연구보고서

장래인구추계 출산력·사망력 추계방법 개선 연구

최종보고서

2016. 10. 31.

김현식
계봉오
김현태

한국인구학회·통계청

<차 례>

제1장 연구 배경 및 내용 ····· 1
제1절 연구 배경 ····· 1
제2절 연구 내용 ····· 2
제3절 연구 방법 ····· 3
참고문헌 ····· 5

제2장 전문가 판단법을 활용한 출산력 추계 ····· 6
제1절 서 론 ····· 6
제2절 선행 연구 개관 ····· 7
제3절 출산력 추계 방법 ····· 9
제4절 전문가 조사 ····· 17
제5절 장래 출산율 추계 결과 ····· 31
제6절 논의 및 제언 ····· 36
참고문헌 ····· 38

제3장 혼인율과 혼인출산율 추이를 활용한 출산율 추계 방법 ····· 40
제1절 서 론: 근접요인 접근법을 활용한 출산율 추계의 필요성 ····· 40
제2절 근접요인 접근법을 활용한 연령별 출산율 추정 기본모형 ····· 42
제3절 장래 출산율 변화 추이에 대한 가정 도출 ····· 43
제4절 분석자료 ····· 48
제5절 1980-2014년 추세 분석 ····· 48
제6절 Lee-Carter 모형을 활용한 과거 변화추이 분석 ····· 59
제7절 장래 연령별 출산율 및 합계출산율 추계 ····· 71
제8절 결론 및 논의 ····· 87
참고문헌 ····· 89

제4장 사망률 개선의 교대현상을 반영한 장기 사망률 예측모형 ···· 96
제1절 서 론 ···· 96
제2절 기존의 사망률 예측모형들 ···· 97
제3절 사망률의 Rotation 문제 ···· 100
제4절 국내 사망률의 추정과 예측 ···· 109
제5절 Coherent rotation모형 ···· 120
제6절 맺음말 ···· 124
참고문헌 ···· 126

제 1 장

연구 배경 및 내용

김현식·계봉오·김현태

제1절 연구 배경

장래인구추계는 각종 연금이나 보험의 재정 추계에 기초자료를 제공하는 등 장기 국가발전계획 수립을 위해 활용되는 핵심자료이므로 정확하고 신뢰성 높은 추계 결과를 생산할 수 있는 방법에 대한 모색이 필요하다. 이런 맥락에서 다양한 인구추계기법이 발전되어 왔으며 국가마다 정부 내에 특정 기관을 선정하여 각국의 상황에 따라 다양한 추계방법을 적용하고 있으며 사회경제적, 그리고 인구학적 조건의 변화에 대응하기 위하여 지속적으로 추계방법을 개발하고 있는 것 또한 그리 놀라운 일은 아니다(이삼식·이지혜·최효진, 2013; 전광희·김태헌·조영태, 2005; Preston, Heuveline, & Guillot, 2001).

이런 점에서 한국의 저출산현상이 심화 및 유지되고 있으며 기대수명이 눈에 띄게 늘어나고 있는 상황에서 인구추계의 핵심적 요소라고 할 수 있는 출산율과 사망률 추계에 대한 선행 연구를 검토하여 현재 사용되고 있는 추계 방법을 개선할 필요가 있다. 예를 들어 우리나라의 출산력은 2001년 합계출산율 1.297을 기록한 이후 합계출산율 1.3을 넘어서지 못하는 수준에서 불규칙한 변동을 보이고 있으므로(통계청, 2016), 과거 시계열 자료만 활용하는 수학적 추계 모형만을 사용하는 것에는 한계가 있는 것으로 판단된다. 특히 출산력 추계에 있어 최근 국제적인 방법으로 널리 사용되고 있는 전문가 판단법에 대한 검토와, 출산에 영향을 주는 혼인상태 분포 및 혼인출산율을 활용한 추계 방법의 도입을 연구할 필요가 있다.

전문가 패널의 의견 교환을 통한 장래출산력 예측은 UN의 인구추계에 있어 참고 자

료로 사용되고 있으며, 캐나다, 영국, 프랑스에서도 적극적으로 받아들이고 있는 방법으로 과거 자료를 통한 단순 시계열 예측을 넘어설 수 있는 대안으로 널리 논의되고 있는 실정이다(2장 참조). 이러한 맥락에서 기존의 시계열 자료를 활용한 추계에 더해 전문가 판단법을 고려하는 추계 방법을 검토하여 복합적으로 활용함으로써 다양한 미래 출산력 시나리오를 그려볼 수 있을 것이다. 또한 출산관련 분야에 상당 기간 연구를 수행한 전문가들을 대상으로 장래출산율 예측에 대한 견해를 청취하여 추계에 반영함으로써 추계 오차를 줄일 수 있을 뿐 아니라 추계에 대한 신뢰도를 높일 수 있을 것으로 판단된다.

이에 더해 인구 고령화 진전으로 인해 사망력 추계의 정확성 제고가 요구되고 있다. 가장 좋은 예측은 조건부 예측이라는 통계학의 금언에 따라 출산의 결정 요인에 대한 이해를 바탕으로 이들 결정 요인에 따른 조건부 예측을 실시함으로써 출산율 예측의 정확성을 높이는 방안을 개발할 필요가 있다. 전통적으로 출산의 결정 요인은 사회경제적 요인과 인구학적인 요인으로 나누어서 살펴볼 수 있으며, 사회경제적 요인은 주기적인 변동폭을 예측하기 어려운 반면 혼인상태 분포 및 혼인출산율과 같은 인구학적인 요인들은 연령에 따라 일정한 패턴을 나타낸다. 따라서, 혼인상태 분포와 혼인출산율을 활용하면, 장래 출산율 추이를 보다 정확하게 예측할 수 있을 것이다. 이러한 인식에 근거하여 3장에서는 혼인율과 혼인출산율 추이를 활용한 출산율 추계 방법을 검토한다.

장래인구추계에서 사망력 역시 금융, 의료, 연금, 재정정책 수립 및 국민의 생활에 직접적인 영향을 미치는 중요한 또 다른 요인이다. 우리나라의 인구 고령화 속도는 세계적으로 전례가 없을 정도로 빨라, 특히 고연령에 대한 사망력의 정확한 추계가 어느 때보다 중요한 시점이라고 할 수 있다. 고연령의 사망력의 추정과 예측은 주어진 시점에서 연령별로 정확한 사망력이 추정되어야 할 뿐 아니라 시간이 흐름에 따라 빠른 속도로 이루어지고 있는 사망력의 개선까지 동시에 반영되어야 한다는 점에서 중요하다. 특히 기존 사망률 예측 모형이 전 연령대에서 유사한 정도의 사망률 개선이 이루어진다는 가정에 근거하고 있기 때문에 연령에 따라 차별적으로 나타나는 개선 속도를 반영하는 모형을 검토하고 제안하는 것이 필요하다. 이런 맥락에서 4장에서는 이러한 사망률 개선의 교대현상을 반영하는 장기사망률 예측모형을 연구한다.

제2절 연구 내용

이 장에 뒤이은 2장에서는 전문가 판단법에 근거한 합계출산율 추계 모형을 살펴본다. 문헌 연구를 통해 전문가의 추계에 근거한 장래출산율 추계 모형을 세 가지로 구분하고, 이 중 두 가지를 시산해 본다. 이러한 모형 추정에 필요한 자료를 구하기 위해 19

명의 인구학 전문가들을 대상으로 수집된 장래출산율 및 관련 설문 결과를 제시한다. 그리고 전문가들의 예측을 단순 평균하거나, 2040년과 2065년의 추계가 이항정규분포를 따른다는 가정 하에 모형화하거나, 베이지안 방법에 근거하여 모형화하는 세 가지 추계 모형을 사용하여 장래출산율을 추계할 것이다. 이러한 결과들을 2010년의 통계청 추계 결과와 2015년의 UN 추계결과와 비교한다.

뒤이은 3장에서는 혼인율과 혼인출산율 추이를 활용한 출산율 추계 방법을 살펴본다. 여기에서는 근접요인 접근법을 활용한 연령별 출산율 추정의 기본모형을 제시하고, 출산율의 반등을 모형화할 수 있는 이론적 논의들을 소개하고 있다. 그리고 1980-2014년 자료를 활용한 과거 추세 분석결과를 제시한 후 Lee-Carter 모형을 활용하여 과거 추세를 요약하고, 장래 출산율 추계를 위한 기초분석 결과를 제시한다. 그리고 마지막으로 출산율 반등에 대한 이론적 모형과 경험적 분석결과를 활용하여 2015-2040년의 연령별 출산율 추계 결과를 몇 가지 가정 하에서 제시한다.

새로운 사망률 추계 모형을 제시하는 4장은, 전통적인 사망률모형은 특정한 시간이 주어졌을 때 각 연령에 대한 사망률 혹은 사망력을 계산해 내는 데 중점을 두었으나, 현대에 이르러 미래의 인구추계를 위해 각 연령의 시간에 따른 개선(진전 혹은 감소라고도 표현함)을 동시에 모형화 하는데 더 큰 관심을 가지게 되었다는 관찰에서 시작한다. 하지만 많은 서구 국가들에서 최근의 사망률개선 속도가 유년층에서는 감소하고 고령층에서는 증가하는, 소위 사망률 개선의 교대(mortality decline rotation)패턴을 보이고 있는데 Lee-Carter(1992) 나 Li-Lee 모형(2005)은 이러한 교대패턴을 표현할 수 없어 장기 사망률예측에서 한계를 가진다고 알려져 있다. 이 장에서는 장기 사망률예측을 보다 안정적으로 추정할 수 있는 coherent logistic 모형을 제안하고 국내 사망률자료에 적합한 결과에 대해 논의한다.

제3절 연구 방법

각 장의 연구 대상이나 내용이 독립적이고 개별 연구자에 의해 진행되었기 때문에 각기 다른 방법이 사용되었다. 우선 2장에서 시도된 전문가 판단법에는 이메일을 통한 전문가 조사라는 서베이 조사 방법에 더해 베이지안 및 frequentist 관점의 통계적 방법을 통한 평균과 신뢰구간 추정 방법을 활용하였다. 특히 신뢰구간을 추정하면서 통상적인 95% 신뢰구간이 아닌 80% 신뢰구간을 사용한 이유는, 통상 인구학에서는 신뢰구간이 없이 고위, 중위, 저위 값을 추계하는데, 이는 특정한 신뢰구간을 의미를 지니지 않으며, 선행 연구에 따라 고위와 저위를 중위 값의 80%의 신뢰구간의 인식하였기 때문이다

(Billari, Graziani, & Melilli, 2012).

혼인률과 혼인출산율을 이용하여 장래 합계출산율을 예측하는 3장에서는 합계출산율을 연령별 혼인율과, 유배우출산율, 그리고 보정요인으로 분해하는 요인분해 방법이 그 뼈대를 이룬다. 이후 이들 세 요인들을 인구학의 사망률 추계에서 가장 널리 사용되는 Lee-Carter 방법(Lee & Cater, 1992)에 의해 특정한 기간까지 추계를 한 후, 다시 이들 요인을 결합하여 합계출산율을 계산하였다. 또한 다양한 시나리오 하에서 세 요인을 추계함으로써 더욱 현실을 반영하는 추계를 시산하고자 하는 노력을 기울였다.

장래 사망력 추계 모형을 연구하는 4장에서는 우선 Lee-Carter 모형과 그 유사 모형, 그리고 이에 더해 일본에서 사용하고 있는 Age-shifting 모형을 개괄한다. 하지만 이들 모형들은 연령에 따라 사망률이 차별적으로 하락한다는 현실을 반영하지 못한다는 점을 지적하면서, 이에 대한 대안으로 등장한 L-Lee-Gerland 모형을 소개한다. 하지만 이 또한 한국의 사망률 변화에 적합한 모형은 아니라는 점을 보여주면서 Logistic 모형을 제안할 것이다. 그리고 성별 모형을 결합하는 Li-Lee 모형처럼 Logistic 모형을 일반화하면서 Coherent logistic 모형을 제안한다. 이 모형으로 과거 자료를 추정하였을 때 나타날 수 있는 문제를 살펴보기 위해 잔차를 검토하면, 이 모형이 실질적인 대안이 될 수 있음을 시사한다.

참고문헌

전광희·김태헌·조영태. (2005). 장래인구추계를 위한 출산·사망 예측모형의 개발에 관한 연구. 통계청·인구학회.

이삼식·이지혜·최효진. (2013). 인구예측모형 국제비교 연구. 한국보건사회연구원.

통계청. 2016. 각종 통계 자료. kosis.kr

Billari, F. C., Graziani, R., & Melilli, E. (2012), Stochastic population forecasts based on conditional expert opinions. Journal of Royal Statistical Society, Series A., 175, Part 2, 491-511.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. Journal of the American statistical association, 87(419), 659–671.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. Demography, 42(3), 575–594.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). Demography: Measuring and modeling population processes. Malden, MA: Blackwell Publishers Inc.

제 2 장

전문가 판단법을 활용한 출산력 추계

김현식

제1절 서 론

전통적으로 인구추계는 코호트요인법(cohort component projection)으로 불리는 방법이 가장 널리 사용되었다(이삼식·이지혜·최효진, 2013; 전광희·김태헌·조영태, 2005; Preston, Heuveline, & Guillot, 2001). 이 방법을 간략하게 설명하자면, 각 코호트들의 출산율, 사망률, 이동률을 추계한 다음 각 코호트의 연령이 성숙해짐에 따라 코호트별 인구가 변해가는 과정을 합쳐 전체 인구의 변화를 추적하는 것이라고 할 수 있다. 그런 만큼 출산력 추계는 인구추계에 있어 중요한 근간을 이루어왔고, 출산력 추계방법에 대한 다양한 논의가 진행되어 왔으며 각 국가의 통계담당 행정부서는 각국의 특색에 맞는 출산력 추계 방법을 사용하고 있다.

예를 들어, 인구추계에 있어 가장 표준적인 방법을 개발하고 사용하는 UN의 경우, 2015년 인구추계에서 출산력을 추계하면서 전세계의 국가를 세 가지 유형으로 나누고 각 유형에 따라 다른 추계방법을 적용하였다(United Nations, 2015a). 고출산력 국가들은 특정한 모형을 설정하지 않았으나, 고출산에서 저출산으로 이동하는 국가들은 double-logistic function으로 모형화하였으며, 이 함수의 모수를 추정하기 위해 베이지안 위계 모형(Bayesian hierarchical model)을 적용하였다. 그리고 저출산 국가들의 출산력 추계는 1계 자기회귀 시계열모형(first-order auto-regressive time series model)을 사용하였으며, 역시 이 모형의 모수를 추정하기 위해서는 베이지안 위계 모형을 적용하였다. UN에서 베이지안 모형이 사용할 수 있는 이유 중의 하나는, 이 연구에서 보여줄 것

처럼, 베이지안 모형이 다양한 국가들의 경험을 서로 참조하여 평균을 하는 과정을 거치기 때문이다. 이에 반해 캐나다는 전문가 조사에 기반하여 향후 특정 수준의 합계출산율을 설정한 다음 Myrskyla와 동료들이 제안한 Lee-Carter 변형 모형을 사용하고 있다(Statistics Canada, 2015). 한국의 통계청의 경우 시계열회귀분석과 보정형 로그감마모형(adjusted generalized log gamma model)에 의해 코호트 출산율 및 평균출산연령을 예측한 다음 이를 특정 연도의 기간합계출산율로 작성하는 방법을 활용하고 있다(통계청, 2016).

하지만 최근 특정 수학적 통계 모형에 의존하는 출산력 추계보다는 전문가들의 견해를 반영하여 미래 출산력을 예측하는 전문가 판단법을 활용한 출산력 추계가 확산되고 있다. 일례로 최근의 European Demographic Data Sheet에 사용되는 출산력 추계는 Wittgenstein Center for Popoulation and Global Human Capital에서 수행한 전문가 조사를 기반으로 작성되었으며(Lutz, Butz, & KC, 2014), 앞서 언급한 캐나다의 경우도 전문가 조사에 기반하여 미래 출산력 수준을 미리 설정한다는 점에서 전문가 판단법을 활용한 추계라고 볼 수 있겠다. 영국의 경우도 추계 전문가 패널 회의를 통하여 향후 각 구성국가들의 미래 출산력을 전망한 후 이들을 합해 영국 전체의 출산률을 예측하는 방법을 활용하고 있다(Office for National Statistics, 2013).

이 장에서는 이러한 전문가 판단법을 활용한 출산력 추계 방법 중 대표적인 몇 가지를 살펴본 후, 향후 출산력 전망에 대한 전문가 의견 조사를 바탕으로, 전문가 판단법에 근거한 몇몇 출산력 추계를 시산한다. 특히 이 연구에서는 전문가들의 응답을 단순 평균하는 방법, 주관적 이변량 정규분포 모형, 그리고 주관적 베이지안 모형을 기술한 후 이러한 방법을 사용하여 2040년과 2065년의 합계출산율을 추계한다.

제2절 선행 연구 개관

인구추계에 있어 전문가 의견은 장래인구 시나리오를 작성하는데 있어 매우 중요한 영역을 담당해 왔다(e.g., Keyfitz, 1982; Sobotka, Zeman, & Basten, 2016). 예를 들어, UN에서는 미래 예측 시나리오를 전망하는 과정에서 전문가 의견을 청취하고 있으며, 캐나다, 영국, 프랑스의 출산력 예측에 전문가 패널의 의견을 청취하여 통계 모형으로 전망된 결과와 비교 및 수정하는 작업을 진행하고 있다(e.g., Office for National Statistics, 2013; Statistics Canada, 2015). 장래 추계를 위한 통계적 모형의 선택 또한 전문가 관점에 기반한 것이라는 점에서, 전문가 의견은 장래 인구 모수에 대한 시나리오 작성이라는 협소한 관점을 넘어서 광범위하게 사용되고 있다. 하지만 통계 기법의 선택

을 고려하는 전문가 판단법은 이 글의 범위를 뛰어 넘는 것이기에 이 글에서는 전문가들의 미래 출산력 예측 값에 기반한 장래 출산력 추계에 초점을 맞추어 논의를 진행하기로 한다.

전문가들의 장래 출산력에 대한 예측 값을 활용한 추계는 매우 다양한 방식으로 이루어졌으나 이들은 넓게 세 가지 방법으로 분류할 수 있을 것으로 판단된다. 첫째로 전문가들이 예상하는 예측값을 사용하여 하나의 시나리오를 구성하는 방법이 있으며, 이때 여러 전문가의 의견을 하나의 값으로 요약하기 위하여, 다른 요인을 고려한 가중평균이나 가중치를 주지 않은 평균을 사용할 수 있다. 이 글에서는 이 방법을 단순 예측 모형으로 부르기로 한다. 두 번째와 세 번째는 모의실험(simulation)을 통한 통계분석 기법을 사용한다는 점에서 동일하지만 한 방법은 전문가의 예측만을 활용해 추계 모형을 설정하지만, 다른 방법은 과거 자료에 기반한 전통적인 수리 추계 모형을 결합하여 사용한다는 점에서 차이점을 보인다. 이 글에서는 전자의 방법을 주관적 예측 모형이라고 부르고 후자의 방법을 결합 베이지안 모형이라고 명명한다.

1. 단순 예측 모형

단순 예측 모형은 베이지안 통계 모형을 사용하지 않고 전문가들의 예측 값을 하나의 요약값으로 제시한다는 점에서 다른 두 방법과 대조된다. 전문가 의견에 기반하여 장래 출산력을 예측한 대부분의 선행 연구들은 이 방법을 사용하였다. 예를 들어, 163명의 전문가들을 대상으로 전 세계 저출산 국가 31개 나라의 2030년과 2050년 기간합계출산율에 대한 예측 값을 조사한 Basten과 동료들(Figure 13 in p.67, 2013; 또한 Sobotka et al.(2016) 참조)은 예측 합계출산율과 80% 신뢰구간을 질문하였다. 예측 값들을 요약하면서, 이들은 각 나라별 인구를 가중값으로 주어 평균하여 합계출산율에 대한 전문가 예측 값을 구하였다. 이들이 제시한 80% 신뢰구간 또한 각 전문가들이 제시한 신뢰구간을 평균하였다는 점에서 단순 예측 모형을 활용한 것으로 평가할 수 있다.

2. 주관적 예측 모형

주관적 예측 모형은 합계출산율, 평균 출산연령, 남녀 기대수명, 남녀 순이동자 수에 대한 전문가들의 조건부 예측 값에 기반하여 장래인구를 추계한 Billari, Graziani, & Melilli(2012, 2014)의 방법을 일컫는다. 이들은 2012년 논문에서 동일한 시간 길이를 가진 특정 두 시기에 대한 예측 값을 제시하기를 요청하면서, 현재와 가까운 시기에는, 예를 들어 합계출산율 수준에 대한 중앙값(central value)과 90% 신뢰도를 가진 상위 값을

질문하였다. 이에 더해 가까운 시기에 합계출산율이 평균 수준의 값일 때와 상위 값일 때 먼 시기에 합계출산율의 중앙값과 상위 값을 예측하도록 요청하였다. 이러한 값들을 이용하여 특정 시기의 합계출산율에 대한 (조건부) 분산을 구하였으며, 다양한 인구학적 모수들의 값들이 다변량정규분포(multivariate normal distribution)를 이룬다는 조건하에 장래 인구학적 예측 값들을 추계한 후 코호트요인법으로 장래인구를 예측한다.

저자들은 2014년 논문에서 명확히 베이지안 관점을 취하면서 이전의 관점을 변화시킨다. 가장 커다란 차이는 이전 논문에서는 한 명의 전문가가 예측하는 평균과 분산을 이용하여 장래 출산율을 추계하였다면, 최근의 논문에서는 여러 명의 전문가가 제시한 추계값들을 통합하는 모형을 설정한다는 것에 있다. 이를 위해 2014년 논문에서 전문가 내 인구학적 모수들의 상관관계와 전문가 사이의 상관관계를 고려하면서, 모든 전문가들을 개별적으로 취급하는 것이 아니라 몇 개의 전문가 집단군으로 취급한다. 이런 접근으로 인해 각 전문가들의 인구학적 모수에 대한 예측 값이 다변량정규분포를 이루지만, 이 전문가들은 몇몇 집단으로 나눌 수 있고, 전체 예측을 이들 집단의 혼합으로 인식하는 혼합모형(mixture model)으로 가정하고 장래 추계를 진행한다. 또한 이전 연구에는 베이지안 통계기법의 핵심이라고 할 수 있는 사전 분포(prior distribution)가 빠져 있었는데 반해 최근 연구에서는 이에 대해 명확하게 밝히고 있으며, 다양한 방법 사이에서 얻어지는 결과를 베이지안 진단법(diagnostics)을 활용하여 비교하고 있다는 점에서 차이를 보이고 있다.

3. 결합 베이지안 모형

결합 베이지안 모형은 유럽의 7개국을 대상으로 입국 이민자수에 대한 전문가 예측 조사를 기반으로 장래 추계를 실시한 Bijak과 Wiśniowski(2010)의 연구를 일컫는다. 이들은, 이민자수에 대한 장래 추계 모형을 네 가지로 설정한 후 각 모형의 모수들에 대한 사전 분포를 전문가의 예측으로 구성하였다. 예를 들어, 향후 입국 이민자수의 추계를 보여주는 다양한 그림을 제시한 후 각 그림의 시나리오가 현실화될 가능성이 얼마나 되는지에 대한 응답을 활용하여 시계열 추계 모형의 자기회귀 모수(auto-regression parameters)에 대한 분포를 설정하였다.

제3절 출산력 추계 방법

1. 이변량 정규분포 모형

이 보고서에서는 2040년과 2065년에 대한 기간합계출산율의 중위값, 고위값, 저위값에 대한 질문과 정규분포 및 80% 신뢰구간에 대한 질문에 대한 응답을 활용하여 단순 예측 모형을 사용한 방법을 재생한다. 이에 더해 주관적 예측 모형 중 2012년의 모형을 이변량 정규분포 모형이라고 명명한 후 이를 시산해 보기로 한다. 앞서 밝힌 것처럼 이변량 정규분포 모형은 한 명의 전문가 추계를 기반으로 발전한 추계 모형이라고 할 수 있다. 하지만 본 연구의 전문가 조사에서는 다양한 전문가가 응답하였기 때문에 이들의 모든 응답을 좀더 잘 활용할 필요가 있다. 이를 위해 전문가가 독립적인 주관적 견해를 피력하였다는 가정 하에 각 전문가들의 관점 하에 이변량 정규분포 모형을 시산하여 이들을 단순 결합한 후 전체 추계를 실시하는 방법 또한 적용하였다.

향후 논의를 위해 이변량 정규분포 모형을 세밀히 살펴보기로 하자. 이 방법은 Billari, Graziani, & Melilli(2012)가 제안한 방법을 일컬으며 다음과 같은 절차를 거쳐 합계출산율을 추계한다.

우리가 추계하고자 하는 합계출산율을 F 로 나타내고, 추계 기간을 $[0, T]$ 로 나타내며, 이 기간은 두 개의 동일한 기간으로 이루어져 있어, 마지막 관측 연도, 첫 번째 마지막 연도와 두 번째 마지막 연도를 각각 t_0, t_1, t_2 로 나타내면, $t_0 = 0$, $t_1 = T/2$, $t_2 = T$ 이다. 물론, 전체 추계 기간을 꼭 두 기간으로 나눌 필요는 없고 임의적인 세 개 이상의 기간으로 나눌 수도 있으며, 이하에서 논의하는 방법을 적용하는데 전혀 문제는 없다.

추계하고자 하는 두 시기의 합계출산율 변수 벡터는 (F_{t_1}, F_{t_2})로 쓸 수 있으며, 이들 값을 구하면 세 시기의 합계출산율 F_{t_0}, F_{t_1}, F_{t_2} 과 선형보간법(linear interpolation)을 활용하여 $[0, T]$ 시기의 모든 합계출산율을 구할 수 있다. 모형을 단순화하기 위해 우리가 구하고자 하는 두 변수 (F_{t_1}, F_{t_2})이 이변량 정규분포(bivariate normal distribution)를 이루고 있다고 가정하면, 이 변수들의 결합분포를 특정하기 위해 다섯 개의 모수, 즉 두 개의 평균, 두 개의 분산, 그리고 두 변수의 상관계수를 추정하면 된다. 이들 다섯 개의 모수를 추정하는 절차는 다음과 같은 몇 가지 단계로 구분할 수 있다.

첫 번째 단계에서 전문가는 t_1 시점의 F 에 대한 중위값과 고위값을 기술하며, 이를 각각 c_1과 h_1 이라고 하자. 그러면 $E(F_{t_1}) = c_1$이고 $var(F_{t_1}) = \sigma_1^2$으로 표현할 수 있

다. 정규분포를 이루는 확률변수 F_{t_1} 의 평균으로 c_1 를 설정하고, 특정 분위값으로 h_1 을 설정하면 표준화를 통해 σ_1^2 을 다음과 같이 구할 수 있다.

$$\sigma_1^2 = \left(\frac{h_1 - c_1}{z_{1-\alpha}} \right)^2$$

위 식에서 $z_{1-\alpha}$ 는 표준정규분포에서 $1-\alpha$ 분위값을 의미한다. α를 어느 수준으로 설정할 것인가는 임의적인 것이지만 저자들은 0.25로 설정하고 있다. 하지만 이 연구에서는 전문가 조사 결과 합계출산율에 대한 예측이 정규분포를 따를 때 80% 신뢰구간 중 상한값이 일반적인 추계의 고위값과 유사하다는 발견에 근거하여, α를 0.1로 정의하였다. 한 가지 덧붙여 여기에서는 고위값에 대한 질문으로 σ_1^2 를 추정하고 있으나 저위값에 대한 질문으로 σ_1^2를 추정할 수도 있다.

두 번째 단계에서 전문가는 t_1 의 F 값이 중위값 혹은 고위값으로 고정되어 있을 때 t_2 시기 (조건부) F 의 중위값에 대해 응답한다. 다시 말해 다음의 질문에 대해 응답한다.

1. 전문가적 관점에서 보았을 때, t_1 시기 F 값이 c_1 이라면, t_2 시기 F 의 중위값은 얼마일 것이라고 생각하십니까?
2. 전문가적 관점에서 보았을 때, t_1 시기 F 값이 h_1 이라면, t_2 시기 F 의 중위값은 얼마일 것이라고 생각하십니까?

이러한 질문에 대한 응답은 이전 시점의 합계출산율이 중위값 혹은 고위값으로 주어졌을 때 F_{t_2} 의 조건부 평균을 기술하는 것으로 볼 수 있다. 즉, 첫 번째 질문에 대한 응답을 $c_2|c_1$이라고 하고, 두 번째 질문에 대한 응답을 $c_2|h_1$ 이라고 하면, $c_2|c_1 = E(F_{t_2}|F_{t_1} = c_1)$ 으로, 그리고 $c_2|h_1 = E(F_{t_2}|F_{t_1} = h_1)$ 으로 놓을 수 있다.

(F_{t_1}, F_{t_2})의 결합분포를 특정하기 위하여 F_{t_1} 이 주어졌을 때, F_{t_2} 의 조건부분포의 분산 $var(F_{t_2}|F_{t_1})$이 필요하기 때문에 다음의 질문을 했다.

3. 전문가적 관점에서 보았을 때, t_1 시기 F 값이 c_1 이라면, t_2 시기 F 의 고위값은 얼마일 것이라고 생각하십니까?

위 질문에 대한 응답을 $h_2|c_1$ 라고 하자. 또한 $var(F_{t_2}|F_{t_1}) = var(F_{t_2}|c_1) = \sigma_2^2|c_1$ 로 추정한다면, 위의 σ_1^2 에 관한 식을 활용하여, $\sigma_2^2|c_1$ 을 다음과 같이 구할 수 있다.

$$\sigma_2^2|c_1 = \left(\frac{h_2|c_1 - c_2|c_1}{z_{1-\alpha}} \right)^2$$

혹은 앞서 구한 F_{t_1} 의 주변분포(marginal distribution)의 분산 σ_1^2 가 F_{t_2} 의 조건부 분포와 동일하다고 가정하는 것, 즉 $var(F_{t_2}|F_{t_1}) = var(F_{t_1}) = \sigma_1^2$ 로 놓는 것도 한 가지 방법이다. 다만 이 경우 추계기간이 달라지는 것을 고려하기 위해 다음과 같은 일반적 공식을 사용하는 것이 바람직하다.

$$var(F_{t_2}|F_{t_1}) = \sigma_1^2 \left(\frac{t_2 - t_1}{t_1 - t_0} \right)^2$$

여기에서는 두 기간의 길이가 같기 때문에 두 번째 항은 1의 값을 갖고, $var(F_{t_2}|F_{t_1}) = \sigma_1^2$ 가 된다.

두 변수 (F_{t_1}, F_{t_2})이 이변량 정규분포(bivariate normal distribution)를 이루고 있다면, 이는 평균인 μ_1 과 μ_2, 분산인 σ_1^2 과 σ_2^2, 그리고 상관계수 ρ 의 다섯 개 모수에 의해 특정할 수 있다. 또한

$$E(F_{t_2}|F_{t_1} = x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

$$var(F_{t_2}|F_{t_1} = x) = (1 - \rho^2)\sigma_2^2$$

이므로 앞서 구한 전문가 의견을 이용하여 다음과 같은 결과를 도출할 수 있다.

1. $\mu_1 = E(F_{t_1}) = c_1$
2. $\mu_2 = E(F_{t_2} | F_{t_1} = \mu_1) = c_2 | c_1$
3. $\sigma_1^2 = \left(\frac{h_1 - c_1}{z_{1-\alpha}} \right)^2$

위의 조건부평균과 조건부분산에 관한 식을 활용하면,

$$c_2|h_1 = E(F_{t_2} | F_{t_1} = h_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(h_1 - \mu_1) = c_2|c_1 + z_{1-\alpha}\rho\sigma_2$$

$$\left(\frac{h_2|c_1 - c_2|c_1}{z_{1-\alpha}} \right)^2 = (1-\rho^2)\sigma_2^2$$

이 두 식을 풀면,

4. $\sigma_2^2 = \frac{(h_2|c_1 - c_2|c_1)^2}{z_{1-\alpha}^2} + \frac{(c_2|h_1 - c_2|c_1)^2}{z_{1-\alpha}^2}$
5. $\rho^2 = \frac{(c_2|h_1 - c_2|c_1)^2}{(h_2|c_1 - c_2|c_1)^2 + (c_2|h_1 - c_2|c_1)^2}.$

만약 $var(F_{t_2} | F_{t_1}) = \sigma_1^2$ 라는 가정 하에 추계한다면,

4-1. $\sigma_2^2 = \frac{(h_1 - c_1)^2}{z_{1-\alpha}^2} + \frac{(c_2|h_1 - c_2|c_1)^2}{z_{1-\alpha}^2}$

5-1. $\rho^2 = \frac{(c_2|h_1 - c_2|c_1)^2}{(h_1 - c_1)^2} \left(1 + \frac{(c_2|h_1 - c_2|c_1)^2}{(h_1 - c_1)^2} \right)^{-1}$

이 보고서에서는 이러한 추계 방법을 다음의 다섯 가지 자료에 적용해 보았다. 첫 번째로, 통계청의 2010년 추계 자료에 적용해 보았다. 즉 통계청이 하나의 전문가라고 설정한 후 이 전문가의 견해에 따른 합계출산율 가정에 위에서 기술한 방법론을 적용하여 2060년까지 합계출산율을 추계하였다. 이때, 다음과 같은 가정을 활용하였다.

1. c_1 은 2035년 중위값이다.
2. $\mu_2 = c_2 | c_1$ 는 2060년 중위값이다(통계청에서는 2045년 이후 중위, 고위, 저위값이 모두 같은 것으로 가정한다).
3. $\sigma_1^2 = \left(\frac{h_1 - c_1}{z_{1-\alpha}} \right)^2$ 로 구하며, h_1 은 2035년 고위값이고, $\alpha = 0.1$로 정의한다.
4. 4-1과 5-1을 활용하여 σ_2^2과 ρ를 추계하며, $c_2 | h_1$ 은 2060년의 고위값과 같다.
5. $var(F_{t_2} | F_{t_1}) = \sigma_1^2 \left(\frac{t_2 - t_1}{t_1 - t_0} \right)^2$

둘째, 전문가 조사에 대한 중위 및 고위값에 대한 응답 평균을 활용하여 2065년까지 합계출산율을 추계하였다. 이때 중위, 고위값에 대한 조건부 응답과 1-5의 식을 활용한다.

셋째, 두 번째의 방법과 동일하지만 모든 응답자들의 평균값을 사용하는 것이 아니라 각 응답자들이 독립적인 분포를 이룬다는 가정 하에 각 응답자의 분포를 구한다음 이 분포를 모두 더하여 전체적인 분포를 도출하였다.

넷째, 전문가 조사에서 평균과 80% 신뢰구간으로서의 고위값에 대한 응답 평균을 활용하여 2065년까지 합계출산율을 추계하였다. 또한 조건부 응답과 식 1-5를 사용하였다.

다섯째, 네 번째의 방법과 동일하지만 모든 응답자들의 평균값을 사용하는 것이 아니라 각 응답자의 분포를 구한다음 이 분포를 모두 더하여 전체적인 분포를 도출하였다.

2. 주관적 베이지안 예측 모형

앞 부분에서 정의한 것처럼 추계하고자 하는 두 시기의 합계출산율 변수 벡터를 (F_{t_1}, F_{t_2})로 나타내고, 두 변수의 평균과 공분산 행렬을 각각 $\mu = (\mu_1, \mu_2)$,

$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}$ 로 나타내자. 그리고 전문가들의 응답을 $y = (c_1, c_2)$라고 표기하자. 베이지안 모형에서는 전문가들의 응답 y를 μ와 Σ에 대한 사전분포가 주어졌을 때 나타나는 하나의 우도(likelihood)로 모형화하며, 실제 이 자료들을 본 후 μ와 Σ가 어떤 분포를 이루는가에 관심을 가진다. 다시 말해, 특정한 변수 x의 확률밀도함수를 $f(x)$로 나타냈을 때, 베이지안 모형에서는 $f(\mu, \Sigma|y)$에 관심을 가지는데, 이를 사후 분포(posterior distribution)라고 한다. 이를 구하기 위해 다음과 같은 관계를 활용한다.

$$f(\mu, \Sigma|y) = \frac{f(\mu, \Sigma, y)}{f(y)} = \frac{f(y|\mu, \Sigma) f(\mu, \Sigma)}{f(y)} \propto f(y|\mu, \Sigma) f(\mu, \Sigma)$$

위의 식에서 세 번째 항까지는 베이즈 정리라고 불리는 것으로 결합 분포와 조건부 분포의 관계로부터 도출되었으며 이들 사이에는 정확한 일치관계가 존재한다. 하지만 마지막 수식에서 $\propto$ 기호는 비례적(proportional)이라는 의미를 가지는 것으로, $f(y)$는 우리가 관심을 가지고 있는 모수들인 μ와 Σ를 포함하고 있지 않기 때문에 하나의 고정된 값으로 볼 수 있으므로, 이들 모수를 포함하고 있는 마지막 두 분포의 곱에 비례적이라고 표현할 수 있다. 마지막에 있는 두 분포의 결합을 구한 후 각 모수들의 확률을 더하여, 그 값을 각 모수들의 값으로 나누어주면 전체 합이 1이 되도록 쉽게 만들 수 있는데, 이를 normalize라고 하며, 이렇게 함으로써 $f(y)$를 고려해야 하는 번거로움을 피할 수 있다.

위의 식에서 마지막에 있는 분포 $f(\mu, \Sigma)$를 흔히 사전 분포(prior distribution)라고 하며, 이는 어떤 특정한 사건, 즉 현재의 맥락에서 전문가들의 견해를 보기 전의 미래 합계출산율에 대한 믿음이다. $f(\mu, \Sigma)$는 μ와 Σ에 대한 결합 분포(joint distribution)이기 때문에, 이를 더 분해하여 $f(\mu, \Sigma) = f(\mu|\Sigma) f(\Sigma)$로 쓸 수 있다.

하지만 우리의 관심은 μ와 Σ의 사후 결합 분포, 즉 $f(\mu, \Sigma|y)$에 있다기보다는 자료가 주어졌을 때 평균의 사후 주변 분포(marginal posterior distribution)인 $f(\mu|y)$에 있다. 이를 구하기 위해서는 위에서 도출한 결합 사후 분포(joint posterior distribution)를 Σ에 대해 적분을 한다. 즉, 다음과 같은 식을 활용한다.

$$f(\mu|y) = \int f(\mu, \Sigma|y) d\Sigma = \int f(\mu|\Sigma, y) f(\Sigma|y) d\Sigma$$

통상 위에서 도출한 결합 사후 분포를 Σ에 대해 적분하는 것은 수학적으로 다루기 어려운 일이기 때문에, 마지막 항에 있는 결과를 활용하여 무작위 변수 추출을 통한 모의실험(simulation)을 실시하게 된다. 즉 $f(\Sigma|y)$를 추출한 후 여기서 추출한 Σ에 기반하여 $f(\mu|\Sigma,y)$를 추출하여 결합 사후 분포인 $f(\mu,\Sigma|y)$를 구한다.

이 보고서에서는 표준적인 베이지안 분석을 따라 다음의 분포를 가정한다.

$$y|\mu,\Sigma \sim N(\mu,\Sigma)$$
$$\Sigma \sim IW(\Lambda_0^{-1},\nu_0)$$
$$\mu|\Sigma \sim N(\mu_0,\Sigma/\kappa_0)$$

위 식에서 $N(\mu,\Sigma)$ 는 평균 μ 와 표준편차 Σ 를 가진 정규분포를 뜻하며, $IW(\Lambda_0^{-1},\nu_0)$는 너비 행렬(scale matrix) Λ_0과 ν_0 자유도를 지닌 inverse-Wishart 분포를 뜻한다. 여기에서 사전 분포를 위와 같이 가정한 이유는 이들이 우도에 대한 동형 분포(conjugate distribution)이며, 이를 사용했을 때 사후 분포가 잘 알려져 있기 때문이다. 즉 위와 같은 분포를 가정하게 되면 μ의 사후 주변 분포는 다음과 같다는 것이 알려져 있다(Gelman et al., 2004).

$$\mu|y \sim t_{\nu_n-d+1}(\mu_n,\Lambda_n/(\kappa_n(\nu_n-d+1)))$$

여기서

$$\mu_n=\frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{y}$$
$$\kappa_n=\kappa_0+n$$
$$\nu_n=\nu_0+n$$
$$\Lambda_n=\Lambda_0+S+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)(\bar{y}-\mu_0)^T$$

$\bar{y}$는 전문가들의 예측 평균이며 $S=\sum(y-\bar{y})(y-\bar{y})^T$이고, n은 응답 전문가 수를 나타낸다. 또한 d는 μ의 차원을 의미한다.

마지막으로 사전 분포에서 가정한 상위 모수들(hyperparameters)－Λ_0, ν_0, μ_0, κ_0－를 결정할 필요가 있다. Λ_0는 출산율의 역사적 변화로부터 과거로 소급하여 구했다. 조금 더 세부적으로 말해서, 2015, 1990, 1970년의 출산율을 활용하여 2015년과 1990년의 출산율을 하나의 변수로, 그리고 1990년과 1970년의 출산율을 다른 하나의 변수로 취급한 후 두 변수의 공분산을 구하여 Λ_0로 설정하였다. ν_0는 $d-1$보다 큰 실수로 설정하면 되는데, 사전 분포의 영향을 최소화하기 위해 $d-1$보다 크지만 가장 작은 정수인 2로 설정하였다. μ_0는 2015년 출산율로 설정하였다. κ_0는 사전 분포에 대한 정보를 얻기 위해 얼마나 많은 조사가 있었는지를 나타내는 모수라고 할 수 있는데, 이 값이 크면 클수록 분산의 정도가 작아지게 된다. 이런 맥락에서 장래인구추계는 여러 번 있었으나 참고한 추계는 2010년의 것 하나뿐이고, 출산율 추계에 대한 불확실성을 나타내는 분산의 크기를 줄이지 않기 위해 이 보고서에서는 1로 설정하였다.

