

확률적 인구추계*

우해봉**

인구추계 분야에서 경험한 최근의 발전 상황은 불확실성에 대한 정확한 이해와 이에 대한 적절한 방법론적 대처가 핵심적인 이슈 중의 하나임을 보여 주고 있다. 이러한 측면에서 본 연구는 1980년대 이후 점차 활용도가 높아지고 있는 확률적 인구추계의 주요 내용과 방법들을 소개하였다. 구체적으로 본 연구는, 첫째, 인구변동 요인들의 미래 추세 예측에 수반된 불확실성의 특성 그리고 이러한 불확실성의 처리 방법으로 전통적인 시나리오 접근과 상대적으로 최근에 주목을 받기 시작한 확률적 접근을 개관하였다. 시나리오 접근의 한계와 관련해서는 확률적 해석의 불가능, 시나리오의 경직성, 예측구간의 비일관성의 문제를 검토하였다. 둘째, 확률적 인구추계의 특성들을 살펴봄과 함께 인구변동 요인들에 대한 확률적 예측을 위해 현재까지 가장 빈번하게 활용되고 있는 전문가 판단법, 통계적 모형에 기초한 방법, 그리고 과거의 예측 오차를 활용한 방법의 주요 내용과 장단점을 살펴보았다. 마지막으로 이러한 논의를 기초로 인구추계에 있어서 확률적 접근이 해결해야 할 과제와 향후 발전 방향 그리고 인구추계 분야에서 이루어진 최근의 발전 상황들이 갖는 시사점을 논의하였다.

핵심단어: 확률적 인구추계, 불확실성, 전문가 판단, 통계적 모형, 과거 예측 오차 분석

I. 서론

우리나라는 현재 출산력 감소와 기대여명 증가로 인하여 인구 구조의 급격한 고령화가 진행되고 있다. 1960년대 초반까지 6.0 수준이었던 합계출산율(TFR)은 1983년에 대체출산율 이하로 떨어졌으며, 2005년 1.08의 최저 수준을 기록한 이후 2008년 현재 1.19 수준에 머물고 있다. 기대수명의 경우 1960년대 전반까지 남녀 모두 60세에도 미치지 못했지만, 2008년 현재 남성 76.5세와 여성 83.3세로 급격히 증가하는 모습을 보였다. 이러한 인구변동 요인들에서의 변화는 결과적으로 인구 구조에 있어서의 급격한 변화를 야기했는데, 1960년대에 전체 인구의 3%

* 본 연구는 국민연금연구원에서 수행한 「장래인구추계의 방법과 현황」의 일부분을 수정 및 발전시킨 것이다. 본 연구에 대해서 유익한 논평을 해 주신 익명의 논평자들에게 감사드린다.

** 국민연금연구원 부연구위원 | haebongwoo@gmail.com

수준에 머물렀던 65세 이상 노인인구의 비율은 2008년(7월 1일 기준)에는 10.3% 수준으로 증가함으로써 다른 사회에 비해 훨씬 빠른 속도로 인구의 고령화 현상을 경험하고 있다.

비록 기대여명의 증가와 같은 최근의 인구학적 변화가 개인들의 삶의 질 향상에 기여하는 긍정적인 측면을 과소평가할 수는 없지만, 현재 우리나라가 경험하고 있는 인구 구조의 급격한 고령화 추세는 국민연금을 포함한 사회보장제도의 재정적 부담을 가중시키는 부정적 함의를 동시에 지니고 있다. 이에 따라 사회보장제도의 안정적 운영을 위한 장기적 계획 수립을 위해서는 장래인구의 규모와 구조에 관한 보다 정확한 인구추계가 필요하다는 인식 또한 증가하고 있다.

비록 정확한 인구추계의 중요성이 강조되고 있음에도 불구하고, 현재와 같이 인구변동 요인들에 관한 정확한 추세 전망이 어려운 시기의 경우 인구추계 결과물에 수반된 불확실성 또한 매우 높은 것이 현실이다. 인구추계의 정확성 그리고 인구추계에 수반된 불확실성에 대한 이해가 이렇듯 중요한 함의를 갖고 있음에도 불구하고 현재까지 인구추계 방법에 대한 국내의 인구학적 관심은 상대적으로 매우 부족했다. 또한 인구추계 방법을 소개하는 문헌의 경우에도 대체로 인구변동 요인들에 관한 단일 혹은 다양한 가정들을 코호트-요인법(cohort-component method)에 적용하는 시나리오 접근법에 국한되는 모습을 보임으로써 인구추계 방법에서 이루어진 최근까지의 발전 상황을 적절히 반영하지 못하는 한계를 보이고 있다¹⁾.

본 연구의 목적은 문헌 고찰을 통하여 1980년대 이후 인구추계 분야에서 그 활용도를 넓혀가고 있는 확률적 인구추계(stochastic demographic and population forecasting)를 소개하는 것이다²⁾. 미래에 대한 다른 모든 예측과 마찬가지로 인구추계의 경우도 불가피하게 불확실성을 수반하고 있는데, 전통적인 시나리오 접근과 확률적 접근이 갖는 기본적인 차이는 인구추계에 수반된 이러한 불확실성을 어떠한 방식으로 처리하는가에 있다고 할 수 있다³⁾.

1) 코호트-요인법이 시나리오 접근법(scenario approach)에서 빈번하게 활용되는 관계로 코호트-요인법을 시나리오 접근법과 동일한 의미로 사용하기도 한다. 그러나 뒤에서 설명하겠지만 코호트-요인법은 시나리오 접근뿐만 아니라 확률적 접근(stochastic approach)에서도 빈번하게 활용된다는 측면에서 두 개념은 구분될 필요가 있다.

2) 본 연구에서는 인구추계의 최종 결과물인 인구 규모와 구조뿐만 아니라 출산력, 사망력, 인구이동과 같은 인구변동 요인들에 대한 확률적 예측을 포함하는 광의의 개념으로 확률적 인구추계를 사용하기로 한다. 이는 본 연구에서 검토하는 확률적 예측 방법들이 최종 결과물인 인구 규모 및 구조보다는 인구변동 요인들의 예측과 보다 밀접히 연관된 것과 관련이 있다.

3) 연구자에 따라 추계와 예측을 엄격히 구분하기도 하지만, 본 연구에서는 인구추계 분야에서 이루어진 최근의 논의처럼 추계와 예측을 엄격히 구분하지 않는다. 이들 두 용어의 사용과 관련된 보다 자세한 내용에 관해서는 우해봉(2009)을 참조할 수 있다.

서론에 이어 제2장에서는 인구추계에서 발생하는 불확실성의 문제를 살펴보고 이러한 불확실성의 처리와 관련하여 전통적인 시나리오 접근이 갖고 있는 문제점을 검토한다. 제3장에서는 확률적 인구추계를 개관함과 함께 현재 가장 널리 활용되고 있는 확률적 예측 방법들의 주요 내용과 장단점을 살펴본다. 본 연구에서 검토될 확률적 인구추계의 주요 방법들은 대체로 최종 결과물인 인구 규모나 구조보다는 인구변동 요인들에 대한 예측과 관련해서 논의되는 경향이 있다. 이러한 점은 시나리오 접근과 마찬가지로 확률적 접근에서도 코호트-요인법이 광범위하게 활용되고 있으며, 인구추계와 관련된 불확실성의 가장 큰 원천이 인구변동 요인들에 있다는 점 그리고 확률적 인구추계에 관한 기존의 방법론적 연구 또한 인구변동 요인들에 대한 확률적 예측에 초점을 맞춘다는 점과 관련이 있다. 마지막으로 제4장에서는 확률적 인구추계의 발전 방향과 함께 인구추계 분야에서 이루어진 최근의 발전 상황들이 갖는 시사점을 논의한다.

II. 연구 배경

1. 불확실성과 인구추계

미래에 관한 다른 모든 예측과 마찬가지로 인구학적 예측 또한 불가피하게 불확실성(uncertainty)을 내포하고 있다. 인구추계의 불확실성은 일정 부분 현 시점의 인구학적 정보가 불확실한 것에 기인한다(예컨대, 기준연도 정보의 부정확성). 그러나 인구추계 과정에서 발생하는 불확실성의 가장 중요한 원천은 인구변동 요인들의 미래 추세, 즉 출산력, 사망력, 인구이동 예측에 수반되는 불확실성이라고 할 수 있다(De Beer, 2000)⁴⁾.

인구변동 요인들을 상호 비교할 때 단기적으로는 인구이동으로 인한 불확실성의 정도가 가장 높은 것으로 지적된다. 그러나 장기적인 관점에서 볼 때 인구이동은 순환적인 패턴을 보인다는 측면에서 예측에 수반되는 불확실성의 정도는 상대적으로 낮아지는 경향이 있다. 반면 출산력의 경우 단기적으로 커다란 변화를 보이지는 않지만 장기적인 불확실성의 정도는 매우 큰 특성을 보인다. 이는 가임 연령대에 위치한 여성의 규모 예측에 수반된 불확실성뿐만 아니라 이들 여성들이 각각 얼마의 자녀를 출산할 것인가에 관한 추가적인 불확실

4) 「인구추계」 그 자체도 추가적인 불확실성을 발생시키는 피드백 효과를 가질 수 있다. 특히, 사회적으로 부정적인 합의를 지닌 인구추계 결과는 이를 수정하기 위한 후속의 정책적 개입을 촉발시키는 중요한 계기가 되기도 한다.

성이 발생하기 때문이다. 출산력이나 인구이동에 비해 사망력은 대체로 방향성을 갖고 점진적으로 변하는 경향이 있기에 추세 예측과 관련된 불확실성의 정도는 상대적으로 낮은 것으로 평가되고 있다(De Beer, 2000).

인구변동 요인들과 관련된 전망이 불확실성을 수반하고 있음을 고려할 때 이를 기초로 하고 있는 인구추계 결과물 역시 불확실성을 수반하고 있음은 당연하다. 빈번하게 인용되고 있는 키피츠(Keyfitz, 1981)의 다음과 같은 지적에서 볼 수 있듯이 인구추계 작업에서 불확실성에 대한 고려는 이미 수십 년 전부터 그 중요성이 인식되어 왔다.

Demographers can no more be held responsible for inaccuracy in forecasting population 20 years ahead than geologists, meteorologists, or economists when they fail to announce earthquakes, cold winters, or depressions 20 years ahead. What we can be held responsible for is warning one another and our public what the error of our estimates is likely to be (Keyfitz, 1981: 579).

