	0.912	0.852	0.748	0.645	0.763	0.701
Coherent	(1, 2)	0.988	0.988	0.983	0.933	0.912	0.925	0.882	0.806	0.835	0.799
FDM	(2, 1)	0.996	0.995	0.993	0.986	0.968	0.852	0.748	0.645	0.763	0.701
	(2, 2)	0.996	0.995	0.993	0.986	0.968	0.925	0.882	0.806	0.835	0.799

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

**Table 3.3.** Mean absolute error of model by fitting period

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.00433	0.00351	0.00343	0.00397	0.00221	0.02070	0.01700	0.01182	0.00609	0.00302
	LM	0.00406	0.00330	0.00319	0.00371	0.00217	0.01594	0.01278	0.00908	0.00544	0.00298
	BMS	0.00318	0.00318	0.00318	0.00346	0.00142	0.01015	0.01015	0.00792	0.00258	0.00258
FDM	1	0.00342	0.00277	0.00253	0.00288	0.00203	0.00763	0.00627	0.00500	0.00445	0.00290
	2	0.00183	0.00159	0.00128	0.00093	0.00098	0.00169	0.00160	0.00163	0.00118	0.00134
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	0.00344	0.00278	0.00261	0.00303	0.00185	0.00076	0.00082	0.00079	0.00058	0.00053
Coherent	(1, 2)	0.00344	0.00278	0.00261	0.00303	0.00185	0.00039	0.00039	0.00043	0.00039	0.00037
FDM	(2, 1)	0.00121	0.00114	0.00108	0.00062	0.00067	0.00027	0.00026	0.00027	0.00026	0.00026
	(2, 2)	0.00121	0.00114	0.00108	0.00062	0.00067	0.00076	0.00082	0.00079	0.00058	0.00053

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

먼저, 각 모형별 자료이용기간에 따른 설명력과 적합력은 Table 3.2와 Table 3.3과 같다. 여기서 FDM은 모형의 주성분의 개수(order)를 1-2까지 사용하였으며, Coherent FDM은 곱(product) 모형을 설명하기 위한 주성분의 개수(order1)와 비(ratio) 모형을 설명하기 위한 주성분의 개수(order2)인 (order1, order2)를 (1, 1)-(2, 2)으로 설정하여 모형을 적합시켰다. LC, LM, BMS, 그리고 FDM은 사망률을 직접 모형화 하여 남자와 여자를 각각 독립적으로 적합시키는 반면 Coherent FDM은 사망률을 직접 모형화하는 대신 곱, 비 함수를 생성하여 적합하므로 Coherent FDM에서 product와 ratio는 곱 모형 및 비 모형을 적합시킨 설명력을 의미한다. FDM과 Coherent FDM은 order가 클수록 적합력이 좋아지나 모형이 복잡해지는 것을 감안하여 적절한 order를 선택하는 것이 바람직하며, 우리나라는 자료의 특성상 사망률이 일정 패턴이 반복적이기 때문에 FDM의 경우 order는 1 또는 2, Coherent FDM의 경우 (order1, order2)는 (1, 2) 또는 (2, 1)-(2, 2) 정도가 적당한 것으로 판단된다.

적합된 모형 대부분 설명력이 남자는 90%, 여자는 80%이상 수준이므로 적합 모형을 통한 연령별 사망률 추정치가 성별에 따른 사망률의 변동요인을 잘 설명하고 있다 (Table 3.2). 시계열 이용기간이 길어질수록 모형의 설명력이 높게 나타나며, 모형별 구분 없이 1970-2010년 자료를 활용한 경우 설명력이 높게 나타나고 시계열 이용기간이 단기간으로 변화됨에 따라 설명력은 감소하는 패턴을 보인다.

**Table 3.4.** MFE of age-specific log mortality (forecast period 2011–2015)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.045	0.044	0.054	0.051	0.011	0.290	0.254	0.183	0.037	-0.026
	LM	-0.019	-0.010	0.002	0.010	-0.006	0.006	0.010	0.012	-0.007	-0.022
	BMS	0.042	0.042	0.044	0.035	-0.036	0.143	0.143	0.107	-0.043	-0.043
FDM	1	-0.071	-0.046	-0.042	-0.067	-0.005	-0.030	-0.025	-0.051	-0.016	-0.021
	2	-0.069	-0.046	-0.042	-0.067	-0.002	-0.020	-0.018	-0.055	-0.016	-0.021
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	-0.040	-0.058	-0.006	-0.001	-0.013	0.015	-0.018	0.000	0.003	-0.009
Coherent	(1, 2)	-0.040	-0.058	-0.006	0.005	-0.010	0.015	-0.017	-0.001	-0.003	-0.012
FDM	(2, 1)	-0.038	-0.061	-0.013	-0.001	-0.006	0.018	-0.021	-0.006	0.003	-0.003
	(2, 2)	-0.038	-0.061	-0.012	-0.005	-0.003	0.018	-0.020	-0.007	-0.003	-0.006

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

**Table 3.5.** MAFE of age-specific log mortality (forecast period 2011–2015)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.101	0.087	0.089	0.086	0.053	0.355	0.311	0.229	0.093	0.086
	LM	0.051	0.047	0.044	0.043	0.050	0.080	0.081	0.081	0.078	0.096
	BMS	0.084	0.084	0.085	0.074	0.074	0.228	0.228	0.171	0.106	0.106
FDM	1	0.083	0.064	0.061	0.079	0.050	0.083	0.082	0.089	0.078	0.097
	2	0.081	0.064	0.061	0.079	0.049	0.086	0.082	0.095	0.078	0.097
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	0.066	0.075	0.045	0.044	0.063	0.088	0.087	0.079	0.076	0.082
Coherent	(1, 2)	0.066	0.075	0.045	0.046	0.066	0.087	0.087	0.079	0.077	0.080
FDM	(2, 1)	0.066	0.078	0.057	0.044	0.051	0.087	0.091	0.089	0.076	0.077
	(2, 2)	0.066	0.078	0.057	0.046	0.054	0.087	0.091	0.089	0.077	0.075

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

Table 3.3은 연령별 사망률에 대해 시계열 이용기간에 따른 모형 적합도로 MAE를 정리한 결과이다. 각 모형별로 시계열 이용기간이 짧을수록 적합력은 높은 것으로 나타났으며, 단기 예측의 경우 모형구분 없이 예측력도 우수할 것이라 기대를 할 수 있다. 위 도표에서 알 수 있듯이 설명력과 적합도는 서로 상반되는 결과를 보인다.

다음으로 로그사망률과 기대수명에 대한 예측력 평가검증이다. Table 3.4와 Table 3.5에 따르면 연령별 로그사망률이 시계열 이용기간에 따라 예측오차가 점진적 감소 경향을 나타낸다. 또한 예측모형들이 로그사망률을 과대 추정하는 경향이 있는 것을 알 수 있으며 MAFE는 시계열 이용기간이 상대적으로 짧은 경우에는 오차가 감소되는 경향을 나타낸다.

두 도표에 따르면 여자가 남자보다 상대적으로 로그사망률에 대한 오차가 큰 것을 알 수 있는데 이는 로그사망률 변동이 여자가 남자보다 크다는 것을 시사한다. 하지만 비모수모형에서는 평활기법을 적용하기 때문에 변동이 상쇄되어 MFE 측면에서는 여자가 남자보다 작은 경우도 보인다.

전반적으로 시계열 이용기간별로 남녀의 예측 로그사망률이 실제 로그사망률과 매우 유사하여 MAFE가 낮은 수준을 보이고 있다. 따라서 모형에 상관없이 로그사망률 예측력이 우수하다 할 수 있으나, 시계열 이용시점이 단기간일수록 오차가 작아지는 경향을 보인다. 또한 단기간 시계열 자료를 이용할수록

**Table 3.6.** MFE of life expectancy at birth (forecast period 2011–2015)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.227	0.156	0.075	-0.004	0.002	-0.026	-0.059	-0.079	-0.053	-0.099
	LM	0.354	0.264	0.168	0.063	0.053	0.233	0.183	0.103	-0.010	-0.122
	BMS	0.170	0.170	0.162	0.057	0.125	0.398	0.398	0.259	-0.131	-0.131
FDM	1	0.780	0.561	0.536	0.728	0.050	0.421	0.371	0.479	0.058	-0.131
	2	0.749	0.561	0.536	0.728	0.167	0.071	0.081	0.654	0.058	-0.131
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	0.396	0.513	0.159	-0.008	-0.073	0.285	0.455	0.211	0.079	-0.014
Coherent	(1, 2)	0.397	0.517	0.156	-0.058	-0.103	0.284	0.454	0.212	0.152	0.023
FDM	(2, 1)	0.286	0.634	0.420	-0.008	-0.032	0.205	0.555	0.452	0.079	0.047
	(2, 2)	0.288	0.638	0.417	-0.058	-0.062	0.205	0.554	0.452	0.152	0.084

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

**Table 3.7.** MAFE of life expectancy at birth (forecast period 2011–2015)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.246	0.197	0.146	0.088	0.076	0.164	0.151	0.121	0.083	0.099
	LM	0.354	0.264	0.173	0.097	0.088	0.233	0.183	0.121	0.055	0.122
	BMS	0.179	0.179	0.173	0.094	0.144	0.398	0.398	0.259	0.131	0.131
FDM	1	0.780	0.561	0.536	0.728	0.087	0.421	0.371	0.479	0.086	0.131
	2	0.749	0.561	0.536	0.728	0.167	0.094	0.104	0.654	0.086	0.131
		product	product	product	product	product	ratio	ratio	ratio	ratio	ratio
	(1, 1)	0.396	0.513	0.167	0.050	0.073	0.285	0.455	0.211	0.101	0.058
Coherent	(1, 2)	0.397	0.517	0.165	0.058	0.103	0.284	0.454	0.212	0.152	0.064
FDM	(2, 1)	0.308	0.634	0.441	0.050	0.039	0.221	0.555	0.453	0.101	0.069
	(2, 2)	0.304	0.638	0.438	0.058	0.062	0.221	0.554	0.454	0.152	0.092

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

모수모형보다는 비모수모형의 예측력이 우수하다.

