

Comparison of Mortality Estimate and Prediction by the Period of Time Series Data Used

Kyunam Jung^a · Jeeseon Baek^b · Donguk Kim^{c,1}

^aStatistics Korea; ^bStatistical Research Institute, Statistics Korea

^cDepartment of Statistics, Sungkyunkwan University

(Received November 8, 2013; Revised December 18, 2013; Accepted December 23, 2013)

Abstract

The accurate prediction of future mortality is an important issue due to recent rapid increases in life expectancy. An accurate estimation and prediction of mortality is important to future welfare policies. The optimal selection of a mortality model is important to estimate and predict mortality; however, the period of time series data used is also an important issue. It is essential to understand that the time series data for mortality is short in Korea and the data before 1982 is incomplete. This paper divides the time series of Korean mortality into two sets to compare the parameter estimates of the LC model and LC model with a cohort effect by the period of data used. A modeling and prediction of the mortality index and cohort effect index as well as the evaluation of future life expectancy is conducted. Finally, some suggestions are proposed for the future prediction of mortality.

Keywords: Period of time series data used, LC model, LC model with cohort effect, mortality, life expectancy.

1. 서론

최근 통계청의 발표에 의하면 2012년 평균수명은 남자 77.9세, 여자 84.6세로 나타났다. 이는 2002년 평균수명에 비해 10년 동안 남자 4.6세, 여자 4.2세가 증가한 것으로 세계 어느 나라보다 그 증가속도가 빠르게 나타나고 있다. 이러한 추세는 향후 꾸준히 유지될 것으로 보여 정부에서는 장기 연금정책 등 각종 복지정책수립을 위한 사망률 예측을 통한 장래 기대수명 추계에 대한 관심이 어느 때보다 높다. 이러한 이유로 사망률 예측에 대한 연구가 활발하게 이루어지고 있으며, 전세계적으로 지난 20여 년간에 걸쳐서 사망률 예측을 위한 모형화는 커다란 발전을 이루어 왔다. 가장 대표적인 모형은 Lee와 Carter (1992), Lee (2000)에서 제시된 LC모형으로 사망률을 포함한 인구학적 변수를 예측하는 효과적인 방법으로 널리 알려져 있다. 이 모형은 미국 센서스국의 기준방법으로 사용되기도 하였으며 우리나라의 경우도 2006년 사망률 예측을 위하여 처음 도입하여 적용한 바 있다. LC모형은 사망률 자료에서 연령 효과와 시간효과를 추정한 후 시간효과를 모형화하고 예측하는 형태이며 이후 Renshaw와 Haberman (2006)과 Cairns 등 (2007, 2009, 2011)에서는 LC 모형에 코호트효과를 포함하는 모형을 토대로 사망률에 대한 연구를 실시하였다. 이밖에도 많은 모형들이 장래 사망률 예측을 위하여 연구되어 왔다.

이와 같이 다양한 모형을 토대로 사망률 예측에 적합한 최적모형을 찾으려는 연구는 매우 중요한 과정이

¹Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 110-745, Korea. E-mail: dkim@skku.edu

나 여기서 간과하지 말아야 할 것은 사망력 시계열 적용기간을 어떻게 설정할 것인가에 대한 연구도 중요하다. Cairns 등 (2009)는 장래 사망력 예측을 위해서는 시계열 적용기간에 따른 모수 추정값의 robustness 문제 또한 중요한 이슈라는 점을 강조하였다. 즉, LC모형과 코호트효과가 포함된 여러 개의 모형을 토대로 모수 추정값의 robustness를 검토한 결과 대부분의 모형에서 robustness에 문제가 있다는 것을 지적하였다.

앞에서 소개한 모형들을 이용한 우리나라 사망력 예측에 대한 연구로는 Park 등 (2005), Baek과 Jeong (2012), Jung (2012) 등이 있으며, 우리나라의 경우 선진국에 비해 사망자료의 시계열이 짧아 1970년 이후 시계열밖에는 이용할 수 없다는 점과 1982년 이전의 자료가 고연령층 부분에서 부족하여 다소 불완전하다는 점, 짧은 기간동안 사망력의 개선이 급속히 이루어졌다는 점 등의 특수성을 고려할 때 사망력 예측을 위한 모형뿐만 아니라, 시계열 적용기간의 적정성 검토도 매우 중요하다고 할 수 있다. 시계열 적용기간에 대해 Baek과 Jeong (2012)은 모수 추정 및 예측 결과에 차이가 있음을 지적한 바 있으며, Jung (2012)은 우리나라 사망력 시계열에 코호트효과가 있음을 확인하고 1983년부터 2005년까지의 시계열자료를 이용하여 LC모형과 코호트효과를 포함한 확장모형에 대한 연구를 수행하였다.

본 논문에서는 사망력 추계를 위해 가장 대표적으로 많이 사용되고 있는 ‘LC모형’과 LC모형의 기본 틀에 코호트효과를 추가하여 사망력 추계를 한 Renshaw와 Haberman (2006)에서 사용한 ‘LC 코호트효과 확장모형’을 기본 모형으로 하여, 사망력에 대한 시계열 적용기간을 다르게 적용할 경우 각 모형별로 모수 추정결과는 어떠한 차이를 보이며, 모형별 사망력지수와 코호트지수에 대하여 적절한 예측모형의 선정결과 및 기대수명에 대한 예측적합력을 비교 분석하고자 한다.

이를 위해서 Jung (2012)에서 1983년부터 2005년까지의 시계열을 이용하여 분석한 결과와 1970년 이후 비교적 사망력 시계열이 안정적 추세를 보이는 1976년부터 1982년까지의 시계열을 추가하여 분석한 결과간의 차이점을 비교분석하여 향후 장래 기대수명 추계시 고려해야 할 사항을 제시한다.

본 논문에서는 제 2절에서 시계열 적용기간을 달리하여 LC모형과 LC 코호트효과 확장모형에 대한 모형별 모수추정결과를 비교하고, 제 3절에서는 두 모형의 사망력지수와 코호트지수에 대한 모형화 및 최적 모형을 통한 모형별 예측값을 산출한다. 제 4절에서는 시계열 적용기간에 따른 모형별 기대수명을 추계하고, 각각에 대한 예측 적합력을 검증한 후, 기대수명간의 차이를 분석한다. 제 5절에서는 향후 장래 기대수명 추계를 위한 제언을 제시한다.

2. 사망력 시계열 적용기간별 모수 추정결과 비교

2.1. 모수 추정모형

사망력 자료는 연령효과, 시간효과, 코호트효과로 구분할 수 있다. 연령효과는 연령에 따른 사망력 수준과 그 변화효과를 의미하며, 시간효과는 시간에 따라 변화하는 사망력지수를 나타낸다. 코호트효과는 특정 출생 코호트가 다른 코호트에서 볼 수 없는 특징을 나타내면서 변화하는 효과를 의미한다. 본 논문에서는 사망력 시계열 적용기간에 따른 분석을 위하여 사망력 추계시 가장 많이 사용하고 연령효과와 시간효과를 모형화한 ‘LC모형’ (Lee와 Carter, 1992)과 LC모형에 일반화된 코호트효과를 확장한 ‘LC 코호트효과 확장모형’ (Renshaw와 Haberman, 2006)을 기본 모형으로 분석한다.

