

경제·사회적 요인을 반영한 장기 인구추계 모형 고도화

- 2022. 8. -

이 연구는 국회예산정책처의 연구용역사업으로 수행된 것으로서,
보고서의 내용은 연구용역사업을 수행한 연구자의 개인 의견이며,
국회예산정책처의 공식 견해가 아님을 알려드립니다.

연구책임자

경희대학교 김현식

경제 · 사회적 요인을 반영한 장기 인구추계 모형 고도화

2022년 8월

연구책임자 김현식 (경희대학교 사회학과)

연구보조원 오혜진 (고려대학교 사회학과)

이 연구는 국회예산정책처의 연구용역사업으로 수행된 것으로서,
본 연구에서 제시된 의견이나 대안 등은 국회예산정책처의 공식
의견이 아니라 본 연구진의 개인 의견임



국회예산정책처
National Assembly Budget Office

요 약

1. 연구 개요

- 2019년의 연구 용역과 2021년 NABO 내국인 인구 시범추계에 기반하여 내국인의 기준 자료를 재작성하고 2020년까지의 자료를 활용해 장래 50년간의 장래인구 추계 필요
- 내국인과 외국인의 추계를 구분하고 각각의 기준 인구를 바탕으로 출산율, 사망률, 국제이동을 추계한 후 코호트조성법으로 장래인구 추계 실시
- 내국인 추계에서 경제사회적 변수들이 출산율과 사망률 변화에 주는 영향들을 여러 수리 모형으로 추정하고 각 모형들의 장단점을 살펴본 다음 적절한 모형을 선택하여 장래 추계에 활용
 - 다양한 경제사회적 변수들을 고려한 후 최종적으로 여자의 경제활동참가율, 청년 실업률, 1인당 실질 GDP 상승률, 지니계수, 공동학교 졸업 인구 비율, 종교 없음 비율을 활용
- 외국인의 추계의 경우 등록센서스 2020년 자료를 기준 인구로 활용하여, Kim(2018)의 연구에 기반해 외국인 출산율을 추정하고, 정연·이나경(2022)의 연구를 바탕으로 외국인의 사망률을 추계하고, 2001년부터 2020년까지 5세별 외국인 국제순이동 자료에 기반해 외국인 국제순이동을 추계한 후 외국인 추계를 실시

2. 추계 자료

- 내국인 연앙인구로 통계청의 장래인구추계에 인구총조사의 내국인 비율을 적용하였고, 연령별 출산율을 구하기 위해 통계청의 인구동향조사 출생아수에서 외국 출생 비율을 제하였으며, 통계청 인구동향조사 사망자수 자료를 이용하여 각세별 사망률을 구하였고, 통계청 국제인구이동통계 자료를 활용하여 내국인 국제인구이동율을 추정

3. 출산율 추계

- 로그이차함수 모형(LL모형), 출산율이 가장 높은 연령을 기준으로 불연속 출산율을 가정하거나 연속 출산율을 가정하는 Peristera-Kostaki 모형(PK 모형), 일반화로그 감마 모형(GLG 모형)으로 출산율 추계
- 각 모수를 추정한 후 연도에 따른 선형으로 모수를 추계한 후 향후 50년간의 합계출산율 변화를 살펴보면, 모든 추계 모형에서 우하향하는 결과가 나타났으며, 로그이차

함수 모형, 일반화로그감마 모형, 불연속 PK 모형, 연속 PK 모형 순으로 하락의 정도가 심화되는 것으로 나타났고, 로그이차함수 모형을 제외한 다른 모형에서는 2050년 전후에 합계출산율이 0으로 하락하는 형태를 보임

4. 사망률 추계

- Li-Lee-Gerland(2013)모형을 활용하여 추계하였으며, $K(t,i)$ 값을 선형으로 으로 추계할 경우 2070년 남자의 기대수명은 94.5, 여자의 기대수명은 98.1년으로 나타남

5. 국제이동 추계

- 통계청의 국제이동 추계방법을 적용하여, 2010년부터 2019년까지 국제인구이동율을 구한 다음 이 중 최댓값과 최솟값을 빼고 나머지 8개년도의 평균을 구하여 국제인구이동율을 사용하였으며, 2020년이 사용되지 않은 이유는 COVID-19으로 인한 외부충격의 효과를 제한하기 위해서임

6. 추계 결과

- 추계 결과는 뒤에서 경제사회적 변수를 고려한 결과와 같이 제시함
- 모든 시나리오에서 내국인은 줄어드는 것으로 나타났으며, 로그이차함수 모형으로 출산율을 추계했을 때 감소의 정도가 가장 적음

7. 경제사회적 변수를 활용한 추계

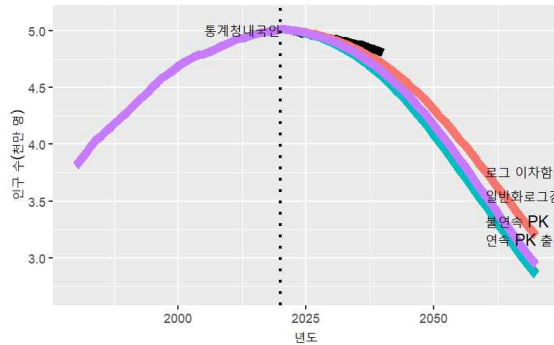
- 출산율과 사망률에 영향을 주는 경제사회적 변수로 성별 경제활동참가율, 실업률, 고용률, 청년실업률, 1인당 실질 GDP 연간 성장률, 소비자물가상승률, 지니계수, 학력별 인구 비율, 취업자의 주당 평균노동시간, 아내의 초혼연령, 기혼여성의 자녀 필요성 질문에 꼭 있어야 한다고 응답한 여성의 비율, 종교가 없다고 응답한 사람들의 백분율을 고려하였으나 변수들과의 높은 상관성이 존재하는 항목들을 제외하고 최종적으로 여자의 경제활동참가율, 청년실업률, 1인당 실질 GDP 연간 성장률, 지니계수, 고등학교 졸업 비율, 종교없음 비율을 변수로 활용
- 이들 변수들의 장래 50년간 추계를 위해 관찰값, 1차 차분값, 2차 차분값의 VAR(vector autoregression) 모형과 ARIMA(autoregressive integrated moving average) 모형을 추정하였으며, 실질 GDP 성장률을 ARIMA 모형을, 그 외의 변수들에는 1차 차분 VAR 모형을 적용

□ 2000년부터 2020년의 자료에 적용해 경제사회적 변수들이 출산율 및 사망률 모수에 주는 영향을 추정한 후 그 회귀값을 2021년 이후의 자료에 적용해 출산율 및 사망률 모수 추계 후 출산율과 사망률 추계 진행

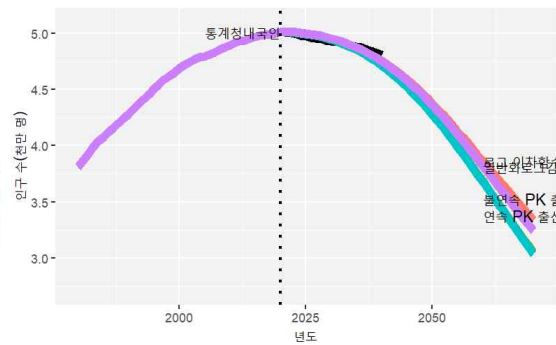
□ 다음 그림들은 경제사회적 변수를 고려하지 않을 때와 고려했을 때의 결과를 보여줌

A. 전체

가. 경제사회적 변수 고려 없음

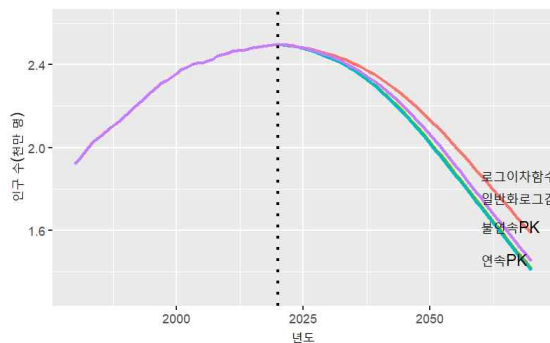


나. 경제사회적 변수 고려

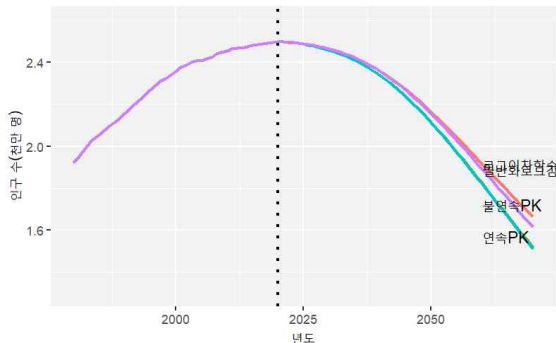


B. 남자

가. 경제사회적 변수 고려 없음

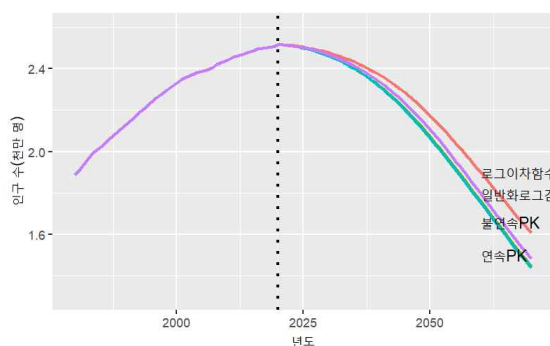


나. 경제사회적 변수 고려

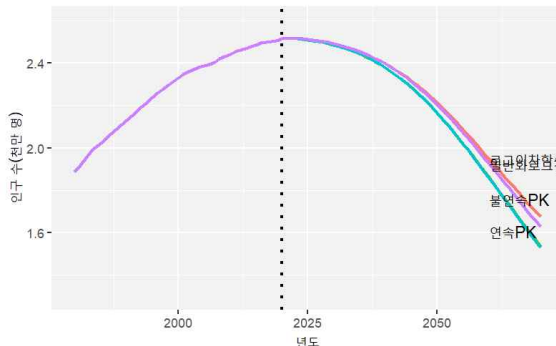


C. 여자

가. 경제사회적 변수 고려 없음



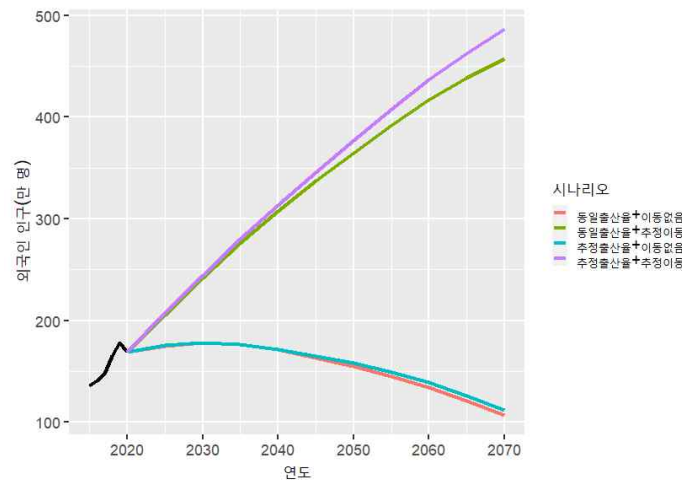
나. 경제사회적 변수 고려



□ 경제사회적 변수들을 고려했을 때 추계 인구수가 약간 많아지는 것으로 타나났으나 그리 큰 차이를 보이지는 않음

8. 외국인 추계

- 2020년의 등록센서스 외국인 수를 기준 인구로 설정하고, 정연·이나경(2022)의 연구에 근거해 외국인의 사망률을 추정하고, Kim(2018)의 연구 및 2020년 등록센서스 무배우자 및 유배우자 비율에 기반하여 외국인 출산율을 추정하고, 2001년부터 2020년의 외국인 국제이동 자료에 근거해 국제이동을 추정함
- 다음 그림은 출산율 가정(동일출산율은 내국인과 동일한 출산율을, 추정출산율은 Kim의 연구에 근거한 출산율 추정)과 국제이동가정(이동없음은 국제이동이 없다는 가정이며, 추정이동은 2001년부터 2020년의 평균값을 이용한 추정)에 따른 추계 결과를 보여줌



- 위 그림을 보면 출산율 가정에 따른 차이는 미미하며 국제이동에 따른 차이가 커 외국인 장래 인구수는 국제이동에 따라 좌우될 것으로 판단함

9. 시사점 및 향후 연구 방향

- 이 보고서에서는 내국인과 외국인에 대한 추계를 실시하였으며, 내국인 추계에 경제 사회적 변수를 고려한 추계를 실시
- 경제사회적 변수를 고려한 추계에서 변수들의 관찰 기간이 2000년부터 시작된다는 면에서 기간이 짧아 다양한 수리 모형을 적용하는데 한계가 있으며, 이로 인해 더 많은 변수들을 고려하기 어렵다는 제한점이 있음
- 외국인 추계에 있어 외국인의 출산율과 사망률에 대한 연구가 축적되어 보다 다양한 관점이 고려되는 추계 방법을 모색하여야 함

차 례

1. 연구 개요	1
가. 연구 목적	1
나. 기대효과	2
다. 활용방안	3
2. 추계 자료	4
가. 내국인 연앙인구	4
나. 모의 각세별 출생아 수	7
다. 각세별 사망자 수	10
라. 내국인 국제인구이동통계	12
마. 성비	13
3. 출산율 추계	15
가. 추계 모형	15
나. 추계 결과	23
4. 사망률 추계	25
가. 문헌 검토	25
나. 자료 분석	27
다. 추계 과정 및 결과	28
5. 국제이동 추계	32
가. 문헌 검토	32
나. 자료 분석	32
다. 추계 모형	34
라. 추계 결과	34
6. 추계 결과	51

7. 경제사회적 변수를 활용한 추계	40
가. 경제사회적 변수 관련 선행 연구	40
나. 자료	41
다. 기초분석	47
라. Vector autoregression 모형을 활용한 설명 변수 추계	51
마. ARIMA 모형을 활용한 설명 변수 추계	59
바. 출산율과 사망률 모수 추계 결과	62
8. 외국인 추계	73
가. 자료 및 방법	73
나. 추계 결과	80
9. 시사점 및 향후 연구 방향	84
참고문헌	86

표 차례

<표 2.1> $\log(-\log(80\text{세 이상 각세별 인구 비율}))$ 과 연령의 선형 관계 추정값	5
<표 2.2> 연도별 분석대상 및 분석미대상 출생아수 분포	8
<표 2.3> 2010년 및 2015년 인구주택총조사의 연령별 외국출생 분포	9
<표 2.4> 연도별 연령기록과 연령미상 사망자 수	11
<표 2.5> 연도별 연령기록과 연령미상 내국인 출입국자 수	13
<표 3.1> 로그 이차함수 모형의 모수 추정값	16
<표 3.2> PK 모형의 모수 추정값	20
<표 3.3> 일반화 로그감마 모형의 모수 추정값	22
<표 7.1> 다양한 자료별 지니계수 추이	43
<표 7.2> 다양한 자료별 주당 평균노동시간	44
<표 7.3> 자녀의 필요성 및 종교 없음의 백분율 변화	46
<표 7.4> 경제사회적 변수들 사이의 상관계수	49
<표 7.5> 출산율 및 사망률 모수들과 경제사회적 변수들 사이의 상관계수	50
<표 7.6> 관측값 VAR(2) 모형의 계수	53
<표 7.7> 1차 차분 VAR(1) 모형의 계수	57
<표 7.8> 2차 차분 VAR(1) 모형의 계수	58
<표 8.1> 등록센서스와 등록외국인 자료를 이용한 사망자 추정	76
<표 8.2> 외국인과 내국인의 유배우자 비율을 활용한 출산율 비 계산	79
<표 8.3> 외국인 국제순이동 추계 값	80

그림 차례

<그림 2.1> 통계청 추계 자료의 2000년 이후 80세 이상의 인구 비율	4
<그림 2.2> $\log(-\log(80\text{세 이상의 인구 비율}))$ 과 연령의 선형관계	5
<그림 2.3> 연도별 및 5세별 외국인 비율:관측값	6
<그림 2.4> 연도별 및 연령별 외국인 비율:추정값	7
<그림 2.5> 여아 1명당 남아 출생아 수 비	14
<그림 3.1> 연령별 출산율 변화	15
<그림 3.2> 로그 이차함수 모형의 모수 추정값 및 추계값	17
<그림 3.3> 불연속 PK 모형의 모수 추정값	19
<그림 3.4> 연속 PK 모형의 모수 추정값	19
<그림 3.5> 일반화로그감마 모형의 모수 추정값	21
<그림 3.6> 출산율 추계 결과: 합계출산율	23
<그림 3.7> 연령별 출산율의 추계 결과	24
<그림 4.1> 성별 10세 연령별 로그 사망률의 변화	27
<그림 4.2> 성별 연령별 로그 사망률의 몇몇 연도별 변화	28
<그림 4.3> 성별 관찰 기대수명 및 추계 기대수명	28
<그림 4.4> 추정 $B(x,t)$	29
<그림 4.5> 추정 $K(t,i)$: LC와 LLG의 $K(t,i)$ 비교	30
<그림 4.6> 성별 연령별 로그 사망률의 변화	31
<그림 5.1> 연령별 내국인 국제순이동	33
<그림 5.2> 내국인 국제순이동률 추계 값	34
<그림 6.1> 추계 인구: 전체	36
<그림 6.2> 추계 인구: 남자	37
<그림 6.3> 추계 인구: 여자	37
<그림 6.4> 출산율 가정에 따른 추계 인구 피라미드	38
<그림 7.1> 경제사회적 변수들의 관측값	48
<그림 7.2> VAR level 모형의 추계 결과	52
<그림 7.3> 1차 차분 값들의 변화	54
<그림 7.4> VAR 1차 차분 모형의 차분 추계 결과	55
<그림 7.5> 1차 차분 모형의 추계 결과	56
<그림 7.6> 2차 차분 모형의 추계 결과	59
<그림 7.7> ARIMA 모형의 추계 결과	61

<그림 7.8> 로그 이차함수 모형의 모수 추계	62
<그림 7.9> 불연속 PK 모형의 모수 추계	63
<그림 7.10> 연속 PK 모형의 모수 추계	63
<그림 7.11> 일반화로그감마 모형의 모수 추계	64
<그림 7.12> 사망률 $k(t)$ 추계	64
<그림 7.13> 합계출산율 추계	65
<그림 7.14> 기대수명 추계	65
<그림 7.15> 경제사회적 변수를 활용한 인구 추계: 전체	66
<그림 7.16> 경제사회적 변수를 활용한 인구 추계: 남자	66
<그림 7.17> 경제사회적 변수를 활용한 인구 추계: 여자	67
<그림 7.18> 경제사회적 변수 고려 전후 추계 인구 비교: 전체	68
<그림 7.19> 경제사회적 변수 고려 전후 추계 인구 비교: 남자	69
<그림 7.20> 경제사회적 변수 고려 전후 추계 인구 비교: 여자	70
<그림 7.21> 출산율 가정에 따른 추계 인구 피라미드	71
<그림 8.1> 등록센서스와 등록외국인 현황 자료 비교	74
<그림 8.2> 외국인 국제이동(2001-2020)	79
<그림 8.3> 외국인 추계 결과	81
<그림 8.4> 외국인 추계 인구 피라미드: 내국인과 동일 출산율 가정	82
<그림 8.5> 외국인 추계 인구 피라미드: 추정 출산율 가정	83

1. 연구 개요

가. 연구 목적

- (1) 2019년의 연구 용역(김현식·유삼현·안재혁, 2019)에 기반하여 기준 자료를 재작성하고 2020년까지의 자료를 활용해 장래 50년간의 장래인구 추계
 - (가) 2019년의 연구 용역에서는 주민등록연앙인구를 사용하여 기준인구를 추정하였으나 이 연구에서는, 김경수·김상미(2021)의 연구를 반영하여, 2020년까지 통계청의 추계인구와 1980년부터 2020년까지의 인구총조사로부터 구한 내외국인 비율을 적용하여 내국인 기준인구를 작성
 - (나) 15-49세 여성의 각세별 출생아수 자료를 이용해 연령별 출산율을 추정하고 2020년부터 2070년까지 출산율을 추계
 - 선행 연구와 마찬가지로 로그 이차함수 모형, Peristera-Kostaki 모형을 변형한 두 가지 모형, 일반화 감마 모형을 이용하여 출산율 추계
 - (다) 인구동향조사의 연령별 사망자 수를 이용하여 과거 사망률을 추정하고 Lee-Carter 모형(1992), Li-Lee 모형(2005), Li-Lee-Gerland 모형(2013)을 검토한 후 LLG 모형을 이용하여 사망률 추계를 진행
 - (라) 통계청의 국제인구이동 가정에 따라 10년의 자료(2011-2019)를 이용하여 최댓값과 최솟값을 제한 후 순이동률 평균을 이용하여 장래 순이동률 추정
- (2) 경제사회적 변수들이 장래인구 변화에 주는 영향들을 여러 수리 모형으로 추정하고 각 모형들의 장단점을 살펴본 다음 적절한 모형을 선택하여 장래 추계에 활용
 - (가) 이 보고서에서 고려한 경제사회적 변수들은 다음과 같음
 - 성별 경제활동참가율, 실업률, 고용률, 청년실업률, 1인당 실질 GDP 연간 성장률, 소비자물가상승률, 지니계수, 학력별 인구 비율, 취업자의 주당 평균노동시간, 아내의 초혼연령, 자녀 필요성 질문에 대한 자녀는 꼭 필요하다고 응답한 유배우 여자 비율, 한국종합사회조사에서 종교가 없다고 응답한 사람들의 비율을 경제사회적 변수들로 고려하였음

- 이중 기초분석을 통하여 출산율 모수에 영향을 주는 변수들로 여자의 경제활동참가율, 청년실업률, 1인당 실질 GDP 상승률, 지니계수, 고등학교 졸업 인구 비율, 종교 없음 비율을 설정
- 사망률 모수에 영향을 주는 변수들로 출산율 모수에 영향을 주는 변수들을 설정하였으나 종교가 건강에 주는 영향을 납득하기 어렵다는 제안에 따라 종교 없음 비율은 제외

(나) 이들 변수들의 2000년부터 2020년의 관측값에 기초하여 2021년부터 2070년까지 50년간 추계 진행

- 관측값, 1차 차분, 2차 차분 값들의 VAR(Vector autoregression) 모형에 더해, ARIMA(autoregressive integrated moving average) 모형을 활용

(다) 이들 변수를 추계한 후 2000년부터 2020년 사이의 경제사회적 변수들과 출산율 및 사망률 모수들의 관계를 이용하여 2021년부터 2070년까지 출산율 및 사망률 모수 추계 후 인구 추계 적용

(3) 등록센서스 2020년 자료를 기준 인구로 활용하고, Kim(2018)의 연구에 기반해 외국인의 출산율을 추계하고, 정연·이나경(2022)의 연구를 바탕으로 외국인의 사망률을 추계하고, 2001년부터 2020년까지 5세별 외국인 국제순이동 자료에 기반해 장래의 외국인 국제순이동을 추계한 후 외국인 추계를 진행

나. 기대효과

- (1) 통계청의 장래인구추계를 기다리거나 정책 목표 집단과 상이한 모집단에 기반한 추계 결과를 사용하기보다 필요한 시기에 적절한 목표 집단에 대한 추계를 실시함으로써 현실적인 재정전망을 진행할 수 있을 것임
- (2) 통계청의 추계 방식을 그대로 답습하지 않고 여러 인구 모수들에서 새로운 추계 방법을 시도하여 결과를 비교해 봄으로써 새로운 방법들의 장단점을 파악할 수 있고, 그 기반위에서 보다 한국의 상황에 적합한 추계 기법을 개발하는데 공헌
- (3) 특히 경제·사회적 요인을 반영한 추계를 실시함으로써 향후 전망에 대한 보다 설득력 있고, 이해가능한 결과를 제시할 수 있음

- (4) 외국인의 출산율 및 사망률 자료를 축적하여 이 분야에 대한 연구를 촉진시킬 수 있으며, 부족한 자료가 있으면 생산해야할 자료 목록을 만들어 향후 연구 방향을 제시하고 새로운 자료 수집을 활성화할 수 있음
- (5) 다양한 시나리오를 작성하여 인구 추계를 진행함으로써 미래 실현 양상이 특정 시나리오에 부합할 시 제시한 인구 전망과 유사하게 전개될 가능성이 크므로 다양한 대처 방안을 구성할 수 있음
- (6) 추계 결과만을 제시하지 않고, 추계 과정을 통계 프로그램 R로 작성하여 제출함으로써, 향후 국회예산정책처의 추계 모형을 개발하는데 일조

다. 활용방안

- (1) 기초연금 재정전망 등 사회보장재정 전망시 기초 자료로 활용
- (2) 연구 과정에서 축적하는 외국인의 출산율과 사망률은 이 분야의 향후 연구를 선도하고, 다양한 정책 형성, 예를 들어, 다문화가족 지원 방안의 기초 자료로 활용할 수 있음
- (3) 향후 국회예산정책처의 추계 모형 개발의 초기 모형으로 활용

2. 추계 자료

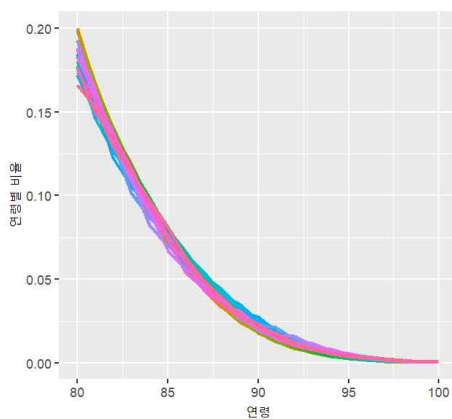
가. 내국인 연앙인구

(1) 통계청 추계 자료 (kosis.kr)

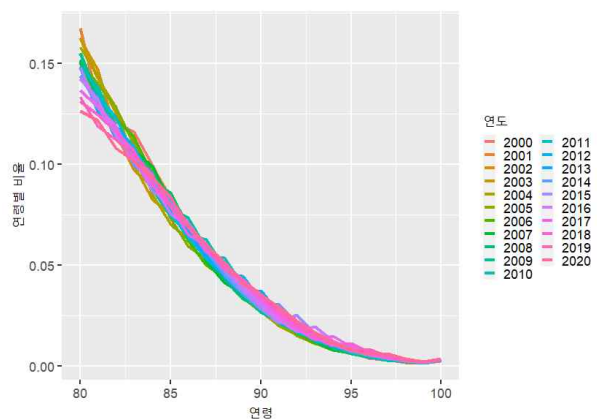
- 통계청의 장래인구추계는 각년도 7월 1일 시점의 연앙인구로 대한민국에 3개월 이상 거주하는 내국인과 외국인을 대상으로 작성
- 내국인만을 구분하기 위해 1980년부터 2020년까지 5년 간격으로 이루어진 인구총조사의 내외국인 수를 활용하여 내국인의 비율을 통계청 추계 자료에 적용
- 통계청 추계 자료는 1980년부터 2020년의 자료를 이용
- 통계청 추계 자료는 1999년까지 79세까지는 각세별로 되어 있으나 80세 이상은 하나의 연령으로 처리되어 있고, 2000년부터는 99세까지 각세별로 되어 있고, 100세 이상이 하나의 연령으로 처리되어 있음
 - 이로 인해 1999년까지 80세 이상의 인구수를 99세까지 각세별로, 100세 이상을 하나의 연령으로 취급하여 나누어줄 필요가 있음
 - 이러한 작업을 하기 위해 2000년부터 80세 이상의 인구 대비 80세 이상 각 연령대의 비율을 살펴봄

<그림 2.1> 통계청 추계 자료의 2000년 이후 80세 이상의 인구 비율

1) 남자



2) 여자

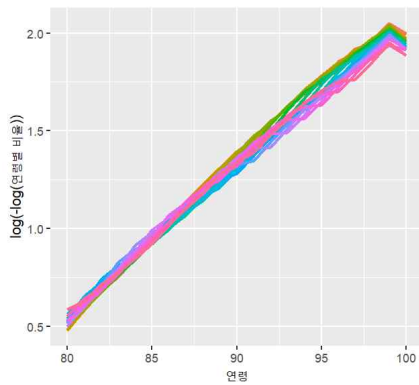


- 위의 그림을 보면 연령과 비율의 관계가 명확해 보이지 않음

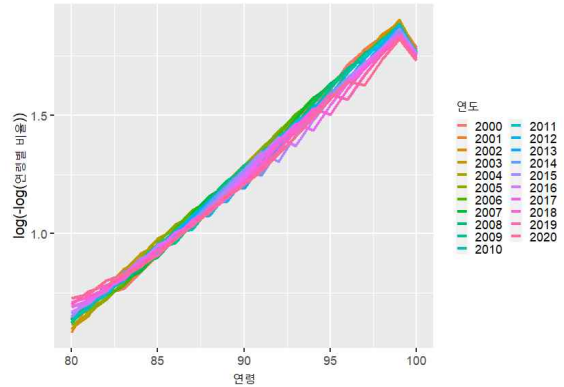
- 하지만 각세별 비율에 로그를 취한 다음 음의 부호를 붙여 다시 로그를 취하면, 즉 $\log(-\log(\text{비율}))$ 을 계산하면 연령과의 선형 관계가 뚜렷이 나타남

<그림 2.2> $\log(-\log(80\text{세 이상의 인구 비율}))$ 과 연령의 선형관계

1) 남자



2) 여자



- 연령을 설명 변수로 설정하고 $\log(-\log(\text{비율}))$ 을 종속 변수를 설정한 후 100세 이상을 제외하고 최소제곱법(ordinary least squares) 모형으로 관계를 추정하면 다음과 같은 결과를 얻음

<표 2.1> $\log(-\log(80\text{세 이상 각세별 인구 비율}))$ 과 연령의 선형 관계 추정값

	남자			여자		
	계수	표준오차	p-값	계수	표준오차	p-값
절편	-5.635	0.026	0.000	-4.581	0.023	0.000
연령	0.077	0.000	0.000	0.065	0.000	0.000
R ²	0.994			0.993		

- 위 결과를 1999년 이전 자료에 적용하여 80세부터 110세까지 비율을 구한 후 100세 이상을 합해 100+ 연령의 비율을 구함
- 100+ 연령을 포함하여 80세 이상의 비율 합이 1이 되도록 재조정 과정을 거침
- 이후 80세 이상의 관측된 인구수를 각 비율로 나누어 주고 정수로 반올림하여 각세별 인구수를 추정

(2) 인구총조사의 내국인 비율 (kosis.kr)

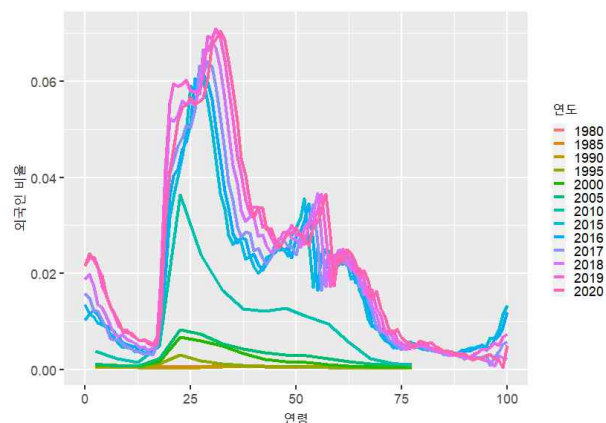
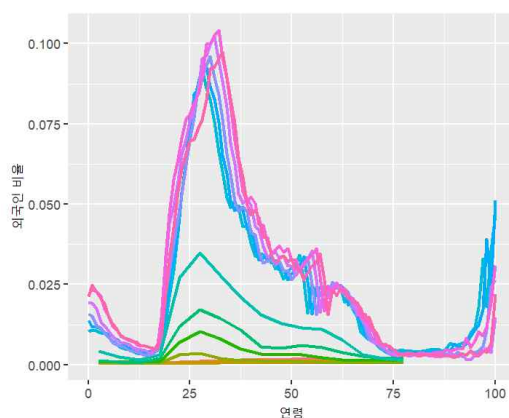
- 1980년부터 5년 간격으로 조사가 이루어진 인구총조사의 내국인 비율을 계산

- 각 조사별 총인구수와 외국인수를 kosis.kr에서 내려 받아 내국인 비율을 계산
- 외국인수 자료는 2005년까지 0세와 1-4세, 그 이후로 5세별로 74세까지 있으며 75세 이상이 하나로 묶여 있음
- 2010년은 0-4세가 하나의 연령으로 묶여 있음
- 2015년부터 2020년은 매년 100+까지 각세별로 외국인 수가 있음
- 총인구수는 2010년까지 85세+까지 5세별로 인구수가 있으며 2015년부터 2020년 자료는 100+까지 각세별로 인구수가 있음
- 2010년까지는 75+가 상한연령이 되도록 5세 간격별로 외국인 비율을 구하고, 2015년부터는 각세별로 100+까지 외국인 비율을 구하였으며, 그 결과는 다음과 같음

<그림 2.3> 연도별 및 5세별 외국인 비율: 관측값

1) 남자

2) 여자

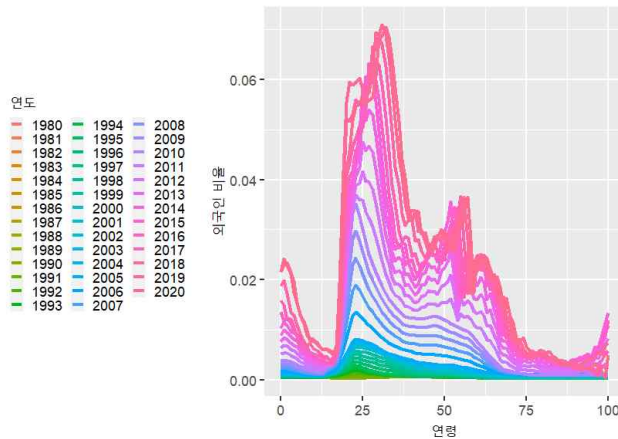
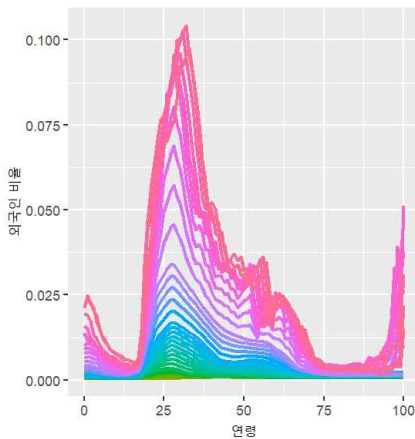


- 2010년까지는 위 자료의 두 인접 5세별 비율을 이용하여 단순 선형모형으로 각세별 비율을 구함
 - 예컨대 0-4세의 비율을 a 라고 하고 5-9세의 비율을 b 라고 하면, 이들의 연령을 각각 2.5세 및 7.5세라고 보고, 3세의 비율을 $a+0.5*(b-a)/5$ 로 구함
 - 0세부터 2세까지의 비율은 2.5세의 비율과 동일하다고 가정
 - 77.5세가 넘는 각세별 연령의 비율은 77.5세의 비율과 같다고 가정
- 인구총조사 자료가 없는 중간연도의 자료는 성별 및 각세별로 두 인접 인구총조사 자료의 단순 선형 모형으로 비율을 구함
- 최종 각세별 비율은 다음과 같음

