



## 제1장

# 장래인구추계를 위한 사망력 모형 연구

백지선 · 정미옥

## 제1절 서론

### 1. 연구배경

장래인구추계(population projection)는 인구주택 총조사, 주민등록인구와 더불어 대표적인 인구정태통계로 미래의 사회적, 경제적 국가 정책 수립에 중요한 국가통계이다. 통계청은 1962년에 1955년 센서스 자료를 기초자료로 1960~1980년 장래인구추계를 시작하여, 현재 2010년 인구주택 총조사 결과를 기반으로 향후 2060년까지 장래인구추계를 작성하여야 하는 시점에 와 있다. 우리나라 장래인구추계는 코호트 요인법에 의해 이루어지며, 신뢰성을 확보하기 위해서는 기준인구와 인구변동요인(출생, 사망, 인구이동)에 대한 정확성과 객관성이 요구된다. 인구변동요인 중 사망은 인구 감소를 가져오는 요인이며 전 연령층에서 발생하여 인구 구조의 변화에 직접적인 영향을 미치는 요인이다. 저출산과 맞물려 전 세계적으로 사망수준 개선과 고령화가 진행되면서 인구 구조의 변화가 일어나고 있으며, 이에 따라 고연령층의 사망력은 중요한 이슈로 부각되고 있는 상황이다.

사망력 예측과 관련하여 장래 사망수준과 기대여명 예측 등은 고령화에 따른 연금, 노동, 보험 등 가까운 미래에 발생 가능한 사회적·경제적 문제점들에 대비한 정책 수립 및 개발의 기초자료가 되므로 예측과정의 객관성뿐만 아니라 예측 결과에 대한 타당성 또한 뒷받침되어야 하는 매우 어렵고 중요한 통계이다. 우리나라 사망력 수준은 의료 기술 발전 속도와 더불어 단기간 동안 매우 급격한 속도로 개선되어 왔으며, 과거 고연

령 사망력 자료의 부족과 개선 속도가 향후 언제까지 지속될 것인지에 대한 의문은 장기예측을 어렵게 하는 요인이다. 따라서 우리나라 사망수준과 기대수명의 변화를 잘 반영할 수 있는 사망률 예측 모형과 더불어 예측결과에 대해 인구학적 측면을 고려한 검토과정이 필요하다고 판단된다. 사망력 예측 모형은 Lee-Carter(1992)가 제안한 확률모형(stochastic model)인 LC 모형이 적용이 용이하고 예측력도 높은 것으로 평가받고 있어 미국, 일본 등의 장래인구추계에서 사용되고 있다. 통계청은 2005년 장래인구추계에서 부분적으로 LC 모형을 적용하였으며, 정승환 외(2011)는 LC 모형을 포함하여 5개 예측모형을 1980~2008년 기간의 0~80세 사망률 자료에 대해 적용하고 예측결과를 비교하였다. 본 연구는 2011년 작성되는 장래인구추계를 앞두고 부족한 고연령 사망력 추정과 LC 모형을 중심으로 한 사망력 예측모형 적용 및 예측결과 비교를 통하여 확률모형에 대한 특성 및 타당성을 진단하고자 한다.

## 2. 연구내용

2005년 장래인구추계에서는 사망자료의 여러 가지 한계로 인해 1970~2005년까지 5세별 사망확률 자료에 대해 0세 및 1~4세는 일본의 2050년 수렴치, 5~74세 연령계급은 LC 모형, 75세 이상에 대해서는 Brass-Logit 모형을 적용하여 전체 0~100세에 대해 5세 단위 사망확률을 예측한 후 Beers 방법을 이용하여 각 세별 장래 사망수준을 산출하였다. 연령구간별로 다른 모형을 적용하였지만, 부분적으로나마 확률모형인 LC 모형에 기반하여 사망수준과 장래기대수명을 예측하였다는 점에 의미가 있다고 보여진다.

사망력 예측은 예측에 사용할 자료와 기간을 설정하고 예측 후보 모형을 적합한 후 장래사망확률을 예측하고 이를 이용하여 기대수명을 산출하는 일련의 과정으로 이루어지며, 일반적으로 50년 정도의 장기예측을 하고 있다. 우리나라는 현재 1970년부터 2010년 까지 약 40년 정도의 연령별 사망 자료 시계열이 수집되어 있으며, 과거에 비해 시계열이 늘어났지만, 향후 50년을 예측하기에는 40년 자료가 충분하다고 볼 수는 없을 것이다. 특히 사망은 신고에 의해 수집되는 인구동태자료로 신고 누락, 지연 등 기본적으로 문제를 포함하고 있으며, 신고율 개선과 함께 품질 개선이 이루어져 오고 있다. 과거 영아 사망의 경우 약 70%정도가 누락되는 것으로 추정되었으나, 1999년 이후부터 영아사망 화장신고 자료를 수집 보완하면서 품질이 크게 향상되었다(박경애, 2011). 통계청은 이러한 사망 관련 시계열을 보완·정비하여 국가통계포털에 1983년 이후의 자료를 제공하고 있다. 장래 사망력 예측을 위해서는 장기시계열이 필요하지만, 그에 못지않게 자료의 품질 또한 중요하며, 1970년 이후 자료와 정비된 1983년 이후 자료를 이용할 것인지에 대해서도 고민이 필요하다고 보여진다.



전반적인 연구방법은 다음과 같다. 장래인구추계를 위해서는 각 세별 사망수준이 산출되어야 하므로, 각 세별 사망률을 사망확률로 변환하고 그레빌 9차항 보정을 적용한 각 세별 사망확률을 이용하였다. 0세의 경우 사망률을 사망확률로 변환하였을 때 생명표 0세 사망확률과는 상당한 차이가 존재하였다. 이것은 과거 누락 등으로 인한 부족분을 보정한 생명표 사망확률이 타당한 것으로 판단되어 생명표 0세 사망확률을 사용하였다. 연도별 부족한 고연령에 대해서는 Brass-Logit 모형, 로지스틱 모형, CK(Coale and Kisker) 모형을 이용하여 75~99세 고연령층의 사망확률을 추정하고 추정결과를 비교하였다.

예측모형으로는 미국, 일본 등의 국가에서 사용하고 있는 LC 모형을 중심으로 검토하였다. LC 모형은 연령별 일반적인 사망률 패턴과 연령별 사망률 변화를 나타내는 연령효과, 사망률에 대한 시간효과로 구성되어 있으며, 적합기간이 주어진 경우 연령효과는 연령별 상수로 볼 수 있으므로 시간효과의 예측이 예측결과에 중요한 영향을 미치게 된다. 일반적으로 시간효과 모형으로 추세를 갖는 랜덤워크모형을 적용하고 있으나, 랜덤워크모형의 경우 과거의 선형 추세가 미래에도 지속됨으로 인하여 장기적 관점에서 사망확률 감소 추세가 느려질 것이라는 인구학적 측면을 설명하지 못한다는 단점을 가지고 있다. 이러한 점을 고려하여 일본 2001~2050년 장래인구추계에서는 사망확률이 지수적으로 감소하는 반-대수 모형을 사용하였다. 따라서 시간효과 예측모형으로 추세를 갖는 랜덤워크 모형과 반-대수모형을 고려하였다. 전체자료에 대한 예측에 앞서 적합기간과 모형의 예측력을 알아보기 위하여 1970~2005년과 1983~2005년 자료를 이용하여 적합하고 2006~2010년 생명표 결과와 비교하였다. 최종적으로 1970~2010년과 1983~2010년 자료를 이용하여 모형을 적합하고 예측결과를 비교하였다.

보고서의 구성은 다음과 같다. 2절에서는 적용 가능한 고연령 사망력 추정 모형을 소개하고 특징 및 추정방법에 대해서 살펴보았으며, 3절에서는 사망력 예측을 위한 모형으로 LC 모형을 중심으로 특성에 대하여 살펴보았다. 4절에서는 우리나라 사망자료와 0세 및 고연령 사망력 추정 및 예측 방법을 포함하여 0~100세 전체 연령을 이용한 사망력 예측 방법과 예측결과를 비교·정리하였다. 5절에서는 연구결과를 정리하였다.

## 제2절 고연령 사망력 모형

사망은 인구동태자료로 대부분의 국가에서 사망신고에 의해서 자료가 관리되고 있으며, 지연, 왜곡, 누락 등으로 인해 자료의 품질에 많은 문제점을 가지고 있다. 특히 노인 인구의 증가와 과거 노인인구 사망에 관련된 자료의 질적·양적 부족은 사망력 예측을

더욱 어렵게 하는 요인이 되고 있다. 하지만 사망력의 감소가 대부분의 사망이 발생하는 고연령에서 발생하고 있고, 개선폭이 점차 가속화될 전망이므로 고연령 사망수준을 설명할 수 있는 모형은 더욱 중요하다고 할 수 있다. 우리나라는 2000년 이후 고령화가 빠른 속도로 진행되고 있으며 고연령의 사망수준은 장래의 기대수명에 중요한 영향을 미치는 요인이 될 것이다. 우리나라의 고연령 사망자료의 최고연령은 연도별로 다르게 집계되어 예측에 필요한 0~100세 연령구간이 부족한 상황으로 예측에 앞서 과거의 고연령 사망확률을 100세까지 확장하는 것이 필요하다.

연령별 사망률에 대해 사망질서(혹은 사망규칙)의 관점에서 수리적 모형으로 Gompertz 모형, Gompertz-Makeham 모형 등 여러 가지 함수가 제안되어 왔다. 또한 사망률에 대한 관계적 모형으로 William Brass(1971)에 의해 제안된 로짓모형이 있으며, Coale and Kisker(1990)가 사망률의 지수 증가율이 상수가 아니라 선형적으로 감소하는 패턴에 근거하여 제안한 CK모형이 있다. 수리적 모형은 연령에 따른 사망규칙을 함수로 표현한 것으로 전반적으로 실측치 적합 결과가 우수하지만, 모형 추정에 어려움이 있으며, CK 모형과 Brass-Logit 모형은 사망확률에 대한 일련의 로그변환을 통한 선형관계에 기반한 모형으로 적합이 용이하다는 장점이 있다. Thatcher, Kannisto and Vaupel(1998)는 13개 국가에 대해 Gompertz 함수와 여러 가지 로지스틱 함수들을 적합하였으며, 분석 결과 Kannisto에 의해 제안된 2-모수 로지스틱 함수가 일관적으로 가장 좋은 결과를 제공함을 보였다. Yi and Vaupel(2003)는 고연령 중국 한족에 대해 연령별 사망률 산출을 위하여 수리적 모형을 적합한 결과 Gompertz 함수는 실제 사망률보다 고연령 사망률을 높게 예측하는 경향이 있다는 것과 2-모수 로지스틱 모형이 가장 우수함을 보였다. Edviges Coelho(2007)는 포르투갈 고연령 사망력에 대해 Gompertz 모형, 로지스틱 모형, CK 모형, Denuit and Goderniaux 방법을 적합하고 비교하여, 전통적으로 사용된 고펜르츠 모형이 부적절함을 보이고 Denuit and Goderniaux(2005)가 제안한 이차함수모형의 결과가 가장 적절함을 보였다. 선행연구로부터 전통적인 Gompertz 모형은 고연령 사망력을 과대추정하여 부적절하며, 2-모수 로지스틱 모형이 사망률이 낮은 국가들의 고연령 사망력 추정에 우수한 결과를 준다는 것을 알 수 있다.

## 1. 수리적 모형

### 가. Gompertz-type 모형

Gompertz-type 모형은 포함된 모수의 수에 따라 2-모수 모형, 3-모수 모형으로 구분된다. 2-모수 모형은 Gompertz(1825)가 제안한 모형으로 사망력이 연령에 따라 지수형태로 증가함을 표현하였으며,  $m_x$  를 연령  $x$ 에서 사망률이라고 할 때, 모형식은 식(1)과 같



다. 이 모형은 전통적인 사망률 모형으로 널리 사용되어 오고 있지만, 저연령에서는 관측사망률보다 과소추정하며, 고연령에서는 과대추정하는 경향이 있는 것으로 알려져 있다.

$$m_x = Be^{qx} \quad (1)$$

Gompertz가 제안한 2-모수 모형을 개선하여 Makeham(1867)이 우발적인 사고에 의한 사망을 나타내는 상수항을 포함하여 식(2)와 같은 3-모수 모형을 제안하였으며, 이 모형은 Gompertz-Makeham 모형이라고도 한다.

$$m_x = A + Be^{qx} \quad (2)$$

#### 나. Logistic-type 모형

Gompertz-type 모형의 고연령 사망률 과대 추정을 개선하기 위하여 로지스틱 형태의 모형들이 제안되었다. 포함된 모수의 수에 따라 4-모수 모형, 3-모수 모형 그리고 2-모수 모형으로 구분된다. 4-모수 모형은 Perks(1932)가 제안한 것으로 모형은 식(3)과 같다. Beard(1963)는 식(4)와 같은 3-모수 모형 제안하였으며, Kannisto(1992)는 식(5)와 같은 2-모수 모형을 제안하였다.

$$m_x = \frac{A + Be^{\beta x}}{1 + Ce^{\beta x}} \quad (3)$$

$$m_x = \frac{Be^{\beta x}}{1 + Ce^{\beta x}} \quad (4)$$

$$m_x = \frac{Be^{\beta x}}{1 + Be^{\beta x}} \quad (5)$$

여기서  $\beta$  는 연령별 사망력의 지수 증가율을 의미하며, 연령  $x$ 가 무한대로 가면 2-모수 모형은 1로 수렴하며, 3-모수, 4-모수 모형은  $B/C$  로 수렴한다. Perk의 4-모수 모형에서  $A = 0$  이고  $C = 0$ 이면 Gompertz 2-모수 모형,  $C = 0$  이면 Makeham 3-모수 모형,  $A = 0$  이면 Beard의 3-모수 모형,  $A = 0$  이고  $B = C$  이면 Kannisto 2-모수 모형을 나타낸다.

## 2. CK 모형

CK 모형은 Coale and Kisker(1990)가 고연령층 사망률 자료가 질적으로 우수하고 매우 낮은 사망률을 가지는 네덜란드, 일본, 프랑스 등에서 사망률의 지수 증가율이 상수가 아니라 선형적으로 감소하는 패턴에 근거하여 제안하였다. 이 모형은 고연령에 대한 예측 사망률의 지수 증가율  $k_x$ 가 목표 연령까지 상수가 아니라 선형적으로 감소함을 나타내며 모형식은 다음과 같다. 참고로 이 모형은 우리나라 생명표 사망확률 산출시에 85세 이상 고연령 사망확률을 보정하는 데 사용되고 있다.

$$m_x = m_{x-1} \cdot e^{k_x} \quad \text{또는} \quad \ln\left(\frac{m_x}{m_{x-1}}\right) = k_x, \quad (6)$$

$$k_x = k_{x_0} + (x - x_0) \cdot s, \quad x = x_0, x_0 + 1, \dots, x_{\max}$$

여기서,  $x_0$ 는 예측 시작연령,  $x_{\max}$ 는 예측 목표연령이다. 식(6)의 CK 모형은  $m_x = m_{x_0-1} \cdot \exp\left[\sum_{y=x_0}^x k_y\right]$ 과 같이 재표현 할 수 있으며, 이 식을 이용하여 모수  $s$ 를 추정할 수 있다. 만약  $k_x = k$  (상수)이면, 다음과 같이 Gompertz 모형과 유사한 형태로 표현된다.

$$\begin{aligned} m_x &= m_{x_0-1} \cdot \exp\left[\sum_{y=x_0}^x k_y\right], \quad x = x_0, x_0 + 1, \dots, x_{\max} \\ &= m_{x_0-1} \cdot \exp[(x - x_0 + 1)k]. \end{aligned}$$

사망률의 지수 증가율이 선형적으로 감소한다는 가정에 의해서  $m_x$ 는 다음과 같이 표현된다.

$$m_x = m_{x_0-1} \cdot \exp\left[\sum_{y=x_0}^x k_y\right] = m_{x_0-1} \cdot \exp\left[\sum_{y=x_0}^x k_{x_0} + (y - x_0) \cdot s\right] \quad (7)$$

따라서,  $x = x_{\max}$ 로 두면 위 식으로부터 사망률 지수 증가율  $k_x$ 의 변화 기울기  $s$ 를 식(8)과 같이 추정할 수 있으며, 모수  $s$ 를 추정하기 위해서는 최고연령(예측 목표연령) 사망률  $m_{x_{\max}}$  값에 대한 가정이 필요하다.



