

# Stochastic population projections on an uncertainty for the future Korea

Jinho Oh<sup>a,1</sup>

<sup>a</sup>Department Mathematical Sciences, HanBat National University

(Received November 21, 2019; Revised December 30, 2019; Accepted February 12, 2020)

---

## Abstract

Scenario population projection reflects the high probability of future realization and ease of statistical interpretation. Statistics Korea (2019) also presents the results of 30 combinations, including special scenarios, as official statistics. However, deterministic population projections provide limited information about future uncertainties with several limitations that are not probabilistic. The deterministic population projections are scenario-based estimates and show a perfect autocorrelation of three factors (birth, death, movement) of population variation over time. Therefore, international organizations UN, the Max Planck Population Research Institute (MPIDR) of Germany and the Vienna Population Research Institute (VID) of Austria have suggested stochastic based population estimates. In addition, some National Statistics Offices have also adopted this method to provide information along with the scenario results. This paper calculates the demographics of Korea based on a probabilistic or stochastic basis and then draws the pros and cons and show implications of the scenario (deterministic) population projections.

Keywords: scenario, deterministic population projection, uncertainty, stochastic population projection

---

## 1. 서론

예전부터 UN, OECD 등과 같은 국제기구는 인구추계를 시나리오(scenario) 방법으로 그 결과를 제시해오고 있다. 최근 통계청 (KOSTAT, 2019)도 출산율(fertility rate), 사망률(mortality rate), 국제이동률(migration rate)의 실현개연성이 가장 높은 예측치(중위(medium))와 복수의 예측치들인 고위(high), 저위(low)의 조합 시나리오 결과를 제공하였다. 시나리오 인구추계의 장점은 통계 이용자들의 해석 용이성과 인구 변화요인 3요소의 목표치(target value)를 가정하여 정책기반 참고 자료로도 제공 가능하다는 것이다. 더불어 정책입안자(policy maker)의 의사결정과 같은 특별 상황을 반영할 수 있고 미래 실현가능성에 대한 수치 표현으로도 용이하다. 이런 이유로 30개 시나리오를 제시하는 우리나라(KOSTAT)를 비롯하여, 영국(ONS), 스위스(BFS), 아일랜드(CSO), EUROSTAT, 캐나다(STATCAN), 르완다(NISR), UN(2001년 이전까지의 인구추계 결과), 일본(NIPSSR, 국립사회보장인구문제연구소), 독일인구연구소(MPIDR), 오스트리아 인구연구소(VID) 등 이외에도 많은 나라들이 시나리오 방식을 채택해 공식통계로 발표하고 있다.

This research was supported by newly appointed professor research fund of HanBat National University 2019.

<sup>1</sup>Department Mathematical Sciences, HanBat National University, 125 Dongseodaero, Yuseong-gu, Daejeon 34158, Korea. E-mail: [jhoh75@hanbat.ac.kr](mailto:jhoh75@hanbat.ac.kr)

하지만 Alho, Lee, Lutz 등을 비롯한 여러 인구학자들이 시나리오 인구추계에 대한 여러 한계점을 지적해 왔다 (Alho, 1997, 2005; Alho와 Spencer, 1985; Alkema 등, 2015; Dunstan과 Ball, 2016; Gerland 등, 2014; Lee, 1998, 2004; Lee와 Tuljapurkar, 1994; Lutz 등, 2014; Lutz와 Goldstein, 2004; Raftery 등, 2012, 2013; UN, 2017, 2019). 이와 관련된 한계점은 4가지로 요약 가능하다.

첫째, 인구추계 결과의 변동성 크기와 양(measure of variability)을 정량적으로 표현할 수 있어야 하지만 시나리오 조합 결과는 단지 미래에 실현될 수 있는 가능한 대안들 중의 하나일 뿐이다. Alho와 Spencer (1985), Keyfitz (1981), Stoto (1983)에 따르면 고위, 저위 시리즈에 의해서 형성된 범위와 실측치를 비교할 경우 실측치가 고위와 저위시리즈의 범위를 벗어나는 경우도 종종 있다고 보고한다. 실제로 2018년 합계출산율 0.98명은 2016년 통계청 장래인구추계의 출산율의 저위수준(1.0명)보다 낮았다. 또한 미래의 불확실성(uncertainty, 불확실성이란 미래에 나타날 수 있는 경우에 대해서 확실하게 재현된다고 확정지을 수 없는 것)에 대해 시나리오 접근은 인구 변수가 미래에 실현될 개연성에 관한 확률적 해석을 내리지 못하는 문제 (Woo, 2010)가 있다. 따라서 Keilman 등 (2002)과 Lee (2004)은 미래는 불확실성을 내포하고 있기에 추계 대상 인구 변수가 미래에 실현될 가능한 범위와 이에 대한 확률을 구체적으로 명시할 필요가 있다고 한다.

둘째, 시나리오의 경직성이다. 경직성은 시나리오의 경우 인구변화 3요인(출산, 사망, 이동)의 동태를 이 모든 연령대, 시간에 걸쳐 동시적으로 결정된다는 가정을 의미한다. 예를 들어 고위에 해당하는 출산율 가정의 경우 추계기간의 모든 연도에 걸쳐 다른 시리즈의 출산율 가정보다 높게 추계된다. 그리고 무이동과 출산율과 사망률을 조합하여 총인구의 고위 시리즈를 구성할 경우에 추계기간에는 높은 수준의 출산율 가정과 항상 낮은 수준의 사망률 가정을 조합하게 되는데, 이런 방식은 개별 인구 변동 요인들 사이에 강한 양의 상관관계가 내포되어 있음을 가정한다 (Lee, 1998, 2004).

셋째, 추계구간의 비일관성(inconsistency)이다. 예를 들어 미래 인구 최대값을 도출하는 조합은 고위 출산율, 고위 국제이동률, 저위 사망률의 조합이다. 하지만 이 조합은 노년부양비(old age dependency ratio 또는 elderly dependency ratio, 65세 이상/15-64세 생산연령인구)가 매우 좁게 형성되는 문제점을 보인다. 다음으로 사회복지나 연금정책에서 고려하는 복지비용 시나리오(고위 3요인)인 저비용 시나리오(low cost scenario)는 고위 출산율은 높은 인구 성장률을, 고위 사망률은 낮은 인구 성장률을 나타내 서로 상쇄되는 효과를 보인다 (Alho와 Spencer, 1985; Lee, 1998, 2004; Lee와 Tuljapurkar, 1994). 따라서 시나리오 접근이 인구 부양비, 복지 비용과 같은 예측에 있어서는 일관성 없는 예측 결과를 제공한다.

넷째, 미래는 불확실하므로 미래 궤적이 균일하게 고위나 저위처럼 지속적으로 유지되지 않을 가능성이 많다. 따라서 이런점 때문에 미래 궤적의 증감은 여러 형태로 나타날 수 있어 단일적인 우상향, 우하향 패턴은 합리적이지 않다. 이처럼 시나리오 기반 인구추계는 여러 한계점이 있으므로, UN (2017, 2019), UNICEF (2017)와 같은 국제기구와 독일 막스플랑크인구연구소 (MPIDR, 2006, 2007; Li 등, 2007), 오스트리아 비엔나인구연구소 (VID, 2016) 등은 이에 대한 대안으로 확률론적(stochastic) 기반 인구추계를 제시하고 있다. 더불어 해외 일부 국가 통계청에서도 선도적으로 확률론적 인구추계 방식을 도입하는 추세이다 (Statistics NZ, 2016; SCB, 2006; ONS, 2009).

따라서 본 논문은 우리나라 인구추계를 확률론적 기반으로 도출해 보고 시나리오 인구추계와 비교한 후, 결과 비교를 통해 시사점을 도출하고자 한다. 그리고 본 연구는 다음과 같은 점에서 선행연구들과 차별된다. 첫째, 확률론적 인구추계와 시나리오 접근과의 차이점을 살펴본 선행 연구는 있으나, 이들을 통계프로그램으로 구현하여 시뮬레이션으로 비교한 연구는 없다. 더불어 시나리오 인구추계에 대한 한계점을 보이고 확률론적 인구추계의 효용성과 활용을 제안한다. 둘째, UN (2017, 2019), Booth (2006), 그리고 Lee (1998)가 제시한 확률론적 인구추계 구현법과는 다르게 먼저 인구변화 3요인(일반화로그

감마모형(generalized log gamma model; GLG) (Kaneko, 2003)의 출산율, LC-ER 모형(Lee-Carter method extended with rotation) (Li 등, 2013)의 사망률, 함수적자료모형(functional data model; FDM) (Hyndman 등, 2013)의 국제이동에 개별 함수를 적용해 예측분포(predictive distribution)를 생성한다. 다음으로 도출된 예측값(궤적)을 확률변수화하여 이 값들을 랜덤으로 선택해 코호트 요인법(cohort component method)에 적용 후 추계인구를 도출하는 과정을 반복함으로써 확률론적 인구 추계를 완성한다. 이런 방법은 확률적 예측에서 일반적으로 사용되는 통계적 방법(statistical method) (Hyndman과 Ullah, 2007; Hyndman과 Booth, 2008; Hyndman 등, 2013; UN, 2017), 전문가 판단법(expert judgment method) (Alders과 De beer, 1999; Luts와 Goldstein, 2004; Luts, 2009; Luts 등, 2014), 과거 예측 오차 분석법(historical forecast error method) (Alders과 De Beer, 2004; De Beer, 2000; Keilman, 2005; Keyfitz, 1981; Stoto, 1983) 중 통계적 방법에 해당된다.