<그림 2.4> 연도별 및 연령별 외국인 비율: 추정값

1) 남자

2) 여자



나. 모의 각세별 출생아 수

(1) 통계청 인구동향조사 출생아수 KOSIS 자료

- 통계청에서 수집 및 공표하고 있는 인구동향조사는 “인구동태사건이 발생한 대한민국 국민 중 국내신고자”가 조사모집단이므로 외국에 거주하는 대한민국 국민이 국내에서 신고할 경우 통계에 포함되어 출생아 수가 과도하게 추정될 여지가 있음
- 외국에서 신고하는 경우와 출생자 주소에서 외국으로 입력하는 경우가 있을 수 있으나, 어떻게 하든 외국에 거주하는 한국국적 부모의 출생아는 모두 부모의 주소지로 신고 후 그 주소지로 등록한다고 하므로, 출생아 수 자료는 국내 출생을 과도하게 집계할 가능성이 큼
- 기준 년도가 10년이 지난 경우 지난 10년 동안 접수한 신고서를 누적 집계하고 10년이 지나지 않은 경우 기준 년도 익년 4월까지 접수된 신고서를 기준으로 집계
- 1981년부터 모의 연령 각세별 출생아 수 자료가 가용
- 전체 출생아 중 모의 연령이 15세~49세인 경우만 이용하기 때문에 이로 인한 문제를 살펴보기 위해 다음의 표에 2000년 이후 분석대상과 미대상의 분포를 제시

<표 2.2> 연도별 분석대상 및 분석미대상 출생아수 분포

연도	전체	분석대상		분석미대상			
		15-49세	백분율	15세미만	50세이상	연령미상	백분율
1981	867,409	866,347	99.9	63	498	501	0.1
1982	848,312	847,550	99.9	3	364	395	0.1
1983	769,155	768,522	99.9	1	254	378	0.1
1984	674,793	674,248	99.9	4	196	345	0.1
1985	655,489	654,977	99.9	3	127	382	0.1
1986	636,019	635,525	99.9	5	69	420	0.1
1987	623,831	623,255	99.9	7	144	425	0.1
1988	633,092	632,445	99.9	5	142	500	0.1
1989	639,431	638,820	99.9	7	139	465	0.1
1990	649,738	649,071	99.9	7	134	526	0.1
1991	709,275	708,503	99.9	6	136	630	0.1
1992	730,678	729,904	99.9	19	149	606	0.1
1993	715,826	715,082	99.9	10	77	657	0.1
1994	721,185	720,430	99.9	11	63	681	0.1
1995	715,020	714,129	99.9	7	69	815	0.1
1996	691,226	690,388	99.9	11	65	762	0.1
1997	675,394	674,455	99.9	7	53	879	0.1
1998	641,594	640,586	99.8	11	69	928	0.2
1999	620,668	618,298	99.6	10	64	2,296	0.4
2000	640,089	637,296	99.6	9	66	2,718	0.4
2001	559,934	557,048	99.5	8	55	2,823	0.5
2002	496,911	495,607	99.7	53	46	1,205	0.3
2003	495,036	493,571	99.7	48	63	1,354	0.3
2004	476,958	475,600	99.7	27	46	1,285	0.3
2005	438,707	437,361	99.7	26	76	1,244	0.3
2006	451,759	450,596	99.7	34	43	1,086	0.3
2007	496,822	495,622	99.8	5	47	1,148	0.2
2008	465,892	465,009	99.8	46	32	805	0.2
2009	444,849	444,044	99.8	27	28	750	0.2
2010	470,171	469,315	99.8	34	62	760	0.2
2011	471,265	470,883	99.9	19	66	297	0.1
2012	484,550	484,423	100.0	29	48	50	0.0
2013	436,455	436,192	99.9	20	28	215	0.1
2014	435,435	435,092	99.9	22	26	295	0.1
2015	438,420	438,153	99.9	16	14	237	0.1
2016	406,243	406,009	99.9	15	11	208	0.1
2017	357,771	357,579	99.9	6	9	177	0.1
2018	326,822	326,608	99.9	8	18	188	0.1
2019	302,676	302,507	99.9	10	18	141	0.1
2020	272,337	272,196	99.9	11	4	126	0.1

□ 위에서 제시한 표를 보면 분석에 포함되지 않는 출생아가 가장 많은 비율을 차지한

2001년 전체 자료의 0.5%이고, 최근으로 올수록 그 비율이 줄어들고 있으므로, 분석 대상이 되지 않는 자료로 인한 추계 문제는 그리 크지 않을 것으로 판단됨

(2) 인구주택총조사 마이크로데이터(<https://mdis.kostat.go.kr/index.do>)

- 외국출생 한국 자녀의 영향을 살펴보기 위해 인구주택총조사 마이크로데이터 2% 샘플을 살펴봄
- 국적에 대한 정보는 2010년과 2015-2020년에 있으나 2016년부터 2020년의 마이크로데이터는 출생지에 대한 정보가 없어 분석이 불가능하여 이전 연구처럼 2010년과 2015년의 자료를 분석
- 태어날 때 국적이 대한민국인 응답자들을 대상으로, 출생지에 대한 정보에 근거하여 한국에서 태어났는지, 외국에서 태어났는지를 구분하였으며 대표성을 담보하기 위해 인구가중치를 사용하여 다음의 표를 작성

<표 2.3> 2010년 및 2015년 인구주택총조사의 연령별 외국출생 분포

연령	2010년				2015년			
	국내출생		외국출생		국내출생		외국출생	
	수	%	수	%	수	%	수	%
0	442,269	99.7	1,124	0.3	423,768	99.8	651	0.2
1	421,748	99.5	1,954	0.5	423,325	99.8	822	0.2
2	454,879	99.3	3,177	0.7	435,148	99.8	1,070	0.2
3	457,758	99.4	2,684	0.6	470,030	99.8	947	0.2
4	425,993	99.4	2,554	0.6	465,224	99.7	1,343	0.3
5	421,426	99.4	2,673	0.6	445,997	99.7	1,477	0.3
6	450,868	99.3	3,234	0.7	428,719	99.7	1,361	0.3
7	474,003	99.4	2,633	0.6	460,187	99.5	2,218	0.5
8	487,238	99.6	1,999	0.4	464,403	99.7	1,497	0.3
9	540,188	99.5	2,977	0.5	431,220	99.7	1,127	0.3
10	617,559	99.7	1,996	0.3	419,498	99.6	1,667	0.4

- 2010년 0세인 인구수는 앞선 표의 2009년과 2010년 출생아 수의 평균보다 적고, 2015년 0세인 인구수는 앞선 표의 2014년과 2015년 출생아 수의 평균보다 적어, 잘 알려진 것처럼, 인구주택총조사의 인구수가 과소추정되었음을 시사
- 외국출생 인구수는 5세까지는 증가하는 것으로 나타나는데, 이는 외국에서 태어난 자

녀들이 귀국하는 한편 한국에서 태어난 자녀들이 외국으로 나가는 동시적인 현상을 반영하는 것으로 판단됨

- 2015년 인구주택총조사는 주민등록인구를 활용한 등록센서스였으므로 외국출생이 과소추정되었을 것으로 판단되므로, 이 연구에서는 2010년 자료를 활용하여 인구동향조사에 보고된 출생아 중 0.3%가 외국에서 출생한 것으로 가정함
- 따라서 국내거주 내국인의 출산율을 구할 때 모든 모의 연령별로 0.3%를 제외하고 출생아수를 구함

다. 각세별 사망자 수

(1) 통계청 인구동향조사 사망자수 KOSIS 자료

- 메타정보에 “총 사망자수는 국내 거주 사망자수만을 집계”하며, “지연신고분을 반영한 사망자수, 사망률 보정”을 실시한다고 적시
- 지연신고분은 기준 년도가 10년이 지난 경우 지난 10년 동안 접수한 신고서를 누적 집계하고 10년이 지나지 않은 경우 기준 년도 익년 4월까지 접수된 신고서를 기준으로 집계하는 것을 의미
- 인구동향조사의 각세별 사망자 수 자료는 1983년부터 2020년까지 100+까지 제공하고 있어 이 자료를 이용
- 다음의 표에서 볼 수 있는 것처럼 연령미상이 매우 적은 수를 차지하고 있으므로 분석에서 제외

<표 2.4> 연도별 연령기록과 연령미상 사망자 수

연도	남자		여자	
	0-100+	연령미상	0-100+	연령미상
1983	148,039	62	106,279	183
1984	137,734	65	98,505	141
1985	140,595	76	99,560	187
1986	139,711	62	99,287	196
1987	140,978	50	102,329	147
1988	136,856	56	98,710	157
1989	137,493	35	99,164	126
1990	138,686	44	102,761	125
1991	138,232	38	103,910	90
1992	134,121	110	101,586	345
1993	133,086	13	101,110	48
1994	137,647	36	104,655	101
1995	137,049	10	105,732	47
1996	136,322	6	104,775	46
1997	137,447	9	107,132	105
1998	137,864	15	107,878	68
1999	137,630	7	110,067	30
2000	137,785	4	110,935	16
2001	135,211	7	108,584	11
2002	135,841	5	111,662	16
2003	135,880	5	110,561	17
2004	136,233	2	109,977	8
2005	135,317	0	110,548	9
2006	134,630	1	109,524	7
2007	135,663	1	110,816	2
2008	136,930	2	109,176	5
2009	137,701	34	109,193	14
2010	142,314	44	113,021	26
2011	143,250	0	114,128	18
2012	147,343	29	119,829	20
2013	146,589	10	119,632	26
2014	147,300	21	120,350	21
2015	150,428	21	125,426	20
2016	152,520	9	128,265	33
2017	154,312	16	131,181	25
2018	161,162	25	137,615	18
2019	160,276	46	134,763	25
2020	165,154	9	139,767	18

□ 이렇게 자료를 활용했을 때 나타나는 문제는 2000년 이전 자료 중 90세가 넘어가는 고연령층에서 사망률이 1보다 큰 경우가 나타난다는 것임

- 연앙인구를 구할 때, 80세 이상 인구를 연령별로 배분하면서 생기는 오차와 추계인구의 일정부분을 외국인으로 가정하여 제외하여 분모의 수가 작아지기 때문에 생기는 오류일 것으로 판단
- 이 문제를 해결하기 위해 사망률이 1보다 큰 경우 1로 대체함

라. 내국인 국제인구이동통계

(1) 통계청 국제인구이동통계 KOSIS 자료

- “입국”은 외국에서 대한민국으로 이동해 체류기간 90일이 초과한 경우로 “출국”은 대한민국에서 외국으로 이동해 체류기간 90일이 초과한 경우로 정의
- 내국인 국제인구이동통계는 2000년부터 제공하고 있으며 최상한 연령은 85+임
- 다음의 표에 연도별로 연령을 식별할 수 있는 출입국자와 연령이 알려지지 않은 출입국자 수를 제시하고 있는데, 이를 보면, 연령 미상의 출입국자 수가 미미하여 분석에서 제외한다고 하더라도 큰 문제는 없을 것으로 판단되기 때문에 제외

<표 2.5> 연도별 연령기록과 연령미상 내국인 출입국자 수

	입국				출국				순이동			
	남자		여자		남자		여자		남자		여자	
	0-85+	미상	0-85+	미상	0-85+	미상	0-85+	미상	0-85+	미상	0-85+	미상
2000	103,730	16	94,430	16	135,527	14	138,823	14	-31,797	2	-44,393	2
2001	107,935	10	101,834	10	145,030	7	151,963	7	-37,095	3	-50,129	3
2002	115,107	6	113,006	6	140,679	5	149,263	5	-25,572	1	-36,257	1
2003	117,937	5	117,656	5	140,936	2	151,597	2	-22,999	3	-33,941	3
2004	121,926	2	122,062	2	156,778	4	164,106	4	-34,852	-2	-42,044	-2
2005	139,231	8	137,274	8	178,260	3	181,874	3	-39,029	5	-44,600	5
2006	157,643	15	153,045	15	194,372	5	197,413	5	-36,729	10	-44,368	10
2007	166,385	3	163,579	3	199,595	1	200,984	1	-33,210	2	-37,405	2
2008	179,890	3	176,583	3	196,282	0	197,078	0	-16,392	3	-20,495	3
2009	181,829	2	176,951	2	168,340	1	169,436	1	13,489	1	7,515	1
2010	169,745	0	169,286	0	174,923	0	178,834	0	-5,178	0	-9,548	0
2011	177,275	2	173,948	2	177,036	1	172,830	1	239	1	1,118	1
2012	178,347	0	164,484	0	179,036	1	167,323	1	-689	-1	-2,839	-1
2013	174,412	-	161,281	-	179,079	-	163,789	-	-4,667	-	-2,508	-
2014	175,315	-	152,802	-	167,313	-	155,694	-	8,002	-	-2,892	-
2015	165,467	1	145,313	1	165,159	0	156,121	0	308	1	-10,808	1
2016	164,832	3	146,985	3	160,116	0	153,587	0	4,716	3	-6,602	3
2017	161,913	2	143,534	2	151,468	1	151,134	1	10,445	1	-7,600	1
2018	164,215	1	158,679	1	148,516	1	148,333	1	15,699	0	10,346	0
2019	160,359	1	150,607	1	144,358	1	147,494	1	16,001	0	3,113	0
2020	233,666	-	206,416	-	112,597	-	86,203	-	121,069	-	120,213	-

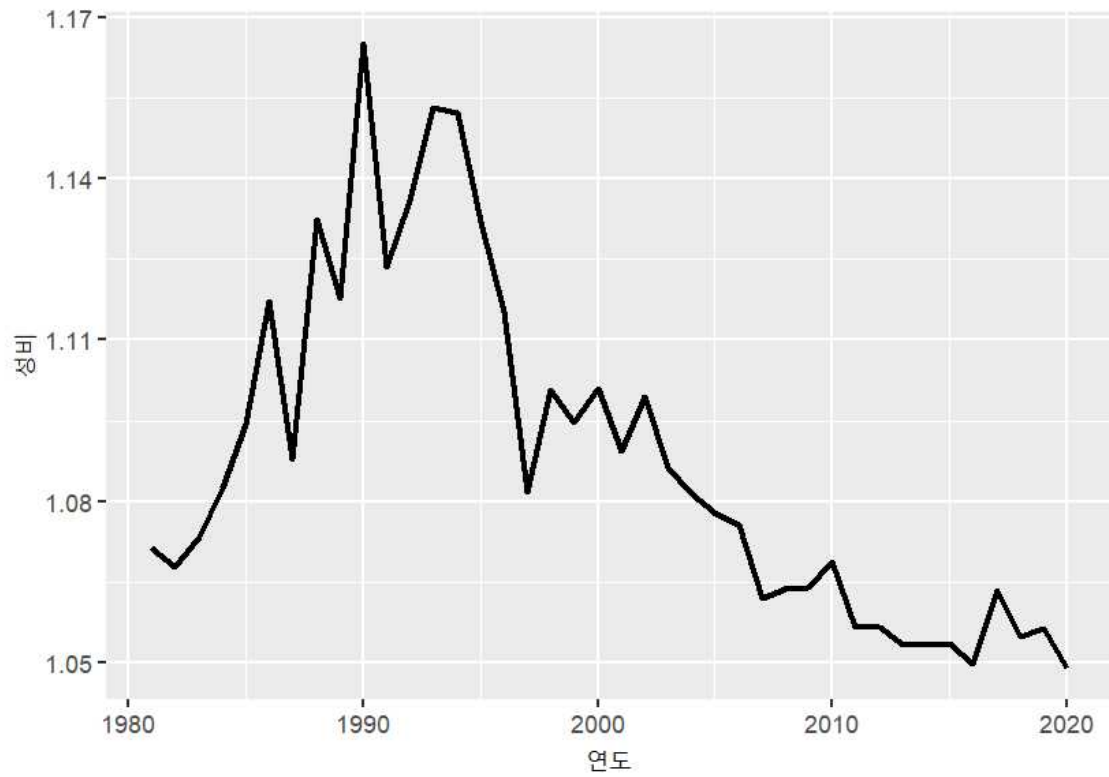
마. 성비

(1) 통계청 인구동향조사 출생아수 KOSIS 자료

□ 다음 그림은 여아 1명당 남아 출생아 수의 변화를 보여줌

□ 최근 통계청 추계에서 최근 3년간의 평균을 사용하였으며, 이 보고서에서도 이를 따라 2018년부터 2020년 출생아 성비(여아 출생 1명당 남아 출생 수)의 평균값 1.05329를 사용

<그림 2.5> 여아 1명당 남아 출생아 수 비



3. 출산율 추계

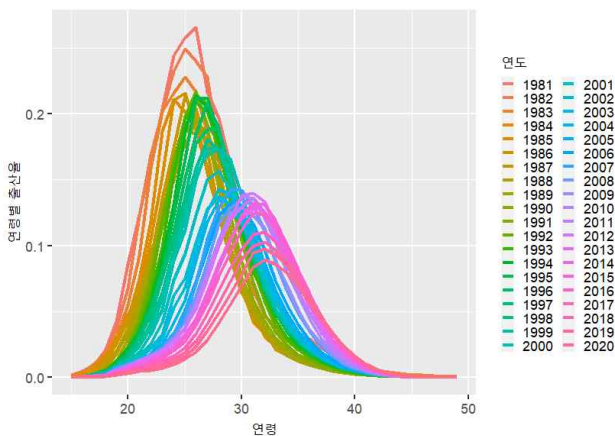
가. 추계 모형

(1) 이 연구에서는 몇몇 모수적 모형을 활용한 출산율 추계를 시도함

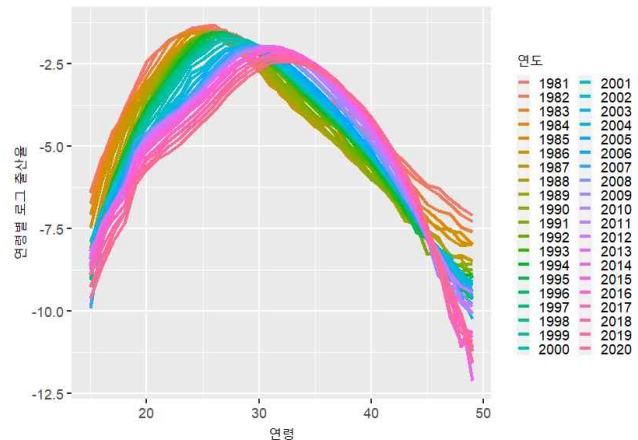
- 모수적 모형은 연령별 출산율이나 연령별 로그 출산율을 특정 함수에 적합시키는 모형이라고 이해할 수 있음
- 모의 연령별 출생아수를 분자로하고, 인구총조사의 내국인 비율을 적용한 주민등록연앙인구의 특정 연령 여성 수를 분모로 하여 구한 1993년부터 2017년까지 각세별 출산율과 각세별 로그 출산율을 <그림 3.1>에 보여주고 있음

<그림 3.1> 연령별 출산율의 변화

1) 연령별 출산율의 변화



2) 연령별 로그 출산율의 변화



□ 로그 이차함수 모형

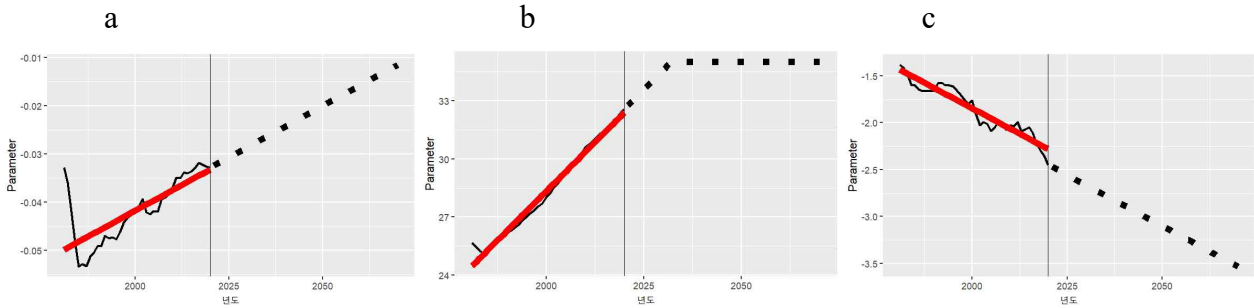
- $\ln(f(x)) = a(x-b)^2 + c$
- $f(x)$ 는 특정 연령 x 세의 출산율을 의미하고, 이 모형에서 a , b , c 는 추정되어야 할 모수들 임
- a 는 2차 함수의 형태를 의미하고, b 는 연령별 출산율이 가장 높은 연령을 의미하며, c 는 연령별 출산율의 수준을 의미함
- R의 nls 함수를 이용하여 위 모형을 직접적으로 추정

○ 추정 결과

<표 3.1> 로그 이차함수 모형의 모수 추정값

년도	a	b	c
1981	-0.03285	25.66277	-1.38179
1982	-0.03605	25.48544	-1.41204
1983	-0.04199	25.25453	-1.48837
1984	-0.04830	25.12031	-1.59282
1985	-0.05338	25.21608	-1.59890
1986	-0.05287	25.42243	-1.64508
1987	-0.05329	25.60721	-1.66472
1988	-0.05136	25.81334	-1.66515
1989	-0.05050	25.96830	-1.66045
1990	-0.04913	26.16547	-1.66367
1991	-0.04914	26.31505	-1.57655
1992	-0.04699	26.46380	-1.57341
1993	-0.04747	26.60249	-1.59865
1994	-0.04730	26.85573	-1.59905
1995	-0.04778	27.00797	-1.61017
1996	-0.04619	27.19505	-1.65876
1997	-0.04423	27.36866	-1.70287
1998	-0.04327	27.54555	-1.75888
1999	-0.04271	27.68886	-1.79384
2000	-0.04172	28.03627	-1.76302
2001	-0.04079	28.22593	-1.90077
2002	-0.03938	28.51100	-2.02622
2003	-0.04204	28.73276	-1.99285
2004	-0.04249	29.03415	-2.01290
2005	-0.04183	29.31426	-2.08940
2006	-0.04197	29.60277	-2.04712
2007	-0.03952	29.80930	-1.96718
2008	-0.03899	30.07196	-2.02939
2009	-0.03839	30.27005	-2.07453
2010	-0.03707	30.58184	-2.02528
2011	-0.03506	30.75617	-2.03399
2012	-0.03497	30.91859	-1.99028
2013	-0.03387	31.16880	-2.09150
2014	-0.03401	31.32658	-2.07260
2015	-0.03361	31.48113	-2.04535
2016	-0.03293	31.65017	-2.10935
2017	-0.03188	31.85144	-2.22772
2018	-0.03215	32.09446	-2.29791
2019	-0.03259	32.35672	-2.35543
2020	-0.03286	32.55865	-2.44781

<그림 3.2> 로그 이차함수 모형의 모수 추정값 및 추계값



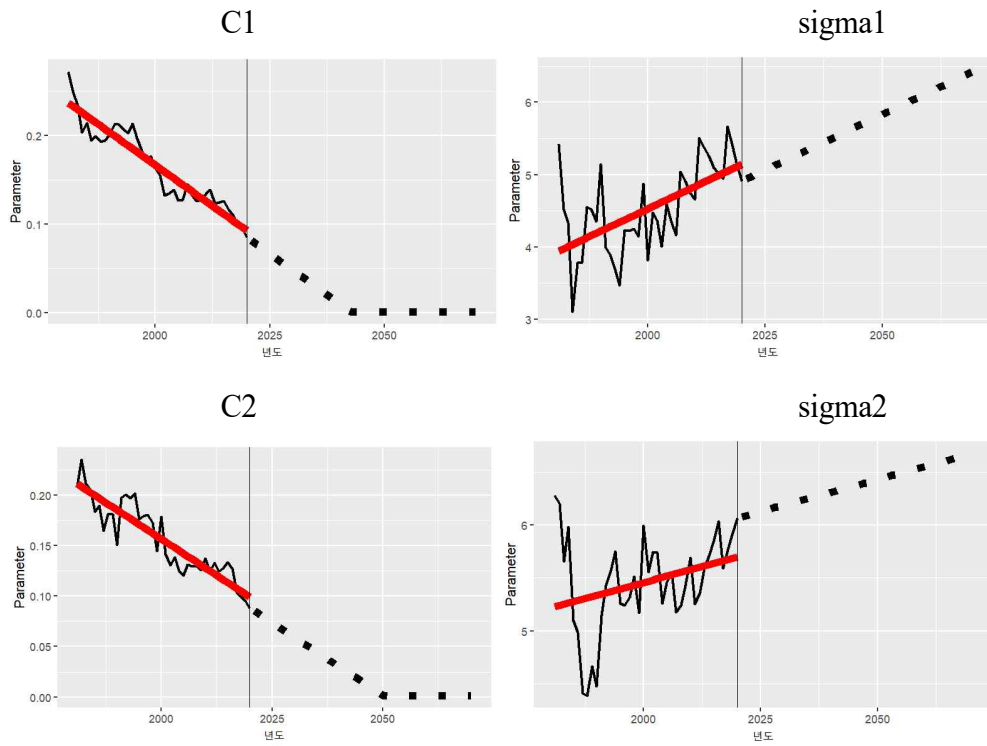
- 2021년 이후의 값들을 추계하는데 있어 2020년과 2021년의 차이가 두드러지게 나타날 수 있는데, 이는 자료 내 회귀선이 자료 전체에 대한 회귀선을 찾고, 이 회귀선의 2021년의 값이 2020년과 연속적으로 나타나지 않기 때문임
- 이는 추계 문헌에서 흔히 jump-off 문제라고 알려져 있으며, 이 보고서에서는 이를 해결하기 위해 2020년의 값을 추계한 후 실제 관측값과의 차이를 구하여 장래 년도의 추계값에서 그 차이를 제하였음
- 다른 모수들의 추계에서도 이런 방식으로 jump-off 문제에 대처하였음
- 또한 출산율이 가장 높은 연령을 뜻하는 b 의 장래값을 추계할 때 40 이상의 연령이 나타나는 경우가 발생하게 되는데 이는 현재로서는 받아드리기 어려운 결과이기 때문에 35세 이상의 경우 모두 35세로 대체하였음

□ Peristera-Kostaki 모형

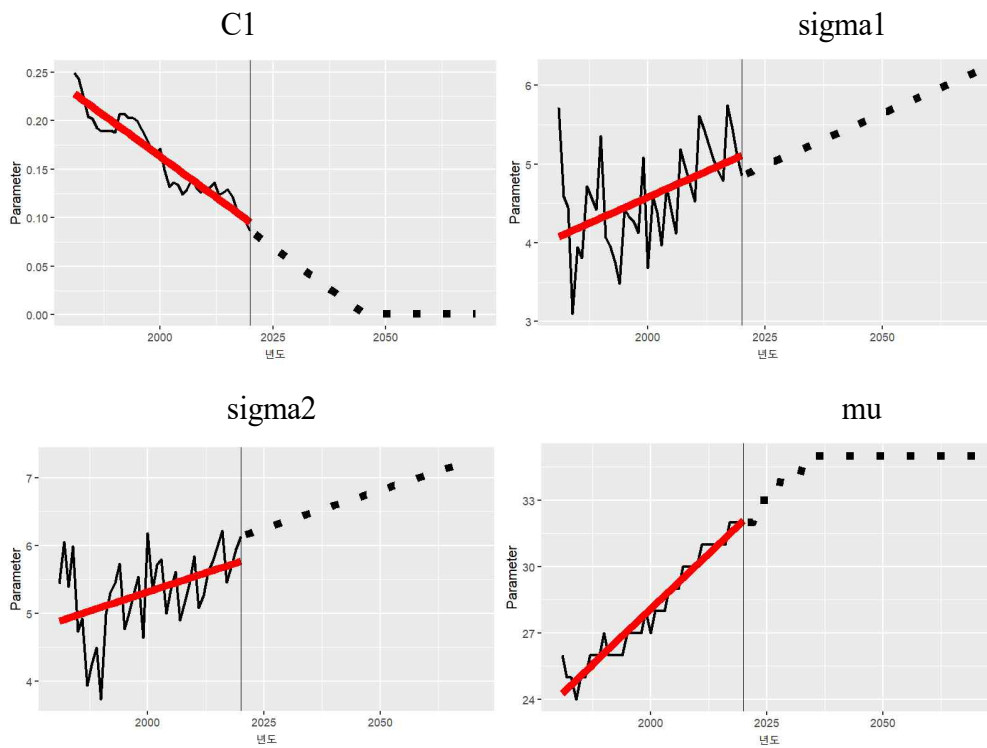
- 이 연구 우선 다음과 같은 Peristera-Kostaki 모형의 변형을 고려함 (Peristera & Kostaki, 2007)
- $$f(x) = c_1 \exp\left(-\left(\frac{x-\mu}{\sigma_1}\right)^2\right)I(x \leq \mu) + c_2 \exp\left(-\left(\frac{x-\mu}{\sigma_2}\right)^2\right)I(x > \mu)$$
- 이 식에서 μ 는 출산율이 가장 높은 연령을 의미하고, c_1 , c_2 , σ_1 과 σ_2 는 추정되어야 할 모수이며, I 는 괄호 안의 조건을 충족하면 1의 값을, 조건을 충족하지 못하면 0의 값을 갖는 조건부 함수임
- 이 모형은 <그림 3.1>에서 볼 수 있는 것처럼 연령별 출산율이 가장 높은 연령을 기준으로 좌우 대칭이 되지 않는다는 점에 착안하여 좌우를 나누지만, 양쪽의 연령별 출산율은 로그 이차함수를 따른다는 관점에 기초하고 있음

- 앞서와 마찬가지로 모든 연도에 있어, μ , c_1 , c_2 , σ_1 과 σ_2 를 구한 후 각 모수들의 경향을 살펴보고 시계열 모형으로 적합시킨 후 향후 값을 추계하고, 특정 연도의 추계값들을 원래 함수에 넣어 연령별 출산율을 추계
- 예산정책처에서는 위 모형의 좌우 함수가 μ 에서 동일한 값을 갖지 않는 문제를 해결하기 위해 다음 모형을 제안
- $$f(x) = c * \exp\left(-\left(\frac{x-\mu}{\sigma_1}\right)^2\right) I(x \leq \mu) + c * \exp\left(-\left(\frac{x-\mu}{\sigma_2}\right)^2\right) I(x > \mu)$$
- 앞선 모형에서는 좌우의 결합점, 즉 출산율이 가장 높은 연령에서의 왼쪽 값과 오른쪽 값이 다르도록 설계되었으나 위 모형에서는 좌우의 결합점이 같도록 설계하는 차이가 있음
- 모수를 추정할 때 μ 의 값이 너무 커지는 것, 예를 들어 40세 정도로 증가하는 것은 현재의 상식으로 맞지 않기 때문에 35를 상한값으로 정함
- 두 모형 모두에서 c 의 값이 음수이면 출산율이 음수이기 때문에 음의 값을 가질 수 없으나 선형으로 추계하면 음의 값을 갖게되는 경우가 있어 0.001보다 작은 값들은 0.001로 대체
- 추정 결과

<그림 3.3> 불연속 PK 모형의 모수 추정값



<그림 3.4> 연속 PK 모형의 모수 추정값



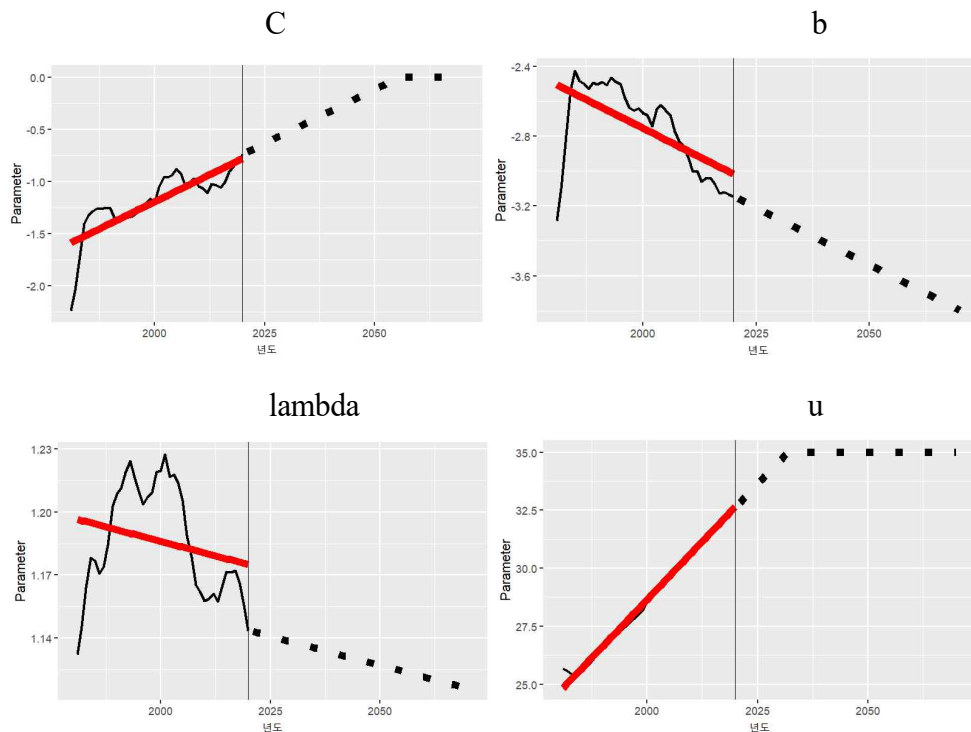
<표 3.2> PK 모형의 모수 추정값

년도	불연속 모형				연속 모형			μ
	c_1	σ_1	c_1	σ_2	c	σ_1	σ_2	
1981	0.27202	5.42091	0.20937	6.28048	0.24930	5.71937	5.44206	26
1982	0.24859	4.52351	0.23548	6.20465	0.24295	4.59053	6.05561	25
1983	0.23432	4.33717	0.21181	5.65716	0.22525	4.44815	5.39038	25
1984	0.20383	3.10532	0.20464	5.98111	0.20426	3.10008	5.98915	24
1985	0.21438	3.78216	0.18386	5.09855	0.20211	3.94901	4.73266	25
1986	0.19477	3.78041	0.18972	4.98641	0.19266	3.81203	4.92752	25
1987	0.19942	4.55522	0.16465	4.41224	0.18909	4.72059	3.93360	26
1988	0.19304	4.51972	0.18096	4.38845	0.18920	4.58003	4.23508	26
1989	0.19465	4.35179	0.18096	4.66411	0.18986	4.42277	4.49532	26
1990	0.20081	5.14170	0.15089	4.47910	0.18770	5.35499	3.73505	27
1991	0.21256	3.99019	0.19786	5.14855	0.20655	4.06782	4.98831	26
1992	0.21256	3.89843	0.20063	5.42980	0.20741	3.96411	5.29752	26
1993	0.20732	3.68998	0.19668	5.57496	0.20250	3.75227	5.45413	26
1994	0.20311	3.47176	0.20174	5.74984	0.20244	3.48066	5.73518	26
1995	0.21335	4.22993	0.17640	5.25849	0.19938	4.43508	4.76609	27
1996	0.19728	4.22757	0.17921	5.24092	0.19016	4.34001	5.00146	27
1997	0.18267	4.25323	0.18068	5.31794	0.18186	4.26687	5.29124	27
1998	0.17116	4.14356	0.17291	5.51186	0.17192	4.13024	5.53531	27
1999	0.17655	4.87833	0.14465	5.17241	0.16568	5.08692	4.63994	28
2000	0.16318	3.81448	0.17863	5.99174	0.17093	3.68383	6.17911	27
2001	0.15477	4.47970	0.14163	5.55563	0.14940	4.58978	5.33818	28
2002	0.13277	4.35904	0.13072	5.74797	0.13186	4.38081	5.71191	28
2003	0.13449	4.00910	0.13838	5.73718	0.13632	3.96805	5.79922	28
2004	0.13891	4.58592	0.12498	5.25867	0.13353	4.71519	5.00282	29
2005	0.12671	4.35905	0.12022	5.45447	0.12396	4.43027	5.33425	29
2006	0.12761	4.16507	0.13132	5.55366	0.12931	4.12214	5.61485	29
2007	0.14544	5.03713	0.12959	5.17835	0.13961	5.18699	4.89975	30
2008	0.13293	4.90832	0.12919	5.23907	0.13146	4.94981	5.17305	30
2009	0.12590	4.75710	0.12553	5.43253	0.12574	4.76176	5.42585	30
2010	0.12756	4.65401	0.13714	5.68939	0.13195	4.52904	5.84263	30
2011	0.13406	5.51211	0.12481	5.25447	0.13068	5.61569	5.08275	31
2012	0.13876	5.37304	0.13320	5.36029	0.13662	5.43526	5.26455	31
2013	0.12304	5.26898	0.12423	5.58514	0.12353	5.25289	5.60684	31
2014	0.12444	5.10148	0.12784	5.71463	0.12591	5.05561	5.77395	31
2015	0.12587	5.02064	0.13364	5.84676	0.12935	4.91666	5.97626	31
2016	0.11646	4.94389	0.12673	6.03781	0.12125	4.79072	6.21735	31
2017	0.11005	5.66551	0.10399	5.59403	0.10771	5.75226	5.45892	32
2018	0.10181	5.39160	0.09868	5.77392	0.10052	5.44194	5.70189	32
2019	0.09453	5.13262	0.09546	5.91748	0.09494	5.11596	5.93933	32
2020	0.08528	4.91072	0.08816	6.06772	0.08661	4.85263	6.13928	32