기대수명의 예측력 평가검증은 로그사망률 결과와 비슷하다. Table 3.6과 Table 3.7에 따르면 연령별 기대수명은 시계열 이용기간에 따라 MFE와 MAFE 모두 점진적인 감소를 보인다. 로그사망률 결과와 비교해 본다면 기대수명의 예측오차가 상대적으로 높다. 이는 각기 다른 모형별로 사망률을 사망확률로 변경하는 방식차이에 기인한 결과이다.

끝으로 연령별 로그사망률 및 기대수명에 대한 구간예측을 검토하고 연령별 로그사망률의 불확실성을 평가한다. 연령별 로그사망률에 대한 구간예측은 예측분산을 이용하여 95% 예측구간을 계산하였고, 기대수명에 대한 구간예측은 Hyndman과 Booth (2008)에서 설명한 로그사망률 시뮬레이션을 이용하였다. 기대수명은 시뮬레이션 된 로그 사망률의 각 세트에 대해 계산되었으며 시뮬레이션 횟수는 200회, 신뢰구간은 95%이다. 연령별 로그사망률에 대한 구간예측의 불확실성(정확도) 평가 방법은 다음과 같다. 예측기간의 각 연도에 대하여 계산된 연령별 로그 사망률의 95% 구간예측안에 실제데이터가 포함되는 비율을 산출했다 (Swanson과 Beck, 1994; Tayman 등, 2007). 이는 계산된 95% 예측구간에 실제 데이터가 포함되는 비율로, 분모는 예측기간의 전체 데이터 수이다(예: 0-100살 5년 =  $101 \times 5 = 505$  관측).

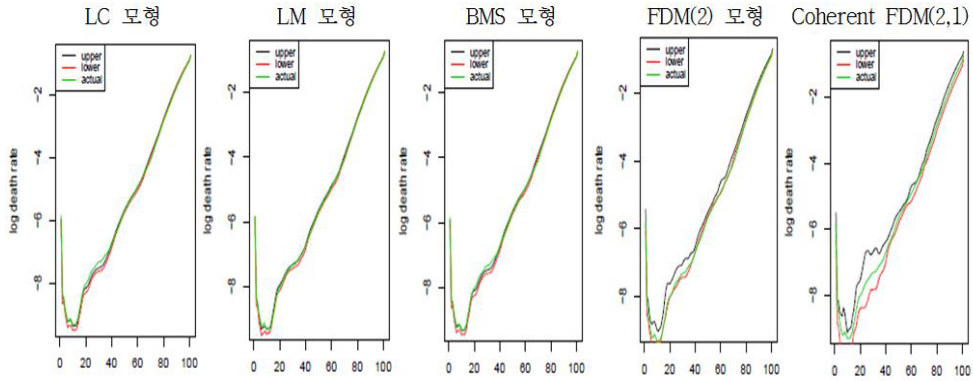


Figure 3.1. Interval forecast of male age-specific log mortality (fitting period 1996–2010, forecast period 2015).

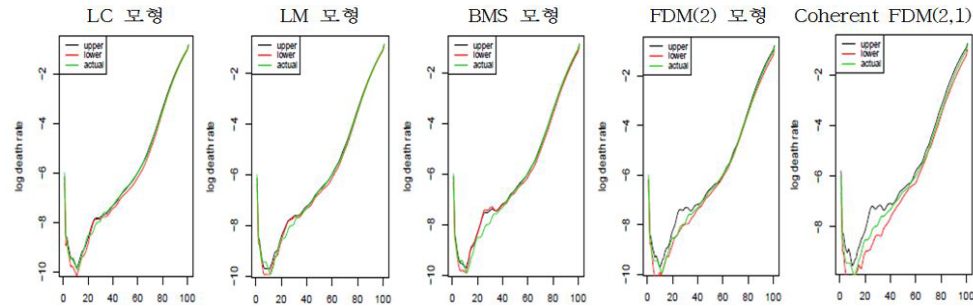


Figure 3.2. Interval forecast of female age-specific log mortality (fitting period 1996–2010, forecast period 2015).

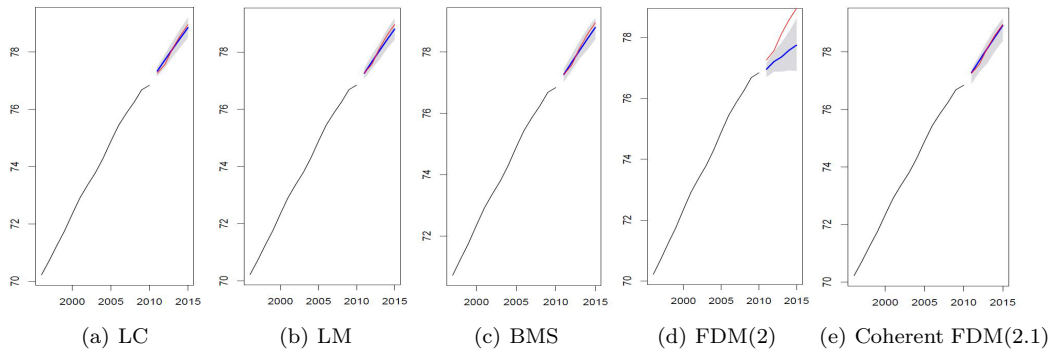
Table 3.8. Interval forecast accuracy of age-specific log mortality (forecast period 2011–2015)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
LC		0.24	0.28	0.29	0.22	0.44	0.06	0.06	0.10	0.34	0.52
LM		0.48	0.48	0.41	0.49	0.58	0.43	0.39	0.34	0.51	0.43
BMS		0.30	0.30	0.30	0.30	0.47	0.05	0.05	0.47	0.47	0.47
FDM	1	0.96	0.98	0.95	0.91	0.89	0.97	0.97	0.94	0.90	0.71
	2	0.87	0.92	0.76	0.62	0.81	0.81	0.81	0.71	0.88	0.71
Coherent FDM	(1, 1)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	(1, 2)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
FDM	(2, 1)	1.00	1.00	0.99	1.00	1.00	0.99	0.96	0.97	0.99	0.99
	(2, 2)	0.99	0.98	0.99	1.00	1.00	0.98	0.94	0.98	0.99	0.99

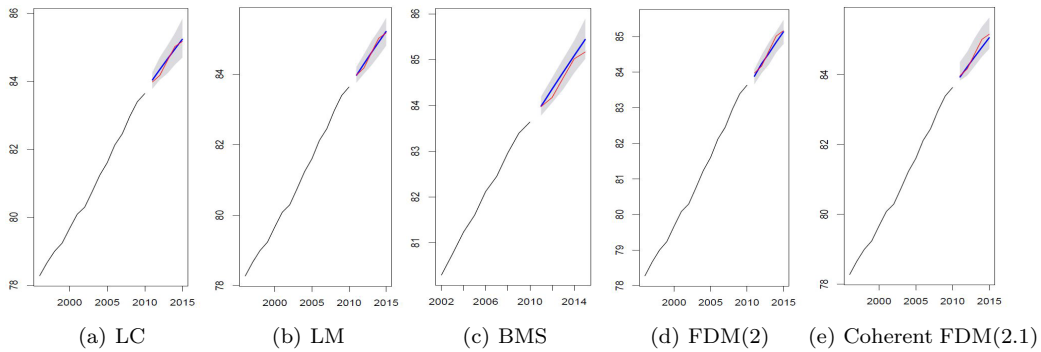
LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

구간예측에 대해서는 본 연구에서 제안하는 회귀모형으로 시계열 이용기간을 선정한 결과인 1996년 이후의 자료를 이용하여 적합시킨 모형의 로그사망률 및 기대수명에 대한 구간예측을 성별로 살펴보았다.