$$s = - \left[ \ln \left( \frac{m_{x_0-1}}{m_{x_{\max}}} \right) + (x_{\max} - x_0 + 1)k_{x_0} \right] \Bigg/ \left[ (x_{\max} - x_0)(x_{\max} - x_0 + 1)/2 \right] \quad (8)$$

예를 들어,  $x_0 = 75$ ,  $x_{\max} = 110$ 이면,  $m_{110}$ 과  $s$ 는 식(7)과 (8)에 의해서 다음과 같이 각각 표현되며, 남성에 대해  $m_{110} = 1.0$ , 여성에 대해  $m_{110} = 0.8$ 으로 가정을 하면 변화기울기  $s$ 를 추정할 수 있다.

$$m_{110} = m_{74} \cdot \exp \left[ \sum_{y=75}^{110} (k_{75} + (y-75)s) \right],$$

$$s = - \left[ \ln \left( \frac{m_{74}}{m_{110}} \right) + 36k_{75} \right] \Bigg/ 630.$$

### 3. Brass-Logit 모형

Brass-Logit 모형은 생명표 작성을 위해 William Brass(1971)에 의해 제안된 것으로 두 개의 다른 연령별 사망패턴의 사망률 로짓들간에 선형관계가 성립함을 가정하는 관계적 모형이다. 이 모형은 표준인구의 연령별 사망률 로짓과  $t$ -연도의 연령별 사망률 로짓간의 선형관계를 나타내며, 모형식은 다음과 같다

$$\text{Logit}(m_{x,t}) = \alpha + \beta \cdot \text{Logit}(m_{x,s}) \quad (9)$$

여기서  $\text{Logit}(m_{x,t}) = 0.5 \cdot \ln(m_{x,t}/m_{x,t}^*)$ ,  $m_{x,t}^* = 1 - m_{x,t}$  이고,  $m_{x,s}$ 는 표준인구의 연령별 사망률이다. 모수  $\alpha$ 는 표준인구 사망수준에 대한 상대적 사망수준을 나타내는 것으로,  $\alpha > 0$ 이면 표준인구보다 높은 사망률을,  $\alpha < 0$ 이면 표준인구보다 낮은 사망률을,  $\alpha = 0$ 이면 표준인구와 같은 사망률을 갖는 것을 의미한다. 모수  $\beta$ 는 표준인구와  $t$ -연도간의 고연령과 저연령의 사망수준 변화에 대한 상대적 크기를 나타내는 것으로  $\beta > 1$ 이면 고연령의 사망수준 변화가 크고,  $\beta < 1$ 이면 저연령의 사망수준 변화가 크다는 의미이다.  $\beta = 1$ 이면 고연령과 저연령의 사망수준 변화가 같다.

관측된  $m_{x,t}$ 와  $m_{x,s}$ 에 의해 선형관계를 만족하는  $\alpha$ ,  $\beta$ 를 추정할 수 있으며, 임의의 모든 생명표에 대해 위의 모형식이 만족되고,  $(\alpha, \beta)$ 가 주어지면 모든 생명표를 생성하는 것이 가능하다. Brass-Logit 모형은 표준인구의 연령별 사망확률과 관측된 사망확률의

차이가 클 때, 선형성 이탈이 커지게 될 수 있으나, 기초자료가 부족한 고령자 사망률 추정을 위해 UN에서 권고하는 방법이다. 2005년 장래인구추계 사망률 예측에서 부족한 고연령 사망확률에 대해 이 모형을 이용하여 75세 이상 사망확률을 예측하였다.

### 제3절 사망률 예측 모형

장래 사망률 예측은 미래에 대한 예측으로 불확실성(uncertainty)을 내포하게 되며, 장기로 갈수록 불확실성은 더욱 커지게 된다. 따라서 이러한 불확실성을 처리하기 위해 확률적인 접근방법으로 여러 가지 통계적 모형들이 제시되고 있다. 통계적 모형은 과거자료에 대한 경험적 분석을 바탕으로 모형을 설정하여 미래값을 예측하고, 예측에 대한 불확실성을 신뢰구간으로 나타내는 객관적 방법론을 제공한다는 장점을 가지고 있으나, 장기예측에 있어서 과거자료에 의해 설정된 모형이 그대로 지속될 것인지에 대해서는 장담하기 어려운 점이 있다. 그러나 통계적 모형은 전문가 판단에 기초한 방법에 비해 가능한 주관적인 측면을 배제하며 장기예측에 대해 적절한 결과를 도출하는 것으로 알려져 널리 사용되고 있다. 사망률 예측 통계모형으로 가장 대표적으로 사용하고 있는 모형은 Lee-Carter(1992)에 의해 제안된 LC모형이며, LC모형을 개선한 LM(Lee-Miller,2001) 모형, BMS(Booth, Maindonald and Smith, 2002)모형 등이 있다. 본 절에서는 LC 모형과 개선모형들에 대해 간략하게 정리하였다.

#### 1. LC 모형

LC모형은 Lee and Cater(1992)가 제안한 모형으로 식(10)과 같이 로그 연령별 사망률(age-specific mortality)에 대하여 연령효과( $a_x, b_x$ )와 시간효과( $k_t$ )로 구성되어 적용이 간단하며, 예측력에 있어서도 우수하다고 알려져 있다.

$$\ln[m(x,t)] = a_x + b_x \cdot k_t + \epsilon_{x,t}, \quad x = 0, 1, \dots, \omega; \quad t = 1, 2, \dots, T \quad (10)$$

여기서  $a_x$ 는 연령별 일반적인 사망률 패턴,  $b_x$ 는 연령별 사망률 변화,  $k_t$ 는 사망률에 대한 시간추세(underlying time trend)를 나타낸다.  $\epsilon_{x,t}$ 는 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따르는 확률변수이다.





LC 모형은 유일한 해(unique solution)을 얻기 위해서  $\sum_x b_x = 1$ 과  $\sum_t k_t = 0$ 이라는 제약조건을 두고 비정칙 분해(singular value decomposition)를 통해 얻어지는 첫 번째 주성분을 이용하여  $a_x, b_x, k_t$  를 추정한다. 추정된  $\hat{k}_t (t = 1, 2, \dots, T)$ 에 대해 로그사망률에서의 오차를 최소화되도록 아래의 식을 만족하는  $k_t$ 를 재추정한다.

$$D(t) = \sum_x [N(x, t) \cdot \exp(\hat{a}_x + \hat{b}_x \hat{k}_t)] \quad (11)$$

여기서  $D(t) = \sum_x D(x, t)$ 로 관측된 총 사망자 수이고  $N(x, t)$ 는 연령별 인구수를 나타낸다.

재추정된  $k_t (t = 1, 2, \dots, T)$ 에 대해 모형을 적합하고 적합된 모형을 이용하여  $k_{T+j} (j = 1, 2, \dots)$ 를 예측한 후 LC 모형에 적용하여 사망률을 예측하게 된다. 이 때  $k_t$ 에 대한 예측모형으로 기본적으로 식(12)와 같은 추세를 갖는 임의확률보행 모형(random walk with drift)를 사용하지만, 다른 시계열 모형을 사용할 수도 있다.

$$k_t = d + k_{t-1} + e_t. \quad (12)$$

여기서  $d$ 는 임의확률보행 과정의 추세를 나타내는 모수이며,  $e_t$ 는 평균이 0이고 분산이  $\sigma_e^2$ 인 정규분포를 따르는 확률변수이다.

LC 모형은 적용이 용이하며, 만족스러운 예측결과와 미래예측에 대한 불확실성을 신뢰구간으로 제공하여 준다는 장점을 가지고 있다. 반면, 연령별 사망률의 변화가 일정하게 유지되는 것으로 가정하고 있으므로 연령층에 따른 변화 속도의 차이를 고려할 수 없다는 단점을 가지고 있다.

## 2. LM 모형

LM 모형은 Lee와 Miller(2001)가 제안한 모형으로 LC 모형에 대해 다음의 3가지를 개선하였다. 첫째, LC 모형은 연령별 사망률 변화  $b_x$ 가 시간에 대해 일정하다는 가정을 하였지만, 미국의 사망자료가 1950년을 기준으로 연령별 사망률 변화가 다른 것을 발견하고 1950년 이후의 자료를 이용하여 모형을 적합하였다. 둘째,  $k_t$ 를 기대수명을 이용하여 재추정하였다. 셋째,  $a_x$ 에 대해 새로운 추계를 시작하는 연도의 연령별 사망률 실

측치로 대체하고 그 해의  $k_t$ 를 0으로 둬으로써 탈선으로 인한 편의(jump-off bias)를 제거하였다.

### 3. BMS 모형

BMS 모형은 Booth-Maindonald-Smith(2002)가 제안한 모형으로  $k_t$ 가 선형이라는 가정 아래 식 (14)의 적합도 검정을 통하여 적합기간을 선택하고 총 사망자수가 아니라 사망에 대한 연령 분포(연령별 사망자수)를 이용하여  $k_t$ 에 대하여 조정을 하였다.

$$Deviance(t) = 2 \sum_x D(x,t) \ln [D(x,t)/D'(x,t)] - (D(x,t) - D'(x,t)). \quad (13)$$

여기서  $D(x,t)$ 는 연령별 사망자 수,  $D'(x,t)$ 는 적합된 사망자 수이다.

### 4. HU 모형

Hyndman과 Ullah(2007)는 로그사망률을 식(14)와 같이 평활함수(smooth function)와 관측오차로 구성하였으며, 평활함수  $f_t(x)$ 를 식(15)와 같이 나타내고 비모수적 평활방법을 이용하여 추정하였다. HU모형은 비정칙 분해에서 얻어진  $(k_t, b_x)$ 에 대한 주성분을 두개 이상을 사용하고  $k_t$ 에 대한 보정은 하지 않는다.

$$\ln m(x,t) = f_t(x) + \sigma_t(x)\epsilon_{x,t} \quad (14)$$

$$f_t(x) = a(x) + \sum_{j=1}^J k_{t,j} b_j(x) + e_t(x) \quad (15)$$

여기서  $a(x)$ 는 연령에 따른 일반적인 사망률 패턴이며,  $k_{t,j}$ 는 주성분  $j$ 의 시간에 따른 사망률 변화와  $b_j(x)$ 는 연령별 사망률 변화를 나타낸다.



## 제4절 장래인구추계를 위한 사망력 예측

### 1. 사망자료

우리나라 사망자료(<표 1-1> 참조)는 1970년부터 연도에 따른 연령별 사망자수( $D_x$ )와 연앙인구( $P_x$ )를 기초로 하여 산출된 각 세별 사망률, 각 세별 사망확률, 5세별 사망확률 자료가 있으며, 각 연도에 따라 집계된 최고연령이 다르다. 1970~2010년에 대해 제공되는 연앙인구는 1992년 이전에 대해서는 추계인구이며, 1992년부터는 주민등록인구로 제공되고 있다. 사망자료에서 5세별 사망확률은 1970년부터 0~80세 이상까지 자료가 구축되어 있으며, 각 세별 사망확률은 2001년부터 0~100세 이상까지 자료가 있다. 사망률은 1970년부터 자료가 수집되어 있으며, 1983년 이후에 대해서 KOSIS(국가통계포털, Korean Statistical Information Service)에 제공되고 있다. 2005년 추계에서는 1970~2005년 5세별 사망확률에 대해 LC모형을 이용하여 2050년까지 사망확률을 예측한 후 75세 이상에 대해서 Brass-Logit 모형을 이용하여 장래 사망확률을 예측하고 장래 간이생명표를 작성한 후 Beers 방법을 적용하여 완전생명표를 작성하였다.

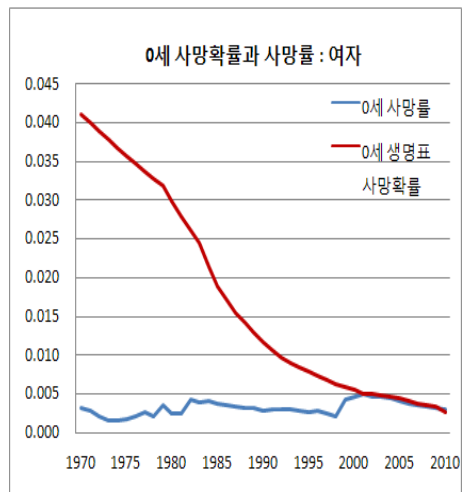
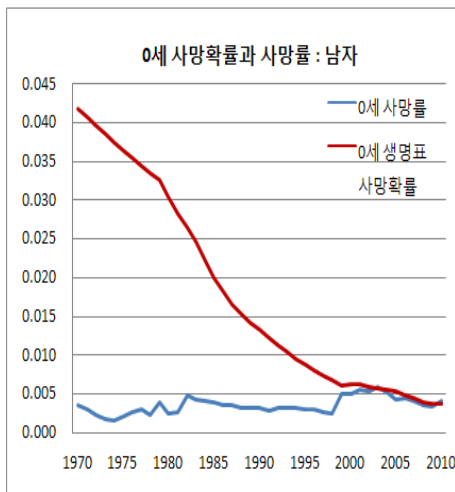
장래인구추계를 위해서는 궁극적으로 각 세별 장래생명표가 필요하므로, 본 연구에서는 각 세별 사망률 자료를 사망확률로 변환하고 이를 이용하여 사망력 예측모형을 검토하였다. [그림 1-1]과 [그림 1-2]에 1970~2010년 각 세별 사망률 시계열 그림을 나타내었다. [그림 1-1]은 남녀별 0세 사망률과 생명표 0세 사망확률 시계열 그림이다. 0세 사망률은 감소하다가 1999년부터 화장장 자료가 추가되면서 점프가 발생한 후 다시 감소하는 패턴을 보이고 있다. 0세 사망률은 출생신고 전 사망(영아사망 미신고) 등으로 인한 누락이 매우 크며, 신고율 개선과 더불어 누락분에 대한 보완을 위하여 화장장 자료 수집 및 행정자료를 활용하면서 자료의 질이 개선되었다. 과거 신고에 의한 영아사망은 과소집계 정도가 매우 크므로 통계청에서는 여러 가지 자료를 활용하여 보다 정확한 영아 사망확률을 추정하여 생명표에 제공하고 있다. 따라서 0세에 대해서는 생명표의 사망확률을 사용하기로 한다. [그림 1-2]는 1970~2010년까지 1~80세 이상 각 세별 사망률 시계열 그림이다. 1970년대 사망률 자료는 변동이 크며 1983년 이후부터는 안정적으로 감소하는 패턴을 보이고 있다.