제4절 전문가 조사

1. 전문가 조사 개요

출산력 추계를 위한 전문가 조사를 위해 통계청의 인구 추계 전문가들과 전문가 명단을 작성하였으며, 설문지를 개발하였다. 인구학회에서 활발하게 활동하고 있는 출산력관련 전문가들에 더해, 2016년과 2015년 사이에 한국의 대표적 인구학 논문집인 <한국인구학>에 출산관련 논문을 게재한 저자를 대상으로 전문가 명단을 작성하였다. 이렇게 만들어진 전문가 명단은 학계를 과다대표하는 한편 정부나 정부관련 연구기관에 종사하는 전문가들을 과소대표할 것으로 우려하여 통계청 인구 추계 담당자들이 확보하고 있는 전문가 명단을 조사에 반영하였다. 하지만 본 연구의 연구진과 통계청 인구 추계 담당자들은 조사 대상에서 제외되었음을 밝힌다.

전문가 조사는 2016년 8월 2일부터 12일까지 2주에 걸쳐 실시되었다. 인구학 및 출산관련 분야에 종사하고 있는 전문가 31명에게 8월 2일 첫 번째 이메일을 보내어 조사 개요에 대한 설명을 담고 있는 설문지를 보내어 응답을 요청하였다. 이어 8월 2일까지 응답을 하지 않은 전문가들에 한하여 8월 8일 재차 설문지를 첨부하여 응답을 요청하였다. 이러한 방식으로 응답을 요청한 결과 19명의 전문가가 조사에 응하였으며, 두 명의 응답자를 제외한 응답자들이 모든 설문 문항에 자신의 견해를 제시하였다.

이들 두 명의 응답자 중 한 명은 설문지의 2.3번 문항인 평균출산연령에 관한 질문 중 고위와 저위값에 대해 응답하지 않았고, 다른 한 명은 4.1번과 4.2번에 해당하는 정규분포 가정 함계출산율 및 조건부 합계출산율 예측에만 응답하지 않았다. 이렇듯 설문지에 견해를 제시하지 않은 응답자들도 소수의 항목에만 응답하지 않았기 때문에, 이 보고서에서는 문항별로 모든 응답자들의 응답을 분석한다. 다시 말해 위에서 제시한 문항에 대한 결과는 그 응답자의 수가 18이 아니라 17이지만 여타의 문항에 대한 응답자 수는 18임을 밝힌다.

2. 응답자의 일반적 특성

다음의 두 표는 응답자들의 일반적 특성과 인구학관련 분야에서 자신의 전문성에 대한 스스로의 평가를 보여주고 있다. 성별에서는 단 세 명의 여성만이 포함되어 있어 여성의 과소대표 현상이 있음을 알 수 있다. 현재 몸담고 있는 기관을 살펴보면 대학이 73.3%로 가장 높으며 정부출연 연구기관이 15.8%로 그 뒤를 잇고 있다. 연령대를 보면 40대가 가장 많은 부분을 이루고 있으며, 그 다음이 30대이지만 50대와 60대도 적지 않은 부분을 차지하고 있다. 인구관련 분야에서의 연구 경력은 평균 16.26년으로 나타나고 있으며, 경험이 가장 적은 전문가는 3년이라고 응답한 반면, 경험이 가장 많은 전문가는 35년에 이르고 있다.

〈표 2-1〉 응답자의 일반적 특성

성별			일하고 있는 기관		
	빈도	백분율		빈도	백분율
남성	16	84.2	대학	14	73.7
여성	3	15.8	정부출연 연구기관	3	15.8
합계	19	100.0	정부 부처	1	5.3
			그 외	1	5.3
			합계	19	100.0

연령대			연구 경력(년)	
30대	5	26.3	평균	16.26
40대	7	36.8	최솟값	3
50대	3	15.8	중간값	13
60대 이상	4	21.1	최댓값	35
합계	19	100.0		

다음의 표는 전문가들의 전문성에 대한 자신의 평가를 담고 있다.

〈표 2-2〉 영역별 전문성에 대한 평가

	출산		사망		인구이동		인구추계	
	빈도	백분율	빈도	백분율	빈도	백분율	빈도	백분율
높다	12	63.2	6	31.6	6	31.6	7	36.8
중간	5	26.3	11	57.9	7	36.8	8	42.1
낮다	2	10.5	0	0.0	4	21.1	1	5.3
경험이 없다	0	0.0	2	10.5	2	10.5	3	15.8
합계	19	100.0	19	100.0	19	100.0	19	100.0

두 명의 전문가를 제외하고 출산에 대한 자신의 전문성이 중간 이상이라고 평가하였으며, 경험이 없다는 응답은 없었다. 이러한 응답은 조사의 대상이 주로 출산관련 분야에 종사하는 전문가들로 이루어져 있다는 사실을 반영한 것으로 풀이된다. 또한 절대 다수의 전문가들이 사망과 인구이동, 인구추계에 있어서 자신의 전문성이 중간 이상이라고 응답하고 있다는 것도 주목할 만하다. 이러한 전문성 평가 분포는 인구학관련 연구

경력이 늘어나면서 인구학의 한 분야에만 종사하기 보다는 다양한 영역에서 경험을 넓혀가는 한국의 연구 환경을 보여주는 것으로 판단된다.

3. 전문가 조사 결과

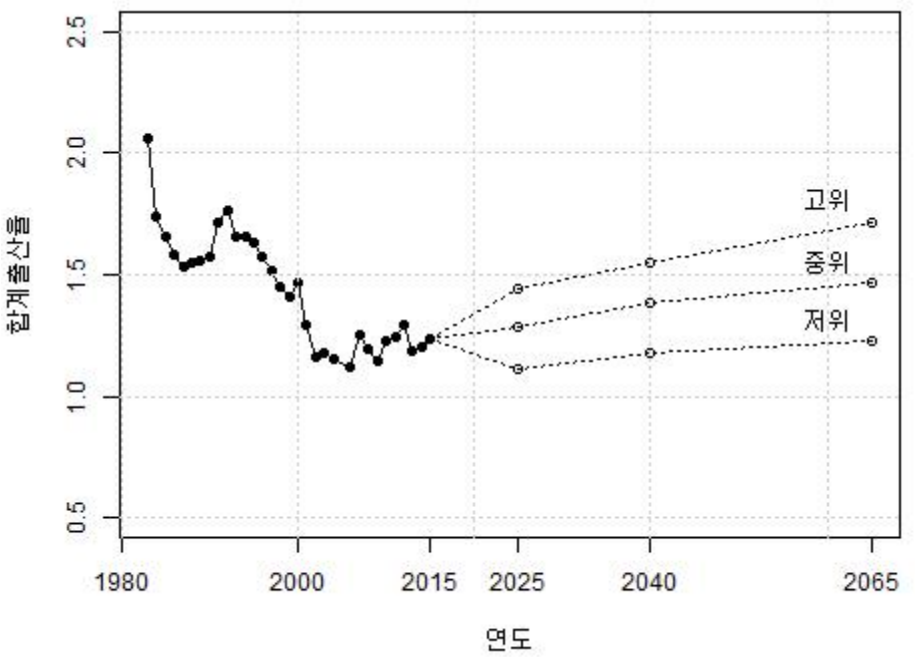
다음의 표는 2025년, 2040년, 2065년의 합계출산율 중위값, 고위값 및 저위값에 대한 전문가 추계 분포를 제시하고 있다.

〈표 2-3〉 합계출산율 추계 결과

		중위값	고위값	저위값
2025년 (2015년의 10년 후)	평균	1.28	1.45	1.11
	표준편차	0.11	0.14	0.12
	최솟값	1.10	1.20	0.90
	중간값	1.30	1.40	1.12
	최댓값	1.50	1.70	1.30
2040년 (2015년의 25년 후)	평균	1.39	1.55	1.18
	표준편차	0.22	0.24	0.21
	최솟값	1.10	1.19	0.90
	중간값	1.40	1.55	1.19
	최댓값	2.00	2.20	1.80
2065년 (2015년의 50년 후)	평균	1.46	1.71	1.23
	표준편차	0.29	0.36	0.29
	최솟값	0.90	1.10	0.70
	중간값	1.50	1.70	1.20
	최댓값	2.20	2.50	2.00

위 표를 보다 쉽게 보기 위해 아래의 그림에서는 각 년도의 중위값, 고위값 및 저위값 평균을 연결한 그림을 제시하였다.

〈그림 2-1〉 합계출산율 추계 결과



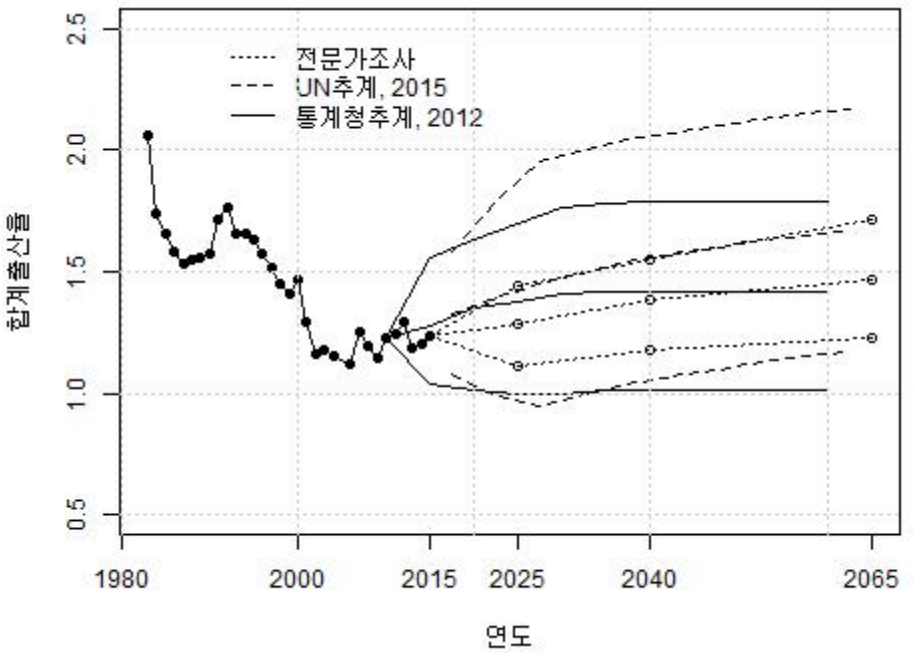
위 그림을 보면 전문가들은 평균적으로 장래 합계출산율이 최근의 초저출산율에서 점차 상승할 것으로 예측하고 있다는 점을 알 수 있다. 예를 들어, 중위값을 보면 2040년이 되면 평균 1.39를 기록해 1.3인 초저출산율을 탈피할 수 있을 것으로 추계하였다. 그리고 그 이후로도 꾸준히 상승할 것이지만, 그 기울기 값이 이전보다 작은 것으로 미루어, 증가 속도는 떨어질 것으로 판단하고 있다. 특히 2065년이 되어도 고위값의 평균이 1.71로 대체출산율이라고 일컬어지는 2.1에 미치지 못할 것으로 예측하고 있다. 물론, <표 2-3>의 각 년도 중위값 중 최솟값이 보여주는 것처럼 장래 출산율이 떨어질 것이라고 예측하는 전문가도 있었기 때문에 이러한 출산율 상승 예측은 평균적 예측이라는 점을 강조할 필요가 있다.

전문가 조사에서는 전문가들의 예측에 대해 간략한 이유를 제시하도록 하였는데, 여러 전문가들이 "템포 효과의 소멸"과 "출산장려정책의 효과"를 출산율 상승의 동력으로 지목하였다. 또한 저출산 세대의 노동시장 진출은 노동력 공급의 감소를 의미하고 이들의 상대적 경제적 지위의 안정화로 인해 현재 청년 세대보다는 많은 자녀 출산이 이루어질 것이라는 전망도 있었다. 그럼에도 불구하고, 출산력 상승이 크지 않거나 현재 수준을 유지할 것이라는 전망의 원인으로는 출산력 장려정책의 효과에 대한 의구심, 여전히 사회에 만연한 성역할 구분에 따른 여성의 부담, 출산 및 가족 가치관의 변화, 젊은

층의 경쟁 강화 및 실업으로 인한 사회경제적 어려움, 양육 및 교육에 대한 부담, 미혼의 증가와 초혼 연령의 상승 등을 거론하였다.

전문가 조사에 기반한 추계가 다른 추계와 어느 정도의 차이를 보이는가 살펴보기 위하여 다음의 그림에서는 2015년 UN의 추계(United Nations, 2015b)와 2012년 통계청 추계(통계청, 2016b)를 본 연구의 전문가 조사 기반 추계와 비교하였다.

〈그림 2-2〉 전문가조사 추계와 다른 추계의 비교



위 그림을 보면 전문가 추계는 장래의 출산율이 상승할 것이라는 점에서 여타의 추계와 유사하지만, 고위값과 저위값의 차이가 상대적으로 크지 않다는 점에서, 전문가들의 지나친 자신감 경향을 반영하고 있는 것으로 판단된다. UN의 중위값이 전문가의 고위값과 거의 유사하다는 점도 매우 특징적이며, UN의 저위값이 전문가의 저위값보다 낮아, 전문가의 고위값과 중위, 그리고 저위값 모두가 UN의 중위와 저위값 사이에 위치하는 것을 볼 수 있다. 이에 반해 통계청의 2012년 추계는 단기 미래에는 출산율이 상승하다 장기적으로 일정 수준을 유지할 것으로 전망하였다는 점에서, 지속적인 상승을 예측한 전문가 조사와 차이를 보이고 있다.

<표 2-3>에서 발견할 수 있는 추계 경향 중의 하나는 먼 미래로 갈수록 예측의 불확

실성이 커지기 때문에 분산의 정도가 커지는 것을 볼 수 있다는 점이다. 즉 중위, 고위, 저위값 모두에서 먼 미래로 갈수록 표준편차가 커지며, 또한 최댓값과 최솟값의 범위 또한 커지고 있다. <표 2-3>과 <그림 2-1>에서 발견할 수 있는 또 하나 언급할 만한 특징은 중위값을 중심으로 저위값과 고위값이 대칭을 이루는 것이 아니라 중위값에서 저위값의 차이가 중위값에서 고위값의 차이보다 더 크다는 점이다. 이러한 경향은 합계출산율이 상승하여도 많이 상승하지는 않을 것이라는 비관적인 전망이 우세하기 때문인 것으로 풀이된다.

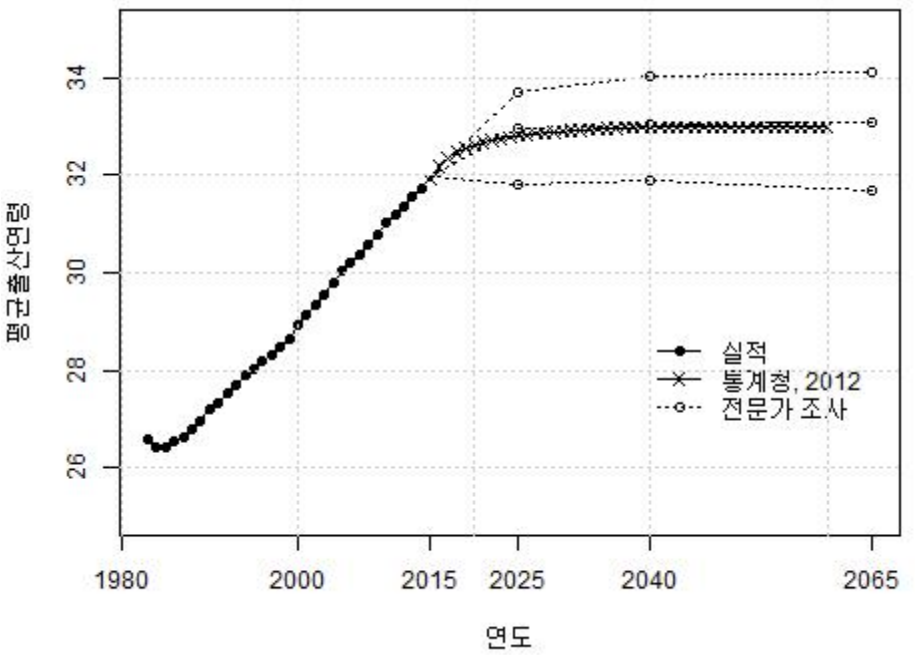
다음의 표는 평균출산연령에 대한 전문가 추계 분포를 보여주고 있다.

〈표 2-4〉 평균출산연령 추계 결과

		중위값	고위값	저위값
2025년 (2015년의 10년 후)	평균	32.95	33.72	31.80
	표준편차	0.79	1.23	1.16
	최솟값	31.00	31.30	29.00
	중간값	33.00	34.00	32.00
	최댓값	34.38	36.00	33.62
2040년 (2015년의 25년 후)	평균	33.04	34.02	31.88
	표준편차	1.83	2.29	2.03
	최솟값	30.00	30.00	28.00
	중간값	33.00	34.12	32.00
	최댓값	37.55	38.63	36.47
2065년 (2015년의 50년 후)	평균	33.10	34.12	31.70
	표준편차	3.51	4.20	3.66
	최솟값	28.00	28.00	26.00
	중간값	33.00	34.43	32.00
	최댓값	42.83	44.45	41.23

다음 그림은 위의 표에서 제시한 중위값, 고위값, 저위값의 평균을 그림으로 보여주고 있다. <그림 2-3>에서는 비교의 목적 아래 위의 추계값 평균에 더하여 2012년 통계청 추계값 또한 제시하였다.

〈그림 2-3〉 평균출산연령 추계



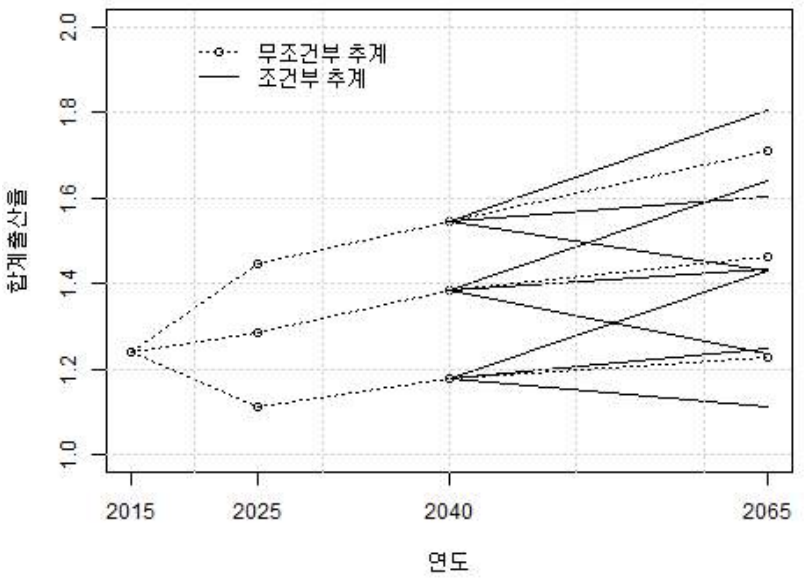
위 그림을 보면 평균출산연령이 향후 10년 후에 33세 정도로 올랐다가 이후 안정화 될 것으로 예측하고 있는 것으로 나타난다. 또한 중위값 추계의 평균이 2012년의 통계청 추계보다 약간 높지만 매우 유사한 것을 볼 수 있다.

다음의 <표 2-5>는 조건부 합계출산율 추계에 대한 기술 통계를 제시하고 있다. 조건부 추계는 "예를 들어, 2040년 합계출산율이 (귀하가 제시한) 중위에 도달했을 때, 2065년 합계출산율에 대한 중위, 고위, 저위값이 얼마인지"를 물었다. 그 다음의 그림은 <그림 2-1>에서 제시한 무조건부 추계에 더하여(점선과 빈 원), 조건부 추계의 평균값을 제시하였다. 이 그림에서 가장 흥미로운 점은 2040년 고위에 도달했다고 가정하면 2065년의 중위값이 무조건부 2065년 중위값보다 상당히 낮다는 점이다. 이에 반해 2040년의 다른 두 점에서 2065년으로 가는 조건부 중위값은 무조건부 중위값과 유사한 것을 볼 수 있다. 이러한 응답 유형은 합계출산율이 상승한다고 하더라도 지속적으로 가파르게 상승하지는 않을 것이라는 예측 때문인 것으로 풀이된다.

〈표 2-5〉 조건부 합계출산율 추계

2040년 합계출산율 실현값		2065년 추계값		
		중위값	고위값	저위값
중위값	평균	1.44	1.64	1.24
	표준편차	0.23	0.29	0.20
	최솟값	1.05	1.09	0.90
	중간값	1.50	1.70	1.30
	최댓값	1.82	2.10	1.63
고위값	평균	1.60	1.81	1.43
	표준편차	0.24	0.31	0.23
	최솟값	1.09	1.14	1.05
	중간값	1.60	1.80	1.38
	최댓값	2.03	2.23	1.87
저위값	평균	1.25	1.43	1.11
	표준편차	0.24	0.28	0.23
	최솟값	0.90	0.96	0.70
	중간값	1.25	1.45	1.12
	최댓값	1.65	1.88	1.56

〈그림 2-4〉 조건부 합계출산율



지금까지의 설문에서 중위 및 고위와 저위는 통계청에서 사용하는 의미로 사용하였다. 통계청에 따르면 중위는 현재의 추세에 따라 인구가 성장할 경우를, 고위는 가장 많이 성장할 경우를, 저위는 가장 낮은 성장을 할 경우를 뜻한다. 전문가 조사에서는 중위와 고위 및 저위값을 정규분포와 80% 신뢰구간으로 파악할 때, 그 추계값을 제시하도록 묻는 문항을 포함하였다. 즉 장래의 합계출산율을 추계할 때, 추계값은 정규분포를 이루고, 중위값은 평균을 의미하며 고위 및 저위값은 80% 신뢰구간의 상하한값을 의미한다는 가정 하에, 중위값 및 고위값을 답하도록 유도하였다. 다음의 <표 2-6>은 이에 대한 응답의 결과 분포를 보여주고 있다.

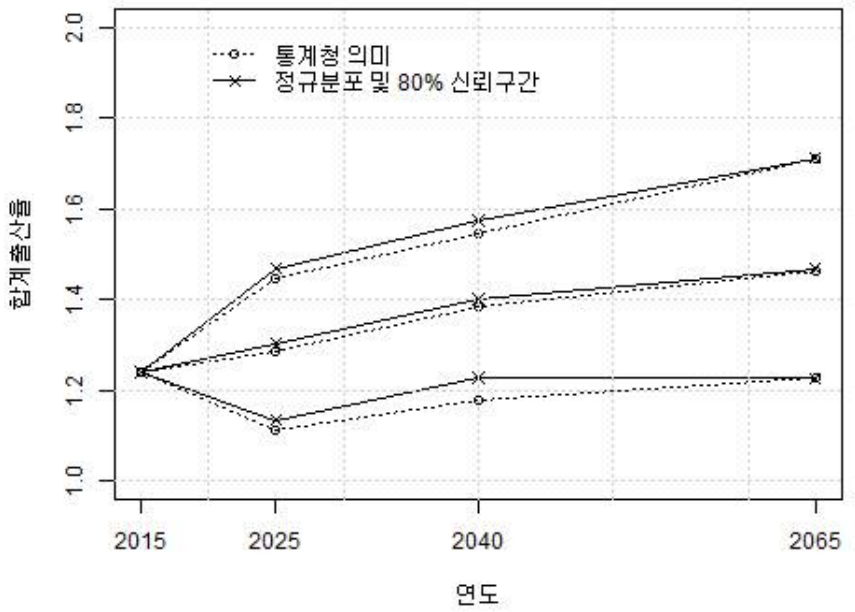
〈표 2-6〉 정규분포 가정 합계출산율 추계 결과

		중위값	고위값
2025년 (2015년의 10년 후)	평균	1.30	1.47
	표준편차	0.11	0.15
	최솟값	1.16	1.18
	중간값	1.30	1.45
	최댓값	1.60	1.80
2040년 (2015년의 25년 후)	평균	1.40	1.57
	표준편차	0.17	0.22
	최솟값	1.10	1.17
	중간값	1.41	1.60
	최댓값	1.80	2.00
2065년 (2015년의 50년 후)	평균	1.47	1.71
	표준편차	0.30	0.36
	최솟값	0.90	1.03
	중간값	1.50	1.77
	최댓값	2.20	2.40

다음의 그림에서는 비교를 위하여 앞서 제시한 <표 2-1>의 통계청에서 사용하는 의미를 차용한 중위, 고위, 저위값에 더하여 <표 2-6>의 결과를 함께 제시하고 있다. 다만 정규분포적 의미에서 하한값을 직접적으로 묻지는 않았으나, 좌우 대칭이 되는 정규분포의 특성을 이용하여 하한값은 평균에서 상한값과의 차이를 빼주어 계산하였다. 이 그림에서는 합계출산율의 중위, 고위, 저위값이 통계청에서 사용하는 의미로 질문을 하든, 정규분포와 80% 신뢰구간으로 묻든 그리 큰 차이가 나지 않는다는 것을 알 수 있다. 다만 정규분포의 의미에서 질문한 평균과 상하한 값들이 중위, 저위, 고위값보다는 약간

높은 것을 발견하게 된다. 그럼에도 불구하고 그 차이는 매우 미미한 것으로 판단된다.

〈그림 2-5〉 정규분포 가정 합계출산율



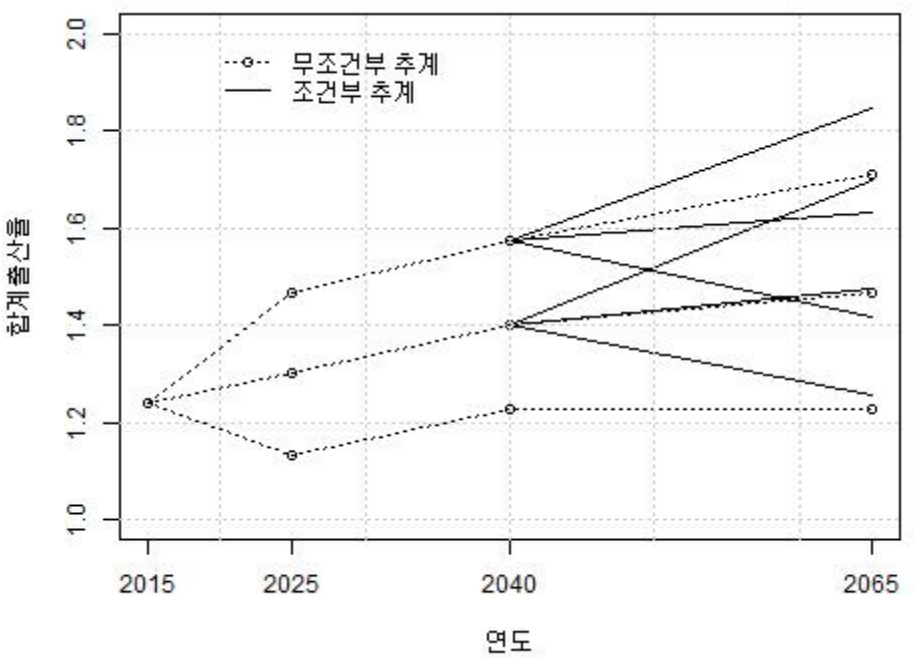
앞서 중위, 고위, 저위에 있어 조건부 합계출산율 추계한 것과 유사하게 정규분포 가정 하에 조건부 합계출산율에 대한 질문을 하였고 그 결과를 다음의 표에 제시하였다.

〈표 2-7〉 정규분포 가정 조건부 합계출산율 추계

2040년 합계출산율 실현값		2065년 추계값	
		중위값	고위값
중위값	평균	1.48	1.70
	표준편차	0.24	0.31
	최솟값	1.10	1.17
	중간값	1.50	1.78
	최댓값	2.02	2.31
고위값	평균	1.63	1.85
	표준편차	0.28	0.37
	최솟값	1.01	1.03
	중간값	1.70	2.00
	최댓값	2.19	2.51

다음의 그림은 정규분포 가정 조건부 합계출산율 추계를 무조건부 합계출산율과 같이 보여주고 있다. 중위, 저위, 고위에 대한 그림과 마찬가지로 2040년의 평균에서 2065년의 평균으로 가는 것은 조건부와 무조건부가 유사하지만, 2040년 조건부 고위의 경우 2065년 중위의 값이 무조건부 중위의 값보다 작은 것을 볼 수 있다.

〈그림 2-6〉 정규분포 가정 조건부 합계출산율

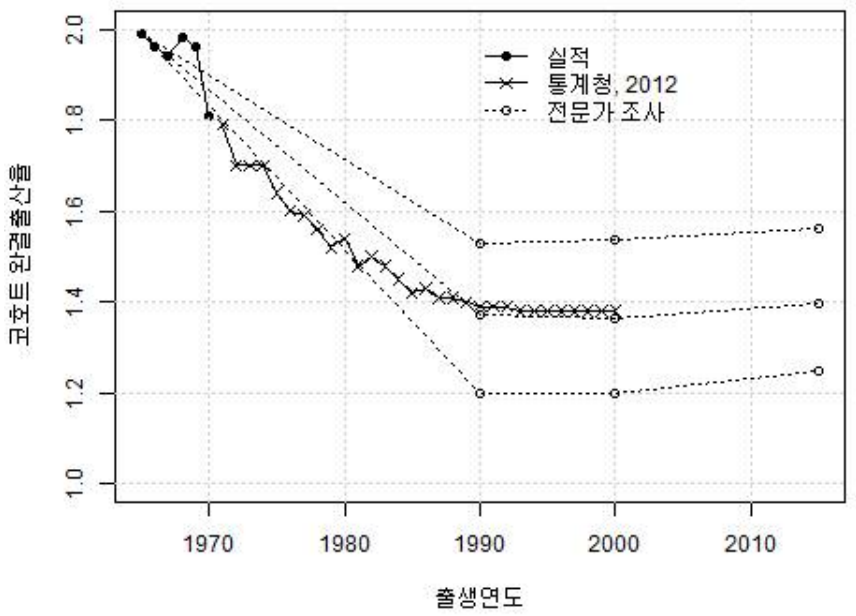


다음의 표와 그림은 코호트 완결출산율에 대한 추계 결과를 보여주는데, 1990년 출생코호트까지는 완결출산율이 떨어지다 그 이후 약간 올라갈 것이라고 예측하고 있다. 이에 더해 이 그림에서는 전문가들의 중위값이 통계청의 추계와 유사하지만 약간 낮은 값을 보여주는 것으로 타나고 있다.

〈표 2-8〉 코호트 완결출산율 추계 결과

		중위값	고위값	저위값
1990년생 완결 출산율	평균	1.37	1.53	1.20
	표준편차	0.21	0.22	0.20
	최솟값	1.10	1.30	0.90
	중간값	1.40	1.53	1.20
	최댓값	2.00	2.20	1.80
2000년생 완결 출산율	평균	1.36	1.54	1.20
	표준편차	0.21	0.24	0.20
	최솟값	1.03	1.10	0.80
	중간값	1.37	1.58	1.17
	최댓값	1.90	2.10	1.70
2015년생 완결 출산율	평균	1.40	1.56	1.25
	표준편차	0.26	0.32	0.28
	최솟값	1.00	0.90	0.70
	중간값	1.40	1.60	1.20
	최댓값	1.95	2.15	1.77

〈그림 2-7〉 코호트 완결출산율 추계

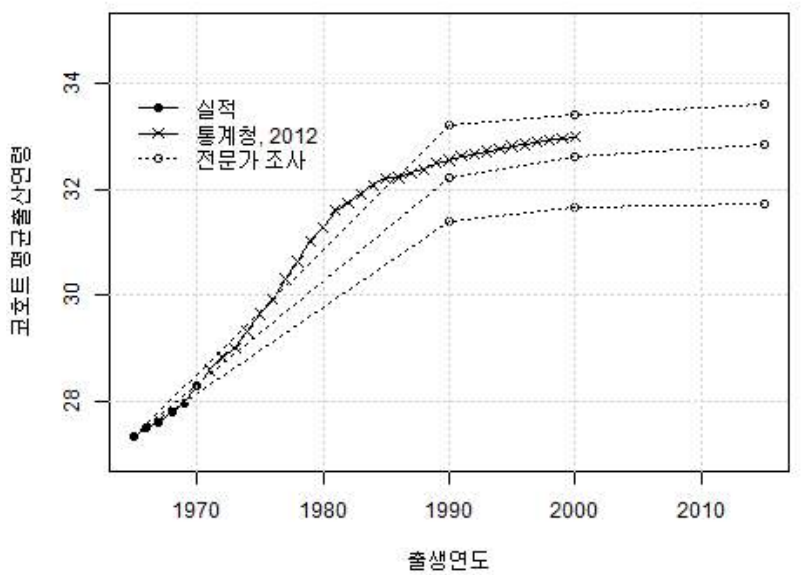


다음 표는 코호트 평균출산연령 추계 결과를, 이어서 <그림 2-8>은 <표 2-9>를 요약적으로 보여주고 있다.

〈표 2-9〉 코호트 평균출산연령 추계 결과

		중위값	고위값	저위값
1990년생 완결 출산율 감안 평균출산연령	평균	32.23	33.21	31.39
	표준편차	1.21	1.59	1.51
	최솟값	30.00	30.06	28.50
	중간값	32.30	33.61	31.05
	최댓값	34.35	35.55	33.25
2000년생 완결 출산율 감안 평균출산연령	평균	32.62	33.40	31.67
	표준편차	1.50	2.26	1.72
	최솟값	29.50	28.80	28.80
	중간값	32.50	34.00	31.00
	최댓값	35.55	36.55	34.55
2015년생 완결 출산율 감안 평균출산연령	평균	32.84	33.59	31.72
	표준편차	2.03	3.00	2.09
	최솟값	29.00	28.00	28.50
	중간값	32.96	34.50	31.05
	최댓값	36.55	37.55	35.55

〈그림 2-8〉 코호트 평균출산연령 추계



위의 그림을 보면 전문가들은 1990년도 출생코호트까지 평균출산연령이 급격하게 증가하다 이후에 안정화되면서 매우 조금씩 증가할 것으로 예상하고 있다. 이에 더해 전문가 추계 중위값이 통계청의 추계보다 낮지만 그리 큰 차이가 나지 않는 수준에서 변화가 이루어지고 있는 것을 본다.

제5절 장래 출산율 추계 결과

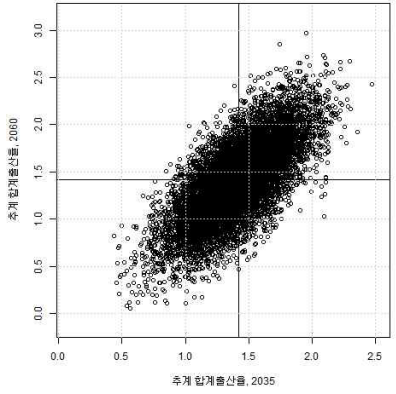
1. 이변량 정규분포 모형 결과

다음의 그림은 이변량 정규분포 모형 및 주관적 베이지안 모형의 추계 결과를 보여주고 있다. x-축에는 첫 번째 추계연도의 합계출산율 예측값을 제시하였으며, y-축에는 두 번째 추계연도의 예측값을 제시하였다. 자료별 비교를 용이하게 하기 위하여 x-축과 y-축의 값을 통일하였다. 이 그림에서 <A 패널>은 2012년 통계청의 인구추계 자료를 이용하여 시산한 추계값이며, <B 패널>과 <C 패널>은 고위값 및 중위값에 대한 예측을 바탕으로 추계한 값들이다. 다만 전자는 모든 전문가 응답의 평균을 사용하여 하나의 전문가가 예측한 듯한 효과를 보이지만, 후자는 각 전문가의 응답을 하나의 분포로 추정한 다음 모든 전문가들의 분포를 합친 것이다. <D 패널>과 <E 패널>은 추계값이 정규분포를 이룬다는 가정 하에 질문한 추계값을 이용한 것이며, 앞서와 마찬가지로, 전자는 전문가들의 평균을 사용하였으며, 후자는 개별 전문가들의 응답을 하나의 분포로 설정한 다음 모든 분포를 결합한 것이다. <F 패널>은 방법론 부분에서 상술한 것처럼 주관적 베이지안 모형을 적용한 결과를 보여주고 있다. <A, B, D, F 패널>의 자료를 생성하기 위해서 10,000 사례를 표집하였으며, <C 패널>에는 각 전문가 당 600 사례씩, 총 10,800 사례를 표집하였고, <E 패널>에서는 각 전문가당 600 사례씩, 총 10,200 사례를 표집하였다. 마지막 자료의 경우 한 명의 전문가가 무응답한 필수 항목이 있었으므로 17명의 전문가가 응답이 쓰였기 때문에 10,200 사례가 표집되었다.

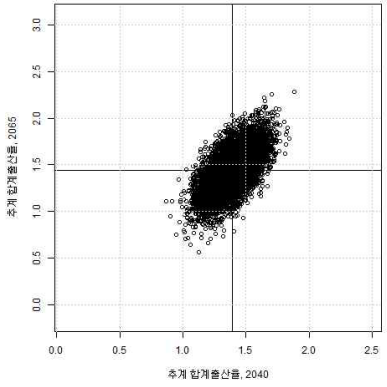
이 결과를 비교해 보면 몇 가지 흥미로운 현상을 관찰할 수 있다. 우선 2040년보다 2065년 추계의 산포도가 더 큰데, 이는 예측 기간이 길어질수록 불확실성이 커진다는 일반적인 현상을 반영하는 것으로 풀이된다. 또한 전문가 조사의 평균값을 사용하여 추계를 한 B와 D 모형 결과의 분산이 가장 작지만 통계청 자료를 활용한 A 모형 결과의 분산은 작지 않다. 이는 전문가들이 자신의 관점에서 추계를 할 때, 고위값과 저위값 그리고 80% 신뢰구간에 대한 과도한 자신감으로 인해 불확실성의 정도를 과소추정하기 때

〈그림 2-9〉 이변량 정규분포 및 주관적 베이지안 모형 추계 결과

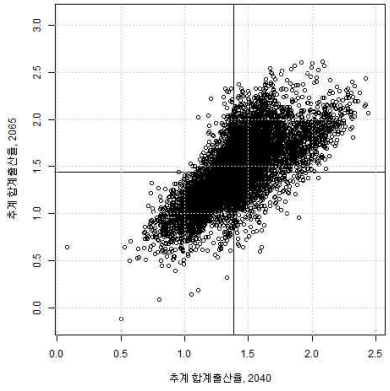
A. 통계청 자료



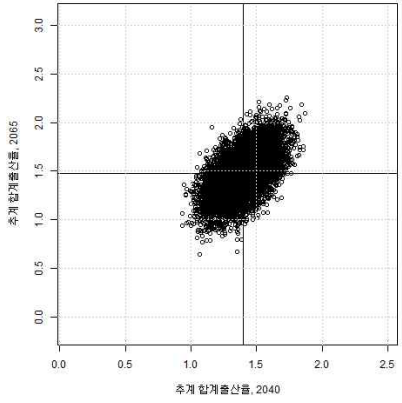
B. 시나리오 평균



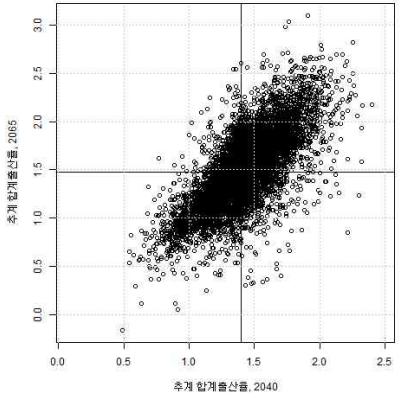
C. 개별 전문가 시나리오



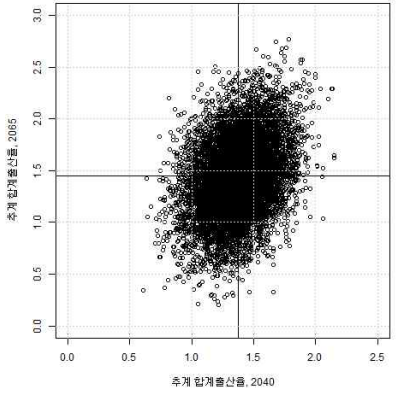
D. 정규분포 평균



E. 개별 전문가 정규분포



F. 주관적 베이지안 모형



문인 것으로 생각된다.

이에 더해, B와 D의 분산이 그리 큰 차이가 나타나지 않는 것은, 앞에서도 살펴 본 것처럼 고위값과 저위값의 분산이 정규분포 80% 신뢰구간의 상한과 하한값과 그리 큰 차이가 나지 않기 때문인 것으로 판단된다.

다른 결과는 이항 정규분포의 형태를 띠고 있으나, C와 E의 결과는 이항 정규분포의 모습과 상당히 다른 것을 볼 수 있는데, 이는 이들의 결과가 다양한 전문가의 분포를 합쳐 놓은 결합분포(mixture distribution)이기 때문이다. 즉 한 명의 전문가는 자신의 이항 정규분포를 이루고 있는데, 이렇게 다양한 이항 정규분포를 결합해 놓았기 때문에 전체 분포는 이항 정규분포에서 멀어지게 되었다. 또한 각 전문가들의 추계값들이 상당한 차이를 보이기 때문에 전체 분포의 분산이 적지 않은 것으로 나타나고 있다.

2. 주관적 베이지안 예측 모형 결과

<그림 2-9>의 마지막 <F 패널>에서는 주관적 베이지안 예측 모형의 결과를 보여주고 있다. 이를 다른 예측 모형들과 비교해 보면 <B 패널>과 <D 패널>의 타원형 모형에 <C 패널>과 <E 패널>의 분산만큼 늘인 것 같은 형태로 나타나는 것을 볼 수 있다. 이는 베이지안 모형에서는 앞선 두 모형의 평균을 모형화하지만, 이들과는 다르게 전문가들 사이의 예측에 있어서 분산의 정도 또한 모형화함으로써 불확실성을 고려하기 때문이다.

다른 한편 2040년의 합계출산율에 대한 예측의 분산정도가 다른 모형들보다 작은 것으로 나타나고 있는데, 이는 사전 분포의 설정과 연관이 있는 것으로 판단된다. 즉 2015년과 1990년의 출산율을 하나의 변수로, 그리고 1990년과 1970년의 출산율을 다른 하나의 변수로 취급한 후 두 변수의 공분산을 구하여 Λ_0로 설정함으로써 2015년과 1990년의 합계출산율 변이가 25년의 출산율 예측에 대한 분산으로 설정되었는데, 두 년도의 합계출산율은 1.24와 1.57로 커다란 차이가 없다.