Figure 3.1과 Figure 3.2는 1996–2010년 자료를 이용하여 5가지 사망률 예측모형에 의해 예측한 남자와 여자의 2015년 연령별 로그사망률의 구간예측 결과(95% 상한(upper), 하한(lower))과 실제값으로, Coherent FDM을 제외하고 남녀 모두 예측구간 간격이 모든 연령대에서 매우 좁게 나타났다.



**Figure 3.3.** Interval forecast of male life expectancy at birth (fitting period 1996–2010). LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.



**Figure 3.4.** Interval forecast of female life expectancy at birth (fitting period 1996–2010). LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

Table 3.8은 로그사망률 구간예측(2011–2015년)의 정확도를 시계열 이용기간별로 정리한 것으로, 예측 기간 각 연도별 구간예측의 정확도를 평균한 것이다. 남, 여 모두 최근의 시계열 자료를 이용할수록 구간예측에 실제값이 포함되는 비율이 높음을 알 수 있다. 특히 모수모형은 점진적 상승을 하는 반면 비모수 모형은 거의 일정하거나 감소하는 경향을 보인다. FDM은 시계열 이용기간과 정확도가 관련이 없어 보이는 반면 Coherent FDM은 거의 일정한 값을 유지하는 것을 알 수 있다. 즉, Coherent FDM은 그룹간의 내적 일관성이 충족되고 그룹별 FDM을 적용할 시 발산하는 경향을 억제하는 공통성(coherent)을 반영한 모형이므로, 즉 Coherent FDM은 FDM보다 변동에 대해서 강건한(robust)모형이기 때문에 이러한 결과가 도출됨을 알 수 있다.

남자의 경우 각 모형별로 시계열 이용기간에 따른 구간예측의 정확성을 살펴보면 FDM(1)을 제외한 다른 모형은 전반적으로 최근의 시계열을 이용할수록 구간예측에 실제값이 포함되는 비율이 높은 것으로 나타났다. 여자의 경우 LC와 BMS가 최근의 시계열을 사용할수록 구간예측의 정확성이 월등히 향상됨을 알 수 있으나 FDM의 경우만 시계열 이용기간이 길수록 구간예측에 실제값이 포함되는 비율이 높은 것으로 나타나 상반된 결과를 얻었다. 이런 결과가 기대수명에는 어떠한 영향을 미치는지 기대수명의 구간예측을 살펴보았다.

Figure 3.3과 Figure 3.4는 1996–2010년 자료를 이용하여 5가지 사망률 예측모형에 의해 예측한 기대수명에 대한 구간예측 결과(95% 상한·하한: 음영, 예측값: 파란색)와 실제값(빨간색)이다.

Figure 3.3에서 남자는 FDM을 제외하고 모든 모형들이 구간예측의 음영부분(상한과 하한)에 실제값이 포함되고, 모형에 따른 구간예측의 간격도 비슷함을 알 수 있다. FDM은 최근의 자료(2000-2010)를 이용하여 적합한 경우만이 실제값이 구간예측 간격 내에 포함되고 그 외의 자료이용기간을 적합한 경우는 실제값이 구간예측 간격 내에 포함되지 않았다. Coherent FDM의 경우 FDM과 다르게 모수모형들과 유사한 구간예측의 결과를 제시한다. Figure 3.4에서 여자는 남자와 달리 5가지 모형의 구간예측 내에 실제값이 포함되어 있다. 특히 FDM의 경우 다른 모형에 의해 생성된 구간예측 간격보다 예측 간격이 좁음을 확인할 수 있었다.

따라서 로그사망률과 기대수명의 구간예측에서 모수모형과 Coherent FDM은 남녀 모두 규칙성과 일관성 있는 결과를 제공하지만 FDM은 그렇지 못했다. 이러한 결과는 우리나라 사망률 자료의 적합과 예측에 있어서 남자, 여자 즉 전체 인구집단의 사망경향 동일성(coherence)을 고려하는 것이 중요한 것임을 시사한다. 여기서 동일성이란 인구사회학적으로 밀접히 연관된 하위 인구집단(예: 남자, 여자)의 사망률 격차가 장기적으로 확대될 가능성은 높지 않으므로 각각의 인구하위집단의 사망경향이 서로 유사함을 의미한다.

지금까지 모형 적합의 설명력과 예측오차, 로그사망률과 기대수명의 평가검증, 구간예측 등으로 시계열 기간에 따른 문제점과 특징을 살펴보았으며, 3가지 사항에 대해서 정리함으로써 다음과 같은 연구결과를 도출할 수 있다.

첫째, 시계열 이용기간을 길게 고려하면 모형 적합 설명력은 좋으나 적합도(MAE)와 예측오차 측면에서는 상반된 결과를 제시한다. 이는 연령별 로그사망률과 기대수명 평가검증에서도 동일하게 도출되었다.

둘째, 모형별 적합도가 좋으면 구간예측의 간격이 좁으며, 적합도가 떨어질수록 넓은 구간예측을 보인다. 예측오차측면에서는 시계열 이용기간이 길수록 오차는 크고 짧을수록 작은 경향을 보인다. 이는 장래 예측이 바로 앞 시점의 정보에 가장 많이 의존한다는 것을 의미하기도 한다.

셋째, 국가통계는 적합도 보다는 설명력에 중점을 두는 경향이 있다. 그 이유는 자료의 완비성 측면에서 과거 자료의 보완과 검증 등으로 장시간의 시계열을 확보하는 것이 국가통계 생산 목적이기 때문이다. 예를 들어 HMD제공 국가 중 266년의 가장 긴 사망률 자료를 제공하는 스웨덴과 최근 통계청의 '1970-2015년 생명표 100세 확장'이 좋은 사례이다.

따라서 본 연구는 연구자가 모형 적합 설명력에 초점이 맞춰져 있다면 장기간의 자료(1970-2010)를 사용하고 적합도, 구간예측과 오차관점에서 장래의 정밀한 예측값을 제시하고자 한다면 단기간 (1996-2010)의 자료를 활용하는 것을 제안한다.

### 3.3. 국가통계와의 비교

앞 절에서는 3가지 기준에 대해서 사망률 시계열 이용기간 선정에 따른 문제점을 살펴보았으며, 5가지 상이한 시계열 이용기간과 모형에 따른 2011-2015년의 예측력을 실측값과 비교 분석하였다.

일반적으로 장래 예측을 위해 예측기간보다 같거나 많은 실측값을 도태로 분석하는 것이 의미있다. 따라서 본 연구에서는 자료이용기간의 최단기간이 15년(2000-2015)이므로 향후 15년을 예측하기로 한다.

본 절에서는 각 모형별, 시계열 이용시점별로 향후 15년(2016-2030년)의 연령에 따른 남녀 사망률 예측을 위해 모형을 적합하고, 예측된 사망률을 바탕으로 예측기대수명을 작성하여 통계청에서 제공하는 장래 연령별 사망률과 기대수명을 비교하였다.

Table 3.9와 Table 3.10은 시계열 이용기간과 예측 모형별로 향후 15년(2016-2030)의 연령별 사망률

**Table 3.9.** MFE of age-specific log mortality (forecast period 2016–2030)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.113	0.123	0.145	0.122	0.066	0.380	0.339	0.255	0.071	0.024
	LM	0.035	0.057	0.084	0.092	0.065	0.111	0.118	0.120	0.073	0.058
	BMS	0.124	0.124	0.124	0.114	0.061	0.106	0.106	0.106	0.106	0.106
FDM	1	0.036	0.041	0.063	0.092	0.064	0.074	0.081	0.079	0.081	0.064
	2	0.046	0.041	0.063	0.092	0.061	0.095	0.098	0.079	0.081	0.064
Coherent	(1, 1)	0.048	0.066	0.070	0.079	0.056	0.057	0.066	0.096	0.104	0.081
	(1, 2)	0.057	0.075	0.076	0.084	0.062	0.048	0.057	0.090	0.098	0.075
FDM	(2, 1)	0.065	0.069	0.072	0.079	0.056	0.074	0.070	0.097	0.104	0.081
	(2, 2)	0.074	0.078	0.078	0.084	0.062	0.065	0.060	0.092	0.098	0.075

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

**Table 3.10.** MAFE of age-specific log mortality (forecast period 2016–2030)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
	LC	0.133	0.135	0.157	0.131	0.104	0.404	0.360	0.271	0.135	0.150
	LM	0.054	0.067	0.092	0.101	0.097	0.120	0.129	0.135	0.131	0.146
	BMS	0.134	0.134	0.134	0.124	0.101	0.145	0.145	0.145	0.145	0.145
FDM	1	0.054	0.057	0.074	0.101	0.096	0.091	0.098	0.104	0.136	0.150
	2	0.068	0.057	0.074	0.101	0.103	0.148	0.149	0.104	0.136	0.150
Coherent	(1, 1)	0.062	0.077	0.081	0.100	0.111	0.075	0.086	0.108	0.126	0.126
	(1, 2)	0.069	0.087	0.086	0.101	0.109	0.066	0.074	0.100	0.126	0.128
FDM	(2, 1)	0.095	0.077	0.082	0.100	0.111	0.117	0.093	0.109	0.126	0.126
	(2, 2)	0.106	0.087	0.088	0.101	0.109	0.105	0.081	0.102	0.126	0.128

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

예측값을 통계청과 비교 정리한 결과이다.