$m_{x,t}$ 를 t 시점의 x 세의 연앙사망률(mid-year death rate)이라고 하면 LC모형은 다음과 같다.

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \epsilon_{x,t}, \quad (2.1)$$

$$\sum_t \kappa_t^{(2)} = 0, \quad \sum_x \beta_x^{(2)} = 1,$$

여기서 $\beta_x^{(1)}$ 는 사망력의 전 연령에 걸친 일반 추세이며, $\beta_x^{(2)}$ 는 $\kappa_t^{(2)}$ 변화에 따른 특정 연령층 사망률의 변화상태를 나타낸다. $\kappa_t^{(2)}$ 는 시간에 대한 선형함수로서 Lee와 Carter (1992)에서와 같이 ‘사망력지수(mortality index)’라 부르기로 한다. $\epsilon_{x,t}$ 은 평균이 0이고 분산이 σ^2 로서, 모형에 반영되지 않는 영향을 나타낸다.

LC 코호트효과 확장모형은 LC모형에 코호트효과를 포함한 모형으로 다음과 같이 표현된다.

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}, \quad (2.2)$$

$$\sum_t \kappa_t^{(2)} = 0, \quad \sum_x \beta_x^{(2)} = 1, \quad \sum_{x,t} \gamma_{t-x}^{(3)} = 0, \quad \sum_x \beta_x^{(3)} = 1,$$

여기서 $\beta_x^{(1)}$ 는 연령효과, $\beta_x^{(2)} \kappa_t^{(2)}$ 는 시간효과, $\beta_x^{(3)} \gamma_{t-x}^{(3)}$ 는 코호트효과를 나타낸다. $\kappa_t^{(2)}$ 는 사망력지수이며 $\gamma_{t-x}^{(3)}$ 는 출생연도 $t-x$ 에 대한 순수한 코호트효과 또는 ‘코호트지수(cohort index)’를, $\beta_x^{(3)}$ 는 연령가중치를 의미한다. 우리나라 사망력에 대한 코호트효과에 대해서는 Jung (2012)을 참고하기 바란다.

2.2. 모수 추정결과 비교

LC모형과 LC 코호트효과 확장모형에 대한 모수들 $\beta_x^{(1)}$, $\beta_x^{(2)}$, $\beta_x^{(3)}$, $\kappa_t^{(2)}$, $\gamma_{t-x}^{(3)}$ 는 최우추정법(MLE)으로 추정되었으며, JPMorgan (2007)에서 개발하여 보급한 LifeMetrics Software Version 2.0을 이용하였다. 우리나라의 시계열자료 중에서 부족한 고연령층의 사망력은 CK모형을 이용하여 연령별 표준 사망률을 추정한 후 관계형 모형을 적용하여 추정하였으며 자세한 내용은 Jung과 Kim (2012)을 참조하기 바란다.

2.2.1. 남자 사망력 모형 추정 기간별(1976~2005년, 1983~2005년) 남자의 사망력 모형 추정 결과는 Figure 2.1과 Figure 2.2와 같다. Figure 2.1은 LC모형 추정결과로서 사망력의 전 연령에 걸친 일반 추세를 나타내는 $\beta_x^{(1)}$ 에 대한 추정 결과를 보면 시계열 추가로 인한 연령별 추세 차이는 거의 없었으며, 사망력지수 $\kappa_t^{(2)}$ 변화에 따른 특정 연령층 사망률의 변동을 나타내 주는 $\beta_x^{(2)}$ 는 50세 이하에서 차이를 보이고 있는 것으로 나타났다. 시간에 대한 선형함수로서 알려진 사망력지수 $\kappa_t^{(2)}$ 는 1976년 이후 자료를 사용한 경우 1976년부터 1982년까지의 추세가 다소 불규칙한 모습을 보이고 있으나 1983년 이후부터는 적용기간에 따라 추정결과 수준 차이는 존재하며 추세는 동일한 모습을 보이고 있다. 따라서 LC모형의 경우 시계열 적용기간에 따른 모수 추정값의 robustness 문제는 크게 나타나지 않았다.

Figure 2.2는 LC 코호트효과 확장모형 추정결과로서 연령효과인 $\beta_x^{(1)}$ 와 사망력지수와 코호트지수의 가중치 역할을 하는 $\beta_x^{(2)}$, $\beta_x^{(3)}$ 의 추정값은 시계열 적용기간에 따른 추정결과간에 차이가 있다. 또한, 사망력지수의 경우 시계열 적용기간과 관계없이 1983년 이후 감소추세를 보이고 있으나 1976년 이후 시계열을 적용했을 경우가 더 가파른 감소추세를 보이고 있고, 코호트지수의 경우도 시계열자료 적용기간에 따라 크게 차이를 보이고 있다. 따라서 코호트효과를 포함한 모형의 경우, 시계열 적용기간에 따른 모수 추정값의 robustness에 문제가 있기 때문에 이에 대한 충분한 검토가 필요하다는 것을 알 수 있다.

2.2.2. 여자 사망력 모형 추정 여자의 사망력 모형 추정 결과는 Figure 2.3과 Figure 2.4와 같다. Figure 2.3은 LC모형 추정결과로서 $\beta_x^{(1)}$ 에 대한 추정값은 사망력 시계열 적용기간에 따른 차이는 거의 없다. $\beta_x^{(2)}$ 의 추정값은 남자와는 달리 전 연령대에서 다소 차이를 보여 주었으며, 사망력지수인 $\kappa_t^{(2)}$ 는 1976년 이후 자료를 사용한 경우 1976년부터 1982년까지의 추세가 다소 불규칙한 모습을 보이고 있는

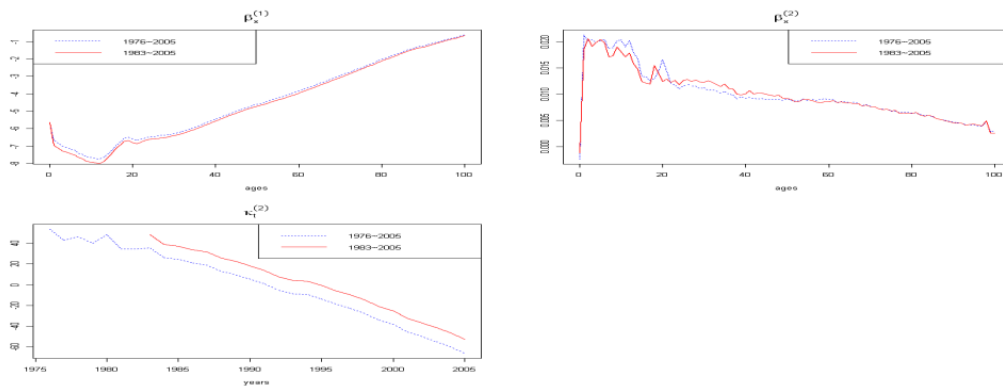


Figure 2.1. Parameter estimates for LC model by the period of data used(male)