정리하면, 장래 사망확률 예측을 위한 기초자료는 1970~2010년의 연앙인구와 사망자수이며, 0세에 대해서는 생명표의 0세 사망확률을 사용하고 그 외 연령에 대해서는 각 세 사망률을 사망확률로 변환하는 일반적인 공식  $q_x = m_x / (1 + 0.5m_x)$ 과 그레빌 9차항 보정을 이용하여 사망확률로 변환하였다. 1970~1999년의 부족한 고연령의 사망확률에

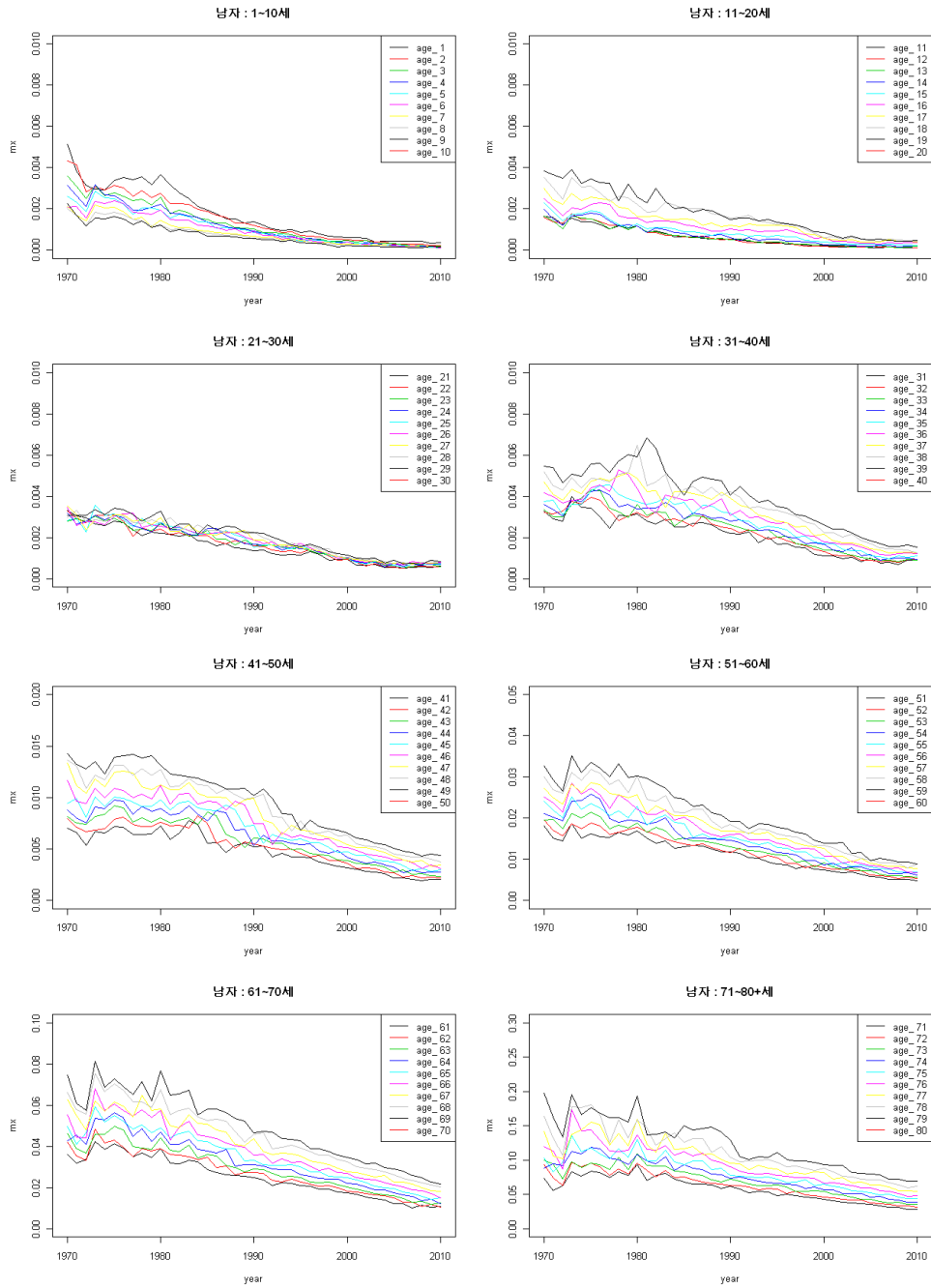
대해서는 Brass-Logit 모형, CK 모형, 2-모수 로지스틱 모형을 각각 적합하고 75세 부터 100세까지로 확장된 사망확률을 추정하여 예측에 필요한 0세부터 100세까지 각 세 사망 확률 자료를 구성하였다.

〈표 1-1〉 우리나라 연도별 사망자료 : 1970~2010년

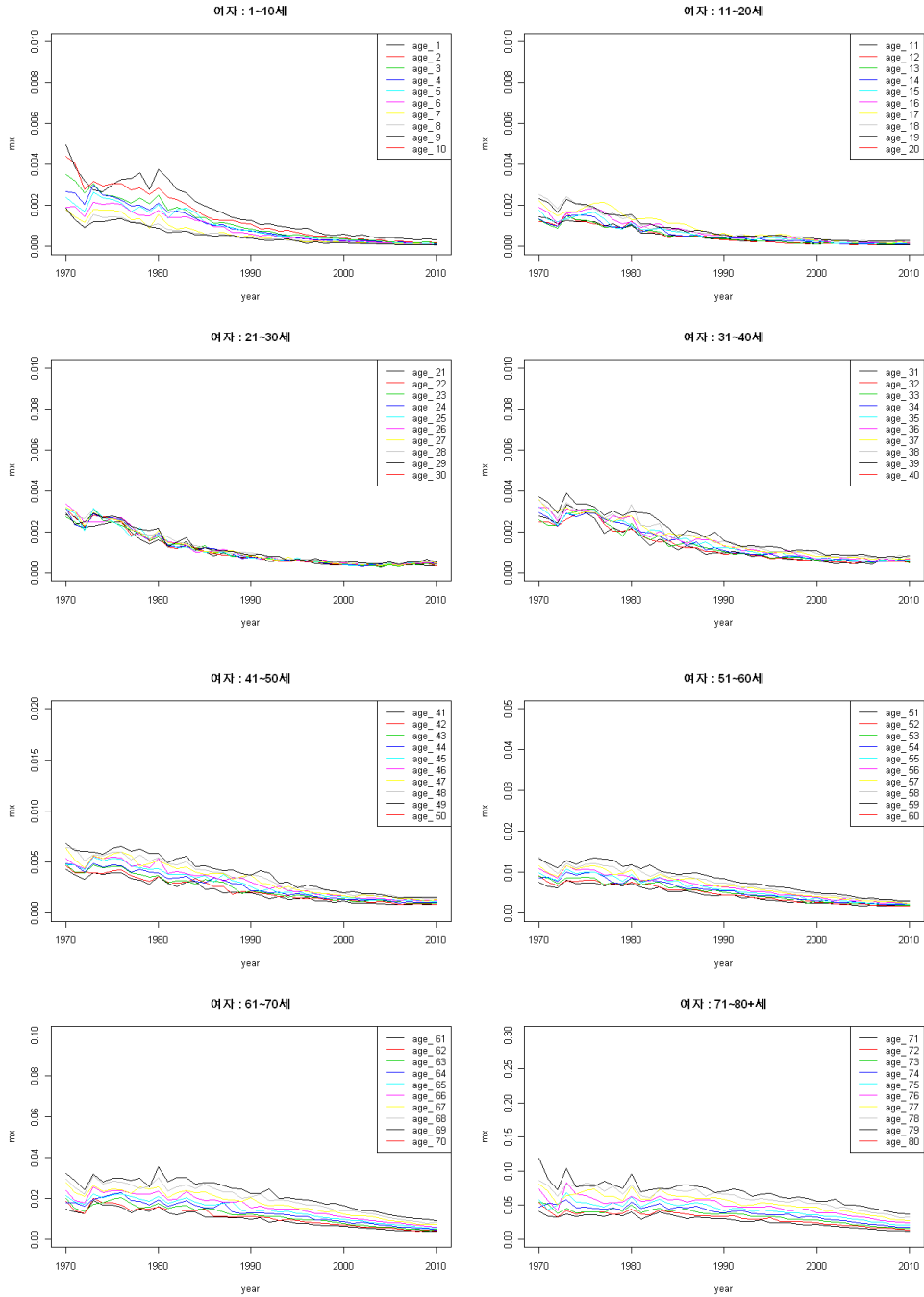
	1970 ... 1982 1983 ... 1992 1993 ... 1997 1998 1999 2000 2001 ... 2010
연앙인구	(1970~2010) 각세별, 5세별, 0~80+
사망률 (사망자수)	(1970~1982) 각세별, 5세별 0~80+ (1983~1999) 각세별, 5세별 0~80+ (2000~2010) 각세별, 5세별 0~100+
각세별 사망확률	(1997) 0~85+ (1999) 0~95+ (2001~2010) 0~100+
5세별 사망확률	(1970~1992) 0,1~80+ (1993~1998) 0,1~85+ (1999~2000) 0,1~95+ (2001~2010) 0,1~100+



[그림 1-1] 남녀 영아 사망률과 생명표 사망확률, 1970~2010년



[그림 1-2] 남녀별 1~80+세 사망률, 1970~2010년 : 남자



[그림 1-2] 남녀별 1~80+세 사망률, 1970~2010년 : 여자 (계속)



## 2. 고연령 사망확률 추정

LC모형을 이용한 사망력 예측에 앞서 부족한 75~100세 이상 고연령 사망확률 추정이 필요하며, 고연령 사망력 모형에 대해서는 2절에서 살펴보았다. 본 연구에서는 고연령 사망확률 추정 모형으로 2-모수 로지스틱 모형, CK 모형, Brass-Logit 모형을 검토하였다. CK모형은 현재 완전생명표에서 85세 이상 고연령 보정에 사용하고 있는 모형이며, Brass-Logit 모형은 2005년 추계에서 사용되었다. 2-모수 로지스틱 모형은 사망률이 낮은 경우에 우수한 추정 결과를 보인다는 선행연구 결과가 있다. CK 모형을 적합하기 위해서는 최고연령의 사망확률에 대한 가정이 필요하므로 선행연구 결과를 근거로 하여 110세에 대해 남자는 1.0, 여자는 0.8을 고려하여 모형을 적합하였다. Brass-Logit 모형은 2010년 잠정 작성된 완전생명표를 표준인구에 대한 생명표(이하 표준생명표)로 하고 (60,75)세의 사망확률 로짓을 이용하여 모형을 적합하였다. 2-모수 로지스틱 모형은 남녀 각각 (65~75)세 사망확률을 이용하여 모형을 적합하였으며, 각 모형별 모수 추정 결과는 <표 1-2>에 수록하였다.

<표 1-2> 모형별 모수추정 결과

연도	Logistic				Brass-Logit				Coale-Kisker	
	남자		여자		남자		여자		남자	여자
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$s$	$s$
1970	-7.716	0.073	-10.778	0.107	0.116	0.770	0.126	0.779	0.0000	-0.0007
1971	-8.357	0.081	-11.365	0.113	0.078	0.775	0.051	0.777	0.0012	0.0013
1972	-7.958	0.074	-11.217	0.109	0.033	0.764	-0.018	0.766	0.0003	0.0021
1973	-8.032	0.080	-10.766	0.107	0.453	0.898	0.240	0.808	-0.0035	-0.0026
1974	-8.232	0.081	-9.995	0.094	0.357	0.879	0.062	0.767	-0.0034	-0.0043
1975	-8.156	0.081	-9.601	0.089	0.355	0.861	0.025	0.748	-0.0028	-0.0039
1976	-8.176	0.081	-9.259	0.084	0.291	0.841	-0.038	0.729	-0.0020	-0.0036
1977	-8.080	0.078	-9.494	0.086	0.182	0.812	-0.098	0.724	-0.0003	-0.0004
1978	-8.124	0.080	-10.223	0.097	0.239	0.825	-0.011	0.747	-0.0001	-0.0004
1979	-8.132	0.079	-9.977	0.093	0.178	0.815	-0.032	0.752	-0.0008	-0.0002
1980	-9.185	0.096	-10.792	0.107	0.423	0.902	0.208	0.818	-0.0013	-0.0006
1981	-9.363	0.096	-10.889	0.105	0.274	0.871	0.085	0.807	-0.0006	-0.0013
1982	-8.716	0.087	-10.806	0.105	0.223	0.847	0.193	0.846	-0.0001	-0.0005

연도	Logistic				Brass-Logit				Coale-Kisker	
	남자		여자		남자		여자		남자	여자
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$s$	$s$
1983	-8.984	0.091	-10.924	0.108	0.396	0.918	0.312	0.878	-0.0016	-0.0021
1984	-8.718	0.086	-10.899	0.106	0.234	0.871	0.204	0.856	-0.0012	-0.0018
1985	-8.969	0.089	-11.150	0.110	0.312	0.913	0.285	0.897	-0.0005	-0.0007
1986	-8.974	0.089	-11.186	0.110	0.364	0.949	0.318	0.916	-0.0017	-0.0006
1987	-9.089	0.090	-11.480	0.114	0.342	0.940	0.350	0.927	-0.0016	-0.0016
1988	-9.156	0.090	-11.441	0.113	0.245	0.911	0.308	0.915	-0.0012	-0.0022
1989	-9.544	0.095	-11.775	0.117	0.247	0.917	0.262	0.904	-0.0013	-0.0033
1990	-9.190	0.090	-11.485	0.112	0.164	0.889	0.146	0.870	0.0004	-0.0010
1991	-9.469	0.093	-11.966	0.119	0.198	0.915	0.226	0.903	0.0003	0.0000
1992	-9.780	0.097	-12.210	0.122	0.241	0.955	0.327	0.959	0.0003	0.0006
1993	-9.392	0.091	-11.784	0.115	0.154	0.906	0.221	0.921	-0.0007	-0.0014
1994	-9.478	0.092	-12.012	0.118	0.136	0.896	0.276	0.944	-0.0003	-0.0018
1995	-9.536	0.093	-12.400	0.123	0.140	0.908	0.348	0.984	0.0002	-0.0015
1996	-9.734	0.095	-12.368	0.122	0.102	0.899	0.358	1.000	0.0001	-0.0015
1997	-9.941	0.097	-12.673	0.126	0.134	0.922	0.328	0.992	-0.0005	-0.0016
1998	-9.852	0.096	-12.684	0.125	0.150	0.935	0.356	1.018	-0.0012	-0.0023
1999	-10.119	0.099	-13.082	0.131	0.171	0.958	0.365	1.022	-0.0009	-0.0022
2000	-10.132	0.099	-13.331	0.134	0.142	0.953	0.371	1.036	-0.0009	-0.0021
2001	-10.311	0.101	-13.328	0.133	0.160	0.974	0.333	1.033	-0.0006	-0.0020
2002	-10.300	0.100	-13.357	0.133	0.115	0.960	0.325	1.032	-0.0011	-0.0024
2003	-10.336	0.100	-13.409	0.133	0.135	0.987	0.333	1.051	-0.0008	-0.0030
2004	-10.354	0.100	-13.481	0.133	0.161	1.012	0.285	1.044	-0.0007	-0.0027
2005	-10.491	0.101	-13.727	0.135	0.183	1.040	0.248	1.040	-0.0013	-0.0026
2006	-10.628	0.102	-13.550	0.132	0.194	1.061	0.232	1.052	-0.0012	-0.0031
2007	-10.850	0.105	-13.749	0.134	0.193	1.067	0.172	1.034	-0.0017	-0.0022
2008	-11.202	0.109	-13.892	0.135	0.125	1.044	0.053	0.998	-0.0001	-0.0015
2009	-11.205	0.108	-13.933	0.134	0.061	1.023	0.066	1.013	-0.0004	-0.0025
2010	-11.632	0.114	-13.844	0.133	0.048	1.015	0.056	1.016	-0.0007	-0.0021





고려한 모형의 사망확률 적합력을 파악하기 위하여 100세 이상까지 사망확률(생명표 사망확률)이 제공되는 2001~2010년에 대해 모형으로 75세 이상 사망확률을 추정하고 생명표 로그사망확률, 관측된 사망률을 변환한 로그사망확률(이하 관측 로그사망확률)을 비교하였다. 2001~2010년에 대해서 모형으로 추정된 로그사망확률, 생명표 로그사망확률 그리고 관측 로그사망확률을 남녀 각각에 대해 [그림 1-3]에 나타내었다. 연도별로 적합정도에 차이가 존재하나 대체적으로 90세 이상에 대해서 모형추정치가 과대 추정되는 경향이 있으며 90세 이전 연령에 대해서는 비교적 적절한 추정치를 제공하고 있음을 알 수 있다. Brass-Logit 모형은 2010년 잠정 완전생명표를 표준생명표로 하였기 때문에 2010년에 가까워지면서 생명표와 매우 근사한 결과를 제공하고 있다. 남자의 경우 고려한 모형 중 로지스틱 모형의 추정치가 생명표 결과와 가장 근사하며, 여자의 경우는 모형 추정치 결과에 뚜렷한 차이가 없다고 보여 진다.

모형별 적합력 비교를 위하여 2001~2010년의 추정된 로그사망확률과 생명표 로그사망확률의 RMSE와 MAPE를 <표 1-3>에 수록하였다. 남자의 경우 로지스틱 모형의 MAPE가 4%, RMSE가 0.083으로 고려한 모형 중 추정결과가 가장 우수하였다. 여자의 경우는 Brass-Logit 모형의 MAPE가 6.8%, RMSE가 0.147로 적합력이 가장 낮으며, 로지스틱모형과 CK모형의 적합력 차이는 크지 않다고 판단된다.