본 논문은 차별화된 연구 수행을 위해 총 4개의 장으로 구성한다. 제 2장에서는 시나리오 인구추계로 대변되는 결정론적 인구추계(deterministic population projections)와 미래에 실현될 가능한 범위와 이런 범위가 갖는 확률을 구체적으로 명시하는 확률론적 인구추계와의 차이를 조명하고, 이들의 선행연구들을 소개한다. 3장에서는 우리나라 미래 인구에 대해 확률론적 인구추계 방법을 적용해 산출하고 시나리오 인구추계와 비교한다. 그리고 시간에 따른 인구구조 차이와 고령화 지수 등으로 이 둘의 차이를 구체적으로 살펴본다. 그리고 4장에서는 결론과 함께 본 연구 한계점과 향후연구방향을 제시한다.

## 2. 결정론적과 확률론적 인구추계

### 2.1. 결정론적과 확률론적 인구추계 차이

Dunstan과 Ball (2016)에 따르면 인구추계는 확률론적 인구추계의 상반된 용어이며, 시나리오 인구추계와 동일한 용어로 통용된다. 이는 출산율, 사망률, 국제이동률의 실현가능성의 결과들을 조합하여 하나의 추계를 산출하는 방법이다. 그리고 다양한 재현가능성 조합의 시나리오 인구추계는 서로 다른 가정의 조합 결과이다. 그러므로 고위, 중위, 저위 가정에 기초한 복수의 시리즈 제공으로 장래인구추계에 내포된 불확실성을 표현하지 못할뿐더러 Figure 2.1의 (b)처럼 정량적인 방식(quantified way)의 범위에 따른 확률(5th percentile, 95th percentile 등)을 보여주지 못한다. 이런 불확실성에 대해 Lee (1998, 2004)는 8가지로 분류하고 있다. 첫 번째, 개인적차원에서 임신, 유산, 심장마비와 같은 생애이벤트는 인구통계학적 관점에서 사건(event)으로 정의되고 결정적(deterministic)이지 않고 확률적(stochastic or probabilistic)으로 발생된다. 이런 개개인의 모인 유한인구(finite population)에서 출산율과 사망률의 합계는 불확실성의 확률적인 요소이다. 두 번째, 인구추계에서 시간 경과에 따른 예측할 수 없는 변동은 인구 예측의 불확실성을 좌우하는 중요한 원인이다. 세 번째, 예측 방법이 무엇이든 간에 추정모형은 선행적 자료에서 시작한다. 하지만 자료는 측정오류를 내재하고 있다. 네 번째, 모형 분석가는 인구 동태통계의 결측치 처리, 5세를 1세로 변환하는 과정 등의 자료 가공뿐만 아니라 이들 자료에 서로 다른 모수, 비모수 모형 등을 적용하여 예측할 때 도출된 결과가 사뭇 다를 수 있다. 일반적으로 자료의 완비성(completeness)이 좋은 경우 모수적 모형, 자료 결측치가 내포되어 있고 과거의 패턴이 미래에도 지속된다는 가정이 합리적이지 못한 경우 비모수 모형을 적용한다. 다섯 번째, 모형과 자료가 선택되면 통계적인 방법으로 모수가 추정된다. 모수 자체는 불확실성을 가정하고 추정되며 이는 예측의 불확실성으로 이어진다. 여섯 번째, 예측 연구자는 다양한 가정으로 인구장려정책 등과 관련된 이벤트성 정책을 고려한다. 예를 들어 정책 변화가 없거나, 확대, 새로운 정책이 추가 등의 다양한 경우를 시뮬레이션한다. 일곱 번째, 사회의 구조적 변화의 불확실성이다. 즉, 20세기 여성의 노동시장으로 진입, 새로운 피임 기술의 개발, 신약 개발 등 근원적인 추세를 변화시키고 예측으로 반영되는 구조적 변화를 말한다. 끝으로 핵전쟁, 쓰나미, 지진 등과 같은 대재앙은 위의 분류와 다른 불확실성이다. 이들 8가지 불확실성 원천

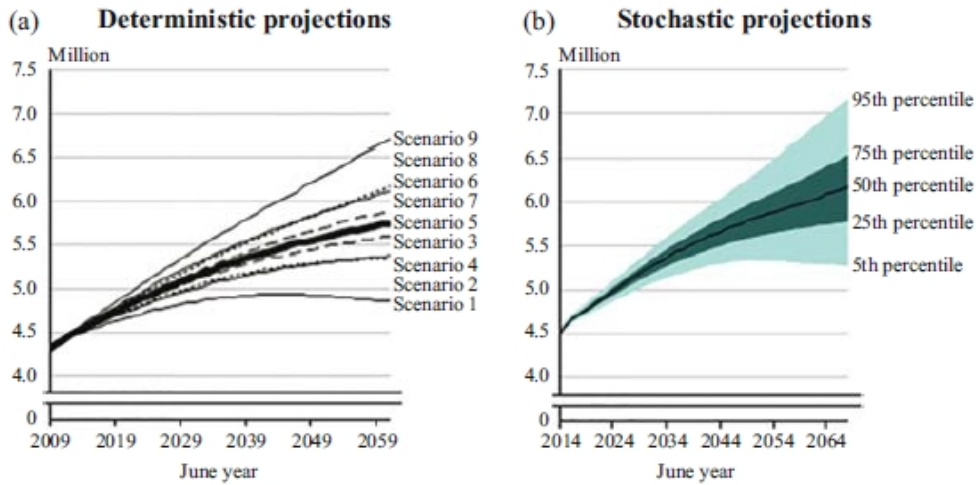


Figure 2.1. Comparisons with deterministic and stochastic population projections (Dunstan and Ball, 2016).

이 출산을 변동이 심한 우리나라에게는 마지막을 제외하고 관련성이 높다고 할 수 있다. 결정론적과 확률론적 인구추계의 주된 차이점은 인구추계 가정에 따른 변동성 크기와 양의 포함여부이다. 즉 범위에 대한 확률표현의 가능성 여부이다. 확률표현을 위한 방법은 다음처럼 가능하다. 먼저 다양한 시뮬레이션 결과가 산출되는데, 이런 산출은 과거 데이터나 전문가 판단법을 활용하여 실험적인 모형으로부터 추출되는 개별 가정의 확률적 분포에 따라 랜덤하게 구현된다 (Booth와 Tickle, 2008; Lutz, 2009). 이러한 가정의 시뮬레이션은 인구 시뮬레이션이나 추계를 위해서 코호트요인법을 적용하여 도출한다. 따라서 확률론적 접근의 진정한 값들은 개별적인 시뮬레이션의 결과가 아닌 수없이 생성된 시뮬레이션과 확률과정을 반복한 결과이다. 이런 결과로부터 인구크기와 다른 인구구조적인 변수들의 확률분포를 생성한다. 이는 Figure 2.1의 (b)와 같이 분위수(percentiles)로 정리가 가능하다. 예를 들어 Figure 2.1의 (b)는 25번째 분위수(25th percentile)는 이 분위수 보다 낮은 결과값으로 도출될 가능성이 25%이라는 것을 의미한다.

하지만 결정론적 인구추계는 미래의 인구변동에 대한 하나의 가능한 전망을 제시한 것이고, 중위가 다른 고위, 저위에 비해 우월하다고는 해석하지 않는다 (Smith 등, 2001, Woo 2010). 따라서 만약 어떤 특정한 시나리오 결과가 다른 결과에 비해 상대적으로 실현 개연성이 높다고 한다면 나머지 시나리오 산출은 이에 대한 실현 개연성이 떨어지는 대안으로 해석될 수밖에 없다. 또 다른 사례를 들어 이들의 차이점을 살펴보자. Figure 2.1(a)에서도 나타나 있듯이 시나리오 조합에 따른 9가지의 인구추계 순위가 오름차순이 아님을 살펴보았다. 이를 확인하기 위해 통계청 장래인구추계 (KOSTAT, 2016)로 살펴보자. 30가지 시나리오 조합 중, 인구 규모 순으로 상위 5개와 하위 5개 조합을 선택하여 2065년까지 3그룹의 인구구조(유소년층(0-14세), 경제활동인구(15-64세), 장년층(65세 이상)) 변화양상을 Figure 2.2처럼 나타내 보았다. Figure 2.1(a)와 동일하게 Figure 2.2에서도 상위 5개와 하위 5개의 순위가 3그룹별로 순위가 다르게 나타나는 것을 확인할 수 있다. 즉, 30가지 시나리오 상위 1, 2, 3, 4, 5와 하위 26, 27, 28, 29, 30순위는 유소년층에서는 3, 2, 1, 4, 5와 26, 27, 30, 29, 28 순으로, 경제활동인구에서는 3, 2, 1, 5, 4와 27, 26, 30, 29, 28 순으로, 끝으로 장년층은 28, 29, 30, 27, 26과 5, 4, 1, 2, 3 순으로 나타났다 (Table 2.1). 특히 유소년층의 5순위와 26순위는 거의 유사한 궤적을 보인다. 더불어 30개 시나리오 조합 결과 중 시간 흐름에 따른 최상위(1)와 최하위(30) 범위(interval range)를 3계층의 간격 범위를 비