□ 일반화 로그감마 모형

- $$f(x) = \frac{C|\lambda|}{b\Gamma(\frac{1}{\lambda^2})} (\frac{1}{\lambda^2})^{(\frac{1}{\lambda^2})} \exp(\frac{1}{\lambda}(\frac{x-u}{b}) - \frac{1}{\lambda^2} \exp(\frac{1}{\lambda}(\frac{x-u}{b})))$$
- Γ 는 감마 함수를 의미하고, u 는 평균 출산 연령을 의미하며, C, λ, b 는 추정되어야 할 모수임
- 이 모형은 로그를 취한 출산율에 음의 부호를 붙여 다시 한번 로그를 취한 후 평균 출산 연령을 기준으로 좌우대칭이 되는 선형 직선으로 표현하는 것으로 이해할 수 있음
- 앞서와 마찬가지로 모든 연도에 있어, u, C, λ, b 를 구한 후 각 모수들의 경향을 살펴보고 시계열 모형으로 적합시킨 후 향후 값을 추계하고, 특정 연도의 추계 값들을 원래 함수에 넣어 연령별 출산율을 추계
- 위의 식에서 gamma 함수는 양의 값을 갖고, 출산율이 양의 값을 갖기 위해서는, C 와 b 는 음의 값을 가져야 하는데, 추계 결과 b 의 값이 양수로 전환하게 되어 -0.001 보다 큰 값은 모두 -0.001 로 대체하였음
- 또한 u 의 최댓값을 35로 정함
- 추정 결과

<그림 3.5> 일반화로그감마 모형의 모수 추정값



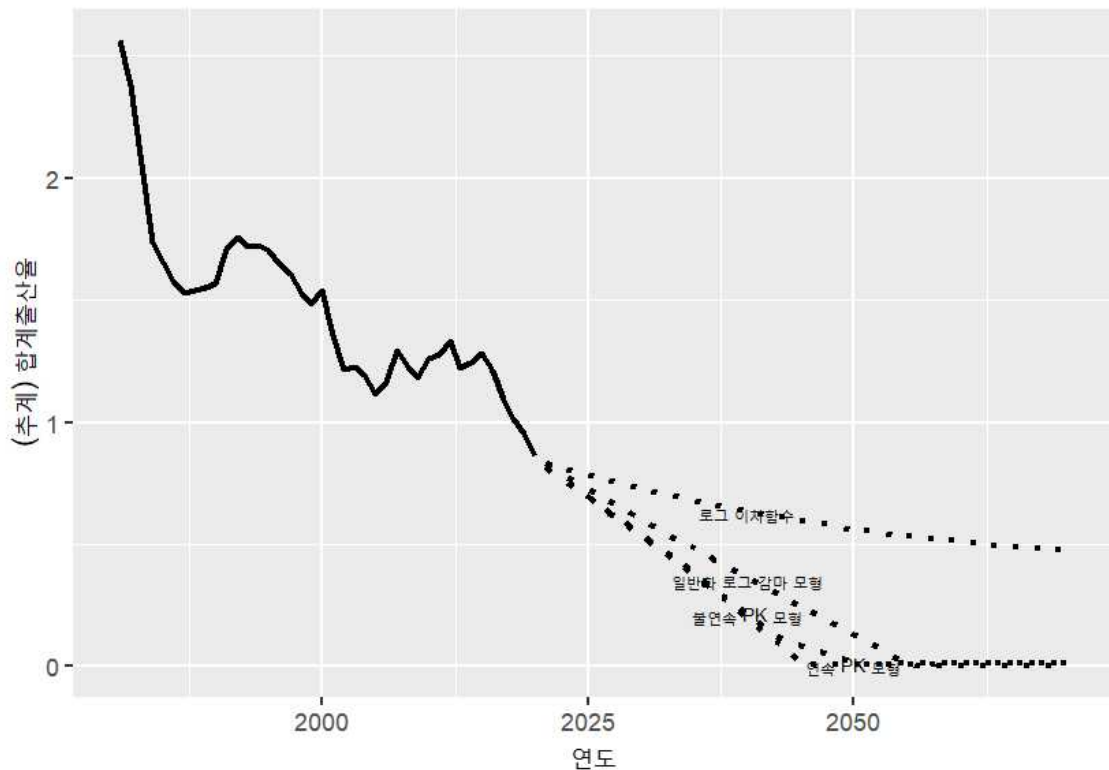
<표 3.3> 일반화 로그감마 모형의 모수 추정값

년도	C	λ	b	u
1981	-2.24028	1.13232	-3.28485	25.67401
1982	-2.04484	1.14546	-3.09864	25.57019
1983	-1.71095	1.16393	-2.79628	25.43569
1984	-1.40291	1.17833	-2.53687	25.37804
1985	-1.32732	1.17679	-2.42554	25.49118
1986	-1.28662	1.17074	-2.48396	25.68379
1987	-1.25739	1.17401	-2.50086	25.86841
1988	-1.26652	1.18423	-2.53071	26.10669
1989	-1.25425	1.20267	-2.49441	26.33607
1990	-1.25802	1.20847	-2.50657	26.56243
1991	-1.36597	1.21140	-2.49041	26.72151
1992	-1.38590	1.21881	-2.50790	26.91536
1993	-1.33312	1.22431	-2.46694	27.09012
1994	-1.34591	1.21650	-2.48594	27.32087
1995	-1.33622	1.20965	-2.50004	27.45760
1996	-1.30708	1.20381	-2.57980	27.62547
1997	-1.27475	1.20688	-2.63695	27.81878
1998	-1.21298	1.20945	-2.65294	27.99912
1999	-1.16402	1.21913	-2.63839	28.19102
2000	-1.21392	1.21947	-2.66972	28.52889
2001	-1.05944	1.22746	-2.67886	28.76776
2002	-0.96097	1.21680	-2.74507	29.00501
2003	-0.95925	1.21777	-2.64612	29.22630
2004	-0.93673	1.21345	-2.62334	29.49749
2005	-0.88056	1.20534	-2.64941	29.73590
2006	-0.93332	1.18883	-2.68007	29.94048
2007	-1.05070	1.17866	-2.77127	30.09436
2008	-1.00840	1.16528	-2.82750	30.29582
2009	-0.97431	1.16216	-2.85457	30.47484
2010	-1.04710	1.15767	-2.92409	30.76455
2011	-1.06544	1.15859	-3.00339	30.94164
2012	-1.11106	1.16108	-3.00060	31.11999
2013	-1.02424	1.15735	-3.06076	31.34381
2014	-1.03444	1.16424	-3.03991	31.54173
2015	-1.06145	1.17156	-3.04394	31.73368
2016	-1.00681	1.17117	-3.07960	31.90238
2017	-0.90870	1.17218	-3.12914	32.10127
2018	-0.84846	1.16606	-3.12190	32.30385
2019	-0.80549	1.15582	-3.13423	32.50733
2020	-0.74056	1.14367	-3.14384	32.62936

나. 추계 결과

□ 다음 그림은 앞서 제시한 세 모형을 통해 추계한 합계출산율 변화를 제시하고 있음

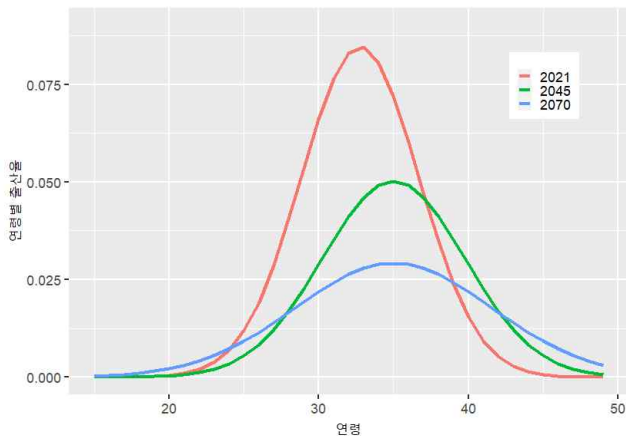
<그림 3.6> 출산율 추계 결과: 합계출산율



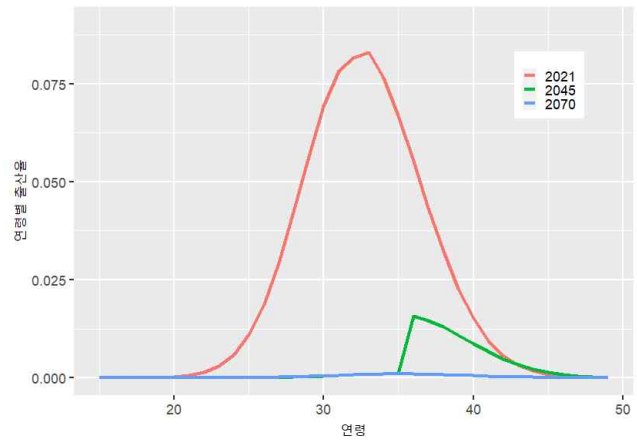
- 모든 방법에서 출산율이 우하향하는 결과를 보여주고 있으나 하락의 정도는 추계방법에 따라 차이가 적지 않게 나타남
 - PK 모형들에서 하락의 속도가 가장 빠르며, 일반화로그감마 모형도 빠른 속도의 하락을 보여주고 있으며, 로그 이차함수 모형도 2063년 합계출산율이 0.5이하로 떨어지고, 2070년에는 0.476인 것으로 나타남
 - 불연속 PK 모형에서는 2045년, 연속 PK 모형에서는 2043년, 일반화로그감마 모형에서는 2052년 이후 합계출산율이 0.1 밑으로 떨어지는 것으로 나타남
- 다음 그림에는 각 추계방법에 따라 달리 나타나는 2021, 2045, 2070년의 연령별 합계출산율의 변화를 보여주며 이들은 모두 설명 변수를 고려하지 않은 결과임

<그림 3.7> 연령별 출산율의 추계 결과

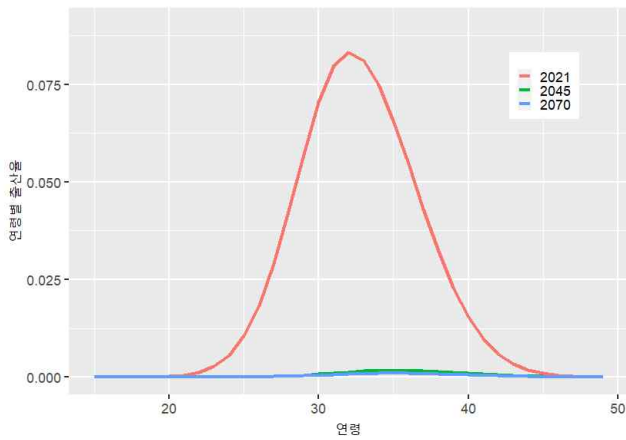
1) 로그 이차함수 모형



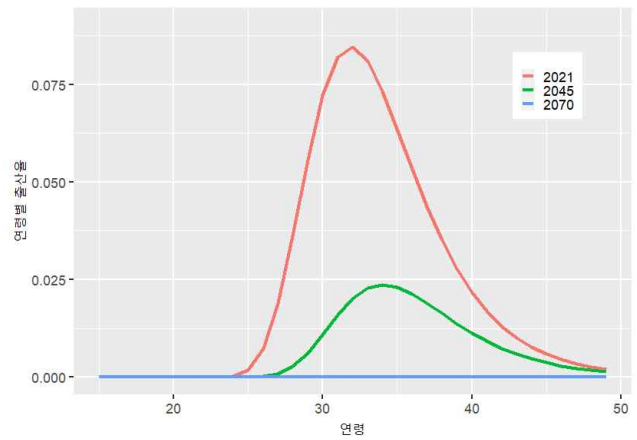
2) 불연속 PK 모형



3) 연속 PK 모형



4) 일반화 로그감마 모형



4. 사망률 추계

가. 문헌 검토

(1) 연령별 사망률이 시간이 지남에 따라 선형으로 하락한다는 것은 매우 오래된 관찰이며(권태환·김두섭, 2011; 김현식, 2013; Preston, Heuveline, & Guillot, 2001), 이러한 관찰에 기반하여 장래 사망률 추계에 적용한 대표적인 모형이 Lee-Carter 모형이며, 향후 모형들은 이 모형을 다양한 방향으로 확장한 것으로 볼 수 있음(Lee & Carter, 1992)

□ 두 가지 대표적인 확장이 있었는데, 하나는 하위집단의 이질적인 사망률 하락 속도를 고려한 Li-Lee 모형이고, 다른 하나는 이질적인 연령별 사망률의 하락 속도를 고려한 Li-Lee-Gerland 모형임(Li & Lee, 2005; Li, Lee, & Gerland, 2013)

□ Lee-Carter 모형

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

- $m(x,t)$ 는 t 시기 연령 x 의 사망률을 나타내고, $a(x)$ 는 모든 시기 연령별 로그 사망률의 평균을 나타내며, $b(x)$ 는 시간에 따라 변하는 로그 사망률의 속도 $k(t)$ 에 반응하는 연령별 로그 사망률의 민감도를 나타내며, $\epsilon(x,t)$ 는 이들로 설명되지 않는 잔차를 나타냄

□ Li-Lee 모형

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

- Li-Lee 모형은 Lee-Carter 모형에 비해 집단을 뜻하는 i 가 들어 있으며, 전체의 변화를 나타내는 $B(x)K(t)$ 와 각 집단의 변화를 나타내는 $b(x,i)k(t,i)$ 가 같이 들어 있다는 특징을 지니고 있음

- $B(x)K(t)$ 를 통해 전체 집단의 사망률 변화를 구분하고, 나머지 변화에서 집단별 변화를 모형화하는 전략으로 이해

□ Li-Lee-Gerland 모형

- $\ln(m(x,t,i)) = a(x,i) + B(x,t)K(t,i) + \epsilon(x,t,i)$

- 원래의 Lee-Carter 모형을 변형하여, 변화 속도에 대한 연령별 로그 사망률의 민감도를 포착하는 $b(x)$ 에 t 를 삽입하여 시간에 따른 사망률 변화 속도를 제어하고자 하는 방법임

- Lee-Carter 모형에서는 $k(t)$ 의 값이 사망률 하락 속도를 결정하고 $b(x)$ 가 하락 속도에 대한 민감도를 나타내기 때문에, 0세의 사망률 하락 속도는 매우 빨랐으나 근래로 올수록 줄어드는데 반해 노년의 사망률 하락 속도는 느렸으나 최근에 빨라진다는 관찰을 모형화할 수 없는 문제가 생김
- 따라서 0세의 사망률 하락을 과도하게 추정할 수 있고, 특정 연령대에서의 사망률 하락이 다른 인접한 연령대의 하락보다 과도하게 높게 되면 최근의 변화와 일치하지 않는 문제들이 생김
- 이러한 이론적인 문제의식에 기초하여 Li-Lee-Gerland 모형은 궁극적으로(ultimately) 특정 연령까지 $b(x)$ 의 값이 같아지고 그 이후부터 하락할 것으로 가정하여, 원래의 Lee-Carter 모형에서 나오는 $b(x)$ 의 값이 궁극적인 $b(x)$ 값으로 수렴할 것으로 가정하며, 그 과정이 매년 조금씩 이루어지므로 $b(x)$ 에 시간을 뜻하는 t 첨자가 포함되었으며, 이런 의미에서 $B(x,t)$ 가 소문자가 아닌 대문자로 표현되었으며, 이 과정을 $b(x)$ 의 rotation이라고 표현함
- Li, Lee, & Gerland(2013)는 처음의 Lee-Carter 모형에서 추정되는 $b(x)$ 가 궁극적인 $B(x,t)$ 로 변화하는 과정을 $B(x,t) = (1 - w_s(t))b_0(x) + w_s(t)b_u(x)$ 로 공식화하였으며, 여기서 원래 Lee-Carter 모형의 $b_0(x)$ 와 사망률 감소율의 궁극적 연령별 패턴 $b_u(x)$ 의 가중 평균으로 표현됨
- 이러한 rotation은 Lee-Carter 모형으로 추계된 사망률에 기반하여 계산한 기대수명이 80세가 될 때부터 시작되며 궁극적 연령별 패턴인 $b_u(x)$ 는 기대수명이 102세가 될 때 도달하는 것으로 가정
- 또한 그 둘의 가중치는 $w_s(t) = [0.5[1 + \sin[\frac{1}{2}(2w(t) - 1)]]]^p$ 이고 이때,

$$w(t) = \frac{e_0(t) - 80}{e_0^u - 80}$$
 이며 $p = 0.5$ 의 값을 제안
- 또 하나 눈여겨보아야 할 것은 $B(x,t)$ 에 성별을 뜻하는 i 가 포함되지 않았다는 것인데, 성별에 따른 변화의 민감도에 차이가 없도록 설정하였다는 것을 알 수 있음
- 다만 변화의 속도를 나타내는 $K(t,i)$ 에 i 를 넣음으로써 성별에 따른 변화 속도의 차이를 고려하였으며, $B(x,t)$ 의 도입으로 Lee-Carter 모형에서 얻은 성별 $k(t)$ 값을 변화시킬 필요가 생김

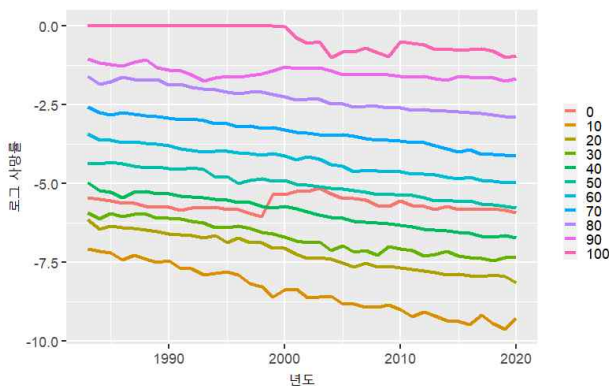
- 이를 위해, 연도별 $B(x,t)$ 를 구한 다음 Lee-Carter 모형에서 얻은 기대수명과 동일하도록 $K(t,i)$ 을 적합
- 이 연구에서는 $K(t,i)$ 값을 얻기 위해 통계 프로그램 R에서 사용가능한 [uniroot] 함수를 활용

나. 자료 분석

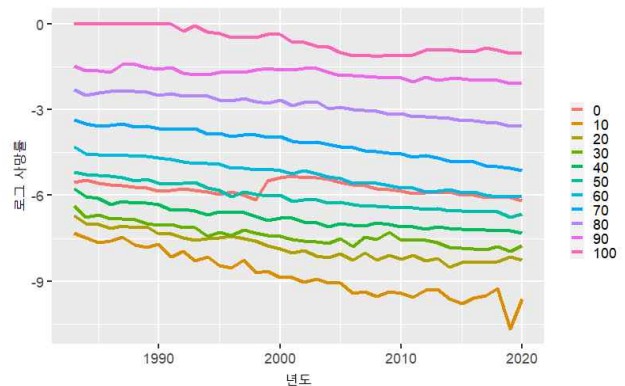
- 이 보고서에서는 이러한 모형을 적용하기 위해 우선 1993년부터 2017년 사이의 각세별 사망률을 그림으로 제시함

<그림 4.1> 성별 10세 연령별 로그 사망률의 변화

1) 남자



2) 여자

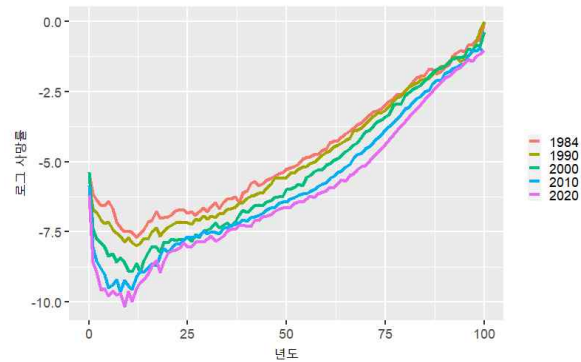
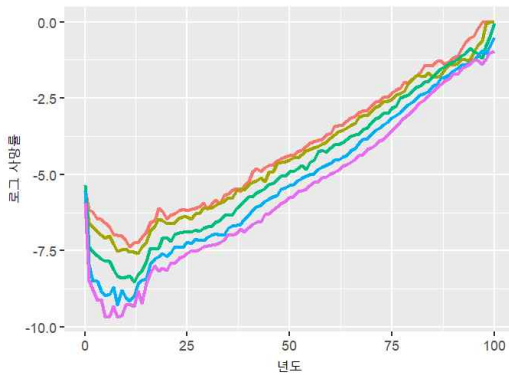


- 성별 및 각세별 로그 사망률의 변화를 그림으로 나타낸 <그림 4.1>을 보면 ① 남성이 나 여성이나 연령별 로그 사망률이 상당히 선형적으로 변화해왔으나, ② y-축 값의 차이가 보여주는 것처럼 성별에 따라 변화의 정도는 약간 달라 여성의 사망률 변화가 더 큰 것으로 보이고, ③ 연령별 기울기도 약간의 차이가 있어 10세에서 기울기가 가장 가파르나 80세 이상의 기울기가 가장 완만하며, ④ 아마도 자료의 연도 기간이 길지 않기 때문에 Li-Lee-Gerland의 논문이 중심으로 다루었던 특정 연령들에서 변화 속도의 변화는 그해 두드러지지 않는 것으로 판단됨

<그림 4.2> 성별 연령별 로그 사망률의 몇몇 연도별 변화

1) 남자

2) 여자

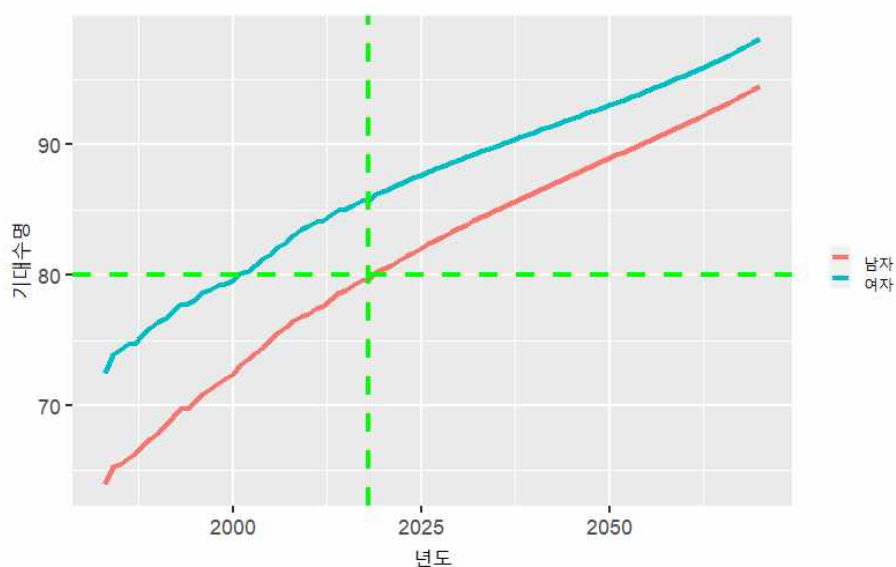


- 위의 그림은 몇몇 연도에서 각세별 사망률의 변화를 보여주고 있음
- 이 그림에서 매우 흥미로운 발견 중 하나는 한국의 경우 사망률이 10-15세에서 가장 낮다는 것인데, 이는 Li-Lee-Gerland 모형에서 가장 낮다고 가정되는 연령대인 15-19세와는 큰 차이를 보임

다. 추계 과정 및 결과

- Li-Lee-Gerland 모형을 활용하여 추계하고자 Lee-Carter 모형으로 성별 사망률을 추계하여 기대수명의 변화를 살펴봄

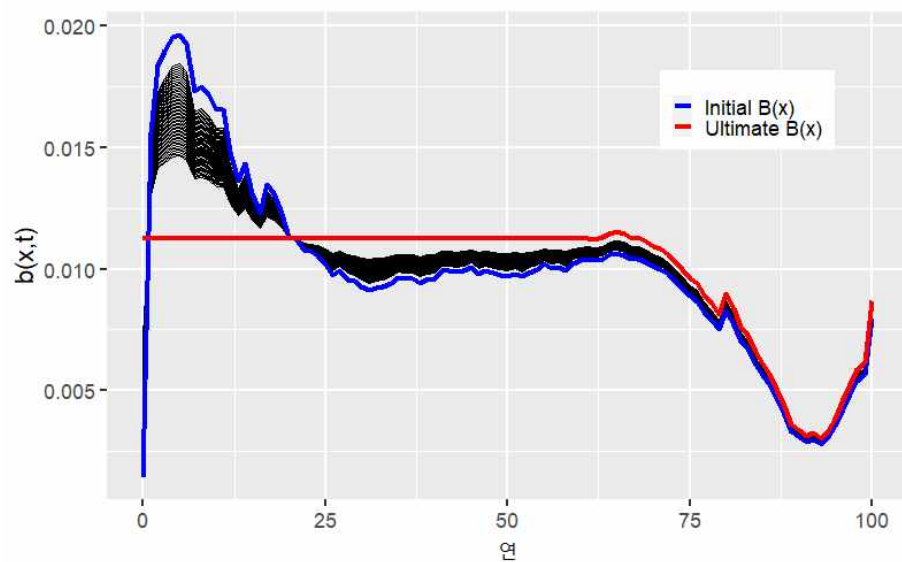
<그림 4.3> 성별 관찰 기대수명 및 추계 기대수명



주. 2020년 이후는 추계 결과

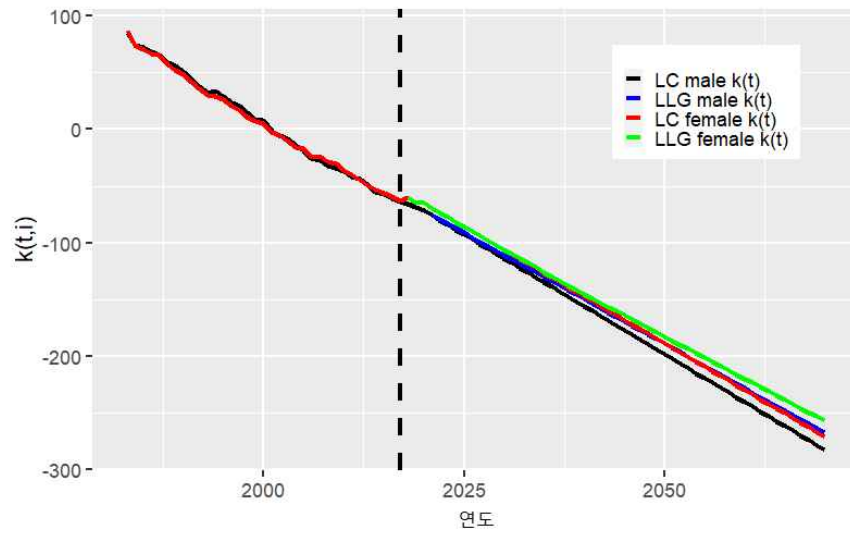
- 기대수명의 변화를 살펴보면 추계가 시작되는 2020년 80세가 넘어 Li-Lee-Gerland에 따르면 rotation이 시작되어야 함
- LLG의 모형은 궁극적 연령 패턴 $B(x,t)$ 에 도달하는 기대수명이 102세라고 설정하였는데, 위 그림의 자료를 보면 추계 기간 내 102세를 넘어서는 때가 없어 이런 가정을 유지
- 아래 그림은 $B(x,t)$ 의 값을 보여줌

<그림 4.4> 추정 $B(x,t)$



- 아래 그림은 $K(t,i)$ 의 값을 보여주며 다음 쪽에 몇몇 연령과 연도의 사망률 추계 결과를 제시함

<그림 4.5> 추정 $K(t,i)$: LC와 LLG $K(t,i)$ 의 비교

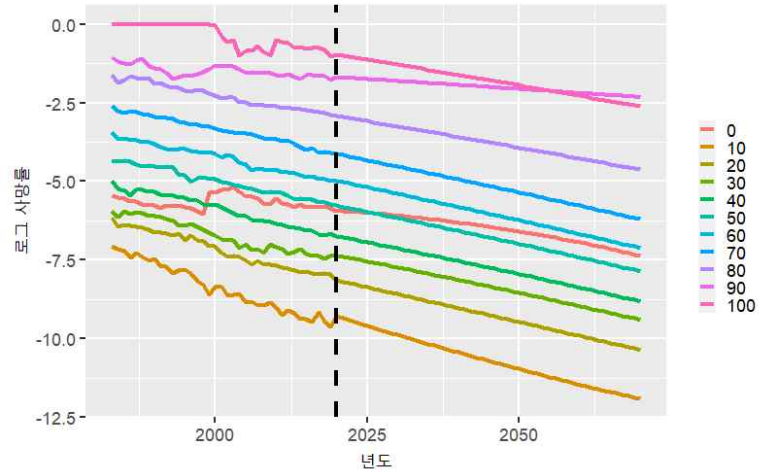


□ 다음 그림은 추계 결과의 연령별 로그사망률을 보여줌

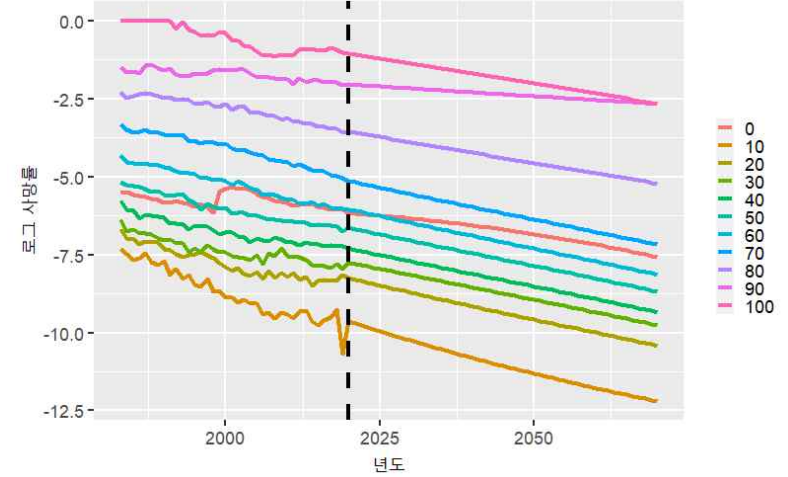
<그림 4.6> 성별 연령별 로그 사망률의 변화

10세
연령
별

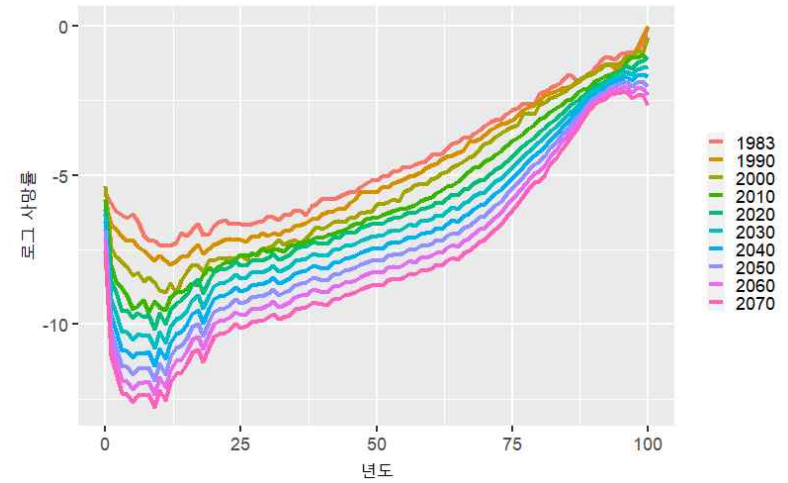
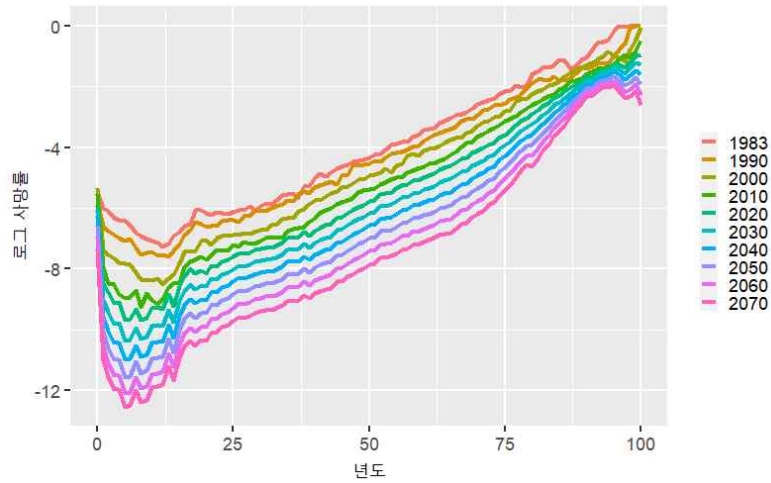
1) 남자