<표 1-3> 2001~2010년 75세 이상 생명표 로그사망확률에 대한 모형 적합력

모형	남자		여자	
	RMSE	MAPE	RMSE	MAPE
Logistic	0.083	0.040	0.138	0.060
Brass-Logit	0.117	0.064	0.147	0.068
CK	0.139	0.068	0.127	0.057

고연령 추정모형에 따른 기대수명 변화를 살펴보기 위하여 <표 1-4>에 각각 추정된 기대수명을 수록하였으며, 2010년 생명표 기대수명은 잠정치이다. 전반적으로 생명표보다 기대수명이 낮게 산출되는데, 이것은 90세 이상의 사망확률이 생명표 사망확률보다 다소 과대추정된 영향으로 판단된다. 각 모형별 MAE(mean absolute error)를 보면 남녀 모두 로지스틱 모형으로 추정 후 산출된 기대수명이 가장 생명표 결과와 근사함을 알

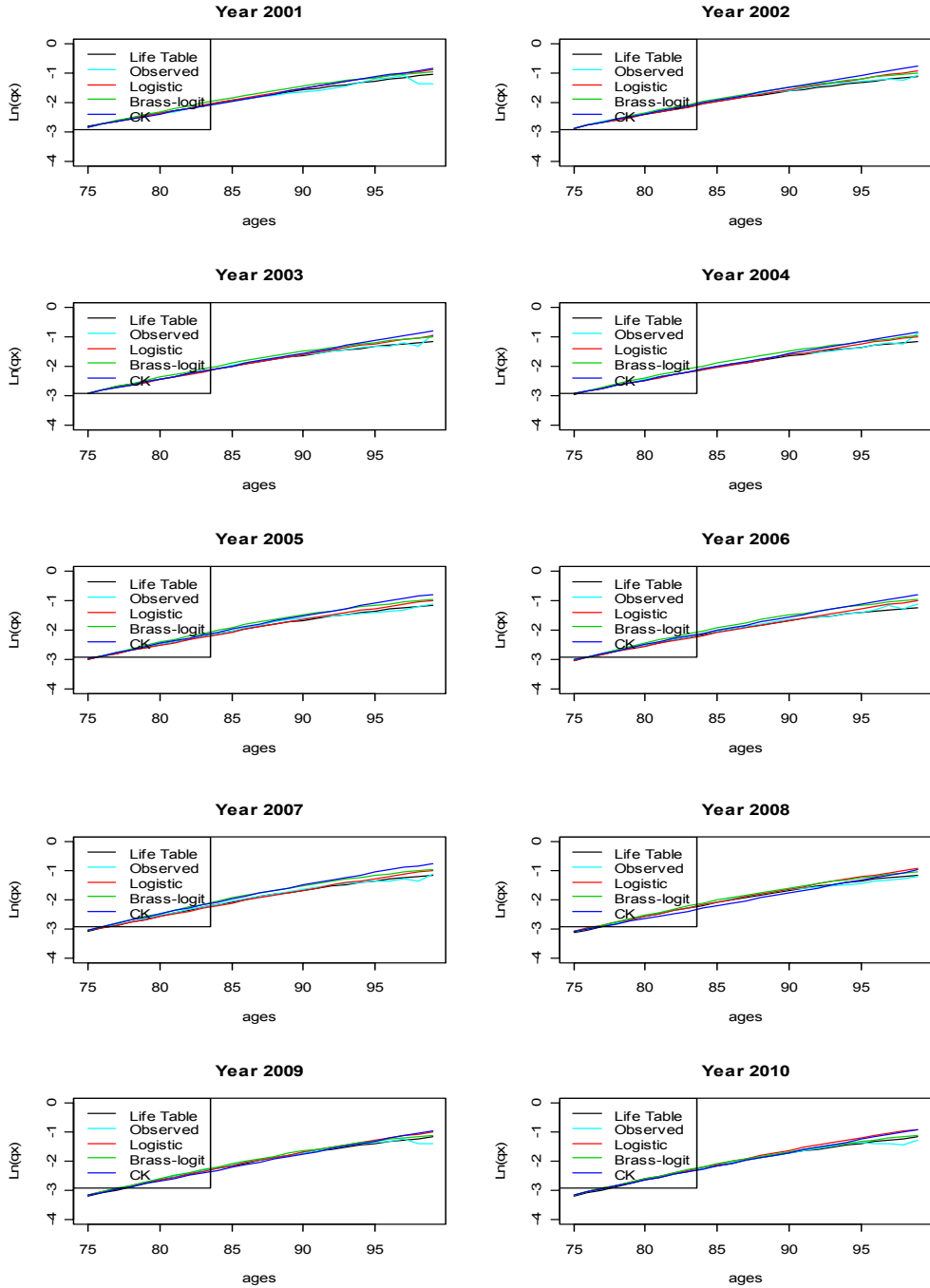
수 있다. 로지스틱 모형으로 추정된 경우 평균적으로 남자는 0.167세, 여자는 0.124세 정도 차이가 발생하며, Brass-Logit 모형으로 추정된 경우가 남자 0.342세, 여자 0.459세로 가장 크게 차이가 남을 알 수 있다.

〈표 1-4〉 2001~2010년 기대수명에 대한 모형비교

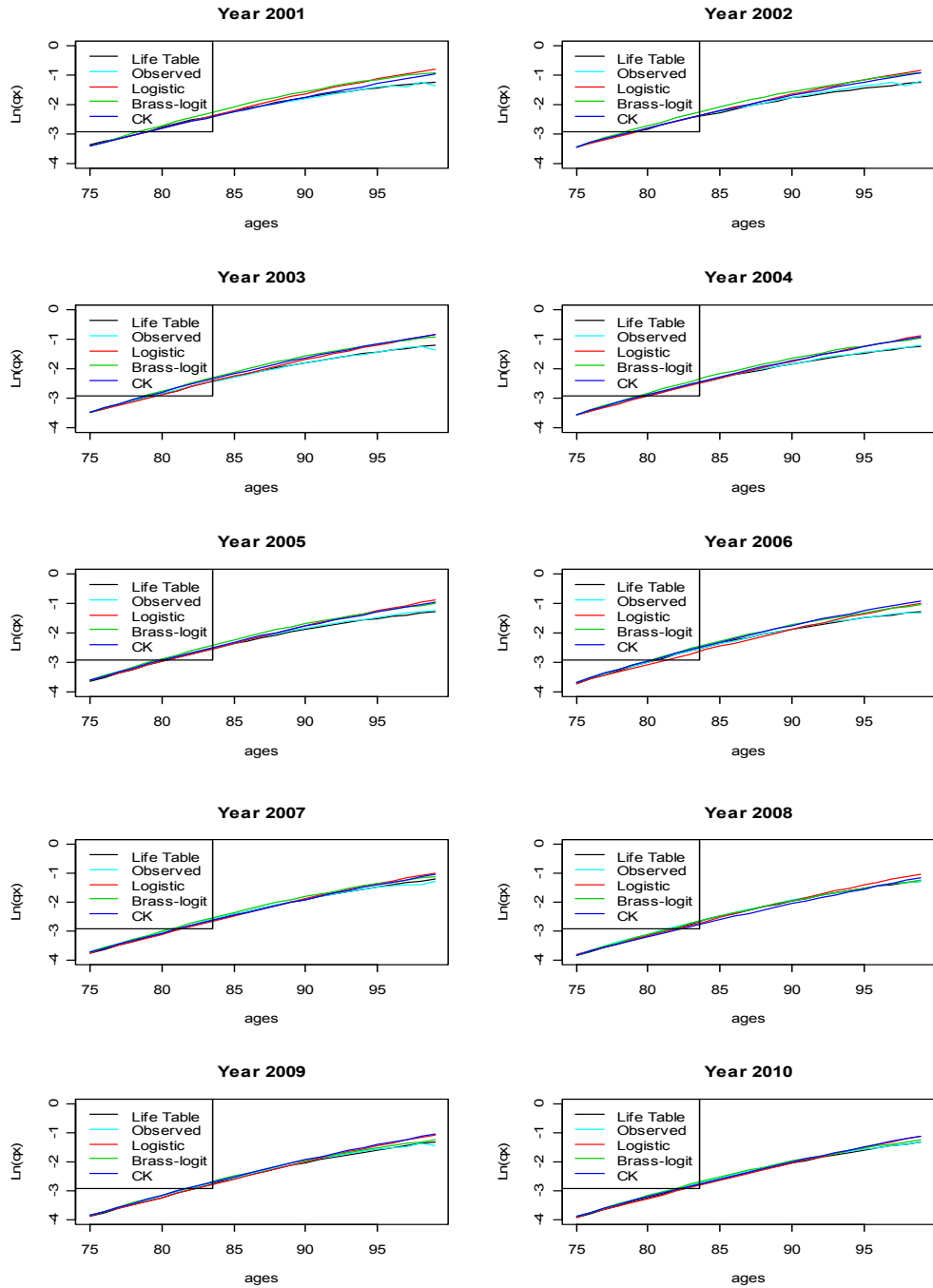
연도	남자				여자			
	생명표	Logistic	Brass-Logit	CK	생명표	Logistic	Brass-Logit	CK
2001	72.82	72.90	72.74	72.94	80.04	79.98	79.66	80.22
2002	73.40	73.46	73.30	73.34	80.45	80.27	79.88	80.25
2003	73.86	73.90	73.68	73.82	80.81	80.69	80.23	80.40
2004	74.51	74.37	74.10	74.30	81.35	81.27	80.83	81.10
2005	75.14	74.98	74.65	74.76	81.89	81.58	81.24	81.48
2006	75.74	75.60	75.23	75.38	82.36	82.47	81.88	81.90
2007	76.13	76.00	75.61	75.63	82.73	82.66	82.28	82.60
2008	76.54	76.24	76.12	76.57	83.29	83.18	83.15	83.55
2009	76.99	76.75	76.63	76.90	83.77	83.56	83.32	83.26
2010	77.18*	76.79	76.83	76.97	84.06*	84.07	83.70	83.82
MAE	-	0.167	0.342	0.199	-	0.124	0.459	0.304

\* 2010년 생명표는 잠정치

앞의 결과로부터 모형으로 추정된 로그 사망확률과 기대수명을 모두 고려하였을 때, 장래 사망확률 예측에 필요한 고연령 사망확률을 추정하기 위하여 로지스틱 모형을 이용하는 것이 적절하다고 판단된다. 그러나 본 연구에서는 고연령 모형에 따른 예측결과 차이 및 영향을 파악하기 위하여 2-모수 로지스틱 모형과 Brass-Logit 모형을 이용하여 추정된 결과 각각을 기초자료로 하여 사망확률 예측에 사용하기로 하였다.



[그림 1-3] 2001~2010년 75세 이상 로그사망확률 추정 결과 : 남자



[그림 1-3] 2001~2010년 75세 이상 로그사망확률 추정 결과 : 여자 (계속)



### 3. LC 모형을 이용한 사망력 예측(I)

#### 가. 개요

##### 1) 예측모형

사망력 예측모형은 LC모형을 기본모형으로 하고 추계가 시작되는 해의 탈선으로 인한 편의(jump-off bias)를 제거하기 위한 조정과 시간효과  $k_t$ 를 조정하는 것을 고려하였다. 원래의 LC모형은 식(11)을 이용하여 로그사망률에서의 오차를 최소화되도록  $k_t$ 를 재추정하며, 재추정을 위해서는 총사망자수와 연령별 인구수가 필요하다. 그러나 추정된 75세 이상 고연령에 대한 연령별 인구수가 없으므로, LM모형과 같이 기대수명을 이용하여  $k_t$ 를 재추정하였다. 본 연구에서는 아래의 3가지 모형을 고려하기로 하였다. [모형1]은 기본적인 LC모형이며, [모형2]는 LC모형에 추계를 시작하는 연도에서 발생하는 탈선으로 인한 편의를 조정한 모형이다. [모형3]은 식(11)의 LC모형에 대해 기대수명을 이용하여  $k_t$ 를 재추정한 것으로 사망확률을 이용하여 모형을 적합하기 때문에 탈선으로 인한 편의 조정이 이루어지게 된다.

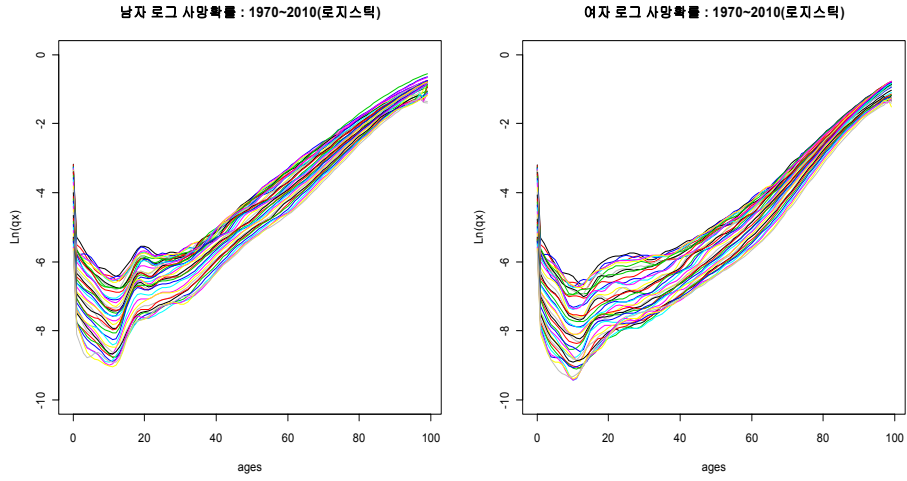
$$[\text{모형1}] \text{ LC모형} : \ln[m(x,t)] = a_x + b_x \cdot k_t$$

$$[\text{모형2}] \text{ LC모형} + \text{jump-off bias조정} : \ln[m(x,t)] = \ln[m(x,2010)] + b_x \cdot (k_t - k_{2010})$$

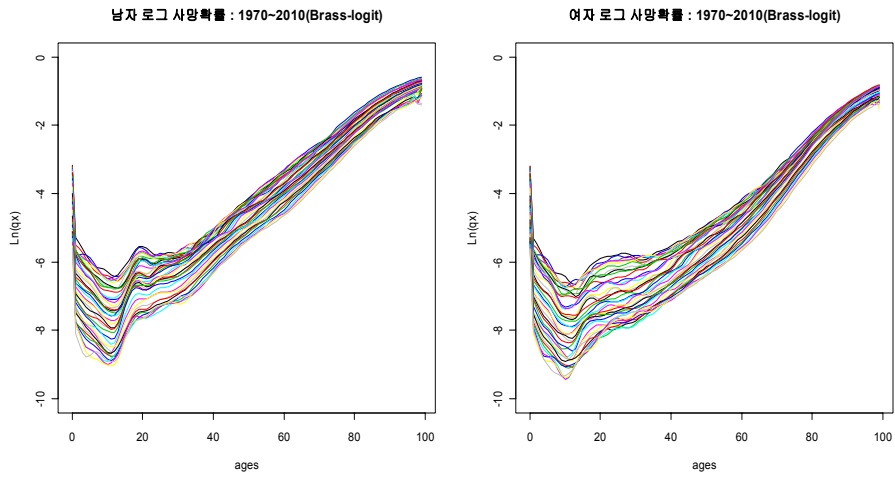
$$[\text{모형3}] \text{ LC모형} + \text{기대수명을 이용한 } k_t \text{ 조정}$$

##### 2) 자료

사망력 예측에 사용할 자료는 0세는 생명표 사망확률, 1~75세는 사망률을 사망확률로 변환하여 보정한 자료, 75세 이상 고연령에 대해서는 로지스틱 모형과 Brass-Logit 모형으로 추정된 자료로 구성된 0~100세 사망확률이다. [그림 1-4]에 1970~2010년의 연령별 남녀 로그사망확률 그림을 나타내었다.