인구추계 과정에 수반된 불확실성의 처리와 관련하여 기존의 연구들은 서로 상이한 두 가지 접근을 취해 왔는데, 시나리오 접근(scenario approach)과 확률적 접근(stochastic approach)이 그것이다. 전통적으로 인구추계에 수반된 불확실성을 처리하는 방식은 실현 개연성이 가장 높은 예측치와 함께 복수의 예측치들을 대안적 시나리오로 제공하는 방식이라고 할 수 있다. 예를 들어, 국내외의 인구추계에서 널리 활용되고 있는 고위(high), 중위(medium), 저위(low) 가정에 기초한 복수의 시리즈 제공은 인구추계에 수반된 불확실성을 처리하는 하나의 방식이라고 할 수 있다. 특히, 국가에 의한 공식적 인구추계의 경우 인구변동 요인들과 관련된 상이한 가정들을 코호트-요인법에 적용하여 복수의 시나리오를 제공하는 방법이 인구추계에 수반된 불확실성을 처리하는 표준적인 방식으로 자리 잡고 있다.

시나리오 접근에서 제공되는 복수의 시리즈에 대해 다음과 같은 두 가지의 해석이 가능하다(Smith, Tayman and Swanson, 2001: 332-333). 첫째, 개별 시리즈는 미래의 인구변동에 대한 하나의 가능한 전망을 제시할 뿐이며, 결과적으로 어떤 특정한 시리즈가 다른 시리즈에 비해 우월하다는 것을 의미하지는 않는다는 해석이다⁵⁾. 복수의 시리즈에 대한 두 번째 해석은, 비록 모든 개별

5) 복수의 추계 시리즈 제공과 관련된 이러한 해석은 과거 미국 센서스국이 취한 입장으로 알려진다. 미국 센서스국은 실현 개연성이 가장 높은 시리즈의 명시적 지정을 거부하는 동시에 인구추계 작

시리즈가 미래에 실현 가능한 하나의 전망을 제시함에도 불구하고, 어떤 특정한 시리즈(대체로, 중위 시리즈)가 다른 시리즈들에 비해 상대적으로 실현 개연성이 높다는 것이다. 이러한 입장에 기초할 경우 선호되는 혹은 실현될 개연성이 가장 높은 어떤 특정한 시리즈가 지정되면 나머지 시리즈들은 이에 대한(실현 개연성이 떨어지는) 대안으로 해석되어진다.

불확실성의 처리와 관련하여 시나리오 접근에서 빈번히 사용되는 고위(high)와 저위(low) 시리즈에 관해서도 상이한 해석이 존재한다(Smith et al., 2001: 333). 우선, 비록 고위와 저위 시리즈에 의해서 형성되는 범위가 엄격한 확률적 의미를 가지는 것은 아니지만, 인구 변수들이 미래에 실현될 값들의 그럴듯한 범위를 이들 고위와 저위 시리즈가 제시한다는 해석이 가능하다. 그러나 이러한 해석과 달리 개별 시리즈는 단지 미래에 실현될 수 있는 (수많은) 가능한 대안들 중의 하나일 뿐이라는 해석을 따를 경우 미래에 실현될 인구 변수의 값이 이들 고위와 저위 시리즈 사이에 위치할 것이라는 기대는 하지 않는다. 비록 암묵적이지만, 인구 변수들이 미래에 실현될 범위(상하한)를 이들 고위와 저위 시리즈가 보여 준다는 해석이 상대적으로 많이 사용되는 경향이 있다.⁶⁾

불확실성의 처리와 관련된 시나리오 접근의 장점은 일반적으로 추계 자료의 사용자들은 (장래에 실현될 개연성이 가장 높은 시나리오를 중심으로 한) 예측 구간(predictive interval)보다는 실현 개연성이 가장 높은 시나리오와는 독립적 이며 내적 일관성을 갖춘 대안들을 필요로 한다는 점이다. 특히, 정치 공동체의 경우 미래에 발생할 개연성이 가장 높은 어떤 특정의 상황에 대한 다양한 대안의 모색에 많은 관심이 주어지는 경향이 있다. 예컨대, 인구 감소가 갖는 부정적인 효과가 관심의 대상일 경우 출산 장려를 위한 각종의 정책들이 어느 정도의 효과를 가질 것인가에 대한 분석은 매우 중요한 함의를 갖는다. 이러한 경우 실현 개연성이 가장 높은 예측과 정책적 효과를 반영한 대안적 예측을 비교 분석할 수 있는 시나리오 접근법이 유용하게 활용될 수 있는 것이다.

인구추계에 내재된 불확실성을 처리하는 또 다른 방식은 본 연구가 초점을 맞추어 검토하고자 하는 확률적인 접근이다. 확률적 접근을 지지하는 연구자들은 전통적인 시나리오 접근이 인구추계에 수반된 불확실성에 대한 충분한 고려

업이 예측을 위해서 이루어지지 않았음을 강조하였다(자세한 사항은 스미스 외(Smith et al., 2001)를 참조). 고위, 중위, 저위 시리즈의 제공 대신 2 혹은 4개의 시리즈 제공, 혹은 수십 개의 시리즈를 제공한 것도 이러한 측면에서 해석해 볼 수 있다.

6) 그러나 고위와 저위 시리즈에 의해서 형성된 범위와 실제치를 비교할 경우 실제치가 고위와 저위 시리즈에 의해서 설정된 범위를 빈번히 벗어나고 있음이 보고된다(Alho and Spencer, 1997: Keyfitz, 1981: Stoto, 1983).

를 하지 않은 채 인구추계를 인구변동 요인들에 관한 일련의 기계적인 가정 설정 작업으로 인식하고 있음을 비판한다. 기본적으로 확률적 접근은 시나리오 접근에 비해 불확실성의 통계적 측정에 초점을 맞추는데, 미래에 대한 다른 예측과 마찬가지로 인구추계 또한 불확실성을 지니고 있기에 추계 대상 인구 변수가 미래에 실현될 가능한 범위(range)와 이러한 범위가 갖는 확률(probability)을 구체적으로 명시할 필요가 있음을 강조한다(Keilman, Pham and Hetland, 2002; Lee, 2004).

2. 시나리오 접근의 문제점

1) 확률적 해석의 불가능

인구추계가 불가피하게 불확실성을 수반한다는 점을 고려할 때 보다 적실성 있는 접근은 예측의 대상이 되는 변수가 미래에 실현될 범위(range)와 이러한 범위가 지시하는 확률(probability)을 명시하는 것이다(Keilman et al., 2002). 앞에서도 언급했듯이 전통적인 시나리오 접근의 경우 인구추계 과정에 수반된 불확실성을 반영하기 위하여 실현 개연성이 가장 높은 중위 시리즈에 추가하여 고위와 저위 등의 추가적인 시리즈를 제공하는 경향이 있다. 그러나 불확실성에 대한 시나리오 접근은 인구 변수가 미래에 실현될 개연성에 관한 확률적 해석을 내리지 못하는 문제가 있다(Booth, 2006; Keilman et al., 2002; Lee, 1998, 2004; Lutz, Sanderson and Scherbov, 1998).

전통적인 시나리오 접근에서 제시되는 상한(고위 시리즈)과 하한(저위 시리즈) 사이의 구간은 명확한 확률적 의미를 갖지 않으며, 단지 인구추계를 담당하는 기관이나 연구자들이 실현 개연성이 높다고 판단하는 범위를 지시함에 불과하다. 결과적으로 인구추계 자료의 이용자들은 고위 및 저위 시리즈를 어떻게 해석해야 할 것인가에 관한 명확한 지침을 받지 않는 한 이들 대안 시나리오들을 효과적으로 사용하기 어려운 것이 현실이며, 인구추계 자료에 기초하여 어떤 후속의 계획을 수립해야 하는 경우 불확실성에 대한 양화가 가능하지 않기에 계획의 유연성 확보에 많은 어려움이 있다.

2) 시나리오의 경직성

시나리오 접근의 경우 일반적으로 개별 시리즈에서 설정된 동태율은 모든 연령대에 걸쳐 그리고 전 추계기간에 걸쳐 동시적으로 결정된다는 가정에 기초하

고 있는데, 결과적으로 인구변동 요인들의 값(계적)은 (정적(+))이든 혹은 부적(-)이든 완벽히 상관(serial correlation)되어 있음을 가정하게 된다(Booth, 2006; Lee, 2004; National Research Council, 2000). 예컨대, 「고위」 시리즈에 해당하는 출산율 가정의 경우 추계기간의 「모든」 연도에 걸쳐 다른 시리즈의 출산율 가정에 비해 항상 높게 설정된다. 마찬가지로 어떤 특정한 연도의 출산율이 낮게 설정될 경우 이는 해당 시리즈의 「모든」 연도의 출산율이 또한 (다른 시리즈들에 비해) 낮게 설정되어 있음을 의미한다.

시나리오를 도출하기 위해 출산력과 사망력을 동시에 고려할 경우 시나리오 접근은 개별 인구변동 요인들이 완전한 시계열 상관을 보일 뿐만 아니라 인구변동 요인들 사이에도 완전한 상관관계가 있음을 가정하게 된다. 예컨대, 출산력과 사망력을 조합하여 총인구 예측의 고위 시리즈를 구성할 경우 추계기간의 모든 연도에 걸쳐 높은 수준의 출산율 가정은 항상 낮은 수준의 사망률 가정과 조합을 이루게 된다.

이러한 시나리오의 경직성은 결과적으로 추계기간 동안 개별 인구변동 요인들에서 나타날 수 있는 유의미한 변화를 수용하지 못하기에 베이비 붐 현상의 발생이나 소멸과 같은 예측하지 못한 「급격한 변동」은 상정되지 않는다. 마찬가지로 인구변동 요인들 사이의 완전한 상관을 상정하기에 인구변동 요인들 사이에 존재할 수 있는 다양한 조합을 예측 과정에서 적절히 반영하지 못하는 결과를 가져 온다(Keilman et al., 2002; Lee, 2004).