2) 여자



10년
별



5. 국제이동 추계

가. 문헌 검토

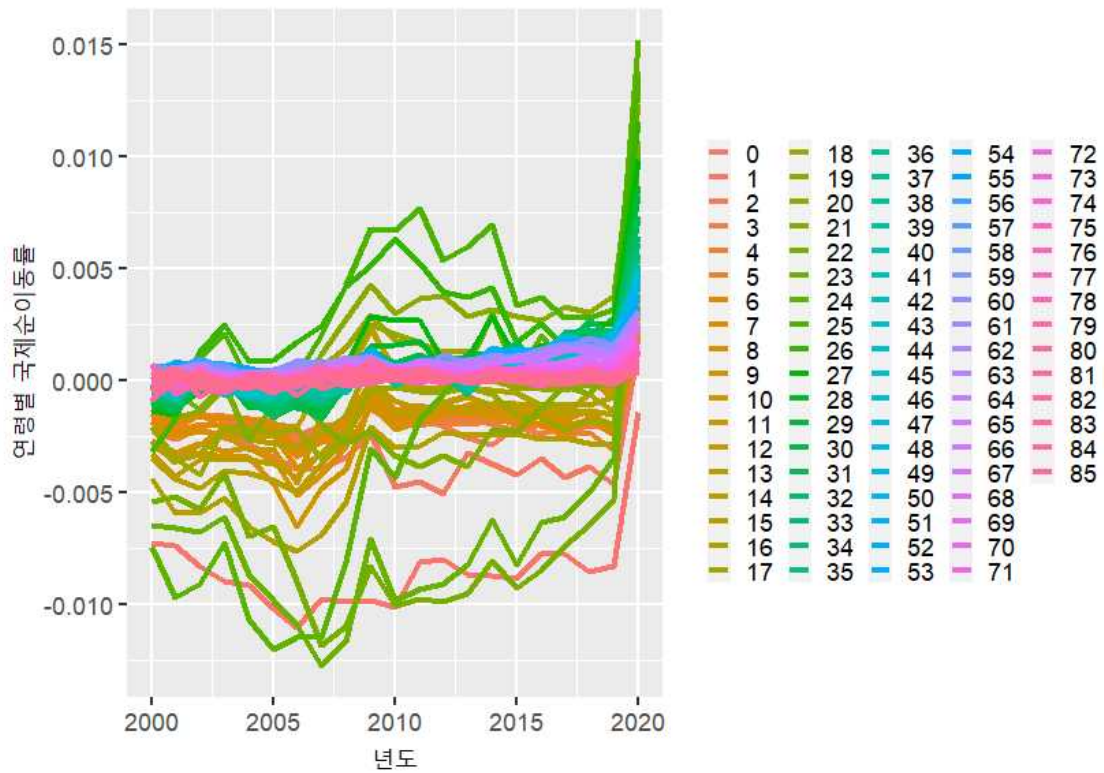
- (1) 국제 이동은 정치적, 경제적, 사회적 변화에 매우 민감하게 반응하는 부분으로 특정한 수리적 모형을 적용하는 것이 쉽지 않음
- (2) 2021년 통계청의 인구추계에서 사용한 방법을 요약하면 다음과 같음(통계청, 2021)
 - 내국인의 국제이동은 최근 10년간(2010-2019년) 국제순이동 최대 및 최저연도를 제외한 후 8개년 평균 성 및 연령별 국제순이동률을 사용
 - 외국인의 경우 체류자격별(정주, 교체순환, 단기) 이동 특성을 반영하기 위해 외국인 입국자와 출국자 수를 체류 유형으로 구분하여 추계

나. 자료 분석

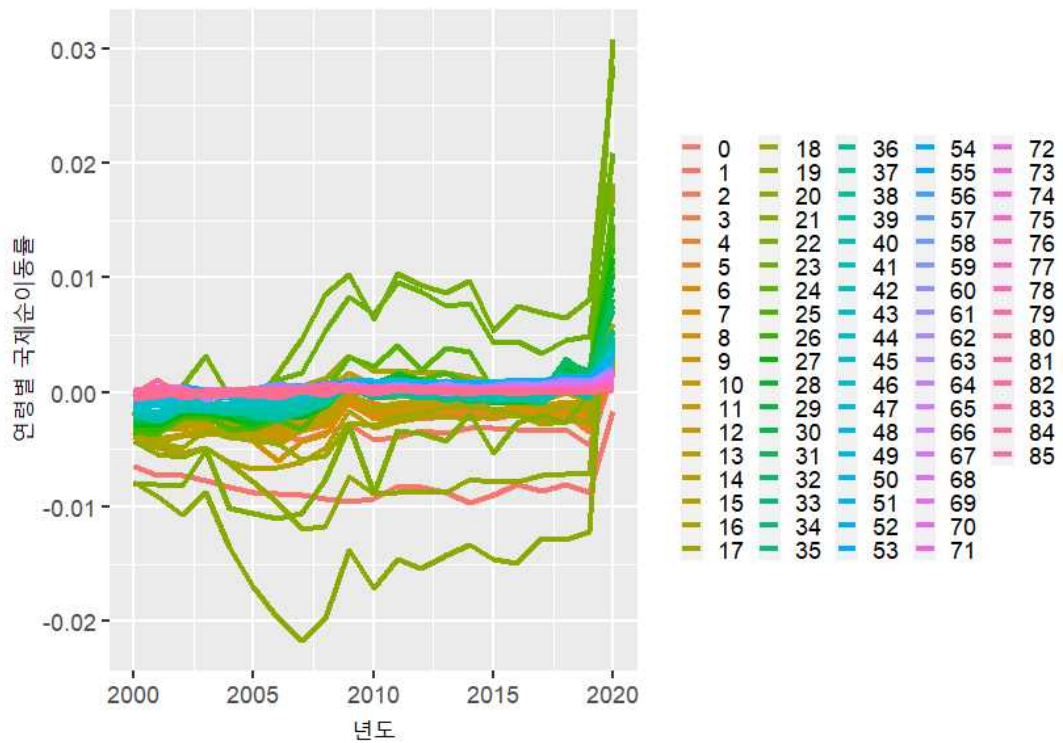
- <그림 5.1>은 연도별 내국인의 연령별 국제순이동을 보여주고 있음

<그림 5.1> 연령별 내국인 국제순이동

1) 남자



2) 여자



- 위의 그림에서 COVID-19 팬데믹으로 인해 2020년의 순국제이동수가 급격히 증가한 것을 볼 수 있는데, 이는 방역 문제로 해외 거주 내국인들이 입국한 것으로 보임
- 내국인의 흐름에 있어 국제순이동이 음의 값을 가지지만, 그리 많지 않은 이동이라고 판단되며, 연도에 따라 뚜렷한 특징들을 찾아내기 어려움

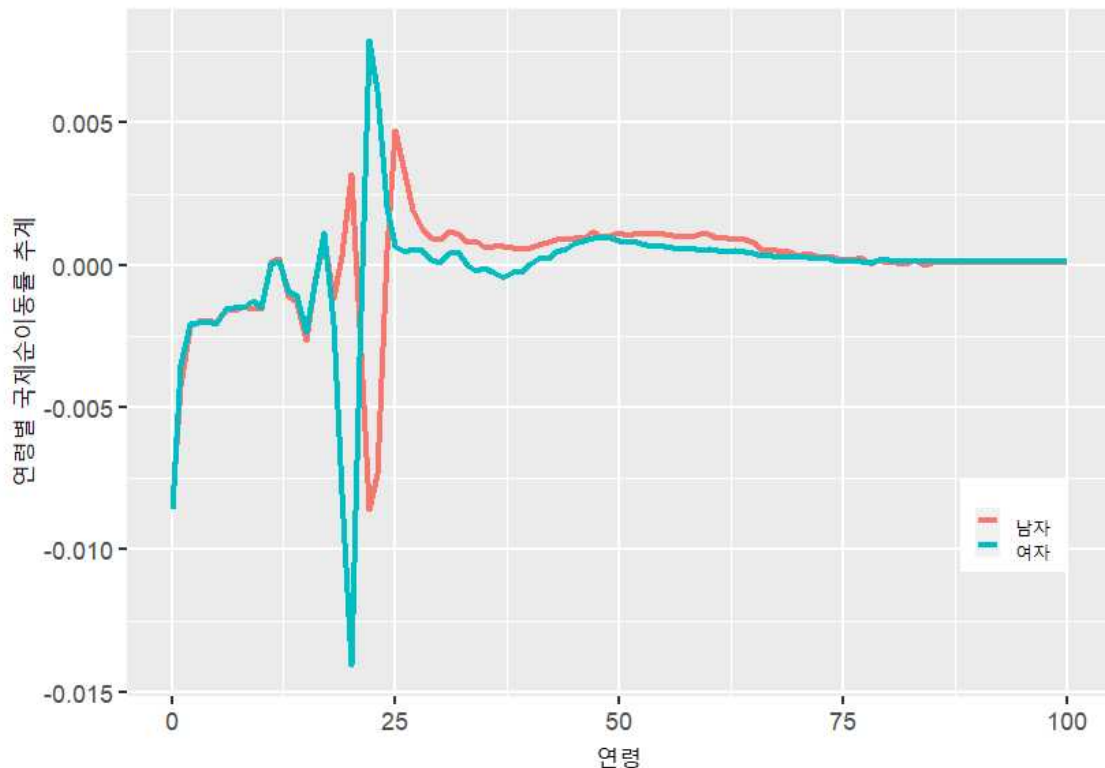
다. 추계 모형

- 이 연구에서도 통계청이 활용한 방법을 활용하여 2010년부터 2019년까지 국제인구이동율을 구한 다음 이 중 최댓값과 최솟값은 빼고 나머지 8개년도의 평균을 구하여 국제인구이동율을 구함

라. 추계 결과

- 다음 그림은 내국인 국제순이동률 추계값을 보여주고 있음

<그림 5.2> 내국인 국제순이동률 추계 값



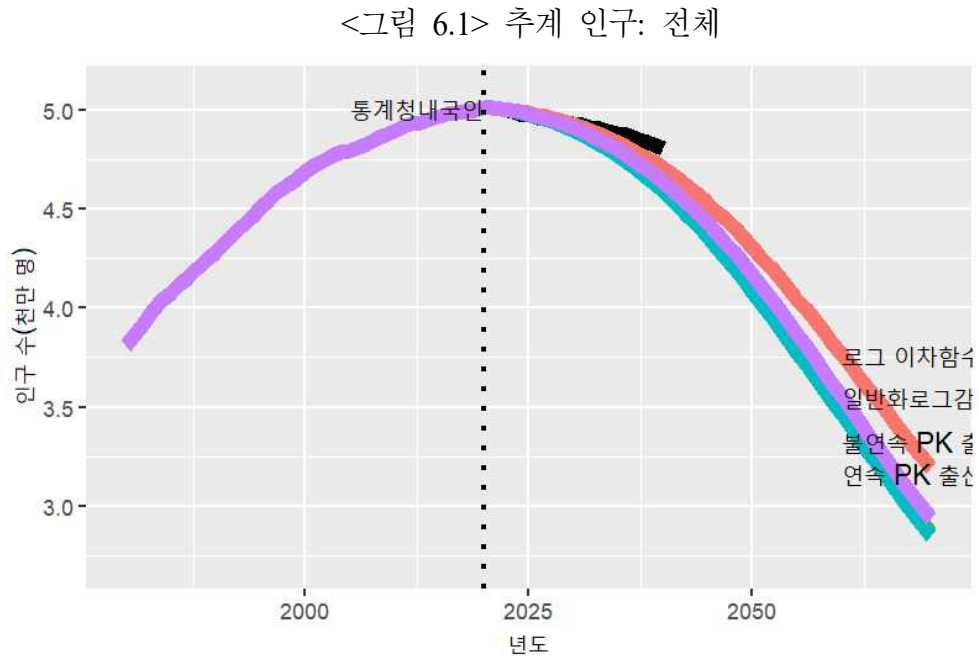
- 내국인 국제이동률을 ‘율’로 추계하는 데 있어 문제가 발생하게 되는데, 앞서 코호트

요인법에서 특정 년도의 인구수를 구하는 식 자체에 특정 년도의 국제이동자 수가 들어 있으나, 율이 주어졌을 때는 인구수를 통해 국제이동자 수는 추정하게 됨

- 다시 말해, 국제이동자 수를 구하기 위해서는 인구수가 있어야 하고 인구수가 있기 위해서는 국제이동자 수가 있어야 하는 순환론적 문제가 발생
- 이러한 문제를 해결하기 위해 여기서는 특정 해의 국제이동률을 이전 해의 인구수에 적용하여 국제이동자 수를 산출

6. 추계 결과

(1) 전체 인구

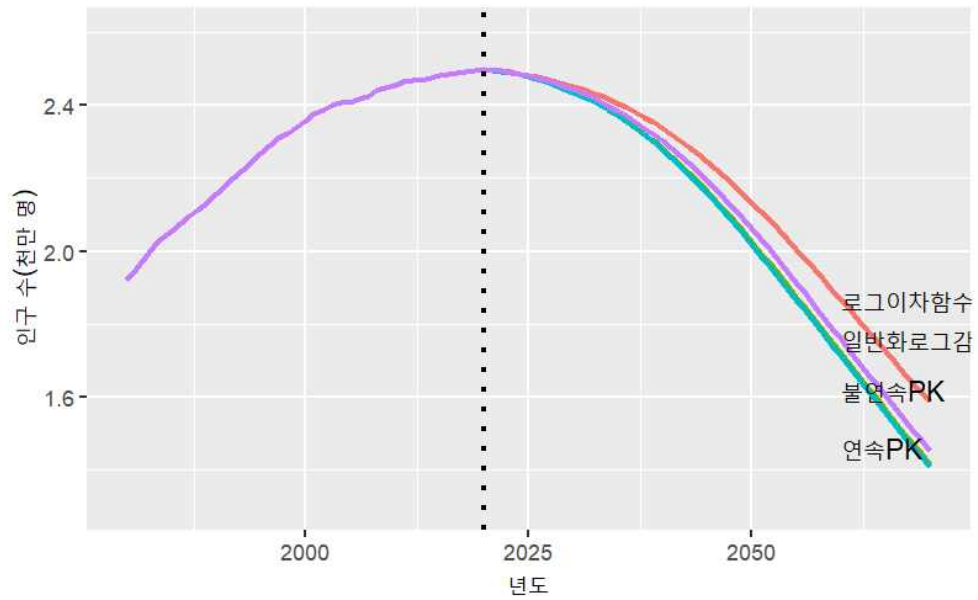


주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

- 위 그림은 모수들을 선형으로 추계한 추계 결과를 보여주고 있음
- 통계청의 내국인 추계는 검은색으로 표시하였으며, 추계가 시작되는 시점에는 거의 동일하지만 시간이 지나면서 통계청의 추계가 더 많은 인구를 보여줌
- 아마도 이는 통계청의 추계에서 출산율이 더 높은 것으로 가정하기 때문으로 보여짐
- 이 보고서에서 실시한 추계의 변화는 출산율의 변화에 따른 것으로 그리 커다란 차이를 보이는 것 같지는 않음

(2) 남자 인구

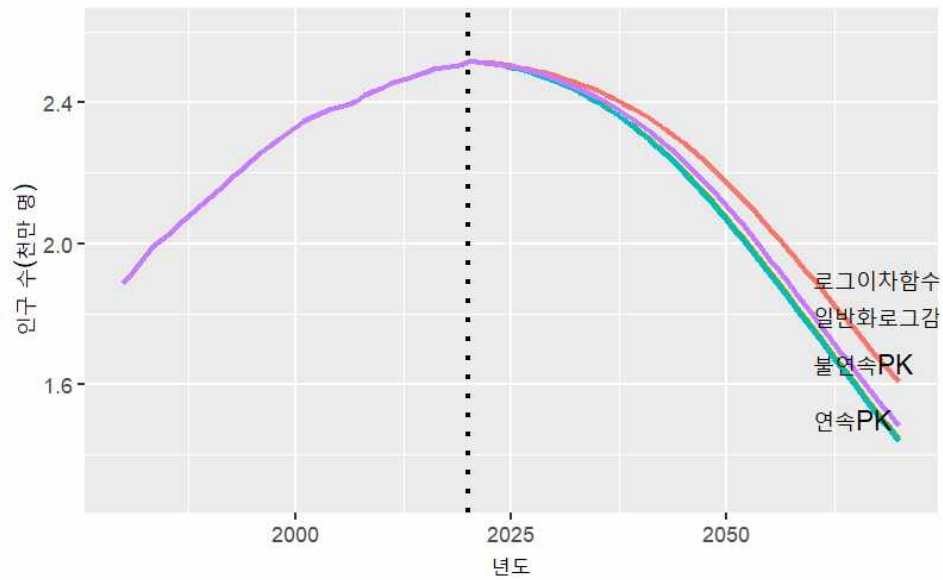
<그림 6.2> 추계 인구: 남자



주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

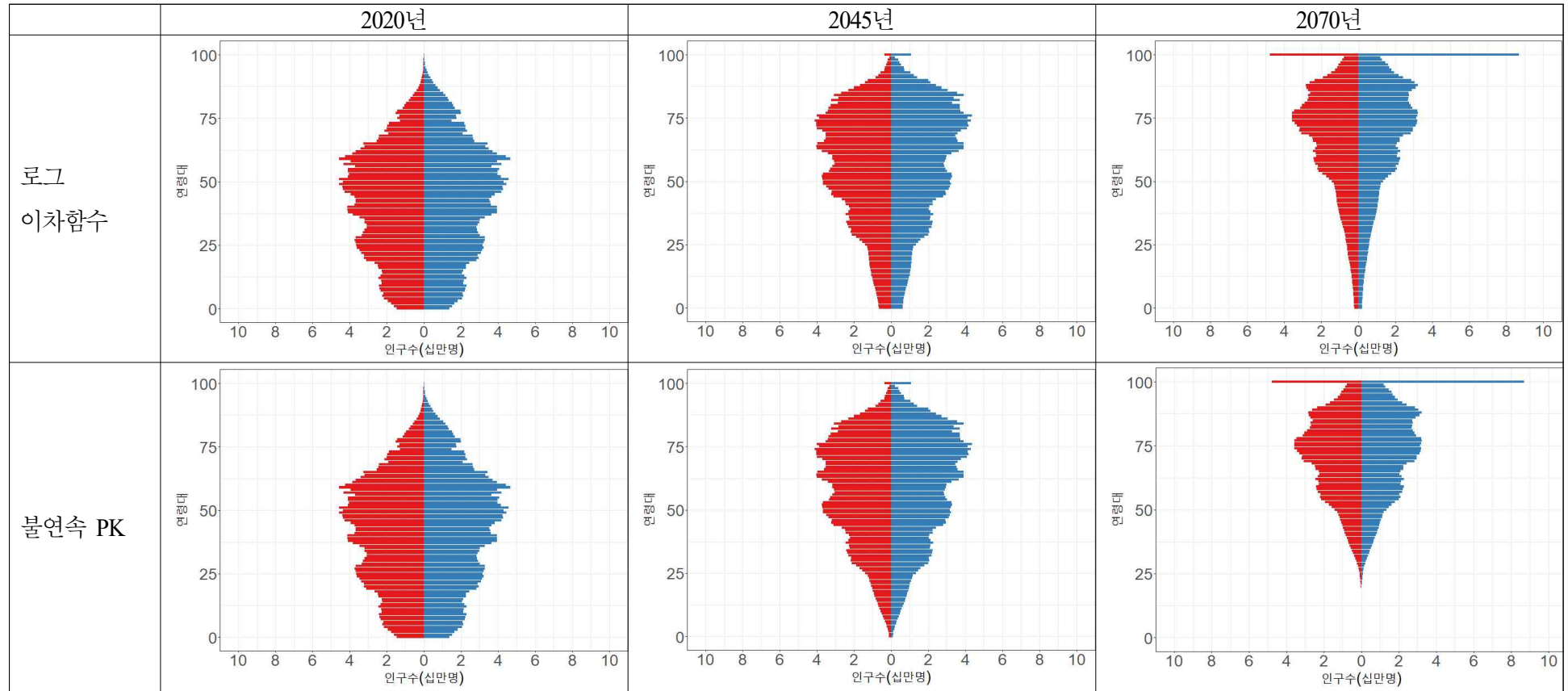
(3) 여자 인구

<그림 6.3> 추계 인구: 여자



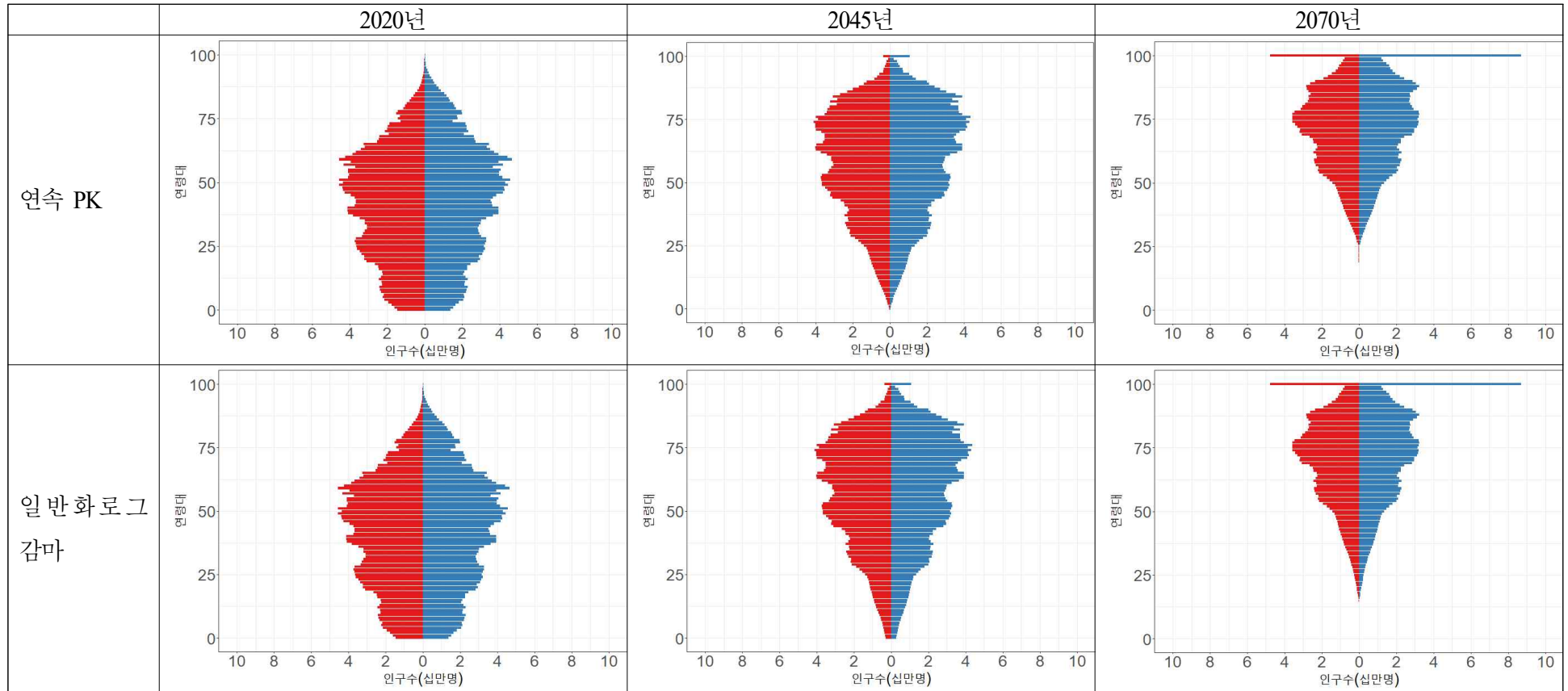
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 6.4> 출산을 가정에 따른 추계 인구 피라미드



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

<그림 6.4> 출산을 가정에 따른 추계 인구 피라미드(계속)



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

7. 경제사회적 변수를 활용한 추계

가. 경제사회적 변수 관련 선행 연구

(1) 출산율과 경제사회적 변수

□ 오유진·박성준(2008)

- 연도별 1인당 GNP, 여성의 고학력 비율, 여성의 임금 증가율, 남성 대비 여성의 임금, 여성의 경제활동참가율, 가구당월평균소득, 가계소득 대비 교육비 지출비율, 청년 실업률, 지가변동률이 출산율에 주는 영향을 추정

□ 정성호(2009)

- 저출산의 경제사회적 요인으로 청년실업율, 경제활동참여율, 초혼연령, 월평균 사교육비, 양성평등 관념의 확산을 들고 있음

□ 권태환·김두섭(2011)

- 매우 다양한 요인들을 검토하고 있는데, 가족가치와 규범, 성선택, 가족주의, 결혼규범, 결혼의 안정성, 경제 조직 등을 언급하였고, 개인의 경제사회적 배경, 예컨대 경제적 지위 및 교육 수준, 거주지 배경이 출산에 영향을 준다고 보았음

□ 김경아(2017)

- 2016년 출산력 자료를 분석하여 연령과 학력, 자가주택 여부가 결혼 여성의 출산에 영향을 주었으나 여성의 직종이나 가구소득은 유의미하지 않음을 보고

□ 국회예산정책처 경제분석국 인구전략분석과(2018)

- 청년층이 노동시장에 참여한 이후의 고용형태와 임금수준, 여성의 근로시간, 보육시설 공급이 출산율에 영향을 준다고 보고 있음

□ 김도균(2021)

- 2012년부터 2018년까지 시군구 단위의 패널자료를 분석한 결과 수도권에서는 높은 집값이 출산에 부정적 영향을 준 반면 비수도권 지역에서는 높은 집값이 출산에 긍정적 영향을 주었으며 노동시장 변수는 유의미하지 않은 것으로 보고

(2) 사망률과 경제사회적 변수

□ 19세기 사망률 하락 논쟁

- 19세기 유럽과 미국에서 눈에 띄는 사망률 하락이 이루어지면서 그 원인에 대한 논쟁

이 있었으며, 항생제의 발견과 다양한 의학기술의 발전, 영양상태 및 생활수준의 향상, 공중보건의 발전과 위생의 향상, 교육 수준의 증가로 인한 건강 행위의 변화 등이 그 배경으로 꼽혔음(김현식, 2013)

□ 이미숙(2001)

- 40대 남성의 사망률에 대한 분석에서 혼인상태, 교육수준, 직업지원 및 사회적 자원이 사망률에 영향을 주는 것을 발견

□ 박은옥(2013)

- 시군구 수준의 심뇌혈관 질환 표준화 사망률을 비교하여 인구규모, 재정자립도, 인구 1000명당 의사수, 인구1인당 자동차 등록대수에 따라 사망률에 차이가 있음을 보고

□ 이원재(2019)

- 세계의 여러 시들에 대한 분석을 통해 미세먼지, 인간개발지수, 소득불평등이 사망률에 영향을 준다는 결과를 보고

나. 자료

(1) 성별 경제활동참가율, 실업률, 고용률

- 경제활동참가율: 15세 이상 인구 중 경제활동인구(취업자+실업자)가 차지하는 비율
- 실업률: 경제활동인구 중 실업자가 차지하는 비율
- 고용률: 15세 이상 인구 중 취업자가 차지하는 비율
- 출처: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2KAA301
- 자료 가용 년도: 2000-2020

(2) 청년실업률

- 15-29세 경제활동인구 중 실업자가 차지하는 비율
- 출처:
https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1YL20531E&conn_path=I3
- 자료 가용 년도: 2000-2021

(3) 1인당 실질 GDP 연간 성장률

- UN에서 작성한 자료를 kosis에서 내려받음
- 출처:

https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2UNS0071&conn_path=I2

□ 자료 가용 년도: 2000-2019

□ 2020년의 경우 World Bank의 자료를 사용함

(<https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG>)

(4) 소비자물가상승률

□ 통계청의 소비자물가조사를 바탕으로 통계청에서 작성한 자료 활용

□ 출처: <https://www.index.go.kr/unify/idx-info.do?idxCd=4226>

□ 자료 가용 년도: 1966-2021

(5) 지니계수

□ 다음의 표는 통계청 홈페이지에서 얻은 가계동향조사와 가계금융복지조사를 이용한 지니계수 산출 결과임

□ 두 조사 모두에서 시장소득을 이용한 지니계수를 활용하였음

□ 가계동향조사의 경우 도시에서 거주하는 2인 이상 가구를 대상으로 지니계수를 작성한 시계열이 가장 길지만 최근의 1인가구의 증가를 고려할 때 이 자료를 사용하는 것은 한계가 있을 것으로 파악됨

○ 동일한 년도 내에서 도시 2인 이상과 전체를 비교해 보면 전체의 지니계수가 크게 나타나는데 2006년에서 2010년 사이는 그 차이가 0.025 정도에 머물다가 이후 0.030을 넘어섬

□ 가계금융복지조사와 가계동향조사의 전체를 비교하면 가계금융복지조사의 지니계수가 높게 나타남

○ 이러한 현상에 대해 이원진·정해식·전지현(2019)은 두 조사 사이의 조사 방식의 차이에 기인하는 문제라고 설명하고 있는데, “연간소득을 면접조사하는 가계금융·복지조사와 비교할 때 가계동향조사는 월간소득 가계부기장을 실시하기 때문에 사적이전 소득 및 지출과 공적이전소득의 누락이 상대적으로 작고 비교적 정확하게 조사되지만, 고소득층의 누락이나 과소보고로 인해 근로소득, 사업소득, 재산소득 등 시장에서 결정되는 소득이 상대적으로 작게 조사될 가능성”이 크기 때문이라고 보고 있음

□ 최종 자료는 다음과 같은 방식으로 산출함

○ 가계동향조사의 전체 가구를 대상으로 작성한 지니계수를 기본으로 사용함

○ 2011년부터 2016년 자료를 이용하여 가계금융복지조사와 가계동향조사의 전체의 차

이를 살펴보면, 가계금융복지조사의 과대추정치가 최근으로 올수록 줄어들고 있고, 2016년의 차이는 0.049임

- 이를 이용하여 2017년부터는 가계금융복지조사 값에서 0.049를 빼서 사용함
- 1990년부터 2005년까지는 가계동향조사의 도시 2인이상에서 0.025를 더하여 최종 자료를 산출

<표 7.1> 다양한 자료별 지니계수 추이

연도	가계동향조사			가계금융 복지조사	최종 자료
	전체	2인 이상 비농가	도시 2인 이상		
1990			0.266		0.291
1991			0.259		0.284
1992			0.254		0.279
1993			0.256		0.281
1994			0.255		0.280
1995			0.259		0.284
1996			0.266		0.291
1997			0.264		0.289
1998			0.293		0.318
1999			0.298		0.323
2000			0.279		0.304
2001			0.290		0.315
2002			0.293		0.318
2003		0.292	0.283		0.308
2004		0.301	0.293		0.318
2005		0.306	0.298		0.323
2006	0.330	0.312	0.305		0.330
2007	0.340	0.321	0.316		0.340
2008	0.344	0.323	0.319		0.344
2009	0.345	0.320	0.320		0.345
2010	0.341	0.314	0.315		0.341
2011	0.342	0.313	0.313	0.418	0.342
2012	0.338	0.311	0.310	0.411	0.338
2013	0.336	0.308	0.307	0.401	0.336
2014	0.341	0.309	0.308	0.397	0.341
2015	0.341	0.307	0.305	0.396	0.341
2016	0.353	0.318	0.317	0.402	0.353
2017				0.406	0.357
2018				0.402	0.353
2019				0.404	0.355
2020				0.405	0.356

(6) 학력별 인구 비율

- 25-64세 인구를 대상으로 고졸 미만, 고졸, 고졸 초과 세 범주의 비율을 보여줌
- 출처: https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1530
- 자료 가용 년도: 1995-2020
- 이 자료에는 2013년의 자료가 누락되어 있어 2013년의 값은 2012년과 2013년의 평균 값으로 대체함

(7) 취업자의 주당 평균노동시간

- 다음의 표는 여러 자료의 주당 평균노동시간에 대한 추정값을 제공

<표 7.2> 다양한 자료별 주당 평균노동시간

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1)											46.0	45.9	45.2	43.9	44.6	43.1	43.9	43.7	43.0	42.8	41.5	40.7	39.0
2)				47.0	46.2	45.9	45.7	45.1	44.2	43.5	42.4	42.3	42.5	41.9	41.4	41.0	40.8	41.1					
3)	53.1	52.8	52.5	51.3	51.1	50.9	50.2																
4)	42.7	42.6	41.6	41.6	41.5	43.2	42.4																
5)	53.1	52.8	52.5	49.9	49.1	48.8	48.6	48.0	47.1	46.4	46.0	45.9	45.2	43.9	44.6	43.1	43.9	43.7	43.0	42.8	41.5	40.7	39.0

- 1)은 경제활동인구조사를 바탕으로 통계청에서 작성한 노동시간임.
 - 이 자료는 신뢰성이 높지만 시계열이 2008년부터 시작되어 가용 기간이 짧음
 - https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=3032
- 2)는 사업체노동력조사를 이용하여 김기민(2019)이 작성한 자료로 10인 이상 사업체의 평균 주당 근로시간을 보여줌
 - 이 자료는 통계청 자료에 비해 노동시간이 적는데 이는 아마도 10인 이상 사업체를 대상으로 작성했기 때문인 것으로 보임
 - 두 자료가 겹치는 2008년부터 2015년의 평균 차이는 2.9시간임
- 3)과 4)는 성재민(2005)이 작성한 것으로 3)은 한국노동패널을 분석한 것이고, 4)는 경제활동인구조사 중 10인 이상 사업체에 종사하는 노동자의 노동시간을 분석한 것임
 - 2)와 3)이 겹치는 2001년부터 2004년까지 두 자료를 비교해 보면 3)이 평균 4.7시간이 많음
 - 동일한 자료를 사용한 1)과 4)를 비교해 보면, 2)의 자료를 볼 때 2004년부터 2008년까지 노동시간이 줄어들었다는 것을 감안할 때 4)의 노동시간이 과소추정되어 있는

것으로 보이는데 이는 10인 이상 사업체에 속해 있는 노동자를 대상으로 했기 때문임
□ 5)는 분석에 사용한 최종 자료의 노동시간을 제시하는데, 이 자료는 다음과 같은 방식으로 작성하였음

- 우선 통계청의 자료가 가용한 시기는 통계청의 자료를 사용함
- 2001년부터 2007년까지는 2)의 자료에 통계청과의 평균 차이인 2.9를 더함
- 1998년부터 2000년까지는 3)의 자료를 사용함

(8) 아내의 초혼연령

- 통계청의 평균초혼연령을 이용하여 아내의 평균초혼연령을 사용
 - 1990년대 남편과 아내의 초혼연령 차이가 3년으로 일정하였으나 2021년 2.27로 줄어들었음
 - 그럼에도 불구하고 두 연령의 상관관계가 높아 여기에서는 아내의 초혼연령만을 사용함
- 출처: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B83A05&conn_path=I2
- 자료 가용 년도: 1990-2021

(9) 기혼여성(15-49세)의 자녀 필요성 질문에 꼭 있어야 한다고 응답한 여성의 비율

- 한국보건사회연구원에서는 3년 주기로 이른바 출산력 조사를 실시하는데, 이 조사에는 15-49세 기혼 여성을 대상으로 질문한 자녀의 필요성에 대한 설문이 들어 있음
- 자녀의 필요성에 대한 선택지는 통상 다음과 같은 4개로 이루어져 있음: 꼭 있어야 한다, 있는 것이 없는 것보다 나을 것이다, 없어도 상관없다, 모르겠다
- 2006년과 2009년의 경우 모르겠다는 선택지가 없었으나 이에 대한 응답은 통상 1%을 넘지 않았기 때문에 그대로 사용함
- 2012년 조사의 경우 15-44세 기혼여성을 대상으로 질문하였으나 별차이가 없는 것으로 간주하고 그대로 사용함
- 2021년의 경우 미혼 포함 19-49세의 남녀 모두에게 질문하여 비교 가능한 수치를 작성하는 것이 쉽지 않으나, 기혼인 경우 42.8%가 꼭 있어야 한다고 응답하고, 남자는 32.9%, 여자는 28.1%라는 결과를 이용해, 혼인상태와 성별이 이 문항 응답에 대한 분포가 독립이라는 가정 하에 $0.428 \times 28.1 / (32.9 + 28.1) \times 2 = 0.394$ 로 계산한 값을 사용
- 자료가 없는 중간 2년은 두 가용한 자료 사이의 단순 선형 변화로 예측

□ 자료: 신창우 외, 2012; 김승권 외, 2012; 이삼식 외, 2015; 이소영 외, 2018; 박종서 외, 2021

<표 7.3> 자녀의 필요성 및 종교 없음의 백분율 변화

연도	자녀는 꼭 있어야 한다		종교 없음	
	원자료	최종자료	원자료	최종자료
2000	58.1	58.1		40.1
2001		57.1		40.1
2002		56.2		40.1
2003	55.2	55.2	40.1	40.1
2004		55.0	40.1	40.1
2005		54.9	40.1	40.1
2006	54.7	54.7	40.1	40.1
2007		54.9	38.8	38.8
2008		55.2	39.3	39.3
2009	55.4	55.4	39.9	39.9
2010		52.4	43.8	43.8
2011		49.3	40.5	40.5
2012	46.3	46.3	42.7	42.7
2013		50.9	43.6	43.6
2014		55.6	42.2	42.2
2015	60.2	60.2		44.0
2016		56.8	45.8	45.8
2017		53.3		46.9
2018	49.9	49.9	48.0	48.0
2019		46.4		50.3
2020		42.9		52.6
2021	39.4	39.4	54.9	54.9

(10) 종교가 없다고 응답한 사람들의 백분율

- 한국종합사회조사에서 “어떤 종교를 가지고 계십니까?”라는 질문에 대한 응답에서 종교 없음으로 응답한 사람의 비율을 구하였음
- 원자료에서 2003년부터 2006년은 해당 연도에 대한 평균을 구하였는데, 여기서는 연도별로 동일하다고 가정
- 2000-2002년은 2003년과 동일하다고 가정하고 이외의 나머지 자료가 없는 년도는 자

료가 있는 가장 근접한 양년도 사이의 단순선형모형으로 추정

□ 자료: 김지범 외, 2022.