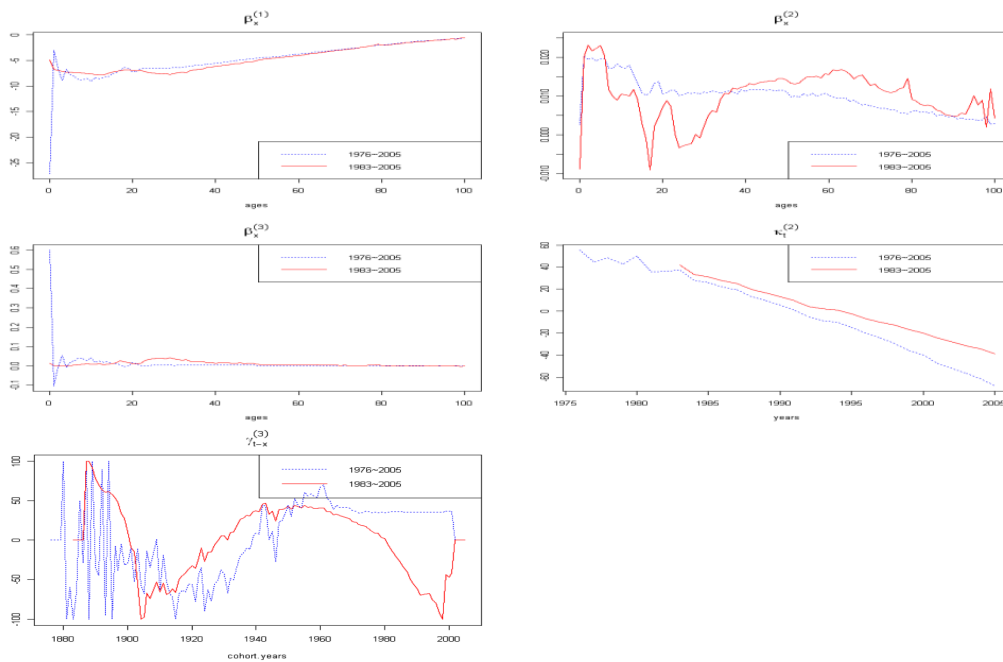


Figure 2.2. Parameter estimates for LC model with cohort effect by the period of data used(male)

나 1983년 이후는 2개의 시계열간에 동일한 추세를 보였다. 따라서 LC모형의 경우 시계열 적용기간에 따른 모수 추정값의 robustness 문제는 크지 않으나 남자의 경우보다는 다소 크게 나타났다.

Figure 2.4는 LC 코호트효과 확장모형 추정결과로서 $\beta_x^{(1)}$, $\beta_x^{(2)}$, $\beta_x^{(3)}$ 의 추정값은 시계열자료의 적용기간에 따라 추정결과가 차이를 보이고 있지만, 남자 시계열 보다는 방향성에 있어서 어느 정도 일치하게 나타났다. 반면, 사망력지수의 경우 1983년 이후 동일한 하향 추세를 나타내고 있으나 추세의 가파른 정도에서는 크게 나타났으며, 코호트지수의 경우 1960년경부터 방향성이 반대로 나타나고 있다. 이에 따라 코호트효과를 포함한 모형의 경우 모수 추정값의 robustness에 문제가 있기 때문에 사망력 시계열 적용기간에 대한 충분한 검토가 필요함을 보여 주고 있다.

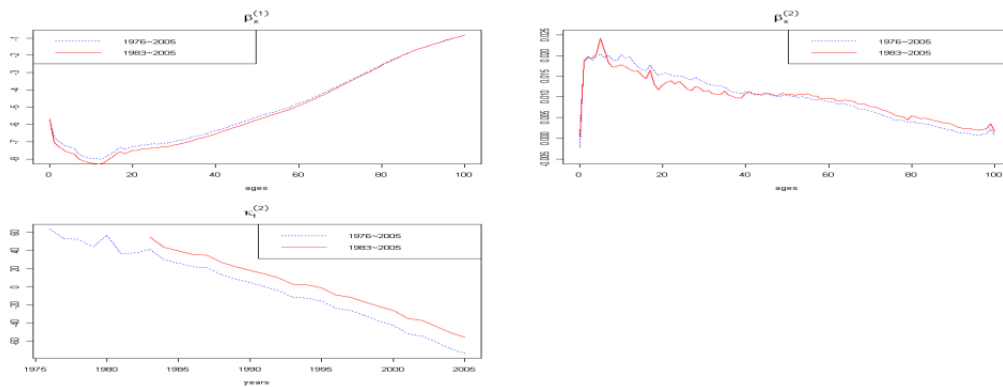


Figure 2.3. Parameter estimates for LC model by the period of data used(female)

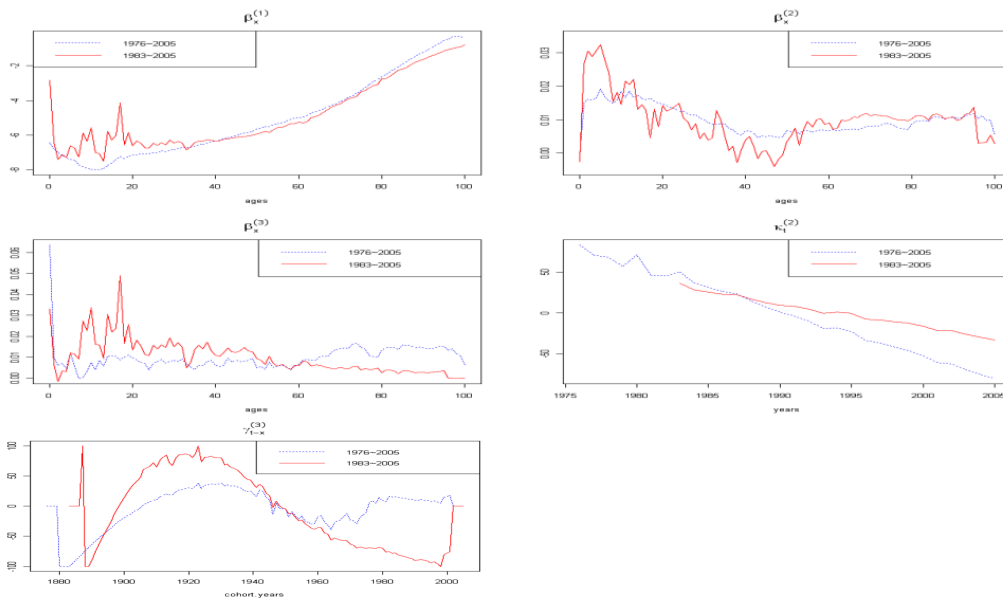


Figure 2.4. Parameter estimates for LC model with cohort effect by the period of data used(female)

3. 사망률지수 및 코호트지수의 모형화 및 예측

모형을 이용한 사망률 예측 및 장래 기대수명 추계를 위해서는 확률변수인 사망률지수 $\kappa_t^{(2)}$ 와 코호트지수 $\gamma_{t-x}^{(3)}$ 에 대한 모형화가 필요하다. 본 절에서는 사망률지수와 코호트지수에 대한 모형화 과정을 살펴보고, 이를 통하여 모형별 예측값을 산출하였다. 분석에 언급된 ARIMA모형들은 intercept가 포함되지 않은 모형이다.