(a) 고연령 모형 : 2-모수 로지스틱모형



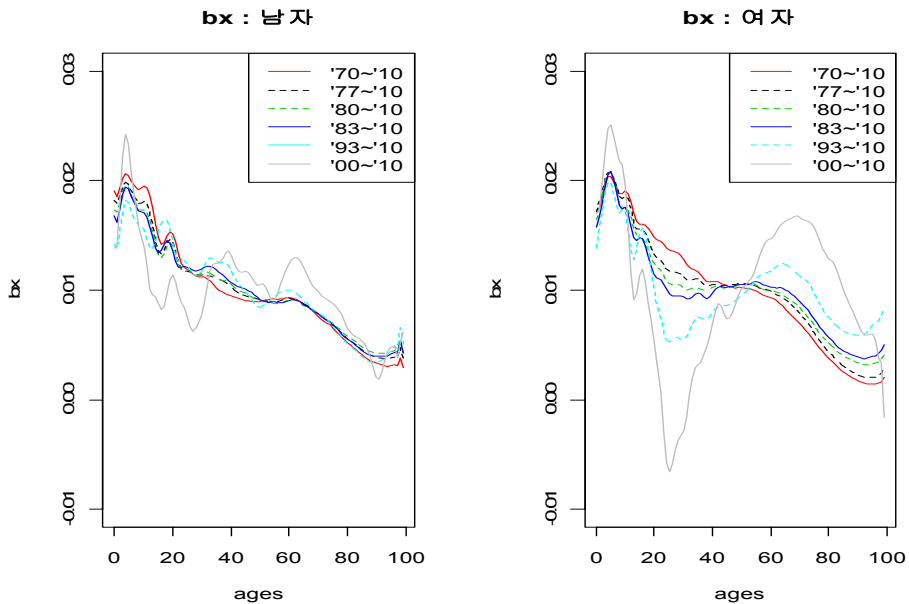
(b) 고연령 모형 : Brass-Logit모형

[그림 1-4] 1970~2010년 남녀별 로그사망확률



## 나. 사망확률 변화 패턴

사망확률 변화패턴을 살펴보기 위해 2-모수 로지스틱 모형으로 75세 이상을 추정한 0~100세 사망확률에 대해 LC모형을 이용하여 기간에 따른 연령별 사망확률 변화  $b_x$ 를 [그림 1-5]에 나타내었다. 여기서  $b_x$ 의 절대값이 크면  $x$ -연령의 사망확률의 변화가 크다는 것을 의미한다. [그림 1-5]에서 1970년부터 2010년까지 연령별 사망확률 변화패턴에서 남자의 경우 적합기간에 따라 크게 다르지 않으며, 여자의 경우는 적합기간에 따라 변화정도에 차이가 있음을 알 수 있다.



[그림 1-5] 기간에 따른 연령별 사망확률 변화

## 다. 모형 예측력

사망력 모형의 예측력을 파악하기 위하여 1970~2005년과 1983~2005년 기간에 대해 0~100세 사망확률을 이용하여 2006~2010년 남녀별 기대수명을 예측하고 생년표 기대수명 및 2005년 추계(1970~2005년 자료이용)결과와 비교하였다(<표 1-5>와 <표 1-6> 참조).

동일한 기간에 대해 고연령 추정 모형에 따른 평균수명 예측 결과에 큰 차이는 없으며, Brass-Logit 모형으로 추정한 경우가 약간 높게 예측되는 경향을 보였다. 대체적으로 고려한 모든 모형에서 생명표 기대수명보다 과소추정(여자의 경우가 과소추정 정도가 더 심함)되며 1970~2005년 자료를 이용한 경우가 예측력이 떨어지는 것을 알 수 있다. <표 1-5> 로지스틱 모형으로 고연령 사망확률을 추정한 경우에서 적합기간이 1970~2005년일 때, 생명표 기대수명과 평균적으로 남자는 0.53~0.98세, 여자는 0.68~2.33세 차이가 발생하였다. 적합기간이 1983~2005년인 경우 남자는 0.14~0.38세, 여자는 0.48~1.23세 차이가 발생하였다. <표 1-6> Brass-Logit 모형 예측결과에서 로지스틱 모형의 MAE와 큰 차이는 없으며, 비슷한 수준으로 나타났다.

5년 정도의 단기예측 측면에서 생명표 기대수명과 비교하였을 때 1983~2005년 자료를 사용하여 예측하는 것이 우수하며, 모형1보다는 모형2와 모형3이 우수하게 나타났다. 모형3의 경우는 기대수명을 이용하여 시간 모수  $k_t$ 를 조정하였기 때문에 단기예측에서 우수한 것으로 보여진다. 모형2는 추계가 시작되는 해에 발생하는 생명표 기대수명과의 차이를 조정한 모형으로 모형1보다는 생명표에 근사한 값을 제공해 주고 있음을 볼 수 있다.

<표 1-5> 2006~2010년 생명표 기대수명과 예측 기대수명 비교(로지스틱 모형)

(a) 적합기간 : 1970~2005년

1970 ~ 2005	남자					여자				
	생명표	2005 추계	모형1	모형2	모형3	생명표	2005 추계	모형1	모형2	모형3
2006	75.74	75.34	74.94	75.27	75.37	82.36	82.09	80.51	82.16	81.97
2007	76.13	75.55	75.22	75.54	75.66	82.73	82.29	80.72	82.37	82.19
2008	76.54	75.75	75.53	75.85	75.98	83.29	82.49	80.91	82.56	82.38
2009	76.99	75.95	75.85	76.17	76.31	83.77	82.68	81.11	82.76	82.58
2010	77.18*	76.15	76.14	76.45	76.61	84.06*	82.88	81.30	82.95	82.78
MAE	-	0.77	0.98	0.66	0.53	-	0.76	2.33	0.68	0.86





(b) 적합기간 : 1983~2005년

1970 ~ 2005	남자					여자				
	생명표	2005 추계	모형1	모형2	모형3	생명표	2005 추계	모형1	모형2	모형3
2006	75.74	75.34	75.33	75.37	75.56	82.36	82.09	81.41	82.04	82.16
2007	76.13	75.55	75.74	75.78	75.97	82.73	82.29	81.73	82.36	82.47
2008	76.54	75.75	76.14	76.18	76.38	83.29	82.49	82.01	82.65	82.77
2009	76.99	75.95	76.54	76.57	76.78	83.77	82.68	82.31	82.95	83.07
2010	77.18*	76.15	76.92	76.96	77.17	84.06*	82.88	82.60	83.23	83.36
MAE	-	0.77	0.38	0.34	0.14	-	0.76	1.23	0.60	0.48

〈표 1-6〉 2006~2010년 생명표 기대수명과 예측 기대수명 비교(Brass-Logit 모형)

(a) 적합기간 : 1970~2005년

1970 ~ 2005	남자					여자				
	생명표	2005 추계	모형1	모형2	모형3	생명표	2005 추계	모형1	모형2	모형3
2006	75.74	75.34	74.96	75.28	75.41	82.36	82.09	80.53	81.94	81.97
2007	76.13	75.55	75.25	75.56	75.71	82.73	82.29	80.74	82.15	82.19
2008	76.54	75.75	75.57	75.88	76.04	83.29	82.49	80.93	82.34	82.38
2009	76.99	75.95	75.89	76.21	76.38	83.77	82.68	81.12	82.54	82.58
2010	77.18*	76.15	76.20	76.51	76.69	84.06*	82.88	81.31	82.73	82.78
MAE	-	0.77	0.94	0.63	0.47	-	0.76	2.32	0.90	0.86

(b) 적합기간 : 1983~2005년

1970 ~ 2005	남자					여자				
	생명표	2005 추계	모형1	모형2	모형3	생명표	2005 추계	모형1	모형2	모형3
2006	75.74	75.34	75.31	75.38	75.55	82.36	82.09	81.37	82.06	82.17
2007	76.13	75.55	75.70	75.78	75.95	82.73	82.29	81.70	82.40	82.51
2008	76.54	75.75	76.13	76.20	76.39	83.29	82.49	82.01	82.72	82.82
2009	76.99	75.95	76.56	76.63	76.83	83.77	82.68	82.33	83.04	83.14
2010	77.18*	76.15	76.97	77.03	77.24	84.06*	82.88	82.64	83.35	83.45
MAE	-	0.77	0.38	0.31	0.15	-	0.76	1.23	0.53	0.42

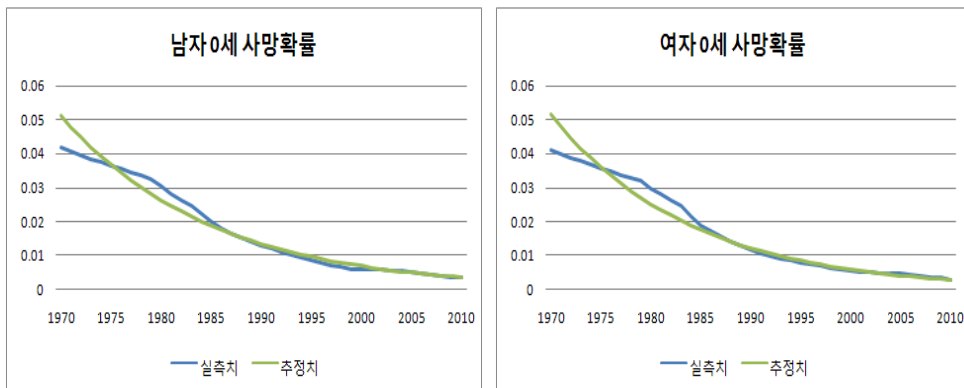
## 라. 영아 사망확률

영아 사망은 출생 후 1년 미만에 사망하는 경우를 의미하는 것으로 사산, 선천적 원인 등으로 인한 사망이 포함되어 다른 연령층의 사망확률 감소 패턴과는 차이가 나게 된다. 영아사망확률은 과거 영아 사망수준이 매우 높았던 경우에 생명표 기대수명에 상당한 영향을 주었으며, 이미 우리나라의 영아 사망률은 의료기술의 발전과 함께 매우 빠른 속도로 개선을 이루어 이미 매우 낮은 수준에 와 있는 현 단계에서는 영아사망확률이 기대수명에 큰 영향을 미치지 못하는 못한다. 그러나 다른 연령층과 같은 속도로 사망확률이 감소하지 않을 것이므로 모형을 달리하여 예측할 필요가 있다. 이러한 이유로 2005년 인구추계 사망력 예측에서 0세에 대해 일본의 2050년 예측결과로 수렴한다는 가정을 하고 사망확률을 예측하였으며, 김기환 외(2006)는 0세 사망률 예측을 위하여 2050년 일본의 0세 사망확률을 이용하여 모형을 적합하고 예측하였다. [그림 1-6]에서 남녀 0세 사망확률 시계열(실측치)은 1970년대에는 서서히 감소하다가 1980년대와 1990년대에 매우 급격하게 감소하는 패턴을 보이며 2000년대에 접어들면서 감소추세가 줄어들고 있다. 1970~2010년 0세 사망확률 시계열에 대해 고펜퍼츠(Gompertz) 모형을 적합한 결과는 아래와 같으며, 실측치와 모형에 의한 추정치를 [그림 1-6]에 나타내었다.

$$\text{(남자)} \ln q_t = -2.90431 - 0.06715 \cdot t$$

$$\text{(여자)} \ln q_t = -2.89456 - 0.07160 \cdot t \tag{16}$$

여기서  $q_t$ 는 0세 사망확률이며,  $t = 1, 2, \dots, 41$  이다.



[그림 1-6] 1970~2010년 영아 사망확률 : 실측치는 생명표 영아 사망확률, 추정치는 고펜퍼츠 모형식(16)에 의해 추정된 영아 사망확률



앞에서 살펴본 바와 같이 0세 사망확률 모형으로 고펜페르츠 모형이 적절한 것으로 보여지나, 적합된 모형식(16)의 지수 감소율이 향후에도 계속 유지될 것으로 보기에 어려움이 있다고 판단된다. 또한 모형식(16)을 이용하여 향후 50년을 예측하는 경우 과소 예측(LC모형의 예측결과와 비슷한 수준)되어지는 것을 확인할 수 있다. 0세 사망확률 변화정도를 알아보기 위하여 10년 단위로 고펜페르츠 모형을 적합하고 그 결과를 <표 1-7>에 수록하였다. 남녀 모두 지수 감소율( $\beta$ ) 추정치를 볼 때, 1970년대에는 서서히 감소하다가 1980년대와 1990년대에 가파르게 감소한 후 2000년대에 접어들면서 그 감소 속도가 줄어들고 있음을 알 수 있으며, 향후 장기적으로 0세 사망확률의 감소 속도는 지금의 감소율보다 더욱 더 느려질 것으로 예상할 수 있다. 본 연구에서는 향후 감소속도에 대해 일본의 2005년 추계결과(2005~2055년)의 0세 사망확률을 이용하여 추정한 결과 남자는 -0.01275, 여자는 -0.01168로 나타났으며, 이를 이용하여 다음과 같이 우리나라 0세 사망확률에 대한 예측모형을 추정하였다.

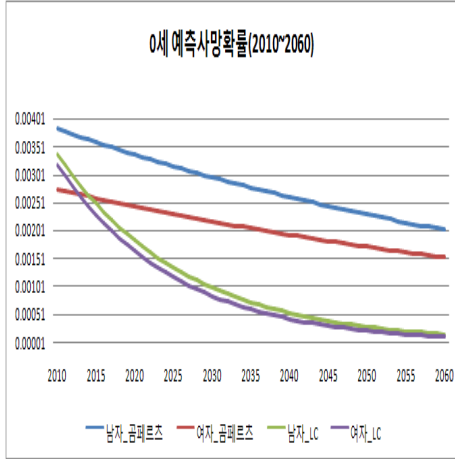
$$(\text{남자}) \ln q_t = -5.55397 - 0.01275 \cdot t$$

$$(\text{여자}) \ln q_t = -5.88608 - 0.01168 \cdot t \quad (17)$$

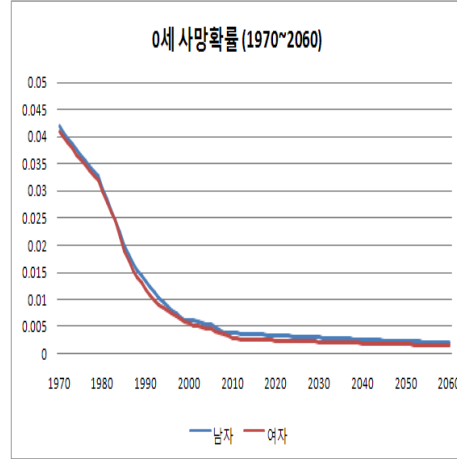
[그림 1-7] (a)는 2010~2010년 0세 예측사망확률 그림으로 적합된 고펜페르츠 모형식(17)과 모형1의 LC모형을 이용한 결과이다. LC모형을 이용한 경우 모형식(17)을 이용한 결과보다 과소 예측됨을 알 수 있다. [그림 1-7] (b)는 2010년 이전은 실측치이며, 2010년 이후는 고펜페르츠 모형을 이용한 예측치에 대한 시계열이다. 본 연구에서는 0세 사망확률 예측치로 고펜페르츠 모형식(17)을 이용하여 예측된 사망확률을 사용하였다.

<표 1-7> 고펜페르츠(Gompertz)모형 적합 결과

기간	남자		여자	
	$\alpha$	$\beta$	$\alpha$	$\beta$
1970~1979	-3.145	-0.028	-3.165	-0.028
1980~1989	-3.382	-0.087	-3.371	-0.098
1990~1999	-4.230	-0.085	-4.390	-0.074
2000~2010	-4.954	-0.058	-4.441	-0.064



(a) 예측치(LC모형과 곰페르츠 모형), 2010~2060년



(b) 실적치와 곰페르츠 모형에 의한 예측치, 1970~2060년

[그림 1-7] 영아 사망확률

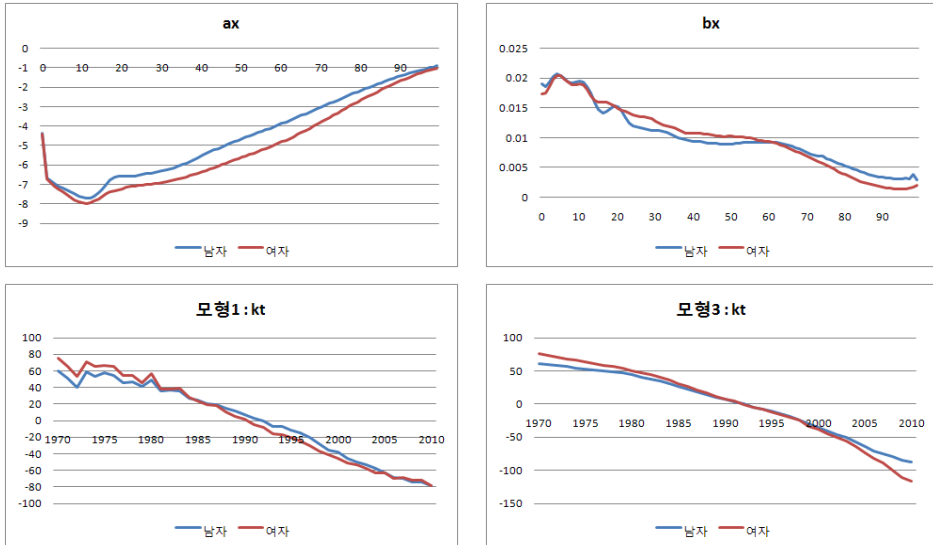
## 4. LC 모형을 이용한 사망력 예측 (II)

### 가. 모형 적합 및 예측 결과 : 1970~2010년

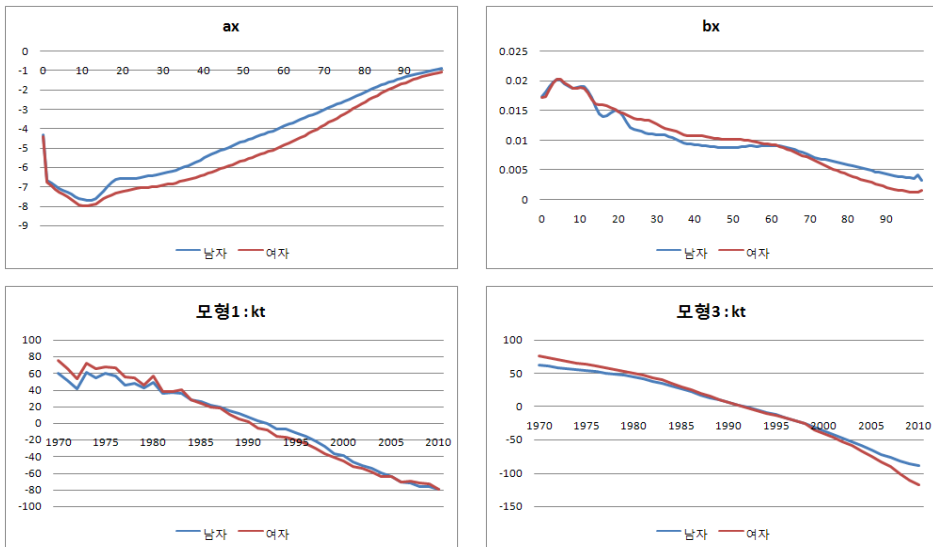
본 절은 1970~2010년 0~100세 사망확률 자료에 대한 모형 적합 및 예측 결과를 정리하였다.