3. 모든 추계 결과 비교

다음의 표는 이 연구에서 수행한 모든 추계 결과에 더해, 더 많은 비교를 하기 위해 통계청의 2012년 추계와 UN의 2015년 추계 결과를 제시하고 있다. 전문가 조사 단순 추계는 전문가들의 응답을 활용하여 평균을 구한 것으로, 예를 들어 2040년의 중위값에 대한 응답 결과의 평균을 활용하여 중위값으로 설정하였다. 이항 정규분포 모형은 앞서 제시한 결과를 활용하여 모든 모형에서 평균과 80% 신뢰구간의 상한값과 하한값을 각각

중위값, 고위값, 저위값으로 제시하였다.

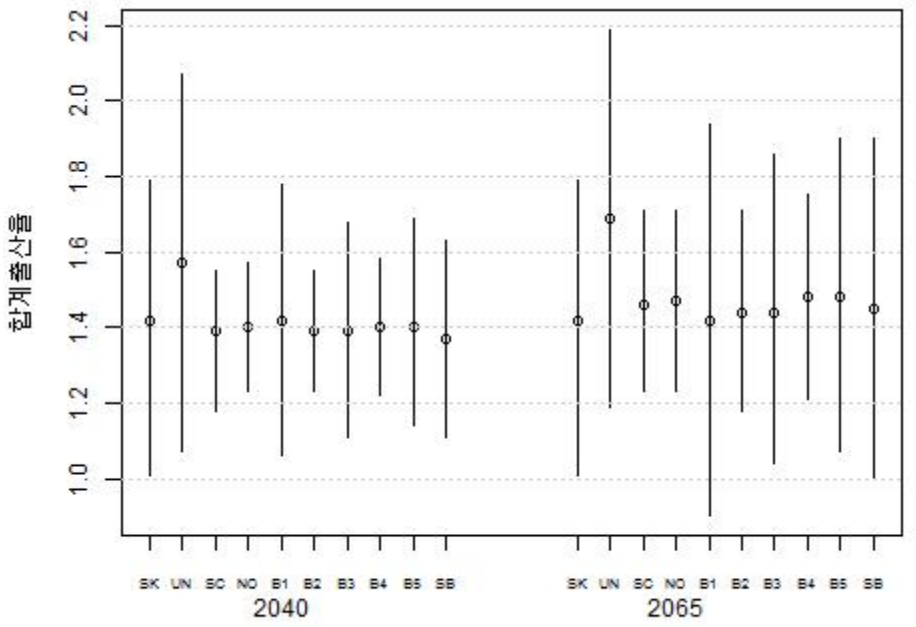
이러한 결과를 보다 쉽게 파악하기 위하여 <그림 2-10>에서는 각 년도의 추계값을 추계방법에 따라 그림으로 표시하고 있다. 이 그림을 통해 다음과 같은 몇 가지 관찰을 할 수 있다. 우선 각 년도에서 UN의 추계가 가장 높은 값을 보여주고 있으며, 다른 방법으로 추계한 값들은 대동소이한 것으로 판단된다. 둘째, 모든 추계에서 2040년에 비해 2065년의 값이 약간 큰 것을 볼 수 있으나 그리 큰 차이를 보이지는 않는다. 보다 정확한 숫자를 대입하여 서술하면, 중위값들이 2040년의 경우 1.4보다 약간 작은 값들을 보이지만 2065년에 대한 추계는 1.4보다 약간 큰 값들을 보이고 있다. 아마도 이는 몇몇 전문가들의 응답에서 나타난 것처럼, 먼 미래의 추계가 불확실할 때 가까운 미래의 값으로 대체하는 인식론적 특성이 반영된 것으로 판단된다. 셋째, 전문가 조사의 평균을 사용한 추계는 고위값과 저위값의 차이가 크지 않은 것으로 나타난다. 이는 전문가들의 불확실성에 대한 과소추정이 하나의 원인이고, 다양한 전문가들의 의견을 평균화하는 과정에서 분산이 큰 값들이 평균화되었기 때문인 것 같다. 이에 반해 개별 전문가 추계를 합친 경우와 주관적 베이지안 모형에서는 불확실성의 정도가 큰 것으로 나타나는데, 이는 전문가들 사이의 편차를 모형화하였기 때문이다.

〈표 2-10〉 모든 추계 결과

		2040년			2065년			비고
	약어	고위	중위	저위	고위	중위	저위	
통계청, 2012	SK	1.79	1.42	1.01	1.79	1.42	1.01	2065년은 2060년 추계값
UN, 2015	UN	2.07	1.57	1.07	2.19	1.69	1.19	2040-2045, 2065-2070값
전문가 조사 단순 추계								
시나리오 평균	SC	1.55	1.39	1.18	1.71	1.46	1.23	
정규분포 평균	NO	1.57	1.4	1.23	1.71	1.47	1.23	
이항 정규분포 모형								
통계청 자료	B1	1.78	1.42	1.06	1.94	1.42	0.90	
시나리오 평균	B2	1.55	1.39	1.23	1.71	1.44	1.18	
개별 전문가 시나리오	B3	1.68	1.39	1.11	1.86	1.44	1.04	
정규분포 평균	B4	1.58	1.40	1.22	1.75	1.48	1.21	
개별 전문가 정규분포	B5	1.69	1.40	1.14	1.90	1.48	1.07	
주관적 베이지안 모형	SB	1.63	1.37	1.11	1.90	1.45	1.00	

주: 정규분포의 경우 중위는 평균을 의미하며, 고위와 저위는 80% 신뢰구간의 상·하한값을 의미한다.

〈그림 2-15〉 추계 결과의 비교



마지막으로 전문가들의 의견을 물을 때 정규분포의 평균과 신뢰구간으로 묻거나 중위, 고위, 저위값으로 묻거나 큰 차이를 보이지 않는다. 물론, 이 연구만을 통해서 고위와 저위값이 80% 신뢰구간과 유사하다는 결론을 내리는 것은 옳지 않은 것 같다. 만약 질문을 달리하여 70% 신뢰구간으로 답을 하도록 요청하였을 때에도 고위와 저위값들과 유사한 값들을 보고하였을 가능성이 매우 크다. 따라서 과연 어느 정도의 신뢰구간이 고위와 저위값들과 유사한 것인가에 대해 더욱 명확히 알기 위해서는 더 많은 연구가 요청된다하겠다.

제6절 논의 및 제언

이 장에서는 전문가들의 개인적인 견해에 기반하여 장래 출산율을 추계하는 전문가 판단법에 관한 학술적인 연구 흐름을 살펴보고, 실제 자료를 수집한 후 몇몇 모형을 설정하여 추계를 실시하였다. 전문가 판단법의 통계적 분석 기법도 다양하여 어떤 방법을 사용하느냐에 따라 예측의 불확실성을 나타내는 신뢰구간이 커진다는 것도 살펴보았다.

그럼에도 불구하고, 어떤 방법을 사용하든 평균 수준에 대한 추계는 크게 차이를 보이지 않았다. 전문가들은 평균적으로 현재에 비해 2040년의 합계출산율은 약간 상승하여 1.4 정도를 기록할 것으로 예측한 반면, 2065년에는 그보다 약간만 더 상승할 것으로 기대하고 있는 것으로 나타났다. 여타의 추계와 비교한다면 UN의 추계(UN, 2015b)만이 두 시점 모두에서 높게 나타나고 있는데, 이는 전세계의 경험을 참조하여 한국의 미래를 예측하였기 때문인 것으로 풀이된다. 예를 들어, 유럽의 다수 국가는 1990년대를 전후하여 짧은 시기의 초저출산을 겪은 후 출산율이 상승하였는데 이런 국가들을 참조함으로써 한국 또한 출산율이 상승하는 것으로 나타나기 때문인 것 같다. 물론 이러한 결과는 UN의 전문가들이 고의로 그런 결과를 산출했다기 보다는 베이지안 모형을 활용하여 추정하는 과정에서 유럽국가의 영향이 반영되었기 때문인 것으로 풀이된다.

이 연구는 출산율 추계에 있어 전문가 판단법을 시도한 최초의 연구라는 점에서 커다란 의의를 가지지만, 그럼에도 불구하고, 아마도 그 최초라는 이유로, 여러 가지 극복되어야할 한계를 지니고 있다. 예를 들어 <그림 2-9>의 <패널 C>를 보면 0보다 낮은 값을 가지는 합계출산율이 있는 것을 본다. 합계출산율은 0이나 양의 값을 가질 수밖에 없다는 점에서 이러한 결과가 나오는 모형은 어느 정도 문제가 있는 것이라고 할 수 있다. 이를 넘어설 수 있는 한 가지 방법은 합계출산율을 모형화하는 것이 아니라 로그를 취한 합계출산율을 모형화한 후 최종값에서는 원래의 값으로 돌려주는 분석을 하는 것이다. 이 보고서에서는 이러한 시도를 해보지 못하였으나 향후 꼭 시도해 볼만한 작업이라고 생각된다.

또한 이 보고서에서는 연구 기간의 한계로 인하여 주관적 베이지안 모형만을 활용하였는데, 문헌 분석에서 제시하였던 것처럼 Billari와 동료들이 2014년에 했었던 혼합 베이지안 모형을 활용하여 추계해 보는 것도 필요하다. 이에 더해 주관적 베이지안 모형만이 아닌 수리 모형과 전문가의 주관적 예측을 결합하는 모형을 발전시킬 필요가 있다. 이러한 방향으로 향후 연구가 전개된다면 출산력 추계의 타당성이나 예측력 향상에 큰 도움이 될 것으로 판단된다.

참고문헌

전광희·김태헌·조영태. (2005). 장래인구추계를 위한 출산·사망 예측모형의 개발에 관한 연구. 통계청·인구학회.

이삼식·이지혜·최효진. (2013). 인구예측모형 국제비교 연구. 한국보건사회연구원.

통계청. (2016a). 장래인구추계 방법 및 가정. kosis.kr

통계청. (2016b). 각종 통계 자료. kosis.kr

Basten, S., Sobotka, T., Zeman, K, Abbasi-Shavazi, M. J., Adsera, A., Van Bavel, J.,..., Zhao, Z. (2013). Future ferility in low fertility countries. Vienna Institute of Demography Working Papers.

Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. Journal of Royal Statistical Society, 173, Part 4, 775-796.

Billari, F. C., Graziani, R., & Melilli, E. (2012), Stochastic population forecasts based on conditional expert opinions. Journal of Royal Statistical Society, Series A., 175, Part 2, 491-511.

Billari, F. C., Graziani, R., & Melilli, E. (2014), Stochastic population forecasting based on combinations of expert evaluations within the Bayesian paradigm. Demography, 51, 1933-1954.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis, 2nd ed. New York: Chapman & Hall/CRC

Keyfitz, N. (1982). Can knowledge improve forecasts? Population and Development Review 8(44), 729-751.

Lutz, W. (2009). Toward a systematic, argument-based approach to defining assumptions for population projections. Interim Report IT-09-037, International Institute for Applied Systems Analysis.

Lutz, W., Butz, W. P., & KC, S. (Eds). (2014). World population and human capital in the 21st Century. Oxford: Oxford University Press.

Office for National Statistics (2013), Background and methodology, 2012-based national population projections.

Preston, S. H., Heuveline, P., & Guillot, M. 2001. Demography: Measuring and modeling population processes. Malden, MA: Blackwell Publishers Inc.

Sobotka, T., Zeman, K., and Basten, S. (2016). The low fertility future? Projections based on different methods suggest long-term persistence of low fertility. Paper presented in Population Association of America, Washington, D.C.

Statistics Canada (2015). Population projections for Canada (2013 to 2063), provinces and

territories (2013 to 2038): Technical report on methodology and assumption (91-620-X).
United Nations, Department of Economic and Social Affairs, Population Division (2015a). World population prospects: The 2015 Revision, methodology of the United Nations population estimates and projections, Working Paper No. ESA/P/WP.242.
United Nations, Department of Economic and Social Affairs, Population Division, (2015b). World population prospects: The 2015 revision, DVD edition.

제 3 장

혼인율과 혼인출산율 추이를 활용한 출산율 추계 방법

계봉오

제1절 서 론: 근접요인 접근법을 활용한 출산율 추계의 필요성

출산에 대한 근접요인 접근법(Proximate determinant approach)은 출산에 영향을 미치는 요인을 간접적인 요인과 직접적인 요인으로 나눈다(Bongaarts 1978). 간접적인 요인은 사회경제적, 문화적, 환경적 요인들을 포함하고, 직접적인 요인은 출산의 근접요인으로 이루어진다 . 근접요인들은 다시 노출 요인, 혼인출산력 통제요인, 자연출산력 요인 등으로 구분된다. 노출요인(exposure factor)은 정기적으로 성관계를 갖는 집단의 비율을 의미하는데, 이 비율이 높으면 출산율이 높아진다. 전통적으로 노출요인은 유배우율(proportion married)로 측정해 왔는데, 혼외 출산이 전체 출산에서 차지하는 비중이 점차 높아진 현대 서구사회에서는 유배우율로 노출요인을 측정하는 것이 적절하지 않다. 그렇지만, 전체 혼인에서 혼외출산이 차지하는 비중이 2-3%로 매우 낮은 한국사회의 경우에는 유배우율로 노출요인을 측정하는 것이 여전히 적실성을 지닌다. 혼인출산력 통제요인은 피임, 인공 유산 등의 요인을 의미하며, 출산통제는 출산율을 낮춘다. 마지막으로 자연출산력 요인은 가임력 및 기간, 모유수유기간, 성관계 빈도, 자연유산 등으로 구성되며, 자연출산력이 높으면 출산율은 높아진다. 근접요인 접근법은 각각의 근접요인들에 대한 지수(index)를 구성하고 이들의 곱으로 합계출산율을 구하는 방식인데, 출산을 구성하는 근접요인들의 변화로 출산율의 변화를 설명하는 방식으로 광범위하게 사용되어 왔다. 예를 들어, 봉가르는 한국의 합계출산율이 1960년에 6.13에서 1970년에 4.05로 34% 하락했는데, 유배우 합계출산율(total marital fertility rate)는 18%, 유배우자

비율 지수는 19% 하락했음을 보여주었다(Bongaarts 1978). 즉, 이 시기 합계출산율의 변화는 유배우 출산율과 유배우자 비율 모두 하락한 결과로 분해해서 설명할 수 있는 것이다. 또한, 이철희(2012)의 연구는 1980년대 이후 한국의 출산율 변화는 유배우 출산율의 변화가 아니라 유배우율의 하락에 기인함을 보여주었다. 이와 같은 근접요인 접근법은 출산율의 변화를 개별 근접요인 변화로 분해하여 과거 출산율 변화를 분석적으로 이해하는 도구로 활발하게 활용되어 왔다.

적절한 인구추계를 위해서는 기준인구(baseline population), 출산율, 사망률, 이동률에 대한 정확한 정보가 필요하다. 기준인구는 기준시점의 인구를 정확하게 파악함으로써 추정할 수 있는 반면에 장래 인구동태율(출산, 사망, 이동율)에 대한 추정치들은 과거 추이를 적절하게 파악할 수 있는 수학적 모형을 개발하고, 장래 변화 추이에 대한 설득력 있는 가정을 결합하여 생산하게 된다. 사망률의 경우에는 Lee-Carter 모형 및 이의 변형이 주로 활용되고 있는데(Lee and Carter 1992; Li and Lee 2005), 이 모형들은 적은 수의 모수치로 사망률의 변화추이를 매우 잘 설명하는 장점이 있다.[1] Lee-Carter 모형이 사망률 추계의 표준적인 방법으로 자리잡은 이유 중의 하나는 사망률의 변화는 대체로 단선적인 패턴을 따른다는 데 있다. 이와 대조적으로 출산율 변천이 완료된 대부분의 국가들에서 출산율은 단선적으로 변화하지 않고, 등락을 거듭하는(cyclically fluctuating) 불규칙한 패턴을 보여주었다(Calot 1993; Foster 1990; Kye 2015; Rindfuss et al. 1988). 이는 사망률이 전반적인 사회경제적 발전에 따라 지속적으로 하락하는 반면에 출산율은 단기적인 경기상황 등에 의해서 영향을 받으면서 등락을 거듭하기 때문이다. 따라서, 출산율의 장래추이에 대한 수학적 모형은 사망률 예측 모형보다 개발이 어렵다고 할 수 있다.

이 연구는 근접요인 접근법에 기초한 접근이 출산율 추계에 유용할 수 있다는 사실에서 출발한다. 아직까지 근접요인 접근법이 장래인구 추계를 위한 출산율 추정치를 개발하기 위해서는 활용되지 않았다. 근접요인 접근법의 기본 가정은 출산율 변화는 개별 근접요인의 변화로 분해해서 분석할 수 있다는 것이다. 이는 출산율 변화 추이에서 관찰되는 불규칙성이 사실은 서로 다른 근접요인들이 서로 다른 방향으로 변화하고 있기 때문에 나타나는 현상일 수 있음을 시사한다. 예를 들어, 혼외출산율은 무시할만하고 유배우율은 지속적으로 하락하고 있으며, 유배우 출산율은 지속적으로 상승하는 경우를 생각해보자. 이 경우에, 합계출산율의 추이는 큰 변화가 없이 유지되고 있을 가능성이 있다. 그렇지만, 이러한 안정성은 두 가지 요인, 즉 유배우율의 하락과 유배우 출산율의 상

1) 물론, Lee-Carter 모형에 대한 비판 역시 존재하며(예: 초고연령대 사망률 추이에 대한 설명력이 높지 않다), 이 과제의 연구목적 중의 하나는 장래인구 추계를 위한 사망률 추정치 개발에 있다. 그럼에도 불구하고, 사망률 추계에서는 Lee-Carter 모형이 표준적인 위치를 점하고 있지만, 출산율 추계와 관련해서는 이와 같은 표준적인 모형이 확립되어 있지 않다.

승이 서로 상쇄하고 있기 때문에 나타나는 현상이라고 할 수 있으며, 출산율의 변화 추이를 보다 정확히 이해하기 위해서 유배우율과 유배우 출산율에 대한 개별적인 추계치를 하고, 이 둘을 결합하는 방식이 바람직할 수 있다. 즉, 현상적으로 드러나는 출산율 변화의 배후에 있는 유배우율과 유배우 출산율의 변화추이를 세밀하게 검토함으로써, 장래 출산율에 대한 적절한 추계치를 개발할 수 있는 것이다. 모든 근접 요인들에 대한 전국적인 수준에서의 측정치가 존재하지 않기 때문에, 이 과제는 유배우율과 유배우 출산율을 활용하여 장래 출산율에 대한 추정치를 개발하고자 한다.

제3장의 구성은 다음과 같다. 제2절에서는 근접요인 접근법을 활용한 연령별 출산율 추정의 기본모형을 제시하고, 제3절에서는 출산율의 반등을 모형화할 수 있는 이론적 논의들을 소개한다. 제4절에서는 분석 자료에 대해서 설명하고, 제5절에서는 1980-2014년 자료를 활용한 과거 추세 분석결과를 제시한다. 제6절에서는 Lee-Carter 모형을 활용하여 과거 추세를 요약하고, 장래 출산율 추계를 위한 기초분석 결과를 제시한다. 제 7절에서는 제3절에서 논의한 출산율 반등에 대한 이론적 모형과 제6절의 경험적 분석결과를 활용하여 2015-2040년의 연령별 출산율 추계 결과를 몇 가지 가정 하에서 제시한다.

제2절 근접요인 접근법을 활용한 연령별 출산율 추정 기본모형

근접요인 접근에 따르면, 합계출산율은 유배우율과 유배우 합계출산율에 의해서 결정된다. 수식으로 표현하면 다음과 같다.

$$TFR = CM \times TM \quad (CM: \text{유배우율 지표; } TM: \text{유배우 합계출산율}) \qquad (1)$$

이 모형은 혼외출산이 일반적인 북미나 유럽의 경우에는 그 적합성이 떨어지지만, 혼외출산이 전체 출산에서 차지하는 비중이 3% 미만인 한국사회에서는 여전히 적합성이 있다고 할 수 있다. 이철희(2012)의 연구에 따르면, 1991년 이후 한국의 유배우 출산율은 안정적인 추세를 보여주며, 한국의 합계출산율 변화는 전적으로 유배우율 변화에 의해서 설명된다. 이는 유배우율과 유배우 출산율에 의해서 결정되는 합계출산율을 두 부분으로 나누어서 살펴보는 것이 매우 유용한 접근임을 보여준다. 특히, 유배우율과 유배우 출산율이 상이한 추이를 보여주는 경우에는 이 둘을 나누어서 살펴보는 것이 매우 유용하다. 합계출산율의 변화가 서로 다른 추이를 보여주는 두 가지 요인에 의해서 결정된다고 하면, 이 둘의 추이를 나누어서 살펴본 후 이를 종합하는 것이 합계출산율의 변

화 추이를 예측하는 데 유용하기 때문이다.

이 연구는 이러한 문제의식에 기반해서 연령별 출산율에 대한 추계치를 추정하기 위해 연령별 유배우율과 연령별 유배우 출산율(age-specific marital fertility rate)을 각각 추정하여 이를 결합하는 방식을 활용한다. 이를 수식으로 표현하면 다음과 같다.

$$ASFR = ASMR*ASMFR*CR \quad (2)$$

[ASFR(age-specific fertiltiy rate): 연령별 출산율,
ASMR(age-specific marriage rate): 연령별 유배우율,
ASMFR(age-specific marital fertility rate): 연령별 유배우 출산율,
CR(correction ratio): 혼외출산 보정 요인]

수식 (2)의 연령별 유배우율(ASMR)은 연령별 초혼이행률, 이혼율, 사별률, 혼인상태별 사망률 등에 의해서 결정된다. 현재 한국사회에서 진행되고 있는 혼인의 지연과 이혼의 증가는 유배우율을 낮추며, 사별의 감소는 유배우율을 높인다고 할 수 있다. 또한, 유배우자의 사망률이 미혼/이혼/사별자보다 낮은 경향이 있는데, 이는 유배우율을 높인다고 할 수 있다. 이 연구는 통계청에서 제공하는 각세별 연령별 출산율(ASFR), 연령별 유배우율(ASMR), 연령별 유배우 출산율(ASMFR)을 기초자료로 활용하여, 혼외출산 보정요인(CR)를 추정하고, 이의 변화 추이(1980년 이후)를 살펴본다. 이 때, 시계열 자료의 변화를 요약하고 향후 변화 추이를 예측하기 위한 모형으로 Lee-Carter 모형을 활용한다. 또한, 이러한 분석결과와 제3절에서 기술할 향후 출산율 변화에 대한 가정을 활용하여 장래인구추계를 위한 연령별 출산율 추정치를 개발하도록 한다.

제3절 장래 출산율 변화 추이에 대한 가정 도출

장래인구 추계를 위한 사망률 추정치의 개발은 주로 Lee-Carter 모형(Lee and Carter 1992)이나 Li-Lee 모형(Li and Lee 2005)를 활용한다. 연령별 사망률은 매우 규칙적인 패턴을 보여주며, 사망률의 하락은 향후에도 지속될 것으로 예상되기 때문에, 장래 사망률 추정치의 개발은 과거 추이를 잘 요약할 수 있는 수리적 모형을 개발하는 작업을 통해서 이루어질 수 있다. 반면, 장래인구 추계를 위한 출산율 추정치의 개발은 보다 복잡한 요인을 고려해야 한다. 이는 출산율의 변화가 단선적인 패턴을 따르지 않고, 비선형적으로 변화하는(fluctuating) 패턴을 따르기 때문이다. 20세기 출산율 변화에 대한 대부분의 연구들은(Calot 1993; Foster 1993; Scheon 2004; Kye 2015) 출산율의 시기 변화가 단선

적인 패턴을 따르지 않았음을 공통적으로 보여주고 있다. 연령-기간-코호트 모형(Age-period-cohort model)에서 사용하는 용어를 빌면, 출산은 기간 효과(period effects)의 영향을 많이 받는데, 기간 효과는 전쟁, 경기변동 등 해당 기간의 특수한 사회경제적 조건에 따라 비선형적으로 변화하며, 이것이 출산율 변화가 비단선적인 패턴을 나타내는 원인이라고 할 수 있다. 따라서, 과거 추이에 대한 분석이 향후 추이를 추정하는 데 주는 함의는 제한적일 수밖에 없다. 앞서 기술한 근접요인 방법에 의한 출산율 분해 역시 출산율의 결정요인을 분석적으로 이해하는 데 큰 도움을 주지만, 이를 바로 장래인구추계를 위한 출산율을 도출하기 위해서 사용하는 데에는 무리가 따른다.

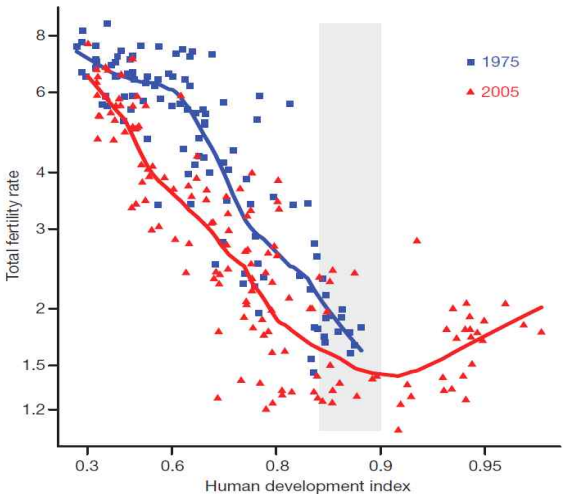
한편, 한국의 경우에는 1960년대 이후 출산율이 지속적, 단선적으로 하락했기 때문에, 출산율 시계열 자료에만 의존하여 장래 출산율 추이를 추정하는 것은 방법론적으로는 사망률 추이를 추정하는 것과 동일하다고 할 수 있다. 문제는 이와 같은 방식을 활용한다면, 출산율의 반등 가능성을 모형화할 수 없다는 데 있다. 즉, 단선적인 하락 추이를 보여주는 자료를 활용하고 별도의 이론적 가정 없이 장래 출산율을 추정한다면 결국은 가까운 장래에 한국의 출산율이 0에 수렴할 것이라는 결론을 내릴 수밖에 없을 것이다. 이는 유배우율과 유배우출산율을 나누어서 분석하고자 하는 이 연구도 피할 수 없는 결론이다. 즉, 유배우율이 지속적으로 하락한다고 가정한다면, 언젠가는 0으로 수렴할 것이고 이는 합계출산율이 0으로 수렴한다는 것을 의미한다. 한편, 유배우출산율이 지속적으로 상승한다고 가정하면, 합계출산율은 비현실적인 수준까지 상승하는 것으로 추정될 것이다.

따라서, 기존의 출산율 변화 추이 자체만을 활용해서 장래 출산율에 대한 유의미한 추정치를 개발할 수 없다. 이는 결국 장래 출산율에 대한 목표 혹은 가정치를 활용하는 것이 불가피하다는 것을 의미한다. 그렇다면, 이러한 목표 혹은 가정치는 어떻게 구성할 수 있는가? 2010년 통계청의 인구추계의 경우에는 합계출산율이 2040년까지 1.42 (중위), 1.80 (고위), 1.08 (저위)로 변화하고 이후에는 동일한 수준으로 유지된다는 가정을 도입했다(통계청 2011). 본 연구자가 이해하기에는 이러한 가정이 어떻게 도출되었는지에 대한 설명은 충분히 제시되지 않았다. 저위의 경우에는 현재의 추세가 지속된다는 가정을 어느 정도 갖고 있는 것 같으며, 중위와 고위의 경우에는 향후 출산율이 일정 정도 상승할 것이라고 가정하고 있지만, 이에 대한 기술은 자세히 제시되어 있지 않다.

이 연구에서는 두 가지 이론적 관점을 활용해서 이론적으로 정당화할 수 있는 향후 출산율의 반등을 모형화할 수 있는 방식을 모색하고자 한다. 우선, 미르스키야 등의 연구(Myrskylä et al. 2009)에 따르면(<그림 3-1>), 합계출산율과 인간개발지수(Human Development Index, HDI)의 관계는 2005년 현재 U자형 관계를 보여준다. 즉, 1970년대에는 인간개발지수가 높은 나라들이 출산율이 낮은 단선적인 관계가 관찰되었지만,

2000년대 중반이 되면 이러한 패턴에 변화가 관찰되고 있다. 즉, 합계출산율은 인간개발지수가 매우 낮은 국가들에서 매우 높으며, 인간개발지수가 높아짐에 따라 떨어지지만, 인간개발지수가 0.8 정도보다 높은 경우에는 오히려 인간개발지수가 높은 국가들의 합계출산율이 높게 나타나는 경향이 나타난다. 이는 사회경제적 발전과 합계출산율이 비선형적인 U자형 관계를 맺고 있음을 보여준다. 즉, 어떤 임계점(tipping point)를 지나게 되면, 사회경제적 발전이 합계출산율의 상승으로 이어질 수 있음을 시사한다. 이는 향후 한국사회에서 출산율의 반등이 일어날 수 있음을 강하게 시사하며, 이를 출산율 추계에 반영하는 것이 가능하다고 할 수 있다.

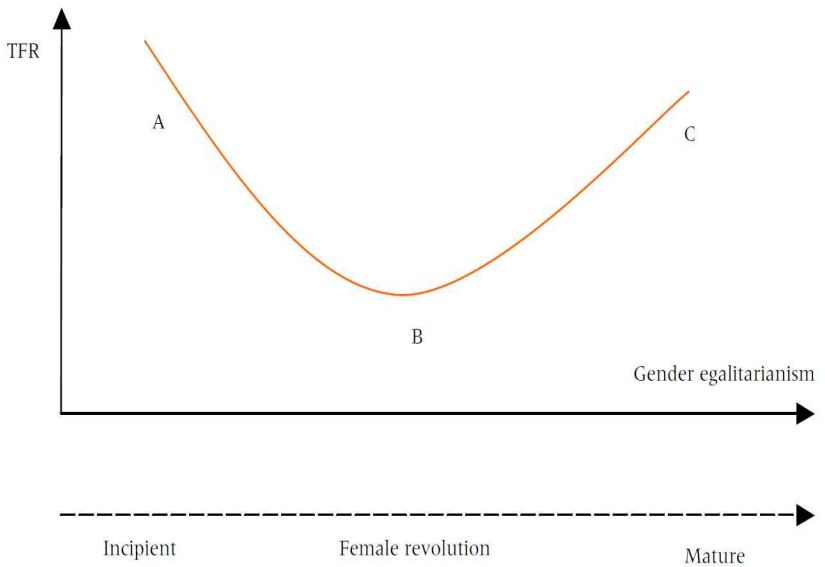
〈그림 3-1〉 합계출산율과 인간개발지수간의 횡단적 관계, 1975년과 2005년



출처: Myrskylä et al. (2009: 741)

또한, 에스핑-앤더슨과 빌라리(Esping-Anderson and Billari 2015)에 따르면(<그림 3-2>), 출산율의 변화 추이는 성평등주의의 발전정도와 U자형 관계를 맺는다. 이는 성평등주의 정도가 낮은 상태에서 상승하는 경우에는 출산율이 하락하지만, 성평등주의가 어느 정도 진전된 이후에는 성평등주의가 발전함에 따라 출산율이 반등함을 시사한다. 이러한 이론에 따르면, 현재 한국사회가 초저출산의 덫에 걸려있는 이유는 성평등주의가 충분히 발전하지 않았기 때문이며, 향후 성평등주의적 제도 및 의식이 확산된다면 출산율이 반등할 수 있음을 시사한다. 서구사회의 성별 가사노동 분담과 관련된 최근의 연구는 성별 가사노동 분담 정도와 합계출산율간의 강한 상관관계가 존재함을 보여주고 있는데, 이러한 연구결과는 성평등주의적 제도 및 의식의 확산이 출산율 반등을 위해서 매우 중요한 요인임을 보여준다(Garcia-Manglano et al. 2015).

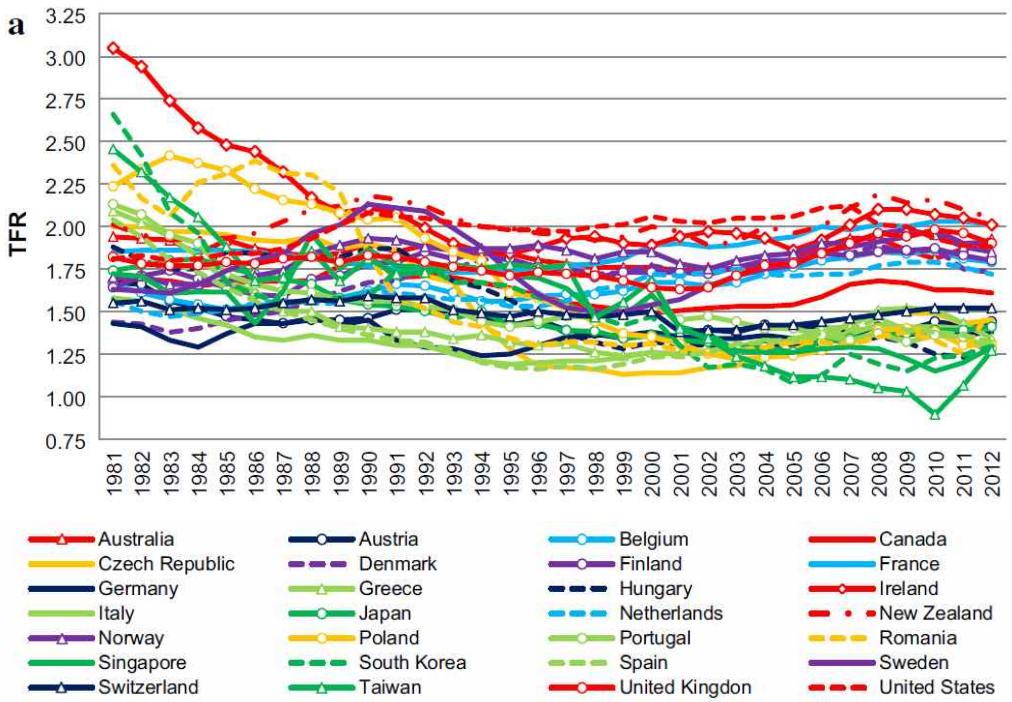
〈그림 3-2〉 다중상태 평형이론에서 가정하는 성평등주의의 확산과 출산율의 관계



출처: Esping-Anderson and Billari (2015: 9)

위의 논의들은 한국사회에서 출산율이 반등할 수 있는 가능성이 존재함을 시사한다. 즉, 사회경제적 발전 혹은 성평등주의의 진전에 따라 한국의 출산율이 반등할 수 있다는 것이다. 그런데, 이와는 대조적으로 한국이 저출산의 덫에서 벗어날 가능성에 대해서 부정적인 전망을 제시하는 연구 또한 존재한다. 최근의 한 연구에 따르면(Rindfuss et al. 2016), 후기 산업국가들의 출산율 변화 추이는 두 그룹으로 구분된다. 한 그룹은 출산율이 그리 낮지 않은 그룹이고(not too low), 다른 그룹은 초저출산 국가 그룹(lowest-low fertility)이다. <그림 3-3>은 이러한 패턴을 보여주고 있는데, 2000년대 이후 합계출산율의 변화 추이는 초저출산을 벗어나서 대체수준에 근접하는 그룹과 합계출산율 1.5 미만에서 벗어나지 못하는 그룹으로 나뉘고 있다. 린드푸스 등은 초저출산의 원인을 특정한 요인에서 찾으려는 시도는 바람직하지 않으며, 경제적/문화적/제도적 요인들을 하나의 묶음(package)로 보는 방식이 필요하다고 주장하고 있다. 이러한 주장은 학술적인 측면에서는 매우 타당하다고 할 수 있지만, 경제적/문화적/제도적 묶음이 단기간에 변화할 가능성이 매우 낮다는 점을 고려하면, 이러한 주장은 출산율의 반등이 단기간에 일어나기 어렵다는 점을 강조한다고 할 수 있다. 이러한 주장에 따르면, 장래 출산율 추계에 출산율의 반등을 고려하는 것은 이론적/경험적 타당성이 낮은 시도라고 할 수 있다.

〈그림 3-3〉 합계출산율 변화 패턴의 분화(divergence)



출처: Rindfuss et al. (2016: 290)

그렇다면, 장래 출산율 추계를 위해서 출산율 반등을 가정하는 것은 이론적/경험적 정당성을 가질 수 있는가? 사회경제적 발전과 성평등주의 발전에 따른 출산율의 반등이 경험적으로 확인되고 있기는 하지만(Esping-Anderson and Billari 2015; Myrskylä et al. 2009), 일부 국가들이 저출산의 늪에서 헤어나지 못하고 있으며 한국이 이러한 국가들의 대표적인 사례라는 연구결과를 고려할 때(Rindfuss et al. 2016), 이는 간단히 답할 수 있는 문제가 아니다. 결국 이 문제에 대한 해답은 경험적인 자료에 근거해야 하는데, 최근의 한 연구는 이와 관련해서 중요한 함의를 제공해준다. 박현준 등의 연구에 따르면(Park et al. 2013), 교육수준에 따른 혼인율 차이의 패턴에 최근 들어 변화가 일어나고 있다. 즉, 예전에는 교육수준이 낮은 집단의 혼인율이 교육수준이 높은 집단보다 높은 경향이 지속적으로 관찰되었는데, 최근 출생코호트의 경우에는 그 차이가 줄어들거나 심지어 역전되는 현상 또한 발견된다. 이를 교육수준의 전반적인 상승과 연결해서 해석하면, 혼인율의 상승을 예상할 수 있다. 즉, 혼인율이 상대적으로 높은 고학력 집단이 오히려 증가함에 따라 전체적인 혼인율은 증가할 가능성이 있으며, 이는 출산율의 상승으로 연결될 수 있기 때문이다.

제4절 분석자료

이 연구는 통계청 인구동향과에서 제공한 다음과 같은 자료를 활용한다. 1) 각세별 연령별 출산율(1980-2014). 2) 각세별 유배우율(1980-2014). 3) 각세별 유배우 출산율(1980-2014). 각세별 연령별 출산율은 15-49세 정보를 활용하는 반면, 각세별 유배우율과 유배우 출산율은 20-49세의 정보만을 활용한다. 이는 최근 자료의 경우 20세 미만의 유배우율이 지나치게 낮아서 이들에 대한 신뢰할만한 추정치가 존재하지 않기 때문이다. 따라서, 15-19세의 연령별 출산율의 경우에는 유배우율과 유배우 출산율을 나누어서 분석하지 않고, 연령별 출산율 정보만을 활용하도록 한다. 최근 들어 20세 미만의 출산율은 거의 무시할만한 수준이기 때문에 이들의 출산율에 대한 추정 방식이 전체 결과에 미치는 영향은 아주 미미할 것으로 판단한다.

제5절 1980-2014년 추세 분석

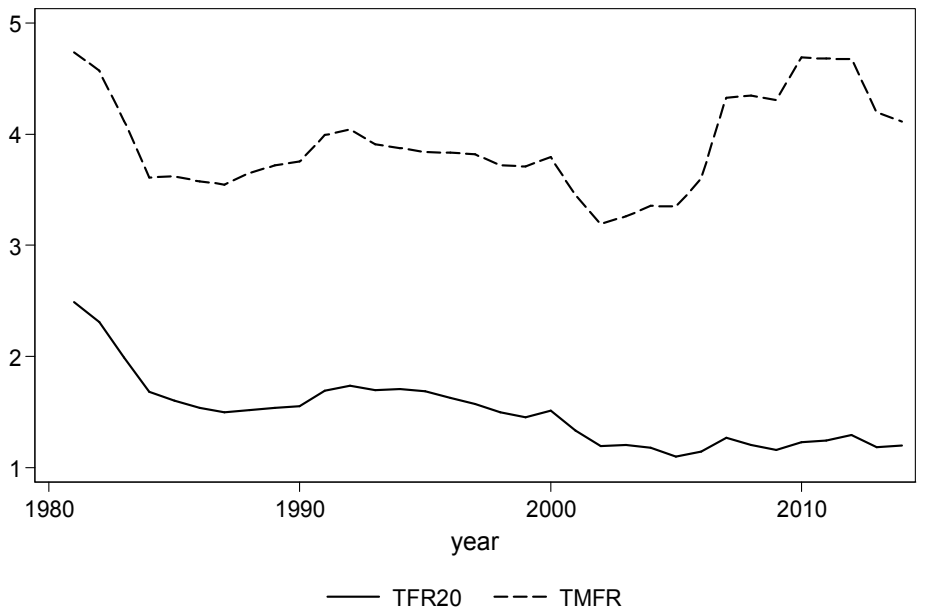
수식 (2)에 따르면(ASFR = ASMR*ASMFR*CR), 연령별 출산율(ASFR)은 연령별 유배우율(ASMR), 연령별 유배우출산율(ASMFR), 혼외출산 보정요인(CR)의 곱으로 표현할 수 있다. 제5절에서는 연령별 출산율을 구성하는 각각의 요인의 추이를 검토하고 Lee-Carter 모형을 활용하여 장래 추이 예측을 위한 기본적인 추정치들을 제시하도록 한다.

1. 합계출산율, 유배우 합계출산율, 유배우율, 혼외출산 보정요인의 변화 추이

<그림 3-4>는 1980-2014년 기간 동안의 합계출산율과 유배우 합계출산율의 추이를 보여주고 있다. 이 때 합계출산율 및 유배우 합계출산율은 20-49세 연령별 출산율과 연령별 유배우출산율을 합한 값을 하는데, 이는 20세 미만의 경우에는 연령별 유배우출산율에 대한 신뢰할만한 추정치를 얻기 어렵기 때문이다. 합계출산율의 경우에는 20세 미만 출산을 포함한 결과와 거의 동일한 결과를 얻을 수 있는 반면, 유배우 합계출산율의 경우에는 15-19세 유배우 여성의 출산율이 매우 높을 것으로 가정할 수 있기 때문에, 결

과에 큰 차이가 나타날 것으로 예상할 수 있다. 그렇지만, 20세 이전의 혼인이 매우 드물다는 점을 고려할 때, 20세 이상의 연령별 출산율 및 연령별 유배우 출산율 정보만을 활용한 분석이 오히려 적절하다고 할 수 있다.