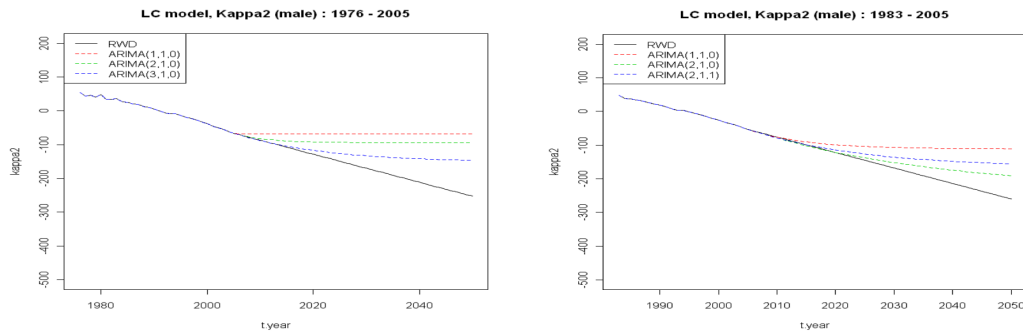
3.1. 사망률지수의 모형화 및 예측

3.1.1. 남자 사망률지수의 모형화 및 예측

LC모형의 장래 사망률 예측을 위한 사망률지수 $\kappa_t^{(2)}$ (이

Table 3.1. Statistics for the fitted models of mortality index by the period of data used(LC model, male)

Period of Data : 1976~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(1, 1, 0)	32.35	-91.58	187.17(4)	189.90(4)
ARIMA(2, 1, 0)	19.38	-84.70	175.40(3)	179.50(3)
ARIMA(3, 1, 0)	15.20	-81.53	171.06(2)	176.53(2)
RWD	17.16	-62.89	129.77(1)	132.51(1)
Period of Data : 1983~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(1, 1, 0)	6.83	-53.18	110.37(4)	112.55(3)
ARIMA(2, 1, 0)	5.15	-50.41	106.82(2)	110.09(2)
ARIMA(2, 1, 1)	5.21	-50.64	109.28(3)	113.65(4)
RWD	3.70	-39.15	82.30(1)	84.57(1)

**Figure 3.1.** Forecast by the fitted model of mortality index(LC model, male)

하 그림에서 ‘kappa2’로 표기)의 최적 모형을 찾기 위하여 자기상관함수(Autocorrelation Function, ACF)와 부분자기상관함수(Partial Autocorrelation Function; PACF) 및 1차 차분 후의 ACF와 PACF 등을 참조하여 구한 적합모형별 통계량을 구하면 Table 3.1과 같으며, 적합모형별로 2050년까지 예측결과를 보면 각각 Figure 3.1과 같다. 이들 모형 중에서 AIC값과 BIC값을 기준으로 보면 사망률의 시계열 적용기간과 관계없이 사망률지수의 모형은 RWD(Random Walk with Drift)가 적합력이 가장 우수한 것으로 나타났다. 다만, 본 논문에서는 사망률에 대한 추정모형의 적합력뿐만 아니라 우리가 궁극적으로 구하고자 하는 장래 기대수명에 대한 예측 적합력도 중요한 요소이기 때문에 제4절에서는 사망률지수와 코호트지수의 적합모형 중 가장 우수한 2개의 모형을 대상으로 비교 분석할 예정이다.

LC 코호트효과 확장모형의 사망률지수 $\kappa_t^{(2)}$ 의 최적모형을 찾기 위하여 LC모형과 동일한 방법으로 적합모형별 통계량을 구하면 Table 3.2와 같으며 적합모형별로 2050년까지 예측결과를 보면 Figure 3.2와 같다. 이들 모형 중에서 AIC값과 BIC값 기준으로 보면 시계열의 적용기간과 관계없이 사망률지수의 모형은 RWD가 적합력이 가장 우수한 것으로 나타났다. 결론적으로 남자의 경우 시계열 적용기간과 관계없이 동일하게 RWD 모형이 선정되었다.

3.1.2. 여자 사망률지수의 모형화 및 예측 LC 모형의 장래 사망률 예측을 위한 사망률지수 $\kappa_t^{(2)}$ 의 최적모형을 찾기 위하여 적합모형별 통계량을 구하면 Table 3.3과 같으며 모형별로 2050년까지 예측결과를 보면 Figure 3.3과 같다. 이들 모형 중에서 AIC값과 BIC값을 기준으로 보면 시계열의 적용기간과

Table 3.2. Statistics for the fitted models of mortality index by the period of data used(LC model with cohort effect, male)

Period of Data : 1976~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(2, 1, 0)	19.33	-84.66	175.31(4)	179.42(4)
ARIMA(3, 1, 0)	15.45	-81.72	171.43(3)	176.90(3)
ARIMA(1, 1, 2)	9.94	-76.35	160.70(2)	166.17(2)
RWD	16.25	-62.48	128.96(1)	131.69(1)
Period of Data : 1983~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(2, 1, 0)	3.43	-45.89	97.79(2)	101.06(2)
ARIMA(3, 1, 0)	3.26	-45.39	98.78(3)	103.14(3)
ARIMA(2, 1, 1)	3.27	-45.60	99.19(4)	103.56(4)
RWD	2.28	-36.36	76.72(1)	78.99(1)

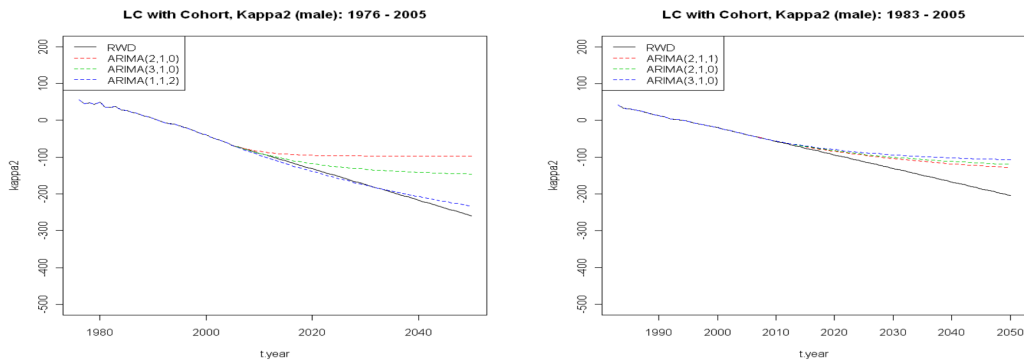


Figure 3.2. Forecast by the fitted model of mortality index(LC model with cohort effect, male)

Table 3.3. Statistics for the fitted models of mortality index by the period of data used(LC model, female)

Period of Data : 1976~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(2, 1, 0)	39.67	-94.77	195.55(4)	199.65(4)
ARIMA(3, 1, 0)	27.88	-90.22	188.44(3)	193.91(3)
ARIMA(1, 1, 2)	14.05	-82.98	173.95(2)	179.42(2)
RWD	30.20	-67.13	138.25(1)	140.99(1)
Period of Data : 1983~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(2, 1, 0)	10.87	-58.28	122.57(3)	125.84(3)
ARIMA(0, 1, 2)	16.60	-62.66	131.32(4)	132.66(4)
ARIMA(3, 1, 0)	8.96	-56.47	120.95(2)	125.31(2)
RWD	6.72	-42.59	89.17(1)	91.44(1)

관계없이 사망력지수의 모형은 RWD가 적합력이 가장 우수한 것으로 나타났다.