3) 예측구간의 비일관성

시나리오에 기초한 전통적인 접근은 또한 인구변동 요인들에 관한 가정의 경직성으로 인해 확률적으로 일관성이 없는 결과를 산출한다(Alho and Spencer, 1985; Lee, 1998, 2004; Lee and Tuljapurkar, 1994). 일반적으로 시나리오 접근의 경우 불확실성을 수용하기 위해 실현 개연성이 가장 높은 중위 가정에 추가하여 고위와 저위의 가정을 도입하는데, 이러한 시나리오를 형성하기 위한 인구변동 요인별 가정들의 조합은 대체로 특정의 인구추계를 산출하는 목적에 의존하는 경향이 있다. 예를 들면, 미국 센서스국(Bureau of Census)은 고위 출산력과 저위 사망력, 그리고 고위 순인구이동을 결합하여 인구 성장의 고위 시리즈(Highest Series)를 그리고 저위 출산력과 고위 사망력, 그리고 저위 순인구이동을 조합하여 인구 성장의 저위 시리즈(Lowest Series)를 구성하는 경향이 있다.

미국 센서스국의 이러한 조합에 비해 사회보장청(Social Security Administration) 산하 계리국(Office of the Actuary)은 고위 출산력과 고위 사망력, 그리고 고위 순 인구이동을 조합하여 저비용 시나리오(Low Cost Scenario)를, 그리고 각 인구변동 요인들의 저위 가정들을 조합하여 고비용 시나리오(High Cost Scenario)를 형성하는데, 이는 이러한 조합이 사회보장 재정추계에서 핵심적인 역할을 하는 노인부양비가 이들 저위 및 고위 시나리오와 밀접히 연계된다는 점과 관련이 있다.

미국 센서스국과 사회보장청이 설정하는 이러한 가정들의 조합을 비교할 때 센서스국의 인구변동 요인들에 대한 가정들의 조합은 노인부양비의 상하한 범위를 과소평가할 개연성이 크다고 볼 수 있다. 이는 센서스국에서 설정한 고위 시리즈의 경우 고위 출산력은(중단기적으로) 인구 구조의 연소화를 그리고 저위 사망력은 인구 구조의 고령화를 가져오기에(노인부양비의 계산에 있어서) 두 효과가 상당한 수준으로 상쇄된다는 것과 관련이 있다.

이에 비해 노인부양비에 초점을 두는 사회보장청 인구추계의 경우 인구 성장률에서 나타나는 변이를 과소평가할 개연성이 높는데, 이는 사회보장청의 저비용 시나리오의 경우 고위 출산력은 높은 인구 성장률을 그리고 고위 사망력은 낮은 인구 성장률을 가져오기에 두 효과가 상쇄되는 것과 관련이 있다. 결과적으로 인구추계 담당 기관의 특수한 목적에 맞추어 인구변동 요인들 사이의 조합을 형성하는 시나리오 접근은 분석 대상(예컨대, 인구성장률) 외의 다른 인구 변수(예컨대, 노인부양비)의 예측에 있어서는 일관성 없는 예측 결과를 산출하는 문제를 갖고 있는 것이다.

Ⅲ. 확률적 인구추계

1. 확률적 인구추계의 개관

비록 1980년대 초에 이미 확률적 인구추계에 관한 연구들이 등장했지만, 인구추계에 수반된 불확실성을 확률적으로 고려하는 방법들이 빈번하게 사용되기 시작한 것은 비교적 최근의 일이다. 사실 인구학의 전반적 발전 상황을 볼 때 인구추계 분야처럼 통계적 원리를 받아들이지 않고 과거의 방식을 지속적으로 사용하는 분야를 찾아보기는 쉽지 않다.

부스(Booth, 2006)는 인구추계 분야에서 확률적 인구추계에 대한 연구가 증가한 것은 확률적 인구추계에 대한 사용자들의 요구가 증가한 결과라기보다는 인구학자들이 통계적 근거가 충분하지 않은 전통적 인구추계 방법들에 대해 비판

적으로 검토하는 동시에 다른 분과 학문에서의 발전 상황을 적극적으로 수용하기 시작한 것과 관련이 있는 것으로 추측하고 있다. 확률적 인구추계에 대한 관심의 증가는 또한 인구추계에 있어서 정확성 향상이 쉽지 않기에 불확실성에 대한 고려가 중요한 이슈로 등장한 것과도 일정 부분 관련이 있다(우해봉, 2009).

인구추계에 있어서 확률적 접근의 필요성은 확률적 접근이 인구추계에 수반된 이러한 불확실성을 명시적으로 측정한다는 점과 밀접히 연관되어 있다. 물론 단기 예측의 경우 인구추계 결과에 수반된 불확실성의 영향력은 상대적으로 제한적이다. 예를 들면, 향후 5년을 예측하는 인구추계의 경우 연구자는 전통적으로 활용된 결정론적 방식으로 비교적 확실성이 높은 인구추계를 수행할 수 있다. 그러나 중장기 예측의 경우 예측에 수반되는 오차의 크기는 추계기간이 길어짐에 따라 급격히 증가하는 경향이 있다. 그러므로 인구변동 요인들에 수반된 불확실성이 무시할 정도가 아닌 한 결정론적인 장기 인구추계는 정책적 시사점의 제시에 있어서 제한적이며, 인구 변수와 관련된 장기 예측의 경우 점 예측(point forecast)보다는 구간 예측(interval forecast)의 효용성이 높아진다(Keilman et al., 2002; National Research Council, 2000).

본 연구의 경우 확률적 접근을 시나리오 접근과 대비하는 방식을 취하고 있다. 앞서서도 언급했듯이 확률적 인구추계를 코호트-요인법과 대비되는 것으로 취급하지 않는 이유는 코호트-요인법이 확률적 인구추계에서도 빈번하게 활용되기 때문이다. 비록 확률적 접근과 시나리오 접근 모두 코호트-요인법에 기초할 수 있지만, 불확실성을 처리하는 방식에 있어서 두 접근은 근본적인 차이를 보인다. 전통적인 시나리오 접근이 복수의 시나리오를 도입함으로써 인구추계에 수반된 불확실성을 고려하는 반면 확률적 인구추계의 경우 인구변동 요인들의 동태율을 확률변수로 취급함으로써 불확실성을 반영한다(Alho and Spencer, 2005). 보다 구체적으로 확률적 인구추계에서 어떤 인구 변수의 미래 예측치들은 전통적으로 예측분포(predictive distribution)라고 불리는 확률분포를 갖는 확률변수로 해석되며, 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. 또한 개별 인구변동 요인들의 분포와 관련하여 세부적인 정보(예를 들면, 인구변동 요인들의 「연령별」 동태율) 대신 합계출산율(total

fertility rate), 기대수명(life expectancy at birth), 순인구이동(net migration)과 같은 요약 지표들의 분포에 초점을 맞추는 경향이 있다. 기준인구(base population) 정보가 완전하다는 가정 또한 빈번히 사용되고 있는데, 일반적으로 이러한 가정은 경제적으로 발전된 국가들의 경우에 수용 가능한 가정이라고 평가할 수 있다(Keilman et al., 2002).

인구변동 요인들의 요약 지표들에 관한 확률적 예측을 수행하기 위해 현재 가장 일반적으로 사용되고 있는 방법들로는 시계열 방법(time series method), 전문가 판단법(expert judgment method), 과거의 예측 오차 분석법(historical forecast error method)이 있다⁷⁾. 이러한 방법들은 상호 배타적이기 보다는 상호 보완적인 관계에 있다고 할 수 있는데, 최근에 이루어진 연구들 또한 하나의 방법보다는 여러 가지 방법들을 병용하는 경향이 있다(예컨대, Keilman et al., 2002).

확률적 인구추계의 경우 예외적으로 장래인구의 규모를 직접 예측하는 경우도 있지만, 보다 일반적인 접근은 시계열 모형 등을 통해 개별 인구변동 요인들의 요약 지표(예컨대, 합계출산율)에 대한 예측을 한 후 코호트-요인법을 통하여 이를 통합하는 방식이다. 일반적으로 코호트-요인법을 통해 예측분포를 도출하는 과정은 시뮬레이션을 통해서 이루어지는데, 인구변동 요약 지표들(예컨대, 합계출산율, 기대여명, 순인구이동)의 예측분포로부터 랜덤으로 도출된 예측값(궤적)을 코호트-요인법에 적용하여 추계인구를 도출하는 과정을 반복함으로써 예측분포를 도출하게 된다(Booth, 2006; Lee, 1998; Lee and Tuljapurkar, 1994)⁸⁾.

2. 전문가 판단법(Expert Judgment Method)

1) 개요

전문가 판단법은 인구추계 과정에서 전문가의 인구학적 지식을 고려하는 특성을 갖는데, 기본적으로 아직 발생하지 않은 사건의 발생 개연성에 관한 전문

7) 전문가에 의한 판단이나 시계열 모형은 점 예측치만을 산출하는 전통적인 시나리오 접근법에서도 사용되어질 수 있다. 본 절에서 검토하는 전문가 판단법이나 시계열 모형은 점 예측치뿐만 아니라 구간 예측치를 산출하는 확률적 접근법을 의미함에 유의할 필요가 있다. 다시 말하면, 비록 전통적인 시나리오 접근법에서 이러한 전문가 판단이나 통계적 모형이 사용된다고 하더라도, 확률적 접근법과 달리 인구 예측에 수반된 불확실성에 대한 양화 작업은 이루어지지 않는다.

8) 예측구간의 도출은 이러한 시뮬레이션 접근(simulation approach) 외에도 보다 강한 가정을 요구하는 해석적 접근(analytical approach)을 통해서도 가능하다. 해석적 접근법에 기초한 사례들로는 알호스펜서(Alho and Spencer, 1985) 그리고 코헨(Cohen, 1986)의 연구가 있다.

가의 의견에 기초하여 예측구간이 갖는 확률을 계산하는 방법이다. 예를 들면, 장기적인 사망률 예측에 수반된 불확실성은 생존율에 유의미한 영향을 미칠 수 있는 과학기술 혁신에 크게 의존하며, 장기적인 인구이동과 관련된 불확실성은 세계경제에서 발생하는 구조적 변동과 밀접한 관련이 있다. 비록 이러한 사건들이 실제로 발생할 개연성은 상대적으로 높지 않다고 하더라도 이에 대한 고려는 인구추계에 수반된 불확실성의 적절한 평가에 있어서 매우 중요한 역할을 할 수 있다(De Beer and Alders, 1999).