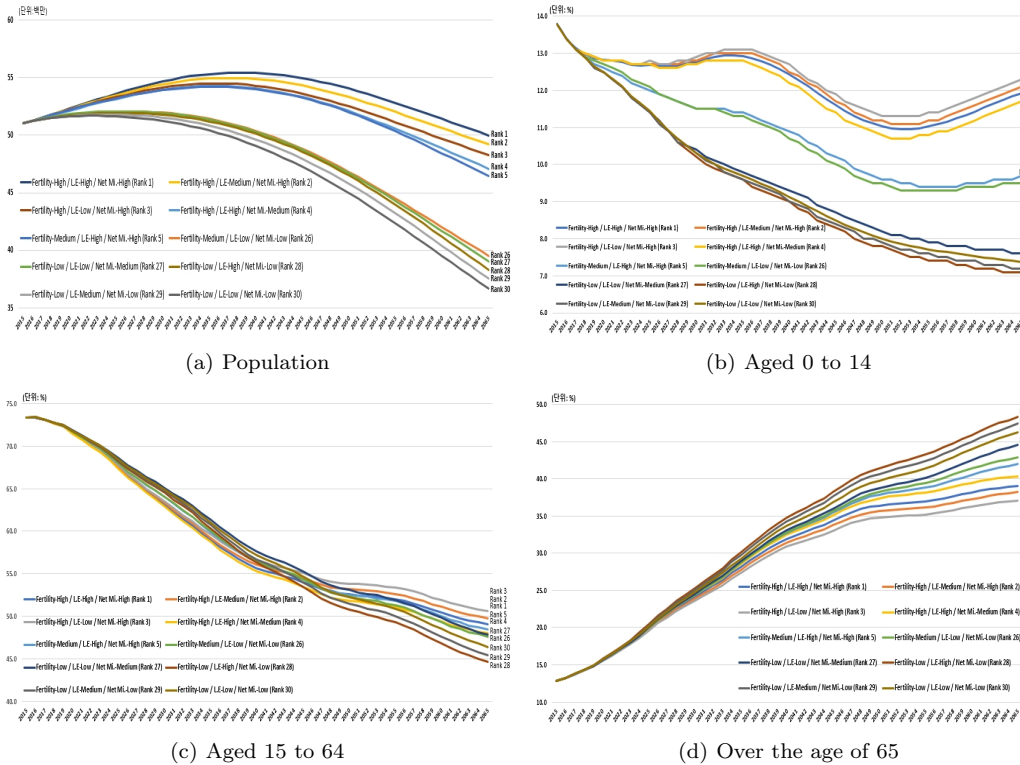


Figure 2.2. Trend on population structure for upper and lower 5-Scenarios.

Table 2.1. Rank of High and Low five scenarios by population group

Population group	Rank of high five					Rank of low five				
	1	2	3	4	5	26	27	28	29	30
Population	R 1	R 2	R 3	R 4	R 5	R 26	R 27	R 28	R 29	R 30
Aged 0 to 14	R 3	R 2	R 1	R 4	R 5	R 26	R 27	R 30	R 29	R 28
Aged 15 to 64	R 3	R 2	R 1	R 4	R 5	R 27	R 26	R 30	R 29	R 28
Over the age of 65	R 28	R 29	R 30	R 27	R 26	R 5	R 4	R 1	R 2	R 3

Source: KOSTAT (2016), Population Projection 2016; “R” 은 Rank를 의미함.

교해보면 매우 다름을 확인할 수 있다. 이처럼 통계청의 장래인구추계도 결정론적 인구추계의 시나리오 접근이므로 Figure 2.1(a)에서 보여준 미래 궤적의 일치성(consistency)을 보이지 않는다. 앞 절에서 결정론적, 확률론적 인구추계 차이점은 변동성 크기와 양 표현 가능여부, 시나리오 경직성 여부, 추계구간의 비일관성, 단일 궤적의 비합리성이다. Figure 2.2는 Figure 2.1(b)처럼 변동성 크기와 양 표현이 어려우며, 고위 조합은 추계 기간의 모든 연도에 걸쳐 다른 시리즈 보다 상위에 위치하는 시나리오 경직성을 나타내고, 추계 기간의 비일관성 순위를 보여주고, 그리고 장래인구추계의 궤적이 단편적이다.

하지만 확률론적 인구추계는 순위의 개념이 아니라 장래인구추계의 미래 불확실성에 대해 추계 대상 인구 변수들이 미래에 실현될 가능한 범위와 확률을 구체적으로 도시한다. 다음 절에는 확률론적 인구추계의 구현법 종류와 결정론적 인구추계와의 차이점을 제시한다.

## 2.2. 확률론적 인구추계 구현 방법

앞 절에서는 실현 재현성과 해석의 용이성 측면에서 선호되지만 여러가지 한계점을 가지고 있는 결정론적 인구추계를 소개하고 살펴보았다. 이번 절은 확률론적 인구추계 구현방법과 비교를 중점적으로 제시하고자 한다. Alho (1997, 2005), Alho와 Spencer (1985), Alkema 등 (2015), Dunstan와 Ball (2016), Lee (1998), Lee와 Tuljapurkar (1994), Lutz와 Goldstein (2004)은 결정론적 대안으로 확률론적 인구추계를 제시한다. 이들이 제시하는 구현 방법은 대략 4가지이다. 첫째, 실측치와 과거 예측 간의 오차를 활용(past projections error)하는 방법으로, 과거 예측 오차 분석법이라 한다. 이는 과거 실시한 인구총조사와 인구추계에서 발표한 수치와 5년이 지난 실측치와 비교할 때 이들 비교에서 인구 성장률 예측의 표준오차를 산출하는 것이다. 일명 'ex post error analysis'라고 한다. 이들 표준오차들은 장래 인구의 예측에 대한 신뢰구간을 도출하는데 사용될 수 있다.

하지만 Woo (2010)은 과거에 관측된 오차의 본질적 측면과 일관된 미래 예측치의 확률분포를 구현할 수 있는 장점이 있으나, 과거 예측 오차 정보가 상대적으로 제한적이라는 문제와 예측 방법의 향상 등에 기인하여 시간의 경과에 따라 예측 오차가 감소했을 개연성이 있음을 지적한다. 둘째, 통계적 모형활용이다. 장래인구추계는 인구변화 직접요인 3요소의 불확실성을 통계적 모형으로 다루는 것으로, 개별 요인들이 확률론적 모형으로 추정되고 그 결과인 예측치를 코호트 요인법에 적용하여 확률론적 인구추계를 산출하는 것이다. 예를 들어 통계청 (KOSTAT, 2016)은 출산율, 사망률, 국제이동률을 서로 다른 확률모형에 적용하여 예측치와 신뢰구간을 도출하고 있다. 개별 요인들에 대한 확률론적 모형으로는 시계열 모형뿐만 아니라 특정 모수적, 비모수적 모형(GLG (Kaneko, 2003); Lee-Carter (LC), 1992; Li와 Lee, 2005; Li 등, 2013; Statistics Canada, 2015; FDM (Hyndman과 Ullah, 2007; Hyndman과 Booth, 2008; Hyndman 등, 2013))이 있다. 이 방법은 과거에 대한 실측치를 통해 예측분포를 도출할 수 있다는 장점은 있으나 모형 도출을 위해 요구되는 자료가 충분치 못한 경우에는 활용성이 떨어지는 제한점이 있고, 미래에 실현될 값들은 과거 실측치를 기반으로 한 모형에 기초하여 발생함을 가정하고 있으므로 미래에도 과거와 유사한 패턴을 유지할 것이라는 가정에 기초한다. 셋째, 미래에 재현될 가능성에 대해 전문가들의 의견을 반영하는 전문가 판단법(expert judgement method or the role of experts in defining uncertainty)이다. 미래에 일어날 수 있는 확률은 전문가의 평균 예측과 관련성이 높다는 선행연구 (Lee, 1998; Billari 등, 2012)가 있다. 이 방법은 통계적 모형과 비교할 때 상대적으로 자료에 대한 요구사항이 크지 않지만 어떤 전문가로 구성해야 하는지와 전문가들의 판단이 대체적으로 보수적일 가능성이 높다는 제한점이 있다. 넷째, Leslie matrix (Smith와 Keyfitz, 1977)처럼 인구추계의 변화 요인 3요소를 확률변수처럼 구현하는 것이다. 즉, 확률변수들은 서로 다른 분포들을 가정하여 생성한다. 이들 과정의 첫 단계는 미래 한 단계의 인구를 예측하기 위해 난수(random number)와 확률변수(random variable) 생성 규칙을 사용한다. 이런 과정은 단계적으로 진행되며 예측 시계열 동안 모의실험을 반복한다. 확률적 반복 실험에서 산출된 개별 요인의 중앙값 또는 평균값은 중위 수준을 제공한다. 본 논문은 서술된 네 가지 방법 중 두 번째와 네 번째 방식을 조합하여 확률론적 인구추계 도출하는 과정을 다음 장에서 소개한다.