Table 3.9에서 시계열 이용기간이 단기간으로 변환됨에 따라 각 모수모형의 MFE는 점진적 감소하고 있으며 남성보다는 여성이 변화폭이 크다. 반면에 비모수모형은 MFE는 미비하게 증가하거나 일정한 값을 보이는 경향을 띄고 성별에 따른 변화폭의 차이는 크지 않게 나타났다. 남녀 모두 시계열 이용기간과 예측모형에 관계없이 일괄되게 연령별 로그사망률을 과소 추정한다.

Table 3.10에서 모수모형은 LM을 제외하고는 단기간으로 변환됨에 따라 MAFE가 점진적 감소 경향을 보인다. 이에 반면 비모수모형은 Table 3.9와 비슷한 경향을 띤다.

위 결과는 앞 절에서 도출된 연구결과와 구분되어야 한다. 즉 예측기간이 길어짐에 따라 모수모형과 비모수모형의 차이를 보여준다. 모수모형보다는 비모수모형이 자료이용기간에 예측력의 변동이 크지 않다는 것을 알 수 있고 이는 비모수의 평활기법에 기인한 결과이다.

남자의 경우 시계열 이용기간이 1970–2015인 경우를 제외하고는 FDM과 Coherent FDM이 통계청에서 제시한 장래 연령별 사망률과 가장 유사하며, 여자의 경우 시계열 이용기간에 관계없이 FDM과 Coherent FDM이 통계청에서 제시한 장래 연령별 사망률과 가장 유사함을 알 수 있다.

Table 3.11과 Table 3.12는 앞에서 예측한 향후 15년의 연령별 사망률을 바탕으로 작성된 기대수명을 KOSIS와 비교 정리한 것이다. Table 3.11의 MFE에 의하면 남녀 모두 시계열 이용기간과 예측모형에



**Table 3.11.** MFE of life expectancy at birth (forecast period 2016–2030)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
LC		-0.169	-0.335	-0.505	-0.694	-0.742	-.385	-0.462	-0.565	-0.696	-0.871
LM		-0.096	-0.280	-0.469	-0.698	-0.759	-.196	-0.296	-0.460	-0.711	-0.909
BMS		-0.614	-0.614	-0.614	-0.636	-0.699	-.638	-0.638	-0.638	-0.638	-0.638
FDM	1	-0.100	-0.173	-0.330	-0.698	-0.755	-0.040	-0.132	-0.264	-0.759	-0.957
	2	-0.330	-0.173	-0.330	-0.698	-0.785	-0.702	-0.778	-0.270	-0.759	-0.957
Coherent	(1, 1)	-0.107	-0.240	-0.502	-0.928	-1.034	-0.047	-0.178	-0.283	-0.619	-0.751
	(1, 2)	-0.213	-0.325	-0.529	-1.000	-1.100	-0.001	-0.152	-0.286	-0.545	-0.690
FDM	(2, 1)	-0.787	-0.386	-0.577	-0.928	-1.034	-0.521	-0.290	-0.348	-0.619	-0.751
	(2, 2)	-0.885	-0.470	-0.604	-1.000	-1.100	-0.481	-0.265	-0.351	-0.545	-0.690

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

**Table 3.12.** MAFE of life expectancy at birth (forecast period 2016–2030)

		Male					Female				
		'70-'10	'76-'10	'83-'10	'96-'10	'00-'10	'70-'10	'76-'10	'83-'10	'96-'10	'00-'10
LC		0.169	0.335	0.505	0.694	0.742	0.385	0.462	0.565	0.696	0.871
LM		0.098	0.280	0.469	0.698	0.759	0.199	0.298	0.460	0.711	0.909
BMS		0.614	0.614	0.614	0.636	0.699	0.638	0.638	0.638	0.638	0.638
FDM	1	0.102	0.173	0.330	0.698	0.755	0.063	0.146	0.271	0.759	0.957
	2	0.339	0.173	0.330	0.698	0.785	0.702	0.778	0.271	0.759	0.957
Coherent	(1, 1)	0.107	0.240	0.502	0.928	1.034	0.055	0.181	0.285	0.619	0.751
	(1, 2)	0.213	0.325	0.529	1.000	1.100	0.044	0.164	0.288	0.545	0.690
FDM	(2, 1)	0.787	0.386	0.577	0.928	1.034	0.521	0.290	0.348	0.619	0.751
	(2, 2)	0.885	0.470	0.604	1.000	1.100	0.481	0.265	0.351	0.545	0.690

LC = Lee-Carter; LM = Lee-Miller; BMS = Booth-Maindonald-Smith; FDM = functional data model.

관계없이 일괄되게 기대수명을 과대 추정한다. 이는 연령별 로그 사망률의 과소추정에 기인한 것이다. 일반적으로 로그사망률이 작을수록 기대수명은 크게 산출된다. Table 3.12의 MAFE에 따르면 연령별 사망률 예측과 반대로 시계열 이용기간이 길어질수록 FDM과 Coherent FDM이 통계청에서 제시한 장래 기대수명과 가장 유사함을 알 수 있다.

지금까지 5가지 모형과 시계열 이용기간별로 예측한 연령별 사망률과 예측기대수명을 국가통계와 비교해 보았으며 연구결과를 토대로 실측치와 예측치의 비교에 대해 다음과 같이 제안해 본다.

먼저 실측치의 경우 시계열 이용기간을 길게 고려하면 모형 적합의 설명력은 좋으나 적합도(MAE)와 예측오차는 떨어지고 모형별 적합도가 좋으면 구간예측의 간격이 좁아진다. 또한 예측오차측면에서는 시계열 이용기간이 길수록 오차는 크고 짧을수록 작은 경향을 보인다. 이러한 결과는 모수 및 비모수모형 구분 없이 일관성을 보인다.

이에 반해 예측치는 모수, 비모수모형과 비교 항목에 따라 실측치 비교와는 다른 결과가 도출된다.

첫째, 로그사망률 예측치를 비교할 경우 시계열 이용기간이 단기간으로 변환됨에 따라 각 모수모형의 MFE가 점진적 감소를 보이고 예측력이 좋아진다. 비모수모형은 MFE 값이 약간 증가나 일정한 값을 보이는 경향을 띄고 성별에 따른 변화폭의 차이는 크지 않게 나타난다.

둘째, 기대수명 예측치를 비교할 경우 시계열 이용기간이 장기간으로 변화됨에 따라 모형 구분 없이

MFE, MAFE가 작아지는 경향을 보인다. 이는 로그사망률의 과소추정에 따른 결과이기도 하다.

셋째, MFE와 MAFE를 기준으로 실측치를 비교할 경우는 모형보다는 시계열 이용기간의 선정이 중요하고 예측치를 비교할 때는 모형과 시계열 이용기간 2가지를 동시에 고려해야 한다는 것을 알 수 있다.

따라서 본 연구는 미래는 불확실성을 내포하고 있으므로 장래 예측치 비교는 실측치의 결과와 구분된 해석과 주의가 있음을 제안한다.

#### 4. 결론 및 제언

HMD에 제공된 사망 자료들의 시계열은 국가별로 상이하다. 스웨덴(266년)처럼 긴 기간을 제공하는 나라가 있는 반면 크로아티아(13년)처럼 단기간을 제공하는 나라도 있다. 우리나라의 경우 이런 나라들에 비해 상대적으로 사망 자료의 시계열이 길지 않은 편이다.

그런데 우리나라는 46년간(1970-2015)의 길지 않은 기간임에도 불구하고 공중위생, 보건 및 의료개선 등으로 고령화가 빠르게 진행되어 사망률 개선이 급속히 이루어진 점을 감안할 때, 사망률 예측모형뿐만 아니라 시계열 이용기간의 적정성 검토도 중요한 고려대상이다.

본 연구는 시계열 이용기간에 대한 평가가 다양한 모형과 비교대상(실측치, 예측치)의 한쪽에 편중되지 않게 모수 및 비모수모형 그리고 실측치와 예측치를 모두 고려하여 분석을 수행하고 정리하였다.

먼저 실측치와 비교한 연구결과를 요약하면 다음과 같다.

첫째, 장기간의 시계열을 사용할 경우 모형 적합 설명력이 우수하게 나타나는 반면 예측오차는 떨어지는 경향을 보인다. 따라서 구간예측과 오차관점에서 정밀한 장래 예측값을 제시하고자 할 경우 장기간보다는 단기간의 시계열을 활용하는 것이 바람직하다. 따라서 연구자의 연구방향과 목적에 따라 적절한 시계열 이용기간의 선정이 중요하다.