만일 미래 예측이 단순히 과거 추세의 연장에 기초하고 있다면 과거 자료에 대한 분석으로부터 과거 추세와 상반된 구조적 변화가 발생할 여지는 없는 것이다(De Beer, 2000). 이러한 측면에서 전문가 판단법을 지지하는 연구자들은 통계적 모형과 같은 방법에 비해 전문가의 판단에 기초한 예측이 인구추계 작업에 수반된 고도의 불확실성을 평가할 수 있는 보다 적절한 방법임을 지적하고 있다.

전문가 판단법을 지지하는 연구자들이 사용하는 「확률」의 의미는 주관적(subjective) 확률을 의미하며, 완벽한 주사위를 던지는 것과 같은 맥락에서 사용되는 이론적 확률의 개념과는 구분된다. 전문가 판단법을 지지하는 연구자들은 미래의 인구학적 변화와 관련된 확률분포를 완벽하게 명시할 수 있는 객관적인 방법은 없기에 인구추계를 담당하는 연구자들이 할 수 있는 최선의 방안은 현 시점에서 이용 가능한 모든 정보를 충분히 활용하여 숙고된 판단(informed judgment)을 내리는 것임을 강조한다(Lutz, Sanderson and Scherbov, 2004).

2) 내용

전문가 판단법은 일정한 기간에 걸친 출산력, 사망률, 인구이동에 관해 미리 설정된(예를 들면, 80%) 예측구간을 전문가 패널(resource experts)에게 구한 후 추계기간 동안 이들 인구변동 지표들의 궤적을 도출하는 방식을 취한다. 예를 들면, 델파이 방법 등을 통하여 미리 설정된 예측구간(예컨대, 80%)의 상한과 하한에 해당하는 인구변동 요약 지표들의 값을 전문가 집단에게 질의할 수 있다. 이와 함께 전문가 집단은 또한 그들의 의견에 대한 근거를 제시할 것이 요구되어진다. 이러한 전문가 패널의 의견은 인구추계를 위한 최종 가정의 설정을 위해 또 다른 전문가 집단(implementation/meta experts)에 의해 재검토되는 과정을 거친다(Howe and Jackson, 2006; Lutz et al., 1998).

전문가 판단법의 경우 예측구간이 갖는 확률은 전문가의 주관적 판단에 기초하여 도출되기에 발생 가능한 편의(bias)의 원천을 고려하는 것이 매우 중요하

다. 특히, 전문가들이 범할 수 있는 오류와 관련하여 인구변동 요인들이 일정 기간 동안 안정적인 추세를 보일 경우 전문가들 사이에 미래를 정확하게 예측할 수 있다는 낙관론이 대두하는 경향이 있음이 지적된다. 개별 전문가의 판단에 의존할 경우에 발생할 수 있는 이러한 위험을 줄이기 위해 일반적으로 특정의 전문가 대신 다수의 전문가들이 예측한 주관적 확률분포들이 결합되어 사용된다(Keilman et al., 2002; O'Neill, Balk, Brickman and Ezra, 2001).

전문가 판단법을 활용한 인구추계는 현재 IIASA(International Institute for Applied Systems Analysis)의 러츠(Lutz)와 그의 동료들에 의해서 주도되고 있다. 현재까지 전문가 판단법은 대체로 세계 및 권역(국가군) 수준의 예측에서 많이 활용된 반면 개별 국가의 경우에 적용된 경우는 제한적이다. 본 절에서는 러츠 외(Lutz et al., 2004)가 세계 인구에 대한 전망에서 전문가 판단법을 통하여 출산력을 예측한 사례를 간략히 살펴보기로 한다. 전문가 판단법이 기본적으로 전문가 집단의 주관적 판단에 기초하기에 이들의 논의에서 출산력의 미래 추세와 관련된 구체적인 모형은 제시되지 않고 있으며, 전문가 집단이 전망한 미래 출산력의 구체적 수준과 이에 대한 논리적 근거의 제시가 주된 위치를 차지하고 있다⁹⁾.

출산력 변화와 관련하여 이들은 개별 국가들을 기준연도(2000년)의 출산력을 기준으로 하여 13개의 권역으로 분류하고 있는데 중국의 경우 별도의 권역으로 검토되고 있다. 권역별 미래의 출산력과 관련해서는 1)출산력 변천이 시작되는 시점, 2)출산력 감소의 속도, 그리고 3)출산력 변천이 완료된 시점에서의 출산력 수준이 핵심적인 논의 사항으로 등장하고 있다. 현재 출산율이 높은 국가들의 경우 인구변천이론(demographic transition theory)에 근거하여 일단 개별 국가들이 출산율 감소를 경험하게 되면 출산율은 지속적으로 대체출산율 수준 혹은 그 이하로 떨어지게 될 것을 전망하고 있다. 출산력 감소의 속도에 관한 논의는 UN(2003)의 전망과 매우 유사하다.

출산력 변천 후의 출산력 수준과 관련하여 13개 권역별 평균 합계출산율(5년 단위)을 기준연도(2000년), 2025-2029년, 2080-2084년의 기간으로 구분하여 제시하고 있으며, 이들 기간 사이의 합계출산율은 내삽법을 통하여 도출하고 있다. 7개의 고위 출산력 권역의 권역별 평균 합계출산율은 2025-2029년 기간의 경우 2.1-3.0 그리고 2080-2084년 기간의 경우 1.5-2.0으로 전망하

9) 결과적으로 전문가 판단법의 경우, 비록 미래의 출산력 수준과 관련된 기본적인 논거는 제시되고 있지만, 전문가들 사이의 의견이 구체적으로 어떻게 조정되는가에 대해서는 보고되지 않는 경향이 있다.

는 동시에 구체적인 출산력 수준은 또한 개별 권역의 인구밀도에 의해서 조정되는 것으로 설정되어 있다. 5개의 저위 출산력 권역의 평균 합계출산율은 2025-2029년 기간의 경우 1.7-1.9 그리고 2080-2084년 기간의 경우 1.6-2.0으로 전망되고 있다. 중국의 합계출산율은 2025-2029년 기간의 경우 1.8 그리고 2080-2084년 기간의 경우 1.7로 설정되어 있다. 마지막으로 이러한 권역별 평균 합계출산율을 중심으로 하여 전문가 집단이 설정한 80%의 예측구간이 13개 권역별로 그리고 시기별로 제시되고 있다.

3) 장단점

전문가 판단법을 지지하는 연구자들은 다른 방법들에 비해 전문가 판단법이 미래에 발생할 수 있는 구조적 변화와 예측하지 못한 사건들을 보다 적절하게 예측할 수 있음을 강조한다(Booth, 2006; De Beer, 2000). 또한 전문가 판단법은 시계열 모형에 비교할 때 상대적으로 자료에 대한 요구사항이 크지 않기에 과거의 추세 자료가 부족한 지역의 경우 확률적 예측치를 생산할 수 있는 유용한 대안임이 지적되고 있다(O'Neill et al., 2001).

그러나 전문가 판단법의 경우 이론적으로 그리고 현실적으로 당면하게 되는 많은 어려움이 존재한다. 첫째, 전문가 판단법에서 중추적인 역할을 수행하는 전문가와 관련하여 어떠한 개인들로 전문가 집단을 구성하여야 하는가와 관련된 문제가 제기된다. 전문가 집단을 구성한 경우에도 미래 추세를 예측하는 과정에서 전문가들이 보일 수 있는(예측치 설정에 있어서의) 보수성을 어떻게 극복할 것인가의 문제가 뒤따른다.

20세기 동안 전 세계적으로 경험한 급격한 사망력 감소와 관련하여 전문가들의 판단이 과도하게 보수적이었음은 주지의 사실이며(사망력 오차의 구체적 분석에 대해서는 알호·스펜서(Alho and Spencer, 1990)를 참조), 출산력 예측의 경우에도 동일한 평가가 내려질 수 있다(이에 대해서는 리(Lee, 1998)의 논의를 참조). 이러한 사례들은 전문가들이 종종 어떤 주어진 구간에 과도하게 높은 수준의 확률을 부여하는 경향이 있음을 시사하고 있다(Keilman et al., 2002).

또한 전문가들의 경우 일반적으로 공통의 지식에 기초하고 있기에 어떤 특정의 전문가가 해당 분야의 다른 전문가들과 매우 상이한 의견을 제시할 개연성은 매우 낮다. 동시에 어떤 특정의 전문가가 전문가 집단 전반의 의견과 매우 상이한 의견을 제시할 경우 해당 전문가의 의견은 전문가 집단 전체에 의해 신뢰되어질 수 없는 의견으로 규정될 개연성이 크다는 점 또한 지적될 필요가 있다(Alho, 2005).

둘째, 리(Lee, 1998)는 인구변동 요인들의 추세와 관련된 추정치들의 예측구간과 관련하여 전문가들이 신뢰수준 사이의 차이를 구분할 수 있는가에 관하여 의문을 제기한다. 예컨대, 90% 신뢰수준과 95% 신뢰수준을 전문가 집단이 적절하게 판별해낼 수 있는가와 관련된 의문이 있을 수 있다. 그는 또한 전문가의 경우도 인구변동 요인들에서 발생할 수 있는 구조적 변화의 개연성을 배제함으로써 불확실성을 저평가하는 측면이 있음을 지적한다. 또한 미래에 발생할 수 있는 구조적 변화를 예측하는 것은 전문가 집단만의 속성은 아님을 지적한다. 예컨대, 과거의 인구추계 오차에 기초한 확률적 모형의 경우도 간접적으로 과거에 발생한 구조적 변화(예를 들면, 세계대전, 베이비 붐, 현대적 피임기술의 전파, 항생제의 출현 등)에 관한 정보를 포함하고 있다는 것이다.

마지막으로 전문가 집단의 판단에 기초한 자료의 경우 전문가의 판단과 실제로 관측되는 추세 사이에 체계적인 간극이 발생할 수 있는 문제가 있다(Booth, 2006). 인구변동 요인들이 어떠한 변화를 보일 경우 전문가는 이러한 변화가 일시적 변동인지 아니면 새로운 추세 형성의 시작인지에 대해서 판단을 내려야 한다. 전문가가 새로운 추세의 시작을 간파하지 못할 경우 이러한 오류는 새로운 추세가 확연하게 나타나는 시점에서야 비로소 교정될 수 있다(assumption drag). 반면 전문가가 일시적 변동을 새로운 추세의 시작으로 판단할 경우 전문가의 예측은 과대반응(overreaction)의 오류를 범할 것이다. 일반적으로 전문가 집단이 인구변동 요인들의 추세를 예측할 때 상대적으로 최근의 추세를 보다 강조하는 경향이 있음을 고려할 때 첫 번째 오류를 범할 개연성이 상대적으로 높은 것으로 지적된다(Keilman, 1997).