LC 코호트효과 확장모형 사망력지수 $\kappa_t^{(2)}$ 의 적합모형별 통계량을 구하면 Table 3.4와 같으며 적합모형 별로 2050년까지 예측결과를 보면 Figure 3.4과 같다. 이들 모형 중에서 AIC값과 BIC값을 기준으로

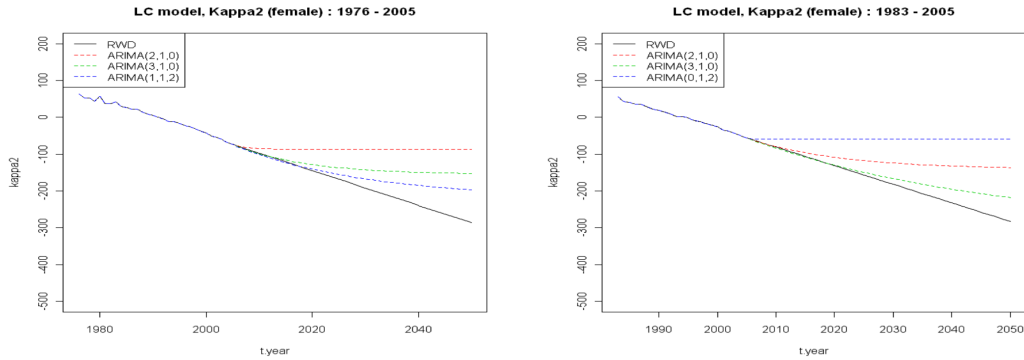


Figure 3.3. Forecast by the fitted model of mortality index(LC model, female)

Table 3.4. Statistics for the fitted models of mortality index by the period of data used(LC model with cohort effect, female)

Period of Data : 1976~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(3, 1, 0)	40.08	-95.41	198.83(4)	204.30(4)
ARIMA(2, 1, 1)	25.54	-90.44	188.87(3)	194.34(3)
ARIMA(3, 1, 2)	15.85	-84.00	180.00(2)	180.00(2)
RWD	41.49	-69.51	143.02(1)	145.75(1)
Period of Data : 1983~2005				
$\kappa_t^{(2)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(1, 1, 0)	9.88	-56.67	117.35(4)	119.53(4)
ARIMA(2, 1, 1)	5.00	-50.88	109.77(2)	114.13(2)
ARIMA(3, 1, 0)	6.83	-53.20	114.41(3)	118.77(3)
RWD	5.27	-41.19	86.39(1)	89.12(1)

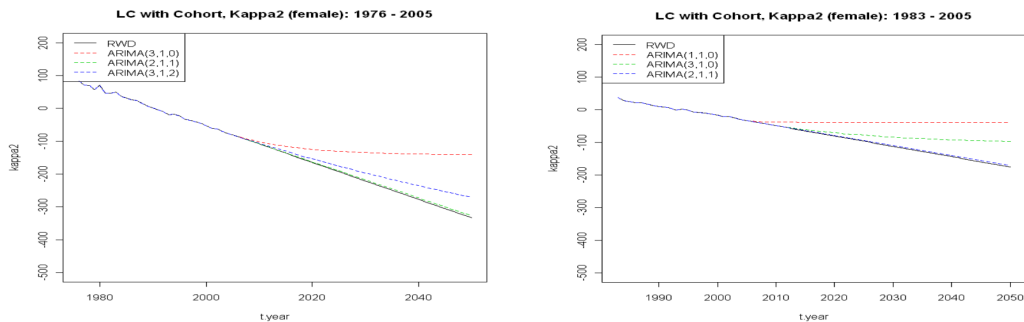


Figure 3.4. Forecast by the fitted model of mortality index(LC model with cohort effect, female)

보면 시계열의 적용기간과 관계없이 사망력지수의 모형은 RWD가 적합력이 가장 우수한 것으로 나타나 여자의 경우도 시계열 적용기간과 관계없이 동일하게 RWD 모형이 선정됨을 알 수 있다.

이상의 것을 종합해 보면, 사망력지수의 경우 성별이나 기본 모형에 관계없이 동일한 모형이 선정되어 시계열 적용기간간의 적합모형 차이는 없는 것으로 나타났다.

Table 3.5. Statistics for the fitted models of cohort index by the period of data used(LC model with cohort effect, male)

Period of Data : 1976~2005				
$\gamma_{t-x}^{(3)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(1, 1, 0)	1196	-605.92	1215.84(4)	1221.45(4)
ARIMA(2, 1, 0)	1046	-597.92	1201.84(2)	1210.25(2)
ARIMA(3, 1, 0)	991.8	-594.77	1197.53(1)	1208.75(1)
ARIMA(3, 1, 2)	973.8	-596.00	1204.00(3)	1220.82(3)
Period of Data : 1983~2005				
$\gamma_{t-x}^{(3)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)
ARIMA(1, 1, 0)	126.8	-441.65	887.30(3)	892.79(2)
ARIMA(1, 1, 1)	121.8	-439.41	884.82(2)	893.05(3)
ARIMA(2, 1, 0)	126.3	-441.41	888.81(4)	897.05(4)
ARIMA(3, 1, 0)	112.1	-434.77	877.54(1)	888.52(1)

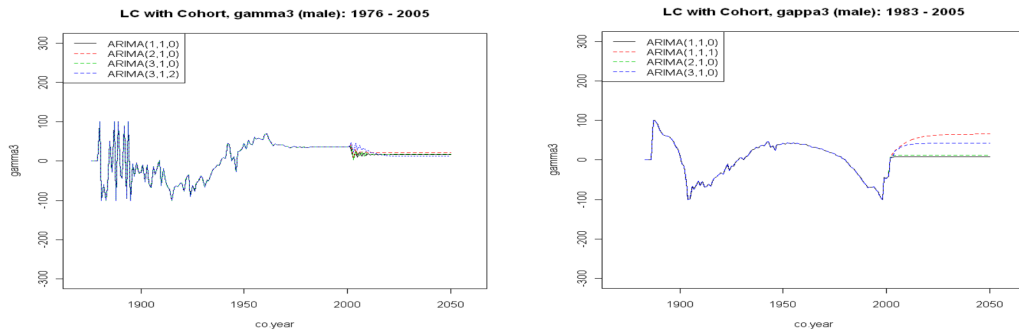


Figure 3.5. Forecast by the fitted model of cohort index(LC model with cohort effect, male)