둘째, 동일한 사망률 예측 모형에도 불구하고 시계열 이용기간에 따라 예측력이 다르게 나타나는 것은 예측의 정확성 측면에서 모형 외에도 시계열 이용기간이 중요한 요인 중의 하나임을 보여준다. 단기간의 기대수명 및 사망률을 예측할 경우 단기 시계열에 기초한(2000-2010) FDM과 Coherent FDM의 예측성과가 상대적으로 우수함을 알 수 있다. 이는 단기 예측을 수행할 때 예측 정확성 평가기간(2011-2015)의 사망률 변화가 모형 구축에 사용된 적합기간(fitting period)의 사망률 변화 패턴과 유사하기 때문으로 볼 수 있다.

셋째, 연령별 사망률 및 기대수명 예측의 정확성에 어떠한 모형이 뚜렷한 비교 우위를 보이는지 살펴본 결과 시계열 이용기간의 조건과 관계없이 일관되게 비교 우위를 보이는 모형은 도출되지 않았다.

다음으로 예측치와 비교한 연구결과는 아래와 같다.

첫째, MFE와 MAFE를 기준으로 실측치를 비교할 경우는 모형보다는 시계열 이용기간의 선정이 중요하지만 예측치를 비교할 때는 2가지를 동시에 고려해야 한다.

둘째, 미래는 불확실성을 내포하고 있으므로 장래 예측치 비교는 실측치의 평가검증 결과와 구분된 해석이 요구된다.

끝으로 본 연구와 관련하여 향후 연구방향을 제안한다.

미래의 기대수명이 빠른 속도로 개선된 과거 추세를 따를 것인지, 둔화될 것인지, 아니면 현 수준을 유지할 것인지는 불확실하다. 따라서 모형에 의한 미래 예측뿐만 아니라 사망률이 향후 어떻게 전개될 것인가에 대해서는 인구전문가의 견해와 데이터의 사전정보를 반영하는 베이지안방법도 추가적으로 논의될 필요가 있다.

부록 A: 성 및 연령별(1세-100+세) 사망률 변화(1970-2015)

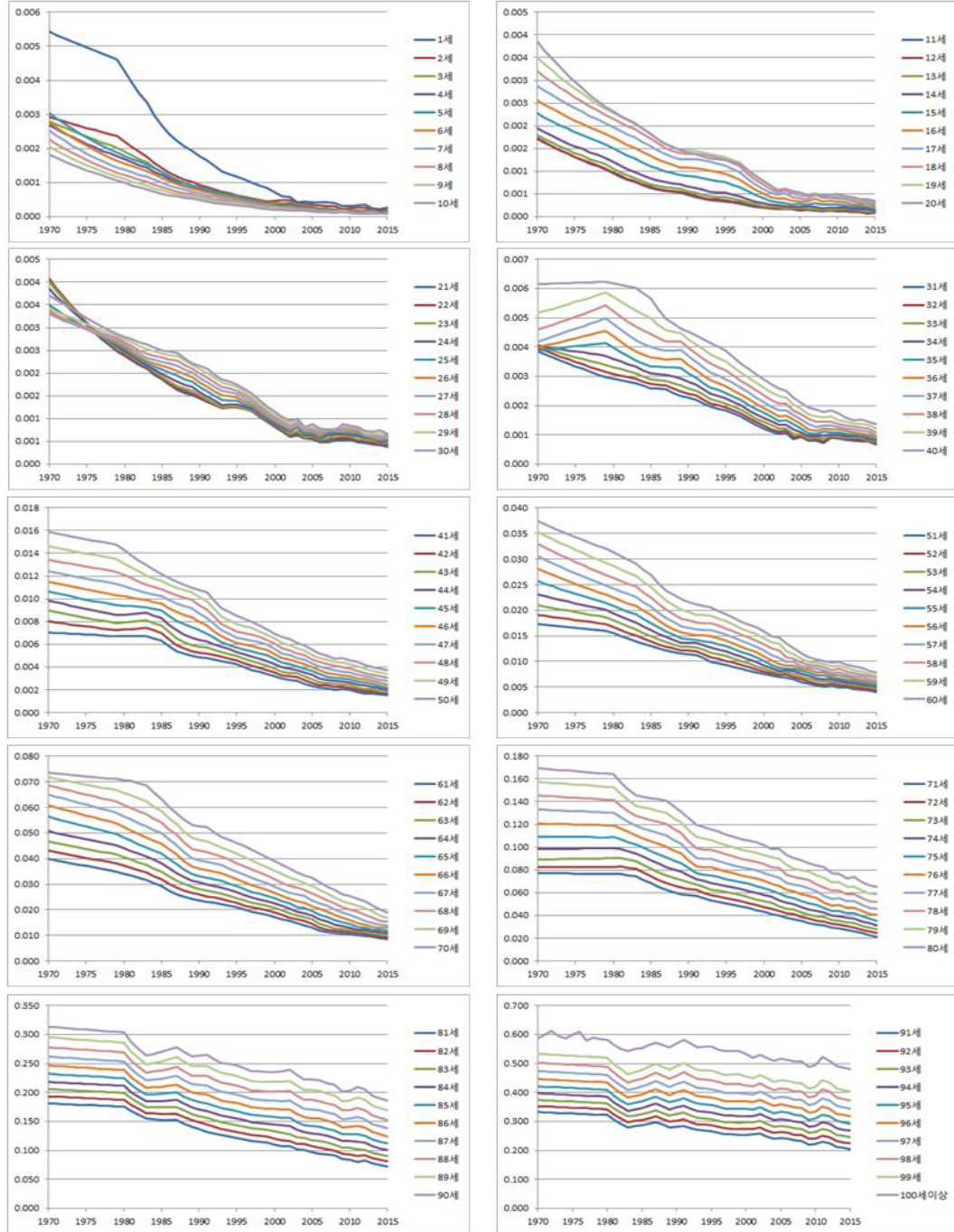


Figure A.1. Male.

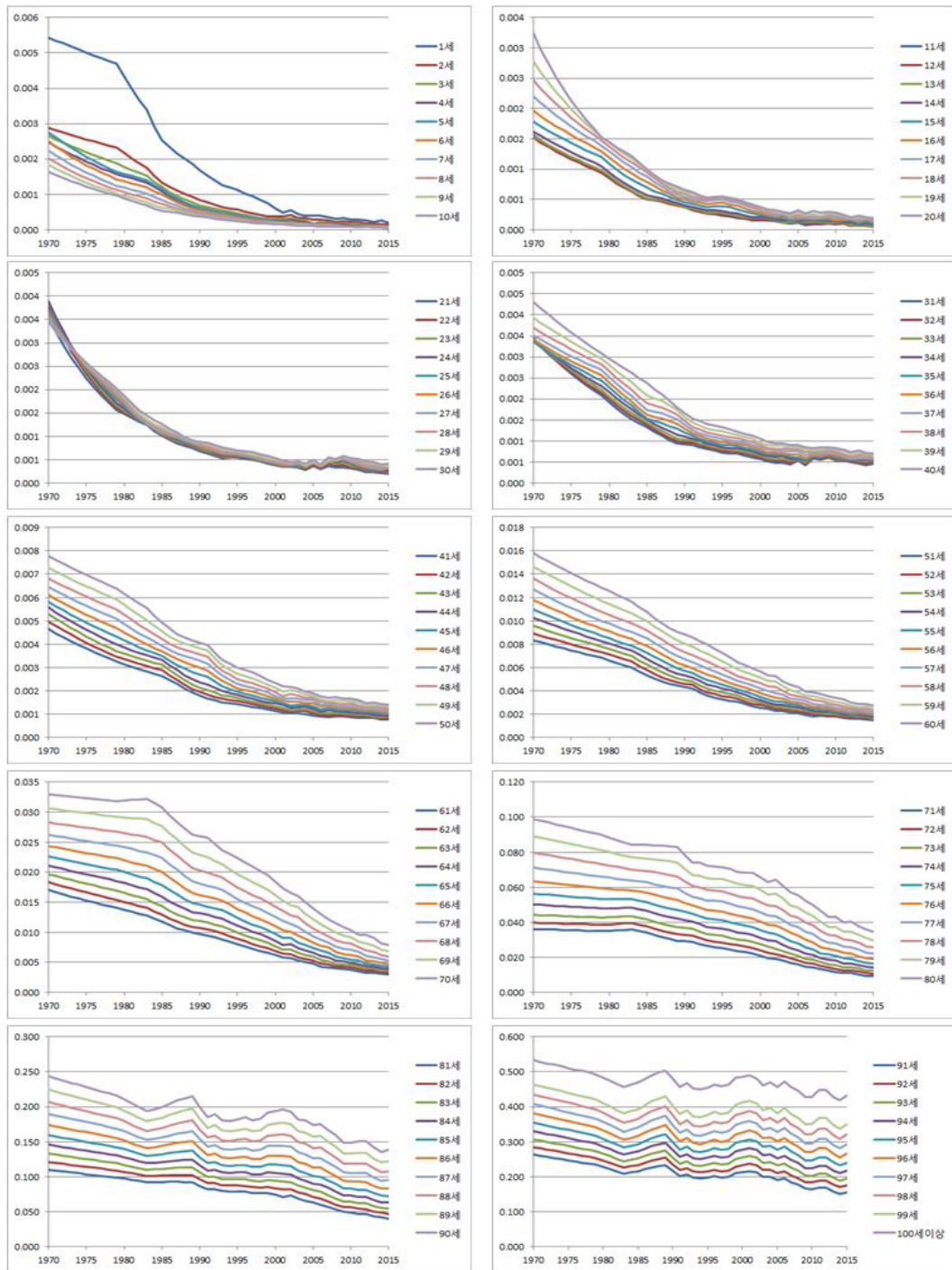


Figure A.2. Female.