#### 1) 모형적합 결과

고려한 LC모형에서  $a_x, b_x$ 는 연령에 대해 상수모수이며,  $k_t$ 는 시간에 따른 사망변화 수준을 나타내는 모수이다. 각 모수 추정 결과를 [그림 1-8]에 나타내었으며, 고연령 모형에 따른 추정결과 차이는 거의 없는 것으로 보인다. 연령별 사망확률 변화  $b_x$ 를 살펴보면 60세를 전후로 하여 남자와 여자의 변화량이 달라지고 있음을 알 수 있다. 즉, 60세 이전은 여자의 변화량이 남자의 변화량보다 크지만 60세 이후부터는 여자의 변화량이 남자의 변화량보다 작아지고 있다. 모형3의 경우는 기대수명을 이용하여 재조정된  $k_t$ 이다.



(a) 고연령 모형 : 로지스틱모형



(b) 고연령 모형 : Brass-Logit모형

[그림 1-8] LC 모형  $a_x$ ,  $b_x$ ,  $k_t$  추정 결과 : 1970~2010년

## 2) 시간모수 $k_t$

사망확률 예측을 위해서는 추정된 시간모수  $k_t$ 에 대한 예측모형이 필요하며, 예측 결과에 많은 영향을 미치게 된다.  $k_t$ 모형으로 기본적으로 추세가 있는 랜덤워크(Random Walk with Draft) 모형을 적합하며, ARIMA모형을 적합할 수도 있다. 일본의 경우는 2001~2050년 장래인구추계(Shigesato Takahashi et al., 2003)에서  $k_t$ 에 대하여 식(18)과 같은 반-대수함수를 적합하여  $k_t$ 를 예측하였는데, 이것은 랜덤워크 모형을 적합하였을 때, 사망확률 변화 수준이 시간에 대해 선형적으로 감소하는 것에 대하여 지수적 감소 측면을 고려하고자 한 것으로 보여 진다.

$$k_t = \beta_1 + \beta_2 \log(t + \alpha) \quad (18)$$

시간모수 예측모형으로 추세를 갖는 랜덤워크 모형과 반-대수모형을 고려하고 예측 결과를 비교하였다. 모형1과 모형3의  $k_t$ 에 대하여 랜덤워크 모형을 적합하였으며 반-대수모형은 모형1의  $k_t$ 에 대해서만 적합하였다. 반-대수 모형의 모수는  $\alpha$ 를 변화시켜 가면서  $\beta_1$ 과  $\beta_2$ 를 추정하였으며, 각 모형에 대한 추정결과는 각각 <표 1-8>과 <표 1-9>에 수록하였다. 기대수명을 이용하여 재조정된  $k_t$ 의 추세모수는 조정하지 않은 경우보다 절대값이 크므로 모형3은 사망확률의 감소 속도를 더 빠르게 만들어 준다는 것을 알 수 있다.

<표 1-8> 추세를 갖는 랜덤워크모형 모수 추정치

RWD	Logistic		Brass-Logit	
	남자	여자	남자	여자
모형1,2	-3.43783	-3.84341	-3.48505	-3.85710
모형3	-3.71471	-4.81966	-3.77554	-4.84503



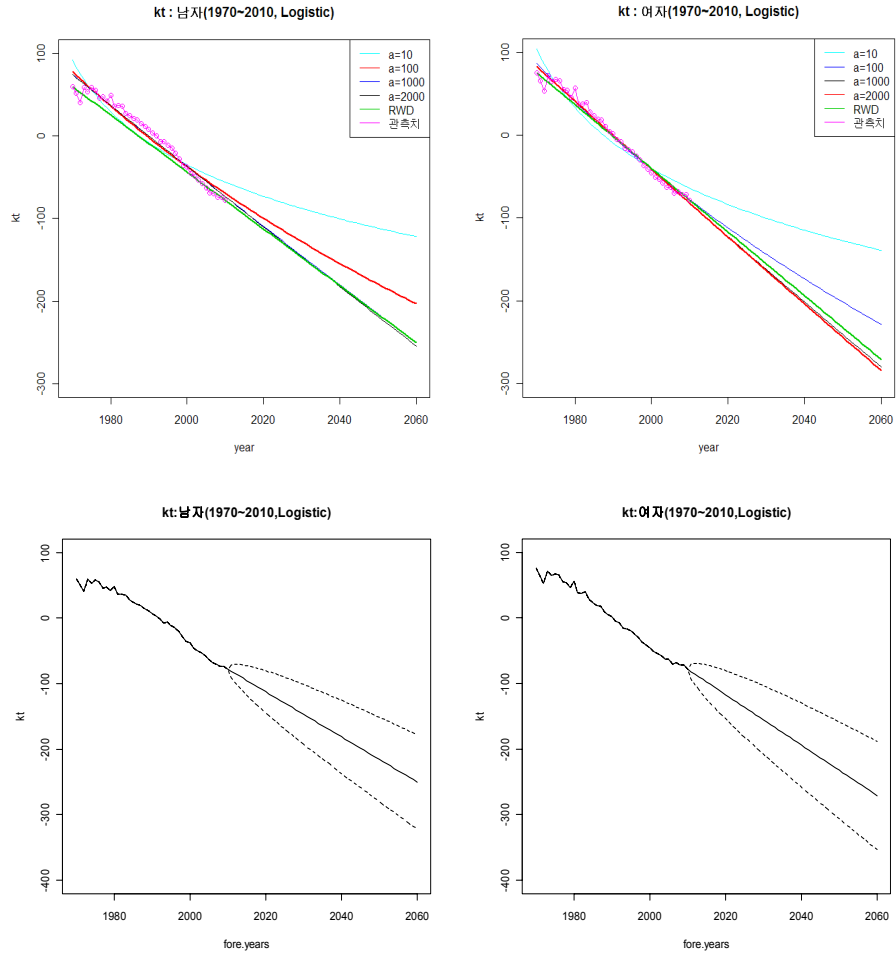
〈표 1-9〉 반-대수모형 모수 추정치

반-대수	남자			여자		
	$\beta_1$	$\beta_2$	$\alpha$	$\beta_1$	$\beta_2$	$\alpha$
모형1,2_Logistic	2108.516	-440.102	100	2367.053	-494.066	100
	26083.556	-3764.692	1000	29148.085	-4207.001	1000
	56745.869	-7455.447	2000	63393.596	-8328.845	2000
모형1,2_Brass-Logit	2146.382	-448.006	100	2380.353	-496.842	100
	26551.462	-3832.225	1000	29316.118	-4231.254	1000
	57763.733	-7589.177	2000	63759.645	-8376.938	2000

〈표 1-10〉에 모형1의  $k_t$ 에 대한 랜덤워크모형과 반-대수모형의 적합력을 살펴보기 위하여 RMSE를 수록하였다. 고연령 모형에 따른 차이는 크지 않으며, 남녀 모두 반-대수모형의  $\alpha$  값이 커질수록 RMSE 값이 감소하는 것으로 나타났다. [그림 1-9]에  $k_t$ 에 대하여 반-대수 모형과 랜덤워크 모형을 적합한 경우의 예측치 그리고 랜덤워크 모형의 95% 신뢰구간을 나타내었다. 시차가 커질수록 두 모형의  $k_t$ 에 대한 예측결과 차이가 커지며, 반-대수 모형의  $\alpha$ 가 1000이상이 되면 랜덤워크 모형과 비슷한 수준으로,  $\alpha$ 가 작아지면 감소속도가 느려져 높게 예측되어진다. 이와 같이 예측모형에 따라 예측결과가 다르게 산출되며, 랜덤워크 모형의 경우 미래의  $k_t$  예측에 대한 불확실성을 [그림 1-9]와 같이 신뢰구간으로 표현할 수 있다.

〈표 1-10〉  $k_t$ (모형1)에 대한 적합력 : 1970~2010년

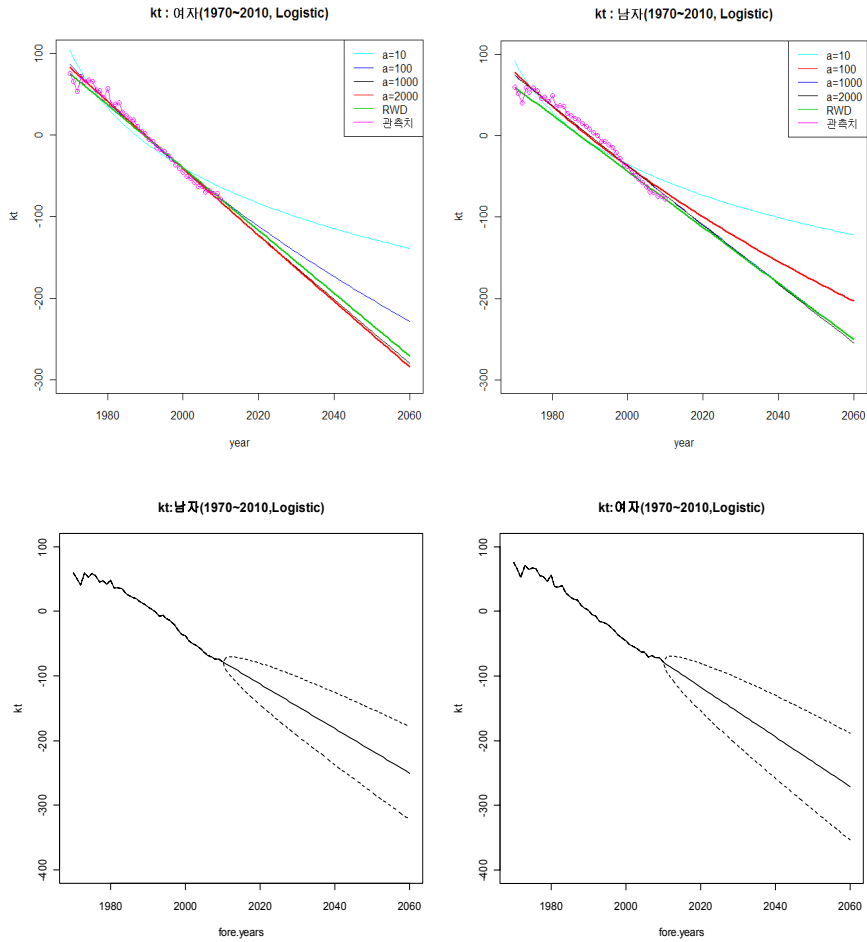
RMSE	Logistic					Brass-Logit				
	RWD	반-대수( $\alpha$ )				RWD	반-대수( $\alpha$ )			
		10	100	1000	2000		10	100	1000	2000
남자	12.643	15.928	9.728	8.256	8.163	12.901	16.220	9.915	8.423	8.329
여자	6.970	13.174	6.905	6.014	5.977	7.289	13.452	7.132	6.207	6.166



(a) 고연령 모형 : 로지스틱모형

[그림 1-9] 모형1의  $k_t$  추정치, 랜덤워크모형과 반-대수모형 예측치,  
 랜덤워크모형의 95% 신뢰구간 : 1970~2010년





(b) 고연령 모형 : Brass-Logit모형

[그림 1-9] 모형1의  $k_t$  추정치, 랜덤워크모형과 반-대수모형 예측치, 랜덤워크모형의 95% 신뢰구간: 1970~2010년 (계속)

### 3) 모형별 적합력과 장래 기대수명

1970~2010년 자료를 이용하여 검토한 모형에 대해 모형별 사망확률 적합력과 장래 기대수명을 비교하였다. 반-대수 모형의 경우 [그림 1-9]에서 살펴본 바와 같이  $\alpha$ 에 따라 미래의  $k_t$ 에 영향을 주게 되며,  $\alpha$ 를 남녀 각각 100과 2000으로 두고 적합하고 예측한 결과를 제시하였다. 이 경우 미래의 시간변화수준  $k_t$ 는 여자가 남자보다 더 낮아지며, 그 차이는 랜덤워크 모형보다 커지게 된다. 1970~2010년에 대하여 각 모형별 추정된 사망확률과 관측 사망확률의 MAPE(mean absolute percent error)를 <표 1-11>에 수록하였으며, 남자는 6~13%, 여자는 9~20% 정도로 나타났다. 모형1에 대해 랜덤워크 모형을 적용한 경우 남자 6%, 여자 9%로 1970~2010년 관측 사망확률에 대한 적합력은 가장 우수한 것으로 나타났다.

<표 1-12>에 각 모형에 따라 산출된 남녀별 기대수명 결과를 수록하였다. 반-대수 모형은 남자의 경우 랜덤워크 모형으로 예측한 결과보다 반-대수모형이 낮은 기대수명 예측치를 주며, 여자의 경우 높은 기대수명 예측치를 준다. 이것은 남녀의 시간에 대한 장래 사망수준을 반-대수 모형을 이용하여 조절하였기 때문으로 당연한 결과로 볼 수 있다. 고연령을 로지스틱 모형으로 추정할 경우에 모형3과 랜덤워크모형을 적용하였을 때, 2060년 남자 88.26세, 여자 90.70세로 가장 높게 예측되었다. <표 1-13>에서 남녀 간 기대수명 차이는 모형에 따라 차이가 있으며 모형별로 2060년에 약 1세에서 5세 정도의 차이를 보이는 것으로 나타났다.