〈그림 3-4〉 합계출산율 및 유배우 합계출산율 추이* (1980- 2014)



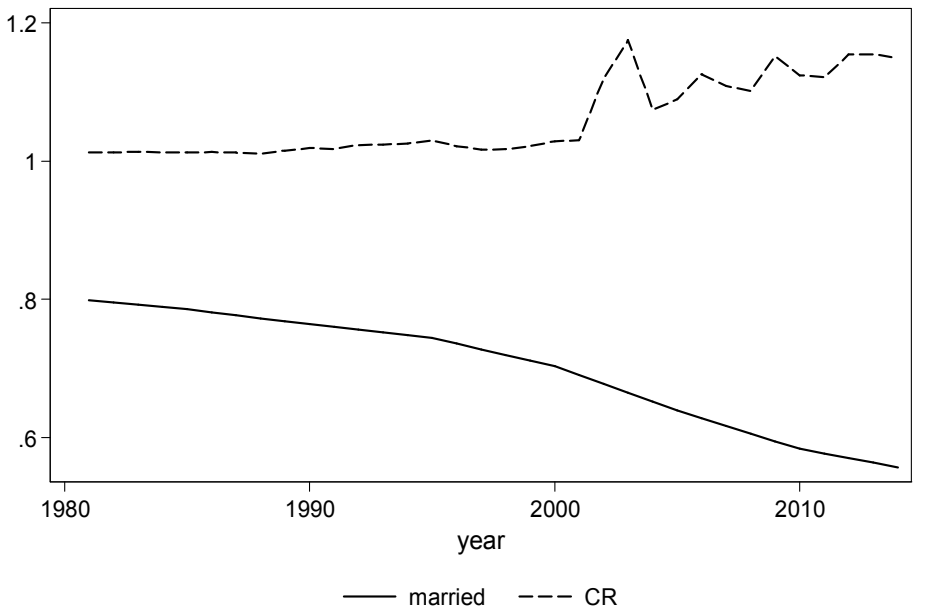
* 20세 이상의 연령별 출산율 및 연령별 유배우 출산율의 합

<그림 3-4>를 통해서 다음과 같은 패턴을 확인할 수 있다. 우선, 2000년대 초반 이전까지는 합계출산율과 유배우 합계출산율의 변화추이가 비슷했다. 변화의 폭에는 차이가 있지만, 유배우 합계출산율이 하락할 때 합계출산율도 하락하며, 유배우 합계출산율이 상승하면 합계출산율도 상승하는 경향을 관찰할 수 있다. 그러나 2000년대 이후에는 이러한 경향이 더 이상 관찰되지 않고 있다. 즉, 유배우 합계출산율은 전반적으로 상승해서 2010년대 초반의 경우에는 1980년대와 비슷한 수준을 보여주고 있는 반면에, 합계출산율은 지속적으로 낮은 수준에 머무르고 있는 것으로 관찰되고 있다. 합계출산율과 유배우 합계출산율의 변화 패턴의 분기(divergence)는 2000년대 들어 저출산의 원인이 유배우 출산율의 하락이 아니라 다른 것에 그 원인을 두고 있음을 강하게 시사한다. 식 (2)에 사용된 용어로 표현하자면, 이러한 변화는 유배우율과 혼외 출산 보정 요인의 변화가

이러한 분기를 설명한다고 할 수 있다.

<그림 3-5>는 유배우율과 혼외 출산 보정요인의 변화 추이를 보여준다. <그림 3-5>에 제시한 유배우율(married)와 혼외출산 보정요인(CR)은 수식 (2)에 사용된 연령별 유배우율과 연령별 혼외출산 보정요인의 각 년도별 단순평균 값이다. 따라서, 유배우율 지표의 경우에는 해당 연도 20-49세 여성의 실제 유배우율과는 다른 값이다. 실제 유배우율과는 다르지만, 이 연구의 목적이 연령구성의 영향을 고려하면서 장래 출생아수에 대한 추정치를 제시하는 것이 아니라 장래 인구추계를 위한 연령별 출산율에 대한 추정치를 제시하는 것이기 때문에, 연령구성의 영향을 배제한 단순평균이 오히려 적절한 정보라고 판단한다.

〈그림 3-5〉 유배우율 및 혼외 출산 보정요인 추이 (1980-2014)



<그림 3-5>를 통해서 다음과 같은 패턴을 확인할 수 있다. 우선, 연령별 유배우율의 단순평균은 1980년에는 80%였지만, 2010년대 들어 60% 미만으로 떨어졌다. 이는 이 기간 동안 혼인의 지연 혹은 포기가 빠른 속도로 이루어져 왔음을 보여준다. 한편, 혼외출산 보정요인의 경우에는 2000년대 초반까지는 대체로 1과 거의 일치하는 값을 보여주고 있는데, 이는 이 기간동안에 혼외 출산의 비중이 매우 낮았음을 반영한다. 그러나,

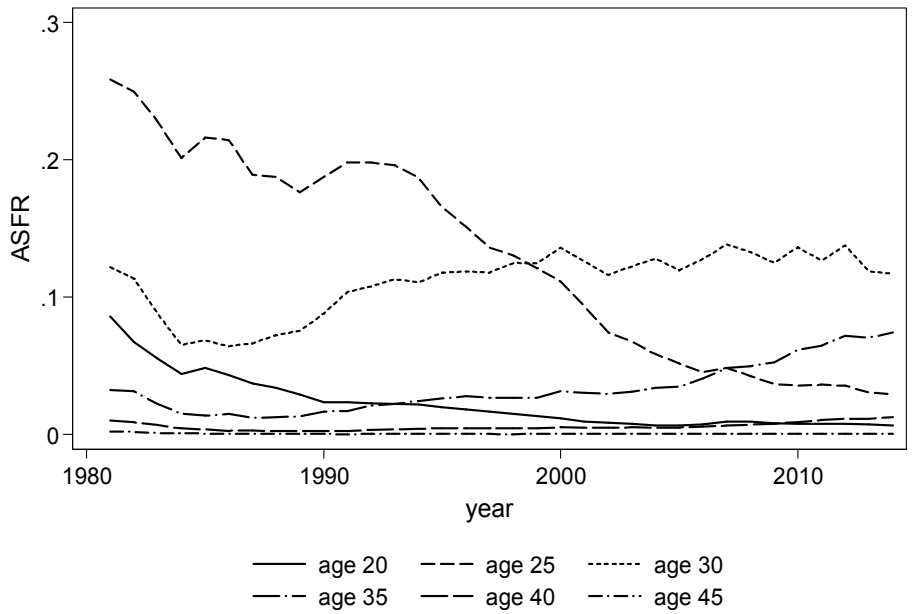
2000년대 이후에는 혼외 출산 보정요인이 등락을 거듭하면서 점차 상승하는 경향을 보여주고 있는데, 이는 이 기간 동안에 혼외출산의 영향력이 점차 증가했음을 의미한다. 이 때, 주의해야 할 것은 <그림 3-5>에 제시된 값이 혼외 출산보정요인의 단순평균이라는 점이다. 전체적으로 보면, 2010년대 들어서도 혼외출산이 전체 출산에서 차지하는 비중은 2-3% 정도밖에 되지 않기 때문에, 혼외출산의 절대적인 비중이 최근 들어 급증하지는 않았다. 그럼에도 불구하고, 2000년대 이후 혼외출산 보정요인의 단순평균값이 증가하는 경향을 보여주는 이유는 일부 연령대(출산율이 낮은 연령대)의 혼외출산 보정요인이 이 기간 동안에 상승했기 때문일 것으로 추정할 수 있다. 이를 종합해서 해석하면, 2000년대 이후 나타나고 있는 합계출산율과 유배우 합계출산유 변화 추이의 분기(divergence)는 기본적으로 2000년대 이후 나타나고 있는 혼외 출산요인의 증가 및 등락에 의해서 상당부분 영향을 받고 있다고 할 수 있겠다.

2. 연령별 합계출산율, 연령별 유배우 합계출산율, 연령별 유배우율, 연령별 혼외출산 보정요인의 변화 추이

앞서 논의한 합계출산율, 유배우 합계출산율, 유배우율, 혼외출산 보정 요인의 변화 추이를 보다 면밀히 검토하기 위해서는 연령별 변화 추이를 살펴보는 것이 필요하다. 우선, <그림 3-6>은 연령별 출산율의 변화 추이를 보여주고 있다. <그림 3-6>은 20, 25, 30, 35, 40, 45세 연령별 출산율의 변화추이를 보여준다. 우선, 40세와 45세 출산율의 경우에는 이 기간 동안 매우 낮은 수준으로 꾸준히 유지되고 있다. 이는 40세 이후의 고령 출산에는 이 기간 동안 큰 변화가 없었음을 보여준다. 한편, 20세와 25세의 출산율은 지속적으로 하락했다. 20세의 경우에는 1980년에는 8% 정도의 수준이었지만, 2010년대 들어서는 거의 무시할만한 수준으로 하락했으며, 25세의 경우에는 매우 빠른 속도로 출산율이 하락했다. 이와 대조적인 패턴을 보여주는 연령 집단은 30세와 35세이다. 이 연령 집단의 경우에는 1980년대 초반에는 출산율이 다소 하락하는 경향을 보여주다가 이후에는 상승하는 패턴을 보여주고 있는데, 30세의 연령별 출산율은 1980년대 후반 빠른 속도로 상승했으며, 1990년대 이후에는 연도별로 다소 등락을 반복하면서 전반적으로 상승하는 패턴을 보여주고 있다. 35세의 경우에는 1980년대 후반 이후 지속적으로 상승하는 패턴을 보여주고 있으며, 그 상승률이 2000년대 후반 이후 다소 빨라지는 경향을 보여주고 있다. 이러한 변화는 40대 출산율의 정체, 30대 출산율의 상승, 20대 출산율의 하락으로 요약할 수 있겠다. 결과적으로 25세 출산율은 1990년대 후반 이후 30세 출산율보다 낮아졌으며, 2000년대 중반 이후에는 35세 출산율보다도 낮아졌음을 확인할 수 있다. 이러한

분석결과는 한국사회에서 출산을 담당하는 주된 연령집단이 지난 30여년 동안 20대에서 30대로 이동했음을 보여주고 있다.

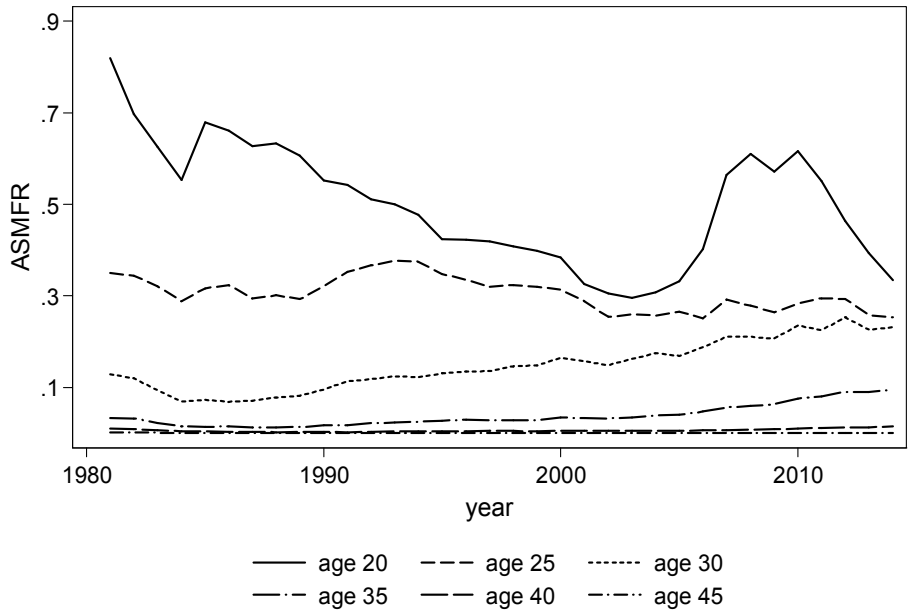
〈그림 3-6〉 연령별 출산율 변화 추이 (1980-2014)



<그림 3-7>은 20, 25, 30, 35, 40, 45세 연령별 유배우 출산율의 변화 추이를 보여주고 있다. 연령별 유배우 출산율의 변화추이는 앞서 살펴본 연령별 출산율과는 확연히 구분된다. 가장 뚜렷한 차이는 연령별 유배우 출산율은 연령이 증가함에 따라 감소하는 패턴을 보여준다는 점이다. 즉, 분석하고 있는 전 기간 동안 연령별 유배우 출산율은 연령이 증가함에 따라 감소하고 있다. 그렇지만, 그 변화 패턴은 연령별로 다르게 나타나고 있다. 20세의 경우에는 2000년대 초반까지 전반적으로 하락하는 추세를 보여주다가 2000년대 후반까지는 급증한 이후에 다시 2010년대 들어 급락하는 패턴을 보여주고 있다. 이러한 불안정한 패턴은 20세 유배우자의 수가 2000년대 들어 매우 적어졌음을 반영한다. 2000년대 들어 이 연령대의 유배우자의 수 자체가 매우 적어졌기 때문에, 몇몇 20세 유배우 여성의 출산이 전체 유배우 출산율에 미치는 영향이 커지고 이것이 20세 유배우 출산율의 심한 등락으로 이어졌다고 할 수 있다. 한편, 25세 유배우 출산율에는 큰 변화가 관찰되지 않고 있다. 25세 유배우 출산율은 1990년대 중반까지는 다소 상승하다가

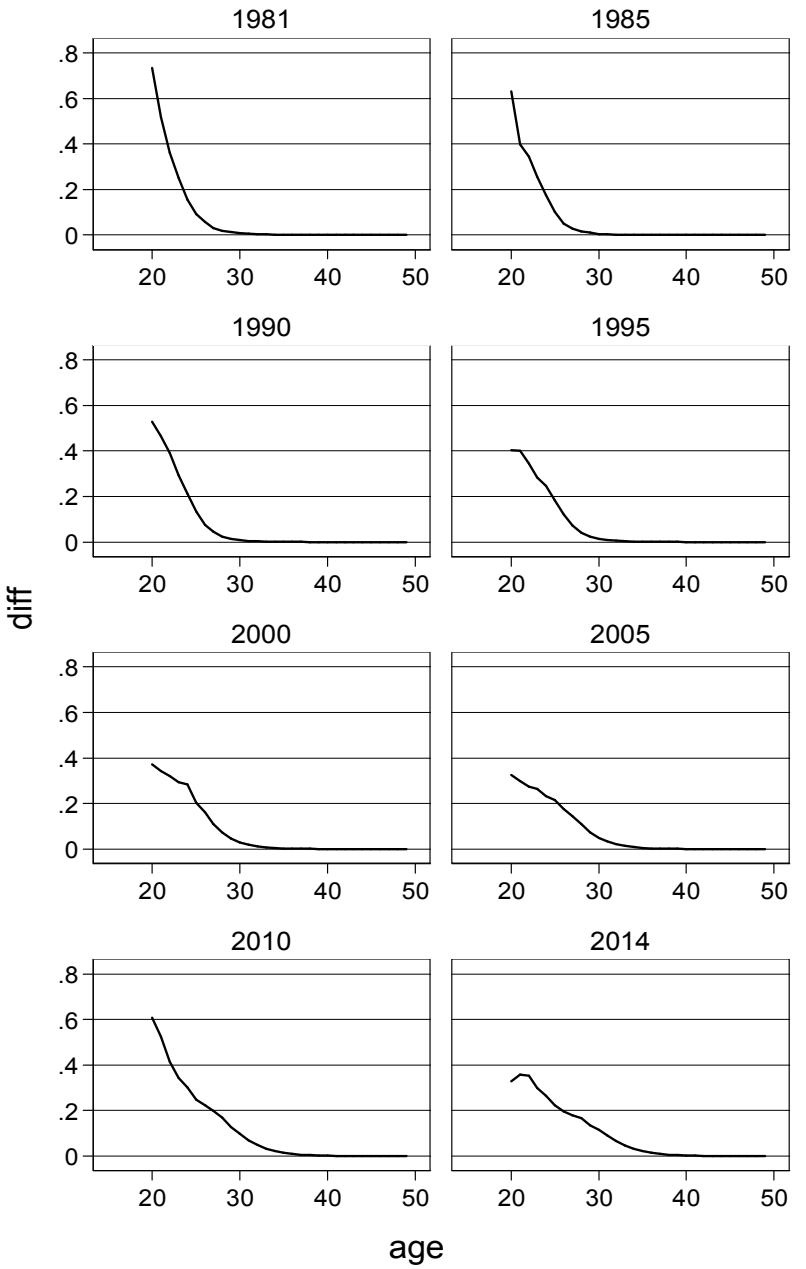
2000년대 초반까지 다소 감소했으며, 이후에는 뚜렷한 패턴 없이 작은 폭의 등락을 거듭하고 있다. 전체적으로 보면, 30% 수준에서 등락을 거듭하고 있다고 할 수 있다. 30세와 35세 유배우 출산율의 경우에는 지속적으로 증가하는 패턴을 보여준다. 30세 유배우 출산율의 경우에는 1980년대 초반에는 10%대 정도에 머물렀지만, 2010년대에 들어 20%를 넘어서고 있으며, 35세의 경우에도 2010년대에는 10%에 육박하고 있다. 이러한 변화는 <그림 3-6>에서 논의한 연령별 출산율의 변화추이와도 연결된다. 연령별 출산율뿐만 아니라 유배우 연령별 출산율에서도 30대가 차지하는 비중이 지속적으로 증가하고 있음을 보여준다고 할 수 있다.

〈그림 3-7〉 연령별 유배우 출산율 변화 추이 (1980-2014)



<그림 3-8>은 연령별 출산율과 연령별 유배우 출산율의 차이 변화를 보여주고 있다. 이 둘의 차이는 연령별 유배율과 연령별 혼외출산 보정요인의 변화에 의해서 설명된다고 할 수 있다. 즉, 유배우율이 낮거나 혼외출산의 비중이 높은 연령집단에서 이 차이가 크게 나타난다. <그림 3-8>에서 다음과 같은 연령별 패턴의 변화를 확인할 수 있다. 첫째, 최근에 올수록 연령별 출산율과 연령별 유배우 출산율이 수렴하는 연령대가 높아진다. 즉, 1980년대에는 30세 이전에 두 수치가 서로 수렴했으나, 2010년대에 들어서면 30

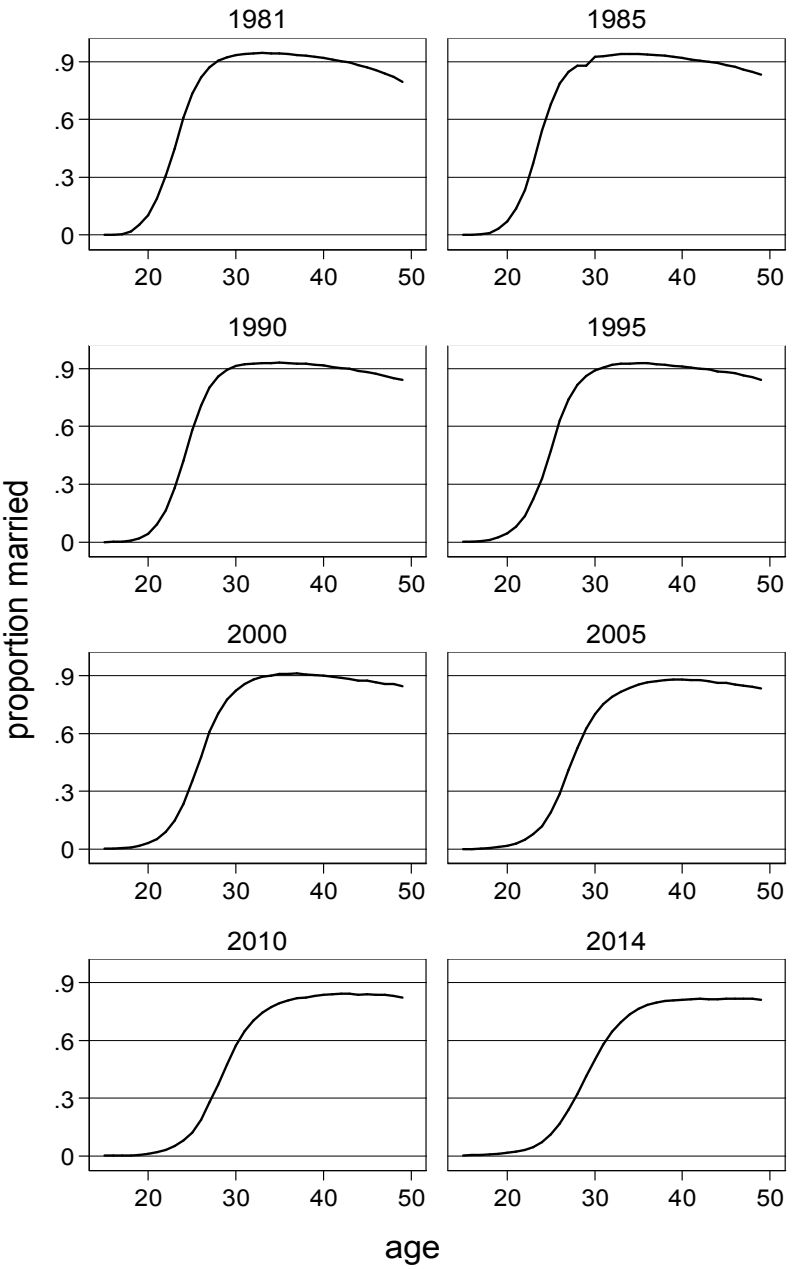
〈그림 3-8〉 연령별 출산율과 연령별 유배우 출산율의 차이 (1981-2014)



대 중후반이 되어서야 둘 간의 차이가 0에 가까운 값을 갖게 된다. 이는 혼인의 지연에 따른 유배우율의 하락에 그 원인이 있다고 할 수 있다. 둘째, 20대 초반의 경우에는 그 차이가 지속적으로 하락하는 추세를 보여주는데, 2010년에는 그 차이가 오히려 1990년 수준으로 급증했다. 20대 초반의 유배우율이 지속적이며 빠른 속도로 하락했다는 점을 고려하면, 이러한 변화는 20대 초반의 유배우 출산율 하락을 반영한다(<그림 3-7> 참고). 단, 2010년에 관찰되는 차이의 증가는 <그림 3-7>에 나타난 대로 20세 연령별 유배우 출산율이 2010년 전후가 갑자기 급증한 것을 반영한다. 이는 앞서 지적했듯이 유배우율의 전반적인 하락에 따른 20대 초반 유배우 여성의 절대적인 수가 급감으로 이 연령대의 유배우 출산율에 대한 안정적인 추정치를 얻기 어려운 상황을 반영한다고 할 수 있다. 또한, 연령별 출산율과 연령별 유배우 출산율의 차이의 변화는 혼외출산의 연령별 패턴의 변화 또한 반영할 가능성이 높은데, 이와 관련해서는 혼외출산 보정 요인에 대해서 살펴보면서 논의하도록 하겠다.

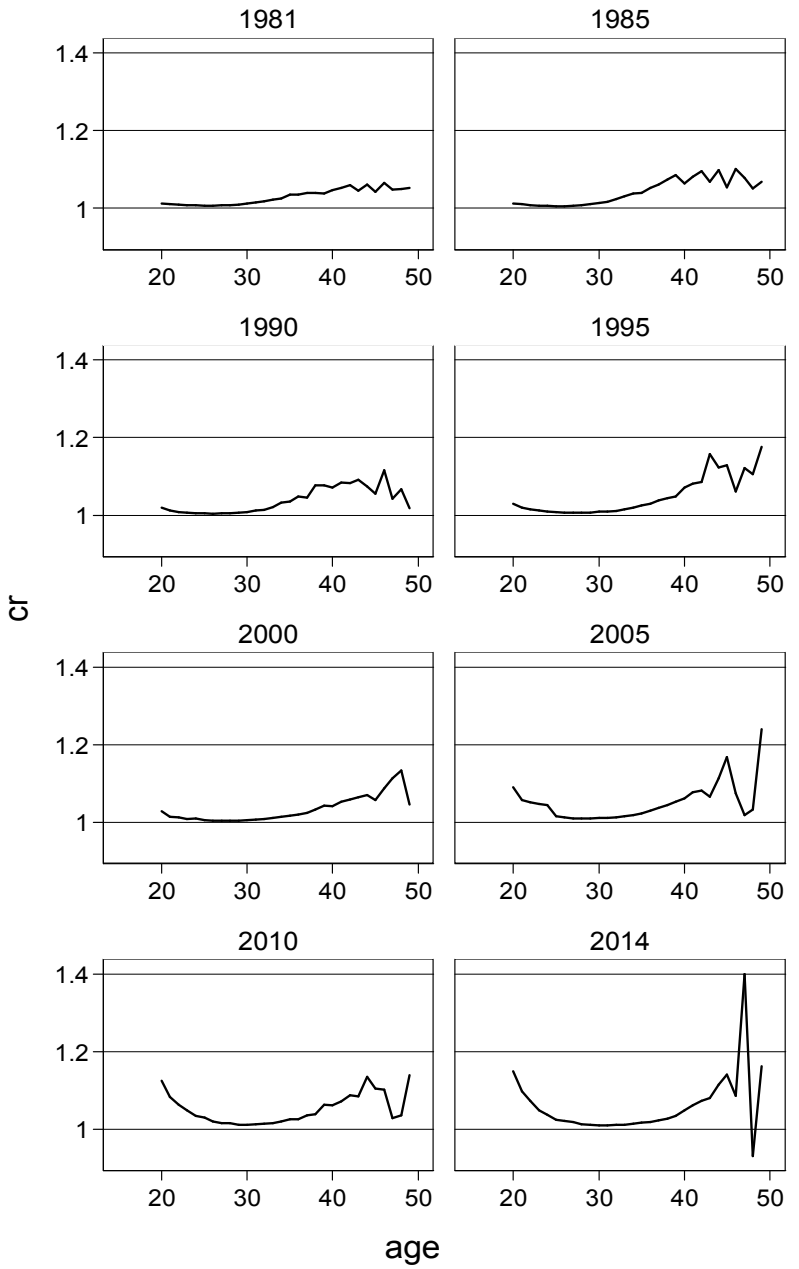
<그림 3-9>는 연령별 유배우율의 변화를 보여주고 있다. <그림 3-9>는 혼인의 지연에 따른 유배우율의 하락 경향을 명확하게 보여준다. 첫째, 1990년대까지는 30세를 전후해서 유배우율이 90%를 넘었으나, 2000년대 이후에는 유배율이 40대 이후에도 90%에 미치지 않고 있다. 연령별 유배우율의 상승 속도 역시 시간이 지나가면서 지속적으로 하락하고 있음을 확인할 수 있다. 결과적으로 2010년대에는 30세에도 유배우율이 60%에 못미치는 현상이 나타나고 있다. 둘째, 유배율이 2005년까지는 일정 연령대까지 증가하다가 이후에 감소하는 경향을 보여주었는데, 2010년대 들어서면 이러한 경향이 나타나지 않는다. 즉, 2010년 이후에는 유배우율이 40세를 전후에 80% 정도에서 최고점에 도달한 이후에 50세까지 뚜렷한 변화를 보여주지 않는다. 이러한 경향은 만혼과 혼인해체(marital disruption) 경향을 반영하는 것으로 해석할 수 있다. 여기서 혼인해체는 이혼과 사별을 포함하는데, 1980년대와 1990년대에는 40대 이상의 사별자 비율이 낮지 않았기 때문에 유배우율을 낮추었다고 해석할 수 있다. 사망률이 하락함에 따라 사별자 비율이 낮아졌지만 이 기간 동안 유배우율이 40대 이후로 비슷한 수준을 유지한 것은 이혼율이 증가했기 때문인 것으로 해석할 수 있다. 한편, 2010년대 이후 유배우율이 40대 이후에 비슷한 수준으로 유지되는 것은 만혼의 증가로 해석할 수 있다. 이 기간 동안에 이혼율은 낮지 않은 수준으로 유지되고 사별자 비율은 꾸준히 하락했는데, 만혼의 증가로 40대 이후 유배우자로 전환한 여성들이 증가함에 따라 40대 이후 유배우율이 하락하지 않은 것으로 해석할 수 있다. <그림 3-9>에 나타난 연령별 유배우율의 변화 패턴을 요약하면, 혼인의 지연으로 유배우율이 전반적으로 하락한 것이 가장 두드러진 특징이라고 할 수 있으며, 사별의 감소 및 만혼의 증가에 따라 최근 들어 40대 이후의 유배우율 감소 현상이 약화된 것 역시 지적할 수 있겠다.

〈그림 3-9〉 연령별 유배우율의 변화 (1981-2014)



<그림 3-10>은 연령별 혼외출산 보정요인의 변화를 보여주고 있다. 혼외출산 보정요인의 값은 연령별 출산율을 연령별 유배우율과 연령별 유배우출산율의 곱으로 나눈 값이다. 따라서, 혼외출산의 비중에 따라 이 값이 달라진다. 이 값이 1이면, 이는 해당 연령대 여성의 혼외출산이 전혀 없다는 것을 의미하며, 1보다 큰 값을 가지면, 이는 혼외출산의 비중이 높다는 것을 의미한다. 시기에 따라 혼외출산 보정요인의 연령 패턴은 변화하고 있음을 확인할 수 있다. 첫째, 1980년에는 혼외출산이 연령이 증가함에 따라 증가한다는 것을 의미하는데, 이는 미혼자의 출산보다 이혼/사별자의 출산이 더 많음을 시사한다. 20대의 경우에는 1990년대까지는 혼외출산 보정요인 값이 거의 무시할만한 수준이지만, 2000년대 중반이후에는 1.1-1.2의 값을 보여주고 있는데, 이는 2000년대 이후 20대 혼외출산의 비중이 늘어났음을 의미한다. 특히, 2010년대에는 이 값이 큰 폭으로 상승하는 것을 확인할 수 있다. 이는 <그림 3-7>에서 나타나는 20세 유배우 출산율의 급격한 상승을 고려할 때 다소 의아한 결과라고 할 수 있다. 즉, 유배우 출산과 혼외출산이 같이 급증했다고 할 수 있는데, 이러한 결과는 혼외출산 패턴 자체의 변화보다는 유배우율 및 연령별 출산율의 하락에 따른 결과로 해석하는 것이 합리적일 것 같다. 즉, 20대 초반 유배우율이 아주 낮은 수준으로 하락함에 따라(<그림 3-9> 참고), 20대 초반의 연령별 출산율 및 연령별 유배우율 자료가 적은 수의 출산에 의해서 큰 영향을 받은 결과로 해석할 수 있다. 이는 또한 <그림 3-8>에서 관찰한 연령별 출산율과 연령별 유배우율의 차이에도 영향을 미쳤다고 할 수 있다. 즉, 유배우율의 하락에 따라 20대 초반의 유배율이 매우 낮은 수준에 머무르게 됨에 따라 이 연령집단의 추정치들의 안정성이 떨어진 것이 연령별 출산율과 연령별 유배우 출산율의 차이를 크게 만든 것으로 해석된다. 둘째, 40대 이후의 혼외출산 보정요인 값은 시간이 지남에 따라 증가하는 추세를 보여줌과 동시에 연령별 진동(fluctuation)이 매우 심하다. 이러한 연령별 진동 역시 이 연령대의 연령별 출산율이 매우 낮기 때문에 나타나는 현상으로 해석할 수 있다. <그림 3-6>에서 확인할 수 있듯이 40대 이후의 연령별 출산율은 전 기간 동안 매우 낮은 수준으로 유지되고 있는데, 결과적으로 적은 수의 유배우 출산과 혼외 출산이 이 연령대에는 큰 영향을 미치기 때문에 연령별 패턴에 심한 등락이 나타나고 있는 것이다. 특히, 2014년의 경우에는 그 진폭이 매우 큰데, 이는 40대 후반의 출산이 매우 드문 현상이기 때문인 것으로 해석할 수 있다.

<그림 3-10> 연령별 혼외출산 보정요인의 변화 (1981-2014)



<그림 3-6>-<그림 3-10>을 통해서 확인한 연령별 출산율, 연령별 유배우 출산율, 연령별 유배우율, 연령별 혼외출산 보정요인의 변화 추이를 요약하면 다음과 같다. 첫째, 연령별 출산율의 변화는 20대 출산율의 하락, 30대 출산율의 상승 40대 출산율의 정체로 요약할 수 있다. 둘째, 연령별 유배우 출산율의 변화는 연령별 유배우 출산율의 변화와 유사한데, 20대 및 40대 유배우 출산율에는 큰 변화가 없고, 30대 유배우 출산율은 지속적으로 상승했다. 단, 연령별 출산율에 비해서 느린 속도로 변화했다. 셋째, 연령별 유배우율은 전 기간 동안 지속적으로 하락했으며, 그 하락 폭은 20-30대에서 더 크게 나타났다. 40대의 경우에는 사별의 감소와 만혼의 증가로 인해서 시기별 변화가 다른 연령대보다 작았다고 할 수 있다. 마지막으로 연령별 혼외출산 보정요인은 20대 초반과 40대에서 증가하는 추세를 보여주고 있는데, 이는 유배우율 및 출산율의 하락에 따라 이들에 대한 안정적인 추정치를 얻기가 어려워진 상황을 반영한다고 할 수 있다.

제6절 Lee-Carter 모형을 활용한 과거 변화추이 분석

제5절에서는 1980-2014년 자료를 활용해서 연령별 출산율, 연령별 유배우 출산율, 연령별 유배우율, 연령별 혼외보정 요인의 변화 추이를 기술했다. 이러한 분석결과에 바탕을 두고 제6절에서는 2015-2040년 기간 동안의 연령별 출산율에 대한 추정치를 개발하고자 한다. 이를 위한 수학적 모형으로는 Lee-Carter 모형을 활용하는데(Lee and Carter 1992), Lee-Carter 모형의 기본식은 다음과 같이 표현할 수 있다.

$$\log(m) = a_x + b_x * k_t + e_{xt} \quad (3)$$

(m: 연령별 율(rate), x: 연령, t: 시간)

Lee-Carter 모형은 원래 연령별 사망률 변화를 추정하기 위해 개발된 모형이다. 수식 (3)에서 log(m)의 연령별 사망률의 자연로그 값을 의미하는데, 연령별 사망률의 자연로그 값은 공통적인 연령별 사망 패턴을 나타내는 a_x, 사망률의 시기적 변화를 나타내는 k_t와 시기별 변화에 해당 연령이 어느 정도 민감하게 반응하는지를 나타내는 b_x의 곱, 그리고 연령 및 시기에 특수한 오차항(e_{xt})의 합으로 추정한다. 이 모형은 연령별 사망률 추정의 기본모형으로 활용되어 왔으며, 최근에 이를 보완한 다양한 모형들이 제시되었는데, 새로운 모형들 역시 그 기본적인 틀은 공유하고 있다고 할 수 있다(Li and Lee 2005). Lee-Carter 모형은 연령별 사망력 추정을 넘어 시계열 자료를 적합하는(fitting) 데 많이 활용되고 있는데, 이 연구 역시 Lee-Carter 모형을 활용하여 연령별 유배우 출산율,

연령별 유배우율, 연령별 혼외출산 보정요인 등에 대한 추정치를 개발하고, 이들을 수식 (2)에 제시된 모형을 활용하여 결합하여 연령별 출산율에 대한 추계치를 제시하도록 한다.

1. Lee-Carter 모형 적용: 연령별 유배우율 (1981-2014)

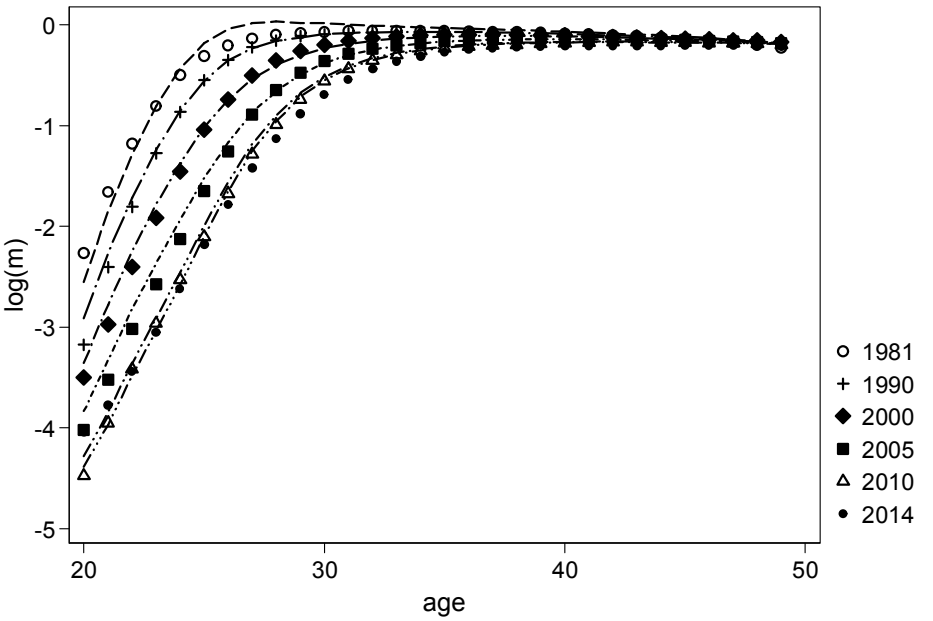
우선 Lee-Carter 모형을 활용하여 향후 연령별 유배우율/유배우출산율/혼외출산 보정요인을 추계하기 위해서는 과거 자료를 활용하여 각각에 대한 a_x, b_x, k_t 값을 확정하는 것이 필요하다. <표 3-1>은 연령별 유배우율 분석에서 추정된 값들을 제시하고 있다. a는 연령이 올라감에 따라 점차 0에 가까운 값을 취하는데, 이는 연령이 상승함에 따라 유배우율이 올라가는 경향을 반영한다. b값은 전반적으로 연령이 증가하면서 0에 수렴하는 경향이 있는데, 이는 유배우율의 전체적인 변화를 나타나는 k값에 대한 반응이 연령이 올라감에 따라 약화되기 때문이라고 할 수 있다. k값의 변화 패턴은 <그림 3-12>를 설명하면서 논의하도록 한다.

표 3-1 Lee-Carter 모형 모수 추정치 (유배우율)

Age	a	b	Year	k
20	-3.421	0.0905	1981	9.592
21	-2.856	0.1044	1982	9.171
22	-2.330	0.1082	1983	8.725
23	-1.864	0.1101	1984	8.248
24	-1.452	0.1060	1985	7.736
25	-1.090	0.0946	1986	7.265
26	-0.805	0.0798	1987	6.765
27	-0.583	0.0626	1988	6.230
28	-0.435	0.0488	1989	5.656
29	-0.329	0.0363	1990	5.034
30	-0.250	0.0280	1991	4.756
31	-0.202	0.0214	1992	4.468
32	-0.169	0.0168	1993	4.167
33	-0.147	0.0137	1994	3.853
34	-0.132	0.0113	1995	3.525
35	-0.122	0.0096	1996	2.899
36	-0.116	0.0083	1997	2.232
37	-0.114	0.0073	1998	1.520
38	-0.114	0.0067	1999	0.753
39	-0.116	0.0060	2000	-0.077
40	-0.117	0.0053	2001	-1.032
41	-0.122	0.0047	2002	-2.080
42	-0.126	0.0041	2003	-3.242
43	-0.130	0.0039	2004	-4.547
44	-0.139	0.0036	2005	-6.038
45	-0.144	0.0028	2006	-6.810
46	-0.152	0.0024	2007	-7.639
47	-0.162	0.0016	2008	-8.532
48	-0.171	0.0010	2009	-9.502
49	-0.184	0.0003	2010	-10.564
			2011	-10.568
			2012	-10.597
			2013	-10.648
			2014	-10.717

모형에서 적합된 값(fitted values)과 관측치들간의 일치 여부를 확인하는 작업 또한 필요한데, <그림 3-11>은 이를 보여주고 있다. 선으로 표시된 수치들은 모형에 따른 추정치이고 점으로 나타난 수치들은 관측값을 나타내고 있다.