## 부록 B: 36개의 회귀계수 결과

		Male				Female			
		Estimate	StdError	t-value	Pr(>  t )	Estimate	StdError	t-value	Pr(>  t )
1970	$\beta_0$	2.14E-02	3.45E-04	61.98	7.90E-80	2.02E-02	3.27E-04	61.71	1.19E-79
	$\beta_1$	-5.36E-04	3.01E-05	-17.84	2.15E-32	-3.93E-04	2.85E-05	-13.80	1.27E-24
2015	$\beta_2$	8.70E-06	7.00E-07	12.42	9.15E-22	5.59E-06	6.64E-07	8.42	3.46E-13
	$\beta_3$	-5.48E-08	4.60E-09	-11.90	1.13E-20	-3.70E-08	4.36E-09	-8.49	2.41E-13
1971	$\beta_0$	2.13E-02	3.36E-04	63.26	1.15E-80	2.03E-02	3.34E-04	60.98	3.70E-79
	$\beta_1$	-5.31E-04	2.93E-05	-18.14	6.13E-33	-4.17E-04	2.90E-05	-14.37	9.30E-26
2015	$\beta_2$	8.67E-06	6.82E-07	12.70	2.35E-22	6.22E-06	6.77E-07	9.20	7.31E-15
	$\beta_3$	-5.49E-08	4.48E-09	-12.24	2.19E-21	-4.13E-08	4.45E-09	-9.29	4.64E-15
1972	$\beta_0$	2.12E-02	3.29E-04	64.31	2.39E-81	2.05E-02	3.42E-04	59.95	1.84E-78
	$\beta_1$	-5.26E-04	2.86E-05	-18.37	2.43E-33	-4.42E-04	2.97E-05	-14.85	1.02E-26
2015	$\beta_2$	8.63E-06	6.67E-07	12.93	7.77E-23	6.87E-06	6.93E-07	9.92	2.02E-16
	$\beta_3$	-5.50E-08	4.39E-09	-12.53	5.36E-22	-4.57E-08	4.55E-09	-10.03	1.18E-16
1973	$\beta_0$	2.10E-02	3.23E-04	65.12	7.29E-82	2.06E-02	3.52E-04	58.69	1.37E-77
	$\beta_1$	-5.20E-04	2.81E-05	-18.51	1.36E-33	-4.67E-04	3.06E-05	-15.25	1.67E-27
2015	$\beta_2$	8.59E-06	6.55E-07	13.11	3.42E-23	7.54E-06	7.13E-07	10.57	7.72E-18
	$\beta_3$	-5.50E-08	4.31E-09	-12.77	1.70E-22	-5.01E-08	4.69E-09	-10.70	4.12E-18
1974	$\beta_0$	2.09E-02	3.19E-04	65.58	3.73E-82	2.08E-02	3.63E-04	57.22	1.49E-76
	$\beta_1$	-5.14E-04	2.77E-05	-18.53	1.25E-33	-4.92E-04	3.16E-05	-15.57	3.91E-28
2015	$\beta_2$	8.53E-06	6.47E-07	13.19	2.33E-23	8.22E-06	7.36E-07	11.17	4.08E-19
	$\beta_3$	-5.49E-08	4.25E-09	-12.92	8.17E-23	-5.47E-08	4.84E-09	-11.31	2.00E-19
1975	$\beta_0$	2.08E-02	3.16E-04	65.69	3.22E-82	2.09E-02	3.76E-04	55.60	2.25E-75
	$\beta_1$	-5.07E-04	2.75E-05	-18.42	1.94E-33	-5.18E-04	3.27E-05	-15.83	1.27E-28
2015	$\beta_2$	8.45E-06	6.41E-07	13.17	2.53E-23	8.93E-06	7.63E-07	11.70	2.97E-20
	$\beta_3$	-5.47E-08	4.21E-09	-12.98	6.14E-23	-5.95E-08	5.01E-09	-11.87	1.35E-20
1976	$\beta_0$	2.06E-02	3.15E-04	65.46	4.49E-82	2.11E-02	3.91E-04	53.85	4.53E-74
	$\beta_1$	-4.99E-04	2.74E-05	-18.20	4.83E-33	-5.45E-04	3.40E-05	-16.01	5.60E-29
2015	$\beta_2$	8.35E-06	6.39E-07	13.07	4.16E-23	9.65E-06	7.93E-07	12.17	3.01E-21
	$\beta_3$	-5.45E-08	4.20E-09	-12.96	6.86E-23	-6.44E-08	5.21E-09	-12.35	1.27E-21
1977	$\beta_0$	2.05E-02	3.16E-04	64.89	1.03E-81	2.12E-02	4.07E-04	52.00	1.20E-72
	$\beta_1$	-4.91E-04	2.75E-05	-17.86	1.99E-32	-5.72E-04	3.55E-05	-16.13	3.25E-29
2015	$\beta_2$	8.24E-06	6.41E-07	12.87	1.07E-22	1.04E-05	8.26E-07	12.59	4.11E-22
	$\beta_3$	-5.41E-08	4.21E-09	-12.85	1.17E-22	-6.94E-08	5.43E-09	-12.78	1.61E-22
1978	$\beta_0$	2.03E-02	3.18E-04	63.88	4.50E-81	2.13E-02	4.26E-04	50.07	4.12E-71
	$\beta_1$	-4.81E-04	2.77E-05	-17.36	1.62E-31	-6.00E-04	3.71E-05	-16.20	2.48E-29
2015	$\beta_2$	8.10E-06	6.46E-07	12.54	5.23E-22	1.12E-05	8.64E-07	12.94	7.61E-23
	$\beta_3$	-5.35E-08	4.24E-09	-12.60	3.78E-22	-7.46E-08	5.68E-09	-13.15	2.78E-23
1979	$\beta_0$	2.02E-02	3.23E-04	62.54	3.39E-80	2.14E-02	4.46E-04	48.08	1.82E-69
	$\beta_1$	-4.70E-04	2.81E-05	-16.74	2.34E-30	-6.29E-04	3.88E-05	-16.20	2.43E-29
2015	$\beta_2$	7.92E-06	6.55E-07	-12.10	4.26E-21	1.20E-05	9.05E-07	13.23	1.87E-23
	$\beta_3$	-5.27E-08	4.30E-09	-12.26	2.01E-21	-8.00E-08	5.95E-09	-13.46	6.40E-24
1980	$\beta_0$	2.00E-02	3.29E-04	60.78	4.99E-79	2.16E-02	4.69E-04	45.98	1.16E-67
	$\beta_1$	-4.58E-04	2.87E-05	-15.97	6.82E-29	-6.58E-04	4.08E-05	-16.13	3.31E-29
2015	$\beta_2$	7.71E-06	6.68E-07	11.54	6.55E-20	1.28E-05	9.51E-07	13.46	6.53E-24
	$\beta_3$	-5.17E-08	4.39E-09	-11.79	1.99E-20	-8.56E-08	6.25E-09	-13.70	2.09E-24