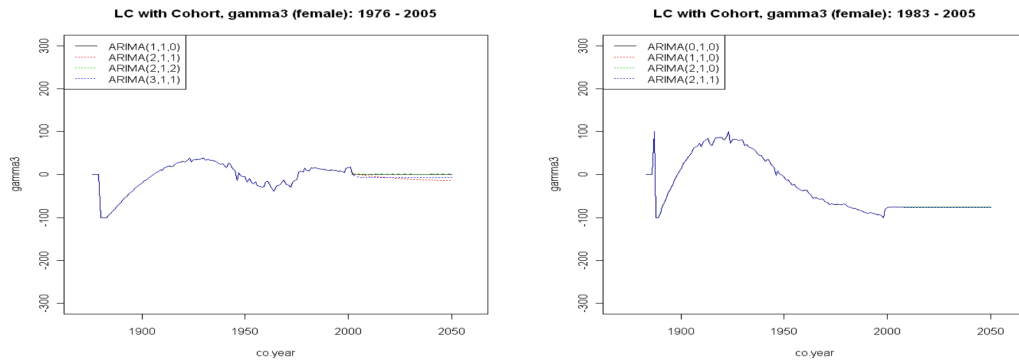
3.2. 코호트지수의 모형화 및 예측

3.2.1. 남자 코호트지수의 모형화 및 예측 Figure 2.2의 LC 코호트효과 확장모형의 코호트지수 $\gamma_{t-x}^{(3)}$ (이하 그림에서 ‘gamma3’로 표기)에 대해 자기상관함수(Autocorrelation Function; ACF)와 부분자기상관함수(Partial Autocorrelation Function; PACF) 및 1차 차분 후의 ACF와 PACF를 구하여 최적의 모형별 통계량을 구하면 Table 3.5와 같으며 적합모형별로 2050년까지 예측결과를 보면 Figure 3.5와 같다. 이들 모형 중에서도 AIC값과 BIC값을 기준으로 보면 시계열의 적용기간과 관계없이 코호트지수의 모형은 ARIMA(3, 1, 0)이 적합력이 가장 우수한 것으로 나타났다. 또한 1983년부터 2005년까지의 시계열을 적용한 사망률 예측결과를 보면 사망률이 증가하는 것으로 예측되어 향후 코호트효과가 기대수명에 음의 영향을 줄 것으로 보인다.

3.2.2. 여자 코호트지수의 모형화 및 예측 Figure 2.4의 LC 코호트효과 확장모형의 코호트지수 $\gamma_{t-x}^{(3)}$ 에 대하여 남자의 경우와 동일한 방법으로 최적의 모형을 찾았다. 그 결과를 보면 적합모형별 통계량은 Table 3.6과 같으며 적합모형별로 2050년까지 예측결과를 보면 Figure 3.6과 같다. 이들 모형 중에서도 AIC값과 BIC값을 기준으로 보면 1976~2005년 시계열의 경우 ARIMA(2, 1, 1) 또는 ARIMA(2, 1, 2), 1983~2005년 시계열의 경우 ARIMA(0, 1, 0)가 우수한 것으로 나타나 시계열 적용기간에 따라 최적 모형이 다르게 나타나는 것을 알 수 있다.

Table 3.6. Statistics for the fitted models of cohort index by the period of data used(LC model with cohort effect, female)

Period of Data : 1976~2005					
$\gamma_{t-x}^{(3)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)	
ARIMA(1, 1, 0)	34.38	-388.89	781.79(4)	787.39(3)	
ARIMA(2, 1, 1)	30.87	-382.56	773.11(2)	784.33(1)	
ARIMA(2, 1, 2)	28.88	-380.60	771.72(1)	785.22(2)	
ARIMA(3, 1, 1)	31.03	-382.82	775.63(3)	789.65(4)	
Period of Data : 1983~2005					
$\gamma_{t-x}^{(3)}$ fitted model	$\hat{\sigma}^2$	LL	AIC(rank)	BIC(rank)	
ARIMA(0, 1, 0)	398.10	-503.00	1008.00(1)	1010.74(1)	
ARIMA(1, 1, 0)	396.60	-502.79	1009.59(2)	1015.06(2)	
ARIMA(2, 1, 0)	396.10	-502.74	1011.48(3)	1019.69(3)	
ARIMA(2, 1, 1)	390.40	-502.05	1012.10(4)	1023.05(4)	

**Figure 3.6.** Forecast by the fitted model of cohort index(LC model with cohort effect, female)

4. 시계열 적용기간 및 모형별 기대수명 비교

4.1. 시계열 적용기간별 기대수명에 대한 예측 적합력 검증

제 3절에서 시계열 적용기간별로 2개의 기본모형에 대한 사망력지수와 코호트지수의 ARIMA모형화를 통하여 후보모형을 선정하였다. 그러나 과거 사망률 시계열을 적용하여 적합된 모형들이 우리가 최종적으로 관심을 갖고 있는 장래 기대수명에 대한 예측력도 우수한 것이라고 기대할 수 없다. 이는 적합모형별로 예측된 사망력으로 통계청의 ‘생명표’ 작성방법(통계청, 각년도)에 따라 장래 기대수명을 추계하기 때문이다. 따라서 시계열 적용기간별로 기본 모형별 적합모형에 대한 예측 적합력을 검증하기 위하여, 제 3장에서 선정된 적합모형별 2006년부터 2010년까지의 기대수명을 추계한 후, 모형별 기대수명과 통계청에서 발표한 기대수명(생명표)에 대한 MAE(Mean Absolute Error)를 산출하고, 모형적합력이 우수하고 MAE값이 적은 적합모형을 Table 4.1과 4.2에 정리하여 수록하였다.