<표 1-11> 모형별 추정사망확률의 관측사망확률에 대한 적합력 : 1970~2010년

MAPE	Logistic					Brass-Logit				
	RDW			반-대수		RDW			반-대수	
	모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
남자	0.059	0.096	0.068	0.106	0.132	0.059	0.104	0.069	0.107	0.132
여자	0.089	0.187	0.124	0.105	0.198	0.091	0.201	0.121	0.108	0.199



〈표 1-12〉 장래기대수명

(a) 남자

연도	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	77.18	76.15	76.39	77.00	77.18	75.63	76.25	76.37	76.91	77.18	75.61	76.16
2015		77.11	77.70	78.29	78.52	76.88	77.48	77.77	78.34	78.62	76.95	77.53
2020		78.04	79.01	79.58	79.87	78.06	78.64	79.14	79.70	80.06	78.17	78.74
2025		78.93	80.24	80.80	81.14	79.14	79.71	80.45	80.99	81.43	79.31	79.87
2030		79.79	81.40	81.94	82.33	80.13	80.69	81.70	82.22	82.73	80.37	80.92
2035		80.60	82.49	83.03	83.46	81.05	81.61	82.89	83.39	83.96	81.36	81.89
2040		81.39	83.53	84.05	84.53	81.91	82.45	84.03	84.52	85.15	82.29	82.81
2045		82.15	84.51	85.02	85.53	82.71	83.24	85.12	85.59	86.28	83.16	83.67
2050		82.87	85.45	85.94	86.49	83.45	83.97	86.17	86.62	87.36	83.99	84.48
2055			86.34	86.82	87.40	84.15	84.66	87.17	87.61	88.39	84.76	85.24
2060			87.19	87.66	88.26	84.81	85.31	88.13	88.55	89.38	85.49	85.96

(b) 여자

여자	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	84.06	82.88	82.16	83.78	84.06	82.37	83.99	82.07	83.78	84.06	82.27	83.98
2015		83.80	83.14	84.73	84.94	83.37	84.95	83.08	84.77	84.98	83.31	84.99
2020		84.68	84.00	85.59	85.81	84.25	85.83	83.97	85.66	85.89	84.23	85.91
2025		85.5	84.79	86.36	86.60	85.05	86.62	84.79	86.47	86.73	85.07	86.74
2030		86.27	85.51	87.07	87.32	85.79	87.34	85.54	87.21	87.49	85.84	87.50
2035		86.99	86.17	87.71	87.98	86.46	88.00	86.24	87.90	88.20	86.55	88.20
2040		87.67	86.78	88.31	88.60	87.08	88.60	86.89	88.53	88.86	87.21	88.84
2045		88.31	87.35	88.87	89.17	87.65	89.16	87.50	89.13	89.47	87.82	89.44
2050		88.92	87.88	89.38	89.71	88.19	89.68	88.07	89.68	90.05	88.39	89.99
2055			88.39	89.87	90.22	88.69	90.16	88.60	90.20	90.59	88.93	90.51
2060			88.86	90.33	90.70	89.16	90.62	89.11	90.68	91.10	89.44	91.00

〈표 1-13〉 남자와 여자의 장래기대수명 차이

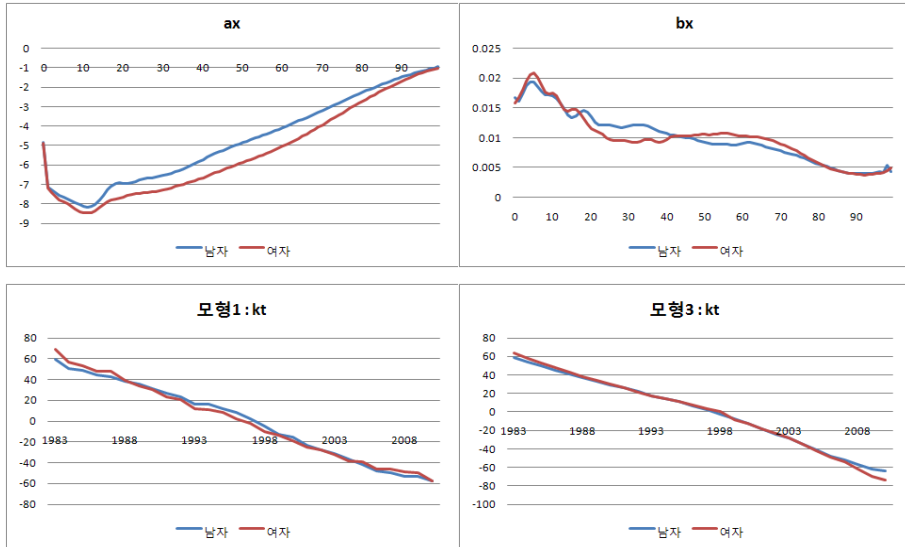
연도	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	6.88	6.73	5.77	6.79	6.88	6.73	7.74	5.71	6.87	6.88	6.66	7.82
2015		6.69	5.44	6.44	6.42	6.48	7.48	5.31	6.42	6.35	6.36	7.47
2020		6.64	5.00	6.00	5.94	6.19	7.19	4.83	5.96	5.83	6.05	7.17
2025		6.57	4.55	5.56	5.46	5.92	6.91	4.34	5.48	5.30	5.75	6.88
2030		6.48	4.11	5.12	4.98	5.65	6.65	3.85	5.00	4.77	5.46	6.59
2035		6.39	3.68	4.69	4.52	5.40	6.39	3.35	4.51	4.23	5.18	6.31
2040		6.28	3.25	4.26	4.07	5.17	6.15	2.86	4.02	3.71	4.91	6.03
2045		6.16	2.84	3.85	3.64	4.95	5.92	2.38	3.53	3.19	4.65	5.77
2050		6.05	2.44	3.44	3.22	4.74	5.70	1.90	3.06	2.69	4.41	5.52
2055			2.05	3.05	2.82	4.54	5.50	1.43	2.59	2.20	4.17	5.27
2060			1.67	2.67	2.44	4.36	5.31	0.98	2.13	1.73	3.95	5.04

## 나. 모형 적합 및 예측 결과 : 1983~2010년

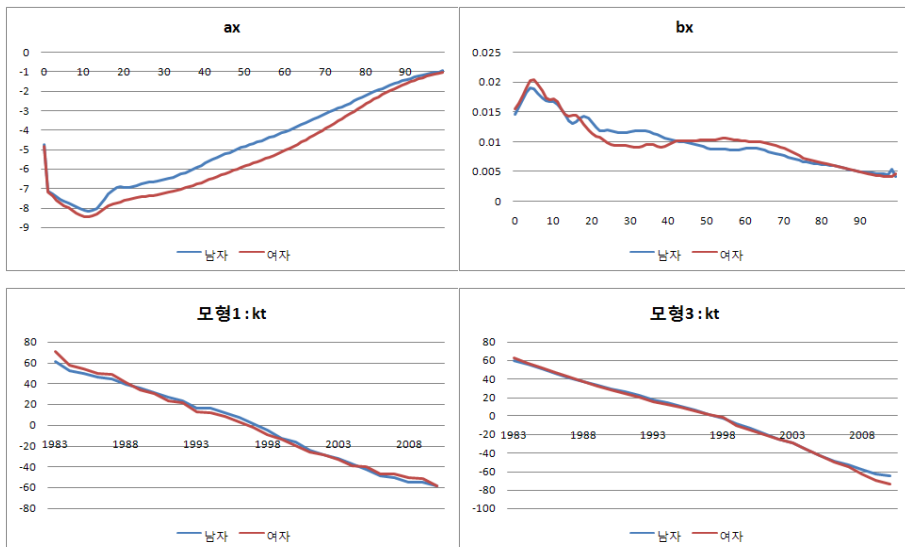
본 절은 1983~2010년 0~100세 사망확률 자료에 대한 모형 적합 및 예측 결과를 정리하였다.

### 1) 모형적합 결과

고려한 LC 모형에서 모수  $a_x$ ,  $b_x$ ,  $k_t$ 에 대한 추정 결과를 [그림 1-10]에 나타내었으며, 고연령 모형에 따른 추정결과 차이는 거의 없는 것으로 판단된다. 연령별 사망확률 변화를 나타내는 모수  $b_x$ 는 1970~2010년 자료를 사용하여 추정된 패턴과 다름을 알 수 있는데, 20~30대에서는 남자의 사망확률 변화량이 크고, 40~70대에서는 여자의 사망확률 변화량이 크다가, 80대 이후에는 사망확률 변화량이 거의 같아지는 패턴을 보이고 있다. 시간모수  $k_t$ 는 남녀간 차이가 크지 않는 것으로 보여진다.



(a) 고연령 모형 : 로지스틱모형



(b) 고연령 모형 : Brass-Logit모형

[그림 1-10] LC 모형  $a_x$ ,  $b_x$ ,  $k_t$  추정 결과 : 1983~2010년

## 2) 시간모수 $k_t$

시간모수 예측 모형으로 1970~2010년과 동일하게 추세가 있는 랜덤워크(Random Walk with Draft) 모형과 반-대수모형을 고려하였다. 각 추정결과는 각각 <표 1-14>와 <표 1-15>에 수록하였다. 기대수명을 이용하여 재조정된  $k_t$ 의 추세모수는 조정하지 않은 경우보다 큰 절대값을 가지므로 모형3은 사망확률의 감소속도 더 빠르게 만들어 준다는 것을 알 수 있다.

<표 1-14> 추세를 갖는 랜덤워크 모형 추정 모수

RWD	Logistic		Brass-Logit	
	남자	여자	남자	여자
모형1,2	-4.34193	-4.68362	-4.40710	-4.76684
모형3	-4.54387	-5.08694	-4.60552	-5.04779

<표 1-15> 반-대수 모형에 대한 추정 모수

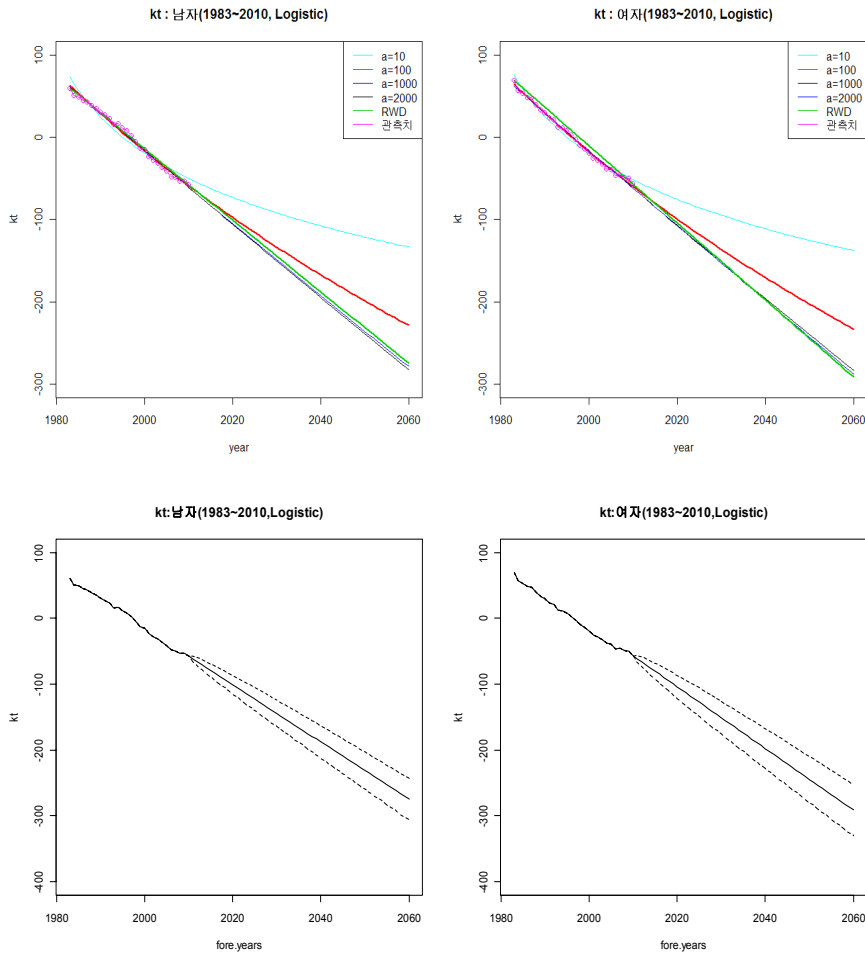
	남자			여자		
	$\beta_1$	$\beta_2$	$\alpha$	$\beta_1$	$\beta_2$	$\alpha$
모형1,2_Logistic	2438.457	-514.651	100	2487.555	-525.014	100
	31718.626	-4582.213	1000	32275.141	-4662.610	1000
	69232.177	-9099.777	2000	70435.743	-9257.972	2000
모형1,2_Brass-Logit	2481.885	-523.817	100	2533.318	-534.672	100
	32280.501	-4663.384	1000	32872.708	-4748.937	1000
	70458.169	-9260.919	2000	71740.372	-9429.450	2000

모형1에 의해 추정된  $k_t$ 에 대하여 고려한 랜덤워크 모형과 반-대수 모형에 따른 RMSE 결과를 <표 1-16>에 수록하였으며, 여자의 경우  $\alpha$ 가 100인 경우에 RMSE가 가장 작으며, 남자의 경우는 RMSE가 계속 감소하고 있다. [그림 1-11]에 반-대수 모형과 랜덤워크 모형을 적합한 경우의  $k_t$  추정치와 예측치 그리고  $k_t$ 에 대한 95% 신뢰구간을 나타내었으며, 반-대수모형의 경우  $\alpha$ 가 100인 경우 시차가 커질수록 감소속도가 느려져,  $k_t$ 가 랜덤워크 모형보다 높게 예측되어진다.



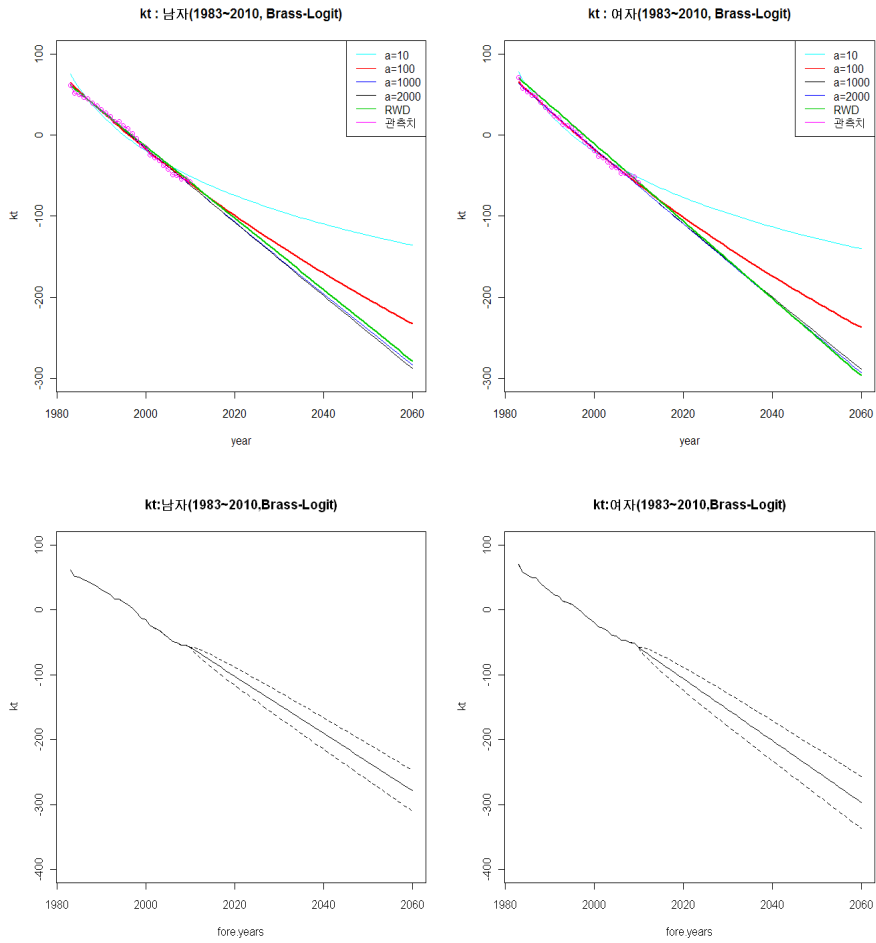
<표 1-16>  $k_t$ (모형1)에 대한 적합력 : 1983~2010년

RMSE	Logistic					Brass-Logit				
	RWD	반-대수				RWD	반-대수			
		10	100	1000	2000		10	100	1000	2000
남자	3.532	7.532	3.674	3.118	3.093	3.672	7.784	3.612	3.065	3.041
여자	7.092	4.670	2.432	3.001	3.046	6.858	4.923	2.514	3.033	3.075



(a) 고연령 모형 : 로지스틱모형

[그림 1-11] 모형1의  $k_t$  추정치, 랜덤워크모형과 반-대수모형 예측치, 랜덤워크모형의 95% 신뢰구간 : 1983~2010년



(b) 고연령 모형 : Brass-Logit모형

[그림 1-11] 모형1의  $k_t$  추정치, 랜덤워크모형과 반-대수모형 예측치,  
 랜덤워크모형의 95% 신뢰구간 : 1983~2010년(계속)





### 3) 모형별 적합력과 장래 기대수명

1983~2010년 자료를 이용하여 검토한 모형에 대한 모형별 사망확률 적합력과 장래 기대수명을 비교하였다. <표 1-17>은 각 모형별 추정된 1983~2010년 사망확률과 관측 사망확률의 MAPE 값을 수록하였다. 여자 모형2에 반-대수 모형을 적용한 경우를 제외하고 MAPE가 10%이하로 적합력이 나쁘지 않음을 알 수 있다. 시간모수를 랜덤워크모형으로 한 경우 남녀 모두 모형1이 가장 적합력이 좋으며, 반-대수 모형의 경우도 모형1이 적합력이 좋은 것을 알 수 있다. 모형 적합력에서는 랜덤워크 모형이 더 작은 MAPE 값을 주고 있다.