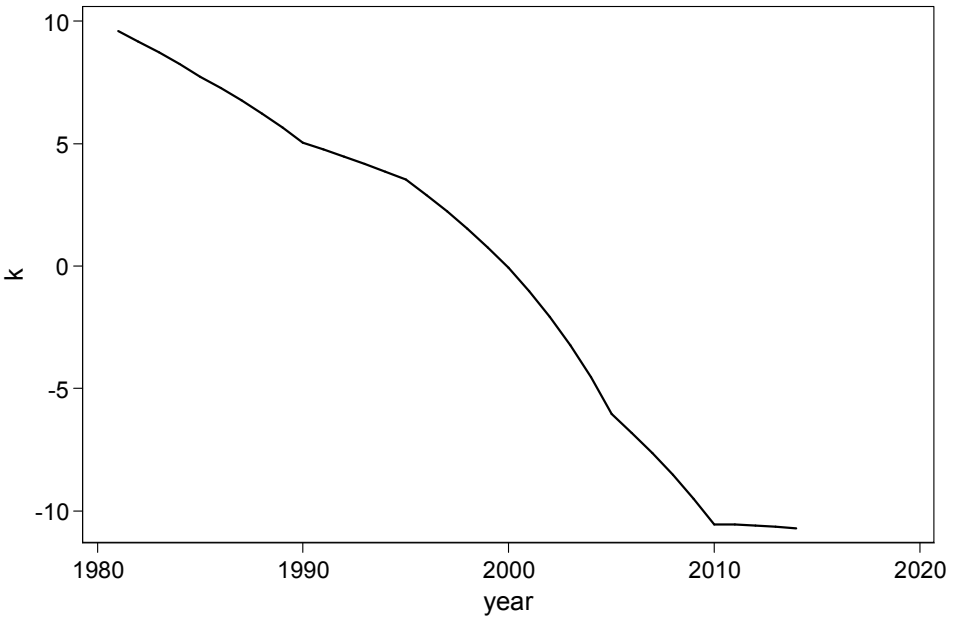
〈그림 3-11〉 Lee-Carter 모형에 따른 연령별 유배우율 추정



분석결과는 모든 연도에 Lee-Carter의 모형 적합도가 매우 높음을 보여주고 있다. 그런데, 시기별로 적합도에 다소 변화가 있음을 확인할 수 있다. 1981년의 경우에는 Lee-Carter 모형이 연령별 유배우율을 다소 과대추정(overestimate)하고 있는데 반해서, 2014년 자료의 경우에는 미세하지만 다소 과소추정(underestimate)하고 있다. 이는 이 모형을 기계적으로 적용하게 되면 장래 유배우율 추이를 과소추정할 수 있는 가능성이 높음을 시사한다. 또한, 유배우율의 변화 속도가 2010년대 들어 다소 느려지고 있음을 확인할 수 있다. 이는 1981-2014년 자료에서 관찰된 유배우율 하락 경향을 기계적으로 적용할 경우에 유배우율 하락 경향을 과대평가할 수 있음을 시사한다. 이는 <그림 3-12>와 <표 3-1>을 통해서 더욱 명확하게 파악할 수 있다. <그림 3-12>는 연령별 유배우율 k값의 변화를 보여준다. k값은 지속적으로 하락하는 패턴을 보여주는데, 변화 양상은 시기별로 다소 다르게 나타나고 있다. 특히 1990년대 후반에서 2010년까지는 이전보다 매우 빠른 속도로 하락하고 있는데, 이는 1997년 경제위기 이후 혼인이 급감했다는 것을 보여준다. 특이한 점은 2010년대 이후 k값이 거의 변화하지 않는다는 점이다. <표 3-1>에 따르면, k는 1980년대에 4.5, 1990년대에 4.0, 2000년대에 9.5 정도 하락했는데, 2010년에서 2014년 사이에는 불과 0.15 하락한 것으로 나타나고 있다. <그림 3-11>에서 나타나

듯이 Lee-Carter 모형이 한국의 연령별 유배우율 변화 추이를 매우 잘 추정하고 있다는 점을 고려할 때, 이러한 변화는 유배우율이 하락 현상이 2010년대 들어 뚜렷이 정체되고 있다는 중요한 경험적 증거라고 할 수 있다. 이는 향후 한국의 유배우율의 변화추이를 추정하는 데 매우 중요한 정보를 제공한다. 즉, 1981-2014 자료에서 관찰되는 k값의 연평균 변화율을 적용하여 향후 유배우율을 예측하면, 유배우율 하락을 과장할 수 있는 것이다. 앞서 검토한 <그림 3-11>은 Lee-Carter 모형이 최근의 유배율을 과소추정한다는 것을 보여주었는데, 이는 1981-2014년 자료에서 발견되는 패턴을 외삽하는 (extrapolating) 방식이 장래 유배우율 하락 추이를 과대추정할 가능성이 높다는 것을 시사한다. 따라서, 장래 유배우율 추계를 위해서는 <그림 3-12>에 나타나는 유배우율 하락의 둔화를 고려한 모형의 개발이 필요하다고 할 수 있다.

〈그림 3-12〉 Lee-Carter kt, 1981-2014 (유배우율)



2. Lee-Carter 모형 적용: 연령별 유배우 출산율 (1981-2014)

<표 3-2>는 Lee-Carter 모형을 적용한 연령별 유배우 출산율 분석에서 추정한 a_x, b_x, k_t 값을 제시하고 있다. 첫째, a값은 분석기간 동안 관찰된 평균적인 연령별 유배우 출산율을 보여주는데, 연령이 올라감에 따라 점차 부정적인 값을 지닌다. 이는 연령의 증가에 따라 유배우 출산율이 낮아지는 경향을 반영한다. 둘째, b값은 20대 초반에는 음의 값을 갖다가 20대 중반부터 양의 값을 가지며, 30대 중반까지 증가한 후 다시 감소하는 패턴을 보여주고 있다. b의 크기와 방향은 유배우 출산율의 시기적 변동을 나타내는 k 값의 변화에 해당 연령이 얼마나 민감하게 반응하는지를 보여준다. 이는 30대 유배우 출산율의 변화 방향이 시기적인 변화 방향과 일치하며 변화의 민감도가 크다는 것을 의미한다. 유배우 출산율이 하락하는 시기에는 30대 유배우 출산율이 가장 많이 하락하고, 반대로 유배우 출산율이 올라가는 시기에는 30대 유배우 출산율이 가장 많이 상승한다는 것을 보여준다. 마지막으로 k값의 변화 패턴에 대해서는 <그림 3-14>를 설명하면서 논의하도록 하겠다.

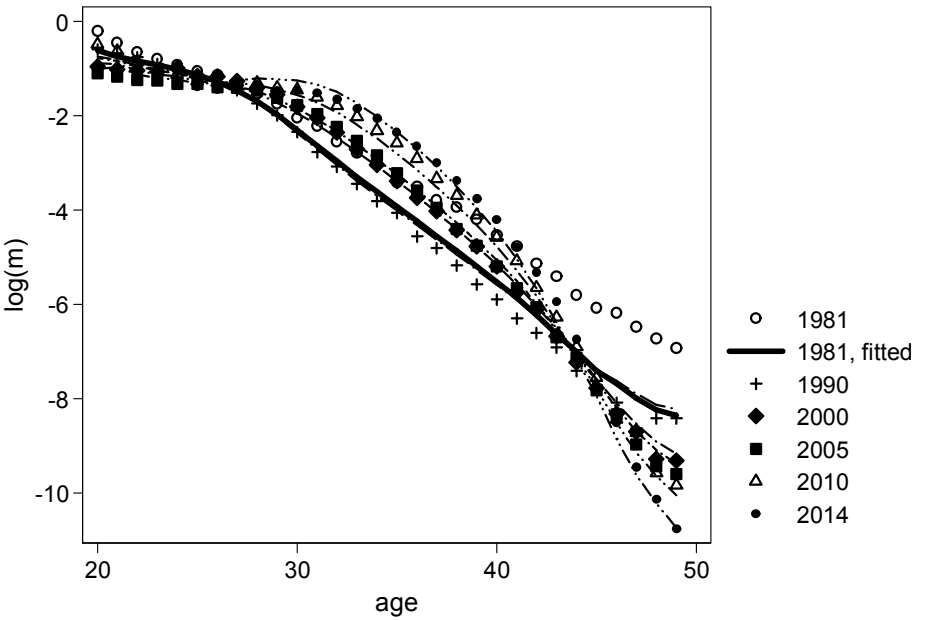
표 3-2 Lee-Carter 모형 모수 추정치 (유배우 출산율)

Age	a	b	Year	k
20	-0.727	-0.044	1981	-2.689
21	-0.813	-0.025	1982	-3.086
22	-0.919	-0.032	1983	-3.661
23	-1.008	-0.032	1984	-4.344
24	-1.076	-0.025	1985	-4.684
25	-1.193	-0.021	1986	-4.745
26	-1.282	-0.004	1987	-4.265
27	-1.396	0.023	1988	-3.851
28	-1.543	0.055	1989	-3.179
29	-1.747	0.085	1990	-3.097
30	-1.983	0.119	1991	-2.352
31	-2.237	0.146	1992	-1.685
32	-2.514	0.166	1993	-0.381
33	-2.820	0.177	1994	-0.846
34	-3.123	0.179	1995	-0.554
35	-3.444	0.180	1996	-0.015
36	-3.762	0.174	1997	0.300
37	-4.103	0.166	1998	0.323
38	-4.450	0.154	1999	0.322
39	-4.820	0.140	2000	0.888
40	-5.199	0.121	2001	0.785
41	-5.610	0.092	2002	0.650
42	-6.058	0.064	2003	0.862
43	-6.551	0.027	2004	1.128
44	-7.051	-0.017	2005	1.664
45	-7.567	-0.061	2006	2.554
46	-8.039	-0.131	2007	2.698
47	-8.487	-0.185	2008	3.172
48	-8.842	-0.224	2009	3.577
49	-9.083	-0.270	2010	3.861
			2011	3.681
			2012	4.833
			2013	6.169
			2014	5.969

유배우율과 마찬가지로 모형에서 적합된 값(fitted values)과 관측치들간의 일치 여부를 확인하는 작업 또한 필요한데, <그림 3-13>은 이를 보여주고 있다. 선으로 표시된 수치들은 모형에 따른 추정치이고 점으로 나타난 수치들은 관측치들이다. 분석결과는 Lee-Carter 모형이 유배우 출산율을 전반적으로 잘 추정하고 있음을 보여주고 있는데, 몇 가지 패턴을 주목할 수 있다. 가장 주목할만한 것은 1981년의 경우 Lee-Carter 모형의 적합도가 상당히 낮다는 것이다. 1981년의 경우 20대 중반 이후 Lee-Carter 모형은

꾸준히 연령별 출산율을 과소추정하고 있으며, 이러한 오차는 고연령대에서 증가하는 경향을 보여주고 있다. 한편, 이러한 오차가 시간이 지남에 따라 줄어들고 모형의 적합도가 높아지는 경향이 관찰된다. 또한, <그림 3-13>은 유배우 출산율이 최근 들어 상승하고 있음을 보여준다. 이러한 경향은 특히 30대에서 두드러지게 나타난다. 1980년대에는 30대 유배우 출산율이 하락하는 경향이 나타나다가 1990년 이후에는 꾸준히 증가하는 경향이 관찰된다. 이는 앞서 논의한대로 한국사회에서 출산의 주 연령대가 20대에서 30대로 넘어가는 경향을 반영한다. 이러한 경향은 장래 출산율을 추정하는 데 매우 중요한 함의를 갖는다. 즉, 30대 유배우 출산율의 증가가 향후 지속될 수 있다면, 이것이 출산율 반등으로 이어질 수 있기 때문이다. 그렇지만, 이러한 예측을 하기에 앞서 두 가지 조건을 고려해야 한다. 첫째, 최근에 나타나고 있는 30대 유배우 출산율 상승은 부분적으로는 혼인의 지연을 반영한다. 즉, 초혼 시기가 30대에 집중되면서 이들의 유배우 출산율이 높아졌기 때문이다. 그런데, 유배우율 변화를 살펴보면서 논의했듯이 유배우율의 하락은 최근 들어 둔화되고 있기 때문에, 30대 여성의 유배우 출산율이 지속적으로 상승할 것이라고 기대하기는 어렵다. 둘째, 유배우 출산율의 상승에는 분명히 수학적인 한계가 존재한다. <그림 3-13>은 유배우 출산율의 자연로그 값을 보여주는데, 자연로그 값이 0과 같은 경우에는 해당 연령의 유배우 출산율이 1과 같다. 즉, 해당 연령대의 모든 유배우 여성이 그 해에 출산하게 될 때 이 값이 0과 같게 된다. 물론, 이는 매우 비현실적인 상황이라고 할 수 있다. 분석자료에서 확인할 수 있는 가장 높은 연령별 출산율은 1981년 20세의 1명당 0.82명이고, 30대의 경우에는 2012년 30세의 1명당 0.25명이다. 생물학적인 가임력 등을 고려할 때, 30대의 유배우 출산율이 1명당 0.5명 이상으로 증가할 가능성은 거의 없다고 할 수 있는데, <그림 3-13>에서 제시된 경향이 지속된다고 가정한다면 향후 30대의 연령별 유배우 출산율이 이러한 한계치를 넘는 값을 가질 가능성이 존재한다. 따라서, 장래 출산율 추계를 위해서는 분석자료에서 나타나는 변화추이를 단순 외삽하는(extrapolating) 방식은 바람직하지 않다고 할 수 있겠다.

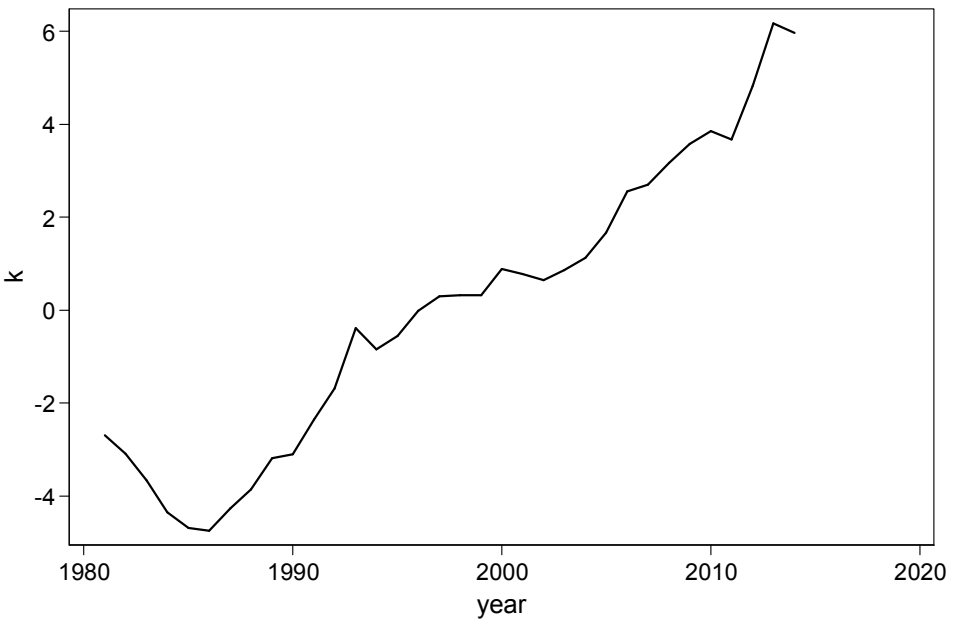
〈그림 3-13〉 Lee-Carter 모형에 따른 연령별 유배우 출산율 추정



<그림 3-14>는 Lee-Carter 모형에서 추정된 k값의 추이를 보여준다. 1980년대 중반까지는 k값이 감소했는데, 이후에는 지속적으로 증가하는 패턴을 보여준다. 1981-2014년 전체 기간을 놓고 볼 때, k값의 증가율은 연평균 0.38인데, 이를 외삽하게 되면 2040년에 이르면 k값이 19에 달한다. 이 경우에는 연령별 유배우 출산율이 1이 넘는 연령대가 발생할 것인데, 앞서 지적했듯이 이는 비현실적이라고 할 수 있다.

<표 3-2>, <그림 3-13>, <그림 3-14>에서 제시한 Lee-Carter 모형을 적용한 연령별 유배우 출산율 분석결과는 연령별 유배우 출산율은 연령별 유배우율과는 대조적으로 증가하는 추세에 있음을 보여준다. 이러한 증가추세를 단순 외삽하게 되면, 연령별 유배우 출산율에 대한 비현실적인 추정을 할 수밖에 없을 것이다. 이는 연령별 유배우 출산율 상승에 대한 현실적인 제약(constraints)을 적용해서 장래 연령별 유배우 출산율을 추계해야 한다는 점을 시사한다.

<그림 3-14> Lee-Carter kt, 1981-2014 (유배우 출산율)



3. Lee-Carter 모형 적용: 혼외출산 보정요인 (1981-2014)

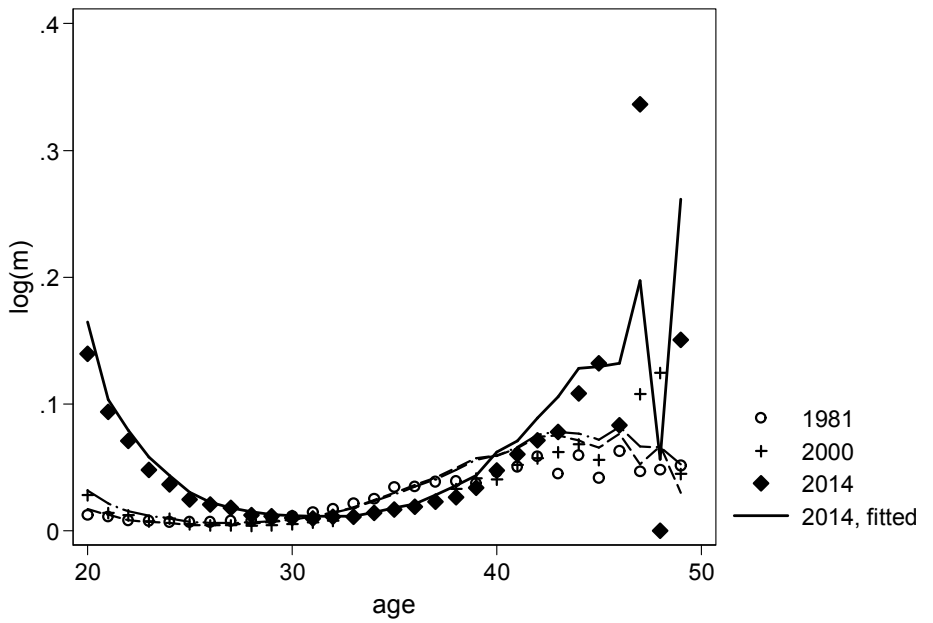
<표 3-3>는 Lee-Carter 모형을 적용한 혼외출산 보정요인 분석에서 추정한 a_x, b_x, k_t 값을 제시하고 있다. 첫째, a값은 분석기간 동안 관찰된 평균적인 혼외출산 보정요인의 연령패턴을 보여주는데, 대체적으로 U자형 형태를 보임을 알 수 있다. 이는 혼외출산이 20대 초반과 40대 이후에 상대적으로 집중되어 있음을 보여준다. 둘째, b값은 20대 초반에는 상대적으로 큰 값을 갖다가 20대 중반에서 40대 중반까지는 0에 가까운 값을 갖는 것을 확인할 수 있다. 20대 초반의 경우에는 시기적인 변화를 나타내는 k값의 변화에 민감하게 반응하는 반면에 다른 연령대에서는 큰 변화가 관측되지 않음을 보여준다고 할 수 있다. 마지막으로 k값의 변화 패턴에 대해서는 <그림 3-16>를 설명하면서 논의하도록 하겠다.

표 3-3 Lee-Carter 모형 모수 추정치 (CR)

Age	a	b	Year	k
20	0.057	0.150	1981	-0.266
21	0.037	0.093	1982	-0.232
22	0.028	0.072	1983	-0.201
23	0.021	0.052	1984	-0.171
24	0.016	0.038	1985	-0.180
25	0.011	0.027	1986	-0.208
26	0.009	0.018	1987	-0.066
27	0.008	0.014	1988	-0.229
28	0.008	0.010	1989	-0.252
29	0.009	0.006	1990	-0.290
30	0.010	0.003	1991	-0.083
31	0.011	0.000	1992	0.055
32	0.014	-0.003	1993	0.143
33	0.017	-0.007	1994	-0.082
34	0.022	-0.010	1995	0.125
35	0.027	-0.012	1996	0.092
36	0.032	-0.015	1997	-0.262
37	0.038	-0.014	1998	-0.130
38	0.046	-0.014	1999	-0.169
39	0.054	-0.014	2000	-0.164
40	0.060	0.003	2001	-0.339
41	0.067	0.006	2002	-0.097
42	0.078	0.015	2003	0.012
43	0.083	0.031	2004	-0.133
44	0.087	0.058	2005	0.249
45	0.083	0.065	2006	0.114
46	0.092	0.056	2007	0.063
47	0.091	0.147	2008	0.231
48	0.064	-0.010	2009	0.536
49	0.093	0.236	2010	0.169
			2011	0.279
			2012	0.173
			2013	0.716
			2014	0.597

앞서 논의한 유배우율, 유배우 출산율과 마찬가지로 모형에서 적합된 값(fitted values)과 관측치들간의 일치 여부를 확인하는 작업 또한 필요한데, <그림 3-15>은 이를 보여주고 있다. 선으로 표시된 수치들은 모형에 따른 추정치이고 점으로 나타난 수치들은 관측값을 나타낸다.

〈그림 3-15〉 Lee-Carter 모형에 따른 연령별 혼외출산 보정요인 추정

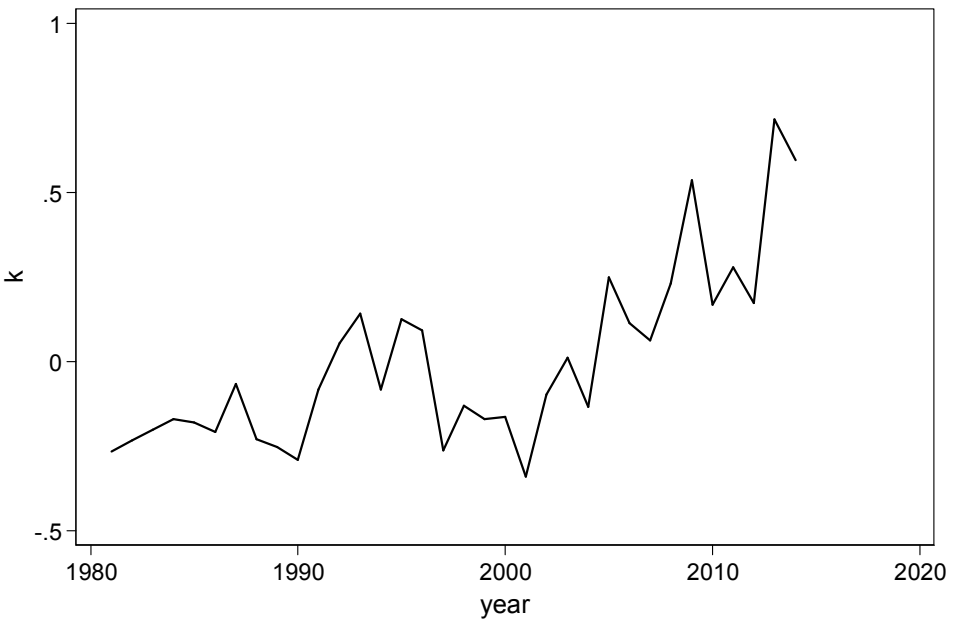


<그림 3-15>에 나타난 결과는 Lee-Carter 모형이 혼외출산 보정 요인을 대체로 잘 추정하고 있지만, 최근 들어 고연령대의 패턴을 추정하는 데 어려움이 있음을 보여준다. 즉, 2014년의 경우 40대 중반 이후 혼외출산 보정요인의 등락이 매우 심하게 나타나고 있는데, 그 결과 혼외출산 보정 요인에 대한 추정값이 관측치와 적지 않은 차이를 보여준다.

<그림 3-16>은 Lee-Carter 모형을 활용한 혼외출산 보정요인 분석에서 추정한 k값의 변화를 보여주고 있다. 이 수치는 전체적으로 보면 혼외출산 보정요인의 변화 추이를 보여준다고 할 수 있는데, 연도가 진동이 매우 심한 것으로 확인할 수 있으며, 전반적으로 혼외출산 보정요인의 값이 증가하고 있음 또한 확인할 수 있다.

연령별 혼외출산 보정요인의 변화 패턴은 연령별 유배우율과 연령별 유배우 출산율의 변화와 비교할 때 매우 불규칙적으로 변화하고 있음을 확인할 수 있으며, 이는 연령별 혼외출산 보정요인의 추계를 어렵게 한다고 할 수 있다. 그렇지만, 상대적으로 a와 b의 절대값이 매우 작고, k값 변화의 절대적인 양 또한 연령별 유배우율과 연령별 유배우 출산율보다 작게 나타나고 있기 때문에, 혼외출산 보정요인에 대한 추계치가 전체 출산율 추계에 미치는 영향을 그리 크지 않을 것으로 예상할 수 있다.

〈그림 3-16〉 Lee-Carter kt, 1981-2014 (혼외출산 보정요인)



제7절 장래 연령별 출산율 및 합계출산율 추계

제6절의 분석결과는 다음과 같이 요약할 수 있다. 연령별 유배우율은 지속적으로 감소하고 있으며, 연령별 유배우 출산율은 다소 증가하고 있는데, 특히 30대의 증가율이 빠르다. 또한, 연령별 혼외출산 보정요인은 연도별로 등락을 거듭하면서 증가하는 경향을 보여준다. 이 연구는 수식 (2)에 제시한 모형에 따라 연령별 출산율을 연령별 유배우율, 연령별 유배우 출산율, 연령별 혼외출산 보정요인으로 분해해서 각각에 대한 추정치를 구한 뒤 이를 종합하여 연령별 출산율에 대한 장래 추계치를 개발하는 것을 목적으로 한다. 그런데, 앞서 지적했듯이 1981-2014년 자료의 분석결과에 기반해서 향후 변화 추세를 단순외삽(simple extrapolation)하는 경우에는 비현실적인 추정치가 생산될 가능성이 존재한다. 이러한 문제점을 피하기 위해서 이 연구는 다음과 같은 몇 가지 시나리오를 상정하고 각각의 시나리오에 따른 추계치를 제시하도록 하겠다.

1. 장래 출산율 추계를 위한 시나리오

첫째, 제6절에서 제시된 추세가 2040년까지 지속된다고 가정한다. 이 때, 각각의 구성요인의 a와 b값은 그대로 적용하며, k값의 경우에는 연도를 독립변수로 사용하고, 각 연도의 k를 종속변수를 하는 회귀식을 통해서 향후 k값 연간 변화율을 추정한 후, 이 변화율이 2040년까지 지속될 것으로 가정한다. <표 3-4>는 연간 변화율 추정치를 보여주고 있다. 분석결과에 따르면, k값의 연간 변화율은 유배우율의 경우에는 -0.685, 유배우 출산율의 경우에는 0.304, 혼외출산 보정요인의 경우에는 0.019로 추정되고 있다. 첫 번째 시나리오에서는 이러한 연간 변화율이 지속된다고 가정한다.[2] 이 방법은 앞서 지적했듯이 비현실적인 연령별 유배우율/유배우 출산율/혼외출산 보정요인 추정치를 생산할 가능성이 높은데, 현실적인 제약을 부여한 결과와 비교하기 위한 참고자료로 활용한다. 이 시나리오는 유배우출산율의 상승을 가장 중요한 특징으로 하기 때문에 "유배우 출산율 상승"이라 칭한다.

표 3-4 K에 대한 회귀분석 결과

	유배우율		유배우 출산율		혼외출산 보정	
	Coef.	s.e.	Coef.	s.e.	Coef.	s.e.
annual change	-0.685	0.022	0.304	0.014	0.019	0.003
s.e.e	1.245		0.821		0.180	
R^2	0.969		0.934		0.537	

둘째, random walk의 방식을 활용하여 k값의 장래 추계치를 생산한다. 이 때 a와 b는 <표 3-1>-<표 3-4>의 값을 그대로 활용하며, k값의 경우에는 <표 3-4>에 제시된 연간 변화율을 사용한다. 그런데, <표 3-4>에서 제시된 연간변화율에는 불확실성이 존재할 수밖에 없다. 이 연구에서는 <표 3-4>에 제시된 표준추정오차(standard error of estimate, see)를 활용하여 이러한 불확실성을 반영하고자 한다. 즉, 이 연구는 불확실성이 과거추세 자료를 OLS 회귀식으로 적합할 때 발생하는 오차, 즉 표준추정오차에 비례한다고 가정한다. 예를 들어, 이는 1981-2014년 연령별 유배우율 자료를 사용한 Lee-Carter 분석의

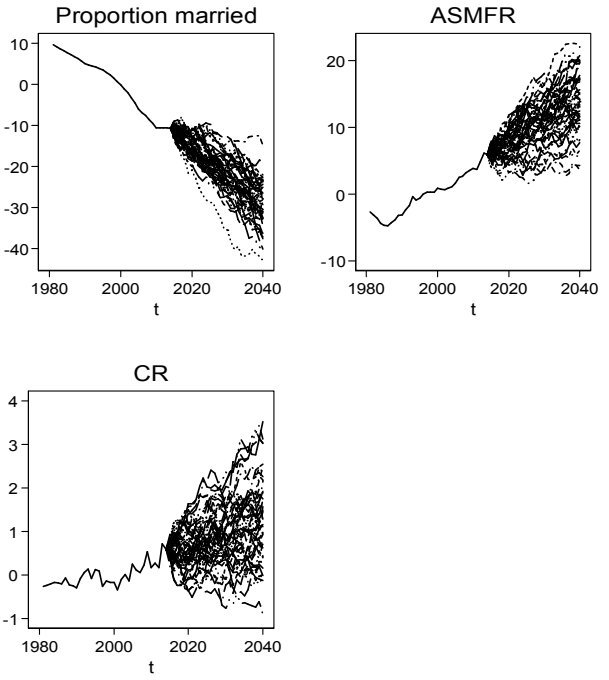
2) 일반적으로 이와 같은 시계열 자료의 경우에는 OLS 회귀식이 아니라 ARIMA 모형을 활용하는 것이 일반적이다. 그런데, 연령별 유배우율과 연령별 혼외출산 보정요인의 ARIMA 모형이 k값의 변화추이를 비교적 잘 추정하고 있는 반면에, 연령별 유배우 출산율의 경우에는 ARIMA 모형이 연간 변화율을 지나치게 높게 추정하고 있다(rate of change = 1.41). 따라서, 이 연구에서는 ARIMA 모형을 활용하지 않고 OLS 모형을 활용하여 k값의 장래 변화율을 추정하도록 한다.

결과에 따르면 k값이 1년에 0.685씩 감소한다. 그렇지만, 이 회귀모형의 표준추정오차는 1.245이며, 이는 회귀식을 활용하여 추정한 k값에 오차가 있음을 의미한다. 이러한 불확실성을 고려하기 위한 방식으로 random walk 방식을 사용할 수 있는데, 이는 다음의 수식 (4)로 표현할 수 있다. 즉, 각각의 k 값은 <표 3-4>에 제시한 변화율, 표준추정오차 및 표준정상분포를 따른다고 가정되는 변수(random variable)의 함수로 추정할 수 있다.

$$k_t = k_{t-1} + \text{annual change} + \tau * \text{s.e.e}, \ \tau \sim N(0, 1) \qquad (4)$$

이러한 방식으로 연령별 유배우율/유배우 출산율, 혼외출산 보정요인에 대한 2015-2040년 기간 동안의 추계치를 생산할 수 있다. Random Walk을 50회 실시한 결과는 <그림 3-17>에 제시하고 있다. 이 결과는 k값의 변화 추이에 대한 예측에 상당한 불확실성이 존재함을 보여준다.

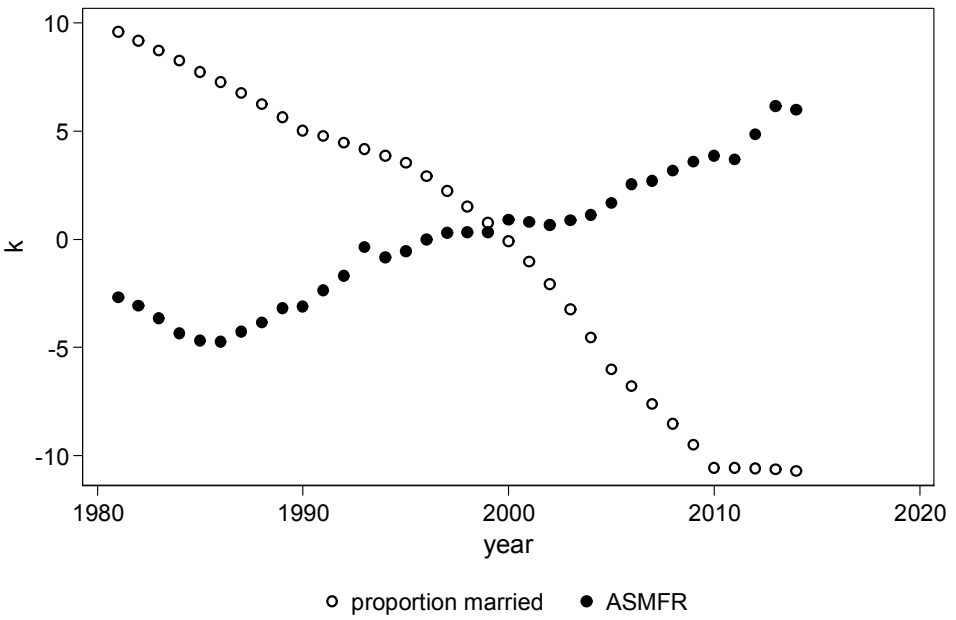
〈그림 3-17〉 유배우율, 유배우 출산율, 혼외출산 보정요인의 k값 추정(random walk)



<그림 3-17>은 k값 추정의 불확실성을 보여준다는 장점이 있지만, 한편으로는 특정

한 추정치를 제공하지 못하는 약점이 있다. 따라서, 이 연구에서는 몇 가지 경우를 조합하여 random walk를 활용한 추계치를 작성하고자 한다. 제6절에서 살펴본 1981-2014년의 유배우율 및 유배우 출산율 추이 분석을 통해서 이 기간 동안 관찰된 추세가 지속적으로 관찰될 것으로 가정하면, 유배우율 및 유배우 출산율의 장래 추계치에 대해서 비현실적인 추정치가 생산될 가능성이 높다는 것을 확인했다. 이는 시뮬레이션 결과에서 평균 혹은 중위값을 선택하는 것이 그리 바람직하지 않음을 시사한다. 시뮬레이션 결과를 선택할 때, 다음과 같은 점을 고려하도록 한다. 우선, 유배우율과 유배우 출산율 간에 부정적인 상관관계가 있음을 고려한다. <표 3-1>과 <표 3-2>은 유배우율 k는 감소하고, 유배우 출산율 k는 증가하는 추세가 있음을 보여준다. <그림 3-18>은 이를 보다 분명하게 보여준다.

〈그림 3-18〉 연령별 유배우율과 유배우 출산율 k의 관계



이러한 음의 상관관계를 고려하여 이 연구에서는 유배우율과 유배우 출산율 시뮬레이션 결과의 백분위 값들을 다음과 같이 조합하여, 연령별 출산율을 추정한다. 우선 각 요수의 k값의 예측치를 식 (4)에 제시된 방식으로 1,000회의 시뮬레이션을 통해서 추정한다. <그림 3-17>에 제시했듯이 이러한 방식은 각각의 k값 변화 추정의 불확실성을 반

영한다. 이 때, <그림 3-18>에 제시한 유배우율과 유배우 출산율을 고려하여, 유배우율 k값으로 큰 값을 택한 경우에는 유배우 출산율 k값은 작은 값을 택한다. 또한, 연령별 유배우율의 경우에는 k값이 감소하는 추세이기 때문에 각 연도의 추정치 중에서 큰 값을 택하고, 연령별 유배우 출산율은 k값이 증가하는 추세이기 때문에 각 연도의 추정치 중에서 작은 값을 택하도록 한다. 이 때, 유배우율과 유배우 출산율 간에 부정적인 상관관계가 있음을 고려하여, 이들의 조합은 <표 3-5>와 같이 한다. 또한, 혼외출산 보정요인의 경우에는 유배우 출산율과 같은 정도로 변화할 것이라 가정한다. 이러한 다섯 가지 조합에 의해서 연령별 출산율 및 합계출산율에 대한 추계치를 생산한다. 이러한 방식은 "random walk"이라 칭한다.

표 3-5 Random walk 조합

	조합				
	1	2	3	4	5
유배우율	99%	95%	90%	75%	50%
유배우 출산율	1%	5%	10%	25%	50%
혼외출산 보정요인	1%	5%	10%	25%	50%

셋째, 제3절에서 논의한 HDI 및 성평등주의와 출산율 관계의 역전을 반영하기 위해서 k값이 증가하는 경우에 출산율을 추정한다. <표 3-4>에 제시한 결과는 유배우율 k가 연평균 0.685 감소했으며, 유배우 출산율 k는 연평균 0.304 증가했음을 보여준다. 따라서, 이 연구에서 활용하고 있는 분해 모형을 통해서 출산율의 반등을 모형화하는 경우의 수는 여러 가지라고 할 수 있다. 즉, 유배우율의 감소 추세가 역전된다고 가정할 수도 있고, 유배우 출산율의 상승 추이가 지속 혹은 가속화된다고 가정할 수도 있다. 이철희(2012)의 연구에 따르면, 1990년대 이후 출산율의 하락은 기본적으로 유배우율의 하락에 기인한다. 따라서, 유배우율의 하락이 유배우 출산율 변화보다 최근의 한국의 출산력 변화를 설명하는 데 그 설명력이 높다고 할 수 있다. 이 연구에서는 이러한 기존 연구 결과를 바탕으로 유배우율의 k값은 2015-2040년 기간 동안 연평균 0.3-0.8 정도의 폭으로 증가하고, 유배우 출산율과 혼외출산 보정요인의 k값은 <표 3-5>에 제시한 조합 1과 조합 2의 값을 갖는 경우의 출산율을 추정한다. 이러한 시나리오는 "유배우율 반등"이라 칭한다.

이상의 네 가지 시나리오를 통해서 연령별 출산율 및 합계출산율을 추정한다. 이 때, 15-19세의 연령별 출산율은 2010-2014년의 평균값을 사용한다. 20세 이전의 연령별 유배우율 및 연령별 유배우 출산율 정보가 부재하기 때문에 이러한 선택을 했는데, 이 연령

대의 출산율이 매우 낮은 수준이고, 향후 증가할 가능성도 매우 낮기 때문에 이 연령대의 출산율에 대한 가정이 합계출산율 추정치에 미치는 영향을 매우 작다고 할 수 있다. <표 3-6>은 각각의 시나리오의 특징을 간략히 제시하고 있다.

표 3-6 출산율 추계 시나리오 특징 요약

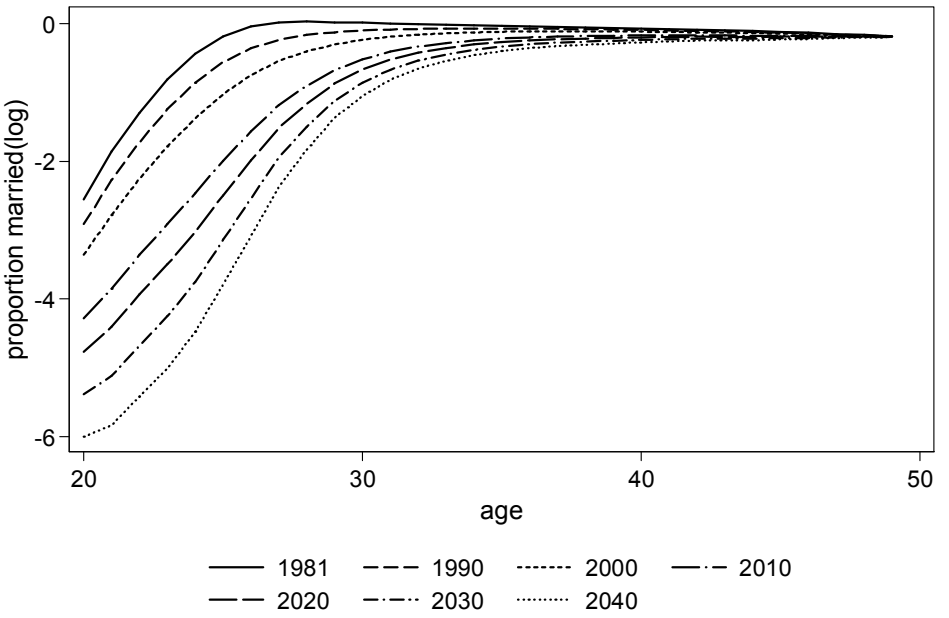
	가정	장점	단점
유배우 출산율 상승	1981-2014 추세가 지속	제공 자료에 충실	비현실적
random walk	1981-2014 자료 활용 불확실성 반영	불확실성 반영	시나리오 조합 자의성
유배우율 반등	2015년 이후 유배우율 반등	출산율 반등 모형화 이론적 정당성	자의성

2. 연령별 유배우율, 연령별 유배우 출산율, 연령별 혼외출산 보정요인 추계결과

1. 연령별 유배우율 추계

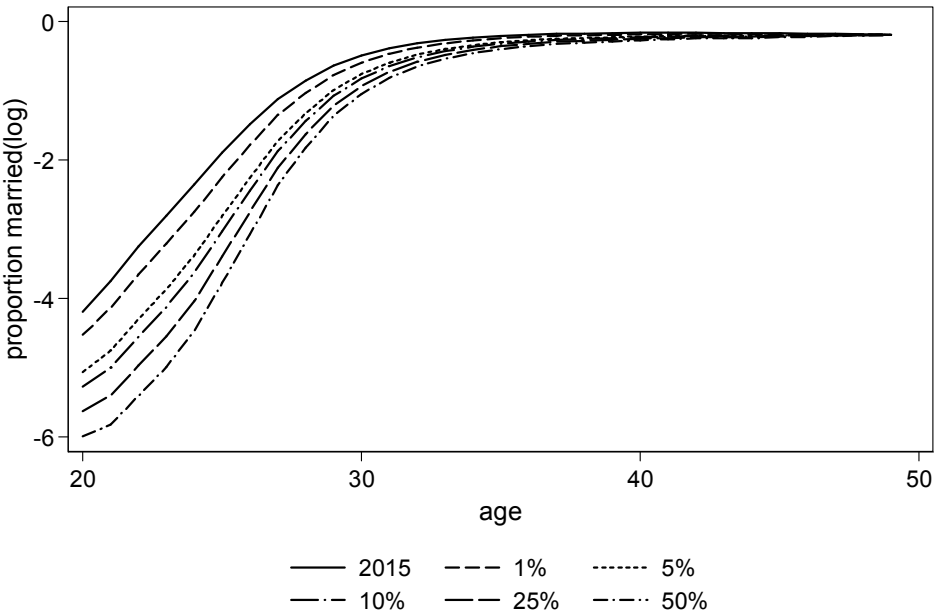
<그림 3-19>는 유배우 출산율 방법을 통해서 추정한 연령별 유배우율의 자연로그값 추이를 보여준다. <그림 3-19>에서 확인할 수 있듯이, 기존의 추세변화가 지속된다고 가정하는 유배우출산율 상승 시나리오에 따르면, 연령별 유배우율은 지속적으로 하락하는 경향을 보여주고 있는데, 특히 20-30대의 하락이 두드러진다.