먼저 남자의 경우 시계열 적용기간에 대한 기본모형의 MAE는 Table 4.1과 같다. 1976년부터 2005년까지의 시계열을 적용한 경우 LC모형은 제3절에서 도출한 과거 사망률 시계열에 대한 적합력이 우수한 모형(RWD)과 기대수명에 대한 예측 적합력이 우수한 모형(ARIMA(3, 1, 0))과는 서로 다른 것을 알 수 있다. 전반적으로 LC 코호트효과 확장모형이 LC모형에 비해 MAE 측면에서 우수한 것으로 나

Table 4.1. MAE for life expectancy by the period of data used and model(male)

	Period of Data : 1976~2005						Period of Data : 1983~2005					
	LC model		LC model with cohort effect				LC model		LC model with cohort effect			
cohort index	-		ARIMA(3, 1, 0)		ARIMA(2, 1, 0)		-		ARIMA(3, 1, 0)		ARIMA(1, 1, 0)	
mortality index	RWD (3, 1, 0)		RWD (1, 1, 2)		RWD (1, 1, 2)		RWD (2, 1, 0)		RWD (2, 1, 0)		RWD (2, 1, 0)	
MAE (life table)	0.484	0.421	0.331	0.132	0.272	0.081	0.348	0.170	0.836	0.868	0.740	0.772

Table 4.2. MAE for life expectancy by the period of data used and model(female)

	Period of Data : 1976~2005						Period of Data : 1983~2005					
	LC model		LC model with cohort effect				LC model		LC model with cohort effect			
cohort index	-		ARIMA(2, 1, 1)		ARIMA(3, 1, 1)		-		ARIMA(0, 1, 0)		ARIMA(2, 1, 0)	
mortality index	RWD (1, 1, 2)		RWD (3, 1, 2)		RWD (3, 1, 2)		RWD (3, 1, 0)		RWD (2, 1, 1)		RWD (2, 1, 1)	
MAE (life table)	1.116	0.963	0.070	0.158	0.059	0.119	0.717	0.600	0.523	0.551	0.522	0.550

타났다. 1983년부터 2005년까지의 시계열을 적용한 경우 LC모형은 모수 적합력(RWD)과 예측 적합력(ARIMA(2, 1, 0))이 우수한 모형은 서로 다른 결과를 보였으며, LC 코호트효과 확장모형의 경우도 과거 적합력과 예측 적합력이 우수한 모형은 일치하지 않았다. 종합적으로, 예측 적합력 측면에서 보면 시계열 1976년부터 2005년을 적용하고자 할 경우에는 LC 코호트효과 확장모형을 사용하는 것이 바람직해 보이며, 시계열 1983년부터 2005년을 적용할 경우에는 LC모형을 선택하는 것이 바람직한 것으로 나타났다.

여자의 경우 시계열 적용기간에 따른 모형별 MAE를 Table 4.2에 수록하였다. 1976년부터 2005년까지의 시계열을 적용한 경우를 보면 LC모형은 과거 사망력 시계열에 의한 적합력이 우수한 모형(RWD)과 기대수명에 대한 예측 적합력이 우수한 모형(ARIMA(1, 1, 2))은 서로 다른 결과를 보였다. 전반적으로 LC 코호트효과 확장모형의 경우가 LC모형에 비해 MAE 측면에서 우수한 것으로 나타났다. 시계열 1983년부터 2005년의 경우 LC모형 적용시 모수 적합력(RWD)과 예측 적합력(ARIMA(3, 1, 0))이 우수한 모형은 서로 다른 결과를 보였으나 LC코호트효과 확장모형 적용시는 모수 적합력과 예측 적합력이 우수한 모형은 전체적으로 일치하게 나타났다. 즉, 예측 적합력 측면만 보면 어떤 시계열을 적용하든 LC 코호트효과 확장모형을 사용하는 것이 바람직해 보인다.

Table 4.1과 Table 4.2로부터 5년 정도의 단기 예측 측면에서 볼 때, LC 코호트효과 확장모형이 LC모형보다 긴 시계열을 이용하는 경우에 예측 적합력이 우수함을 알 수 있다. 즉, 1976~2005년 자료에 대해 코호트 모형을 이용한 경우, 남녀 모두 5년 단기 예측에서 통계청 기대수명 실측치와 가장 근사하게 예측하고 있는 것으로 나타났다.

4.2. 사망력 시계열 적용기간에 따른 장래기대수명

시계열 적용기간에 따른 장래 기대수명을 비교하기 위하여 2020년과 2030년의 기대수명을 추계한 결과를 Table 4.3과 Table 4.4에 수록하였다. 남자의 경우 시계열 1976년부터 2005년의 경우에 있어서 Figure 3.5에서 알 수 있듯이 코호트효과가 장래 사망력에 크게 영향을 주지 않는 것으로 나타났다. LC모형과 LC 코호트효과 확장모형간의 기대수명간에는 큰 차이를 보이지 않았으며, 다만 LC 코호트효과 확장모형으로 추계한 기대수명이 1세 이내에서 LC모형으로 추계한 기대수명보다 높게 나타나는

Table 4.3. Life expectancy by the period of data used and model(male)

	Period of Data : 1976~2005						Period of Data : 1983~2005					
	LC model		LC model with cohort effect				LC model		LC model with cohort effect			
cohort index	-		ARIMA(3, 1, 0)		ARIMA(2, 1, 0)		-		ARIMA(3, 1, 0)		ARIMA(1, 1, 0)	
mortality index	RWD (3, 1, 0)		RWD (1, 1, 2)		RWD (1, 1, 2)		RWD (2, 1, 0)		RWD (2, 1, 0)		RWD (2, 1, 0)	
2020	79.98	79.11	80.16	80.67	80.17	80.68	80.55	80.50	79.39	78.60	80.18	79.34
2030	82.89	80.31	83.12	83.23	83.12	83.23	83.73	82.66	80.94	79.48	82.75	80.97

Table 4.4. Life expectancy by the period of data used and model(female)

	Period of Data : 1976~2005						Period of Data : 1983~2005					
	LC model		LC model with cohort effect				LC model		LC model with cohort effect			
cohort index	-		ARIMA(2, 1, 1)		ARIMA(3, 1, 1)		-		ARIMA(0, 1, 0)		ARIMA(2, 1, 0)	
mortality index	RWD (1, 1, 2)		RWD (3, 1, 2)		RWD (3, 1, 2)		RWD (3, 1, 0)		RWD (2, 1, 1)		RWD (2, 1, 1)	
2020	84.48	84.34	89.18	88.30	89.18	88.30	85.61	85.55	87.10	86.98	87.09	86.97
2030	86.03	85.26	94.75	93.02	94.72	92.99	87.71	87.12	90.40	90.22	90.39	90.22

것을 알 수 있다. 또한 시계열 1983년부터 2005년의 경우에 있어서는 Figure 3.5에서 알 수 있듯이 코호트효과가 장래 사망률을 높이는 역할을 하는 것으로 나타나 오히려 LC모형으로 추계한 기대수명이 1세 내외에서 높게 나타나고 있다. 따라서 시계열 적용기간에 따라 장래기대수명의 수준뿐만 아니라 모형의 선택에 신중해야 함을 알 수 있다.

여자의 경우 Figure 3.6에서 알 수 있듯이 코호트효과가 장래 사망률을 줄여주는 역할을 하게 되므로 시계열 적용기간과 관계없이 LC 코호트효과 확장모형의 2020년과 2030년의 기대수명이 LC모형으로 추계한 기대수명보다는 1~5세 높게 나타나는 것을 알 수 있다.

5. 결론

사망력에 대한 정확한 예측을 위한 모형연구는 꾸준히 지속되어 왔으나 상대적으로 사망력의 시계열 적용기간에 대한 연구는 많지 않다. 특히, 우리나라와 같이 시계열이 짧고 과거 시계열에 다소 불완전한 경우에는 시계열 적용기간에 대한 연구는 중요한 이슈라고 판단된다.