3. 통계적 모형(Statistical Method)

1) 개요

인구추계에 수반되는 불확실성은 또한 일련의 가정에 기초한 통계적 모형을 통해서도 추정될 수 있다. 비록 통계적 모형에 기초하여 예측구간을 설정하는 다양한 방법들이 제안되었지만, 현재까지 시계열 모형(time series model)을 통한 예측구간의 설정이 가장 빈번하게 활용되는 방법이라고 할 수 있다(예컨대, Alho and Spencer, 1990; Lee and Carter, 1992; De Beer, 1997; Hyndman and Ullah, 2007). 시계열 모형은 기본적으로 예측 대상 변수의 과거 값들이 (연구자에 의해) 설정된 시계열 모형에 의해서 생성되었으며, 미래에 실현될 값들도 이러한 모형에 기초하여 발생됨을 가정하고 있다.

인구추계에서 시계열 모형은 장기 예측의 경우에는 적합하지 않다는 의견도 있지만, 최근 장기 예측에서도 시계열 모형의 사용이 증가하고 있다. 비록 시계열 모형을 통하여 장래인구의 규모와 성장률을 직접적으로 예측하는 연구도 있지만(예컨대, Cohen, 1986), 시계열 모형의 보다 일반적인 활용은 개별 인구변동 요인들에 대한 추세 전망이라고 할 수 있다.¹⁰⁾ 현재 시계열 모형 중에서 가장 빈번하게 활용되는 모형 중의 하나는 ARIMA(Autoregressive Integrated Moving Average) 유형의 모형인데, 사망력 예측과 관련된 Lee-Carter 방법과 이를 확장한 방법들이 대표적인 예라고 할 수 있다(예컨대, Booth, Maindonald, and Smith, 2002; Lee, 2000; Lee and Carter, 1992; Lee and Miller, 2000).

2) 내용

인구변동 요인들을 확률적으로 예측하는 다양한 통계적 모형들이 제시되고 있지만, 본 연구에서는 현재 가장 광범위하게 활용 및 확장되고 있는 Lee-Carter 모형(이하 LC 모형)을 간략히 살펴보기로 한다(Lee and Carter, 1992). 연령별 사망력 예측과 관련하여 1992년에 발표된 LC 모형은 인구추계 분야에서 이루어진 가장 중요한 진척 중의 하나로 평가되고 있는데, 이해하기 쉬운 동시에 경험적으로도 정확하게 사망력 변화를 기술하는 것으로 평가되고 있다. 또한 LC 모형이 예측의 불확실성을 모형화할 수 있는 장점이 있는 동시에 다른 모형들에 비해 보다 정확한 예측치를 도출한다는 연구 결과도 있다(예컨대, Bell, 1997).

LC 모형의 독특한 특징은 시간의 경과에 따른 사망력의 변화를 단일의 (연도별) 지표(k)로 축소시켰으며, 이러한 지표를 모형화하고 예측하기 위하여 시계열 모형을 사용하고 있다는 점이다. LC 모형은 행렬의 차원에서 다음과 같이 표시될 수 있다(Lee and Carter, 1992: 661).

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}.$$

여기에서 벡터 $m_{x,t}$ 는 t 시점에서 x 연령의 사망률, a 는 연령별 사망률의 일반적 패턴을 기술하는 연령별 상수항의 집합, k 는 시간에 따른 사망력 수준을 나타내는 지표, b 는 각각의 연령에서 사망력의 상대적 변화 속도를 나타내는 연령별 상수항의 집합을 각각 나타낸다.

10) 확률적 시계열 모형을 사용하여 총인구나 인구 성장률을 「직접」 예측하는 방법은 총인구의 연령별 구성에 관한 정보를 제공하지 못하는 단점이 있다.

LC 모형은 방정식의 오른쪽에 예측변수가 없으며, 그 대신 추정해야 할 모수 그리고 알려지지 않은 지표(k)만이 존재하기에 통상의 회귀모형을 통해 추정할 수는 없다. 리(Lee)와 카터(Carter)는 미국의 1933년에서 1987년까지의 사망률 자료를 가지고 SVD(Singular Value Decomposition) 방법을 사용하여 a , b , 그리고 k 를 추정하였다. 이들은 또한 사망률의 로그값을 추정하는 데 있어서 발생한 연령 집단별 가중치의 왜곡을 조정하기 위하여 a , b , k 가 추정된 다음 k 값을 재추정하는 절차를 거쳤는데, 이러한 재추정 절차를 통해서 k 는 연도별로 총 사망자 수를 정확히 예측하게 된다.

위의 방정식에서 나타나듯이 a 와 b 는 시간(t)의 경과에 따라 값이 변하지 않는 것으로 설정되어 있기에 미래의 추세를 예측하는 경우 단지 k 만이 관심의 대상이 된다. 시간 성분(time component) k 는 로그 사망률에 있어서 시간의 경과에 따른 모든 연령대의 전반적인 사망률 추세를 포착하며, 연령 성분(age component) b 는 시간의 경과에 따른 주된 추세를 수정하는 역할을 하는데, 이로 인해 어떤 특정 연령대에서의 사망률 변화가 주된 추세보다 빠르거나 혹은 느린 결과로 나타나게 된다.

LC 모형은 생성된 일련의 k 값들을 기초로 여러 가능한 모형들의 적합도를 비교 평가하여 가장 적합성 있는 시계열 모형을 선정한 다음 과거 자료를 사용하여 모형의 모수들을 추정하는 절차를 거친다. 대체로 LC 모형의 경우 ARIMA 시계열 모형을 사용하여 추세를 예측한다. 이 때 ARIMA 시계열 모형을 통해 산출된 표준오차가 예측과 관련된 불확실성을 반영하며, 이를 기초로 k 에 관한 (그리고 사망률 및 기대여명에 관한) 확률적 예측구간을 구성하게 된다. LC 방법의 경우 기본적으로 과거의 자료를 기초로 미래를 예측하기에 초고령기의 경우(예컨대, 95세 이상) 과거 자료의 제약(혹은 부재)으로 인해 그 활용이 제한적일 수 있다(LC 모형의 문제점과 관련해서는 리(Lee, 2000)의 논의를 참조).

LC 모형은 초기에 사망률 예측을 위해 주로 사용되었으나, 이후 출산력 예측에도 활용됨으로써 그 적용 범위를 넓혀가는 경향이 있다. 또한 초기의 LC 모형에서 설정되었던 일련의 가정을 완화하는 방향으로 LC 모형의 정교화가 이루어지고 있다. 예컨대, 부스 외(Booth et al., 2002)는 연령 성분이 시불변(time-invariance)이며 시간 성분이 선형(linear)이었던 초기의 LC 모형을 확장하여 (총 사망 수 대신) 연령별 사망 분포와 일치하도록 k 를 조정하였으며, 비선형성(non-linearity)의 문제와 관련해서는 선형적 모형이 적용될 수 있는 최적의 예측구간을 결정할 수 있는 방법을 제시하고 있다. 또한 리·리·툴자파카(Li, Lee, and Tuljapurkar, 2004)는 우리나라처럼 사망률 예측을 위한 충분한 자료가

구축되지 못한 국가들의 경우에 LC 모형을 적용하는 방법을 논의하고 있다¹¹⁾.

3) 장단점

시계열 모형과 같은 통계적 모형에 기반을 두고 있는 확률적 예측 방법은 잘 확립된 기존의 통계적 이론과 프로그램을 사용할 수 있는 동시에 분석 결과의 타당성이 다른 연구자들에 의해서도 검증될 수 있다는 장점을 지니고 있다. 또한 미리 설정된 모형 선정의 기준을 기초로 하여 가능한 여러 대안들을 경쟁적으로 상호 비교할 수 있는 장점이 있다. 시계열 모형은 또한 전적으로 전문가의 판단에 기초하고 있는 방법에 비해 확률적으로 내적 일관성을 유지한 상태에서 과거에 대한 경험적 분석을 통하여 예측분포를 도출할 수 있는 장점이 있다. 비록 과거 시계열 자료에 대한 통계적 분석 또한 전문가의 판단이 요구되는 측면이 있지만(예컨대, 적합한 모형의 선택), 주관적인 측면과 객관적인 측면의 조합에 있어서 그 무게중심은 객관적 방법론 쪽에 있다고 볼 수 있다(Lee, 1998; Lee and Tuljapurkar, 1994). 또한 시계열 모형은 미래의 예측에서 발생하는 오차가 과거의 예측 오차와 유사할 것이라는 가정을 요구하지 않는다(De Beer and Alders, 1999).

그러나 시계열 모형은 또한 여러 가지 단점을 지니고 있는데, 우선, 시계열 모형의 사용을 위해 요구되는 자료를 보유하지 못한 경우에 그 활용은 제한적이다. 또한 시계열 모형은 대상 변수의 과거 값들이 설정된 모형에 의해서 발생하였으며, 미래에 실현될 값들도 이러한 모형에 기초하여 발생함을 가정하고 있다. 그러나 이러한 가정이 미래에도, 특히 장기 예측의 경우에도, 성립할 것인지는 불확실하다(National Research Council, 2000).

또한 시계열 예측 모형은 기준연도에서 발생하는 오차, 모형의 명시화 과정에서 발생하는 오차, 모수 추정에서 발생하는 오차, 모형에서 설정된 통계적 관계와는 상이한 구조적 변화로 인해 발생하는 오차 등을 고려할 필요가 있는데(Lee, 1998), 결과적으로 여러 다른 시계열 예측 모형들이 선택지로 고려될 수 있으며, 선택된 모형에 따라 각기 상이한 예측구간으로 이어질 수 있다. 현재까지 매우 제한된 연구들만이 「시계열 모형을 통해 도출된 예측구간이 불확실성을 적절히 측정하고 있는가」의 문제를 검토하였음이 지적될 필요가 있다(예를

11) LC모형의 다양한 확장에 관해서는 리(Lee, 2000)와 부스·힌드만·티클·드중(Booth, Hyndman, Tickle, and De Jong, 2006)의 논의를 참조할 수 있다. 비록 확률적 접근의 일환으로 사용된 것은 아니지만 LC모형을 사용한 국내 연구들로는 김종면·김우철(2007), 김태현(2006), 정우진(2005)의 연구가 있다.