다. 기초분석

(1) 모든 자료들이 가용한 2000년부터 2020년의 자료를 이용하여 변수들 사이의 상관계수를 구하였음

(2) 다음의 표에서는 각 경제사회적 변수들간 상관계수를 보여주고 있으며 상관계수의 절대값이 0.9보다 큰 두 변수를 예측 변수로 넣으면 둘 중 하나의 예측력이 거의 없을 것으로 판단하므로 이런 변수를 제거할 필요가 있는 것으로 판단하여 절대값이 0.9보다 큰 변수들의 상관계수는 회색바탕으로 처리하였음

□ 높은 상관관계를 가진 변수들을 제외하면 남자 경제활동참가율, 여자 경제활동참가율, 남자 실업률, 여자 실업률, 청년실업률, 1인당 실질 GDP 상승률, 소비자물가상승률, 지니계수, 고등학교 인구비율, 자녀의 필요성, 종교 없음 비율 등이 잠재적 활용가능한 변수로 판단됨

(3) 그 다음 쪽에서는 추계해야할 모수들과 경제사회적 변수들 사이의 상관계수를 살펴보고있음

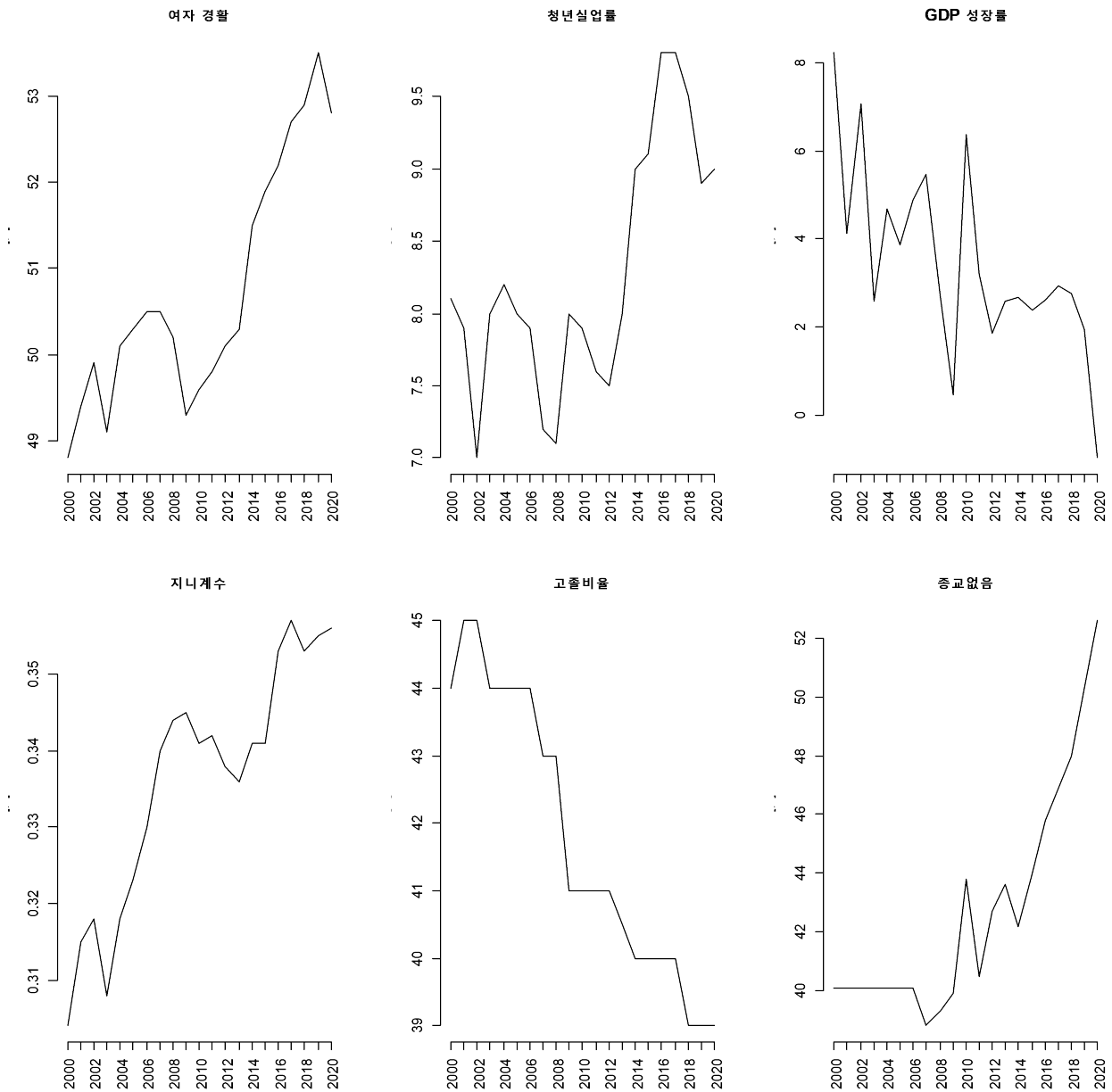
□ 이 표에서는 절대값이 0.8보다 높은 상관계수를 회색 바탕으로 처리하였음

□ 이 표를 보면 여자 경제활동참가율, 여자 고용률, 지니계수, 중학교 이하 인구 비율, 고등학교 인구 비율, 대학교 이상 인구 비율, 주당 노동시간, 아내 평균혼인연령, 종교 없음 비율이 출산율과 사망률의 모수들과 높은 상관관계를 가지고 있음

(4) 여자 경제활동참가율, 청년실업률, 1인당 실질 GDP 상승률, 지니계수, 고등학교 인구비율, 종교 없음 비율을 경제사회적 변수로 사용하였음

□ 출산율 추계에서는 이들 6개의 변수를 사용하였으나, 종교 없음 비율이 사망률과 관계가 없을 것이라는 조언에 따라 사망률 추계에서는 종교 없음 비율은 제외하였음

<그림 7.1> 경제사회적 변수들의 관측값



<표 7.4> 경제사회적 변수들 사이의 상관계수

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.000	-0.315	0.126	-0.182	0.926	-0.309	-0.196	0.578	0.332	-0.702	0.634	0.737	-0.690	0.709	-0.676	0.664	-0.588
2	-0.315	1.000	-0.277	0.551	-0.222	0.991	0.769	-0.491	-0.747	0.767	-0.855	-0.747	0.833	-0.820	0.831	-0.407	0.849
3	0.126	-0.277	1.000	0.385	-0.248	-0.364	0.054	0.401	0.163	-0.433	0.492	0.356	-0.451	0.520	-0.478	0.280	-0.103
4	-0.182	0.551	0.385	1.000	-0.336	0.440	0.805	-0.273	-0.541	0.210	-0.394	-0.464	0.414	-0.354	0.421	-0.216	0.685
5	0.926	-0.222	-0.248	-0.336	1.000	-0.183	-0.241	0.411	0.272	-0.541	0.451	0.604	-0.519	0.511	-0.497	0.526	-0.546
6	-0.309	0.991	-0.364	0.440	-0.183	1.000	0.700	-0.492	-0.719	0.792	-0.861	-0.733	0.834	-0.830	0.831	-0.409	0.805
7	-0.196	0.769	0.054	0.805	-0.241	0.700	1.000	-0.405	-0.683	0.519	-0.684	-0.677	0.675	-0.573	0.676	-0.086	0.723
8	0.578	-0.491	0.401	-0.273	0.411	-0.492	-0.405	1.000	0.349	-0.615	0.690	0.642	-0.686	0.721	-0.688	0.506	-0.530
9	0.332	-0.747	0.163	-0.541	0.272	-0.719	-0.683	0.349	1.000	-0.512	0.698	0.706	-0.688	0.648	-0.706	0.300	-0.753
10	-0.702	0.767	-0.433	0.210	-0.541	0.792	0.519	-0.615	-0.512	1.000	-0.929	-0.866	0.933	-0.915	0.911	-0.500	0.686
11	0.634	-0.855	0.492	-0.394	0.451	-0.861	-0.684	0.690	0.698	-0.929	1.000	0.950	-0.996	0.969	-0.993	0.520	-0.796
12	0.737	-0.747	0.356	-0.464	0.604	-0.733	-0.677	0.642	0.706	-0.866	0.950	1.000	-0.969	0.923	-0.972	0.535	-0.799
13	-0.690	0.833	-0.451	0.414	-0.519	0.834	0.675	-0.686	-0.688	0.933	-0.996	-0.969	1.000	-0.973	0.995	-0.553	0.816
14	0.709	-0.820	0.520	-0.354	0.511	-0.830	-0.573	0.721	0.648	-0.915	0.969	0.923	-0.973	1.000	-0.973	0.663	-0.828
15	-0.676	0.831	-0.478	0.421	-0.497	0.831	0.676	-0.688	-0.706	0.911	-0.993	-0.972	0.995	-0.973	1.000	-0.563	0.822
16	0.664	-0.407	0.280	-0.216	0.526	-0.409	-0.086	0.506	0.300	-0.500	0.520	0.535	-0.553	0.663	-0.563	1.000	-0.642
17	-0.588	0.849	-0.103	0.685	-0.546	0.805	0.723	-0.530	-0.753	0.686	-0.796	-0.799	0.816	-0.828	0.822	-0.642	1.000

1: 남자 경제활동참가율, 2: 여자 경제활동참가율, 3: 남자 실업률, 4: 여자 실업률, 5: 남자 고용률, 6: 여자 고용률, 7: 청년실업률, 8: 1인당 실질 GDP 상승률, 9: 소비자 물가상승률, 10: 지니계수, 11: 중학교 이하 인구 비율, 12: 고등학교 인구 비율, 13: 대학교 이상 인구 비율, 14: 주당 노동시간, 15: 아내 평균혼인연령, 16: 자녀의 필요성, 17: 종교 없음 비율

<표 7.5> 출산율 및 사망률 모수들과 경제사회적 변수들 사이의 상관계수

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a	-0.640	0.775	-0.437	0.375	-0.488	0.772	0.653	-0.555	-0.681	0.838	-0.915	-0.935	0.922	-0.896	0.943	-0.489	0.791
b	-0.691	0.839	-0.469	0.409	-0.514	0.842	0.657	-0.678	-0.702	0.925	-0.994	-0.966	0.998	-0.980	0.997	-0.583	0.827
c	0.505	-0.844	0.315	-0.467	0.390	-0.837	-0.575	0.667	0.588	-0.753	0.803	0.713	-0.802	0.862	-0.792	0.669	-0.861
c1	0.519	-0.843	0.342	-0.479	0.396	-0.833	-0.623	0.655	0.648	-0.768	0.838	0.758	-0.838	0.877	-0.826	0.608	-0.869
sig1	-0.530	0.581	-0.601	0.026	-0.311	0.625	0.358	-0.515	-0.310	0.803	-0.787	-0.753	0.780	-0.777	0.796	-0.481	0.471
c2	0.410	-0.787	0.429	-0.308	0.255	-0.802	-0.457	0.664	0.424	-0.741	0.752	0.620	-0.739	0.820	-0.730	0.653	-0.724
sig2	-0.170	0.446	0.292	0.693	-0.292	0.362	0.617	-0.087	-0.686	0.109	-0.282	-0.367	0.303	-0.263	0.317	-0.058	0.639
c	0.482	-0.836	0.393	-0.411	0.340	-0.838	-0.562	0.678	0.563	-0.772	0.819	0.713	-0.813	0.872	-0.802	0.641	-0.823
sig1	-0.458	0.527	-0.592	-0.039	-0.242	0.576	0.276	-0.484	-0.208	0.744	-0.708	-0.658	0.697	-0.708	0.712	-0.475	0.391
sig2	-0.211	0.352	0.256	0.609	-0.315	0.275	0.552	-0.069	-0.658	0.099	-0.268	-0.365	0.292	-0.235	0.300	-0.005	0.559
mu	-0.666	0.803	-0.541	0.297	-0.462	0.821	0.574	-0.680	-0.589	0.938	-0.975	-0.929	0.973	-0.967	0.974	-0.602	0.745
c1	-0.167	0.640	-0.160	0.381	-0.112	0.632	0.385	-0.503	-0.336	0.433	-0.465	-0.319	0.451	-0.543	0.429	-0.503	0.619
lambda	0.848	-0.536	0.510	-0.087	0.642	-0.567	-0.298	0.613	0.475	-0.890	0.858	0.870	-0.880	0.879	-0.867	0.606	-0.582
b1	0.705	-0.765	0.434	-0.369	0.550	-0.764	-0.628	0.585	0.680	-0.864	0.932	0.956	-0.944	0.922	-0.959	0.542	-0.806
u	-0.664	0.858	-0.461	0.434	-0.491	0.859	0.685	-0.675	-0.714	0.917	-0.995	-0.963	0.996	-0.977	0.998	-0.570	0.838
TFR	0.346	-0.734	0.272	-0.388	0.247	-0.732	-0.451	0.606	0.440	-0.613	0.642	0.512	-0.634	0.715	-0.613	0.599	-0.723
kt male	0.647	-0.858	0.491	-0.399	0.465	-0.863	-0.677	0.684	0.697	-0.930	0.998	0.952	-0.995	0.976	-0.995	0.539	-0.809
kt female	0.634	-0.844	0.502	-0.402	0.448	-0.849	-0.690	0.669	0.705	-0.910	0.995	0.954	-0.991	0.966	-0.995	0.524	-0.804
LE male	-0.674	0.842	-0.486	0.386	-0.492	0.848	0.659	-0.690	-0.691	0.936	-0.998	-0.958	0.998	-0.978	0.996	-0.554	0.806
LE female	-0.698	0.816	-0.486	0.376	-0.516	0.822	0.645	-0.690	-0.674	0.935	-0.994	-0.966	0.997	-0.976	0.996	-0.557	0.793

1: 남자 경제활동참가율, 2: 여자 경제활동참가율, 3: 남자 실업률, 4: 여자 실업률, 5: 남자 고용률, 6: 여자 고용률, 7: 청년실업률, 8: 1인당 실질 GDP 상승률, 9: 소비자 물가상승률, 10: 지니계수, 11: 중학교 이하 인구 비율, 12: 고등학교 인구 비율, 13: 대학교 이상 인구 비율, 14: 주당 노동시간, 15: 아내 평균혼인연령, 16: 자녀의 필요성, 17: 종교 없음 비율

(4) 다음 그림은 해당 변수들의 1차 차분한 값들의 시계열 변화를 보여줌

라. Vector autoregression 모형을 활용한 설명 변수 추계

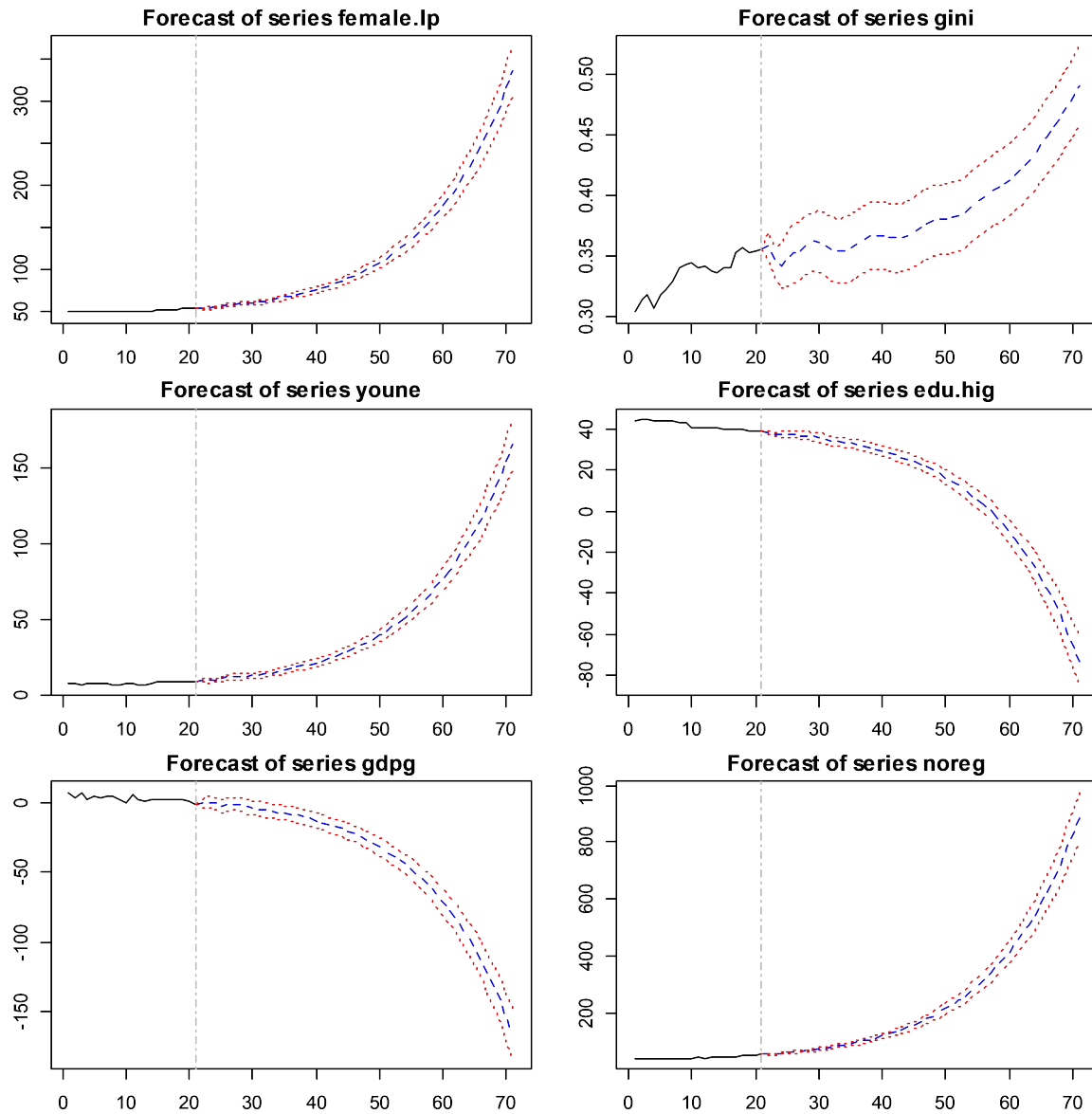
(1) Level 모형

- 특정 t 시기 두 변수 Y_t 와 X_t 가 있다고 할 때 VAR(p) 모형은 다음과 같이 나타낼 수 있음 (Pfaff & Stigler, 2021)

$$\begin{aligned} Y_t &= \beta_{10} + \beta_{11} Y_{t-1} + \dots + \beta_{1t-p} Y_{t-p} + \gamma_{11} X_{t-1} + \dots + \gamma_{1p} X_{t-p} + u_{1t} \\ X_t &= \beta_{20} + \beta_{21} X_{t-1} + \dots + \beta_{2t-p} X_{t-p} + \gamma_{21} Y_{t-1} + \dots + \gamma_{2p} Y_{t-p} + u_{2t} \end{aligned}$$

- 과거 자료를 이용하여 계수들을 추정한 후 현재까지 관측된 자료로 미래의 값을 예측할 수 있는데, 예컨대, 앞의 식에 t 에 $t+1$ 을 대입하면 1년 후의 값을 예측할 수 있음
- 현재 연구의 문제는 변수당 사례수가 2000년부터 2020년까지 21개에 지나지 않아서 $p=1$ 또는 $p=2$ 의 경우 외에는 추정할 수 없다는 것임
- VARselect 함수를 이용하여 최적의 p 를 고르면, AIC나 BIC 모두 $p=2$ 의 값을 추천하여 $p=2$ 로 모형 설정
- 다음 그림은 이 모형의 추계 결과를 보여주고, 그 다음 표는 이 모형의 추정 계수들을 보여줌
- R의 vars 패키지에서는 모형 추정 후 autocorrelation이 없다는 영가설을 검정하는 Portmanteau 검정을 활용할 수 있는데, 이 경우 자유도의 부족으로 검정 결과를 산출할 수 없음

<그림 7.2> VAR level 모형의 추계 결과



주. 1열의 첫 행부터 여자 경제활동참가율, 청년실업률, GDP 증가율이며 2열의 첫 행부터 지니계수, 고졸 비율, 종교 없음 비율임

<표 7.6> 관측값 VAR(2) 모형의 계수

설명 변수	여자 경제활동참가율				청년실업률				GDP 성장률			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	0.181	0.581	0.311	0.766	-0.088	0.792	-0.111	0.916	-1.836	1.727	-1.063	0.329
청년실업1	0.491	0.325	1.510	0.182	0.534	0.444	1.203	0.274	1.263	0.968	1.306	0.240
GDP1	0.032	0.132	0.241	0.817	0.001	0.179	0.004	0.997	-0.392	0.391	-1.001	0.356
지니계수1	-25.71 2	27.633	-0.930	0.388	-48.23 0	37.678	-1.280	0.248	64.287	82.161	0.782	0.464
고졸비율1	-0.231	0.408	-0.567	0.591	0.371	0.556	0.667	0.530	-0.648	1.212	-0.535	0.612
종교없음1	-0.068	0.144	-0.475	0.652	-0.169	0.196	-0.863	0.421	0.134	0.427	0.315	0.764
여자 경활2	0.451	0.617	0.731	0.492	0.842	0.841	1.001	0.355	-0.370	1.834	-0.202	0.847
청년실업2	0.053	0.460	0.115	0.912	-0.521	0.627	-0.832	0.437	2.044	1.367	1.495	0.186
GDP2	-0.030	0.146	-0.203	0.846	-0.199	0.199	-0.999	0.356	-0.012	0.433	-0.027	0.980
지니계수2	-18.30 2	37.724	-0.485	0.645	5.811	51.437	0.113	0.914	-76.87 3	112.16 4	-0.685	0.519
고졸비율2	-0.127	0.372	-0.341	0.744	-0.665	0.508	-1.310	0.238	1.060	1.107	0.958	0.375
종교없음2	0.096	0.155	0.623	0.556	0.068	0.211	0.323	0.758	-0.253	0.460	-0.550	0.602
절편	43.133	21.307	2.024	0.089	1.959	29.052	0.067	0.948	80.658	63.352	1.273	0.250

	지니계수				고졸비율				종교없음 비율			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	-0.007	0.008	-0.868	0.419	-0.553	0.505	-1.094	0.316	1.130	1.507	0.750	0.482
청년실업1	0.001	0.004	0.194	0.852	-0.307	0.283	-1.083	0.320	-0.973	0.844	-1.152	0.293
GDP1	0.001	0.002	0.820	0.444	0.095	0.114	0.833	0.437	-0.871	0.341	-2.550	0.044
지니계수1	0.571	0.358	1.597	0.161	8.599	24.039	0.358	0.733	38.566	71.700	0.538	0.610
고졸비율1	0.000	0.005	-0.063	0.951	-0.382	0.355	-1.077	0.323	-1.665	1.058	-1.574	0.167
종교없음1	-0.002	0.002	-0.809	0.450	0.096	0.125	0.772	0.470	0.823	0.373	2.207	0.069
여자 경활2	0.014	0.008	1.790	0.124	0.180	0.537	0.336	0.748	-1.405	1.601	-0.878	0.414
청년실업2	-0.001	0.006	-0.241	0.818	0.758	0.400	1.896	0.107	2.002	1.193	1.678	0.144
GDP2	0.000	0.002	-0.211	0.840	0.142	0.127	1.122	0.305	0.258	0.378	0.684	0.520
지니계수2	-0.479	0.489	-0.980	0.365	-96.96 4	32.817	-2.955	0.025	-23.37 8	97.883	-0.239	0.819
고졸비율2	-0.006	0.005	-1.325	0.233	0.535	0.324	1.652	0.150	1.645	0.966	1.703	0.139
종교없음2	-0.001	0.002	-0.622	0.557	-0.118	0.135	-0.877	0.414	0.001	0.401	0.003	0.998
절편	0.316	0.276	1.147	0.295	79.698	18.536	4.300	0.005	10.535	55.286	0.191	0.855

주. 첫 열은 설명 변수를 보여주며, 설명 변수의 숫자는 p 값임

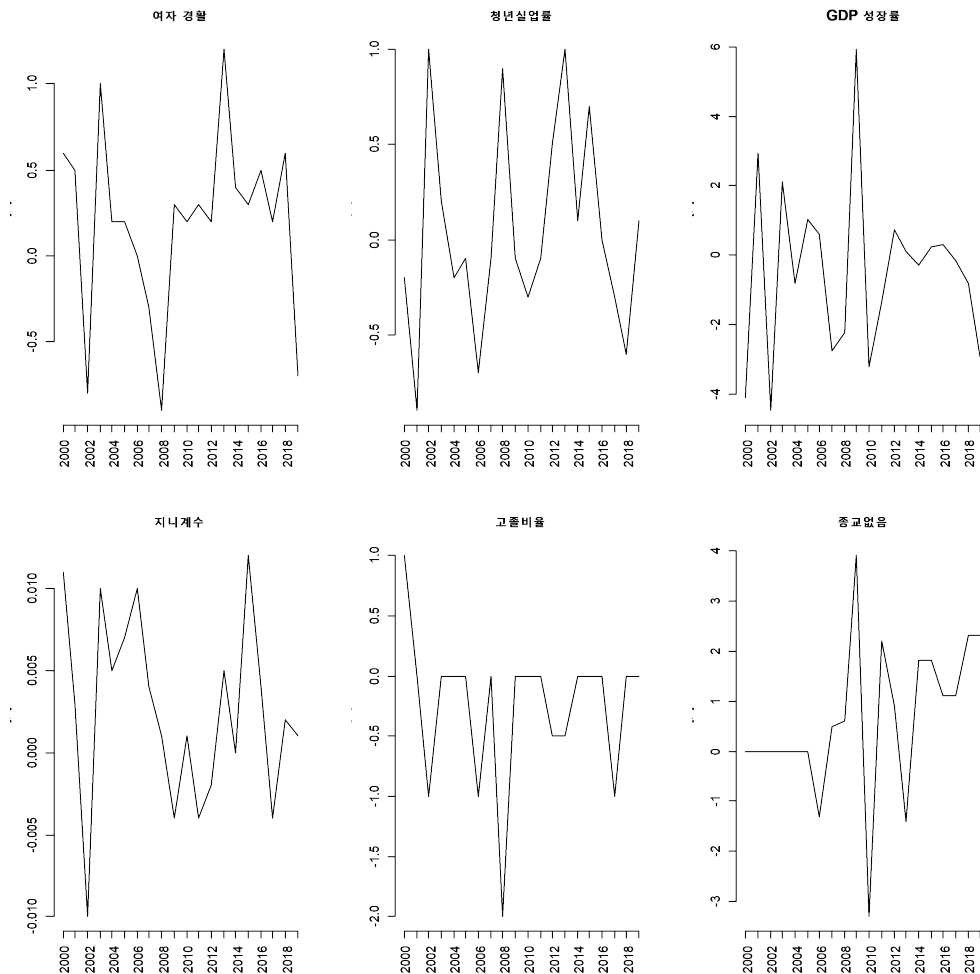
(2) 차분 모형

- 현재의 자료와 같이 변수값이 정상성(stationarity)이 없으면 차분을 하여 사용하는데 이 경우 다음과 같은 식으로 나타낼 수 있음

$$\begin{aligned}\Delta Y_t &= \beta_{10} + \beta_{11}\Delta Y_{t-1} + \dots + \beta_{1t-p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \dots + \gamma_{1p}\Delta X_{t-p} + u_{1t} \\ \Delta X_t &= \beta_{20} + \beta_{21}\Delta X_{t-1} + \dots + \beta_{2t-p}\Delta X_{t-p} + \gamma_{21}\Delta Y_{t-1} + \dots + \gamma_{2p}\Delta Y_{t-p} + u_{2t}\end{aligned}$$

- 다음 그림은 차분 값들의 변화를 보여줌

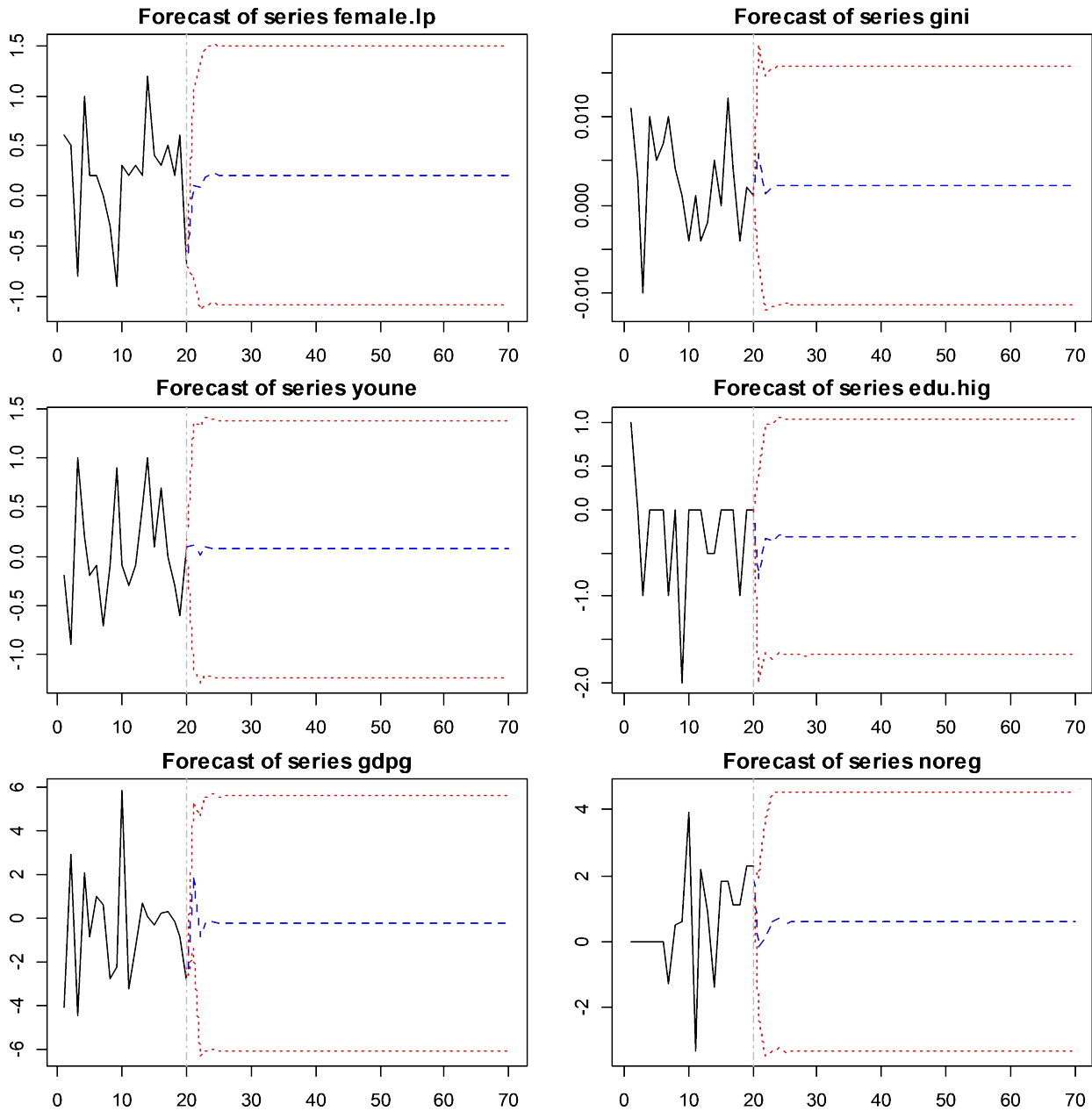
<그림 7.3> 1차 차분 값들의 변화



- Level 모형과 마찬가지로 관측 사례수가 적어 $p=1$ 경우 외에는 추정할 수 없음

□ 다음 그림은 차분값들의 추계를 보여줌

<그림 7.4> VAR 1차 차분 모형의 차분 추계 결과

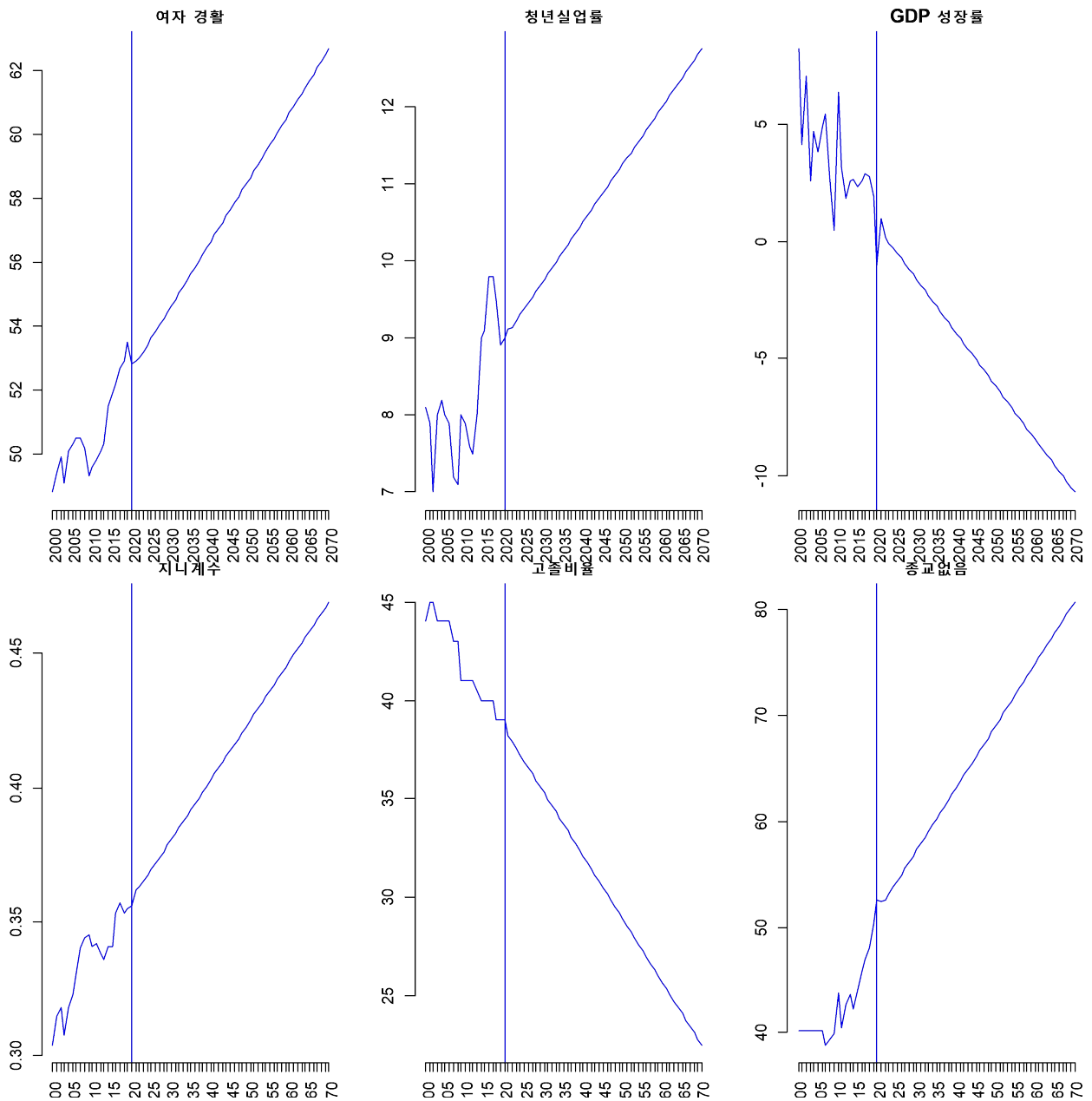


주. 1열의 첫 행부터 여자 경제활동참가율, 청년실업률, GDP 증가율이며 2열의 첫 행부터 지니계수, 고졸 비율, 종교 없음 비율임

□ 1차 차분 모형에 대한 autocorrelation이 없다는 영가설을 검정하는 Portmanteau 검정을 해 보면 p-값이 0.001보다 작아 영가설을 기각하여 autocorrelation이 있다고 해석할 수 있음