제 2절에서 제 4절에 걸쳐 살펴 본 바와 같이 사망력 시계열 적용기간을 달리하게 되면 모수 추정값 등 장래 기대수명 추계에 영향을 주는 여러 요소들간에 차이가 나타나는 것을 알 수 있었다. 따라서 본 논문을 통하여 얻어진 결과를 토대로 모수추정값의 robustness 측면, 사망력지수 및 코호트지수의 모형화 및 예측, 그리고 기대수명에 대한 예측 적합력 측면으로 나누어서 장래 기대수명 추계시 고려해야 할 점을 제안하고자 한다.

장래 기대수명 추계를 위해서 우선적으로 검토가 필요한 것은 모수 추정값에 대한 robustness의 적정성이다. 제2절에서 살펴본바와 같이 시계열 적용기간을 2개의 시계열로 나누어서 모수 추정값의 robustness 측면을 살펴보면, Jung (2012)과 Figure 2.3, Figure 2.4에서 볼 수 있듯이 사망력 시계열에 코호트효과가 존재하지만, 남녀 모두 LC모형이 LC코호트효과 확장 모형보다 모수 추정에서 더 안정적임을 알 수 있다. 다만, 2개의 모형 모두 완벽하게 모수 추정값의 robustness를 만족하지 못하므로 보다 정확한 사망력 예측을 위해서는 추가적인 연구가 필요하다.

사망력지수 및 코호트지수의 모형화 및 예측 측면에서 보면, 우선 사망력지수의 경우 시계열 적용기간과

관계없이 동일한 적합모형(RWD)이 선택되었으나, 코호트지수의 경우 남자는 시계열의 적용기간과 관계없이 코호트지수의 모형은 ARIMA(3,1,0)이 가장 적합력이 우수한 것으로 나타났다. 그러나 여자의 경우 1976~2005년 시계열의 경우 ARIMA(2,1,1), 1983~2005년 시계열의 경우 ARIMA(0,1,0)가 우수한 것으로 나타나 시계열 적용기간에 따라 최적 모형이 다르게 나타나는 것을 알 수 있다. 이와 같이 시계열 적용기간에 따라 예측을 위한 적합모형이 다르게 선택되면 결과적으로 장래 기대수명도 차이가 발생하게 되어 결과에 대한 신뢰성에 문제가 될 수 있으므로 시계열 적용기간에 따른 사망력지수 및 코호트지수의 모형화 및 예측방법에 대한 충분한 검토가 필요하다.

마지막으로 기대수명에 대한 예측 적합력 측면에서 보면 남자의 경우 1976년부터 2005년까지의 시계열을 적용하고자 할 경우에는 LC 코호트효과 확장모형을 사용하는 것이 바람직해 보이며, 1983년부터 2005년까지의 시계열을 적용할 경우에는 LC모형을 선택하는 것이 바람직해 보인다. 여자의 경우는 어떤 시계열을 적용하든 LC 코호트효과 확장모형을 사용하는 것이 바람직해 보인다. 이는 시계열 적용기간에 따라 예측 적합력이 높은 모형이 다르게 나타나는 것으로 이에 대한 충분한 검토를 필요하다는 것을 시사하고 있다. 또한 모형과 시계열 적용기간을 선택할 때 사망력에 대한 모수 적합력이 우수한 모형이 우리가 최종적으로 추계하고자 하는 기대수명에 대한 예측 적합력이 우수한 모형과 일치하지 않는 경우도 종종 발생한다는 점을 유념할 필요가 있다.

미래에 대한 사망력을 정확하게 예측하는 것은 쉬운 일은 아니다. 아무리 국제적으로 우수한 모형도 우리나라 자료에는 적합이 잘 되지 않을 수 있으므로 추정모형 선택시 다양한 모형에 대한 검증작업이 충분하게 이루어져야 할 것이다. 이와 더불어 시계열의 적용기간에 따른 사망력의 예측결과와 이에 따른 기대수명의 차이가 크게 발생한다면 국가통계에 대한 신뢰도의 손상뿐만 아니라 국가적으로 막대한 재정 낭비를 초래할 수 있다. 따라서 보다 정확한 사망력 예측 및 기대수명 추계를 위해서 시계열 적용기간 선정은 무엇보다도 중요한 요소임을 유의해야 할 것으로 사료된다.

References

- Baek, J. S. and Jeong, M. O. (2012). *Study on Mortality Models for Population Projection*, Statistical Research Institute.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. and Balevich, I. (2007). A Quantitative comparison of stochastic mortality models using data from England and Wales and the US, Discussion Paper, Available at <http://www.ma.hw.ac.uk/~andrewc/papers/ajgc50.pdf>.
- 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.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D. and Epstein, D. (2011). Mortality density forecasts: An analysis of six stochastic mortality models, *Insurance: Mathematics and Economics*, **48**, 355-367.
- JPMorgan (2007). LifeMetrics Software user guide version 2.0.
- Jung, K. (2012). *Estimation of mortality rates at older ages and extension of the LC model with the cohort effect*, Ph. D. thesis, Sungkyunkwan University.
- 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.
- Lee, R. D. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications, *North American Actuarial Journal*, **4**, 80-93.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality, *Journal of American Statistical Association*, **87**, 659-675.
- 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.
- Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the LC model for mortality reduction factors, *Insurance: Mathematics and Economics*, **38**, 556-570.

시계열 적용기간에 따른 사망력 추정 및 예측결과 비교 - LC모형과 LC 코호트효과 확장모형을 중심으로 -

정규남^a · 백지선^b · 김동욱^{c,1}

^a통계청, ^b통계개발원, ^c성균관대학교 통계학과

(2013년 11월 8일 접수, 2013년 12월 18일 수정, 2013년 12월 23일 채택)

요약

최근 급격한 기대수명의 증가에 따라 미래 복지정책 등에 커다란 영향을 주는 장래 사망력의 정확한 예측은 중요한 이슈가 되고 있다. 사망력의 정확한 예측을 위하여 최적의 추정모형의 선택도 중요하지만 사망력에 대한 시계열 적용기간도 매우 중요한 이슈다. 이는 우리나라의 사망률 시계열이 짧고, 특히 1982년 이전 자료가 다소 불완전해서 이에 대한 고려가 필수적이기 때문이다. 본 논문에서는 우리나라 사망력 시계열을 기간에 따라 2개의 그룹(1976~2005년, 1983~2005년)으로 나누어서, 남녀별로 LC모형과 LC 코호트효과 확장모형에 대한 모수 추정값, 사망력지수와 코호트지수의 모형화 및 예측, 장래 기대수명의 예측 적합력을 각각 분석한 후 향후에 장래 기대수명 추계시 고려할 시사점을 제시하고자 한다.

주요용어: 시계열 적용기간, LC모형, LC 코호트효과 확장모형, 사망력, 기대수명.

¹교신저자: (110-745) 서울시 종로구 성균관로 25-2, 성균관대학교 통계학과, 교수. E-mail: dkim@skku.edu