들면, Cohen, 1986; Lee, 1974; Smith and Tayman, 2004).

마지막으로, 최근 장기 예측의 경우에도 시계열 모형을 사용한 연구가 점차 증가하는 경향이 있지만(Booth, 2006), 일반적으로 시계열 모형은 장기 예측보다는 단기 예측을 위해서 사용되는 것이 보다 적절한 방법임이 지적되고 있다(De Beer and Alders, 1999; Keilman, 2005; Lee, 1998). 장기 예측에서 시계열 모형을 활용할 경우 추계기간이 길어짐에 따라 예측구간이 과도하게 넓어지는 문제가 발생할 수 있는데 이로 인해 확률적 예측의 효용성이 떨어지는 문제가 발생할 수 있다는 것이다.

4. 과거 예측 오차 분석법(Historical Forecast Error Method)

1) 개요

과거의 예측 오차를 사용하는 방법은 과거의 인구추계에서 나타난 예측 오차들의 분포를 사용하여 미래 예측의 불확실성을 양화하는 방식이다. 과거의 인구추계에서 발생한 오차를 활용하는 방법은 이미 1980년대 초에 총인구 및 인구 성장률 예측의 정확성과 관련하여 키피츠(Keyfitz, 1981)와 스토토(Stoto, 1983)에 의해서 소개되었으며, 최근 NRC(National Research Council, 2000)에 의해 세계 인구와 지역 및 개별 국가들의 인구추계에 활용되었다. 기존 연구들은 또한 총인구나 인구 성장률에 추가하여 합계출산율, 기대수명, 인구가동 등의 예측과 관련해서도 과거 예측 오차 분석법을 사용하였다(예컨대, De Beer, 1997; Keilman, 1997).

시계열 모형과 전문가 판단법이 기대치를 중심으로 한 모수의 분포를 구성함에 비해 과거 오차 분석법은 일반적으로(기대오차가 영(zero)임을 가정하여) 영(zero)을 중심으로 한 분포를 명시한다. 과거 오차 분석법에 있어서 모수의 기댓값이 내생적으로 도출되는 것이 아니라 결정론적 방식으로 산출된 외생의 예측치를 그대로 사용한다는 점에서 시계열 모형이나 전문가 판단법과는 차이를 보인다(Keilman, 2005). 예컨대, 키피츠(Keyfitz, 1981)는 미국 센서스국의 인구추계 그리고 NRC(National Research Council, 2000)는 세계 인구에 관한 UN의 인구추계에서 도출된 기댓값(중위 시리즈)을 중심(기댓값)으로 하여 예측구간을 구성하고 있다.

2) 내용

예측 오차 분석법은 일반적으로 과거의 인구추계에서 발생한 예측 오차가 근

사적으로 정규분포를 이룬다는 가정과 함께 미래에 나타나는 예측 오차의 분포 또한 과거와 동일하다는 가정에 기초하여(새롭게 이루어진) 어떤 인구추계에서 특정의 예측구간이 갖는 확률을 계산하는 데 사용되어진다. 앞에서도 언급했듯이 불확실성을 추정하는 데 있어서 전문가 판단법이나 시계열 모형에 비해 과거 오차 활용법은 미래의 기댓값을 산출하는 대신 별도로 생성된 어떤 예측구간에 그에 상응하는 확률을 부착하는 방식을 취한다(Alders and De Beer, 2004; De Beer and Alders, 1999).

과거 인구추계에서 발생한 오차에 관한 정보를 활용하여 새로운 인구추계가 갖는 불확실성을 양화하기 위해서는 예측 오차를 발생시키는 원인을 파악하는 것이 매우 중요하다(Alders and De Beer, 2005). 과거의 예측 오차에 대한 분석이 미래의 새로운 인구추계에 수반되는 불확실성과 관련하여 얼마나 유용한 정보를 제공하는가는 과거에 발생한 오차와 미래에 발생할 오차 사이의 유사성에 의존한다고 할 수 있다. 과거의 인구추계에서 발생한 오차들의 분포가 시간의 경과에 따라 안정적임을 전제로 과거의 오차 분포를 사용하여 미래의 어떤 예측구간이 갖는 확률을 계산할 수 있는 것이다(Stoto, 1983).

과거의 인구추계에서 발생한 오차에 기초하여 현 시점에서 새롭게 추정된 인구추계의 불확실성을 기술하는 다양한 절차들이 사용되고 있는데, 상대적으로 간단한 적용은 과거에 발생한 예측 오차를 경험적으로 분석하여 도출한 표준오차를 사용하여(결정론적 방식으로 얻어진) 현 인구추계의 중위 시리즈를 중심으로 예측구간을 구성하는 것이다¹²⁾. 그러나 과거의 예측 오차에 대한 「경험적 분석」에 기초한 방법은 분석에 사용될 예측 오차에 관한 과거 자료가 부재하거나 매우 제한적으로만 존재하기에 예측구간을 설정하는 방법으로 제약이 심하다. 이러한 측면에서 장기 예측과 관련된 예측 구간의 설정은 예측 오차에 관한 「통계적 모형」에 기초하는 경우가 보다 일반적이다(De Beer, 1997).

통계적 모형에 기초한 과거 예측 오차 분석에 대한 하나의 사례로 드비어(De Beer, 1997)는 네덜란드의 인구이동 예측에서 발생한 오차를 통계적(시계열)

12) 이러한 방법은 키피츠(Keyfitz, 1981)의 경우처럼 과거 오차 분석과 관련하여 초기부터 제안된 방법이다. 키피츠(Keyfitz, 1981)에 의해서 소개된 예를 살펴보면, 그는 인구 성장률이 낮은 국가들의 경우 연 인구 성장률 추정 과정에서 발생한 예측 오차가 0.29임을 얻었는데, 이를 이용하여 미국 센서스국이 1980년에 예측한 2000년 인구 예측치의 예측구간을 구성하고 있다. 미국 센서스국은 미국 인구가 1980년 222백만 명에서 2000년 260백만 명(저위 및 고위 예측치는 각각 246백만 명과 280백만 명)으로 증가할 것을 예측하였는데, 이는 연 인구 성장률의 차원에서는 0.79에 해당한다. 이를 기초로 연 인구 성장률에서 68%에 해당하는 예측구간을 계산하면 이는 0.79 ± 0.29 혹은 $[0.5, 1.08]$ 에 해당한다. 이를 기초로 2000년 미국 인구의 68% 예측구간을 계산하면 $[222 \times (1.0050)^{20}, 222 \times (1.0108)^{20}]$ 혹은 대략 [245백만 명, 275백만 명]의 예측구간으로 나타난다.

모형을 통해 분석함으로써 예측구간을 구성하고 있다. 확률적 예측 방법으로서 앞(제3절)에서 언급한 시계열 방법이 과거 인구변동 요인들의 실제치(예컨대, 순인구이동)를 대상으로 한 모형임에 비해 본 절에서 언급하고 있는 방법은 과거의 예측 오차(예컨대, 순인구이동 예측에서 발생한 오차) 자료에 기초하여 예측 구간을 구성하고 있다는 점에서 차이가 있다.

드비어(De Beer, 1997)가 과거 예측 오차를 활용하여 네덜란드의 순인구이동을 확률적으로 예측하는 데 사용한 통계적 모형은 자기회귀(autoregressive) 모형이다. Z_{t+i} 를 t 시점에서 i 년 후의 예측 오차 그리고 e_{t+i} 를 평균이 0, 분산이 σ^2 인 백색잡음계열(white noise sequence)의 오차항이라고 할 때, 1975년 이후 네덜란드에서 이루어진 순인구이동 예측에서 발생한 오차를 AR(1) 모형을 통하여 추정된 결과는 $Z_{t+i} = .89Z_{t+i-1} + e_{t+i}$ 로 나타나고 있다(자기회귀 계수의 표준오차는 0.07 그리고 오차항의 표준오차는 14,000). 이 때 자기회귀 계수를 ϕ 라고 할 때 j 년 후 예측 오차의 표준오차는 다음과 같다.

$$\sigma \sqrt{\sum_{i=1}^j \phi^{2i-2}}$$

이를 통하여 5년과 15년 후의 68% 예측 구간을 각각 계산해 보면 이는 대략 $\pm 25,000$ 과 $\pm 30,000$ 수준(천단위 반올림)으로 10년의 기간 동안 예측구간이 대략 20% 정도 넓어진 것을 살펴볼 수 있다.

3) 장단점

과거의 예측 오차를 사용하는 분석은 과거에 관측된 오차의 본질적 측면과 일관된 미래 예측치의 확률분포를 구성할 수 있는 장점이 있다. 또한 과거 오차에 대한 경험적 분석의 경우 시계열 모형에서 사용하는 것과 같은 일련의 가정에 의존함이 없이 경험적 예측 오차를 계산할 수 있다. 그러나 과거의 예측 오차를 사용하는 방법은, 첫째, 많은 경우에 있어서 분석에 사용될 수 있는 과거 예측 오차 정보가 상대적으로 제한적이라는 문제를 갖고 있다(Keilman, 2005). 또한 과거에 이루어진 예측 오차 정보가 존재하는 경우에도 해당 정보가 대체적으로 단기간이라는 문제를 갖고 있다.

둘째, 과거의 예측 오차를 사용하는 분석은 또한 예측 방법의 향상 등에 기인하여 시간의 경과에 따라 예측 오차가 감소했을 개연성이 있음이 지적되고 있다(Alders and De Beer, 2004; Cohen, 1986; Keilman, 2005; Keilman et al.,

2001, 2002). 만일 시간의 경과에 따라 예측 방법의 향상 등에 기인하여 예측 오차가 감소하였다면 과거 오차를 통하여 미래 예측치가 갖는 불확실성을 평가하는 작업은 상대적으로 오래 전에 수행된 예측 작업의 영향을 강하게 반영함으로써 편의된 결과를 산출할 수 있다. 이는 기본적으로 최근에 수행된 예측 작업의 경우 정확성과 관련된 오직 단기간의 정보만이 활용 가능한 반면 오래된 예측 작업의 경우 상대적으로 많은 정보가 활용 가능하기 때문이다(De Beer and Alders, 1999)¹³⁾.