## 3. 확률론적 인구추계 구현과 비교

### 3.1. 통계적 모형을 활용한 확률론적 인구추계

이번 절은 확률론적 인구추계의 두 번째 방법인 방법인 인구변화 3요인의 출산, 사망, 국제이동에 대한 통계적 모형과 그 결과를 소개한다. 먼저 출산율이다. 출산율은 GLG 모형 (Kaneko, 2003)을 적용한다. 이 모형은 출산순위별 출산모형에 확률분포의 개념을 적용한다. 이는 확률분포모형을 사용하

게 되면 모수 추정에 이론적 기반이 있는 통계적 방법을 이용할 수 있는 장점이 있다. 그동안 연령별 출산율(age-specific fertility rate; ASFR)의 패턴을 설명하기 위해 다수의 모수화 모형이 소개되어 왔다. 이들 모수화 모형은 Hoem 등 (1981)가 제안한 베타함수를 제외하고는 모두 지수족(exponential family)을 형태를 띠고 있다. 이는 ASFR의 대략적인 형태가 평균출산연령을 중심으로 좌우 대칭을 띠고 있기 때문이다. 통계청 (KOSTAT, 2016)은 ASFR 추정시 GLG 모형을 적용한다. 식 (3.1)은 GLG 모형인데 모수 4개( $C_i, \mu, b, \lambda$ )를 포함하고 있다.

$$f_i(x) = \frac{C_i |\lambda|}{b \Gamma(1/\lambda^2)} \left( \frac{1}{\lambda^2} \right)^{\lambda^{-2}} \exp \left[ \frac{1}{\lambda} \left( \frac{x - \mu}{b} \right) - \frac{1}{\lambda^2} \exp \left( \lambda \left( \frac{x - \mu}{b} \right) \right) \right], \quad (3.1)$$

여기서  $f_i(x)$ 는 연령  $x$ 세의 출산율,  $C_i$ 는 특정의 출생코호트가 가입연령동안 출산순위  $i$ 번째 자녀의 출산을 경험할 확률이다.  $\mu$ 와  $b$ 는 출산연령의 평균과 표준편차,  $\lambda$ 는 분포형태를 나타내는 모수이다. 따라서 GLG 모형은 결혼의 출생순서(birth-order)에 대한 연령패턴의 규칙성(regularity)을 수학적으로 표현한 것 (Kaneko, 2003)으로 이해할 수 있다. 최종적으로 로그감마모형으로 모수를 추정된 후 이 추정된 4개 모수들을 시계열 모형에 적합하여 예측하고 이들 값을 근거로 미래의 연령별 출산율을 산출한다. 다음으로 사망률이다. 최근 사망률이 낮은 선진국과 우리나라는 예전에 비해 유소년(infant and child) 사망률 개선은 점진적으로 감소하고, 고령층의 사망률 개선은 점점 빨라지는 사망률 개선 교대(declined mortality rate with rotation)현상이 발생하고 있다 (Horiuchi와 Wilmoth, 1995; Li와 Gerland, 2011; Li 등, 2013; Kim과 Oh, 2017). 이것은 유소년층과 노년층의 사망률 감소 패턴이 교대(rotation)되는 현상으로 기대수명이 증가함에 따라 발생하고 있으며 LC 모형 (Lee와 Carter, 1992)에서 연령별 사망률 변화를 나타내는  $b_x$ 가 시간의 경과에 따라 변하고 있음을 의미한다. Li와 Gerland (2011)은 이와 같은 현상을 사망률 개선 교대라 명명하고 시간의 변화에 따라 연령별 사망률 개선 교대현상을 반영하기 위해 LC 모형 ( $\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{t,x}$ )의  $b_x$ 에 강건한 순환(robust rotation)을 도입하여 주관적이고 강건한 수정을 간주하는 LC-RR (Lee-Carter method with robust rotation) 모형을 제안하였다. 그리고 2년 뒤에 Li 등 (2013)은 LC-RR 모형의 회전모형을 유연하게 변화를 줄 수 있도록 아래 식 (3.2)와 같은 LC-ER 모형을 제안하였다. 즉, 이는 LC모형에서 시간의 변화에 관계없이 일정한 시간에 변수를 고려한 형태로 변환한 모형이다. 그리고 통계청 (KOSTAT, 2016)은 사회, 경제적 조건이 유사한 인구 부집단(예: 시도별)의 사망률을 고려하는 LL모형과 LC-ER를 조합한 LL&LC-ER을 제안하였다. 본 연구는 전체인구를 산정하는 것이므로 LC-ER모형으로 한정지어 사망률 개선 교대현상을 설명하고자 한다. 이와 같은 사망률 개선 교대현상을 반영하기 위해 Li 등 (2013)은 식 (3.2)와 같이 LC-ER 모형을 제안하였다.

$$\ln(m_{x,t}) = a_x + B_{x,t} K_t + \epsilon_{t,x}, \quad (3.2)$$

여기서  $B_{x,t}$ 는 교대 이전의 연령별 사망률인 LC 모형의  $b_x$ 와 사망률 개선의 최고점 사망률  $b_x^u$ 의 선형 가중평균으로 식 (3.3)과 같이 정의되며,  $e_0^u$ 는 교대 현상이 끝나는 시점의 기대수명을 나타낸다.

$$B_{x,t} = \begin{cases} b_x, & e_0^t < 80, \\ (1 - w_s(t))b_x + w_s(t)b_x^u, & 80 \leq e_0^t < e_0^u, \\ b_x^u, & e_0^t > e_0^u, \end{cases} \quad (3.3)$$

$$b_x^u = \begin{cases} \bar{b}_{15-64}, & 0 \leq x \leq 64, \\ b_x \times \frac{b_{u,60-64}}{b_{65-70}}, & 65 \leq x, \end{cases}$$

$$w(t) = \frac{e_0^t - 80}{e_0^u - 80}; \quad w_s(t) = \left[ 0.5 \left\{ 1 + \sin \left( \frac{\pi}{2} (2w(t) - 1) \right) \right\} \right]^p.$$

이는 LC 모형 계열에서 연령별 사망률 개선 패턴이 매년 동일( $b_x$ )하다는 단점을 보완한 것이다. 우리나라 유소년층의 사산, 영아사망, 그리고 질병 등과 같은 사망률 개선은 과거와는 다른 의료개선으로 그 속도가 예전과 다르게 감소하고 있고, 인간 수명의 연장에 대한 관심 증대와 의학의 비약적인 발전을 가져온 게놈프로젝트(genome project, 유전정보 분석)와 인공지능 등이 개발됨에 따라 고령층 사망률 개선은 상대적으로 빠르게 진행되고 있다. 이는 젊은층과 고령층의 사망률 개선이 예전 패턴과는 다른 사망률 개선 교대와 유사하다. LC-ER 모형은 선진 20개국의 사망률 분석에 근거해서 몇 가지 실험적인 결과 값을 이용한 몇 가지 가정을 활용한다. 식 (3.3)에 나타나 있듯이 연령별 사망률 개선 패턴은 연령 80세에 도달 했을 때 교대하기 시작해서 102세까지 지속되며, 이후에는 일정하다고 간주한다. 이는 65세 미만 모든 연령에서 사망률 개선패턴은 동일하며 그 이후에 연령이 증가할수록 작아짐을 의미한다. 그리고  $B_{x,t}$ 의 임의의 선형변환으로  $K_t$ 는 기대수명이 교대 없이 얻은 LC의  $k_t$ 와 비교하여 둘의 차가 가장 작은  $K_t$ 를 구하기 위해 반복적인 계산으로 도출한다. 이런 과정을 거치면  $b_x^\mu$ ,  $e_0^\mu$ 와  $B_{x,t}$ 를 도출한 후 성별과  $K_t$ 값으로  $e_0^\mu$ 을 예측할 수 있다. 그리고 식 (3.3)을 성별에 따라 확장하면 식 (3.4)와 같다.

$$\begin{cases} \text{남} : \ln(m_{x,t,m}) = a_{x,m} + B_{x,t}K_{t,m} + \epsilon_{x,t,m}, \\ \text{여} : \ln(m_{x,t,f}) = a_{x,f} + B_{x,t}K_{t,f} + \epsilon_{x,t,f}, \end{cases} \quad (3.4)$$

여기서  $m$ 과  $f$ 는 각각 남성(male), 여성(female)을 뜻한다.

끝으로 국제이동이다. Ramsay와 Silverman (2005)과 Hyndman과 Ullah (2007), Hyndman과 Booth (2008), Hyndman 등 (2013)은 함수적 자료 분석 패러다임을 사용하여 출산율, 사망률, 국제이동률 등을 모델링하고 예측하기 위한 비모수적 방법인 FDM을 제안하였다. 그들은 관측치에 존재하는 측정오차와 질병, 기아, 전쟁 등으로 인구동태 자료에서 나타나는 불규칙적인 패턴을 교정하기 위해 함수적 자료분석을 이용하여 식 (3.5)와 같은 모형을 구축하고 비모수 평활기법을 이용한다 (Kim과 Oh, 2017; Oh, 2018). 제안된 FDM 모형의 구조는 식 (3.5)와 (3.6)과 같다.

$$f_t(x) = \begin{cases} \frac{1}{\lambda} \left( f_t^*(x)^\lambda - 1 \right), & 0 < \lambda < 1, \\ \ln(f_t^*(x)), & \lambda = 0, \end{cases} \quad (3.5)$$

여기서  $f_t^*(x)$ 는 시간  $t$ 와 연령  $x$ 에서 국제이동률을 의미한다.  $f_t^*(x)$ 의 Box-Cox (1964)변형은  $f_t^*(x)$ 의 값에 따라 증가하는 변동을 줄여주거나 정규화과정으로  $\lambda$ 는 Box-Cox 변형에서 강도를 뜻한다.

$$\begin{aligned} f_t(x) &= s_t(x) + \sigma_t(x)\epsilon_{t,x}, \\ s_t(x) &= \mu(x) + \sum_{j=1}^J \beta_{t,j}\phi_j(x) + e_t(x), \end{aligned} \quad (3.6)$$

여기서  $\mu(x)$ 는  $\sum_{t=1}^n s_t(x)/n$ 에 의해 추정된 평균함수로 평활된 연령에 따른 로그국제이동률평균이고,  $\beta_{t,j}\phi_j(x)$  ( $t = 1, \dots, n$ ,  $j = 1, \dots, J$ )는 함수적 주성분분석을 사용하여 추정되어지며  $J < n$ 는 사용된 주성분 수이다.  $\Phi = \{\phi_1(x), \dots, \phi_J(x)\}$ 는  $J$ 개의 함수적 주성분의 집합으로 직교 기저함수(orthogonal basis function)이고  $B = \{\beta_{t,1}, \dots, \beta_{t,J}\}$ 는 비상관 주성분 점수(uncorrelated principal component scores)들의 집합으로 시계열 계수를 의미한다.