□ 다음 그림은 관측값들의 추계를 보여줌

<그림 7.5> 1차 차분 모형의 추계 결과



<표 7.7> 1차 차분 VAR(1) 모형의 계수

설명 변수	여자 경제활동참가율				청년실업률				GDP 성장률			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	0.115	0.348	0.329	0.748	-0.110	0.472	-0.233	0.820	-0.601	1.269	-0.473	0.644
청년실업1	0.664	0.235	2.830	0.015	0.034	0.318	0.108	0.916	2.126	0.855	2.487	0.029
GDP1	0.002	0.064	0.038	0.970	0.032	0.087	0.371	0.717	-0.473	0.234	-2.023	0.066
지니계수1	-30.17 4	22.635	-1.333	0.207	-31.26 7	30.670	-1.019	0.328	101.06 5	82.479	1.225	0.244
고졸비율1	0.113	0.275	0.410	0.689	0.107	0.373	0.288	0.778	-0.442	1.003	-0.441	0.667
종교없음1	-0.046	0.092	-0.502	0.625	-0.027	0.124	-0.221	0.829	0.341	0.334	1.022	0.327
절편	0.259	0.225	1.149	0.273	0.221	0.305	0.725	0.482	-0.938	0.821	-1.143	0.275

	지니계수				고졸비율				종교없음 비율			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	-0.002	0.005	-0.493	0.631	0.628	0.452	1.389	0.190	2.365	0.793	2.981	0.011
청년실업1	0.004	0.003	1.124	0.283	0.091	0.304	0.300	0.769	-0.869	0.534	-1.627	0.130
GDP1	0.000	0.001	-0.046	0.964	-0.055	0.083	-0.659	0.522	-0.639	0.146	-4.370	0.001
지니계수1	0.081	0.302	0.268	0.793	-3.362	29.367	-0.114	0.911	57.866	51.533	1.123	0.283
고졸비율1	0.002	0.004	0.670	0.515	-0.444	0.357	-1.242	0.238	-2.594	0.627	-4.140	0.001
종교없음1	0.000	0.001	0.302	0.768	0.044	0.119	0.366	0.721	0.325	0.209	1.558	0.145
절편	0.003	0.003	0.931	0.370	-0.624	0.292	-2.136	0.054	-1.115	0.513	-2.174	0.050

(3) 2차 차분 모형

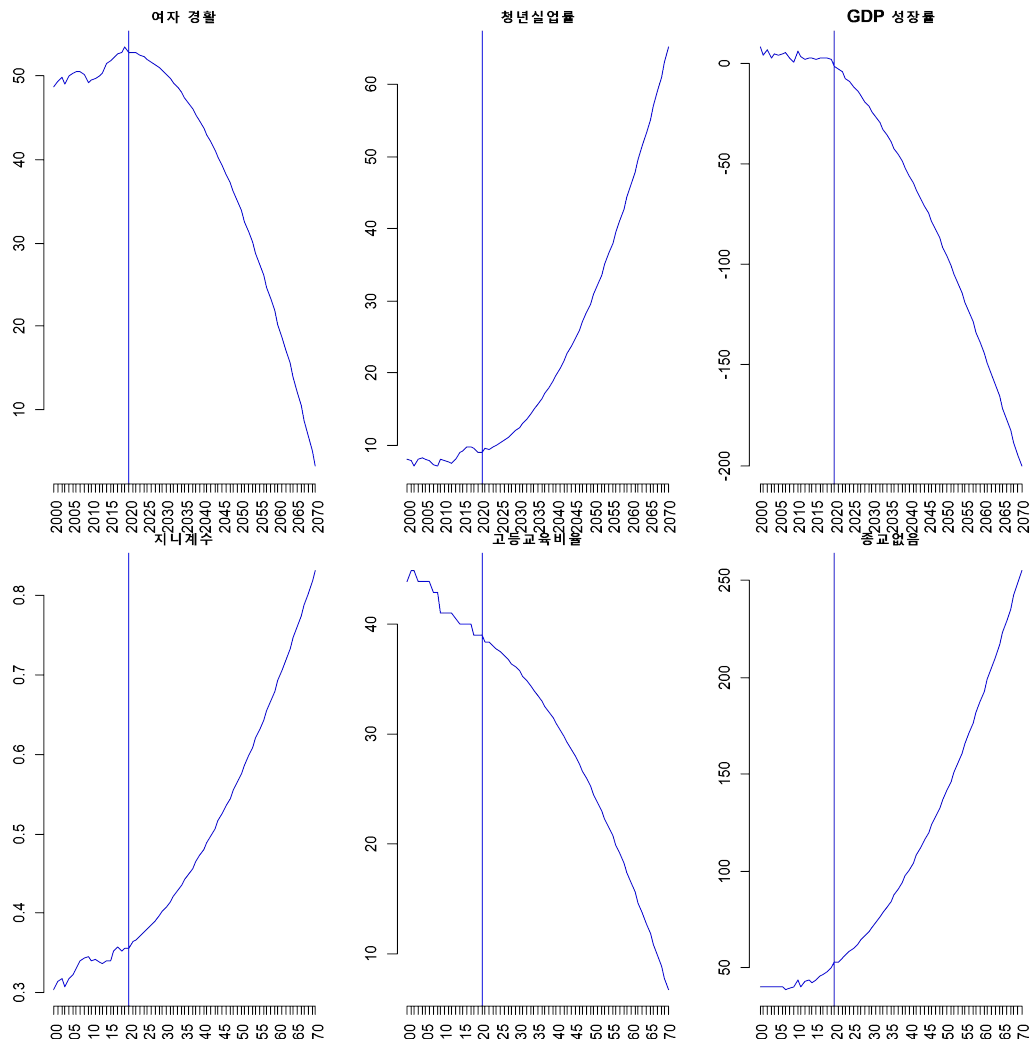
- 앞선 결과들이 모두 흡족하지 않기 때문에 2차 차분 변수를 활용하여 VAR 모형을 추정
- 그 결과는 다음에 제시되어 있음
- 2차 차분 모형에 대한 autocorrelation이 없다는 영가설을 검정하는 Portmanteau 검정을 해 보면 p-값이 0.001보다 작아 영가설을 기각하여 autocorrelation이 있다고 해석할 수 있음

<표 7.8> 2차 차분 VAR(1) 모형의 계수

설명 변수	여자 경제활동참가율				청년실업률				GDP 성장률			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	-0.361	0.498	-0.725	0.483	-0.583	0.581	-1.004	0.337	0.612	1.682	0.364	0.723
청년실업1	0.561	0.287	1.954	0.077	-0.430	0.335	-1.283	0.226	1.425	0.971	1.468	0.170
GDP1	0.031	0.078	0.401	0.696	0.051	0.091	0.560	0.587	-0.716	0.262	-2.733	0.019
지니계수1	-11.38 8	28.647	-0.398	0.699	-7.128	33.448	-0.213	0.835	112.58 9	96.802	1.163	0.269
고졸비율1	-0.027	0.299	-0.091	0.929	0.194	0.349	0.558	0.588	-2.036	1.009	-2.018	0.069
종교없음1	-0.042	0.107	-0.394	0.701	-0.025	0.125	-0.200	0.845	0.749	0.361	2.074	0.062
절편	-0.062	0.162	-0.380	0.711	0.047	0.189	0.249	0.808	-0.316	0.548	-0.577	0.575

	지니계수				고졸비율				종교없음 비율			
	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값	계수	표준 오차	t-값	p-값
여자 경활1	-0.003	0.006	-0.607	0.556	0.499	0.658	0.759	0.464	1.859	0.979	1.899	0.084
청년실업1	0.003	0.003	0.927	0.374	-0.087	0.380	-0.228	0.824	-1.261	0.565	-2.232	0.047
GDP1	0.000	0.001	-0.212	0.836	-0.088	0.103	-0.859	0.409	-0.604	0.152	-3.962	0.002
지니계수1	-0.187	0.325	-0.576	0.576	24.929	37.866	0.658	0.524	83.458	56.321	1.482	0.166
고졸비율1	0.003	0.003	0.941	0.367	-0.834	0.395	-2.114	0.058	-2.376	0.587	-4.049	0.002
종교없음1	0.001	0.001	0.545	0.597	0.116	0.141	0.824	0.427	0.129	0.210	0.614	0.552
절편	0.000	0.002	-0.006	0.995	-0.035	0.214	-0.162	0.874	0.103	0.319	0.322	0.753

<그림 7.6> 2차 차분 모형의 추계 결과



마. ARIMA 모형을 활용한 설명 변수 추계

- (1) 앞선 VAR 모형의 결과들은 1차 차분 결과를 제외하면 통념을 벗어나는 결과들이 많고, 1차 차분 결과 중 GDP 성장률 같은 경우 심하게 우하향하는 결과를 보여주어 사용하기 어려운 측면이 있음
- (2) 여기에서는 ARIMA 모형의 추정 결과를 제시하는데, R의 forecast 패키지에 있는 auto.arima 함수를 이용하여 모형을 추정함

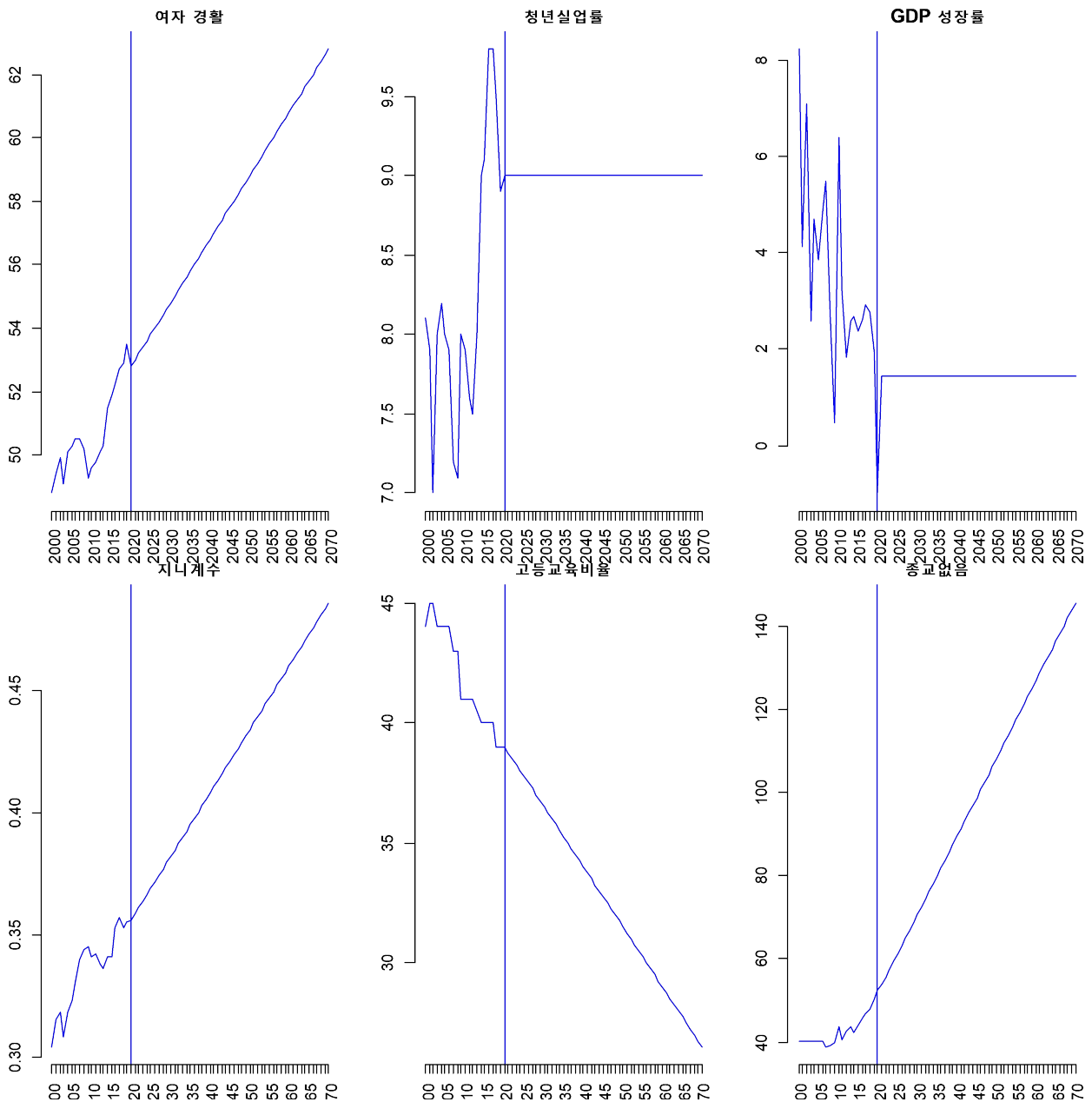
□ 여자 경제활동참가율은 ARIMA(0,1,0) with drift 모형으로 추정되었으며 drift의 계수

는 0.200이었고 표준오차는 0.117이었음

- 청년실업률은 ARIMA(0,1,0) 모형으로 추정되었음
- GDP 증가율은 ARIMA(0,1,1) 모형으로 추정되었으며 drift의 계수는 MA의 계수는 -0.691이었고 표준오차는 0.46이었음
- 지니계수는 ARIMA(0,1,0) with drift 모형으로 추정되었으며 drift의 계수는 0.003이었고 표준오차는 0.001이었음
- 고졸비율 또한 ARIMA(0,1,0) with drift 모형으로 추정되었으며 drift의 계수는 -0.250이었고 표준오차는 0.135이었음
- 종교없음 비율은 ARIMA(0,2,2) 모형으로 추정되었으며 MA의 계수는 1차가 -1.268이었고 표준오차는 0.249, 2차가 0.504, 표준오차가 0.210이었음

(3) 다음 그림은 결과를 보여줌

<그림 7.7> ARIMA 모형의 추계 결과



(3) 1차 차분 모형의 결과와 비교해 보면,

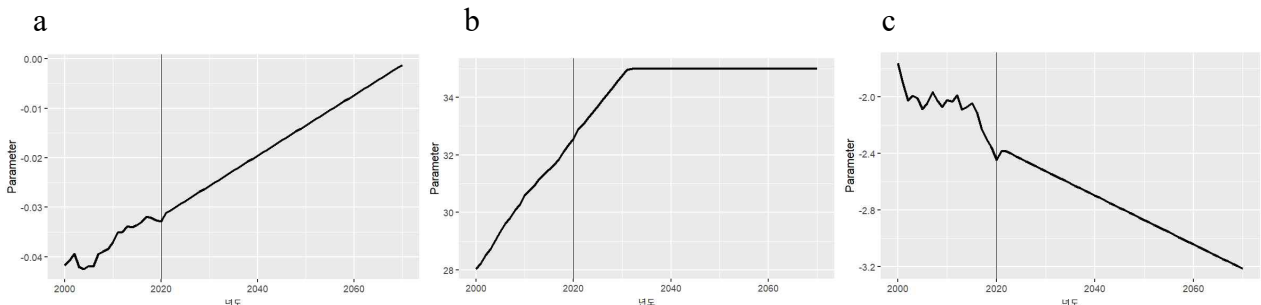
- 여자 경제활동참가율의 경우 상당히 유사한 추계 결과가 나왔음
- 청년 실업률의 경우 1차 차분 모형은 우상향하지만 그리 크게 증가하지는 않아 조금 더 현실적인 결과를 보여주는 것으로 판단
- GDP 성장률의 경우 1차 차분 모형이 지나치게 우하향하므로 장기적으로 균형을 이루는 ARIMA 모형이 더 적합한 것으로 판단함

- 지니계수도 두 모형이 유사한 값을 보여주지만 1차 차분 모형이 덜 급격히 증가하는 것으로 나타남
- 고등교육 비율 또한 1차 차분 모형이 덜 우하향하는 것으로 판단함
- 종교 없음의 비율은 ARIMA 모형 결과에서 100%가 넘어가는 결과가 나와 1차 차분 모형 결과가 더 나은 것으로 판단함
- 이러한 관찰을 바탕으로 1차 차분 모형의 결과를 활용하되 GDP 성장률만 ARIMA 모형의 결과를 활용

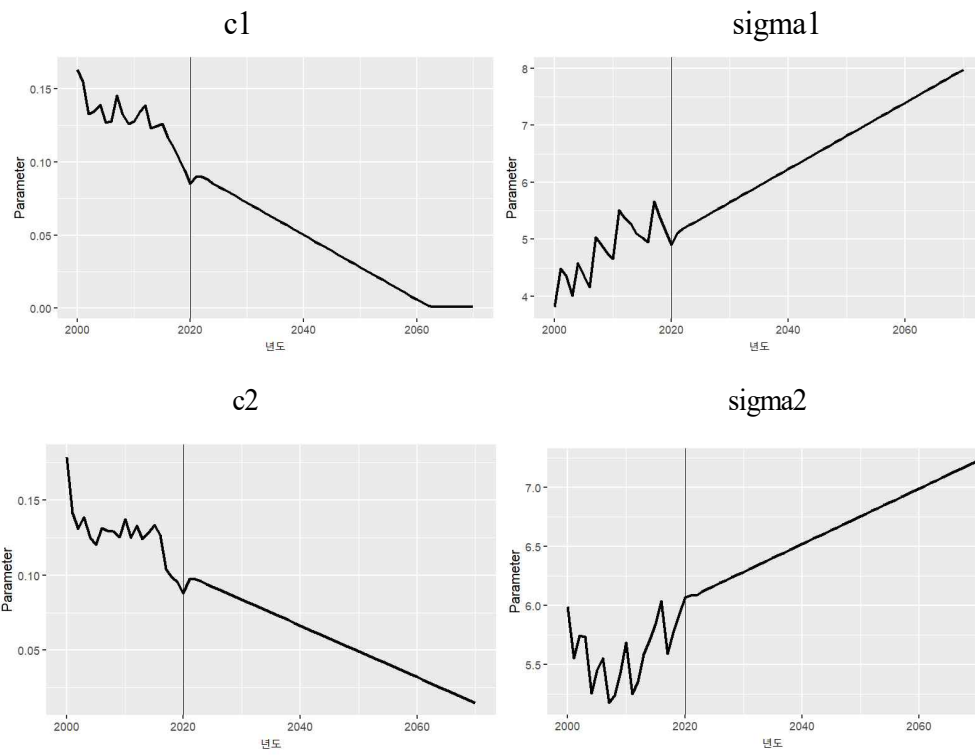
바. 출산율과 사망률 모수 추계 결과

- 위에서 추계한 경제사회적 변수들의 추계값을 활용하여 출산율과 사망률의 모수를 추계하였음
- 추계 과정에서 경제사회적 변수들을 고려하지 않은 모형에서 제기되었던 다양한 문제들이 나타났음
 - 출산율 로그 이차함수 모형에서 a 값이 양수로 추계되는 지점이 존재하는데 이는 연령별 출산율이 위로 볼록한 모양을 띠어한다는 사실에 위배되므로, -0.001 보다 큰 값은 모두 -0.001 로 대체하였으며 b 값이 35세를 넘게 되는 경우 모두 35세로 대체
 - 출산율 관련 두 개의 PK 모형에서 c 모수들이 0.001 보다 작으면 모두 0.001 로 대체
 - 출산율이 가장 높은 연령이나 출산 평균연령이 35세보다 클 수 없다는 가정하에 35세가 넘으면 35세로 대체
- 다음의 그림들은 추계 결과값들을 제시함

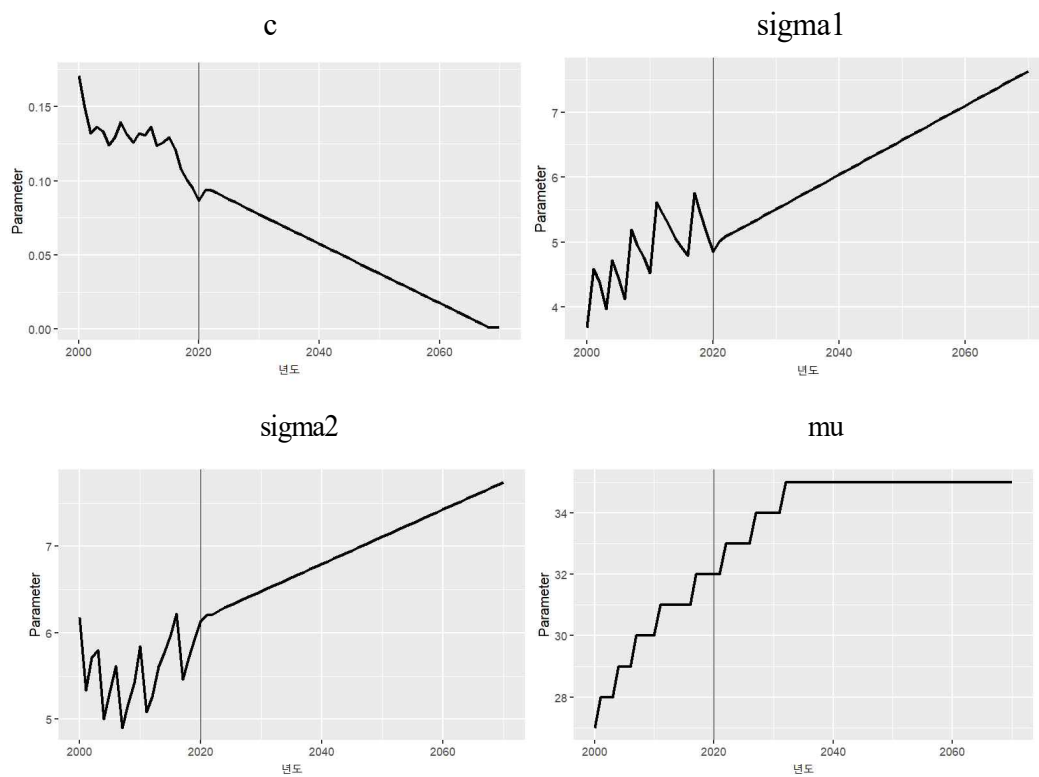
<그림 7.8> 로그 이차함수 모형의 모수 추계



<그림 7.9> 불연속 PK 모형의 모수 추계

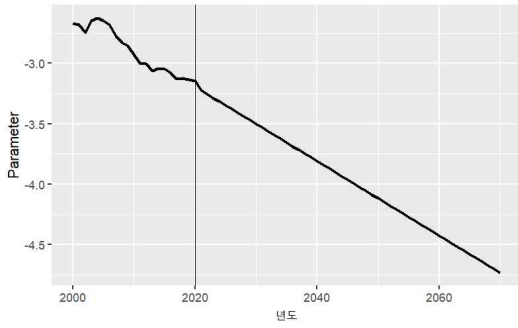


<그림 7.10> 연속 PK 모형의 모수 추계



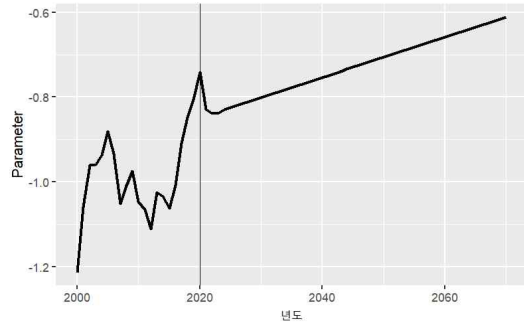
<그림 7.11> 일반화로그감마 모형의 모수 추계

b

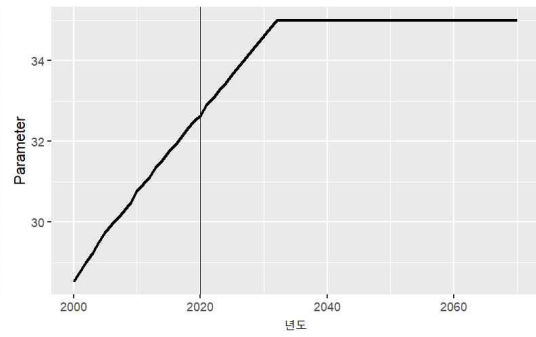
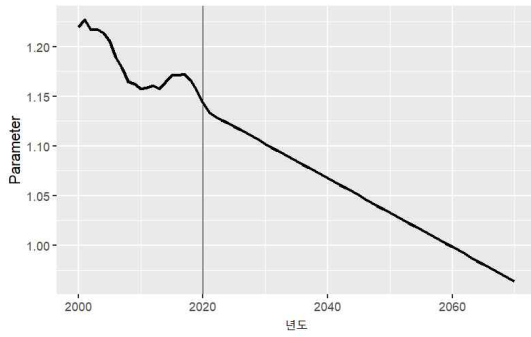