마지막으로 과거의 예측 오차를 활용하는 경우 과거의 인구추계와 새로운 인구추계가 서로 다른 기간(period)을 준거로 설정하고 있음이 지적될 필요가 있다(Alders and De Beer, 2005). 이러한 이유 때문에 케일만(Keilman, 1997)은 어떤 특정 기간(예컨대, 인구변동 요인들이 급격하게 변동하는 기간)이 다른 기간에 비해 예측하기 어려움을 고려할 때 기간효과(period effect)를 명시적으로 모형에 고려할 필요가 있음을 지적한다. 그러나 드비어·앨더스(De Beer and Alders, 1999)는 인구변동 요인들에 있어서 상대적으로 변화가 크지 않는 기간이 반드시 인구추계의 용이성과 연결되는 것은 아님을 지적한다. 예를 들면, 출산력 변천을 완료한 국가에 비해 출산력 변천을 시작한 국가의 경우 출산력 예측이 보다 용이할 수 있다.

5. 확률적 인구추계의 적용 현황

비록 제한적인 리뷰이지만, 현재까지 확률적 인구추계 방법을 사용하여 국가 단위의 인구추계를 수행한 사례들로는 호주(Wilson and Bell, 2004), 오스트리아(Lutz and Scherbov, 1998), 중국(Li, Reuser, Kraus, and Alho, 2007), 핀란드(Alho, 2002, 2003), 네덜란드(Alders and De Beer, 2004, 2005; De Beer and Alders, 1999; Keilman, 1997), 노르웨이(Keilman, 1997; Keilman and Pham, 2000; Keilman et al., 2001, 2002), 폴란드(Matysiak and Nowok, 2007), 미국(Lee and Tuljapurkar, 1994) 등이 있다. 확률적 인구추계는 또한 개별 국가를 넘어서 세계 인구(Lutz et al., 2004) 혹은 국가군(Alho, Alders, Crujisen,

13) 비록 기준인구와 관련된 자료의 경우 일반적으로 시간의 경과에 따른 질적 향상이 기대되지만, 최근까지 보고되고 있는 과거 인구추계의 정확성에 대한 평가 결과들은 과거의 인구추계에 비해서 최근의 인구추계가 보다 정확해졌다는 일관된 결과를 제시하지 않는다(예컨대, 우혜봉, 2009; Keilman, 1997). 스미스·신시치(Smith and Sincich, 1988)의 분석 결과도 (절대치로 계산된) 과거 예측 오차의 분포가 시간의 경과에 따라 상대적으로 안정적이기에 미래의 오차를 예측하는 데 있어서 유용한 정보를 제공할 수 있음을 지적한다.

Keilman, Nikander, and Pham, 2006; Keilman and Pham, 2004; Scherbov and Mamolo, 2006)을 대상으로 하여 수행되기도 하였다.

비록 확률적 인구추계가 일반 연구자들에 의해 최근 활발하게 검토되고 있는 상황이지만, 현재까지 국가의 공식적인 인구추계는 대부분 결정론적인 방식으로 이루어지고 있다. 국가의 공식적인 인구추계에서 확률적 인구추계를 실시한 경우로는 네덜란드(Statistics Netherlands)의 사례가 있다¹⁴⁾. 그러나 최근 공식적 인구추계의 경우에도 확률적 방법의 사용을 검토하는 경우가 증가하고 있음이 지적될 필요가 있다(예컨대, 미국 센서스국의 입장과 관련해서는 롱·홀만(Long and Hollmann, 2004)을 그리고 스웨덴 Statistics Sweden의 확률적 인구추계 논의와 관련해서는 하트만·스트란델(Hartmann and Strandell, 2006)을 참조).

IV. 논의 및 맺음말

비록 제한적이지만 본 연구는 1980년대 초에 소개된 이후 1990년대에 들어 활용도가 높아지기 시작한 확률적 인구추계의 주요 내용과 방법들을 살펴보았다. 확률적 인구추계의 경우 현재 발전 과정에 있기 때문에 여러 가지 한계점과 추가적인 연구가 필요한 영역이라고 할 수 있다. 한계점과 관련하여, 우선, 확률적 접근을 통해 도출된 예측구간이 과연 미래의 인구변동 요인들이 보이는 불확실성을 적절히 측정하는 지표인가에 대해서 의문이 제기되기도 한다(Smith and Tayman, 2004). 이와 관련하여 서로 다른 예측 모형들을 비교한 기존 연구들은 사용된 모형에 따라 매우 상이한 수준의 예측구간으로 이어질 수 있음을 지적하고 있다(Cohen, 1986; Lee, 1974).

또한 확률적 인구추계를 위해서는 장기간에 걸친 질 높은 연도별 실제치 자료 그리고 이러한 실제 자료와의 비교를 통해 예측 오차의 경험적 분포를 추정할 수 있는 장기간의 추계 자료가 필요하다. 그러나 현실적으로 이러한 요건을 갖춘 경우는 드물기에 충분한 자료를 구축하지 못한 국가들의 경우에 확률적 인구추계를 어떻게 적용할 것인가는 중요한 연구 과제로 남아 있다(Keilman, 2005).

14) 네덜란드의 확률적 인구추계는 Statistics Netherlands의 엘더스(Alders)와 드비어(De Beer)같은 연구자들에 의해서 주도되고 있다. 네덜란드의 확률적 인구추계의 경우 불확실성에 대한 평가는 과거 예측 오차 분석, 통계적 모형, 전문가 판단법을 병용하는 방식을 취한다. 출산력에 관한 확률적 접근의 구체적인 사례로는 엘더스·드비어(Alders and De Beer, 2004)를 참조할 수 있다. 캐나다의 공식적 인구추계가 확률적 접근법에 기초한다는 지적이 있지만 현재 Statistics Canada의 공식적인 인구추계 방법은 코호트-요인법이다. 캐나다에 대한 보다 자세한 내용에 대해서는 George(2001)와 Statistics Canada(2005)를 참조할 수 있다.

현실적으로 활용 가능한 자료의 제약을 고려할 때 확률적 인구추계의 발전 방향으로 다음과 같은 사항들이 고려되고 있다. 첫째, 다양한 확률적 예측 방법들을 병용하는 방법이다. 앞서서도 지적했듯이 확률적 인구추계의 경우 어떠한 방법의 사용이 다른 방법의 사용을 배제할 필요는 없다. 예를 들면, 시계열 모형이나 과거의 예측 오차 분석법은 단기 혹은 중기 예측에 보다 적절한 방법일 수 있으며, 장기 예측과 관련해서는, 비록 전문가들에 의해 제시되는 예측구간이 과도하게 좁아지는 현상을 어떻게 극복할 것인가의 문제는 있지만, 전문가의 판단을 활용하는 방법을 고려할 수 있을 것이다(De Beer 2000; De Beer and Alders, 1999; Keilman, 2005).

둘째, 확률적 인구추계를 위해 필요한 과거의 인구추계 자료가 부재할 경우 케일만(Keilman, 2005)은 단순한(naive) 모형을 통해서 예측 자료를 산출한 후 이를 과거의 예측 자료로 대체하는 방법을 제안하고 있다. 이는 인구변동 요인들에 관한 과거의 예측 방법이 대체로 단순한 가정(예컨대, 기대여명이 연도별로 일정한 크기로 증가한다는 가정)에 기초하고 있음을 고려할 때, 이러한 접근이 실제의 예측 자료에 대한 유용한 대안이 될 수 있다는 것이다.

확률적 인구추계의 경우 현재 많은 개별 연구자들의 관심을 받고 있으며, 또한 실제의 인구추계에서도 활용도를 넓혀가고 있다. 그러나 국가적 차원의 공식적 인구추계에서 확률적 방법의 사용은 보다 큰 어려움을 부과하는 측면이 있다. 우선, 개별 연구자들은 연구자 개인이 관심을 갖는 변수에 초점을 맞출 수 있지만, 국가에 의한 공식적 인구추계의 경우 보편적이며 포괄적인 인구추계 결과를 제공할 것이 요구되기에 시간 혹은 비용의 측면에서 커다란 제약에 직면할 수 있다.

또한 국가의 공식적 인구추계의 경우 사용자들의 필요도(욕구)를 충족하여야 하기 때문에 사용자들이 쉽게 이해할 수 있는 방식으로 결과물을 제시하여야 한다는 추가적 제약이 따른다고 할 수 있다. 특히, 확률적 인구추계에 비해 기존의 결정론적 인구추계가 자료의 이용자들이 활용하기에 편리한 특성들을 갖고 있기에 확률적 인구추계의 도입을 검토할 경우 이용자의 자료 활용에 관한 특별한 고려가 필요할 것이다.

전반적으로 인구추계 분야에서 이루어진 최근까지의 발전 상황은 인구추계에 수반되는 불확실성에 대한 인식과 이에 대한 적절한 방법론적 대처가 핵심적인 이슈임을 보여 주고 있다. 그럼에도 불구하고 현재까지 인구추계 자료의 이용자들은 불확실성의 문제보다는 미래에 실현될 개연성이 가장 높은 단일의 예측치(중위 시리즈)를 선호하는 경향이 강하다. 그러나 인구추계에 수반된 불확실

성을 고려하기 위해서는 인구추계 자료를 생산하는 기관에 의해서 제공되는 모든 시리즈를 추가적으로 살펴보는 것이 바람직하다. 마찬가지로 시나리오 접근을 통하여 인구추계 자료를 생산하는 연구자들도 인구변동 요인들과 관련된 가정을 보다 면밀히 검토하여 인구추계에 수반되는 불확실성을 적절히 반영할 수 있도록 가정을 설정할 필요가 있으며, 불확실성의 측정과 관련된 시나리오 접근의 한계를 자료의 이용자들에게 명확히 제시할 필요가 있다.

그러나 전통적인 시나리오 접근에 의해서 제공되는 모든 추계 시리즈를 고려하더라도 불확실성을 적절히 양화하는 데는 일정한 한계가 있다. 본 연구가 검토하고 있듯이 인구추계에 수반되는 불확실성은 인구 변수에 관한 확률분포를 명시화함으로써 보다 효과적으로 처리될 수 있다. 만일 인구추계 결과가 어느 정도의 불확실성을 갖고 있는가에 관한 정보가 제공될 경우 인구추계 자료의 이용자들은 이를 통하여 보다 쉽게 계획의 유연성을 확보할 수 있을 것이다. 이러한 측면에서 현재의 시나리오 접근에 대한 보완으로 혹은 중장기적 대안으로 인구추계의 불확실성을 명시적으로 고려하는 확률적 인구추계의 활용 가능성을 검토해 볼 필요가 있다.