식 (3.6)에서  $f_t(x)$ 는 시간  $t$ 의 연령  $x$ 에 대한 관찰된 로그국제이동률  $\ln f_t(x)$ 이고,  $s_t(x)$ 는 평활함수(smooth function),  $\epsilon_{t,x}$ 는 독립적이고 동일하게 분포된 표준정규 확률변수이고,  $\sigma_t(x)$ 는 시간  $t$ 의 연령  $x$ 에 따라 변하는 잡음의 양이다. 즉,  $\sigma_t(x)\epsilon_{t,x}$ 는 관측된 로그국제이동률과 평활된 곡선의 차이인 관측오류를 의미한다. 식 (3.6)의 두 번째 식은 시간에 따라 변화하는  $s_t(x)$ 의 변화를 설명하는 부분으



로 하나 이상의 주성분을 사용하고 FPCA를 사용하여 평활된 곡선  $s_t(x)$ 를 직교함수 주성분과 비상관 주성분 점수로 분해한 것이다. FDM은 첫 번째 주성분에 직교하는 고차원 주성분에 대해서는 다른 시계열 모형들의 주성분 점수가 도출된다. 모든 성분에 FDM 방법은 최적 시계열 모형을 AIC 등과 같은 모형 판별 기준에 의거하여 선택한다. 이 모형에 대한 보다 자세한 설명은 Hyndman과 Ullah (2007), Hyndman과 Booth (2008), Hyndman 등 (2013), Kim과 Oh (2017), Oh (2018)를 참조하면 된다.

그리고 실측치  $Z = \{f_1(x), \dots, f_n(x)\}$ 와  $\Phi = \{\phi_1(x), \dots, \phi_J(x)\}$ 의 조건부로  $f_{n+h}(x)$ 의  $h$ 단계 예측치를 구할 경우 식 (3.7)에 의해서 도출 가능하다.

$$\hat{f}_{n+h|n}(x) = E[f_{n+h}(x)|Z, \Phi] = \hat{\mu}(x) + \sum_{j=1}^J \hat{\beta}_{n+h|n,j} \phi_j(x), \quad (3.7)$$

여기서  $\hat{\beta}_{n+h|n,j}$ 은 Hyndman과 Booth (2008)에 의한 지수 평활 또는 ARIMA 모형과 같은 일변량 시계열모형을 활용하여 도출한  $\beta_{n+h,j}$ 의  $h$  단계 예측을 의미한다. 그리고 모든 주성분들은 직교성을 보이므로 예측분산을 식 (3.8)과 같이 도출 가능하다.

$$\hat{v}_{n+h|n}(x) = \text{Var}[f_{n+h}(x)|Z, \Phi] = \sigma_{\mu}^2(x) + \sum_{j=1}^J u_{n+h|n,j} \phi_j^2(x) + v(x) + \sigma_t^2(x), \quad (3.8)$$

여기서  $\sigma_{\mu}^2(x)$ 는  $\hat{\mu}(x)$ 의 분산,  $u_{n+h|n,j}$ 은  $\beta_{n+h,j}|\beta_{1,j}, \dots, \beta_{n,j}$ 의 분산이다. 그리고  $v(x)$ ,  $\sigma_t^2(x)$ 는 각각  $e_t(x)$ 와  $\sigma_t(x)$ 의 분산이다. 본 논문은 FDM으로 분석하기 위해 통계프로그램 R에서 제공하는 demography패키지를 활용한다.

지금까지 소개된 3가지 모형으로 인구변동 3요인 출산, 사망, 이동에 대해 2067년까지 예측값을 도출하고, 각 예측값에 대해서 R의 rnorm과 replicate함수를 활용하여 미래 불확실성을 표현 [예측값 + rnorm (확률변수 수, 평균, 표준편차)]한다. 한 예로 출산율의 모수  $C$ 에 대한 R프로그램의 일부를 소개하면, 예측값의 중앙값(50th percentile)에 대해서 확률변수를 예측구간 크기로 랜덤하게 생성한 후  $c$ (중앙값)+rnorm( $n=\text{length}(c)$ (중앙값)), $\text{mean}=0$ , $\text{sd}=0.1\sim 0.2$ ), 이러한 무작위 실험(randomized experiment)을  $n$ 번 반복하여 그 결과값을 벡터로 저장한다. 아래는 R코드 일부를 소개한다.

---

```
temp_c ← replicate(n,c(중앙값)+rnorm(n=length(c(중앙값)),mean=0,sd=0.1~0.2))
n은 반복수인 미래 궤적(trajectories)을 의미
```

---

### 3.2. 확률론적 인구주계 시뮬레이션

이번 절에서는 미래의 불확실성을 표현하는 확률론적 장래인구를 산출한다.

먼저 우리나라 출산율을 고려하는 4개 모수의 GLG 모형에 대해 각 모수별로 R의 replicate와 rnorm함수로 예측값에 불확실성을 반영한다. 그 결과 Figure 3.1과 같이 나타난다. 이들 결과를 GLG 모형에 대입하여 2018-2067년 ASFR을 산정해 total fertility rate (TFR)을 도출한 결과는 Figure 3.2와 같다.

연도에 따른 미래 TFR를 선으로 이은 하나의 임의 곡선과 이런 과정을 50번 반복하면 Figure 3.2 상단 그림으로 나타난다. 이 결과는 앞 장에서 소개한 Figure 2.1(b), Figure 3.2 하단의 UN의 WPP2019 결과와 유사함을 보여준다. 만약 50번 반복횟수를 증가시키면 확률적 범위가 선명하게 도시되며, Figure 2.1(b)처럼 정량적인 범위의 형태를 띠게 된다.

다음으로 LC-ER모형의 사망확률( $1 - \text{생존확률}$ ), FDM 국제이동자수 예측값을 통계적 모형으로 산출하고, 출산율에 적용한 방법으로 이들에 대해 불확실성을 반영한다. 이들의 결과는 Figure 3.3과 같다.

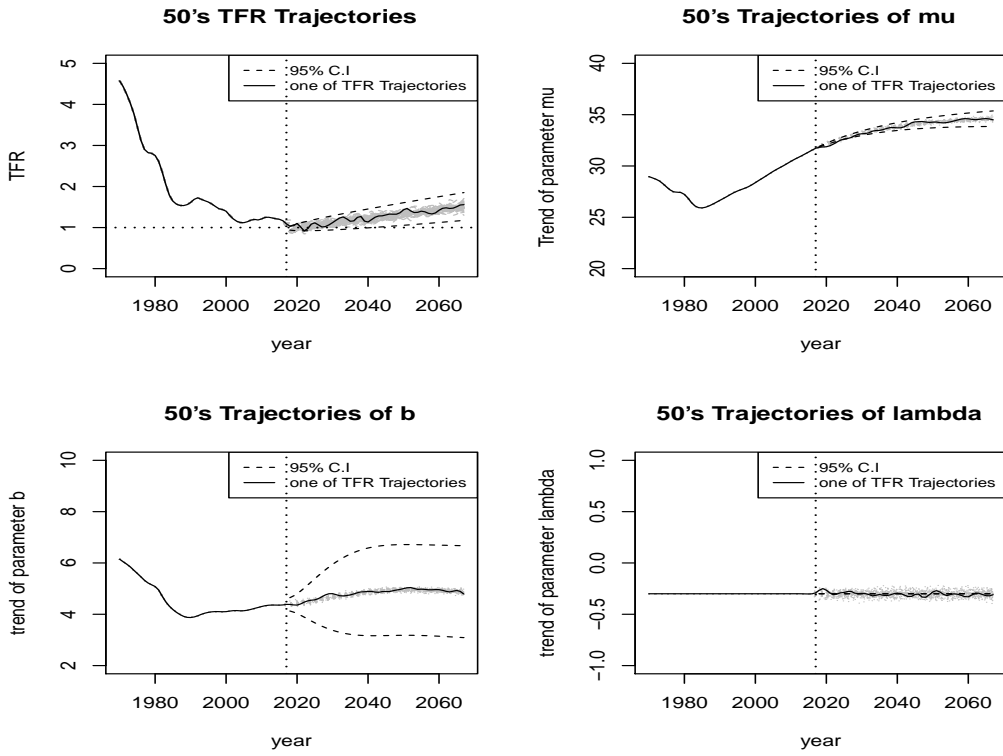


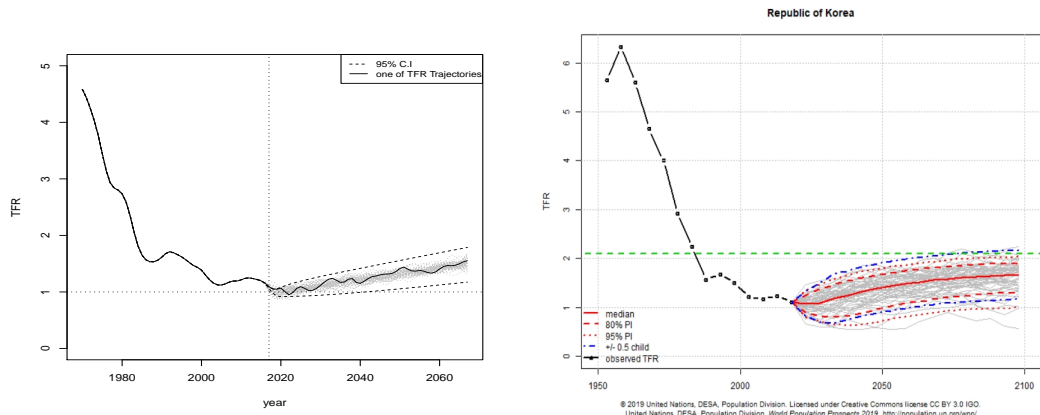
Figure 3.1. Stochastic projection on 4 parameters of generalized log gamma model.