〈그림 3-19〉 연령별 유배우율 추계치 (유배우 출산율 상승)



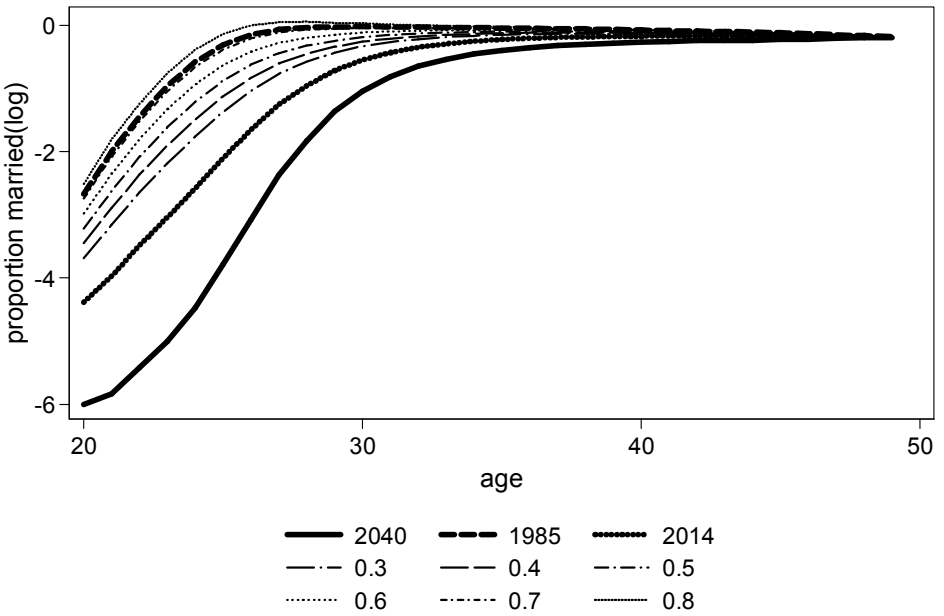
<그림 3-20>은 random walk 방식을 활용한 시뮬레이션을 통해서 추정한 2015년과 2040년의 연령별 유배우율 추계치를 보여준다. 실선은 2015년의 유배우율 추계치를 보여주고 있다. 나머지 선들은 2040년의 추계치를 시나리오 별로 보여주고 있다. 유배우율 변화가 가장 작다고 가정하는 1% 시나리오의 경우에는 2015년과 비교적 유사한 값을 보여주는 반면, 그 변화율이 중위값이라고 가정하면 유배우율의 하락이 상당히 가파르다는 것을 확인할 수 있다. 중위값의 경우에는 <그림 3-19>에 제시한 2040년의 추계치와 매우 유사한 값을 갖는데, 이는 1981-2014년에 관찰된 유배우율 감소 추세가 앞으로도 지속되는 경우에는 유배우율을 급속히 감소시키고 이는 출산율의 하락으로 연결될 것임을 시사한다.

〈그림 3-20〉 연령별 유배우율 추계치, 2015년과 2040년 (random walk)



<그림 3-21>은 유배우율 반등을 가정한 결과를 보여준다. 가장 아래에 위치한 굵은 선으로 표시된 결과는 <그림 3-19>에 제시한 유배우 출산율 상승 시나리오를 적용했을 때의 유배우율 추이를 보여주고, 굵은 선으로 표시된 다른 두 결과는 1985년과 2014년의 결과를 각각 보여준다. 1985년 결과를 제시하는 이유는 1980년대 중반 한국의 합계출산율이 대체 수준에 근접했기 때문이다. 점선으로 표시된 다른 결과들은 연령별 유배우율 k값이 연평균 0.3 - 0.8 상승한다는 가정 하에서의 2040년 연령별 유배우율 추계치를 보여준다. 이 결과를 통해서 알 수 있는 것은 유배우율이 연평균 0.7-0.8 상승한다면, 연령별 유배우율이 1985년의 수준으로 회복됨을 보여준다. 현실적으로 이와 같은 유배우율 반등이 가능한지는 확신하기 어렵다. 오히려 이러한 결과는 연령별 유배우 출산율에 큰 변화가 없었다는 선행연구 결과(이철희 2012)를 고려할 때, 대체수준 출산율을 회복하기 위해서는 혼인율이 향후 급격히 반등해야 함을 보여준다고 할 수 있다.

〈그림 3-21〉 연령별 유배우율 추계치 1985, 2014, 2040년 (유배우율 반등 시나리오)

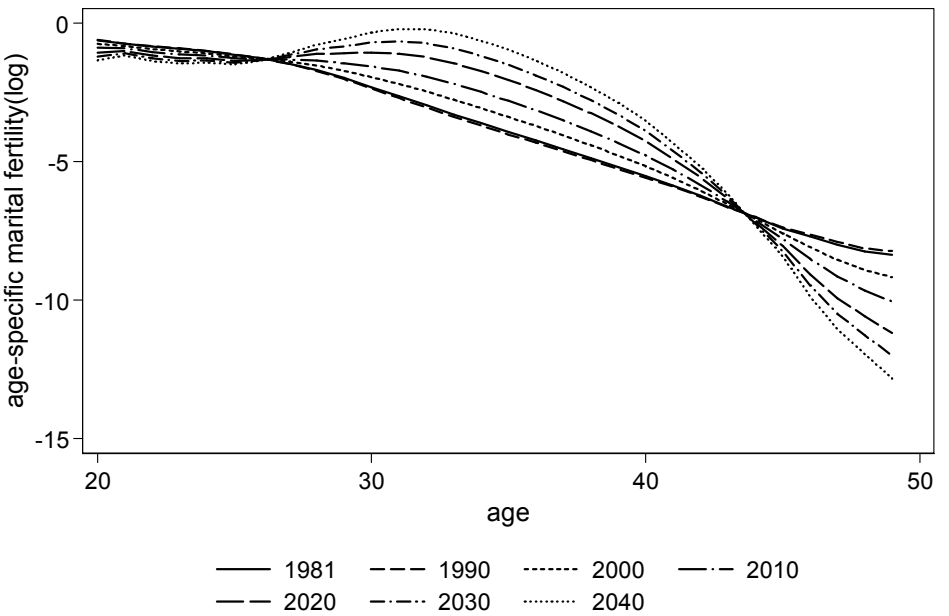


2. 연령별 유배우 출산율 추계

<그림 3-22>는 유배우 출산율 상승 시나리오를 통해서 추정한 연령별 유배우 출산율의 자연로그값 추이를 보여준다. <그림 3-22>에서 확인할 수 있듯이, 유배우 출산율 상승 시나리오에 따르면, 연령별 유배우율은 지속적으로 상승하는 경향을 보여주고 있는데, 특히 30대의 상승이 두드러진다. 즉 1981-2014년에 관찰된 추세가 지속된다고 가정한다면, 30대의 출산율은 가파르게 증가하는 패턴을 보여주는데, 이는 두 가지 요인에 기인한다. 첫째, <그림 3-14>과 <그림 3-17>에서 제시했듯이 유배우 출산율 k값은 증가하는 경향이 있는데, 이것이 유배우 출산율의 전반적인 상승으로 연결된다고 할 수 있다. 둘째, <표 3-2>에 제시한 유배우 출산율 b값의 연령 패턴을 보면, b값이 30대에 상대적으로 큰 양의 값을 갖고 있는데, 연령패턴(b)과 시기별 변화추세(k)가 상호작용하면서 이 연령대의 출산율이 매우 높은 수준으로 증가하는 패턴을 보여준다. 물론, 지난 30여 년간 관찰된 결과는 30대 출산율 및 유배우 출산율이 증가하여 30대가 출산의 주연령대가 된다는 것을 보여주고 있지만, <그림 3-22>에 제시된 것과 같은 2020년대 이후 30대 유배우 출산율의 수준은 비현실적으로 높다고 할 수 있다. 이러한 유배우 출산율 상승은 물론 유배우율의 하락과 동시에 진행된다고 할 수 있기 때문에, 이 연령대의 출산율의 증가는 <그림 3-22>에 나타난 것보다 작다고 할 수 있지만, 이러한 예측은 현실성이 다

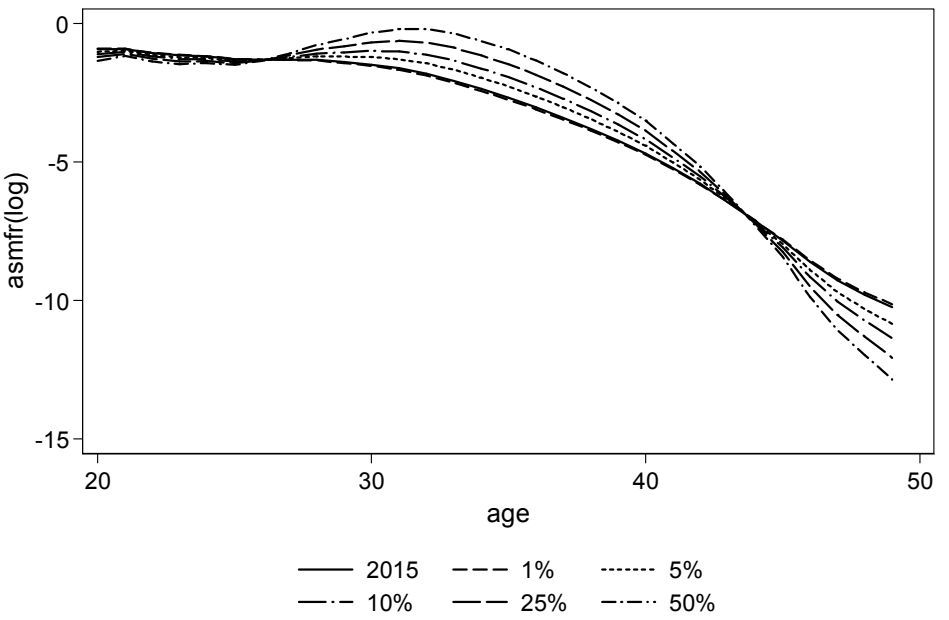
소 떨어진다고 할 수 있다.

〈그림 3-22〉 연령별 유배우 출산율 추계치 (유배우 출산율 상승)



<그림 3-23>은 random walk 방식을 통해서 추정한 유배우 출산율의 변화를 보여준다. <그림 3-23>은 <그림 3-20>에서와 같이 2015년 추계치와 서로 다른 시나리오별 2040년의 추계치를 보여준다. 연령별 유배우 출산율의 변화를 가장 작게 추정하는 1% 시나리오의 경우에는 2015년의 값과 매우 유사한 값을 보여주고 있으며, 중위값의 경우는 <그림 3-23>에 제시한 유배우 출산율 상승 시나리오에 따른 2040년 추계치와 유사한 결과를 보여준다. 현실적으로 보았을 때, 유배우 출산율의 증가추이가 <그림 3-23>의 중위값 혹은 유배우 출산율 상승 시나리오(<그림 3-22>에 제시한 것처럼 가파를 가능성은 높지 않다고 할 수 있다. 따라서, <그림 3-23>의 1-10% 정도의 시나리오가 현실성이 높다고 할 수 있다. 한편, 유배우 출산율과 관련해서는 출산율 반등 시나리오를 적용하지 않는다. 이는 유배우 출산율이 증가하는 추세를 보여주고 있기 때문에, 새로운 가정을 사용하여 유배우 출산율의 반등을 고려할 필요성이 낮기 때문이다.

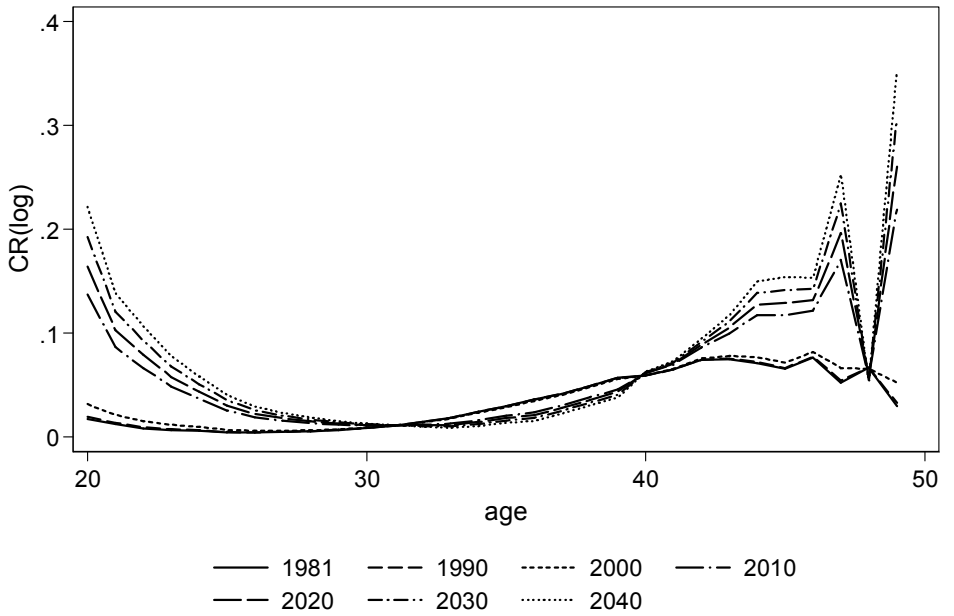
〈그림 3-23〉 연령별 유배우 출산율 추계치 (random walk)



3. 연령별 혼외출산 보정요인 추계

<그림 3-24>는 유배우 출산율 상승 방법을 통해서 추정한 혼외출산 보정요인의 자연로그값 추이를 보여준다. <그림 3-24>에서 확인할 수 있듯이, 혼외출산 보정요인 역시 유배우 출산율과 마찬가지로 증가하는 추세를 보여주는데, 이러한 증가추세는 20대와 40대 이상에서 두드러진다. 그렇지만, 출산이 30대에 집중되고 있고 이 연령대의 출산율이 낮은 수준에 머무르고 있다는 점을 고려할 때, 혼외출산 보정요인의 변화가 출산율에 큰 영향을 미치지는 않을 것으로 예상할 수 있다. 따라서, 혼외출산 보정요인의 경우에는 random walk 및 출산율 반등에 대한 추계결과 역시 별도로 제시하지 않는데, 이는 혼외출산 보정요인이 연령별 출산율에 미치는 영향이 크지 않기 때문이다.

〈그림 3-24〉 혼외출산 보정요인 추계치 (유배우 출산율 상승)

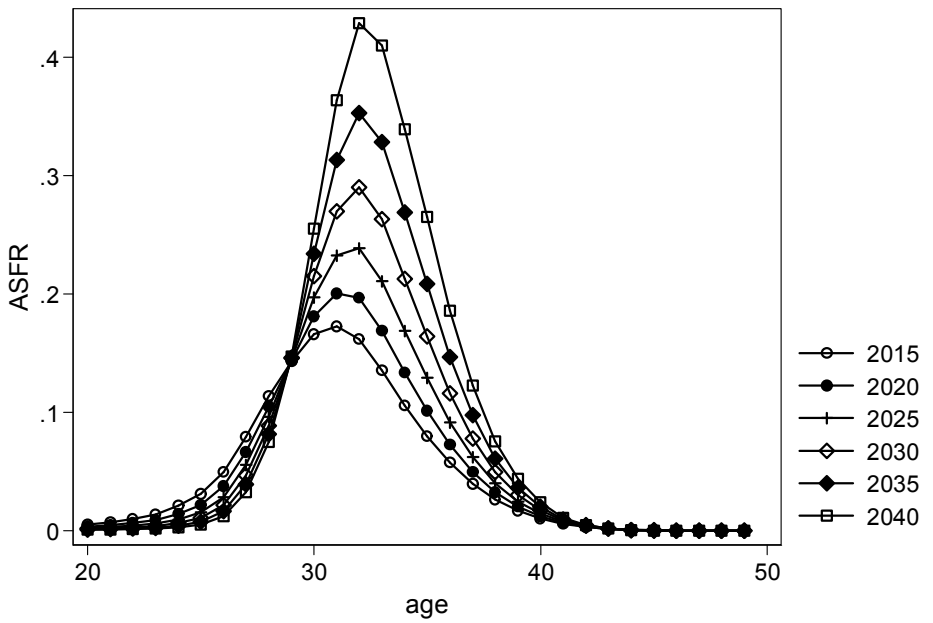


2. 연령별 출산율 및 합계출산율 추계결과

1. 연령별 출산율 추계결과

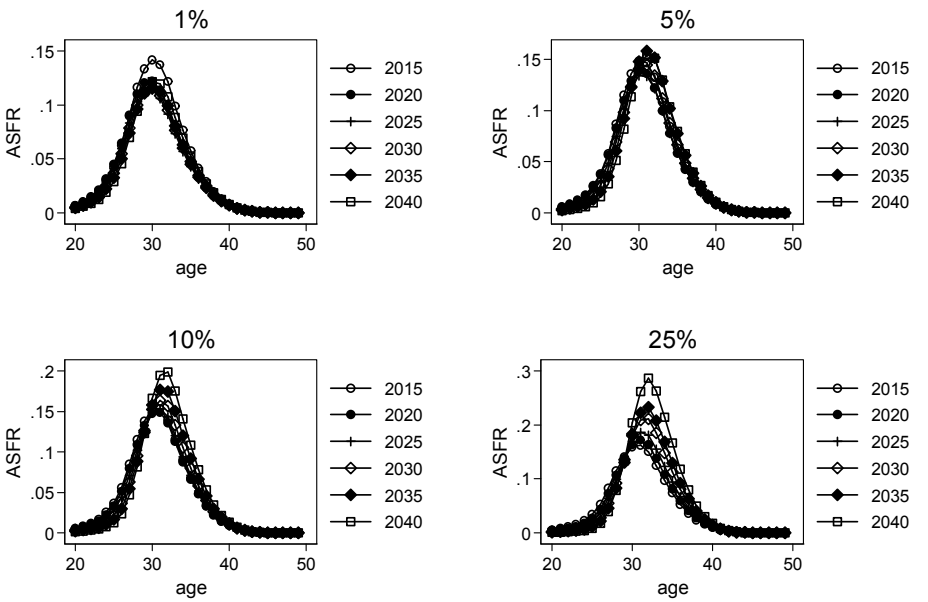
장래 연령별 출산율은 <표 3-1>-<표 3-3>에 제시한 연령별 유배우율, 연령별 유배우 출산율, 연령별 혼외출산 보정요인의 a와 b값과 앞서 논의한 k값에 대한 다양한 추정치를 결합하여 추정했다. 2015-2040년의 연령별 출산율 추정치는 5년 단위로 <부표 3-1>-<부표 3-7>에 제시했다. 우선, <그림 3-24>는 유배우 출산율 상승 방법을 통해서 추정한 연령별 출산율과 합계출산율 변화 추이를 보여준다. 유배우 출산율 상승 방법에 따르면 연령별 출산율은 지속적으로 증가하는 것으로 나타나는데, 특히 30대 초반의 경우에는 이 증가속도가 매우 빠르다. 이러한 변화는 앞서 논의한 연령별 유배우 출산율의 상승 때문인데, 특히 연령별 출산율의 b값이 이 연령대에서 큰 양수 값을 갖는 것이 원인이라고 할 수 있다.

〈그림 3-24〉 연령별 출산율 추계치 (연령별 출산율)



<그림 3-25>는 random walk 방식을 활용한 연령별 출산율 추계치를 보여준다. 연령별 유배우율, 연령별 유배우 출산율, 연령별 혼외출산 보정요인의 변화를 가장 작게 가정하는 1% 시나리오의 경우에는 연도별로 연령별 출산율의 변화 폭이 크지 않은 것으로 나타나고 있는 반면에, 각각의 요소가 random walk 시뮬레이션의 중위값을 따라 변화한다고 가정하는 경우에는 그 변화 폭이 매우 크다. 중위값 시나리오의 추계결과는 <그림 3-24>에 제시한 유배우 출산율 상승 방법과 비슷하기 때문에 제시하지 않는다. 유배우 출산율 상승 방법과 random walk의 중위값 시나리오는 출산율의 상승, 특히 30대 초반의 출산율 상승을 매우 높게 추정하고 있는데, 결과적으로 30대 초반의 연령별 출산율은 2040년에 이르면 0.4보다 큰 값을 갖는다. 이는 30대 초반 여성의 40%가 해당 연도에 출산한다는 것인데, 이는 비현실적으로 높은 값이라고 할 수 있다. 따라서, 연평균 변화율을 보다 낮게 추정하는 방식이 보다 현실성있는 추계치라고 할 수 있다.

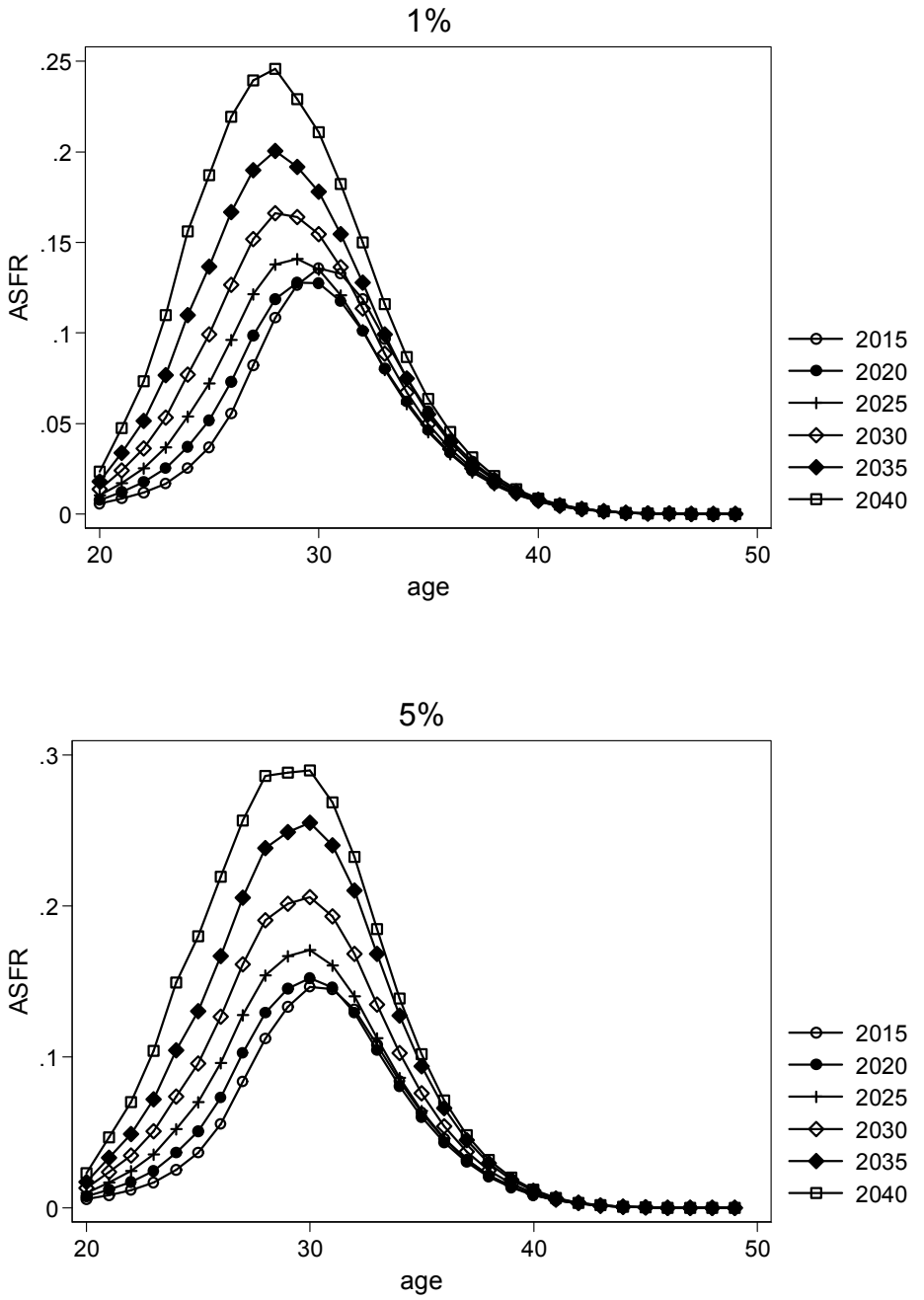
〈그림 3-25〉 연령별 출산율 추계치 (Random walk)



마지막으로 <그림 3-26>은 유배우율 반등 시나리오에 따른 결과를 보여준다. <그림 3-26>에 제시한 결과는 유배우율은 연간 0.7 상승하고, 연령별 유배우 출산율와 연령별 혼외보정 출산요인의 random walk의 1% 및 5%에 따라 변화한다는 가정 하에 산출한 수치들이다. 두 시나리오 모두 출산율의 상승을 예측한다는 점에서 <그림 3-25>에 제시한 결과와 공통점을 갖는다. 단, random walk 방식에서는 유배우 출산율의 상승, 특히 30대 유배우 출산율의 상승이 출산율 반등의 핵심적인 요인인 반면, 이 방식에서는 유배우율 상승이 출산율 반등의 핵심적인 요인이다. <그림 3-26>에 제시한 결과는 유배우 출산율과 혼외출산 보정요인의 변화율이 매우 1%, 5%로 매우 작다고 가정되기 때문이다.

2. 합계 출산율 추계결과

〈그림 3-26〉 연령별 출산율 추계치 (유배우율 반등)



<표 3-7>은 각각의 시나리오를 통해서 추정한 합계출산율의 변화 추이를 보여준다. 기존에 관찰된 추세 변화가 지속된다고 가정하는 유배우율 상승 시나리오에 따른 추계 방법의 경우에는 합계출산율이 점차 증가하여 2040년에는 2.8 정도 수준에 달할 것으로 추정하고 있다. 반면, random walk의 1%는 출산율이 지속적으로 감소하여 2040년에는 1.08 수준으로 하락할 것으로 추정하고 있으며, 5% 시나리오는 출산율이 1.2-1.3 수준에서 등락하는 것으로 나타나며, random walk 10% 시나리오의 경우에는 합계출산율이 1.5 수준으로 상승할 것으로 추정하고 있다. 한편, random walk 25% 시나리오에 따르면, 합계출산율이 점차 상승하여 2040년에는 대체수준에 근접할 것으로 추정하고 있다. 유배우율 k값이 연간 0.7 상승하는 것으로 가정하는 유배우율 반등 시나리오에 따르면, 합계출산율이 대체 수준을 상회하는 2.5까지 상승할 것으로 추정하고 있다.

표 3-7 합계출산율 추이

	추계방식					
Year	유배우율 상승	RW 1%	RW 5%	RW 10%	RW 25%	반등 1%
2015	1.468	1.290	1.335	1.366	1.415	1.224
2016	1.491	1.281	1.293	1.325	1.403	1.214
2017	1.516	1.224	1.278	1.325	1.415	1.188
2018	1.543	1.208	1.259	1.323	1.422	1.192
2019	1.573	1.148	1.262	1.314	1.421	1.181
2020	1.605	1.143	1.250	1.312	1.436	1.215
2021	1.639	1.148	1.251	1.308	1.449	1.256
2022	1.676	1.130	1.257	1.302	1.458	1.302
2023	1.715	1.099	1.240	1.297	1.467	1.306
2024	1.756	1.085	1.213	1.291	1.484	1.339
2025	1.801	1.066	1.198	1.308	1.500	1.364
2026	1.847	1.043	1.209	1.330	1.527	1.411
2027	1.897	1.012	1.234	1.360	1.543	1.427
2028	1.949	1.049	1.253	1.328	1.562	1.501
2029	2.004	1.022	1.230	1.364	1.622	1.541
2030	2.061	1.059	1.242	1.343	1.629	1.638
2031	2.122	1.072	1.217	1.314	1.641	1.690
2032	2.186	1.099	1.221	1.332	1.655	1.803
2033	2.253	1.054	1.250	1.378	1.684	1.830
2034	2.324	1.032	1.273	1.376	1.689	1.871
2035	2.397	1.055	1.303	1.406	1.712	1.990
2036	2.475	1.033	1.302	1.417	1.772	2.047
2037	2.556	1.086	1.262	1.425	1.793	2.157
2038	2.640	1.026	1.256	1.476	1.883	2.217
2039	2.729	1.040	1.307	1.505	1.915	2.324
2040	2.822	1.084	1.252	1.513	2.005	2.480

RW: random walk

반등: 유배우율 반등(0.7) + 유배우 출산율 및 혼외출산 보정요인 RW 1%

제8절 결론 및 논의

제7절에서 제시한 출산율 추계결과는 근접요인 접근법과 Lee-Carter 모형을 활용하여 출산율에 영향을 미치는 유배우율, 유배우 출산율, 혼외출산 보정요인의 변화추세를 추정하고, 각각의 요소의 향후 변화 추이에 대한 가정을 결합하여 생산한 것이다. 현 단계에서 어떠한 추계치가 보다 적합성을 갖는지를 판단하는 것은 쉽지 않은 작업이다. 그럼에도 불구하고, 몇 가지 사항을 고려하면 제시된 결과들의 적합성 혹은 개연성에 대해서 평가할 수 있다. 첫째, 연령별 출산율 추계치의 적절성을 생각해 볼 수 있다. 유배우 출산율 상승 방법과 random walk의 중위값을 취하면, 30대 초반의 연령별 출산율이 1명당 0.4 이상의 값을 갖는다. 해당 연령대 여성의 40%가 동일한 해에 출산을 한다는 가정은 매우 비현실적이기 때문에, 추세연장법이나 random walk의 중위값은 적절하지 않다고 할 수 있다. 둘째, 출산율의 반등 가능성이다. 앞서 검토했듯이 우리보다 출산율 변화를 먼저 경험한 유럽 국가들에서는 2000년대 이후 출산율의 반등 경향이 분명히 관찰된다(Myrskylä et al. 2009). 이는 한국사회에서도 이러한 반등이 분명히 가능할 수 있음을 시사한다. 물론, 출산율 변화 패턴에 특정한 레짐이 존재한다는 주장(Rindfuss et al. 2016) 또한 존재하는데, 이에 따르면 초저출산의 덫에 걸린 한국사회에서 출산율 반등은 어려운 일이라고 할 수 있다. 이러한 두 가지 상반되는 이론적, 경험적 근거를 고려하면, random walk의 1%, 5%, 10%와 같이 최근의 추세가 안정적으로 지속될 것이라는 시나리오나 유배우율 반등에 따라 출산율이 상승할 수 있다는 시나리오 중 어떤 것이 더 적합한 것인지를 판단하기는 쉽지 않다. Random walk 10% 시나리오에 따르면, 2040년에 합계출산율이 1.5 수준으로 추정되는데, 이러한 시나리오가 그간의 출산율 변화 추이를 고려할 때, 합리적인 예측치라고 할 수 있을 것이다. 또한, 이 결과는 2010년의 통계청의 인구추계의 출산율 추계와도 어느 정도 비슷하다고 할 수 있다. 그럼에도 불구하고, 이 연구의 분석결과에 근거해서 합계출산율이 현재 1.2 수준에서 1.5 수준으로 반등할 것으로 예측하는 것은 다소 성급한 결론이라고 할 수 있다. 셋째, random walk의 25% 시나리오와 유배우율 반등 시나리오는 출산율의 반등을 예측하는데, random walk와 유배우율 반등 시나리오 가운데 어떤 것이 더 개연성이 있는지를 평가할 수 있다. random walk 방식에서 나타나는 출산율의 반등은 기본적으로 유배우 출산율의 상승에 기반하며, 유배우율 반등 시나리오는 유배우 출산율은 안정적으로 유지되면서 유배우율이 상승하는 것을 가정한다. 유배우율의 변화가 한국의 출산율 변화에 주된 원인이었다는 기

존 연구결과(이철희 2012)와 유배우 출산율의 수준이 그리 낮지는 않다는 이 연구의 분석결과를 종합할 때, 유배우율 반등 시나리오의 적합성이 보다 높을 것으로 예상된다. 즉, 유배우 출산율이 현재보다 더 높은 수준으로 지속적으로 상승할 것으로 가정하기는 어렵다고 할 수 있다. 이는 출산율 반등의 전제 조건이 유배우율의 상승이라는 점을 다시 한 번 시사한다.

이 연구는 기본적으로 출산율이 유배우율과 유배우 출산율의 함수라는 기본적인 사실이 근거해서, 유배우율과 유배우 출산율에 대한 장래 예측치를 각각 생산하고 이를 결합하는 방식으로 장래 연령별 출산율 및 합계출산율에 대한 추정치를 다양한 시나리오에서 검토했다. 이러한 분석을 통해서 과거의 추세변화가 지속된다는 유배우 출산율 상승 시나리오와 random walk 50% 시나리오의 경우에는 장래 출산율에 대해서 지나치게 높은 추정치를 제시하고 있음을 확인할 수 있었다. 이는 유배우 출산율이 과거 상승한 결과로 해석할 수 있는데, 특히 <표 3-2>에서 제시했듯이 30대 초반의 유배우 출산율 상승에 기인한 것으로 해석할 수 있다. 그런데, 이는 상당 부분 유배우율의 하락, 특히 혼인의 지연에 기인하는 것으로 해석할 수 있다. 즉, 혼인의 지연에 따라 30대에 초혼하는 여성이 증가한 결과, 이러한 패턴이 나타나는 것으로 해석할 수 있다. 이는 근접요인 접근 방법의 기본적인 가정인 유배우율과 유배우 출산율의 독립성이 다소 문제가 있다는 것을 시사한다. 현실적으로 둘 간에 부정적인 상관관계가 존재하기 때문에 (<그림 3-18> 참고), 이 연구에서는 random walk 방법에 따른 다양한 시나리오를 검토함으로써 보다 현실성있는 추정치를 제시하려고 했다. 결과적으로 RW 10% 시나리오가 가장 합리적인 추정치를 제시하는 것으로 나타났다. 물론, 이러한 방식은 충분히 만족스럽다고 할 수는 없다. 이러한 결과는 유배우율과 유배우 출산율간에 존재하는 부정적인 상관관계를 보다 정교하게 반영하는 모형의 개발이 필요하다는 것을 시사한다.

참고문헌

이철희. (2012). 한국의 합계출산율 변화요인 분해 : 혼인과 유배우 출산율 변화의 효과. 한국인구학 35(3): 117-144.

Bongaarts, J. (1978). A framework for analyzing the proximate determinants of fertility. Population and Development Review 4(1): 105-132.

Calot, G. (1993). Relationships between cohort and period demographic indicators: Translation revisited. Population (English edition), 5, 183–222.

Esping-Anderson and G. F. Billari. (2015). Re-theorizing family demographics. Population and Development Review 41(1): 1-31.

Foster, A. (1990). Cohort analysis and demographic translation: A comparative study of recent trends in age specific fertility rates from Europe and North America. Population Studies, 44(2), 287–315.

Garcia-Manglano, J., Nollenberger, N. and Sevilla, A. (2015). Gender, time-use, and fertility recovery in industrialized countries. International Encyclopedia of the Social & Behavioral Sciences (2nd edition) 9: 775-780

Kye, B. 2015. Cross-sectional average fertility (CAF): Properties and comparison with other measures. The Journal of Population Association of Korea, 38(3), 113-139.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality: Age distribution, range, and ultimate level," Journal of the American statistical association

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography

Myrskylä, M. , Kohler, H.-P., and Billari. F. C. (2009). Advances in development reverse fertility declines. Nature, 460(6), 741-743.

Park, H., Lee, J. K., and Jo, I. (2013). Changing relationship between education and marriage among Korean women. Korean Journal of Sociology, 47(3), 51-76.

Rindfuss, R., Choe, M. K., and Brauner-Otto, S. R. (2016). The emergence of two distinct fertilty regimes in economically advanced countries. Population Research and Policy Review, 35, 287-304.

Schoen, R. (2004). Timing effects and the interpretation of period fertility. Demography, 41(4), 810–819.

부록 표 3-1 연령별 출산율 (유배우 출산율 상승

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0050	0.0035	0.0024	0.0017	0.0012	0.0008
21	0.0073	0.0050	0.0034	0.0023	0.0016	0.0011
22	0.0099	0.0066	0.0044	0.0029	0.0019	0.0013
23	0.0139	0.0091	0.0060	0.0039	0.0026	0.0017
24	0.0212	0.0142	0.0096	0.0064	0.0043	0.0029
25	0.0313	0.0220	0.0154	0.0108	0.0076	0.0053
26	0.0497	0.0377	0.0285	0.0216	0.0164	0.0124
27	0.0795	0.0666	0.0557	0.0466	0.0390	0.0327
28	0.1136	0.1046	0.0963	0.0887	0.0817	0.0752
29	0.1431	0.1440	0.1448	0.1456	0.1464	0.1473
30	0.1663	0.1812	0.1974	0.2150	0.2342	0.2551
31	0.1727	0.2005	0.2327	0.2700	0.3134	0.3637
32	0.1619	0.1967	0.2390	0.2904	0.3529	0.4288
33	0.1353	0.1689	0.2109	0.2632	0.3286	0.4102
34	0.1059	0.1337	0.1687	0.2129	0.2688	0.3392
35	0.0799	0.1016	0.1291	0.1642	0.2087	0.2653
36	0.0574	0.0726	0.0918	0.1161	0.1468	0.1857
37	0.0396	0.0496	0.0622	0.0780	0.0979	0.1227
38	0.0263	0.0325	0.0401	0.0495	0.0610	0.0753
39	0.0169	0.0205	0.0248	0.0300	0.0363	0.0439
40	0.0105	0.0124	0.0147	0.0173	0.0204	0.0241
41	0.0059	0.0066	0.0075	0.0085	0.0097	0.0109
42	0.0032	0.0035	0.0038	0.0041	0.0045	0.0049
43	0.0016	0.0016	0.0017	0.0017	0.0018	0.0018
44	0.0007	0.0007	0.0007	0.0007	0.0006	0.0006
45	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002
46	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
47	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
48	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.4678	1.6048	1.8005	2.0614	2.3973	2.8222

부록 표 3-2 연령별 출산율 (Random walk, 1%)

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0066	0.0067	0.0060	0.0053	0.0046	0.0040
21	0.0099	0.0102	0.0091	0.0080	0.0070	0.0061
22	0.0140	0.0148	0.0132	0.0116	0.0101	0.0087
23	0.0199	0.0212	0.0190	0.0168	0.0146	0.0126
24	0.0298	0.0316	0.0285	0.0253	0.0222	0.0194
25	0.0424	0.0447	0.0408	0.0369	0.0328	0.0291
26	0.0626	0.0645	0.0596	0.0549	0.0500	0.0457
27	0.0904	0.0894	0.0832	0.0788	0.0739	0.0699
28	0.1166	0.1100	0.1028	0.0996	0.0961	0.0941
29	0.1336	0.1207	0.1132	0.1122	0.1109	0.1122
30	0.1417	0.1220	0.1142	0.1154	0.1166	0.1216
31	0.1369	0.1136	0.1062	0.1088	0.1120	0.1196
32	0.1217	0.0983	0.0918	0.0951	0.0992	0.1077
33	0.0988	0.0787	0.0735	0.0768	0.0806	0.0885
34	0.0766	0.0609	0.0570	0.0597	0.0630	0.0695
35	0.0575	0.0456	0.0428	0.0449	0.0475	0.0525
36	0.0416	0.0334	0.0314	0.0330	0.0349	0.0385
37	0.0291	0.0235	0.0223	0.0233	0.0246	0.0270
38	0.0198	0.0163	0.0155	0.0161	0.0170	0.0185
39	0.0130	0.0109	0.0104	0.0108	0.0114	0.0123
40	0.0083	0.0071	0.0068	0.0070	0.0073	0.0078
41	0.0049	0.0044	0.0042	0.0043	0.0044	0.0046
42	0.0028	0.0026	0.0025	0.0025	0.0026	0.0027
43	0.0015	0.0014	0.0014	0.0014	0.0014	0.0014
44	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
45	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
46	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
47	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
48	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.2904	1.1428	1.0655	1.0589	1.0546	1.0840

부록 표 3-3 연령별 출산율 (Random walk, 5%)

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0060	0.0056	0.0046	0.0039	0.0030	0.0022
21	0.0090	0.0085	0.0069	0.0058	0.0044	0.0032
22	0.0126	0.0119	0.0097	0.0080	0.0060	0.0044
23	0.0178	0.0170	0.0138	0.0114	0.0085	0.0062
24	0.0267	0.0256	0.0211	0.0176	0.0133	0.0099
25	0.0384	0.0371	0.0313	0.0266	0.0207	0.0159
26	0.0581	0.0561	0.0487	0.0430	0.0353	0.0284
27	0.0867	0.0833	0.0749	0.0692	0.0609	0.0516
28	0.1154	0.1100	0.1018	0.0984	0.0923	0.0820
29	0.1363	0.1288	0.1224	0.1234	0.1229	0.1138
30	0.1487	0.1389	0.1346	0.1410	0.1481	0.1415
31	0.1469	0.1360	0.1340	0.1445	0.1585	0.1553
32	0.1328	0.1221	0.1217	0.1342	0.1518	0.1514
33	0.1088	0.0998	0.1002	0.1120	0.1290	0.1302
34	0.0846	0.0777	0.0784	0.0881	0.1022	0.1040
35	0.0636	0.0584	0.0592	0.0667	0.0778	0.0795
36	0.0459	0.0424	0.0430	0.0484	0.0563	0.0577
37	0.0320	0.0296	0.0301	0.0337	0.0390	0.0400
38	0.0216	0.0201	0.0204	0.0227	0.0260	0.0266
39	0.0141	0.0133	0.0134	0.0148	0.0167	0.0171
40	0.0090	0.0084	0.0085	0.0092	0.0103	0.0104
41	0.0052	0.0050	0.0050	0.0053	0.0057	0.0058
42	0.0029	0.0028	0.0028	0.0030	0.0031	0.0031
43	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015
44	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
45	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003
46	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
47	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
48	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.3348	1.2502	1.1981	1.2423	1.3031	1.2516

부록 표 3-4 연령별 출산율 (Random walk, 10%)

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0058	0.0050	0.0041	0.0032	0.0024	0.0018
21	0.0086	0.0075	0.0060	0.0046	0.0035	0.0025
22	0.0119	0.0104	0.0083	0.0063	0.0046	0.0033
23	0.0168	0.0148	0.0118	0.0089	0.0065	0.0046
24	0.0253	0.0224	0.0181	0.0139	0.0104	0.0075
25	0.0367	0.0330	0.0273	0.0215	0.0166	0.0124
26	0.0561	0.0513	0.0440	0.0365	0.0297	0.0236
27	0.0851	0.0792	0.0712	0.0627	0.0546	0.0469
28	0.1152	0.1088	0.1019	0.0950	0.0882	0.0815
29	0.1381	0.1322	0.1287	0.1264	0.1244	0.1229
30	0.1531	0.1477	0.1485	0.1526	0.1581	0.1659
31	0.1531	0.1487	0.1534	0.1635	0.1765	0.1942
32	0.1396	0.1363	0.1432	0.1566	0.1742	0.1985
33	0.1149	0.1126	0.1197	0.1331	0.1507	0.1752
34	0.0895	0.0880	0.0941	0.1053	0.1202	0.1411
35	0.0673	0.0663	0.0712	0.0801	0.0919	0.1085
36	0.0486	0.0480	0.0515	0.0578	0.0662	0.0780
37	0.0337	0.0333	0.0357	0.0400	0.0456	0.0533
38	0.0227	0.0225	0.0240	0.0266	0.0300	0.0348
39	0.0148	0.0146	0.0155	0.0171	0.0191	0.0218
40	0.0093	0.0092	0.0097	0.0105	0.0115	0.0129
41	0.0054	0.0053	0.0055	0.0058	0.0063	0.0068
42	0.0030	0.0030	0.0030	0.0032	0.0033	0.0035
43	0.0015	0.0015	0.0015	0.0015	0.0015	0.0016
44	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
45	0.0004	0.0003	0.0003	0.0003	0.0003	0.0003
46	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
47	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000
48	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.3661	1.3119	1.3080	1.3427	1.4059	1.5129