1983~2010년 자료에 대해 고연령 추정 모형과 LC모형 그리고 시간모수  $k_t$  예측모형을 고려하여 산출된 남녀별 기대수명 결과를 <표 1-18>에 수록하고 남녀간 기대수명 차이를 <표 1-19>에 수록하였다. 각 모형별로 1970~2010년 자료를 이용하여 예측한 경우보다 기대수명이 높게 예측됨을 알 수 있다. 반-대수 모형은 남녀 모두  $\alpha$ 를 100으로 두고 적합한 모형에 대한 예측 결과이며, 남녀 모두  $k_t$ 의 감소속도를 지수형태로 만들어 주는 반-대수모형이 랜덤워크 모형보다 낮은 기대수명 예측치를 준다. 고연령을 로지스틱 모형으로 추정한 자료에 대해 모형1과 랜덤워크모형을 적용하여 예측한 경우에 2060년 기대수명은 남자 90.34세, 여자 93.99세이며, 모형1과 반-대수모형을 적용한 경우는 남자 88.04세, 여자는 91.84세로 예측된다. 남녀 간 기대수명 차이는 모형에 따라 차이가 존재하며, 2060년에 3.65~4.30세 정도의 차이를 보이는 것으로 나타났다.

<표 1-17> 1983~2010년에 대한 모형별 사망확률 적합력

MAPE	Logistic					Brass-Logit				
	RDW			반-대수		RDW			반-대수	
	모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
남자	0.030	0.051	0.032	0.052	0.080	0.031	0.057	0.033	0.053	0.080
여자	0.042	0.086	0.051	0.063	0.128	0.043	0.093	0.050	0.064	0.128

<표 1-18> 장래기대수명

(a) 남자

연도	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	77.18*	76.15	76.59	77.00	77.18	76.70	77.10	76.56	76.91	77.18	76.71	77.06
2015		77.11	78.33	78.69	78.93	78.27	78.63	78.43	78.77	79.06	78.40	78.74
2020		78.04	80.00	80.35	80.63	79.74	80.10	80.20	80.53	80.88	79.96	80.29
2025		78.93	81.56	81.90	82.23	81.08	81.43	81.89	82.19	82.61	81.41	81.72
2030		79.79	83.04	83.36	83.73	82.32	82.65	83.50	83.77	84.25	82.75	83.04
2035		80.60	84.42	84.73	85.14	83.45	83.77	85.03	85.28	85.81	84.00	84.27
2040		81.39	85.74	86.02	86.47	84.50	84.80	86.49	86.72	87.30	85.16	85.41
2045		82.15	86.98	87.24	87.73	85.48	85.77	87.89	88.09	88.72	86.25	86.48
2050		82.87	88.16	88.41	88.92	86.39	86.66	89.21	89.40	90.06	87.27	87.49
2055			89.28	89.51	90.05	87.24	87.50	90.47	90.64	91.32	88.23	88.43
2060			90.34	90.55	91.11	88.04	88.29	91.66	91.81	92.51	89.12	89.31

\*2010년 생명표는 잠정치

(b) 여자

연도	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	84.06*	82.88	82.96	83.78	84.06	83.14	83.97	82.98	83.78	84.06	83.17	83.97
2015		83.80	84.44	85.25	85.52	84.41	85.22	84.59	85.37	85.63	84.55	85.33
2020		84.68	85.79	86.60	86.89	85.53	86.34	86.07	86.85	87.13	85.78	86.56
2025		85.50	87.04	87.85	88.17	86.55	87.37	87.46	88.23	88.54	86.91	87.69
2030		86.27	88.21	89.01	89.36	87.49	88.30	88.77	89.54	89.86	87.96	88.73
2035		86.99	89.31	90.10	90.48	88.35	89.15	90.01	90.76	91.11	88.93	89.69
2040		87.67	90.35	91.13	91.54	89.15	89.94	91.19	91.91	92.29	89.83	90.57
2045		88.31	91.33	92.09	92.54	89.89	90.68	92.30	92.99	93.39	90.66	91.40
2050		88.92	92.27	93.01	93.48	90.58	91.36	93.34	94.00	94.42	91.44	92.16
2055			93.15	93.86	94.37	91.23	91.99	94.32	94.95	95.38	92.17	92.87
2060			93.99	94.67	95.20	91.84	92.59	95.24	95.83	96.28	92.86	93.53

\*2010년 생명표는 잠정치



〈표 1-19〉 남자와 여자의 장래기대수명 차이

연도	생명표	2005 추계	Logistic					Brass-Logit				
			RWD			반-대수		RWD			반-대수	
			모형1	모형2	모형3	모형1	모형2	모형1	모형2	모형3	모형1	모형2
2010	6.88	6.73	6.36	6.79	6.88	6.44	6.87	6.42	6.87	6.88	6.47	6.91
2015		6.69	6.11	6.55	6.59	6.14	6.58	6.17	6.60	6.57	6.15	6.58
2020		6.64	5.79	6.25	6.26	5.79	6.25	5.87	6.32	6.25	5.82	6.27
2025		6.57	5.47	5.95	5.94	5.47	5.94	5.57	6.04	5.93	5.51	5.97
2030		6.48	5.17	5.66	5.64	5.17	5.65	5.28	5.76	5.61	5.21	5.69
2035		6.39	4.89	5.38	5.35	4.90	5.39	4.99	5.48	5.30	4.93	5.42
2040		6.28	4.61	5.11	5.08	4.65	5.14	4.70	5.19	4.98	4.67	5.16
2045		6.16	4.36	4.85	4.82	4.41	4.91	4.41	4.89	4.67	4.41	4.91
2050		6.05	4.11	4.60	4.56	4.19	4.69	4.13	4.60	4.37	4.18	4.67
2055			3.87	4.36	4.32	3.99	4.49	3.85	4.31	4.06	3.95	4.44
2060			3.65	4.12	4.09	3.80	4.30	3.58	4.02	3.77	3.73	4.22

## 제5절 결 론

### 1. 요약

본 연구에서는 사망력 예측모형을 이용한 장래 사망수준에 대한 예측방법을 검토하기 위하여 부족한 고연령 사망확률 추정, 적합기간, LC모형에 기반한 예측모형 등을 고려하여 비교 분석을 실시하였다. 분석을 위한 기초자료는 1970~2010년의 각 세별 사망률 자료를 사망확률로 변환한 후 Greville 9차항 방법을 이용하여 보정한 자료를 사용하였다. 0세에 대해서는 누락분 등이 조정된 생명표의 사망확률을 사용하였으며, 부족한 75세 이상 고연령 사망확률은 2-모수 로지스틱 모형과 Brass-Logit 모형 그리고 Coale-Kisker 모형으로 추정하고 비교하였다. 고연령 사망확률에 대한 모형 추정값과 100세까지 제공되고 있는 2000~2010년 생명표 사망확률과 기대수명을 비교한 결과 로지스틱 모형이 적절한 것으로 평가되었으며, 고연령 추정모형에 따른 예측결과 차이를 파악하기 위하여 2-모수 로지스틱 모형과 Brass-Logit 모형으로 추정한 사망확률을 각각 사용

하였다.

모형 적합에 사용되는 자료의 기간에 대해서는 1970년 이후와 자료가 정비되어 현재 국가통계포털에 제공되고 있는 1983년 이후로 구분하고, 사망력 예측모형으로는 예측력과 적용의 용이함으로 높게 평가 받고 있는 LC모형에 기반한 3가지 모형을 적용하였다.

적합기간과 모형의 예측력을 평가하기 위하여 1970~2005년, 1983~2005년 자료를 이용하여 2006~2010년을 예측하고 예측결과를 비교한 결과 모든 모형에서 생명표 기대수명보다는 다소 과소 추정되는 경향이 있으며, 5년 단기 예측에 있어서는 1983~2005년 자료에 대해 모형2 또는 모형3을 이용하여 예측한 기대수명이 생명표 기대수명에 더 근사한 결과를 제공함을 알 수 있었다.

영아 사망확률은 다른 연령의 개선패턴과 다소 다른 양상을 가지며 우리나라는 이미 세계 최저 수준에 도달되어 있는 상태이다. 향후에도 영아 사망확률은 다른 연령층과 같은 속도로 개선될 것으로 볼 수 없으므로, LC모형을 이용하지 않고 일본의 2005~2055년 영아 사망확률 추계결과와 고펀페르츠 모형을 이용하여 예측하였다. 장래 사망확률 예측을 위하여 적합기간을 1970~2010년과 1983~2010년으로 구분하고, LC모형에 대해 시간모수  $k_t$ 의 예측모형으로 추세를 갖는 랜덤워크모형과 일본 인구추계에서 사용된 반-대수모형을 적용하였다. 장래 기대수명 결과는 적용한 모형과 적합기간에 따라 차이가 존재하였다. 적합기간으로 1983~2010년을 이용한 경우의 장래 기대수명이 1970~2010년을 이용한 경우보다 더 높게 예측되었으며, 고연령 모형으로 Brass-Logit 모형을 적용한 경우가 로지스틱 모형보다 약간 높게 예측되었다. 시간모수 예측모형으로 반-대수모형을 적용하였을 때, 시간에 따른 장래 사망변화수준을 지속적으로 감소시킴으로써 선형적으로 감소하는 랜덤워크 모형을 적용하였을 때보다 기대수명이 낮게 예측되었다.

## 2. 결론 및 제언

LC모형을 기반으로 적용한 장래 사망력 예측에 대한 지금까지의 연구 결과를 바탕으로 정리하면 다음과 같다. (1)부족한 고연령에 대한 추정모형으로 3가지 모형을 검토한 결과 적합력과 기대수명 측면에서 2모수 로지스틱 모형이 적절하다고 판단되었다 (2) 고연령 모형 추정에 따른 예측결과 차이는 단기적으로는 크지 않으나, 장기적으로는 1~2세 정도의 차이를 보였으며 Brass-Logit 모형으로 추정한 경우의 기대수명이 높게 예측되었다. 본 연구에서는 Brass-Logit 모형 추정시 표준생명표로 2010년 잠정 생명표를 사용하였으며, Brass-Logit 모형을 적용하기 위해서는 표준생명표의 확정이 필요할 것이다 (3)적합기간에 대해서는 10년 또는 20년 정도의 단기적 관점에서 1983~2010년 자료



를 이용하여 장래 사망수준을 예측하는 것이 바람직한 것으로 판단되나, 50년 후 예측이라는 장기적 관점에서 최근 사망수준 감소 속도가 지속될 것이라는 전망 또한 불확실하므로 최대한 장기 시계열을 사용하는 것도 바람직하다고 판단된다 (4)장래 사망수준과 기대수명 예측에 대한 여러 가지 불확실성에 대해 LC모형은 예측 결과에 대해 신뢰구간을 제공하며, 가능한 주관을 배제한 객관적인 방법에 근거한 예측을 가능하게 한다는 장점을 가지고 있으므로 예측모형으로 적절하다고 판단된다 (5)고려한 LC모형에 따라 예측 결과의 차이가 존재하였으며, 시간모수 모형을 포함한 예측모형 선택에 있어서 향후 사망수준 개선 속도에 대한 인구학적 측면을 고려하지 않을 수 없다고 보여진다.

우리나라의 41년(1970~2010년) 사망자료 시계열은 50년 장기 예측을 하기에 충분하지 않으며, 단기간 동안 급속한 사망수준 개선이 이루어져 왔다는 측면은 장기 예측을 더욱 어렵게 하는 요인이다. 본 연구는 사망력 예측을 위하여 확률 모형 중 비교적 좋은 평가를 받고 있는 LC모형을 기반으로 하여 예측방법을 검토하고 비교한 것으로 실제 예측을 하는 과정에서는 여러 가지 선택의 문제가 발생하게 되며, 이에 따라 예측 결과에도 영향을 미치게 될 것이다. 또한 본 연구에서 고려한 LC모형 외에도 관련된 여러 인구의 사망력을 통합하고 각 개별인구의 특성을 고려하는 방법으로 확장된 Coherent LC모형(Li and Lee, 2005)도 고려해 볼 수 있을 것이다. 김수영(2011)은 Coherent LC모형을 우리나라 사망자료에 적용하였다.

한 국가의 장래 사망수준은 국가 정책 수립에 중요한 기초자료이며, 그에 따른 사회적·경제적 파급효과 또한 매우 크다. 앞서 언급하였듯이 사망력 추계는 방법에 있어 가능한 주관을 배제하고 객관적 측면에서 접근되어야 하며, 장래 예측이라는 불확실성(uncertainty)과 확률모형(stochastic model) 선택에 따른 예측 결과 차이 등의 고려와 더불어 예측결과에 대한 타당성 또한 확보되어야 할 것이다. 본 연구결과가 2011년 장래인구추계를 앞두고 있는 시점에서 과거보다 개선된 사망력 추계 방법 적용과 국가 인구통계 품질 제고에 도움이 되기를 기대한다.

## 참고문헌

- 김기환, 정승환 (2006). 우리나라 생명표의 연령구간 확장 및 기대여명 예측, *Journal of the Korean Data Analysis Society*, 8, 5, 1723-1733.
- 김수영 (2011). 국외훈련보고서
- 박경애 (2011). 장래인구추계의 이론과 실무, 통계교육원
- 정승환, 김기환 (2011). 평균수명을 이용한 사망확률 예측모형 비교연구, *응용통계연구*, 24, 1, 115-125
- Beard, R.E. (1963). A Theory of mortality based on actuarial, Biological and Medical Considerations, In *Proceedings of International Population Conference, New York 1961*, 1, 611-625.
- Booth, H., Maindonald, J. and Smith L. (2002). Applying Lee-Carter under conditions of variable mortality decline, *Population Studies*, 56, 325-336.
- Brass, W. (1971). *On the scale of Mortality, biological Aspects of Demography*, Taylor and Francis, London.
- Coale, A and Kisker, E. (1990). Defects in data on old age mortality in the United States : new procedures for calculating approximately accurate mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4, 1-31.
- Denuit, M. and Goderniaux, A. (2005). Closing and projecting life tables using log-linear models. *Bulletin de l'Association Suisse des Actuaries*, 1, 29-49.
- Edviges Coelho (2007). Mortality and longevity projections for the oldest-old in Portugal. Eurostat/UNECE Work Session on Demographic Projections, WP.10.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Philosophical Transactions of Royal Society(SeriesA)*, 115, 513-583.
- Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rate : A functional data approach, *Computational Statistics & Data Analysis*, 51, 4942-4956.
- Kannisto, V. (1992). Development of oldest-old mortality, 1950-1990; Evidence from 28 developed countries. Odense University Press.
- Lee, R.D. and Cater, L.R. (1992). Modeling and forecasting the time series of U.S. mortality, *Journal of the American Statistical Association*, 87, 659-671.
- Lee, R.D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography*, 38, 537-549.
- Li, N., and R. Lee. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method, *Population Association of America*, 42, 3, 575-594.
- Makeham, W.M. (1867). On the law of mortality, *Journal of the Institute of Actuaries*, 8, 301-310; *Journal of the Institute of Actuaries*, 13, 325-358.



- Perks, W. (1932). On some experiments in the graduation of mortality statistics, of the Institute of Actuaries, 63, 12-40.
- Shigesato Takahashi, Akira Ishikawa, Hisakazu Kato, Miho Iwasawa, Ryuichi Komatsu, Ryuichi Kaneko, Masako Ikenoue, Fusami Mita, Akiko Tsuji, and Moriizumi (2003). Population Projections for Japan 2001-2050 with Lpng-Range Population Projections: 2051-2100, Journal of Population and Social Security (Population), 1, 1.
- Thatcher, A., Kannisto, V. and Vaupel, J. (1998). The force of mortality at ages 80 to 120. Odense Monographs on Population Aging, Odense University Press, Odense.
- Zeng Yi and James W. Vaupel (2003). Oldest-Old Mortality in China, 8, 215-244.