		Male				Female			
		Estimate	StdError	$t$ -value	$\Pr(>  t )$	Estimate	StdError	$t$ -value	$\Pr(>  t )$
1981	$\beta_0$	1.98E-02	3.37E-04	58.86	1.05E-77	2.17E-02	4.91E-04	44.06	5.95E-66
	$\beta_1$	-4.44E-04	2.93E-05	-15.12	3.00E-27	-6.87E-04	4.28E-05	-16.05	4.62E-29
2015	$\beta_2$	7.46E-06	6.84E-07	10.91	1.47E-18	1.36E-05	9.97E-07	13.65	2.67E-24
	$\beta_3$	-5.05E-08	4.49E-09	-11.24	2.87E-19	-9.10E-08	6.55E-09	-13.90	8.21E-25
1982	$\beta_0$	1.97E-02	3.44E-04	57.15	1.69E-76	2.17E-02	5.13E-04	42.34	2.38E-64
	$\beta_1$	-4.28E-04	2.99E-05	-14.31	1.24E-25	-7.13E-04	4.46E-05	-15.97	6.53E-29
2015	$\beta_2$	7.17E-06	6.97E-07	10.27	3.44E-17	1.44E-05	1.04E-06	13.81	1.25E-24
	$\beta_3$	-4.89E-08	4.58E-09	-10.67	4.85E-18	-9.62E-08	6.84E-09	-14.06	3.78E-25
1983	$\beta_0$	1.94E-02	3.49E-04	55.61	2.18E-75	2.17E-02	5.33E-04	40.79	7.22E-63
	$\beta_1$	-4.11E-04	3.04E-05	-13.52	4.95E-24	-7.37E-04	4.64E-05	-15.89	9.31E-29
2015	$\beta_2$	6.83E-06	7.09E-07	9.63	8.44E-16	1.51E-05	1.08E-06	13.94	6.65E-25
	$\beta_3$	-4.69E-08	4.66E-09	-10.07	9.33E-17	-1.01E-07	7.10E-09	-14.20	2.03E-25
1984	$\beta_0$	1.92E-02	3.53E-04	54.30	2.07E-74	2.17E-02	5.51E-04	39.4	1.74E-61
	$\beta_1$	-3.92E-04	3.08E-05	-12.75	1.91E-22	-7.58E-04	4.80E-05	-15.80	1.42E-28
2015	$\beta_2$	6.43E-06	7.17E-07	8.97	2.22E-14	1.57E-05	1.12E-06	14.04	4.25E-25
	$\beta_3$	-4.45E-08	4.71E-09	-9.44	2.15E-15	-1.05E-07	7.35E-09	-14.28	1.37E-25
1985	$\beta_0$	1.90E-02	3.59E-04	52.92	2.33E-73	2.17E-02	5.69E-04	38.14	3.35E-60
	$\beta_1$	-3.75E-04	3.12E-05	-12.01	6.73E-21	-7.80E-04	4.96E-05	-15.74	1.82E-28
2015	$\beta_2$	6.07E-06	7.27E-07	8.35	4.88E-13	1.63E-05	1.16E-06	14.15	2.51E-25
	$\beta_3$	-4.23E-08	4.78E-09	-8.84	4.25E-14	-1.09E-07	7.59E-09	-14.38	8.69E-26
1986	$\beta_0$	1.88E-02	3.65E-04	51.47	3.13E-72	2.18E-02	5.88E-04	37.00	5.27E-59
	$\beta_1$	-3.59E-04	3.18E-05	-11.31	2.04E-19	-8.05E-04	5.12E-05	-15.72	2.05E-28
2015	$\beta_2$	5.75E-06	7.40E-07	7.76	8.55E-12	1.70E-05	1.19E-06	14.27	1.44E-25
	$\beta_3$	-4.03E-08	4.86E-09	-8.28	6.90E-13	-1.14E-07	7.84E-09	-14.48	5.45E-26
1987	$\beta_0$	1.86E-02	3.74E-04	49.73	7.83E-71	2.18E-02	6.09E-04	35.84	9.51E-58
	$\beta_1$	-3.46E-04	3.26E-05	-10.63	5.99E-18	-8.32E-04	5.30E-05	-15.70	2.24E-28
2015	$\beta_2$	5.49E-06	7.59E-07	7.23	1.10E-10	1.78E-05	1.24E-06	14.41	7.59E-26
	$\beta_3$	-3.88E-08	4.99E-09	-7.78	7.89E-12	-1.19E-07	8.12E-09	-14.62	2.96E-26
1988	$\beta_0$	1.85E-02	3.88E-04	47.65	4.19E-69	2.19E-02	6.32E-04	34.68	1.82E-56
	$\beta_1$	-3.36E-04	3.38E-05	-9.96	1.64E-16	-8.64E-04	5.50E-05	-15.70	2.21E-28
2015	$\beta_2$	5.32E-06	7.87E-07	6.77	9.97E-10	1.87E-05	1.28E-06	14.59	3.38E-26
	$\beta_3$	-3.80E-08	5.17E-09	-7.36	6.01E-11	-1.25E-07	8.43E-09	-14.80	1.27E-26
1989	$\beta_0$	1.83E-02	4.03E-04	45.52	2.90E-67	2.21E-02	6.60E-04	33.43	4.86E-55
	$\beta_1$	-3.28E-04	3.50E-05	-9.35	3.38E-15	-8.98E-04	5.74E-05	-15.63	3.02E-28
2015	$\beta_2$	5.21E-06	8.17E-07	6.38	6.01E-09	1.97E-05	1.34E-06	14.69	2.14E-26
	$\beta_3$	-3.76E-08	5.37E-09	-7.01	3.16E-10	-1.31E-07	8.80E-09	-14.92	7.47E-27
1990	$\beta_0$	1.82E-02	4.20E-04	43.20	3.73E-65	2.22E-02	6.93E-04	32.07	1.95E-53
	$\beta_1$	-3.22E-04	3.66E-05	-8.79	5.59E-14	-9.34E-04	6.03E-05	-15.47	6.15E-28
2015	$\beta_2$	5.18E-06	8.53E-07	6.07	2.50E-08	2.07E-05	1.41E-06	14.70	2.04E-26
	$\beta_3$	-3.77E-08	5.60E-09	-6.73	1.20E-09	-1.38E-07	9.24E-09	-14.95	6.34E-27
1991	$\beta_0$	1.80E-02	4.39E-04	41.02	4.35E-63	2.24E-02	7.22E-04	31.02	3.76E-52
	$\beta_1$	-3.19E-04	3.82E-05	-8.35	4.80E-13	-9.69E-04	6.28E-05	-15.42	7.87E-28
2015	$\beta_2$	5.25E-06	8.91E-07	5.89	5.55E-08	2.16E-05	1.46E-06	14.76	1.53E-26
	$\beta_3$	-3.86E-08	5.86E-09	-6.59	2.31E-09	-1.44E-07	9.62E-09	-15.01	4.97E-27

		Male				Female			
		Estimate	StdError	t-value	Pr(>   t  )	Estimate	StdError	t-value	Pr(>   t  )
1992	$\beta_0$	1.79E-02	4.59E-04	39.09	3.58E-61	2.25E-02	7.43E-04	30.28	3.15E-51
	$\beta_1$	-3.23E-04	3.99E-05	-8.08	1.77E-12	-1.00E-03	6.47E-05	-15.48	5.92E-28
2015	$\beta_2$	5.48E-06	9.30E-07	5.89	5.59E-08	2.24E-05	1.51E-06	14.89	8.52E-27
	$\beta_3$	-4.06E-08	6.11E-09	-6.63	1.86E-09	-1.49E-07	9.91E-09	-15.08	3.54E-27
1993	$\beta_0$	1.78E-02	4.73E-04	37.62	1.16E-59	2.26E-02	7.67E-04	29.47	3.38E-50
	$\beta_1$	-3.26E-04	4.12E-05	-7.91	4.20E-12	-1.03E-03	6.68E-05	-15.44	7.02E-28
2015	$\beta_2$	5.71E-06	9.60E-07	5.94	4.41E-08	2.32E-05	1.56E-06	14.92	7.26E-27
	$\beta_3$	-4.25E-08	6.31E-09	-6.73	1.17E-09	-1.54E-07	1.02E-08	-15.08	3.52E-27
1994	$\beta_0$	1.77E-02	4.81E-04	36.72	1.04E-58	2.26E-02	7.83E-04	28.80	2.47E-49
	$\beta_1$	-3.28E-04	4.19E-05	-7.83	6.04E-12	-1.05E-03	6.82E-05	-15.38	9.44E-28
2015	$\beta_2$	5.91E-06	9.76E-07	6.05	2.67E-08	2.37E-05	1.59E-06	14.90	7.94E-27
	$\beta_3$	-4.42E-08	6.41E-09	-6.89	5.57E-10	-1.57E-07	1.04E-08	-15.01	4.86E-27
1995	$\beta_0$	1.76E-02	4.93E-04	35.65	1.52E-57	2.25E-02	7.95E-04	28.32	1.07E-48
	$\beta_1$	-3.42E-04	4.29E-05	-7.96	3.32E-12	-1.07E-03	6.92E-05	-15.49	5.65E-28
2015	$\beta_2$	6.43E-06	1.00E-06	6.42	4.92E-09	2.43E-05	1.61E-06	15.09	3.49E-27
	$\beta_3$	-4.80E-08	6.57E-09	-7.31	7.62E-11	-1.60E-07	1.06E-08	-15.14	2.72E-27
1996	$\beta_0$	1.75E-02	5.13E-04	34.21	6.24E-56	2.25E-02	7.99E-04	28.11	2.02E-48
	$\beta_1$	-3.71E-04	4.47E-05	-8.32	5.63E-13	-1.11E-03	6.96E-05	-15.89	9.42E-29
2015	$\beta_2$	7.41E-06	1.04E-06	7.12	1.86E-10	2.53E-05	1.62E-06	15.59	3.56E-28
	$\beta_3$	-5.51E-08	6.84E-09	-8.05	2.06E-12	-1.66E-07	1.07E-08	-15.60	3.41E-28
1997	$\beta_0$	1.76E-02	5.45E-04	32.29	1.09E-53	2.26E-02	8.08E-04	27.95	3.31E-48
	$\beta_1$	-4.18E-04	4.75E-05	-8.80	5.28E-14	-1.16E-03	7.04E-05	-16.45	8.13E-30
2015	$\beta_2$	8.81E-06	1.11E-06	7.96	3.24E-12	2.66E-05	1.64E-06	16.25	1.97E-29
	$\beta_3$	-6.47E-08	7.27E-09	-8.90	3.22E-14	-1.75E-07	1.08E-08	-16.20	2.45E-29
1998	$\beta_0$	1.78E-02	6.02E-04	29.66	1.95E-50	2.29E-02	8.22E-04	27.84	4.70E-48
	$\beta_1$	-4.90E-04	5.24E-05	-9.36	3.28E-15	-1.23E-03	7.15E-05	-17.16	3.88E-31
2015	$\beta_2$	1.08E-05	1.22E-06	8.89	3.41E-14	2.84E-05	1.67E-06	17.01	7.36E-31
	$\beta_3$	-7.83E-08	8.02E-09	-9.76	4.40E-16	-1.85E-07	1.10E-08	-16.84	1.48E-30
1999	$\beta_0$	1.82E-02	6.73E-04	27.08	5.08E-47	2.32E-02	8.43E-04	27.55	1.17E-47
	$\beta_1$	-5.73E-04	5.86E-05	-9.79	3.84E-16	-1.29E-03	7.34E-05	-17.65	4.80E-32
2015	$\beta_2$	1.31E-05	1.36E-06	9.58	1.09E-15	2.99E-05	1.71E-06	17.51	8.76E-32
	$\beta_3$	-9.28E-08	8.97E-09	-10.35	2.30E-17	-1.93E-07	1.12E-08	-17.22	3.00E-31
2000	$\beta_0$	1.89E-02	7.57E-04	24.92	5.88E-44	2.37E-02	8.77E-04	27.03	6.02E-47
	$\beta_1$	-6.71E-04	6.59E-05	-10.18	5.46E-17	-1.36E-03	7.63E-05	-17.86	1.99E-32
2015	$\beta_2$	1.55E-05	1.53E-06	10.09	8.58E-17	3.15E-05	1.78E-06	17.69	3.97E-32
	$\beta_3$	-1.08E-07	1.01E-08	-10.72	3.70E-18	-2.02E-07	1.17E-08	-17.31	2.04E-31
2001	$\beta_0$	1.94E-02	8.40E-04	23.12	3.00E-41	2.40E-02	9.03E-04	26.62	2.23E-46
	$\beta_1$	-7.54E-04	7.31E-05	-10.31	2.90E-17	-1.41E-03	7.86E-05	-17.92	1.52E-32
2015	$\beta_2$	1.75E-05	1.70E-06	10.28	3.27E-17	3.25E-05	1.83E-06	17.75	3.17E-32
	$\beta_3$	-1.21E-07	1.12E-08	-10.80	2.52E-18	-2.08E-07	1.20E-08	-17.31	2.05E-31
2002	$\beta_0$	1.98E-02	9.02E-04	21.93	2.25E-39	2.40E-02	9.06E-04	26.55	2.82E-46
	$\beta_1$	-8.04E-04	7.85E-05	-10.24	4.02E-17	-1.41E-03	7.88E-05	-17.88	1.80E-32
2015	$\beta_2$	1.87E-05	1.83E-06	10.24	4.15E-17	3.26E-05	1.84E-06	17.73	3.42E-32
	$\beta_3$	-1.28E-07	1.20E-08	-10.68	4.70E-18	-2.09E-07	1.21E-08	-17.31	2.03E-31