끝으로 출산, 사망, 이동의 인구변동 3요인 미래 불확실성에 대한 예측분포로부터 하나씩 랜덤으로 선택한 후, 그 값을 코호트요인법에 대입해 추계인구를 산출하는 과정을 반복하여 미래 인구의 궤적을 도식한다.

예측분포란 전통적인 시나리오 접근이 복수의 시나리오를 도입함으로써 인구추계에 수반된 불확실성을 고려하는 반면 확률적 인구추계의 경우 인구변동 요인들의 동태율을 확률변수로 취급함으로써 불확실성을 반영 (Alho와 Spencer, 1985)한다. 보다 구체적으로 확률적 인구추계에서 어떤 인구 변수의 미래 예측치들은 전통적으로 예측분포라고 불리는 확률분포를 갖는 확률변수로 간주되며, 이러한 예측분포의 평균 혹은 중앙값을 점 예측치로 하여 확률구간을 도출하는 것이 확률적 인구추계의 근간을 이룬다. 결과적으로 확률적 접근의 경우 인구변동 요인들과 관련된 모수들의 결합분포(joint distribution)를 도출하는 것이 중요한 고려사항이 된다.

그러나 일반적으로 모든 모수들의 결합분포를 명시하는 대신 출산력, 사망력, 인구이동 사이의 독립성 가정에 기초하여 적은 수의 변수들만을 고려한 소규모의 분포로 분리하는 방식을 취한다 (Woo, 2010). 따라서 본 논문에서는 각 요인에 대해 통계적 모형으로 예측값을 산출하고 이들 예측값의 오차를 R의 `rnorm`과 `replicate` 함수를 활용하여 미래 불확실성을 표현하여 예측분포를 생성한다. 따라서 이러한 일련의 계산과정은 R프로그램의 `replicate`, `rnorm`과 Leslie 행렬로직 (Leslie, 1945, 1948)을 활용한다.

이런 3단계 계산 방식을 거쳐 최종적으로 성별에 따른 2067년까지 확률론적 장래인구추계 결과는 Figure 3.1의 하단과 같다. 이들 결과는 앞 절에서 언급된 결정론적 인구추계의 한계점인 확률적 정량적 양



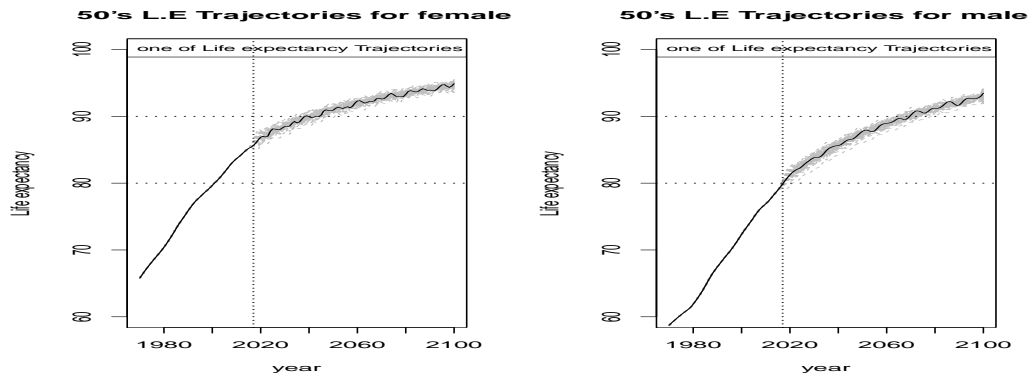
(a) One of TFR trajectories and 50th times replication (b) Fertility rate for Republic of Korea of WPPs 2019

Figure 3.2. Stochastic projections on total fertility rate.

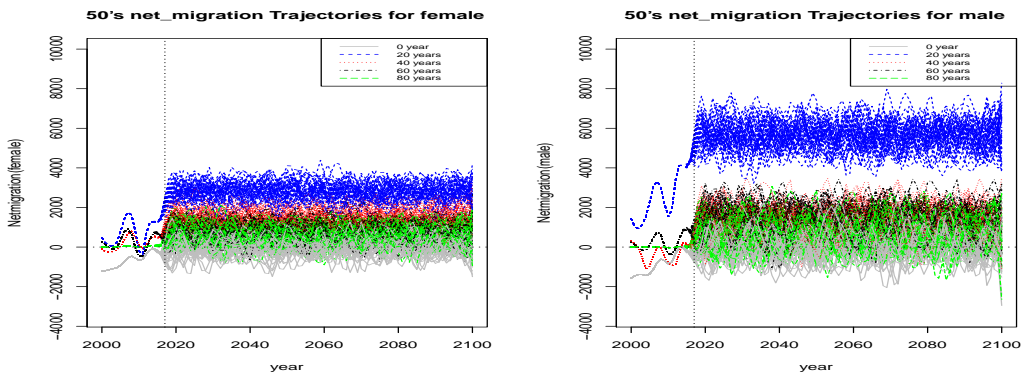
의 미표현, 시나리오 경직성, 미래궤적의 순서 불일치 등을 해결하고 있으며, 동시에 다양한 미래 궤적과 확률범위를 보여주고, 측정값들의 전후의 자기상관관계를 보이지 않고 있음을 알 수 있다.

본 논문은 이들 결과를 바탕으로 결정론적, 확률론적 인구추계를 시나리오 인구추계 한계점으로 제시한 정량적인 방식의 범위에 따른 확률표현 유무, 시나리오 선택, 미래 궤적의 불일치성, 자기상관관계 등의 관점에서 비교해본다. 첫째, 확률포함이다. 시나리오 인구추계는 고위, 저위가 경계지점으로 고위 이상의 값과 저위 이하의 값을 논하지 않는다. 하지만 미래는 불확실성이므로 이 경계 밖에 도출될 확률이 존재할 수도 있다. 확률론적 인구추계는 이런 경우를 풍부한 궤적으로 포함하고 있다. 둘째, 주관적인 시나리오 선택이다. 시나리오는 의사결정이나 정책 입안자들의 미래 목표치를 기준으로 선정하는 것이 일반적이다. 이런 작업은 주관적이고 자의적인 경우가 많다. 하지만 확률론적 인구추계는 확률적인 접근이므로 근본적으로 다르다. 셋째, 미래 궤적의 일방향 문제이다. 인구변화 3요소에서 여러 조합을 선정해 코호트 요인법으로 미래인구를 산출해서 궤적을 그려보면 거의 일방향이 대부분이다. 직전연도의 고위 출산율은 당해연도, 차기연도 고위 출산율으로 이어지고, 사망률, 국제이동률도 동일하다. 이런 결과에서 하나를 선정한 인구추계도 일방향 결과이다. 이러한 결과는 미래의 불확실성을 표현한다고 보기 어렵다. 실제로 출산율을 살펴보면 2000년대 이후 증감을 반복하고 있고 최근에는 0.98명 수준을 보이고 있다. 하지만 확률론적 인구추계는 Figure 3.3처럼 다양한 궤적을 보여주고 있다. 즉, 일방향이 아닌 증감과 같은 다양한 궤적을 보여주고 있다. 넷째, 예측오차들 간의 상관관계와 자기상관 문제이다. 세 번째와 유사한 것인데 인구변화 3요소의 확률론적 추계결과에는 오차가 존재한다. 각 예측치에는 오차들이 내포하고 있다. 이들 예측치를 조합해서 인구추계를 산출하면 그 결과 역시 오차들이 만들어 지는데 이들 오차들 간에도 상관관계나 자기상관이 존재한다. 하지만 확률론적 인구추계는 임의로 선정해서 오차들 간의 상관관계나 자기상관을 상쇄하는 결과를 제공한다.

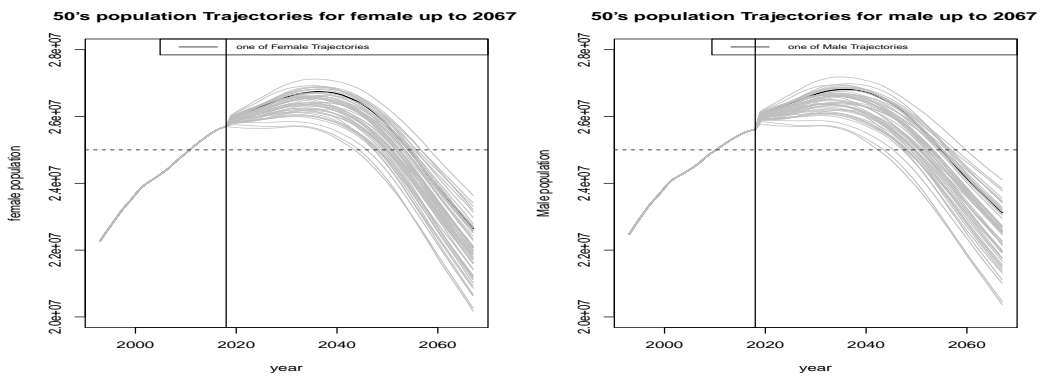
하지만 확률론적 인구추계는 위에서 언급한 제한점을 내포하고 있지 않다. 부가적으로 인구부양비와 인구구조 변화 관점에서 두 추계방법을 살펴보자. Figure 3.4는 인구부양비와 3계층(유소년층, 경제활동인구, 65세 이상)의 인구구조변화를 나타낸 것이다. 그림에서도 알 수 있듯이 시나리오 인구추계와 다른 결과로 보여짐을 알 수 있다. 이처럼 확률론적 인구추계는 결정론적 인구추계에서 표현하지 못하는 제한적인 정보를 표현하고 여러 단점을 보완하고 있다.