부록 표 3-5 연령별 출산율 (Random walk, 25%)

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0054	0.0042	0.0032	0.0024	0.0017	0.0012
21	0.0080	0.0061	0.0046	0.0033	0.0023	0.0017
22	0.0110	0.0082	0.0062	0.0044	0.0029	0.0021
23	0.0155	0.0115	0.0087	0.0061	0.0041	0.0028
24	0.0234	0.0178	0.0136	0.0097	0.0066	0.0047
25	0.0342	0.0267	0.0211	0.0156	0.0111	0.0083
26	0.0532	0.0437	0.0363	0.0286	0.0217	0.0173
27	0.0827	0.0723	0.0639	0.0547	0.0452	0.0393
28	0.1148	0.1063	0.0997	0.0924	0.0824	0.0783
29	0.1408	0.1376	0.1363	0.1358	0.1300	0.1338
30	0.1595	0.1632	0.1697	0.1805	0.1834	0.2036
31	0.1624	0.1722	0.1862	0.2086	0.2223	0.2619
32	0.1500	0.1633	0.1816	0.2112	0.2330	0.2866
33	0.1244	0.1376	0.1556	0.1851	0.2085	0.2630
34	0.0971	0.1082	0.1233	0.1480	0.1684	0.2145
35	0.0731	0.0819	0.0938	0.1133	0.1298	0.1663
36	0.0526	0.0589	0.0673	0.0810	0.0926	0.1180
37	0.0364	0.0406	0.0462	0.0552	0.0629	0.0794
38	0.0244	0.0270	0.0304	0.0359	0.0405	0.0503
39	0.0158	0.0173	0.0193	0.0224	0.0250	0.0305
40	0.0099	0.0107	0.0117	0.0134	0.0147	0.0174
41	0.0056	0.0059	0.0064	0.0070	0.0075	0.0085
42	0.0031	0.0032	0.0034	0.0036	0.0038	0.0041
43	0.0016	0.0016	0.0016	0.0016	0.0016	0.0017
44	0.0007	0.0007	0.0007	0.0007	0.0007	0.0006
45	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
46	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
47	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
48	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.4149	1.4361	1.4997	1.6294	1.7119	2.0050

부록 표 3-6 연령별 출산율 (출산율 반등, 1%)

age	2015	2020	2025	2030	2035	2040
15	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
16	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
17	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
18	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
19	0.0048	0.0048	0.0048	0.0048	0.0048	0.0048
20	0.0057	0.0077	0.0103	0.0136	0.0179	0.0235
21	0.0085	0.0121	0.0170	0.0239	0.0337	0.0474
22	0.0119	0.0175	0.0253	0.0361	0.0515	0.0732
23	0.0169	0.0252	0.0369	0.0532	0.0766	0.1099
24	0.0254	0.0373	0.0538	0.0769	0.1098	0.1561
25	0.0367	0.0519	0.0721	0.0993	0.1366	0.1873
26	0.0555	0.0731	0.0962	0.1267	0.1668	0.2196
27	0.0823	0.0986	0.1213	0.1517	0.1900	0.2396
28	0.1083	0.1187	0.1379	0.1661	0.2007	0.2459
29	0.1265	0.1277	0.1408	0.1640	0.1918	0.2292
30	0.1358	0.1275	0.1352	0.1547	0.1779	0.2109
31	0.1325	0.1174	0.1207	0.1362	0.1547	0.1821
32	0.1187	0.1009	0.1015	0.1134	0.1278	0.1499
33	0.0968	0.0804	0.0798	0.0886	0.0992	0.1159
34	0.0753	0.0620	0.0610	0.0672	0.0747	0.0867
35	0.0566	0.0463	0.0453	0.0497	0.0549	0.0635
36	0.0411	0.0338	0.0330	0.0360	0.0395	0.0453
37	0.0288	0.0238	0.0232	0.0252	0.0275	0.0312
38	0.0196	0.0164	0.0161	0.0173	0.0188	0.0211
39	0.0129	0.0110	0.0108	0.0116	0.0124	0.0138
40	0.0083	0.0072	0.0070	0.0074	0.0079	0.0086
41	0.0049	0.0044	0.0043	0.0045	0.0047	0.0051
42	0.0028	0.0026	0.0026	0.0027	0.0028	0.0029
43	0.0015	0.0014	0.0014	0.0014	0.0015	0.0015
44	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
45	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
46	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
47	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
48	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
49	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TFR	1.2235	1.2150	1.3639	1.6375	1.9899	2.4805

제 4 장

사망률 개선의 교대현상을 반영한 장기 사망률 예측모형

김현태

제1절 서 론

인간의 사망률에 대한 정확한 측정은 오랜 역사를 가지고 있으며 기초적인 국가통계 뿐 아니라 보험, 보건, 의료, 조세 및 재정 정책을 수립함에 있어 기초자료로서 사용된다. 전통적인 사망률모형은 특정한 시간이 주어졌을 때 각 연령에 대한 사망률 혹은 사망력을 계산해 내는 데 중점을 두었으나, 현대에 이르러 미래의 인구추계를 위해 각 연령의 시간에 따른 개선(진전 혹은 감소라고도 표현함)을 동시에 모형화 하는데 더 큰 관심을 가지게 되었다. Lee-Carter (1992) 모형과 이를 다중인구로 확장한 Li-Lee (2005) 모형이 이러한 모형들 중 국내외에서 널리 사용되고 있는 대표적인 모형들이다. 그러나, 많은 서구 국가들에서 최근의 사망률개선 속도가 유년층에서는 감소하고 고령층에서는 증가하는, 소위 사망률 개선의 교대(mortality decline rotation)패턴을 보이고 있는데 Lee-Carter 나 Li-Lee 모형은 이러한 교대패턴을 표현할 수 없어 장기 사망률예측에서 한계를 가진다고 알려져 있다. 본 보고서에서는 장기 사망률예측을 보다 안정적으로 추정할 수 있는 새로운 모형들을 소개하고 국내 사망률자료에 적합한 결과에 대해 논의한다.

제2절 기존의 사망률 예측모형들

1. Lee-Carter 계열 모형

Lee-Carter 모형 (이하 LC 모형)은 단일한 지표(k)를 사용하여 시간의 경과에 따른 사망률의 변화를 설명하고 예측한다. LC모형은 t 년도 x세의 중앙사망률(central death rate) m(x,t)를 다음과 같은 형태로 모형화 한다.

$$\ln[m(x,t)] = a(x) + b(x)k(t) + \epsilon(x,t)$$

여기서 a(x) 는 각 연령에서의 기본 사망률 (base mortality)을 나타내고 k(t)는 시간의 경과에 따른 사망률의 지표 (mortality index)로써 시간의 경과에 따라 감소하는 (선형적) 패턴을 보인다. b(x) 는 각 연령에 대해 상수로써 사망률 지표 k(t)에 대한 민감도를 나타낸다. 우변의 마지막 항은 평균 0, 분산 σ_ϵ^2을 가지는 오차항으로 모형에 의해 설명되지 않는 사망률의 랜덤성을 반영한다.

LC모형의 모수들은 방정식의 최소제곱법(least squares method)를 이용해 구하는데, 방정식의 우변에 독립변수가 존재하지 않아 일반적인 최소제곱법으로는 해를 구할 수 없다. 이를 해결하기 위해 LC는 ln(m(x,t))에서 시간 경과에 따른 행평균 열벡터인 a(x)를 제거한 후 남은 행렬에 SVD (singular value decomposition)의 1차원소 혹은 Newton-Rapson 방법을 통해 b(x)와 k(t)를 각각 추정하게 된다. 여기서 $\sum_x b_x = 1$, $\sum_t k_t = 0$의 정규화 제약조건을 설정하여 identifiability 문제를 피한다. 이론상으로는 SVD의 2,3차 요소를 반영할 수도 있으나, 고차항에 대한 설명력이나 모형의 복잡성을 고려하여 많이 사용하지는 않는다.

위와 같은 과정으로 얻은 k(t)에 대해 Lee-Carter 는 적절한 ARIMA 시계열 모형 중 random walk with drift 모형 혹은 ARIMA(0,1,0)

$$k(t) = k(t-1) + e(t)$$

로 적합하여 미래의 사망률을 예측한다.

2. Li-Lee 모형

1992년 JASA 에 게제된 Lee-Carter모형은 이후 많은 일반화와 변형이 이루어졌고 단일인구내에서 사망률의 적합과 예측에 대표적으로 널리 사용되는 모형들의 근간이 되어 왔다. 그러나 주어진 인구를 관심이 있는 어떤 특성을 기준으로 나누어 (예를 들어, 성별 또는 지역) 몇 개의 그룹으로 고려한다면 LC모형은 그 자체로 사용하기 어려운 측면이 있다. 그 이유는 다중인구에 대해 LC모형을 적용하는 경우 각 그룹에 LC모형을 개별적으로 적합하게 되면 그룹간의 k(t) 값들이 서로 달라 내적인 정합성을 잃어버리고 장기예측에서의 평균 사망률의 추세가 확연히 다르게 되기 때문인데, 이같은 현상은 비슷한 사회경제적 조건하에서 개별 그룹의 사망률들이 장기적으로 비슷한 추세이어야 한다는 기본적인 믿음에 반하는 것이다. 이를 극복하기위한 방법으로 Li and Lee (2005)는 LC모형을 다중인구로 확장한 모형을 제시하였다 (이하 LL모형).

우선 주어진 인구가 몇 개의 그룹으로 나누어질 수 있는 다중인구라고 가정하고, 그 중 i번째 (남녀로 나눈다면 i=1,2 이다)그룹의 중앙사망률을 m(x,t,i) 라고 하자. 이 기호는 LC모형과 같지만 그룹의 인덱스가 더해진 것이다. 이때 LL모형은

$$\ln[m(x,t,i)] = a(x,i) + B(x)\,K(t) + b(x,i)k(t,i) + \epsilon(x,t,i)$$

로 주어진다. LC모형과 비교할 때, 가장 큰 차이는 우변의 둘째 항 B(x)K(t)인데, 이는 모든 그룹에 공통적으로 적용되는 사망률의 추세라고 볼 수 있다. 여기서 a(x,i)는 LC모형에서와 같이 각 성별 기본사망률을 행평균으로 구한 후 B(x)K(t)는 최소제곱법으로 추정할 수 있다. 나머지 잔차에 각 성별로 SVD 혹은 Newton-Rapson 방법을 통해 b(x,i)k(t,i)를 추정한다.

3. 일본의 Age-Shifting 모형

현재 일본에서는 사망률의 예측모형으로서 Age-Shifting 모형으로 불리는 모형을 사

용하고 있는데, 이 모형은 다음과 같은 단계를 거친다.

1. 사망력 $\mu(x,t)$를 다음과 같은 3모수 로지스틱함수로 적합한다.

$$\mu(x,t) = \frac{1}{1+a(t)^{-1}\exp(b(t)*x)} + \gamma(t),$$

2. 위에서 추정된 모수들에 기반하여 시간에 의존하는 S(t) 를 계산한다.

$$S(t) = -\frac{\ln a(t)}{b(t)}$$

3. 위의 모수를 이용하여 원래의 연령 x를 새로운 연령 y로 변환하는데 그 관계식은 x=f(y) 이며 변환은 다음과 같은 구간별 선형변환에 따른다. 여기서 t_0는 기준연도로서 2005로 고정되어 있다.

$$f(y) = \begin{cases} y & (y \le 25) \\ \left\{ \frac{\beta_{t_0}}{\beta_t}(50 - S_{t_0}) + S_t - 25 \right\} \frac{y-25}{50-25} + 25 & (25 \le y \le 50) \\ \frac{\beta_{t_0}}{\beta_t}(y - S_{t_0}) + S_t & (50 \le y \le S_{t_0}) \\ y - S_{t_0} + S_t & (y \ge S_{t_0}) \end{cases}$$

4. 이렇게 만들어진 연령과 시간의 새로운 쌍에 대해 Li-Carter 모형을 적합한다.

위에서 보듯 Age-Shifting 모형은 원래의 연령을 적절한 구간에 따라 서로 다른 선형식을 통해 새로운 연령 y를 구해내는 것이 핵심이다. 그러나 이 모형은 25세 이하에서는 변환을 하지 않으며 25세 이상에서도 구간의 결정이 자의적인 측면이 있어 다른 국가의 사망률에 적용하기 위해서는 좀 더 많은 연구가 필요하다고 하겠다. 또한 현재 일본을

제외한 다른 국가에서는 이 모형을 사용하고 있지 않은 것으로 알려져 있어 아직 보편적으로 받아들여지는 사망률 모형으로 보기는 어렵다.

제3절 사망률의 Rotation 문제

Li-Lee 모형을 포함한 LC 계열의 모형들은 주어진 연령 x에 대해 b(x)를 시간에 무관한 상수로 가정하고 있다. 이러한 상수 가정은 단,중기적인 사망률 예측에는 문제가 없으나, 장기적으로 예측하는데 한계가 있다고 알려져 있다. 그 이유는 많은 서구 국가에서 최근의 사망률개선 속도가 유년층에서는 감소하고 고령층에서는 증가하는, 소위 사망률 개선의 교대 (mortality decline rotation) 패턴을 가지고 있어 b(x)을 상수로 보는 LC계열의 모형들은 이 같은 현상을 표현할 수 없기 때문이다.

사망률개선의 교대현상에 관련된 대표적인 연구들은 Li, N, R.Lee and P. Gerland (2013), Andreev, K, D. Gu, and P. Gerland (2013), Statistics Canada (2015) 등이 있다. 본 연구에서는 이러한 선행연구들을 토대로 국내 사망률 자료로부터 b(x)값의 시간에 따른 과거추이와 여기에 기반하여 다음과 같은 접근방법들을 고려한다.

(1) Li, Lee and Gerland 의 모형을 적용하는 방법
(2) 새로운 방법으로서 사망률 개선지표를 감안하되, 로지스틱 비선형 함수로서 b(x,t)를 추정하는 방법과 이를 확장하여 성별 적합성을 유지할 수 있는 coherent 한 사망률 예측모형의 개발

1. Li-Lee-Gerland Model

회전 rotation 문제와 관련된 LC 모형의 주요 이슈는 $b(x)$가 각 연령에서 상수로 주어지기 때문에 시간에 따른 의존성을 잡아낼 수 없다는 것이다. 시간에 따른 사망률의 감소나 증가는 $b(x)k(t)$에 의존하는 형태이기 때문에 완전히 $k(t)$에 의해 조절된다. 그러므로 만약 사망률 감소 민감도 $b(x)$가 시간에 따라 변할 수 있게 하려면 $b(x)$에 시간변수를 더해 $b(x,t)$ 형태로 확장하는 것이 자연스러울 것이다. 이러한 아이디어를 LC모형에 적용한다면 다음과 같이 나타낼 수 있다.

$$\ln(m(x,t)) = a(x) + b(x,t)k(t) + \epsilon(x,t).$$

여기서 다양한 $b(x,t)$가 가능하겠으나 수많은 후보들 중에서 이해하기 쉽고 사용하기에 비교적 간단한 모수적인 함수를 선택해야 한다. 최근 Li, Lee and Gerland (2013)은 (1)에서의 $b(x,t)$를 특정한 형태의 함수로 주어 장기적 사망률에서의 회전 문제의 해결 방법으로 제안한다. 더 구체적으로 이 모형은 (LLG라고 하겠다) 다음과 같은 비선형 함수 $b(x,t)$를 가정한다.

$$b(x,t) = (1 - w_s(t))b(x) + w_s b_u(x),$$

여기서 우변은 본래의 LC모형의 $b(x)$와 사망률 감소율의 최종적 연령별 패턴 $b_u(x)$의 가중 평균으로 표현된다. 가중치 $w_s(t)$는 시간에 대한 함수이며 시간이 지남에 따라 1에 가까워지고 따라서 이 식 전체인 $b(x,t)$도 점점 $b_u(x)$에 가까워진다. $b_u(x)$의 값은 $0 \le x \le 65$에서 동일하다고 가정하며 그 이후 연령에 대해서는 연령에 따라 감소한다고 가정한다. 이는 20 국가 인구의 과거 사망률의 경험적 분석에 근거하여 65세까지 모든 연령대에서 사망률 감소 속도가 각 인구마다 정도는 다르지만 궁극적으로 하나의 일반적인 수준으로 수렴한다는 것을 의미한다. 최종적 감소율을 다음과 같이 제안한다. 먼저, $0 \le x \le 65$에서 중간 값 $\tilde{b}_u(x)$는 중간 연령에서의 $b(x)$평균값의 상수로 설정한다.

$$\tilde{b}_u(x) = \frac{1}{50}\sum_{x=15}^{65} b(x), \quad 0 \le x \le 65,$$

$x \ge 70$의 연령에서는 $b(x)$을 비례적으로 조정하여 $\tilde{b}_u(x)$를 계산한다. 구체적으로, 비례상수는 $\tilde{b}_u(70) = \tilde{b}_u(65)$를 만족하도록 계산이 되며, 이를 공식으로 표현하면 $\tilde{b}_u(x) = b(x) \times \tilde{b}_u(65)/b(70)$과 같이 계산된다. 따라서, 모든 연령에서의 최종적 감소율은 다음과 같이 얻을 수 있다.

$$b_u(x) = \frac{\tilde{b}_u(x)}{\sum_{x=0}^{\omega} \tilde{b}_u(x)},\ 0 \le x \le \omega,$$

LC모형 형식에서 요구되는 대로, 이는 더해서 1이 된다.

다음으로 가중 함수 $w_s(t)$의 계산은 두 단계를 거친다. 먼저, t시점에서 0세의 기대 수명 $e_0(t)$에 근거한 선형 가중치 함수가 다음과 같이 제안된다.

$$w(t) = \frac{e_0(t) - 80}{e_0^u - 80},$$

여기서 e_0^u는 최종 기대 수명이고 80은 그들이 고려한 20 국가 인구 기대 수명의 대략적인 중앙값이다. 그러므로 rotation 은 $e_0(t) = 80$일 때 $w(t)$=0 에서 시작하며, $e_0(t) = e_0^u$일 때 1까지 변하게 된다. 저자들은 이 값을 102로 설정하였다. 두 번째 단계로서, 이렇게 정의된 가중함수에 기반하여 좀 더 매끄러운 가중치 함수를 다음과 같이 $w(t)$의 삼각 함수로 정의한다.

$$w_s(t) = \left\{0.5[1 + \sin(\frac{\pi}{2}(2w(t) - 1))]\right\}^p,$$

여기서 p는 회전의 속도를 통제하는 모수이고 0.5로 고정한다.

그러나 미래 사망률을 예측하는 데 고려해야할 한 가지 요소가 더 존재한다. 모형(1)에서 사망률 지표 $k(t)$가 수정되어야 한다. 그렇지 않으면 새로운 $b(x,t)$값이 $k(t)$에 곱해지게 되고 이는 기대 수명과 같은 미래 사망률의 특성을 왜곡시킨다. LLG 모형은 각 미래 시점에서 LC모형과 LLG모형의 기대 수명이 같도록 $k(t)$를 바꾸어 이러한 문제를 해결한다. 이를 설명하기 위해 LC 모형으로부터 추정된 미래 t시점의 기대 수명

$e_0(t)$ 값과 LLG모형으로부터 추정된 t시점의 기대 수명을 같게 하려한다고 해 보자. LLG모형이 각 t시점에서 추정된 $b(x,t)$와 $a(x)$을 가지고 있기 때문에, 결국 LC모형의 $k(t)$가 아닌 새로운 사망률 지표 $k^*(t)$를 찾아서 LC모형의 기대 수명과 같도록 하는 것이 유일한 방법이다. 그러므로 $k^*(t)$는 LC모형의 $k(t)$와 다른 값을 갖는다. 이는 LLG모형이 사망률 지표 $k^*(t)$에 의해 도출되고 시간에 따라 변하는 변수 $b(x,t)$가 $k(t)$가 아닌 $k^*(t)$에 의해 설명된다는 중요한 결과를 갖는다.

LLG모형은 현재 $b(x)$와 최종적 $b_u(x)$의 가중 평균치를 가지며 시간에 따라 가중치가 점점 증가한다는 점에서 합리적인 반면, 사용자에 의해 지정되어야 하는 수동 입력 값이 필요하고 이는 실제로 정확하게 찾아내는 것이 항상 간단하지는 않다. $k(t)$를 구하기 위해 여전히 SVD를 수행해야 하지만 이는 오직 새로운 $k^*(t)$로 변환하기 위한 것뿐이다. 뒤에서 $k(t)$를 $k^*(t)$로 변환하는 것은 회전 초기에 로그 사망률의 갑작스러운 변화를 초래하는 것을 볼 수 있다. 현재로서 LLG모형의 가장 큰 한계점이라고 한다면 다중인구모형으로의 확장이 쉽지 않다는 점일 것이다. 물론 그런 확장이 불가능한 것은 아니겠으나 현재 모형의 구조하에서는 사망률 지표 $k^*(t)$가 기대수명을 유지하는 제약식에 의해 도출되는 값이기 때문에 기대수명이 서로 다른 다중인구 (혹은 그룹)에서는 하나의 사망률 지표 $k^*(t)$가 아니라 그룹의 개수만큼 구해진다. 따라서 통일된 공통의 사망률지표가 필요한 경우 어떻게 처리할지가 불분명하다. 또한 비슷한 이유로 rotation을 다중인구하에서 어떻게 내적인 정합성을 유지하면서 처리할 것인지도 어려운 문제이다. 가능한 하나의 방법은 남녀를 합친 자료에 LLG를 적용한 후 각각의 성별로 다시 Li-Lee 모형과 같은 모형을 구현하는 것이다. 그러나 이 경우 성별로 별도의 튜닝과정이 필요하고 실제 어떻게 작동할지는 지금으로서는 알기 어렵다. 향후 LLG모형가 다중모형으로 확장되고 검증이 된 다음 다시 연구해 볼 사항이다.

2. 새로운 모형: 로지스틱 사망률 예측모형

시간에 따라 변화하는 사망률 감소 속도를 모형화하는 방법으로서 이 절에서는 LC모형과 비슷하지만 다소 다른 요소들을 가지는 새로운 모형을 제안한다. 이 모형은 다음과 같이 정의된다.

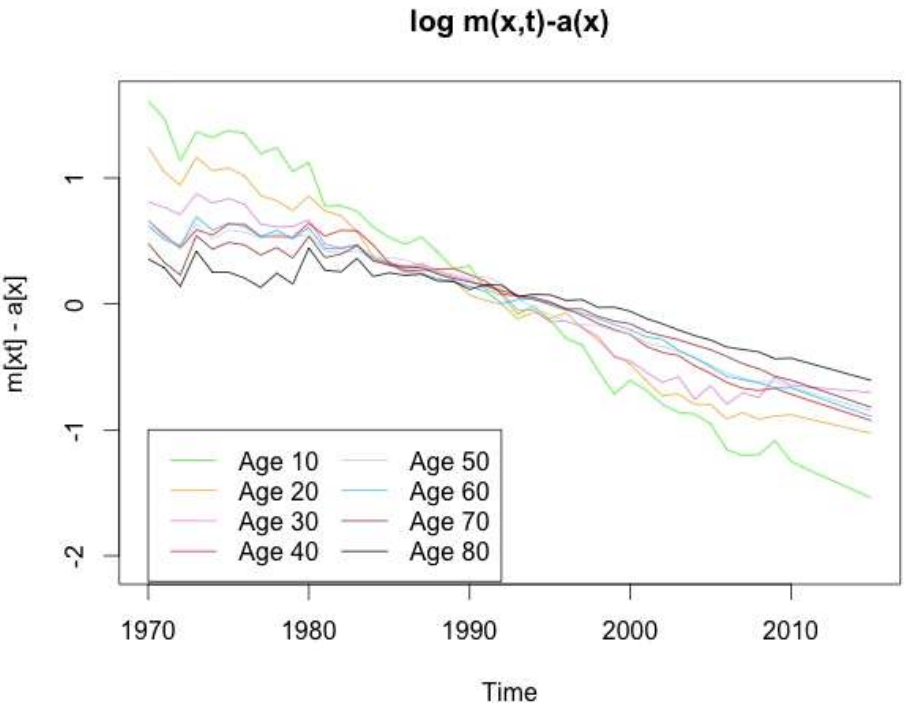
$$\ln m(x,t) = a(x) + \theta(t) + g(x,t) + \epsilon(x,t),$$

여기서 $a(x)$는 LC모형의 기본 사망률과 같지만 나머지 부분은 다르다. $\theta(t)$는 전체적인 사망률 지표를 의미하고 LC모형의 $k(t)$와 같은 역할을 하지만 의미나 계산은 동일하지 않다. LC 모형에서는 최적화를 통해 $b(x)$와 $k(t)$를 동시에 추정하며, 공통개선지표 $k(t)$에 대해 서로 다른 연령별 민감도 $b(x)$가 곱해져 사망률을 결정한다. 따라서 $b(x)$가 달라지면 $k(t)$ 값에 영향을 미치고 그 반대도 마찬가지이다. 그러나 위의 모형에서는 공통사망률지표 $\theta(t)$는 모든 연령에 대해 같은 민감도 1을 암묵적으로 가정한다는 점에서 서로 다르다. 다시 말해, 이 모형에서 $\theta(t)$는 모든 연령에 대해서 t시점의 평균 사망률 증가 지표이지만 모든 연령에서 동일한 강도로 적용된다. 그러나 몇몇 연령에서 사망률은 평균보다 빠르게 혹은 느리게 증가하고, 각 연령에서 개별적인 사망률의 차등을 두는 것이 당연하다. 이는 $g(x,t)$에 의해 조절되는데, 이 함수는 연령과 시간의 함수이고 비선형적 형태로 주어짐을 짐작할 수 있다.

제안된 방법이 실현가능한지 설명하기 위해서 <그림 4-1>에서 남녀를 합친 국내 사망률 자료의 선택된 연령에 대해 시간에 따른 $\ln m(x,t) - a(x)$의 값을 보였다. 앞서 언급한 대로, LC 방법은 그림에서 각각 다른 기울기를 갖지만 공통적인 사망률 향상 지표 $k(t)$에 의해 조절되는 선형적 형태를 모형화 한다는 것을 언급했다. 실제로 그림에서 모든 연령에 대한 곡선들은 시간이 흐를수록 하강하고 사망률은 모든 연령에 대해 각각 다른 속도로 줄어든다는 것을 알 수 있다. 그러나 육안으로 보기에 완벽하게 선형적 감소는 아니라는 것을 또한 알 수 있다. 몇몇 연령에서는 지난 수십 년간 비선형적인 감소가 발생하였다. 예를 들어 10세는 가장 빠른 향상을 보이며 가파른 기울기의 선형적 함수를 나타내는 반면에 고연령에서는 기울기가 점점 가팔라지는 비선형적 감소를 보인다.

이를 더 분명히 알 수 있는 방법은 <그림 4-1>의 곡선들로부터 시간에 따르는 공통 사망률지표를 제거하는 것이다. 만약 LC모형이 유지될 때, 제거 후에 모든 곡선이 다시 (기울기는 다르겠으나) 선형이 되는 결과를 예상할 수 있다. 그러나 이 기울기들은 평균 속도에 대한 사망률 향상 속도의 정도에 따라서 다른 부호를 가진다. 공통 사망률 시간 지표를 $\theta(t)$로 표기하고 다음과 같이 비모수적으로 정의한다.

〈그림 4-1〉 국내 로그사망률 값, $\ln m(x,t)-a(x)$

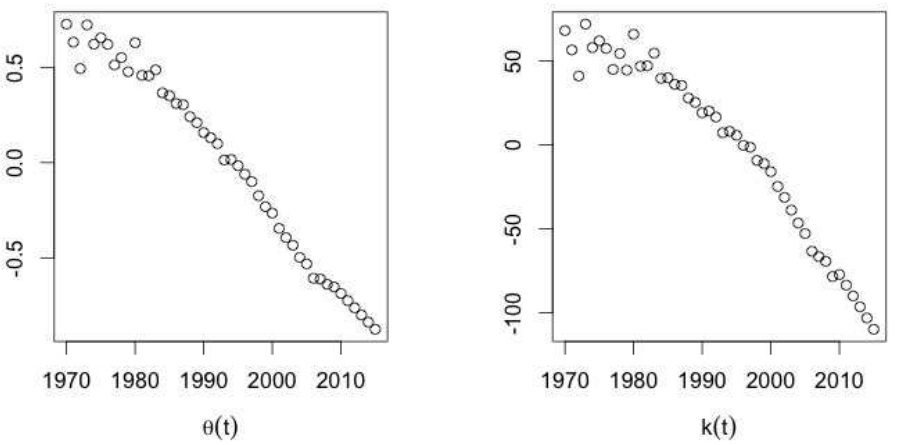


$$\theta(t)=\frac{\sum_{x=x_1}^{x_m}(\ln m(x,t)-a(x))}{x_m-x_1+1},\ t=t_1,\ldots,t_n$$

이는 기본적으로 특정 연령의 기본 사망률 $a(x)$이 설명된 후 t시점에서의 평균 사망률이다. $\theta(t)$의 추세는 $k(t)$와 유사하게 시간이 흐를수록 사망률이 얼마나 향상되는지를 보여주지만 $\theta(t)$에 부속적인 계수를 붙이지 않는다. 이는 LC모형에서 $k(t)$와 $b(x)$가 곱으로 연결된 것과 대비되며, 또한 $k(t)$와 달리 $\theta(t)$에 대한 식별성 문제를 갖지 않는다. <그림 4-2>는 국내 사망률 데이터 (남녀 공통)에 대해 $\theta(t)$와 $k(t)$를 나란히 비교하고 있다. 이 둘이 단위를 제외하고는 큰 차이를 보이지 않는다는 것을 알 수

있다.

〈그림 4-2〉 로지스틱모형(좌)과 LC모형(우)의 공통사망률지표의 시간에 따른 값



이를 이용하여 <그림 4-3>에서는 위의 식을 이용하여 국내 사망률자료에 $\ln m(x,t)-a(x)-\theta(t)$를 계산한 결과를 보여준다. 이는 로그사망률에 연령에 대한 평균적인 값과 시간에 대한 평균적인 값을 제외한 잔차라고도 볼 수 있다. 이 그림으로부터 우리는 다음의 몇 가지를 관측할 수 있다.

- 각 연령의 잔차 곡선은 예상한 대로 상승과 하강의 추세를 모두 보이는 것을 알 수 있다. 저연령의 경우 아래로 하강하는 추세를 보이는데 이는 저연령 사망률의 개선(혹은 감소)이 모든 연령의 평균 $\theta(t)$ 보다 더 빠름을 뜻하며, 따라서 전체적인 사망
- 률 향상이 가속화된다. 반대로 고연령의 상승추세는 사망률 개선이 평균보다 늦음을 반영하고 있다.
- 그럼에도 불구하고 잔차들의 증가 감소는 항상 선형적이지는 않다. 특히 10세 80세와 같이 매우 낮거나 높은 연령을 제외하고 대부분의 중간 연령 곡선은 1980년부터 평평해지며 점점 수평이 된다. 수평이 되는 것은 몇몇의 연령에서 점진적으로 일어나며 다른 연령대에서는 급작스럽게 일어난다. 이 선형성의 위반은 사망률 개선의 교대현상과 연결될 수 있다.
- 80세와 같은 초고연령의 경우 잔차의 패턴은 직선이 아님이 분명해 LC계열의 모형

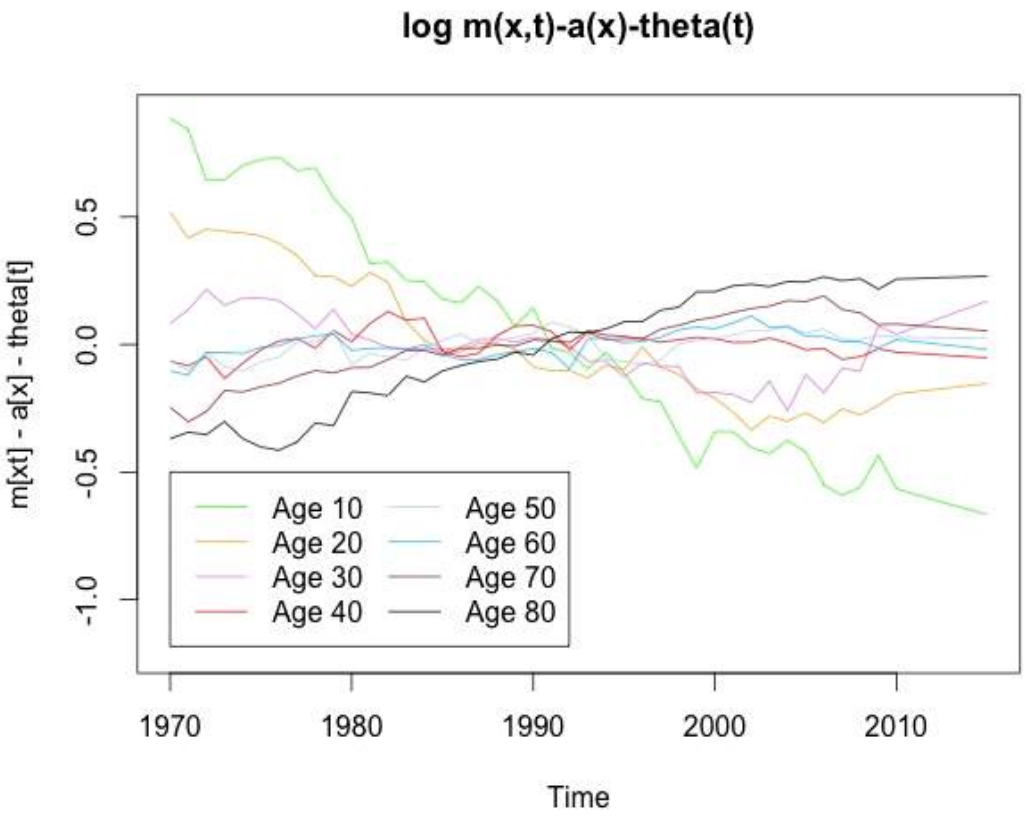
의 적절하지 않음을 알 수 있다. 특히, 2000년이후 잔차의 증가가 현저히 둔화되는데, 이는 $\theta(t)$를 늦추는 힘이 점점 없어진다는 의미로, 결국 사망률 개선속도가 점점 빨라짐을 뜻한다. 반대로, 저연령(예를 들어 20, 30세)에서는 잔차의 감소가 현저히 둔화되는데, 이는 $\theta(t)$에 더해졌던 가속이 더 이상 없음을 뜻해서 사망률의 개선속도가 감소함을 의미한다. 이는 국내에서도 rotation 이 일어나고 있음을 시사한다.

따라서 위의 논의에 입각해 <그림 4-3>의 잔차를 $g(x,t)$라는 함수로서 모형화 한다고 가정한다면, 전체적으로 상승과 하강을 허용하면서도 시간에 따라 점진적으로 수평이 되는 로지스틱함수를 사용하는 것이 자연스러워 보인다. 더 구체적으로 우리는 다음과 같은 4모수 로지스틱 혹은 S자 곡선을 고려한다.

$$g(x,t) = \frac{v(x)}{1+\exp(-(t-\mu(x)/s(x))} + c(x),$$

여기서 점근선 $v(x)$, 스케일 $s(x)$, 대칭점 $\mu(x)$와 $c(x)$는 x가 주어질 때 상수이다. 여기서 $s(x)$는 스케일 모수라고 부르지만 로지스틱 함수가 분포 함수가 아니라 수학적 함수이기 때문에 음수가 될 수 있다. 로지스틱 함수의 표준적 형태는 $c(x)$를 포함하지 않지만 여기서는 반드시 필요하다; 이 부분이 없다면 로지스틱함수의 최솟값은 항상 0일 것이다. 따라서 위의 수식에 의하면 로지스틱 함수는 점근적으로 $|v(x)|+c(x)$의 최대 한계 값으로 수렴한다. 이는 <그림 4-3>에서 대부분의 곡선에서 볼 수 있는 것과 일치한다. 또한 중간 부분에서 대략적으로 선형이며 이는 육안으로 점검한 것으로부터 요구된 것과 같다. 마지막으로 이는 $v(x)s(x)$의 곱의 부호에 따라 감소하거나 증가하고 낮은 연령과 높은 연령 둘 다 통합된 형태로 모형화할 수 있다. 10세와 같은 매우 낮은 연령대에서 $g(x,t)$가 상수로 수렴하는 것이 뚜렷하게 나타나지 않지만 큰 절대 값의 스케일 모수 $s(x)$를 갖는 로지스틱을 적용하면 단기 예측에서는 선형적인 패턴을 유지하지만 최종적으로는 수렴한다.

〈그림 4-3〉 국내 로그사망률의 값, $\ln m(x,t)-a(x)-\theta(t)$



로지스틱 함수의 모수추정은 비선형 최소제곱법을 사용한다. 이를 위해 통계프로그램인 R에서 SSlogis 함수를 이용하였는데, 이 함수 또한 때때로 수치적 불안정이라는 어려움을 가지고 있고, 또 이 함수는 $v(x)$, $s(x)$, $\mu(x)$의 세 모수 로지스틱 함수에만 적용될 수 있기 때문에 주의를 요한다. 이를 해결하기 위해 $c(x)$에 대해 가능한 범위 (grid)를 만들고 범위 안에서 각각의 $c(x)$값에 대해 SSlogis를 사용하여 나머지 세 모수를 추정하였다. 이 과정을 범위 안의 모든 $c(x)$값에 대해 적용한 후 최소 오차 제곱 합을 가지는 모수값을 선택하였다.

마지막으로, 그림에는 보이지 않지만 80대 중반 이후 의 초고령자에 대해서는 $\ln m(x,t)-a(x)-\theta(t)$에 로지스틱함수보다 직선식의 잔차제곱합이 더 작은 패턴을 보이는데 이는 초고령에서는 사망률개선이 약화됨을 의미한다. 부연설명을 하자면, 사망률개선 속도가 고령층에서는 증가하는 사망률 개선의 교대 패턴이 80세이후의 초고령

에서는 보이지 않는다는 점이다. 이는 LLG 모형에서도 초고령인구에 대해서는 교대패턴을 적용하지 않는다는 점과 같은 맥락이다. 이를 해결하기 위해 logistic 사망률모형에서 다음의 알고리즘을 이용한다. 먼저 $\ln m(x,t)-a(x)-\theta(t)$에 직선과 로지스틱함수 둘 다 적합한 후 전자의 잔차제곱합이 더 작게 되는 임계연령 u를 찾는다. 국내자료의 경우 84세부터 이런 현상을 보인다. 그러면 [u, w] 의 연령구간에서는 g(x,t)를 직선과 로지스틱함수의 다음과 같은 가중평균으로 계산한다.

$$g(x,t)=\frac{x-u}{1+w-u}f_{Lin}(x,t)+\left(1-\frac{x-u}{1+w-u}\right)f_{Logistic}(x,t)$$

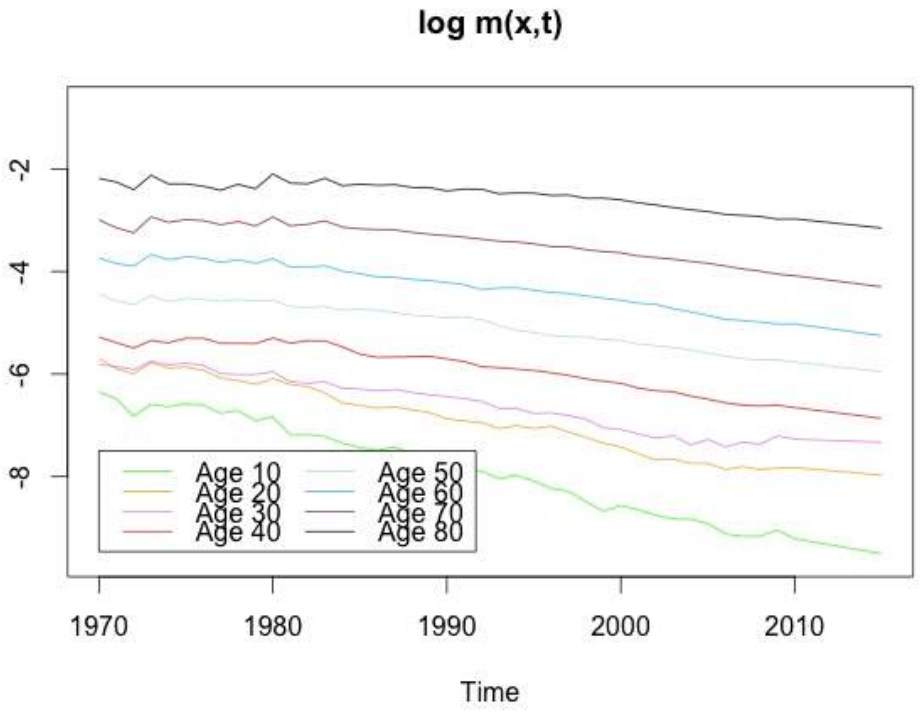
이 공식에 의하면 최고가능 연령 w에 접근할수록 선형함수의 가중치가 늘어나 초고령의 사망률은 교대패턴이 없는 형태가 된다.

제4절 국내 사망률의 추정과 예측

이 절에서는 국내 사망률 데이터를 사용하여 지금까지 소개된 주요모형들을 적합한다. 여기서는 남녀를 합친 데이터를 살펴볼 것이다. 주어진 자료는 0-100세까지의 연령을 1970년부터 2015년간 45년의 걸친 로그사망률이다.