lambda

c

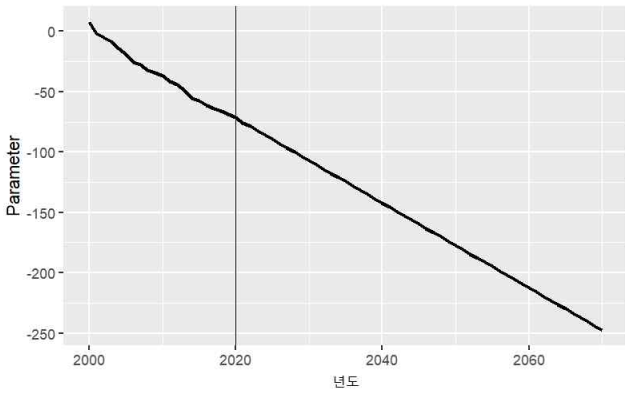


mu

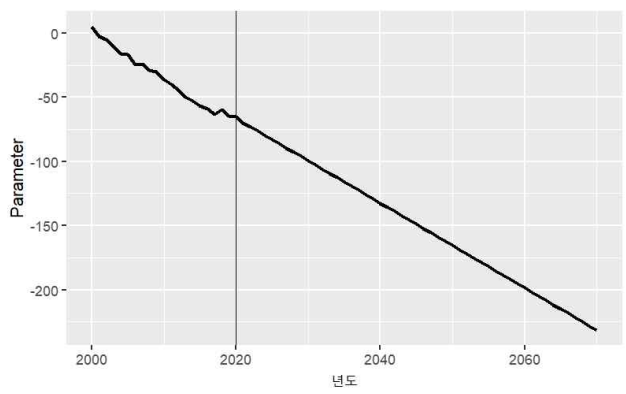


<그림 7.12> 사망률 $k(t)$ 추계

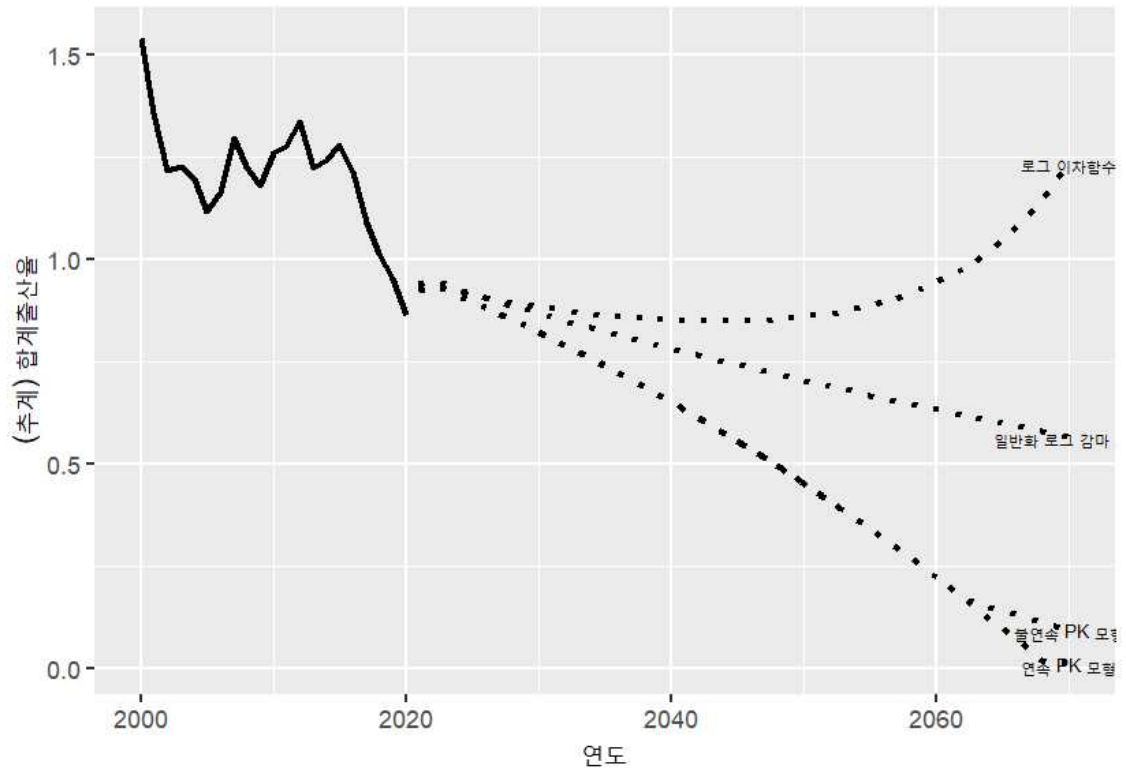
남자



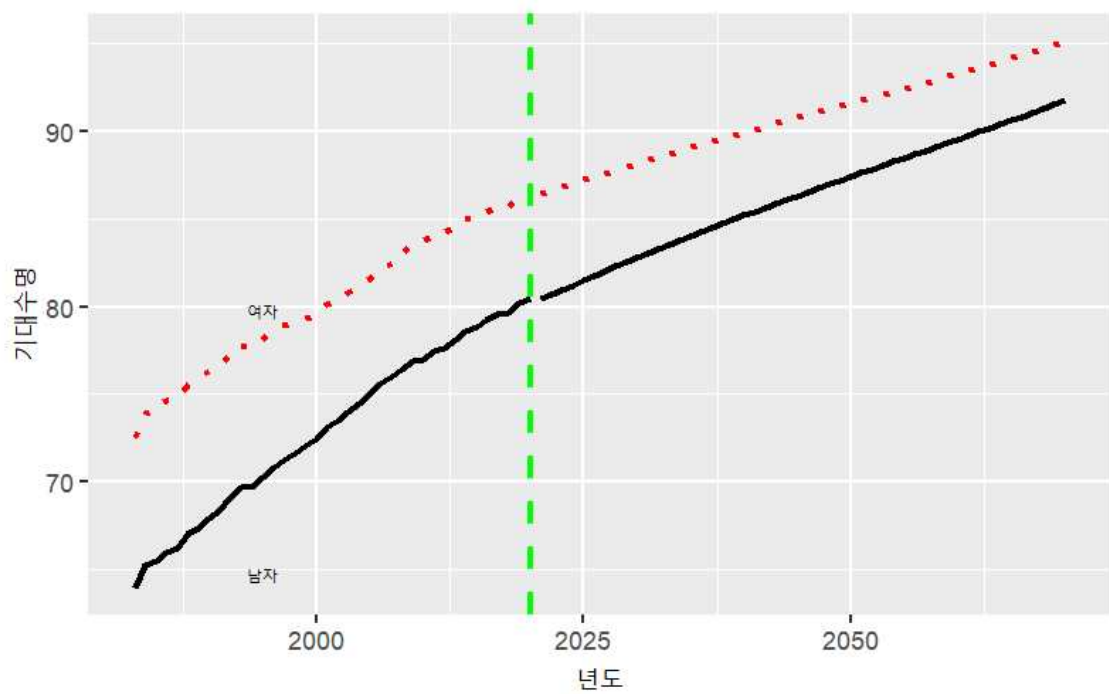
여자



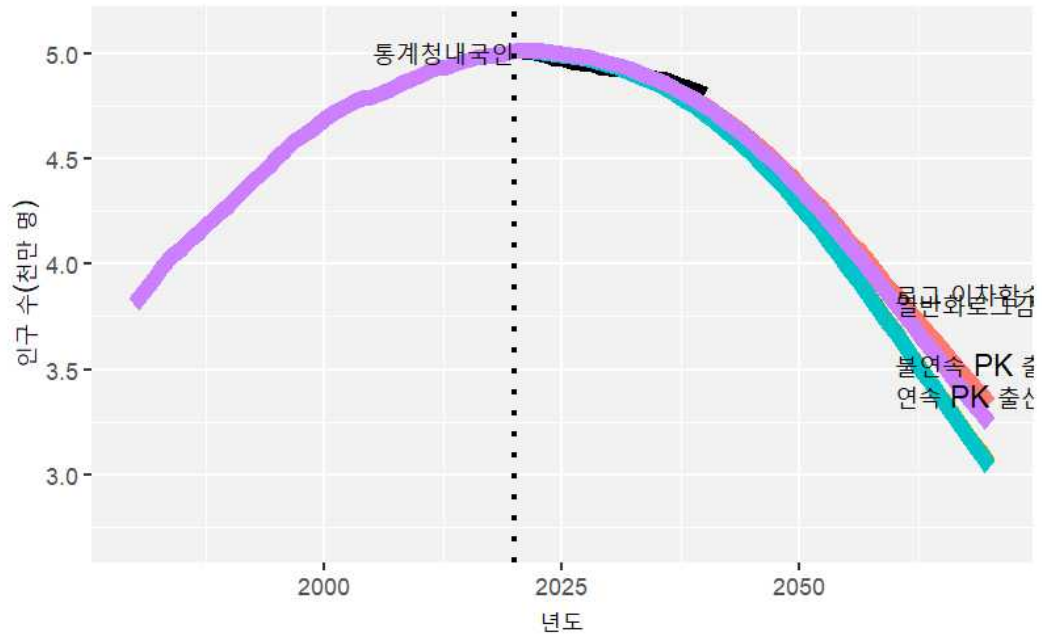
<그림 7.13> 합계출산율 추계



<그림 7.14> 기대수명 추계

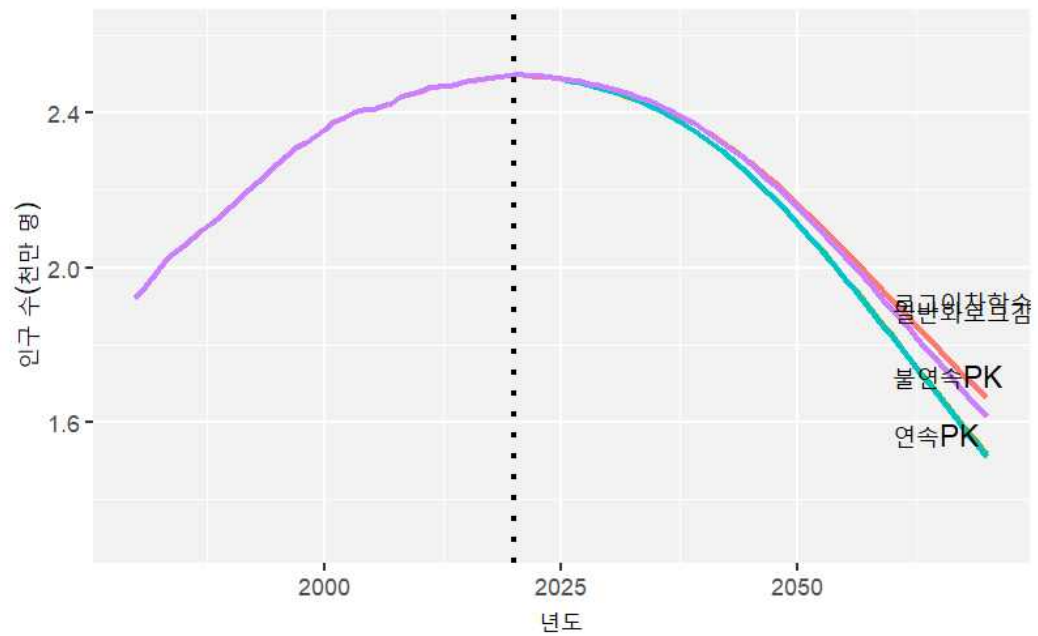


<그림 7.15> 경제사회적 변수를 활용한 인구 추계: 전체



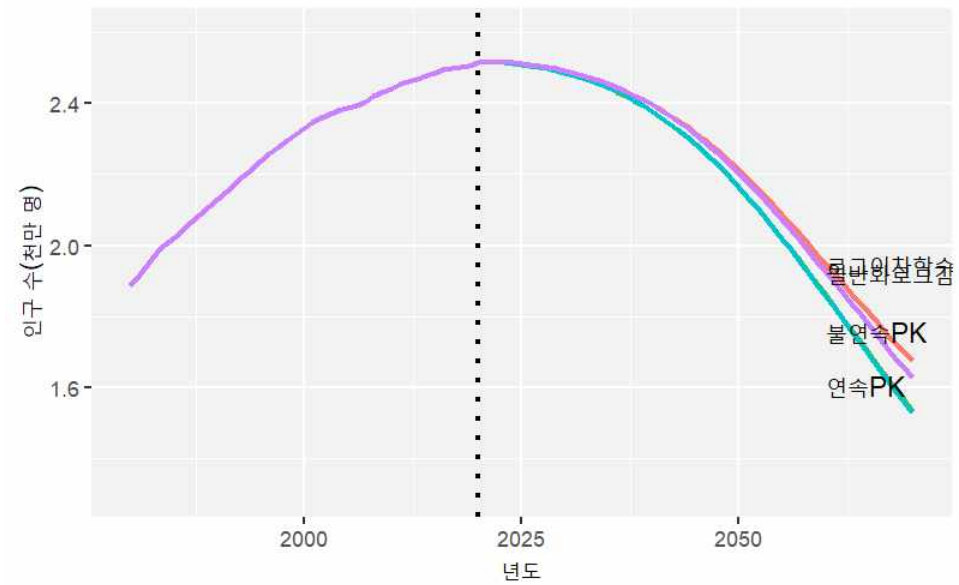
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 7.16> 경제사회적 변수를 활용한 인구 추계: 남자



주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

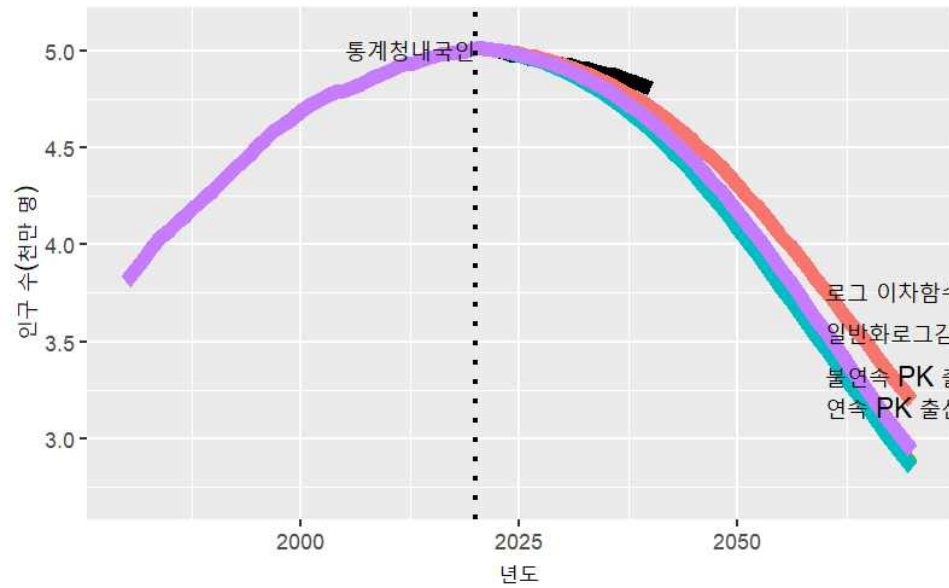
<그림 7.17> 경제사회적 변수를 활용한 인구 추계: 여자



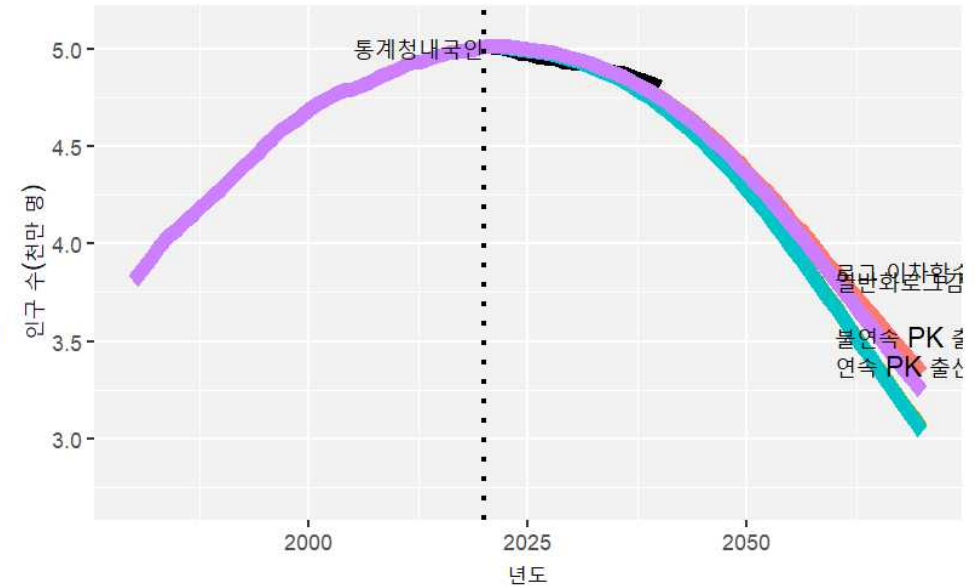
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 7.18> 경제사회적 변수 고려 전후 추계 인구 비교: 전체

A. 선형 변화



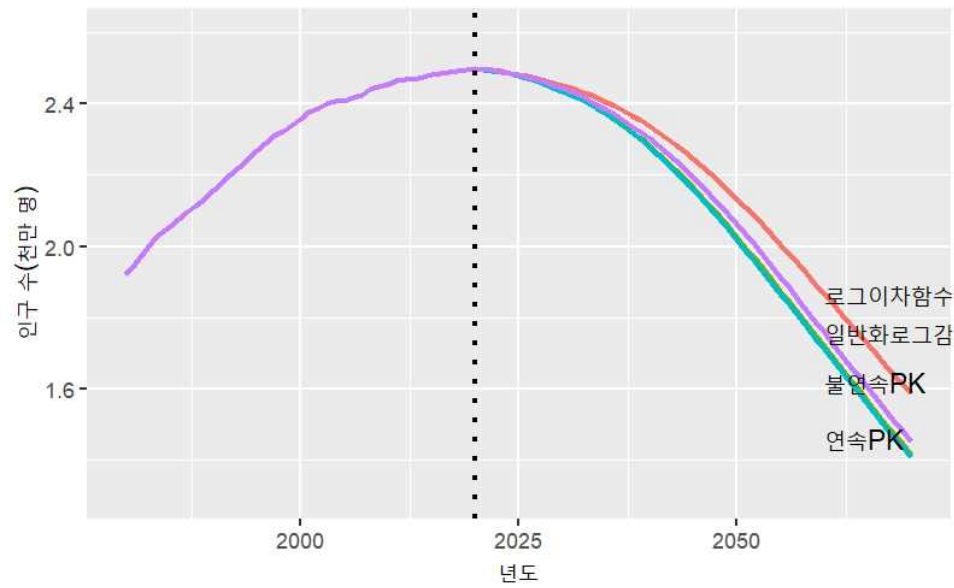
B. 경제사회적 변수 고려



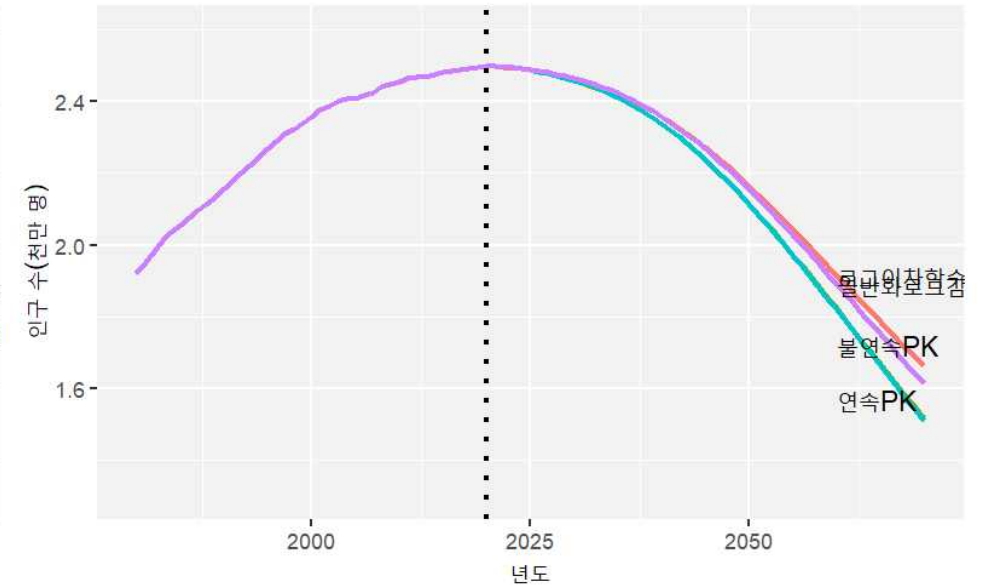
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 7.19> 경제사회적 변수 고려 전후 추계 인구 비교: 남자

A. 선형 변화



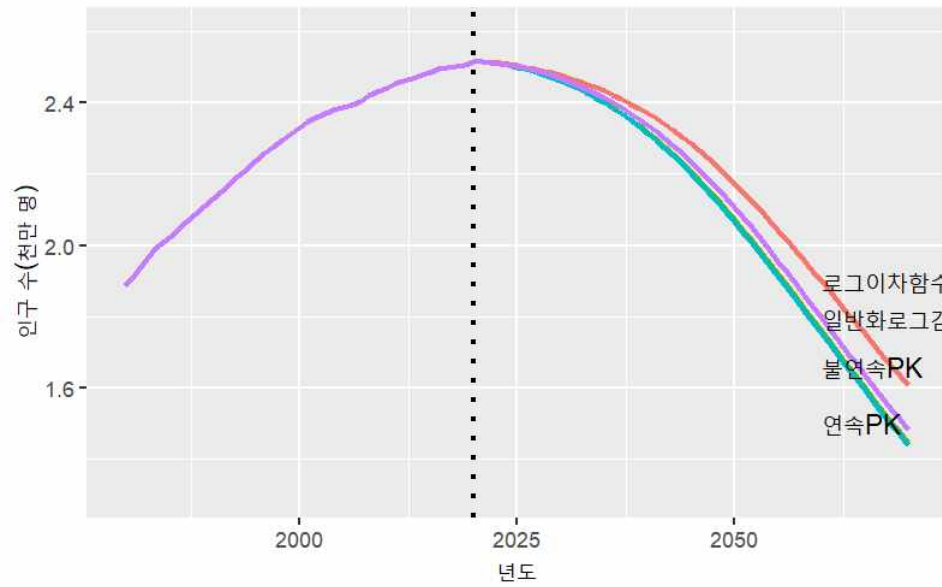
B. 경제사회적 변수 고려



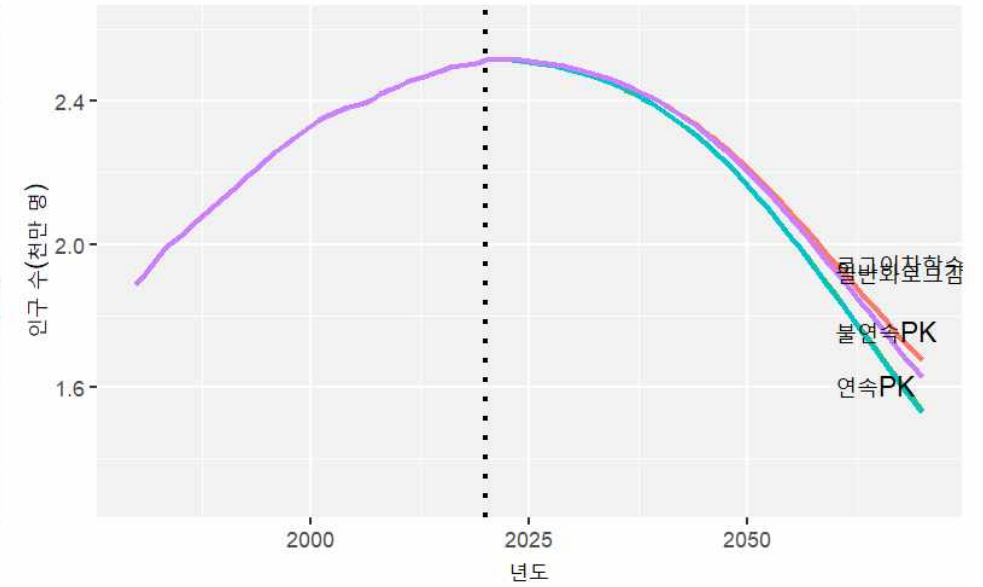
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 7.20> 경제사회적 변수 고려 전후 추계 인구 비교: 여자

A. 선형 변화

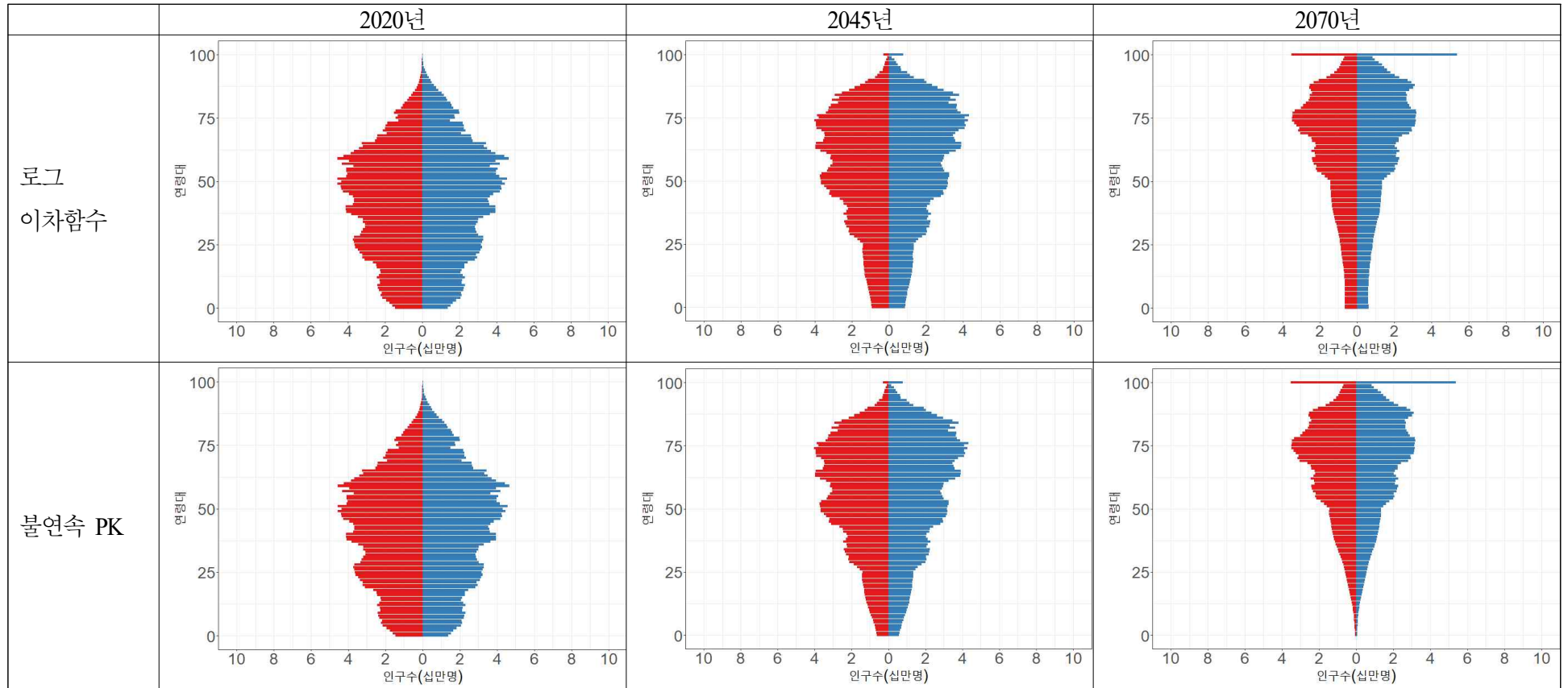


B. 경제사회적 변수 고려



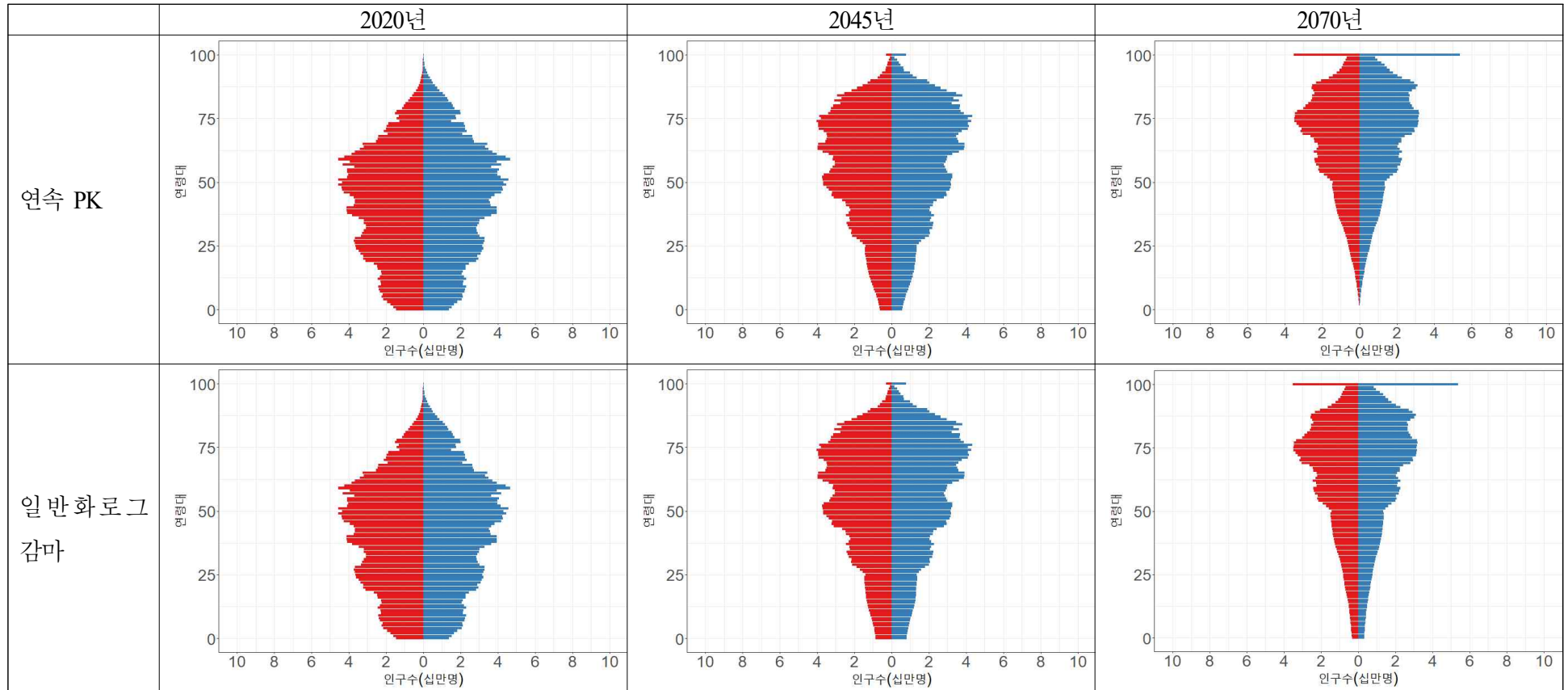
주. 불연속 PK 모형과 연속 PK 모형은 거의 값이 같아 덧붙여짐

<그림 7.21> 출산율 가정에 따른 추계 인구 피라미드



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

<그림 7.21> 출산을 가정에 따른 추계 인구 피라미드(계속)



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

8. 외국인 추계

가. 자료 및 방법

- (1) 특정 시점의 외국인 저량을 파악할 수 있는 자료로 세 가지가 있으나 2016년 이후 실질적으로는 2가지로 통합됨

□ 등록센서스 자료

- 2015년부터 2021년까지 0-4, 5-9,...,75+ 세의 남녀별 외국인 수 자료가 가용하며 여기에서는 2020년까지의 자료만을 활용
- 매년 11월 1일 기준 자료임
- 국내체류 3개월 이상인 대상자만 선정하며, 출입국 자료상으로 조사 대상이지만 해당년도 외국인등록자료에는 없고 다른 년도에 존재하는 외국인을 추가보완
- KOSIS 주제별 통계 -> 인구 -> 인구총조사 -> 인구부문 -> 총조사인구(2015년 이후)에서 찾을 수 있음
- 이 자료의 2020년 자료를 추계의 기준 인구로 사용하였으며, 0세부터 5세별로 작성되어 있고, 상한연령이 75+인 점에 주의할 필요가 있음

□ 행정안전부의 지방자치단체외국인주민현황

- 2016년부터 2020년까지 0-9, 10-19,..., 70+ 세의 남녀별 외국인 수 자료 (10세별에 주의)
- 이 자료의 조사 범위는 국내에 거주하고 있는 한국국적을 가지지 않은 자로 통계기준일 현재 3개월 이상 국내미체류자는 제외
- 2016년부터 등록센서스 결과에서 외국인주민을 추출하여 작성하고 등록센서스를 조사기준을 적용하였기 때문에 등록센서 자료와 동일함
- 2007년부터 2015년의 자료도 있으나 연령별 자료를 찾을 수 없어 2016년부터 사용 (이 시기의 자료는 행정안전부에서 자체 조사를 하였음)
- KOSIS 기관별 통계 -> 중앙행정기관 -> 행정안전부에서 찾을 수 있음

□ 법무부의 등록외국인 현황

- 2009년부터 2020년까지 0-4, 5-9,...,80+ 세의 남녀별 외국인 수 자료
- 등록외국인은 외국인이 입국한 날부터 90일을 초과하여 대한민국에 체류하기 위해 체류지를 관할하는 지방출입국·외국인관서의 장에게 외국인등록을 한 외국인임

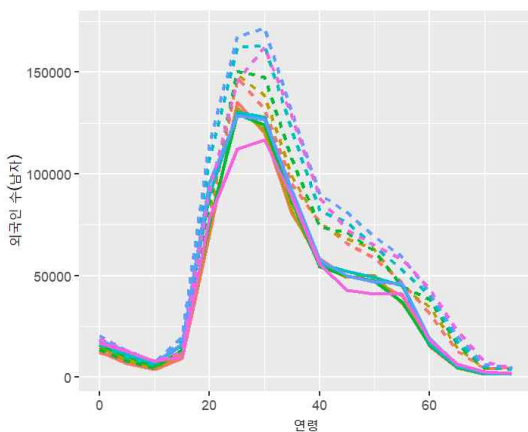
○ KOSIS 기관별 통계 -> 중앙행정기관 -> 법무부 -> 출입국자및체류외국인통계 체류외국인통계에서 얻을 수 있음

□ 등록센서스와 법무부 자료의 비교

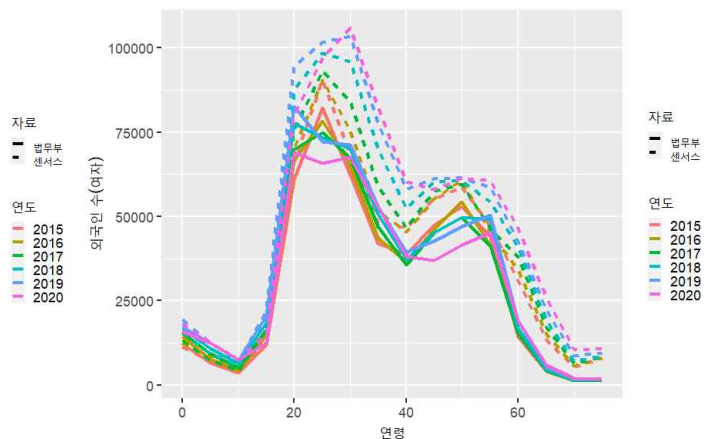
○ 다음 그림은 2015년부터 2020년 기간의 등록센서스와 등록외국인 현황 자료를 비교하여 보여고 있는데, 남자와 여자의 y-축 단위가 다르다는 점에 주의

<그림 8.1> 등록센서스와 등록외국인 현황 자료 비교

1) 남자



2) 여자



- 전반적으로 등록센서스 자료의 외국인이 등록외국인 현황 자료의 외국인보다 많음
- 두 자료 모두에서 3개월 이상의 체류 기간 조건을 적용하고, 등록센서스의 주요 자료원이 출입국 자료라는 점을 고려할 때 등록센서스의 외국인 수가 많은 것은 좀 의아하지만, 아마도 등록센서스의 설명에 적혀 있듯, “출입국 자료상으로 조사 대상이지만 해당년도 외국인등록자료에는 없고 다른 년도에 존재하는 외국인을 추가보완”했기 때문인 것 같음

(2) 외국인의 국제이동 수는 입출국 자료를 이용하여 2000년 이후 외국인의 성 및 연령별(각 세별) 입국자와 출국자로부터 구할 수 있음

□ 이 자료의 보고 대상은 우리나라 영토에 출입국하는 체류기간 90일을 초과하는 내국인 및 외국인이며 여기에서는 외국인만을 사용하였음

(3) 사망률 자료

□ 저량 자료와 유량 입출국 자료를 이용하여 2015년부터 2020년 사이 사망자 및 사망률

을 구할 수 있음

- 사망자를 구하기 위해 다음과 같은 균형방정식(balancing equation)을 사용

$$P_{t+5} = P_t + B_{t,t+5} - D_{t,t+5} + NM_{t,t+5}$$

- 위의 식을 활용하면,

$$D_{t,t+5} = P_t + B_{t,t+5} + NM_{t,t+5} - P_{t+5}$$

- 예컨대, 2020년의 남자 5-9세의 사망자 수는, 이 연령대에서 출생이 0일 수밖에 없기 때문에 ($B=0$), 2015년의 0-4세의 인구에서 두 관측년도 사이의 순국제이동자 수를 더한 후 2020년의 인구수를 빼주면 됨
- 국제순이동자 수는 2016년부터 2020년의 자료를 이용하였는데, 예컨대 2015년 5-9세 연령대 수가 2020년 10-14세 연령대 수로 변할 때, 국제순이동자는 2016년 6, 7, 8, 9, 10세, 2017년 7, 8, 9, 10, 11세, 2018년 8, 9, 10, 11, 12세, 2019년 9, 10, 11, 12, 13세, 2020년 10, 11, 12, 13, 14세이었던 사람들의 순이동자를 더하였음
- 다음 쪽의 표는 센서스 자료와 등록외국인 자료 및 국제이동 자료를 이용하여 이러한 방식으로 구한 추정 사망자 수 및 추정 사망률을 보여주고 있음
- 이 표에서 순국제이동 수는 두 자료 모두에 동일하게 적용하였기 때문에 센서스 자료 부분에서만 순국제이동 수를 보여줌
- 사망자 수는 음의 값을 가질 수 없으나 센서스 자료를 이용하였을 때는 남녀 모두에서 추정 사망자 수가 음의값을 가지는 경우가 많이 관측되므로 이 자료를 신뢰하기 어렵고, 따라서 따라 사망률을 구하지는 않았음
- 등록외국인 자료를 이용하였을 때는 추정 사망자 수가 음의 값을 가지는 경우가 없지 않으나 한 사례만 음의 값을 보여 추정 사망자 수를 이용하여 사망률을 구하였고, 이 때 사망률은 다음과 같은 식으로 구하였음

<표 8.1> 등록센서스와 등록외국인 자료를 이용한 사망자 추정

A. 남자

연령	센서스 자료 이용				법무부 등록외국인 자료 이용			
	2015 인구	2020 인구	2016-2020 순국제이동	추정 사망자	2015 인구	2020 인구	추정사망자	추정 사망률
0-4	11,885	18,920	1,471		12,592	16,748		
5-9	7,683	12,999	3,417	2,303	6,788	12,720	3,289	0.25988
10-14	4,868	7,839	3,945	3,789	3,756	7,943	2,790	0.37879
15-19	10,961	12,519	11,921	4,270	9,374	9,967	5,710	0.83218
20-24	78,919	91,499	75,639	-4,899	69,742	77,967	7,046	0.16134
25-29	147,001	145,827	44,910	-21,998	135,256	112,094	2,558	0.02814
30-34	131,891	162,583	-6,436	-22,018	119,910	116,555	12,265	0.09741
35-39	95,078	128,536	-17,445	-14,090	80,452	88,403	14,062	0.13501
40-44	76,341	88,112	-14,645	-7,679	58,519	55,444	10,363	0.15251
45-49	65,943	72,431	-7,564	-3,654	49,564	42,628	8,327	0.16465
50-54	58,868	65,035	-2,599	-1,691	47,625	41,053	5,912	0.13048
55-59	46,436	58,159	-5,771	-5,062	39,981	40,944	910	0.02055
60-64	31,421	43,433	-7,192	-4,189	15,295	18,151	14,638	0.50361
65-69	12,666	22,654	-7,085	1,682	4,496	6,325	1,885	0.17438
70-74	4,248	7,629	-1,390	3,647	1,479	2,608	498	0.14020
75+	4,454	4,444	1,505	1,309	1,436	1,810	1,174	0.71389