(a) Life expectancy



(b) Net migration for 0, 20, 40, 60, 80 years



(c) Stochastic population projections by gender

**Figure 3.3.** The result of stochastic projection on life expectancy, net migration, population by gender.  
 자료: 통계청 KOSIS, 출산율(1970-2017년), 사망률(1970-2016년), 국제순이동자수(2000-2017년). 주: 출산율은 GLG 모형과 시계열 적용, 사망률은 LC-ER 방법 (Li 등, 2013)을 적용, 국제이동자수는 FDM 적용.

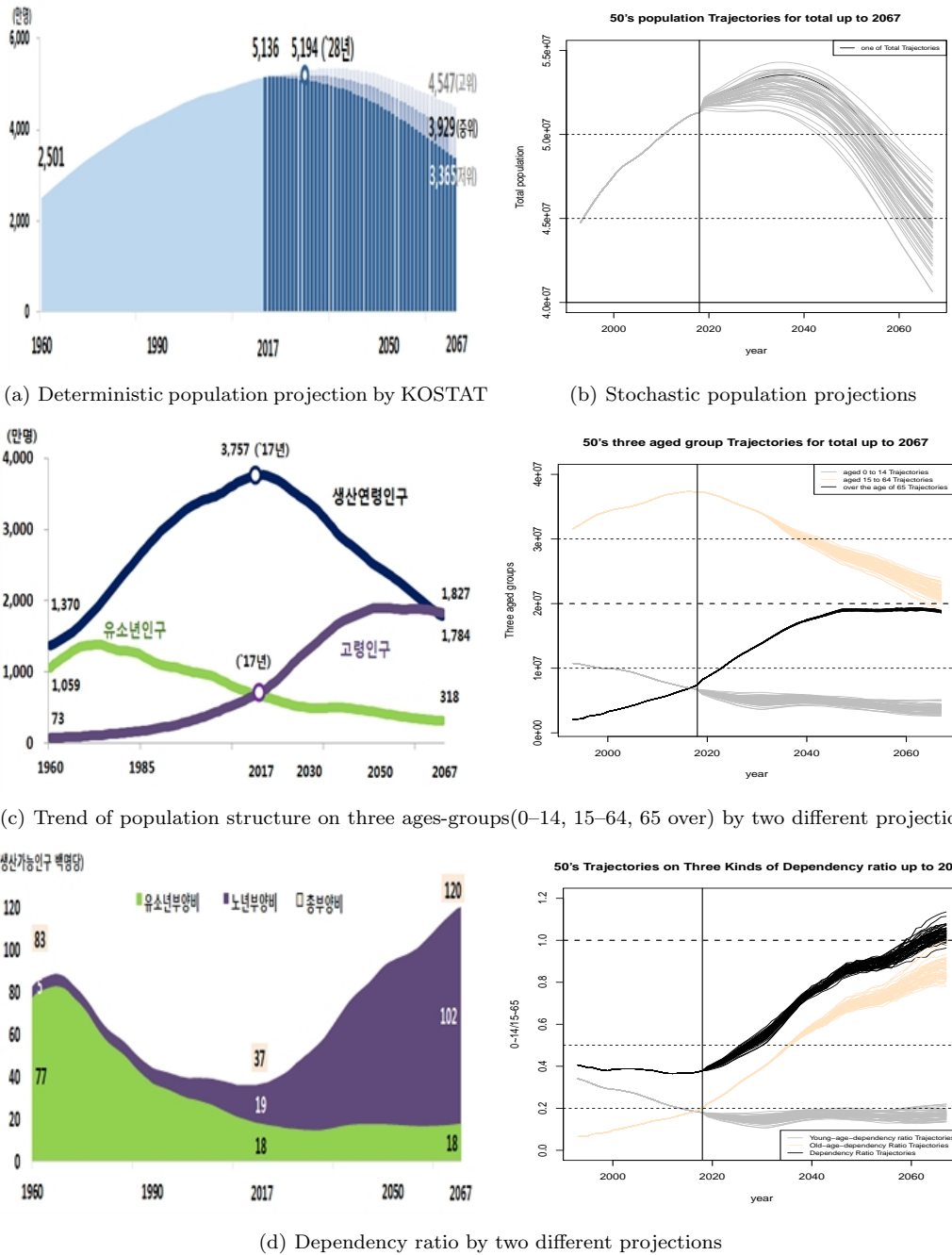


Figure 3.4. Population, trend of population structure for three ages-groups, and Dependency ratio.  
 주: 유소년부양비 = 0-14세/경제활동인구, 노년부양비 = 65세이상/경제활동인구, 총부양비 = 유소년부양비 + 노년부양비.  
 자료: 통계청 (KOSTAT, 2017, 2019).

그런데 확률론적 인구추계도 국가통계인 공식통계로 가기에는 몇 가지 제한점이 있다. 먼저 Woo (2010)에 따르면 확률론적 인구추계는 개별 연구자들은 연구자 개인이 관심을 갖는 변수에 초점을 맞추는 경향이 있지만 국가에 의한 공식적 인구추계의 경우 보편적이며 포괄적인 인구추계 결과를 제공해야 한다. 다음으로 통계사용자들의 수요를 충족해야하기 때문에 쉽게 이해할 수 있는 방식으로 결과물을 제시하여야 한다는 제약이 있다. 누구나 확률적 인구추계에 비해 시나리오 기반의 결정론적 인구추계가 정책입안자나 자료의 이용자들이 활용하기에 편리한 특성들을 갖고 있는 점은 부정할 수는 없으며, 인구추계는 미래의 불확실성 처리보다는 미래에 실현될 가능성 즉 개연성이 높은 단일의 예측 시나리오를 선호하는 경향이 강하다. 끝으로 결정론적 인구추계는 계산 방식이 간단하고 정책 입안자의 목표치 설정 구현이 용이한 반면 확률론적 인구추계는 목표치보다는 일어날 가능성에 대한 정량적인 방식의 범위에 따른 확률개념이고, 산출과정이 상대적으로 단순치 않다.

#### 4. 결론 및 제언

지금까지 결정론적, 확률론적 인구추계의 방법과 장단점을 살펴보았다. 시나리오 기반 결정론적 인구추계는 통계 사용자 편의와 해석의 용이성 등으로 보편적이며 포괄적이라는 점에서 공식통계로 공표되고 있다. 하지만 살펴본 바와 같이 변동성 크기와 양 표현 가능여부, 시나리오 경직성 여부, 추계구간의 비일관성, 단일 궤적의 비합리성 등과 같은 한계점이 존재한다.

그러므로 본 연구는 이런 제한점과 최근 인구추계의 동향이 확률론적 인구추계 추세에 초점을 맞춰 우리나라 장래인구추계를 확률론적 인구추계로 구현해 결과를 제시하고, 결정론적 인구추계의 한계점을 보완하고 있음을 확인하였다.

하지만 본 연구의 한계점도 존재한다. 첫째 결정론적 인구추계의 장점인 보편성과 해석의 용이성 측면이다. 확률론적 인구추계는 계량 인구학자 또는 통계학자들이 주로 사용하는 방법이고 통계학 지식이 있는 경우에만 올바른 해석이 가능하다. 이런 점에서 보편성을 떨어지며, 해석에 주의가 요구된다. 둘째 자료제한점이다. 우리나라는 출산율, 사망률 자료는 1970-2017년의 47년, 국제이동은 15년의 시계열을 보유하고 있지만, 예측기간은 이들 자료 시계열 기간보다 크다. 스웨덴, 프랑스처럼 80년 이상의 자료가 마련된다면 보다 정확하고 실현재현성과 불확실성 구현이 더욱 구체적일 것으로 판단된다. 셋째 결정론적 인구추계는 미래의 불확실성 처리보다는 미래에 실현될 가능성 즉 개연성이 높은 단일의 예측 시나리오이며, 이를 선호하는 경향이 강하다. 이런 측면에서 아직도 확률론적 인구추계가 일부 통계청(네덜란드 통계청, 영국, 스웨덴), UN, 선진 인구연구소(MPIDR, VID 등)에서만 공식통계로 발표하고 있다.

결론적으로 위와 같은 상황을 고려한다면 한 가지 제안을 이끌어 낼 수 있다. 즉 결정론적, 확률론적 인구추계를 조합하여 제시하는 것이다. 해석의 용이성과 실현재현성을 위해 결정론적 인구추계의 중위수준과 미래 불확실성을 나타내기 위해 풍부한 궤적을 보여주는 확률론적 인구추계 결과를 혼용하여 제시하는 것이다. 본 논문에서 도출된 시사점은 통계청의 30가지 시나리오 결과 제시보다 중위수준 결과에 확률론적 인구추계 정보를 추가적으로 제공하는 것이 시나리오 형태의 결정론적 인구추계의 단점을 보완할 수 있는 방안이라고 판단된다.