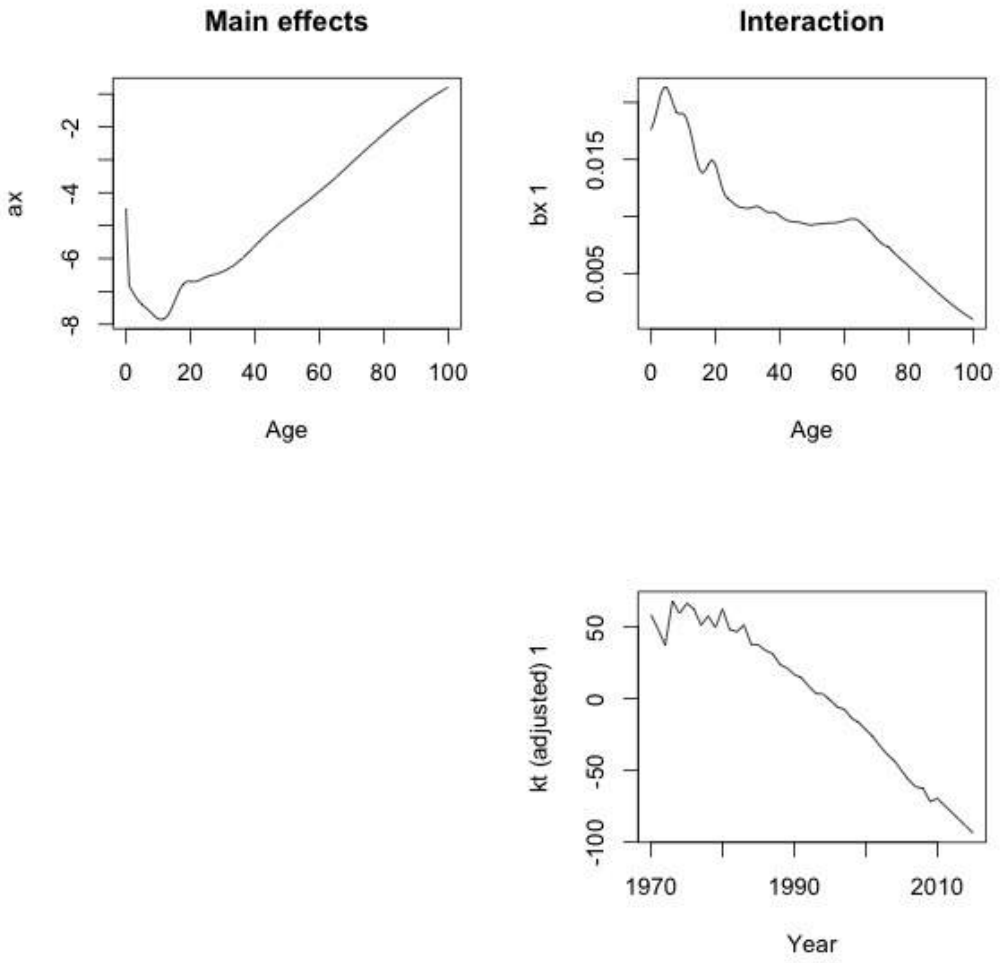
주요연령에 대한 로그사망률의 시간별 값을 플롯으로 나타내면 <그림 4-4>와 같으며 그림으로부터 연령이 높을수록 로그사망률의 수준이 높으며, 모든 연령에 대해 시간이 흐를수록 사망률이 우하향의 패턴을 보여 개선되고 있음을 알 수 있다. 그러나 모든 연령의 사망률 개선속도 (혹은 기울기)가 같지는 않음을 알 수 있다.

〈그림 4-4〉 국내 로그사망률의 값



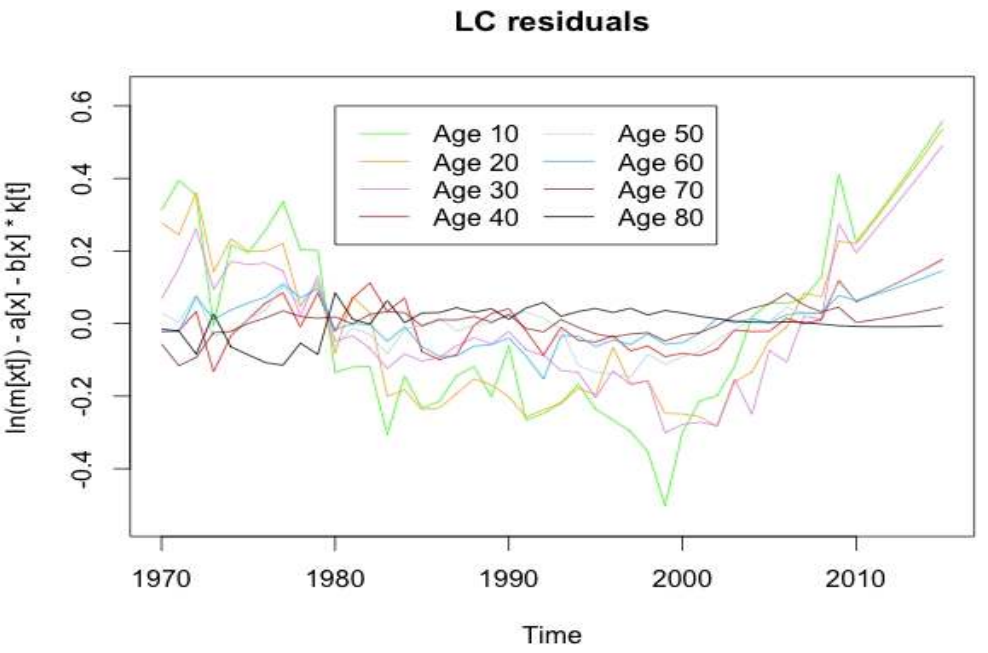
다음 그림은 Lee-Carter를 적합한 결과이다.

〈그림 4-5〉 Lee-Carter 모형의 적합결과



이 그림에서 특이한 사항은 b(x)의 값이 40-60세 사이에서 흔히 볼 수 있는 아래로 들어간 패턴이 보이지 않는다는 점과 1970년대 k(t)의 값이 불안정하다는 점이다. LC로 적합한 후 잔차를 주요연령별로 그리면 다음과 같다.

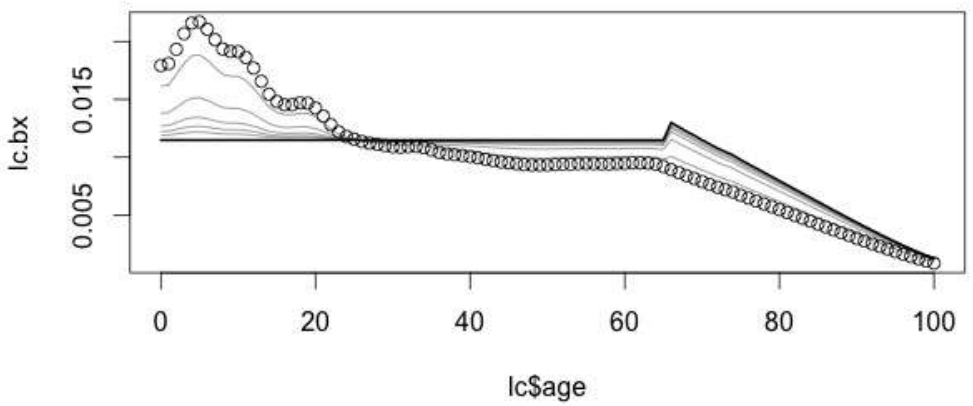
〈그림 4-6〉 Lee-Carter 모형의 잔차



이상적으로 잔차그림은 세로축에 0을 기준으로 랜덤하게 퍼져있어야 하지만 위 그림에서 보듯 잔차들이 시간에 따라 랜덤하지 않은 패턴을 보이고 있음을 알 수 있다. 특히 저연령에서의 잔차는 아래로 볼록한 곡선을 보여주어 LC 모형이 모든 불확실성을 제거하지는 못함을 알 수 있다.

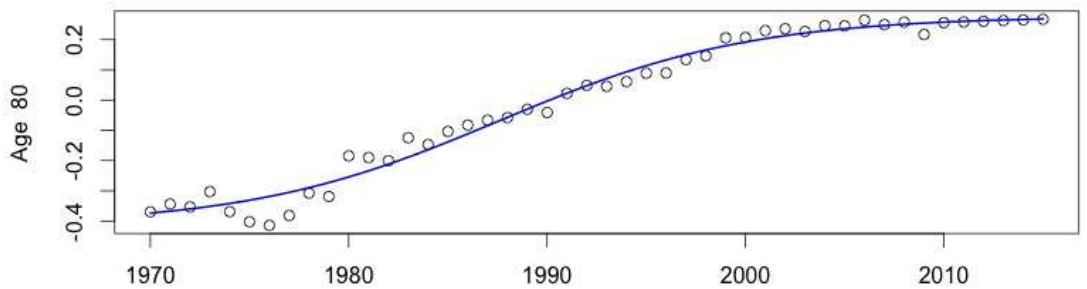
다음은 LLG 모형을 한국데이터에 적용했을 때 b(x)가 시간에 따라 어떻게 다르게 예측이 되는지를 보여주는 그림이다. 동그라미로 표현되는 값이 LC 모형으로부터 계산된 b(x)값이고 이 값을 출발점으로 하여 rotation 이 시작되면 시간이 지남에 따라 점점 평평한 모양의 곡선으로 바뀌어지며 궁극적으로는 65세 이하 모든 연령에 대해 b(x) 값이 기울기가 0인 상수함수로 수렴하게 되는 것이 요점이다. 그러나 앞에서 기술한 바와 같이 다중인구로의 확장이 쉽지 않기 때문에 더 이상 LLG모형에 대한 논의는 하지 않기로 한다.

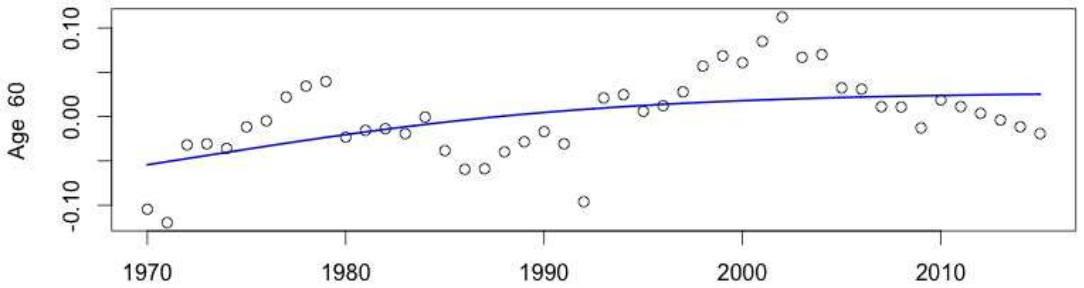
〈그림 4-7〉 LLG 모형의 교대현상 예측값

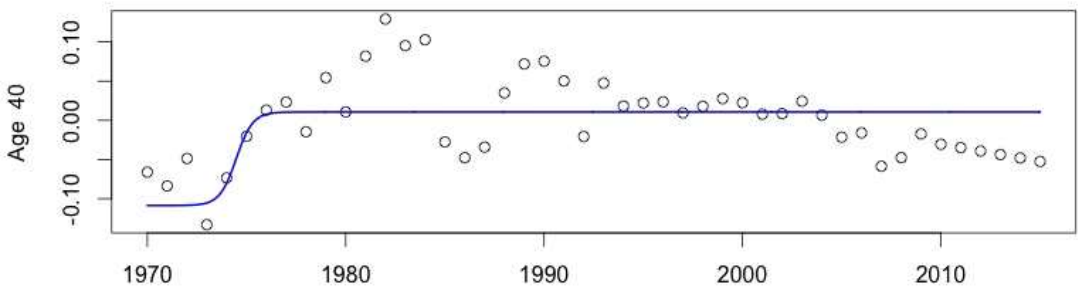


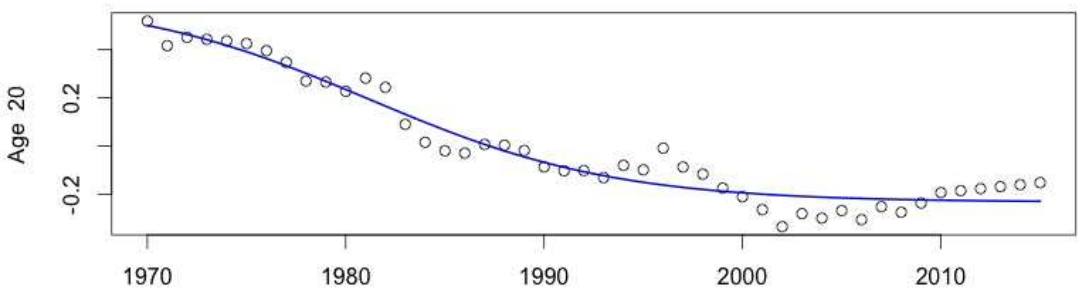
이제 로지스틱모형을 이용한 결과를 보도록 하겠다. 우선 잔차 $\ln m(x,t)-a(x)-\theta(t)$를 로지스틱함수인 g(x,t) 로 적합하는 것이다. 모든 연령에 대해 적합을 하였지만 예를 들어 보이기 위해 주요연령에서 관측값과 추정된 로지스틱 함수곡선을 아래에 그림으로 실었다.

〈그림 4-8〉 $\ln m(x,t)-a(x)-\theta(t)$ 에 대한 로지스틱함수의 적합



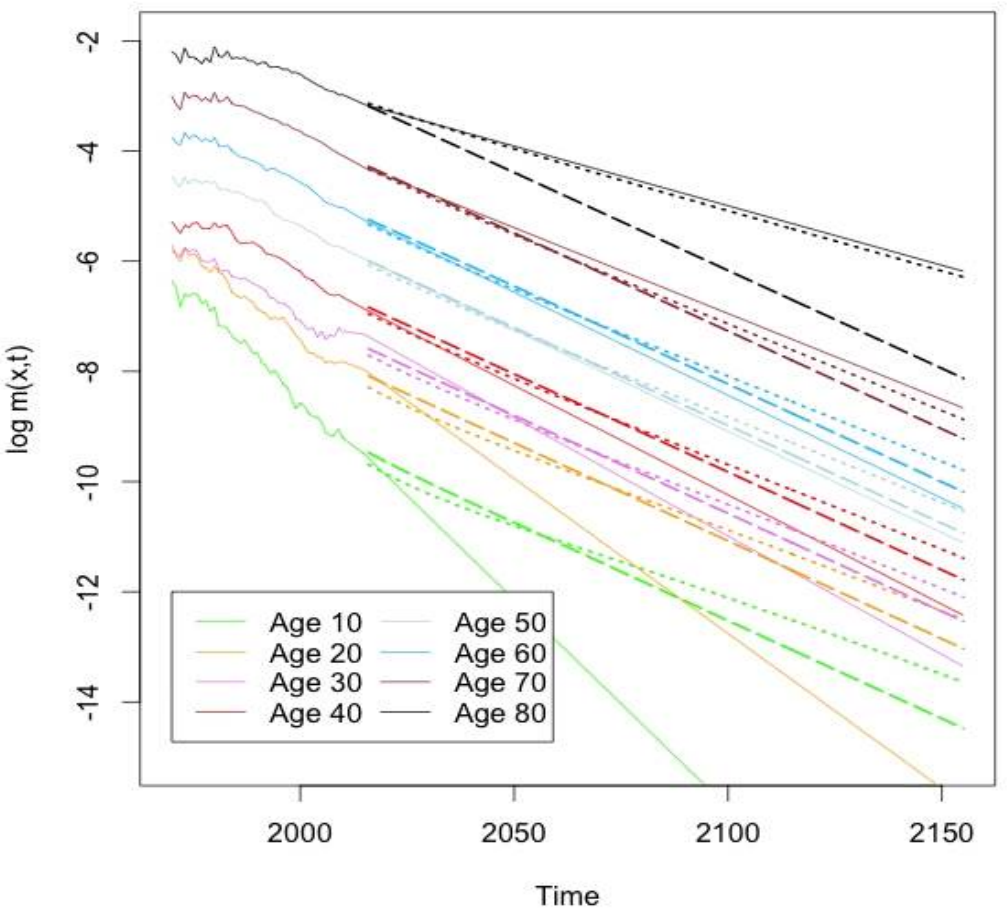

Age 60
0.10
0.00
-0.10
1970
1980
1990
2000
2010


Age 40
0.10
0.00
-0.10
1970
1980
1990
2000
2010


Age 20
0.2
-0.2
1970
1980
1990
2000
2010

다음은 이렇게 적합된 로지스틱모형을 기반으로 했을 때 미래의 로그사망률이 어떻게 예측이 되는지를 보여주는 그림이다. 그림에서 과거 (2015년 이전)의 값은 관측값들이고 이후의 값들은 세 개의 모형, 즉 LC (실선), LLG (점선), Logistic (파선) 모형으로부터 예측이 되는 값들을 나타낸다.

〈그림 4-9〉 $\ln m(x,t)$ 의 예측값 비교: LC (실선), LLG (점선), Logistic (파선)

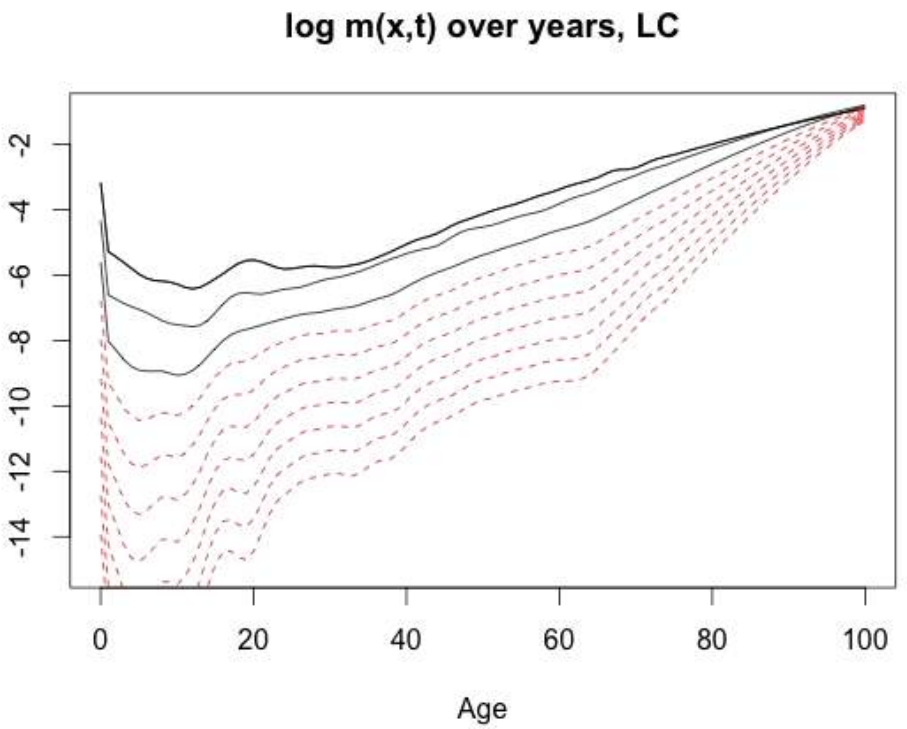


이 그림에서 다음을 알 수 있다.

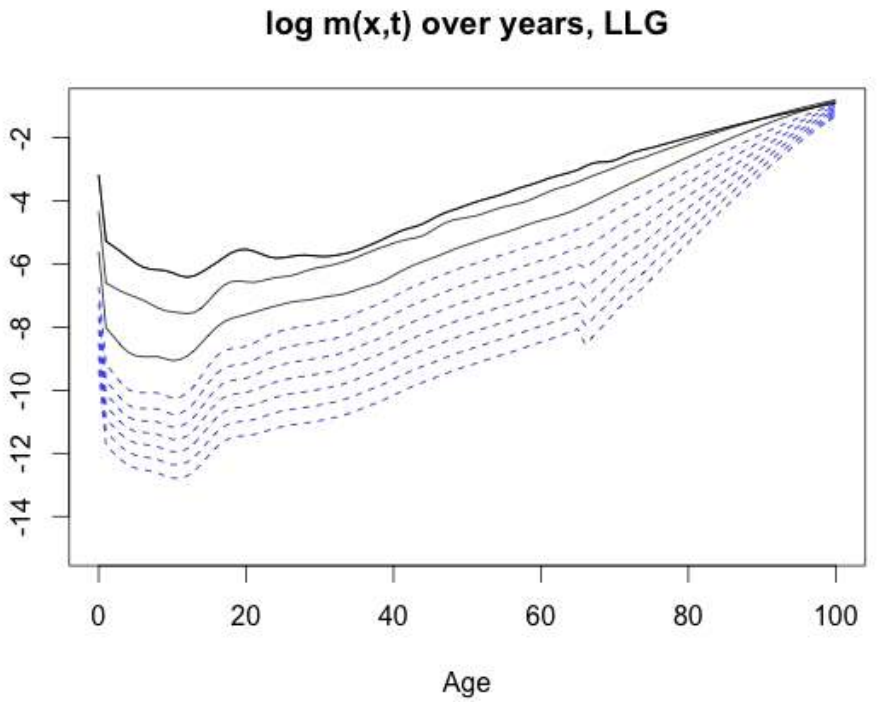
- LC모형의 경우 모든 연령에 대해 과거추세의 평균을 그대로 이어서 미래를 예측하고 있음을 알 수 있다.
- LLG 모형의 경우 LC모형과 비교했을때 저연령에서는 사망률 개선을 늦추고 고연령에서는 개선을 빠르게 함으로써 rotation을 제어한다. 그러나 고연령에서의 개선속도 증가는 저연령에서의 완화폭에 비해 상대적으로 약하다. 그리고 사망률 지표 $k^*(t)$ 가 기대수명을 유지하는 제약식에 의해 도출되는 값이기 때문에 이로 인해 rotation 이 시작되는 시점에서 위쪽 (고연령) 혹은 아래쪽 (저연령)으로의 점프가 있다. 여기에 대한 설명은 앞에서 기술하였다.
- Logistic 모형의 경우 초고연령 (80세)을 제외한 거의 모든 연령에 대해 LLG와 상당히 근접한 값들을 보여준다. 80세에 대해서는 최근의 추세를 강조하여 외삽을 하고 있음을 알 수 있다. 이 모형의 경우 저연령은 개선속도를 줄이고 고연령은 늘리는 패턴을 줌으로서 rotation을 제어하는데 이 점은 LLG와 동일하다. 그리고 궁극적으로 모든 연령의 사망률개선속도 (기울기)가 동일한 값에 수렴함도 알 수 있다.

위에서는 대표연령인 10, 20, ..., 80세에 대한 log m(x,t)값의 시간에 따른 값을 그림으로 표현했는데 이제 모든 연령에 대해 시간에 따른 log m(x,t)의 예측값을 세 개의 사망률 모형, 즉, LC, LLG 그리고 Logistic 하에서 어떻게 나타나는지 살펴보겠다. 그림에서 실선은 과거의 값들이고 점선은 모형에서 예측된 값을 나타낸다. 각 선은 1970년부터 2150년까지 매 20년을 주기로 예측하였다.

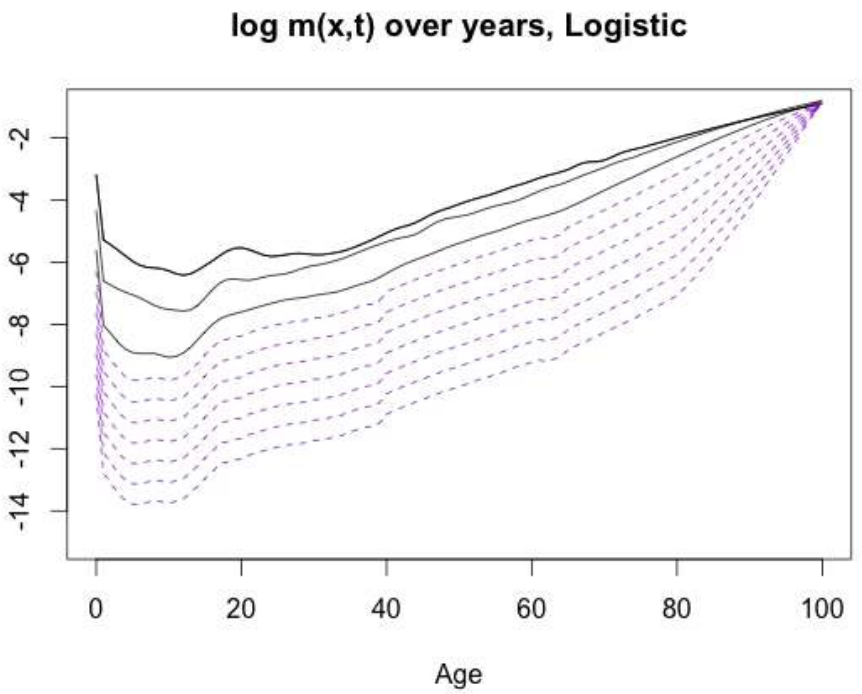
〈그림 4-10〉 로그사망률의 예측값, Li-Carter모형



〈그림 4-11〉 로그사망률의 예측값, LLG모형



〈그림 4-12〉 로그사망률의 예측값, 로지스틱모형



위의 그림으로부터 다음을 관찰할 수 있다.

- LC모형의 경우 모든 연령에 대한 사망률개선속도는 과거의 속도를 그대로 반영한다. 따라서 저연령에서는 미래에도 빠른 속도로 계속적인 사망률개선이 일어나며, 고연령에서는 미래에도 상대적으로 느린속도로 개선이 일어남을 알 수 있다. 따라서 사망률교대의 패턴을 바르게 표현하지 못함을 확인할 수 있다.
- LLG 와 Logistc 모형의 경우 둘 다 LC 의 문제를 극복하고 있다. 특히 저연령에서는 개선의 속도가 시간이 흐를수록 상대적으로 둔화되며 고연령에서는 반대로 속도가 증가함을 볼 수 있어 교대의 패턴을 바르게 구현해 내고 있음을 알 수 있다. 그러나 전반적으로는 로지스틱모형하에서 사망률의 개선이 좀 더 빠르게 나타나, 로지스틱

모형이 LC 와 LLG의 중간정도의 개선을 예측함을 알 수 있다.

제5절 Coherent rotation모형

지금까지 로지스틱함수를 사용하여 사망률의 교대현상을 모형화하는 방법에 대해 설명하였다. 하지만 남녀 각 성별에 대해 이 모형을 독립적으로 적용한다면 성별로 $\theta(t)$ 함수가 서로 다르게 추정이 될 것이다. 따라서 장기인구추계의 관점에서 보면 성별 정합성 (coherency)이 떨어지게 된다. 따라서 LL모형이 LC 모형에 성별정합성을 부여하는 것처럼, 앞 절에서 주어진 로지스틱모형에 성별 정합성을 줄 수 있는 모형을 생각해 보는 것은 당연하다.

이를 위해 다음과 같은 단계를 고려한다.

1. 남녀를 합친 자료 (total)을 이용하여 로지스틱 교대모형을 적합을 한다. 남녀를 합친 자료에 대해 적용하였으므로 여기서 얻어지는 $\Theta(t)$는 남녀공통의 항이 된다. 이는 마치 LL모형에서 남녀를 합친 자료로부터 공통항인 B(x)K(t) 를 계산해 내는 것과 동일하다.

2. 위에서 구한 $\Theta(t)$을 이용하여 각 성별로

$$\ln m(x,t) - a(x) - \Theta(t)$$

을 계산한 후 이 값에 성별로, 연령별로 로지스틱함수를 독립적으로 적합한다. (단, 독립적인 적합은 정해진 임계연령 u 미만에만 적용한다. 아래 항 참고)

3. 여기서 주의할 점은 다음과 같다. 앞서 직선과 로지스틱함수 둘 다 적합한 후 정해진 임계연령 u에 대해 [u, w] 의 연령구간내에서 g(x,t)를 직선과 로지스틱함수의 다음과 같은 가중평균으로 계산함을 설명하였다.

$$g(x,t) = \frac{x-u}{1+w-u} f_{Lin}(x,t) + \left(1 - \frac{x-u}{1+w-u}\right) f_{Logistic}(x,t)$$

남녀 각각에 대해 이를 적용하게 되면 이 식의 우변의 첫 항인 선형직선의 기울기가

서로 독립적으로 추정되므로 [u, w] 의 고연령구간내에서 우리가 원하는 정합성을 잃어버리게 된다. 그러나 로지스틱함수는 결국 상수에 수렴하므로 이러한 문제가 없다. 따라서, 성별 정합성을 모든 연령에서 유지하기 위해서는 위 식의 우변 첫 항의 선형함수는 남녀를 합친 자료로부터 각 연령에 대해 직선 추정을 하고 둘째항이 로지스틱함수는 남녀성별로 서로 독립적으로 각 연령에 대해 추정을 하는 것이 논리적으로 맞음을 알 수 있다.

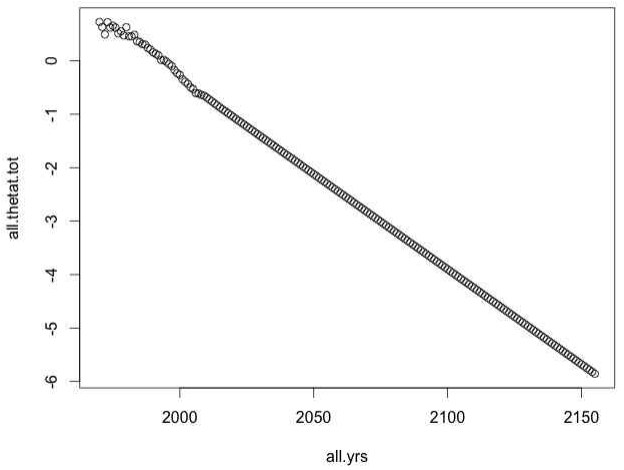
위의 세 단계를 거쳐 추정되는 모형식은 다음과 같이 요약할 수 있다. (아래 식에서 i=1, 2 는 남녀 성별로 따로 모형화됨을 의미하며 i없이 (x,t)로만 표현되는 함수들은 남녀합친 자료로 모형화됨을 의미한다.)

$$\ln m(x,t,i) = a(x,i) + \Theta(t) + g(x,t) + \epsilon(x,t,i)$$

이고

$$g(x,t,i) = \frac{x-u}{1+w-u} f_{Lin}(x,t) + \left(1 - \frac{x-u}{1+w-u}\right) f_{Logistic}(x,t,i)$$

〈그림 4-13〉 남녀공통 사망률지표 $\Theta(t)$

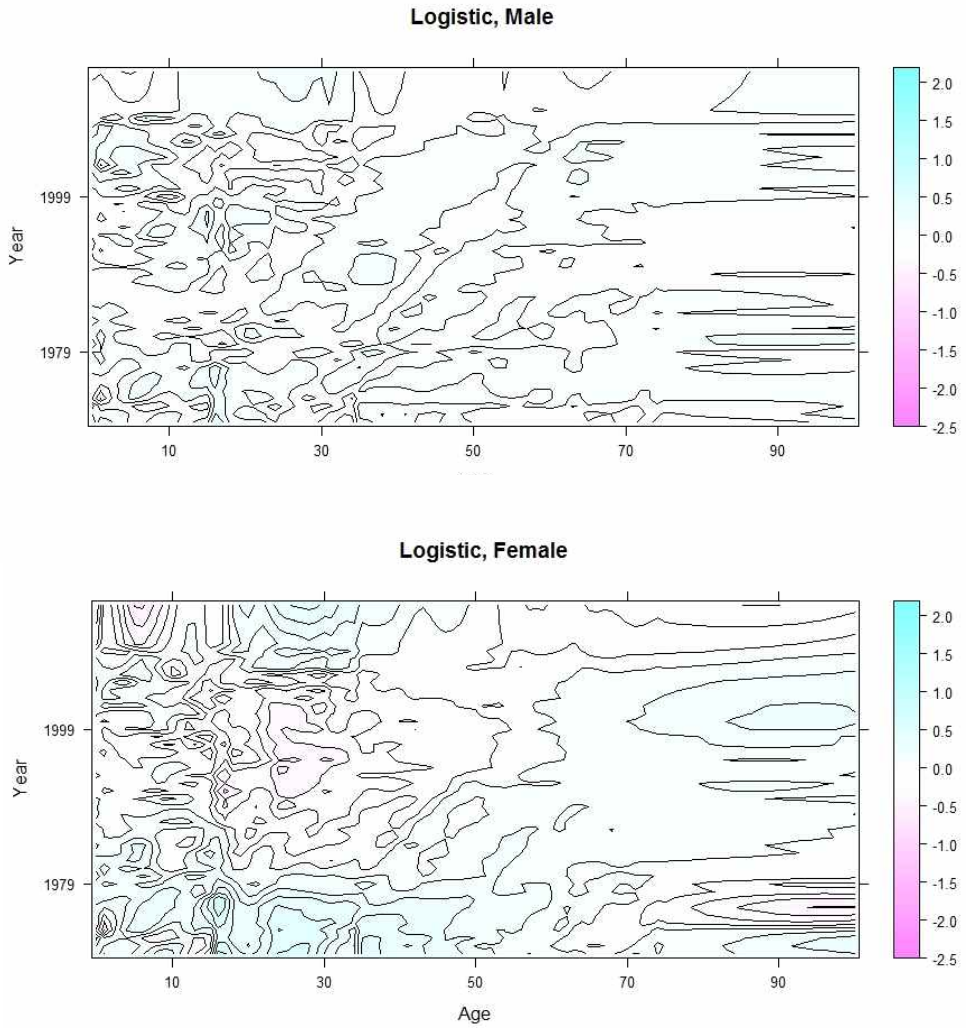


국내 자료에 위 모형을 적합한 결과는 적절하다고 판단되며 계산된 $\Theta(t)$의 적합된

값과 예측값은 <그림 4-13>과 같다.

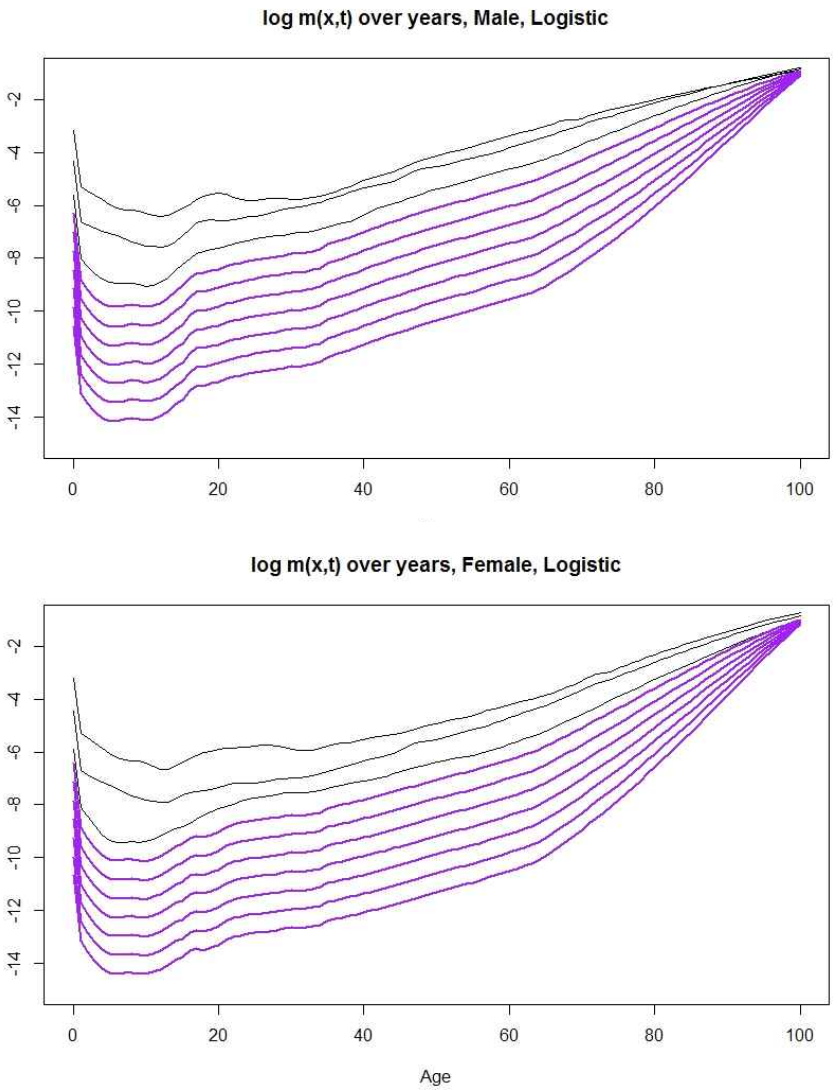
잔차의 그림은 남녀 각각에 대해 다음과 같이 주어진다. 잔차그림으로부터 특별한 패턴을 보이지 않음을 알 수 있고, 코호트효과 역시 보이지 않음을 알 수 있어 로지스틱 모형이 비교적 잘 작동하는 것을 볼 수 있다.

〈그림 4-14〉 로지스틱 사망률모형의 성별 잔차



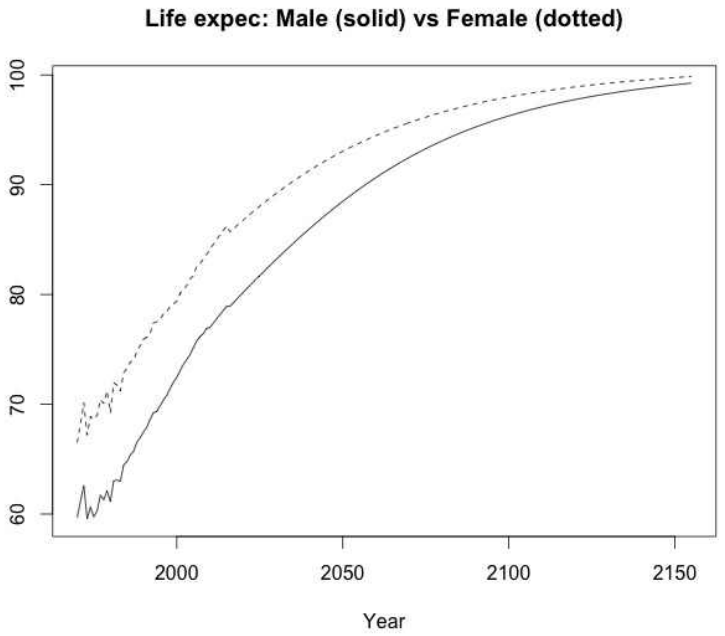
성별정합성을 유지한 로지스틱모형의 적합으로부터 도출된 성별 로그사망률의 예측값은 <그림 4-15>에 주어졌다.

〈그림 4-15〉 로그사망률의 성별 예측값



마지막으로 국내 자료에 위 모형을 적합한 결과 계산된 기대여명에 대한 결과는 <그림 4-16>에 실었다.

〈그림 4-16〉 성별 기대여명에 대한 예측값



제6절 맺음말

본 보고서에서는 로지스틱함수를 이용한 사망률의 적합과 예측을 위한 새로운 모형을 제안하였다. 모형의 구조는 Li-Carter 모형과 비슷한 측면이 있지만 공통 사망률지표를 반영한 후 남는 잔차에 대해 로지스틱함수로서 적합한다는 면에서 새로운 모형이다. 국내 자료에 대해 제안된 모형을 적용한 결과 사망률의 적합과 예측에 있어 대체적으로 만족스러운 값들을 얻었고, 성별 사망률예측에 있어 정합성을 유지할 수 있는 구조로의 일반화 역시 잘 작동함을 알 수 있었다. 그럼에도 불구하고 제안된 모형은 다음과 같은

잠재적인 한계점들이 있어 보인다.

첫째, 로지스틱함수 자체의 한계로 인해 종종 점근값 (Ultimate value) 이 가장 최근의 값과 비슷하게 결정이 된다. 이 같은 경향은 특히 중간연령 (30-60세)에 두드러지게 나타나는데 이는 미래의 사망률이 공통 개선지표인 $\Theta(t)$에 전적으로 의존한다는 의미로서 쉽게 표현한다면 최근자료에 큰 가중치를 주어 이 값이 계속 될 것이라고 예측하는 것과 비슷한 의미이다. 저연령과 고연령에서는 어느 시점까지는 지속적으로 증가함을 보여 이러한 경향이 현저히 감소한다.

둘째, 특정 연령들에 대해서는 로지스틱함수의 적합이 썩 좋지 않은 경우가 있다. 이는 공통 사망률지표를 반영한 후 남는 잔차의 시간에 따른 변화패턴이 로지스틱형태의 함수로 표현하기엔 무리가 있음을 시사한다. 이를 극복하기 위해 더 유연한 (flexible)함수를 사용할 수 있겠지만 (예를 들어 두 로지스틱함수들의 혼합을 생각할 수 있다) 적합의 정도가 높다고 하여도 결국 위의 첫 번째 문제점은 여전히 대두되므로 모수추정의 어려움을 고려할 때 더 많은 모수를 가진 함수를 사용하는 것이 바람직하다고 보여지지는 않는다. 그럼에도 불구하고 로지스틱함수의 적합이 잔차에만 적용이 되기 때문에 전체 사망률에 대해서 본다면 아주 큰 오차라고 볼 수 없으며 Lee-Carter 모형과 비교할 때 과거 자료에 대한 적합은 여전히 더 좋다.

마지막으로, 사망률개선의 교대현상은 비교적 최근에 이슈가 되는 문제로서 아직 학계에서 모두가 동의하는 모형이 존재하지 않는다. 보고자가 제안한 로지스틱모형 역시 교대현상을 반영하는 새로운 시도의 일환으로 보는 것이 합당하며, 향후 좀 더 많은 연구와 실증적인 분석이 이루어져야 할 것이다. 이러한 점에서는 현재 사망률개선의 교대현상을 모형화하는 대표적 모형으로서 이미 몇몇 국가에서 사용되고 있는 LLG모형과 새롭게 제안한 로지스틱모형을 좀 더 면밀히 비교하는 것도 의미가 있을 것이라 생각된다.

참고문헌

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. Journal of the American statistical association, 87(419), 659–671.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. Demography, 42(3), 575–594.

Li, N., Lee, R., and Gerland, P. (2013). Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. Demography, 50(6), 2037–2051.