		Male				Female			
		Estimate	StdError	$t$ -value	$\Pr(>  t )$	Estimate	StdError	$t$ -value	$\Pr(>  t )$
2003	$\beta_0$	1.94E-02	8.99E-04	21.54	9.65E-39	2.28E-02	9.29E-04	24.60	1.73E-43
	$\beta_1$	-7.70E-04	7.83E-05	-9.84	3.01E-16	-1.27E-03	8.08E-05	-15.76	1.72E-28
2015	$\beta_2$	1.81E-05	1.82E-06	9.92	2.01E-16	2.95E-05	1.88E-06	15.65	2.76E-28
	$\beta_3$	-1.25E-07	1.20E-08	-10.43	1.57E-17	-1.90E-07	1.24E-08	-15.37	9.87E-28
2004	$\beta_0$	2.05E-02	9.97E-04	20.59	3.47E-37	2.25E-02	1.03E-03	21.71	5.05E-39
	$\beta_1$	-9.02E-04	8.68E-05	-10.40	1.87E-17	-1.19E-03	9.01E-05	-13.23	1.88E-23
2015	$\beta_2$	2.11E-05	2.02E-06	10.44	1.52E-17	2.72E-05	2.10E-06	12.95	7.18E-23
	$\beta_3$	-1.43E-07	1.33E-08	-10.80	2.58E-18	-1.74E-07	1.38E-08	-12.65	3.10E-22
2005	$\beta_0$	2.02E-02	9.34E-04	21.61	7.45E-39	2.11E-02	1.17E-03	18.01	1.05E-32
	$\beta_1$	-8.52E-04	8.12E-05	-10.49	1.17E-17	-9.57E-04	1.02E-04	-9.40	2.61E-15
2015	$\beta_2$	1.97E-05	1.89E-06	10.40	1.87E-17	2.12E-05	2.37E-06	8.93	2.75E-14
	$\beta_3$	-1.33E-07	1.24E-08	-10.69	4.32E-18	-1.36E-07	1.56E-08	-8.70	8.35E-14

## References

- Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline, *Population Studies*, **56**, 325–336.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A Quantitative comparison of stochastic mortality models using data from England and Wales and the US, *North American Actuarial Journal*, **13**, 1–35.
- Hyndman, R. J. (2010). demography: Forecasting mortality, fertility, migration and population data. R package version 1.07. Contribution from Heather Booth and Leonie Tickle and John Maindonald.
- Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration, *International Journal of Forecasting*, **24**, 323–342.
- Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models, *Demography*, **50**, 261–283.
- Hyndman, R. J. and Ullah, S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach, *Computational Statistics & Data Analysis*, **51**, 4942–4956.
- Jeong, S. and Kim, K. W. (2011). A comparison study for mortality forecasting models 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.
- Jung, K. and Kim, D. (2012). An estimation of an old age mortality rate using CK model and relational model, *Communications of the Korean Statistical Society*, **19**, 859–868.
- Kang, J. C., Lee, D. S., and Shung, J. H. (2006). A study on the methods for forecasting mortality considering longevity risk, *The Journal of Risk Management*, **17**, 153–178.
- 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.
- Park, K. A. (2011). *Theory of practice of population projections*, Statistics Training Institute.
- Park, Y. S., Kim, K. W., Lee, D. H., and Lee, Y. K. (2005). A comparison of two models for forecasting mortality in South Korea, *The Korean Journal of Applied Statistics*, **18**, 639–654.
- Swanson, D. A. and Beck, D. M. (1994). A new short-term county population projection method, *Journal of Economic and Social Measurement*, **20**, 25–50.
- Tayman, J., Smith, S. K., and Lin, J. (2007). Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models, *Population Research and Policy Review*, **26**, 347–369.
- UNPD (2015). World Population Projections 2015 Revision.



# 시계열 이용기간에 따른 사망률 예측 비교

김순영<sup>a,1</sup> · 오진호<sup>a</sup> · 김기환<sup>b</sup>

<sup>a</sup>통계청 통계개발원, <sup>b</sup>고려대학교 응용통계학과

(2017년 9월 5일 접수, 2017년 11월 13일 수정, 2017년 12월 15일 채택)

---

## 요약

우리나라의 경우 선진국에 비해 짧은 기간 동안 사망률 개선이 급속히 이루어짐에 따라 사망률 예측에 있어 모형의 선택뿐만 아니라 시계열 이용기간의 선정 또한 중요한 고려사항이 될 수 있다. 따라서 본 연구에서는 시계열 이용기간의 선택 관점에서 회귀모형을 이용하는 방법을 제안하였다. 또한 Lee-Carter (LC) 모형, LC류 (Lee-Miller (LM), Booth-Maindonald-Smith (BMS)) 그리고 비모수 모형(functional data model (FDM), Coherent FDM)을 토대로 시계열 이용기간을 다르게 적용할 경우 어떠한 문제가 발생되며, 연령별 사망률과 기대수명 예측력에 어떠한 차이를 보이는지 살펴보았다. 분석결과를 바탕으로 5개의 모형별 2030년까지 남녀의 연령별 사망률과 예측기대수명을 작성하고 통계청(Korean Statistical Information Service; KOSIS)에서 제공하는 장래 연령별 사망률과 기대수명과 비교하였다.

주요용어: 사망률 예측 모형, 기대수명, 연령별 사망률, Lee-Carter 모형, FDM, Coherent FDM.

---

연구는 통계청의 공식견해가 아니며 저자의 개인적인 연구결과임을 밝힙니다.

<sup>1</sup>교신저자: (35220) 대전광역시 서구 한밭대로 713 통계센터 6층, 통계개발원 통계분석실.

E-mail: sykim0213@korea.kr