B. 여자

연령	센서스 자료 이용				법무부 등록외국인 자료 이용			
	2015 인구	2020 인구	2016-2020 순국제이동	추정 사망자	2015 인구	2020 인구	추정사망자	추정 사망률
0-4	11,388	18,244	1,605		12,279	16,044		
5-9	7,218	12,680	3,868	2,576	6,566	12,490	3,657	0.29529
10-14	4,564	7,529	4,017	3,706	3,575	7,626	2,957	0.41671
15-19	12,573	15,040	15,209	4,733	11,768	12,557	6,227	0.77201
20-24	66,253	79,989	56,084	-11,332	61,071	68,937	-1,085	-0.02689
25-29	90,391	97,154	21,569	-9,332	82,318	65,717	16,923	0.26695
30-34	68,685	105,916	8,083	-7,442	62,358	67,693	22,708	0.30275
35-39	48,195	82,721	8,461	-5,575	42,075	52,781	18,038	0.31333
40-44	46,565	60,255	7,253	-4,807	38,891	38,114	11,214	0.27969
45-49	55,273	57,939	6,074	-5,300	47,607	36,886	8,079	0.21323
50-54	58,603	61,336	1,042	-5,021	52,907	41,565	7,084	0.15888
55-59	47,318	60,566	-4,340	-6,303	43,965	45,132	3,435	0.07007
60-64	31,095	46,647	-5,502	-4,831	14,356	19,078	19,385	0.61498
65-69	13,482	25,760	-4,407	928	4,077	5,844	4,105	0.40644
70-74	5,561	10,508	212	3,186	1,317	1,997	2,292	0.75469
75+	7,885	10,740	3,442	-1,737	1,692	1,719	3,040	2.00264

$$m_{t,t+5} = \frac{D_{t,t+5}}{\frac{(P_t + P_{t+5})}{2}}$$

- 사망률을 구한 결과 믿기 어렵게 높은 사망률이 추정되었는데, 예컨대 한 연구에 따르면(정연·이나경, 2022), 이주노동자의 사망만인율(만명당 사망자 수)이 1.39라는 추정
에 비추어 볼 때, 남자 15-19세의 사망률(여기서는 1명당 사망자 수로 해석 가능)
0.832는 지나치게 높은 사망률을 보여주는 것임
- 국회예산정책처에서는 MDIS 사망원인통계의 국적구분 자료가 있어 이를 활용하기를
제안하였으나 연구진이 찾아본 바로는 국적구분이 출생한국인, 귀화한국인으로 되어
있어 외국인 여부를 알 수는 없음
- 정연·이나경(2022)의 연구에 의하면 이주노동자의 사망만인율은 1.39인데 전체 취업
자의 사망만인율은 0.77로 이주노동자의 사망만인율이 1.805배 높았음
 - 이러한 연구 결과를 이용하기 위해 2019년 기준 외국인 입국자의 체류자격을 구분하
여, 단기취업, 구직, 기술연수, 선원, 특정활동, 비전문취업, 관광취업, 방문취업을 이주
노동자와 관련된 항목으로 선정하여 해당 항목의 외국인 입국자 수 108,494명을 구함
 - 체류자격에서 공무와 사증면제, 관광통과를 제외한 전체 입국자 수를 344,572명으로
추정할 때, 이주노동자는 $108,494/344,572=0.315$ 비율인 것으로 파악
 - 외국인 중 이주노동자가 아닌 경우 내국인과 사망률이 동일하다고 가정하고, 이들 이
주노동자의 사망률은 1.805배 높다고 가정하여 내국인 사망률*0.685+내국인 사망률
*0.315*1.805=내국인 사망률*1.254로 외국인 사망률을 추계
 - 이러한 외국인 사망률은 20-64세에 적용하고, 나머지 연령대에서는 내국인과 동일한
사망률을 적용
 - 사망률 자료를 5세별로 맞추고 상한연령이 75+가 되도록 맞춰 인구 추계를 진행하기
위해 이와 동일한 연령 구분에 맞는 생명표를 작성할 필요가 있는데, 기존의 MortCast
패키지에 있는 life.table 함수는 0세 및 1-4세를 구분하기 때문에 이 함수를 적용할 수
없음
 - 따라서 외국인 추계 프로그램에 lifet5라는 이름의 별도의 생명표 프로그램을 작성하
였으며, 이 프로그램은 nax라고 알려진 특정 연령대의 사망자의 평균 생존연수를 2.5
년으로 설정한 것이 특징이며, 나머지는 Preston et al. (2001)의 제안을 따랐음

- 이 함수를 2020년 자료에 적용했을 때, 내국인 남자와 여자의 기대수명은 83.5세 및 91.9세 였으나 외국인 남자 및 여자의 기대수명은 82.7세 및 91.4세로 크게 차이가 나지는 않는데, 이는 20-64세의 연령별 사망률이 낮기 때문인 것으로 파악됨

□ 사망률 자료는 2020년의 자료를 활용하여 이 사망률이 변하지 않는다고 가정

- (4) 추정 출산율은 내국인의 출산율과 외국인의 출산율 비교에 대한 연구를 활용하여, 내국인 출산율의 일정 비율로 구함

□ 여기에서는 Kim(2018)의 연구를 활용

- 이 연구는 2012년 전국 다문화가족 실태조사 자료와 2012년 전국 출산력 및 가족보건·복지실태조사 자료를 결합하여 분석한 후 혼인부부의 경우 첫째 자녀 출산율을 비교했을 때 내국인은 연간 0.403명을 출산하며, 내국인 남편-외국인 아내의 연간 출산율은 0.340명, 내국인 아내-외국인 남편의 연간 출산율은 0.190명, 부부 모두 외국인의 연간 출산율은 0.206명이라는 결과를 보고 (p. 53)
- 외국인 중 내국인 남편-외국인 아내에 속하는 비율을 추정해 보면, 2020년 기준 15-49세 외국인 여자의 수가 499,014였고, 2020년 결혼이민자 중 여자의 수가 137,878이므로, 약 27.6%로 추정가능
- 내국인 아내-외국인 남편의 수가 무시할 정도로 작다는 가정하에, 외국인의 (유배우) 출산율은 $0.340 \times 0.276 + 0.206 \times 0.724 = 0.243$ 로 추정 가능
- 이 연구는 혼인 여성의 결과를 보여주는 것이므로 여자 인구 중 유배우자의 비율을 구한 후 출산율을 곱해주는 과정이 필요
- 이를 위해 2020년 등록센서스 자료를 이용하여 아래와 같은 연령별 내외국인의 유배우율(C와 G 열)을 계산
- D열 출산율은 $0.243 \times C / 100$ 으로 구하였고, H열 출산율은 $0.403 \times G / 100$ 으로 구하였는데, 이는 무배우자 여자의 출산율은 0이고 유배우자 여자의 출산율은 앞선 수치를 따른다는 가정임
- 이후 두 출산율의 비인 I 열을 계산
- 2020년 내국인 출산율 자료를 이용하여 5세별 출산율에 I 열의 비를 적용하여 외국인 연령별 출산율을 계산하였고, 연령별 출산율을 더한 후 (5세별 연령별 출산율이므로) 5를 곱한 합계출산율을 구하면 1.257정도가 도출됨(내국인의 합계출산율은 0.856임)

<표 8.2> 외국인과 내국인의 유배우자 비율을 활용한 출산율 비 계산

	외국인				내국인				비 I=D/H
	무배우자 A	유배우자 B	백분율 C	출산율 D	무배우자 E	유배우자 F	백분율 G	출산율 H	
15 ~ 19세	14,093	968	6.4	0.016	1,162,333	784	0.1	0.000	57.495
20 ~ 24세	57,627	20,767	26.5	0.064	1,494,483	28,197	1.9	0.007	8.626
25 ~ 29세	48,763	49,115	50.2	0.122	1,330,878	278,457	17.3	0.070	1.749
30 ~ 34세	32,153	73,513	69.6	0.169	699,736	753,196	51.8	0.209	0.809
35 ~ 39세	18,630	64,403	77.6	0.188	484,811	1,267,689	72.3	0.292	0.647
40 ~ 44세	11,808	48,550	80.4	0.195	410,043	1,438,737	77.8	0.314	0.623
45 ~ 49세	10,484	47,402	81.9	0.199	456,864	1,614,852	77.9	0.314	0.633

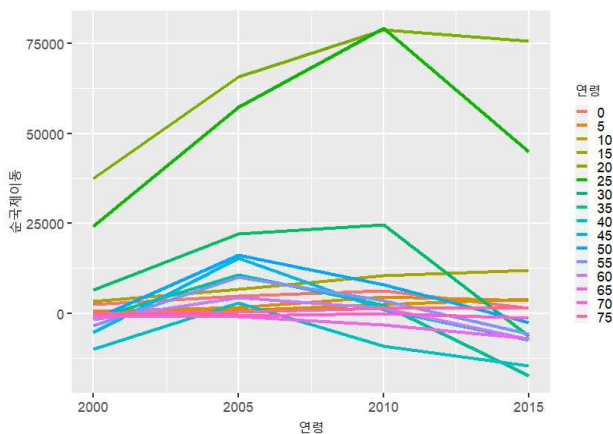
□ 출생아의 성비는 내국인 추계와 마찬가지로 1.05329를 사용

(5) 외국인의 국제이동 추계

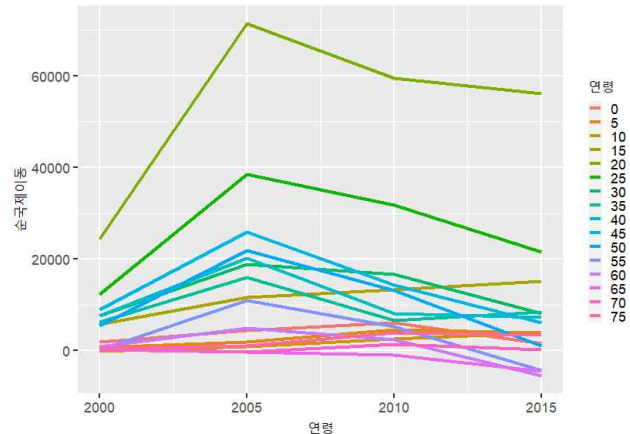
□ 2001년부터 2020년의 각세별 외국인 국제이동 자료를 이용하여 5년간 순국제이동을 파악하여 다음의 그림에 제시하였음

<그림 8.2> 외국인 국제이동 (2001-2020)

1) 남자



2) 여자



○ 이 그림에서 연도는 기준이 되는 연도를 의미하여, 예컨대 2000은 2001년부터 2005년까지의 순국제이동을 보여줌

○ 앞서도 설명한 것처럼, 2000년의 10-14세는 2001년의 11, 12, 13, 14, 15세, 2002년의 12, 13, 14, 15, 16세, 2003년의 13, 14, 15, 16, 17세, 2004년의 14, 15, 16, 17, 18세, 2005년의 15, 16, 17, 18, 19세를 순국제이동을 더한 수치임

□ 이러한 순국제이동자 수를 ARIMA 모형으로 추계하려고 하면, auto.arima 함수에서

여자 10-14세 및 여자 65-69세를 제외한 모든 연령대에서 0의 값으로 추계하는 결과를 보여주어 신뢰하기 어려움

- 이 자료에서 각 연령대에 관측값이 4개에 지나지 않기 때문에 이들의 평균을 이용하여 추계값으로 활용

<표 8.3> 외국인 국제순이동 추계 값

	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75+
남자	3,736	2,536	1,726	8,088	64,418	51,375	11,648	-1,498	-7,714	860	4,960	1,109	-782	-3,080	-633	746
여자	3,558	2,830	1,880	11,476	52,819	25,993	12,823	9,361	10,815	13,834	10,377	3,061	689	-1,252	375	2,280

나. 추계 결과

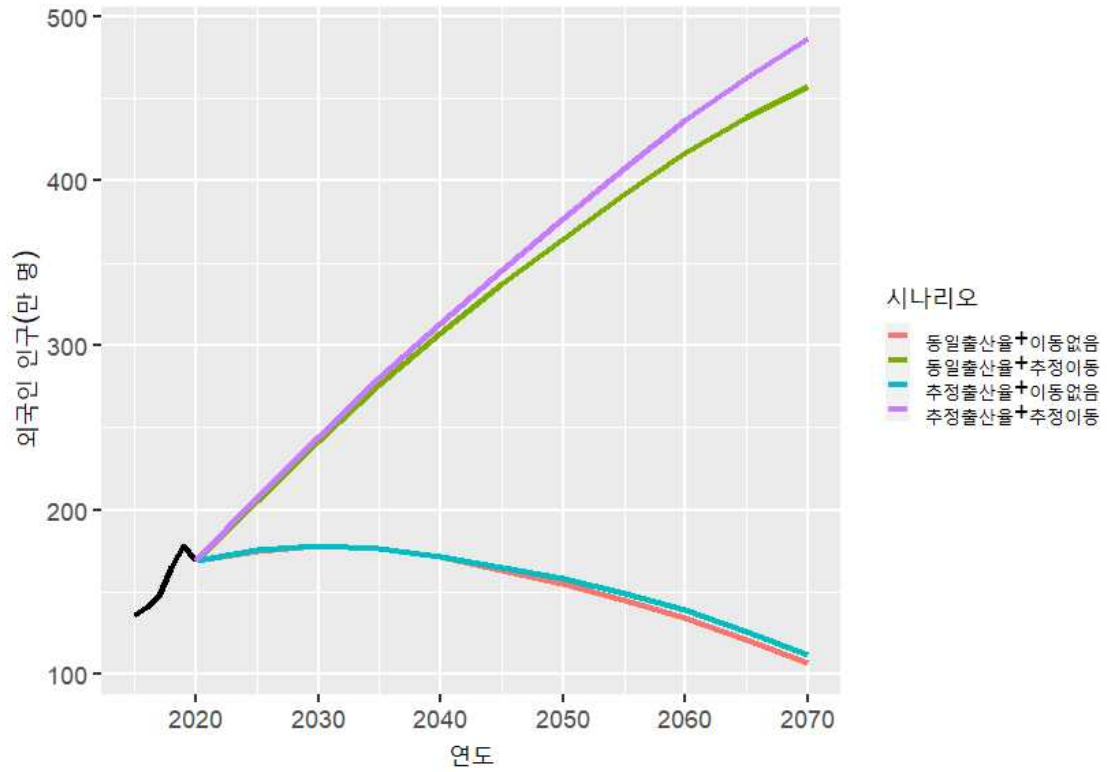
(1) 다음과 같은 시나리오를 작성하여 추계 진행

- 기준 인구는 2020년의 5세별 센서스 외국인 인구
- 사망률은 위에서 구한 2020년 외국인 사망률 적용
- 출산율은 1) 위에서 구한 외국인의 출산율을 적용할 때와 2) 내국인의 출산율을 적용할 때(동일 출산율)로 구분
- 외국인 순국제이동은 1) 4개 관측값의 평균을 사용할 때와 2) 순국제이동이 0일 때(이동 없음)로 구분

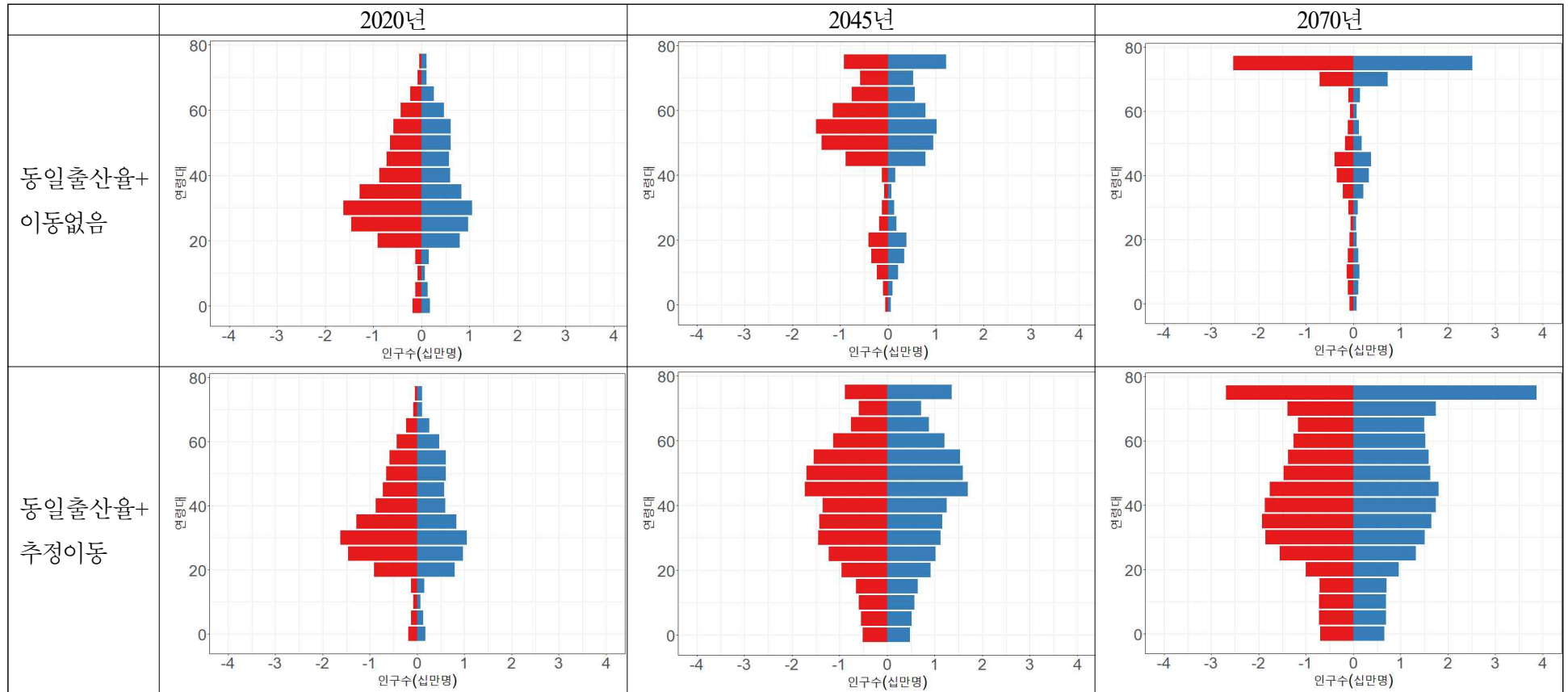
(2) 추계 결과

- 다음 그림은 2015년부터 2020년 사이 관측 외국인 수와 이후 50년간 시나리오별 추계 인구를 제시
- 이 그림을 보면 출산율 시나리오에 따른 변화는 매우 미미하지만 국제이동 가정에 따른 변화는 작지 않은 것으로 판단함
- 그 다음의 그림들은 시나리오에 따른 2020년, 2045년, 2075년의 인구 피라미드를 보여줌

<그림 8.3> 외국인 추계 결과

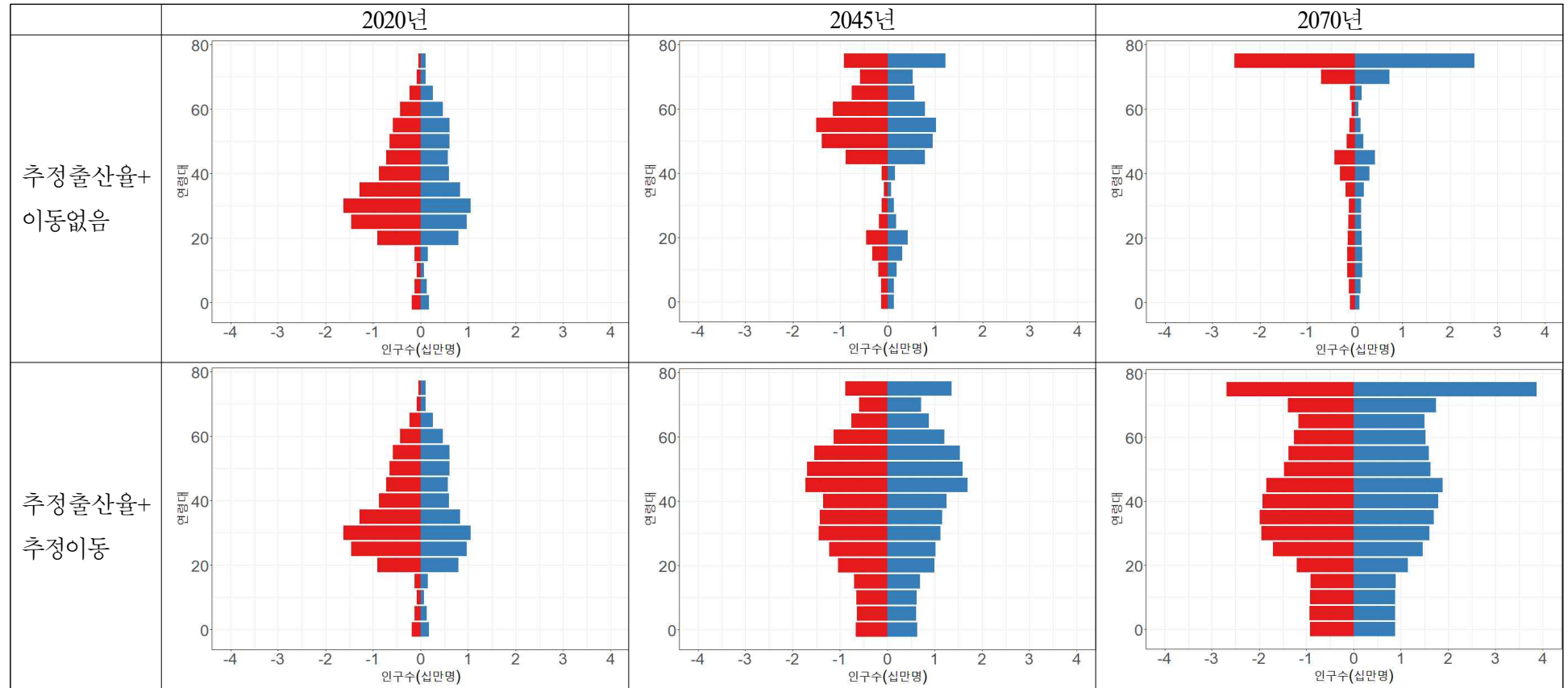


<그림 8.4> 외국인 추계 인구 피라미드: 내국인과 동일출산율 가정



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

<그림 8.5> 외국인 추계 인구 피라미드: 추정 출산율 가정



주: 왼쪽은 남자, 오른쪽은 여자. 2020년은 관측값임

9. 시사점 및 향후 연구 방향

- 이 연구에서는 국내거주 내국인 및 외국인에 대한 추계를 실시하였으며, 그 연구 성과도 적지 않으나 다양한 부분에서 개선의 여지가 있음
- 선행 연구들에 근거하여 인구및주택총조사, 출생 및 사망관련 인구동향조사, 통계청의 추계인구, 법무부의 외국인 출입국 자료 등을 이용하여 내국인과 외국인을 구분하는 작업을 진행한 점에서 다양한 자료를 정리하고 추계에 활용하려고 하였으나 여전히 더해져야 할 연구들이 있어 이를 간략하게 제시하고자 함
 - 인구및주택총조사나 사망 자료와 관련하여 고령자들의 각세별 자료가 없는 경우 로그-로그 모형으로 연령별 분포를 추정하였으나 이에 다른 대안이 있는지에 대한 탐색이 필요함
 - 외국인 추계에 있어 등록센서스를 기준 연도로 사용하였으나, 등록센서스의 기준 시점이 11월 1일이라는 것을 고려할 때 7월 1일이 기준이 되는 인구로 변환하는 작업을 통해 개념적으로 정합적인 연구를 진행할 필요가 있음
 - 외국인 출산율과 사망률 추계를 특정 연구들에 기대 추정해 보았으나, 좀 더 많은 연구들을 참조하여 보다 균형잡힌 추계를 진행하고, 단순히 하나의 연령별 스케줄이 미래에도 지속된다는 가정보다는 특정하게 변화하는 현실을 반영하여 이들 인구학 모수들도 변화하는 모형을 도입할 필요가 있음
- 출산율 추계에서 4가지 모형을 살펴보고 적용한 점에서 이전 연구들에 비해 차별점이 있으나, 모수들을 추계할 때 다양한 문제점들이 나타났으며 이를 개선할 수 있는 방법을 모색할 필요가 있음
 - 예컨대, 로그 이차함수 모형에서 a 값이 양의 값으로 바뀌는 점이 존재한다거나, PK 모형에서 c 의 값들이 음수가 나오는 것은 가능하지 않은 것들이므로 이런 문제들이 나타나지 않는 모형들, 예컨대 점근선이 있는 로지스틱 모형을 변형하는 모형들을 개발할 필요가 있음
- 사망률 추계 모형으로 LLG모형을 활용하였다는 점에 연구의 의의가 있으나, 이 모형이 한국 상황에 적합한가에 대한 의문이 제기되어 좀 더 현실적합성이 있는 모형 개발이 필요한 것으로 판단됨
- 인구이동의 경우도 최근 9년간의 자료만을 사용하여 최솟값과 최댓값을 뺀 나머지로 평균을 취하여 추계하였으나, 인구이동율이 일정할 것이라고 가정하는 것은 비판의 소지가 많

이 있으므로 이를 개선하는 연구 방법 개발이 시급

- 여기에서는 출산과 사망에 영향을 주는 경제사회적 변수로 17가지의 변수를 고려하였으나 이 이외에도 다양한 변수들이 출산과 사망에 영향을 줄 수 있음
 - 예컨대 남성과 여성의 임금격차나 어린이집과 같은 보육시설의 증가 및 국공립 보육시설의 비율, 최근 이슈가 되었던 공교육 연령 등이 출산에 영향을 줄 수 있으나 이 보고서에서는 이러한 변수들을 고려하지 못하였음
 - 따라서 향후 연구에서는 출산율과 사망률에 영향을 주는 더 많은 변수들을 찾아 실제 검정해 볼 필요가 있음
 - 하지만 여러 경제사회적 변수 자료가 2000년 이후부터 가용하다는 점이 더 한계점으로 판단함
 - 이러한 문제로 인해 고차원의 lag을 활용하는 VAR 모형을 사용할 수 없는 한계가 있어 이를 이러한 문제를 완화할 방안을 강구할 필요가 있음
- 이러한 한계들에도 불구하고, 이 연구는 국내거주 내국인 및 외국인을 추계함으로써 국회의 재정 및 세금 계산에 필요한 인구 수를 추계한 첫 연구라는 점에서 커다란 의의를 지님
 - 특히 통계청의 내외국인 추계는 전체 추계를 진행한 후 일정 비율로 외국인과 내국인을 나누는 형태로 내외국인 추계가 진행되는 한계가 있으나, 이 연구에서는 내외국인을 별도로 추계함으로써 보다 이론정합적인 추계를 진행한 것임
- 또한 다양한 경제사회적 변수들을 찾아 이 변수들과 인구학 모수들과의 관계를 살펴보고 추계에 적용한 것은 작지 않은 의미를 지님
- 국회예산정책처의 인구관련 연구부서는 이 연구를 통해 개발한 다양한 모형과 방법론을 활용해, 좀 더 현실적합적인 가설을 설정한 후 인구를 추계하는 모형을 발전시킬 수 있을 것임

참고문헌

- 계봉오. (2013). 제1장. 출산과 재생산. 계봉오·김중백·김현식·이민아·이상림·조영태, 인구나 보건의 사회학: 건강한 사회를 위하여. 다산출판사. pp. 3-20.
- 국회예산정책처. (2019). 2019~2028년 8대 사회보험 재정전망. 국회예산정책처.
- 국회예산정책처 경제분석국 인구전략분석과. (2018). 우리나라 저출산의 원인과 경제적 영향. 국회예산정책처.
- 권태환·김두섭. (2011). 인구의 이해. 서울대학교 출판문화원.
- 기획재정부. (2017). 제25차 재정전략협의회 개최.
- 김경수·김상미. (2021). NABO 내국인 인구 시범추계: 2020~2040년. 국회예산정책처.
- 김경아. (2017). 결혼 여성의 출산율에 영향을 미치는 요인 분석. 문화와 융합, 39(6), 895-924.
- 김기민. (2019). 사업체의 근로시간 실태 현황: 2015년 사업체패널조사의 근로시간 설문을 활용하여. 패널브리프 제17호 (2019-01).
- 김도균. (2021). 지역별 출산율 변동에 관한 연구: 패널데이터 분석. 한국컴퓨터정보학회논문지, 26(5), 77-86.
- 김승권·김유경·김혜련·박종서·손창균·최영준·김연우·이가은·윤아름. (2012). 2012년 전국 출산력 및 가족보건·복지실태조사. 연구보고서 2012-54. 한국보건사회연구원.
- 김지범·강정환·김석호·김창환·박원호·이윤석·최슬기·김솔이. (2022). 한국종합사회조사 2003-2021. 성균관대학교 출판부.
- 김현식. (2013). 제3장. 사망률변천과 차별사망률. 계봉오·김중백·김현식·이민아·이상림·조영태, 인구나 보건의 사회학: 건강한 사회를 위하여. 다산출판사. pp. 39-60.
- 김현식. (2017). 전문가 판단법을 활용한 출산력 추계. 통계연구, 22(2), 1-25.
- 김현식·계봉오·김현태. (2016). 장래인구추계 출산력·사망력 추계방법 개선 연구. 통계청연구보고서. 한국인구학회·통계청
- 김현식·우해봉·안재혁. (2017). 교육수준별 장래인구추계 개발. 동국대학교 산학협력단·통계청.
- 김현식·유삼현·안재혁. (2019). 장기인구전망 모형 구축. 국회예산정책처.
- 박유성·김기환·김성용. (2010). 우리나라 공식인구의 신뢰성 및 문제점에 대한 고찰. 조사연구, 11(2), 71-95.
- 박유성·김미리·김성용. (2013). 확률적 출산율 모형과 한국의 미래인구 구조. 조사연구, 14(3), 49-78.
- 박유성·이선미. (2017). 2015 등록센서스와 장래인구추계의 신뢰성에 대하여. 조사연구, 18(2), 1-37.
- 박은옥. (2013). 시군구 단위 지역사회 심뇌혈관 질환 표준화 사망률과 관련 요인 분석. 보건과 사회과학, 34(2013.12), 257-271.
- 박종서·임지영·김은정·변수정·이소영·장인수·조성호·최선영·이혜정·송지은. (2021). 2021년도 가족과 출산조사. 연구보고서 2021-50. 한국보건사회연구원.
- 법무부. (2018). 제3차 외국인정책 기본계획 (2018년 ~ 2022년). 법무부.
- 법무부·통계청. (2018). 2018년 이민자 체류실태 및 고용조사 지침서. 법무부·통계청.
- 서우석·이명진. (2008). 주민등록자료와 인구주택총조사의 연령별 비교에 관한 연구. 조사연구, 9(3), 23-53.
- 성재민. (2005). 한국노동패널조사에서 나타난 근로시간의 특징과 법정근로시간 단축의 포괄범위. 노동리뷰 2005년 4월호.
- 신창우·이삼식·이난희·최효진. (2012). 출산력 시계열 자료 구축 및 분석. 연구보고서 2012-47-26. 한국보건사회연구원.
- 양경진·김형석. (2008). 행정자료를 이용한 외국인 추계방법 개발. 통계개발원.
- 양경진. (2009). 제3장 인구주택총조사 외국인 조사방법 개선. 2009 상반기 연구보고서 제Ⅱ권, pp. 1-58. 통계개발원.
- 오유진·박성준. (2008). 저출산의 경제학적 분석. 한국경제학보, 15(1), 91-112.
- 우해봉·양지윤·조성호·안형석. (2016). 인구추계 방법론의 현황과 평가. 한국보건사회연구원.
- 이미숙. (2001). 40대 남성사망률: 사회적 관련 요인에 대한 탐색. 한국사회학, 35(4), 189-212.
- 이삼식·박종서·이소영·오미애·최효진·송민영. (2015). 2015년 전국 출산력 및 가족보건·복지실태조사. 연구보고서 2015-31. 한국보건사회연구원.
- 이소영·김은정·박종서·변수정·오미애·이상림·이지혜. (2018). 2018년 전국 출산력 및 가족보건·복지 실태조사. 연구보고서 2018-37. 한국보건사회연구원.

- 이원재. (2019). 사망률에 대한 미세먼지, 인간개발지수, 소득불평등도의 영향 대한보건연구, 45(3), 83-89.
- 이원진·정해식·전지현. (2019). 소득조사 마이크로데이터 비교 분석: 가계동향조사와 가계금융복지조사를 중심으로. 한국보건사회연구원 연구보고서(수시) 2019-13.
- 전광희·김태현·조영태. (2005). 장래인구추계를 위한 출산·사망 예측모형의 개발에 관한 연구. 통계청·인구학회.
- 정연·이나경. (2022). 이주노동자 산업안전보건 현황과 정책 과제. 보건복지포럼, 2022. 02, 51-65.
- 중앙선거관리위원회. (2017). 제19대 대통령선거총람 (2017.5.9. 시행). 경기: 중앙선거관리위원회.
- 통계청. (2017a). 2015~2065 장래인구추계: 2016년 12월 추계. 대전: 통계청.
- 통계청. (2017b). 2015~2045 장래인구추계 시도편: 2017년 6월 추계. 대전: 통계청.
- 통계청. (2018) 사망원인통계 통계정보 보고서. 대전: 통계청.
- 통계청. (2022a). 각종 통계 자료. kosis.kr
- 통계청. (2022b). 2020~2070 장래인구추계(2020 12월 추계). 통계청.
- 통계청. (2022c). 2021년 장래인구추계를 반영한 내외국인 인구전망: 2020~2040년. 통계청 보도자료.
- Aldridge, R.W. et al. (2019) Global patterns of mortality in international migrants: a systematic review and meta-analysis. *Lancet* 392:2553-66.
- Baker, B. (2018). Population estimates - Illegal alien population residing in the United States: January 2015." Office of Immigration Statistics, U.S. Department of Homeland Security.
- Bijak, J. and Wisniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society*, 173(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.
- Caswell, H. (2001). *Matrix population models: Construction, analysis, and interpretation*, 2nd ed. Sunderland, MA: Sinauer Associates, Inc. Publishers.
- Cho, L.-J., Retherford, R. D., and Choe, M. K. (1986) *The Own-Children Method of Fertility Estimation* Honolulu, HI: The East-West Center.
- Jandl, M. (2004). The estimation of illegal migration in Europe. *Migration Studies*, 153, 141-155.
- Kim, H. S. (2018). Fertility differentials between Korean and international marriage couples in South Korea. *Asian Population Studies*, 14(1), 43-60
- Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American Statistical Association*, 87(419), 659 - 671.
- Lemaitre, G. and Thoreau, C. (2006). Estimating the foreign-born population on a current basis. OECD, Paris.
- 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.
- OECD. (2005). Counting immigrants and expatriates in OECD countries - a new perspective" *Trends in International Migration: SOPEMI - 2004 Edition*, OECD, Paris.
- Pfaff, B., and Stigler, M. (2021). Package 'vars'. <https://cran.r-project.org/web/packages/vars/vars.pdf>
- Passel, J. S. and Cohn, D. (2018). U.S. unauthorized immigrant total dips to lowest level in a decade: Pew Research Center. November 27, 2018. https://www.pewhispanic.org/wp-content/uploads/sites/5/2019/03/Pew-Research-Center_U-S-Unauthorized-Immigrants-Total-Dips_2018-11-27.pdf
- Peristera, P., & Kostaki, A. (2007). Modeling fertility in modern populations. *Demographic Research*, 16(6), 141-194.
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Malden, MA: Blackwell Publishers Inc.
- Smith, K. E., & Favreault, M., M. (2013). *A primer on modeling income in the near term, version 7 (MINT7)*. Washington, D.C.: The Urban Institute.

- United Nations. (1983). Manual X: Indirect Techniques for Demographic Estimation. Department of International Economic and Social Affairs. Population Studies. No 81.
- Van Hook, J., Zhang, W., Bean, F. D., & Passel, J. S. 2006. Foreign-born emigration: A new approach and estimates based on matched CPS files. *Demography*, 43(2), 361-38