#### References

- Alho, J. M. (1997). Scenario, uncertainty and conditional forecasts of the world population, *Journal of Royal Statistical Society*, **160**, 71-85.
- Alho, J. M. (2005). Remarks on the use of probabilities in demography and forecasting. In N. Keilman (Ed),

- Perspectives on Mortality Forecasting (Vol 2. Probabilistic Models, 27–38), Swedish Social Insurance Agency, Stockholm.
- Alho, J. M. and Spencer, B. D. (1985). Uncertain population forecasting, *Journal of the American Statistical Association*, **80**, 306–314.
- Alders, M. and De Beer, J. (2004). Assumptions on fertility in stochastic population forecasts, *International Statistical Review*, **72**, 65–79.
- Alkema, L., Gerpand, P., Raftery, A., and Wilmoth, J. (2015). The United Nations Probabilistic Population Projections: an introduction to demographic forecasting with uncertainty, *Foresight (Colch)*, **37**, 19–24.
- Billari, F. C., Graziani, R., and Melilli, E. (2012). Stochastic population forecasts based on conditional expert opinions, *Journal of the Royal Statistical Society*, **175**, 491–511.
- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review, *International Journal of Forecasting*, **22**, 547–581.
- Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: a review of methods, *Annals of Actuarial Science*, **3**, 3–43.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26**, 211–252.
- De Beer, J. (2000). Dealing with uncertainty in population forecasting, *Statistics Netherlands*. Available from: [www.cbs.nl/nr/rdonlyres/7dc466f9-fe4c-48dc-be90-30216b697548/0/dealingwithuncertainty.pdf](http://www.cbs.nl/nr/rdonlyres/7dc466f9-fe4c-48dc-be90-30216b697548/0/dealingwithuncertainty.pdf)
- Dunstan, D. and Ball, C. (2016). Demographic projections: user and producer experiences of adoption a stochastic approach, *Journal of Official Statistics*, **32**, 947–962.
- Gerland, P., Raftery, A. E., Ševčíková, H., *et al.* (2014). World population stabilization unlikely this century, *Science*, **346**, 234–237.
- Hoem, J. M., Madsen, D., Nielsen, J. L., Ohlsen, E. M., Hansen, H. O., and Rennermalm, B. (1981). Experiments in modelling recent Danish fertility curves, *Demography*, **18**, 231–244.
- Horiuchi, S. and Wilmoth, J. R. (1995). *Aging of Mortality Decline*, Rockefeller University, New York.
- Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration, *International Journal of Forecasting*, **24**, 323–342.
- Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratation method with functional time series models, *Demography*, **50**, 261–283.
- Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach, *Computational Statistics & Data Analysis*, **51**, 4942–4956.
- Kaneko, R. (2003). Elaboration of the Coale-McNeil nuptiality model as the generalized log gamma distribution: a new identity and empirical enhancements, *Demographic Research*, **9**, 223–262.
- Keilman, N. (2005). Erroneous population forecasts. In *Perspectives on Mortality Forecasting: Vol II Probabilistic Models*, N. Keilman (Ed), Swedish National Social Insurance Board, Stockholm.
- Keilman, N., Pham, D. Q., and Hetland, A. (2002). Why population forecasts should be probabilistic - illustrated by the case of Norway, *Demographic Research*, **6**, 409–454.
- Keyfitz, N. (1981). The limits of population forecasting, *Population and Development Review*, **7**, 579–593.
- Kim, S. Y. and Oh, J. H. (2017). A study comparison of mortality projection using parametric and non-parametric model, *The Korean Journal of Applied Statistics*, **30**, 701–717.
- KOSTAT (2016). Population projection 2015~2065.
- KOSTAT (2019). The Special population projection 2017~2067.
- Lee, R. D. (1998). Probabilistic approaches to population forecasting, *Population and Development Review*, **24**, 156–190.
- Lee, R. D. (2004). Quantifying our ignorance: stochastic forecasts of population and public budgets, *Population and Development Review*, **30**, 153–175.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, **87**, 659–671.
- Lee, R. D. and Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: beyond high, medium, and low, *Journal of the American Statistical Association*, **89**, 1175–1189.
- Leslie, P. H. (1945). The use of matrices in certain population mathematics, *Biometrika*, **33**, 183–212.

- Leslie, P. H. (1948). Some further notes on the use of matrices in population mathematics, *Biometrika*, **35**, 213–245.
- Li, N. and Gerland, P. (2011). Modifying the Lee-Carter Method to Project Mortality Changes up to 2100, the Population Association of America 2011 Annual meeting-Washington, DC, Session 125, formal Demography I: Mathematical Models and Methods.
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method, *Demography*, **42**, 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**, 2037–2051.
- Li, Q., Reuser, M., Kraus, C., and Alho, J. (2007). *Aging of a giant: a stochastic population forecast for China, 2001-2050*, MPIDR Working Papers WP-2007-032, Max Planck Institute for Demographic Research, Rostock, Germany.
- Lutz, W. (2009). *Toward a Systematic, Argument-Based Approach to Defining Assumptions for Population Projections* (Interim Report IR-09-037), International Institute for Applied Systems Analysis, Austria. Available from: <http://pure.iiasa.ac.at/9115/1/IR-09-037.pdf> (accessed 19 May 2016).
- Lutz, W., Butz, W. P., and KC, S. (2014). *World Population & Human Capital in the Twenty-first Century: Executive Summary*, IIASA, Laxenburg, Austria.
- Lutz, W. and Goldstein, J. R. (2004). Introduction: How to deal with uncertainty in population forecasting?, *International Statistical Review*, **72**, 1–4.
- MPIDR (2006). Stochastic forecast of the population of Poland, 2005-2050. MPIDR working paper WP-2006-26.
- MPIDR (2007). Aging of a giant: a stochastic population forecast for Poland, 2005-2050. MPIDR working paper WP-2007-32.
- Oh, J. H. (2018). A comparison between the real and synthetic cohort of mortality for Korea, *The Korean Journal of Applied Statistics*, **31**, 427–446.
- ONS (2009). 「progress report on developing stochastic population forecasts for the United Kingdom」.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed), Springer-Verlag, New York.
- Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries, *Proceedings of the National Academy of Sciences*, **109**, 13915–13921.
- Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries, *Demography*, **50**, 777–801.
- SCB (2006). *Stochastic Population Projections for Sweden*, Research and Development, Methodology reports from Statistics Sweden.
- Smith, D. P. and Keyfitz, N. (1977). *Mathematical Demography: Selected Papers*, Springer Verlag, Berlin.
- Smith, S. K., Tayman, J., and Swanson, D. A. (2001). *State and Local Population Projections: Methodology and Analysis*, Kluwer Academic/Plenum Publishers, New York.
- Stoto, M. (1983). The accuracy of population projections, *Journal of the American Statistical Association*, **78**, 13–20.
- Statistics Canada (2015). Population Projections for Canada (2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions.
- Statistics NZ (2016). National Population Projections: 2016(base)-2068. Available from: <http://www.stats.govt.nz/browse-for-stats/population/estimates-and-projections/NationalPopulationProjectionsHOTP2016.aspx> (accessed 25 October 2016)
- UN (2017). World Population Prospects 2017.
- UN (2019). World Population Prospects 2019.
- UNICEF (2017). Level & Trends in Child Mortality.
- VID (2016). 「40 years of the Vienna Institute of Demography 1975-2015」.
- Woo, H. B. (2010). Stochastic demographic and population forecasting, *Korea Journal of Population Studies*, **33**, 161–189.



# 미래의 불확실성에 대한 확률론적 인구추계

오진호<sup>a,1</sup>

<sup>a</sup>한밭대학교 공과대학 기초과학부

(2019년 11월 21일 접수, 2019년 12월 30일 수정, 2020년 2월 12일 채택)

---

## 요약

예전부터 시나리오 인구추계(scenario population projection)는 미래 실현개연성이 높은 상황 반영과 통계적 음해석 용이성으로 각광을 받아왔다. 통계청 (2019)도 특별 시나리오를 포함한 30가지 조합 결과를 공식통계로 제시하고 있다. 하지만, 이런 결정론적(determinant) 인구추계는 미래의 불확실성(uncertainty)에 대해 제한적으로 정보를 제공하고, 시나리오 기반 예측치이므로 확률적이지 않으며, 시간에 따라 인구변동 3요소(출산, 사망, 이동)들의 완벽한 자기상관을 보이는 등 여러 한계점이 있다. 따라서 국제기구 UN, 독일 막스플랑크 인구연구소(MPIDR), 오스트리아 비엔나인구연구소(VID) 등은 확률론적(stochastic) 기반 인구추계를 제시하고 있다. 더불어 해외 일부 국가 통계청에서도 이 방식을 도입해 시나리오 결과와 함께 정보를 제공하고 있다. 본 논문은 우리나라의 인구추계를 확률론적 기반으로 산출한 후, 시나리오(결정론적) 인구추계 결과와 비교해 장·단점과 시사점을 도출해본다.

주요용어: 시나리오, 결정론적 인구추계, 불확실성, 확률론적 인구추계

---

이 논문은 2019년 한밭대학교 신입교수연구비의 지원을 받았다.

<sup>1</sup>(34158) 대전광역시 유성구 동서대로 125, 한밭대학교 공과대학 기초과학부. E-mail: jhoh75@hanbat.ac.kr