<참고문헌>

- 김종면·김우철 (2007) 《보건·의료부문 장기재정 모형 구축》 한국조세연구원.
- 김태현 (2006) “장래인구추계를 위한 사망률 예측” 《한국인구학》 29(2): 27-51.
- 우해봉 (2009) “우리나라 인구추계의 정확성과 시사점” 《조사연구》 10(2): 71-96.
- 정우진 (2005) “인구고령화에 따른 국민건강보험 진료비 및 정부재정 지원규모 중·장기 예측” 한국개발연구원 인구고령화와 보건·의료 세미나 발표문.
- Alders, M. and J. De Beer (2004) “Assumption on Fertility in Stochastic Population Forecasts” *International Statistical Review* 72(1): 65-79.
- _____ (2005) “An Expert Knowledge Approach to Stochastic Mortality Forecasting in the Netherlands” pp. 39-64 in *Perspectives on Mortality Forecasting: Vol II Probabilistic Models*, edited by Nico Keilman, Stockholm: Swedish National Social Insurance Board.
- Alho, J. M. (2002) “The Population of Finland in 2050 and Beyond” Discussion Paper no. 826, The Research Institute of the Finnish Economy.
- _____ (2003) “Experiences from Forecasting Mortality in Finland” pp. 29-40 in

- Perspectives on Mortality Forecasting: Vol I Current Practice*, edited by Tommy Bengtsson and Nico Keilman, Stockholm: Swedish National Social Insurance Board.
- _____ (2005) "Remarks on the Use of Probabilities in Demography and Forecasting" pp. 27-38 in *Perspectives on Mortality Forecasting: Vol II Probabilistic Models*, edited by Nico Keilman, Stockholm: Swedish National Social Insurance Board.
- Alho, J. M., Alders, M., Cruijsen, H., Keilman, N., Nikander, T. and D. Q. Pham (2006) "New Forecast: Population Decline Postponed in Europe" *Statistical Journal of the United Nations ECE* 23: 1-10.
- Alho, J. M. and B. D. Spencer (1985) "Uncertain Population Forecasting" *Journal of the American Statistical Association* 80(390): 306-314.
- _____ (1990) "Error Models for Official Mortality Forecasts" *Journal of the American Statistical Association* 85(411): 609-616.
- _____ (1997) "The Practical Specification of the Expected Error of Population Forecasts" *Journal of Official Statistics* 13(3): 203-225.
- _____ (2005) *Statistical Demography and Forecasting* New York: Springer.
- Bell, W. R. (1997) "Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates" *Journal of Official Statistics* 13(3): 279-303.
- Booth, H. (2006) "Demographic Forecasting: 1980-2005 in Review" *International Journal of Forecasting* 22(3): 547-581.
- Booth, H., Hyndman, R. J., Tickle, L. and P. De Jong (2006) "Lee-Carter Mortality Forecasting: A Multi-Country Comparison of Variants and Extensions" *Demographic Research* 15(9): 289-310.
- Booth, H., Maindonald, J. and L. Smith (2002) "Applying Lee-Carter Under Conditions of Variable Mortality Decline" *Population Studies* 56(3): 325-336.
- Cohen, J. E. (1986) "Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-Based and Empirical Approaches" *Demography* 23(1): 105-126.
- De Beer, J. (1997) "The Effect of Uncertainty of Migration on National Population Forecasts: The Case of the Netherlands" *Journal of Official Statistics* 13(3): 227-243.
- _____ (2000) "Dealing with Uncertainty in Population Forecasting" Department of Population, Statistics Netherlands.
- De Beer, J. and M. Alders (1999) "Probabilistic Population and Household Forecasts

- for the Netherlands” Paper Presented at the European Population Conference EPC39, Hague, August 30 - September 3, 1999.
- George, M. V. (2001) “Population Forecasting in Canada: Conceptual and Methodological Developments” *Canadian Studies in Population* 28(1): 111-154.
- Hartmann, M. and G. Strandell (2006) *Stochastic Population Projections for Sweden*, Statistics Sweden, Research and Development Report.
- Howe, N. and R. Jackson (2006) “Long-Term Immigration Projection Methods: Current Practice and How to Improve It” CRR WP 2006-3, Center for Retirement Research at Boston College.
- Hyndman, R. J. and M. S. Ullah (2007) “Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach” *Computational Statistics & Data Analysis* 51: 4942-4956.
- Keilman, N. (1997) “Ex-Post Errors in Official Population Forecasts in Industrialized Countries” *Journal of Official Statistics* 13(3): 245-277.
- _____ (2005) “Erroneous Population Forecasts” pp. 7-26 in *Perspectives on Mortality Forecasting: Vol II Probabilistic Models*, edited by N. Keilman, Stockholm: Swedish National Social Insurance Board.
- Keilman, N. and D. Q. Pham (2000) “Predictive Intervals for Age-Specific Fertility” *European Journal of Population* 16(1): 41-66.
- _____ (2004) “Empirical Errors and Predicted Errors in Fertility, Mortality and Migration Forecasts in the European Economic Area” Discussion Papers No. 386, Statistics Norway, Social and Demographic Research.
- Keilman, N., D. Q. Pham and A. Hetland (2001) “Norway’s Uncertain Demographic Future” *Social and Economic Studies* 105, Statistics Norway.
- _____ (2002) “Why Population Forecasts Should Be Probabilistic - Illustrated by the Case of Norway” *Demographic Research* 6(15): 409-454.
- Keyfitz, N. (1981) “The Limits of Population Forecasting” *Population and Development Review* 7(4): 579-593.
- Lee, R. D. (1974) “Forecasting Births in Post-Transition Populations: Stochastic Renewal with Serially Correlated Fertility” *Journal of the American Statistical Association* 69(347): 607-617.
- _____ (1998) “Probabilistic Approaches to Population Forecasting” *Population and Development Review* 24(Supplement: Frontiers of Population Forecasting): 156-190.

- _____ (2000) "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications" *North American Actuarial Journal* 4(1): 80-93.
- _____ (2004) "Quantifying Our Ignorance: Stochastic Forecasts of Population and Public Budgets" *Population and Development Review* 30(Supplement: Aging, Health, and Public Policy): 153-175.
- Lee, R. D. and L. R. Carter (1992) "Modeling and Forecasting U.S. Mortality" *Journal of the American Statistical Association* 87(491): 659-671.
- Lee, R. and T. Miller (2000) "Assessing the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality" Paper Presented at the Annual Meetings of the Population Association of America, Los Angeles, March 23-25, 2000.
- Lee, R. D. and S. Tuljapurkar (1994) "Stochastic Population Forecasting for the United States: Beyond High, Medium, and Low" *Journal of the American Statistical Association* 89(428): 1175-1189.
- Li, N., Lee, R. D. and S. Tuljapurkar (2004) "Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data" *International Statistical Review* 72(1): 19-36.
- Li, Q., Reuser, M., Kraus, C. and J. Alho (2007) "Aging of a Giant: A Stochastic Population Forecast for China, 2001-2050" MPIDR Working Paper WP 2007-032, Max Planck Institute for Demographic Research.
- Long, J. F. and F. W. Hollmann (2004) "Developing Official Stochastic Population Forecasts at the US Census Bureau" *International Statistical Review* 72(2): 201-208.
- Lutz, W., Sanderson, W. C. and S. Scherbov (1998) "Expert-Based Probabilistic Population Projections" *Population and Development Review* 24(Supplement: Frontiers of Population Forecasting): 139-155.
- _____ (2004) "The End of World Population Growth" pp. 17-83 in *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation & Sustainable Development*, edited by W. Lutz, W. C. Sanderson, and S. Scherbov, London: Earthscan.
- Lutz, W. and S. Scherbov (1998) "An Expert-Based Framework for Probabilistic National Population Projections: The Example of Austria" *European Journal of Population* 14(1): 1-17.
- Matysiak, A. and B. Nowok (2007) "Stochastic Forecast of the Population of Poland, 2005-2050" *Demographic Research* 17(11): 301-338.
- National Research Council (2000) *Beyond Six Billion: Forecasting the World's*

- Population*, The National Academy Press.
- O'Neill, B. C., Balk, D., Brickman, M. and M. Ezra (2001) "A Guide to Global Population Projections" *Demographic Research* 4(8): 203-288.
- Scherbov, S. and M. Marmolo (2006) "Probabilistic Population Projections for the EU-25" European Demographic Research Papers, Vienna Institute of Demography.
- Smith, S. K. and T. Sincich (1988) "Stability Over Time in the Distribution of Population Forecast Errors" *Demography* 25(3): 461-474.
- Smith, S. and J. Tayman (2004) "Confidence Intervals for Population Forecasts: A Case Study of Time Series Models for States" Paper Presented at the Annual Meeting of the Population Association of America, Boston, April 1-3, 2004.
- Smith, S. K., Tayman, J. and D. A. Swanson (2001) *State and Local Population Projections: Methodology and Analysis*, New York: Kluwer Academic/Plenum Publishers.
- Statistics Canada (2005) Population Projections for Canada, Provinces and Territories 2005-2031, Ottawa, Ontario, Canada: Statistics Canada.
- Stoto, M. A. (1983) "The Accuracy of Population Projections" *Journal of the American Statistical Association* 78(381): 13-20.
- United Nations (2003) *World Population Prospects: The 2002 Revision*, New York, USA: United Nations.
- Wilson, T. and M. Bell (2004) "Australia's Uncertain Demographic Future" *Demographic Research* 11(8): 195-234.

Stochastic Demographic and Population Forecasting

Haebong Woo

Dealing with uncertainty has been a critical issue in demographic and population forecasting since 1980. This study reviews methodological developments in demographic and population forecasting over the last several decades. First, this study reviews the important issue of the uncertainty surrounding demographic forecasts. Several limitations of the traditional scenario approach to dealing with uncertainty are also discussed. Second, in forecasting demographic processes such as mortality, fertility, and migration, three approaches of stochastic forecasting are identified and discussed: expert judgment, statistical modeling, and analysis of historical forecast errors. Finally, this study discusses the current issues and directions for future research in stochastic demographic forecasting.

Key Words: Stochastic demographic and Population forecasting, Uncertainty, expert judgment, Statistical modeling, Analysis of historical forecast errors