

우리나라 소지역 인구 추계를 위한 방법론 연구: 추세외삽법을 중심으로

이상일* · 조대현**

Trend Extrapolation Methods for Small Area Population Projections in Korea

Sang-Il Lee* · Daeheon Cho**

요약: 본 논문의 주된 연구 목적은 우리나라 소지역 인구 추계 방법론으로서 추세외삽법이 가지는 가능성과 한계를 검토하는 것이다. 추세외삽법은 크게 단순외삽법, 복잡외삽법, 비율외삽법으로 범주화되며, 활용되는 과거 정보의 양과 미래 경향성의 형태에 대한 가정에 따라 모두 11개의 세부 기법으로 나뉜다. 세부 기법의 실행성을 평가하기 위해 서울시 노원구와 송파구에 대한 미래 인구 추계를 실행하였다. 중요한 결과는 다음과 같다. 첫째, 미래로 갈수록 추계 기법간 편차가 증가한다. 둘째, 매우 빠르게 인구가 증가하거나 감소하는 지역의 경우 다항곡선적합법이나 로지스틱곡선적합법은 상대적으로 그 적절성이 떨어진다. 셋째, 상위 공간단위의 인구 증감과 그 방향이나 정도에서 다른 특성을 보이는 하위 지역의 경우는 비율외삽법에 속하는 기법의 적절성이 상대적으로 낮다. 넷째, 과거 인구 변화 패턴이 복잡한 지역의 경우, ARIMA 모형의 장점이 가장 잘 드러난다. 소지역 인구 추계를 위한 추세외삽법의 가능성을 극대화하기 위해서는 다음의 두 가지가 고려될 필요가 있다. 첫째, 특정 추계 기법의 결과에 전적으로 의존할 것이 아니라 여러 추계 기법의 결과에 대해 일종의 절충적 관점을 채택할 필요가 있다. 둘째, 추세외삽법에 의한 소지역 인구 추계값은, 독립적인 차상위 공간단위의 추계값을 통한 통제 과정을 반드시 거쳐야 한다.

주요어: 인구 추계, 추세외삽법, 단순외삽, 복잡외삽, 비율외삽, 코호트-요인법

Abstract: The main objective of this paper is to provide a detailed discussion on possibilities and limitations of various trend extrapolation methods for small area population projections. Trend extrapolation methods are grouped into three categories, simple extrapolation, complex extrapolation, and ratio extrapolation, and are subdivided into 11 techniques in terms of how much past information is utilized and how to depict the future trend. Those 11 techniques are applied to a case study for two districts in Seoul, Nowon-gu and Songpa-gu. Main findings are as follows. First, the more distant future the population projection is done for, the more variable results the different extrapolation techniques yield. Second, polynomial curve fitting and logistic curve fitting techniques appear to be relatively less efficient for small areas having experienced a rapid population increase or decrease. Third, some ratio extrapolation methods seem to be relatively less suitable for small areas which have been largely deviated from their higher-level areas in terms of the amount and direction of population growth. Forth, ARIMA models work better than any other methods for small areas having experienced a complex population growth pattern. The study offers two lessons to improve the extrapolation methods' practicability for small area population projections. One is that an eclectic perspective is needed to harness projection results from multiple methods rather than depending exclusively upon a single method. Second, small area population projections from any extrapolation method need to be controlled in accordance with projection results for higher-level areas.

Key Words: Population projection, Trend extrapolation method, Simple extrapolation, Complex extrapolation, Ratio extrapolation, Cohort-component method

*서울대학교 지리교육과 교수(Professor, Department of Geography Education, Seoul National University, si_lee@snu.ac.kr)

**가톨릭관동대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Catholic Kwandong University, dhncho@gmail.com)

I. 서론

소지역에 대한 인구 추계의 필요성은 계속 대두되고 있지만, 데이터의 문제, 적용 방법론의 문제, 실행 소프트웨어의 문제 등으로 인해 실질적으로 인구 추계가 이루어지는 경우는 전 세계적으로도 많지 않다. 우리나라 통계청은 인구 센서스 결과와 연동하여 전국과 시도 단위의 인구 추계를 실시하고 있다(통계청, 2019a; 2019b). 따라서 시도 수준이 우리나라에서 인구 추계가 이루어지는 공식적인 공간단위 수준이다. 그러나 최근 시도 단위 미만 공간단위의 인구 추계에 대한 수요가 지방자치단체나 민간 분야에서 증대되고 있다. 특히 시군구 기초자치단체나 시도의 정책연구원들에서 인구 추계 수요는 매우 높다(김재국 등, 2006). 실질적으로 시군구 단위의 인구 추계가 개별적으로 이루어지고 있기도 하다. 그러나 자체적인 시군구 단위 인구 추계 시 개발계획상의 인구유입을 과다 계상하는 경향이 강해 국가 전체나 시도 수준에서 행해지는 인구 추계의 합리성을 따라가지 못하는 것이 일반적인 상황이다(이상일, 2014).

학술적인 차원에서 소지역 인구 추계가 하나의 연구 주제로 부상하고 있다(Davis, 1995; Swanson *et al.*, 2010; Wilson, 2011; Baker *et al.*, 2013; 2014; Wilson, 2015; Inoue, 2017). 예를 들어 미국의 경우는 주(state) 단위 인구 추계를 벗어나 카운티(county) 및 센서스 구역(census tract) 단위의 인구 추계를 위한 방법론들이 제안되고 있다(Smith and Rayer, 2012; Rayer and Smith, 2014; Hauer, 2019; Baker *et al.*, 2020). 우리나라의 시군구 단위 인구 추계에 대한 기존 연구는 많지 않다. 순이동 코호트-요인법을 적용한 연구(김태현 등, 2006; 이상일, 2014; 전광희 등, 2014)와 해밀턴-페리(Hamilton-Perry) 기법을 적용한 연구(이상림·조영태, 2005; 이상복, 2007; 김정희, 2008; 이상일, 2014)가 거의 전부를 차지한다. 최근 우리나라 통계청 역시 시군구 단위의 인구 추계를 위한 체계적인 방안을 수립하려는 노력을 하고 있다(통계청 보도자료, 2017). 시군구보다 더 작은 공간단위(예를 들어 읍면동 단위)에 대해서는 최소한 학문적인 장에서 소개된 적은 전무한 상황이다.

이러한 측면에서 시도의 하위 행정단위에 적용 가능한 인구 추계 기법에 대해 검토해 보는 작업은 매우 의미 있는 일이 아닐 수 없다.

인구 추계 기법 중 가장 널리 사용되고 있는 것은 코호트-요인법(cohort-component method)이다(Smith *et al.*, 2013). 코호트-요인법은 인구 변동의 세 가지 구성요소에 대한 개별적 추정이 가능하다는 점뿐만 아니라 성, 연령 등에 의해 구분되는 코호트들에 대한 개별적인 추정이 가능하다는 점에서 널리 사용되고 있다(이상일·조대현, 2012). 주로 한 국가 전체나 최상위 행정단위 수준의 인구 추계에 널리 사용되고 있으며, 우리나라의 통계청의 인구 추계 역시 이 기법에 기반을 두고 있다(통계청, 2019a; 2019b). 최근 들어 행정단위 간의 인구가동을 실질적으로 추계 과정에 포함시키는 완전한 형태의 다지역 코호트-요인법(multiregional cohort-component method)이 실행되고 있다(조대현·이상일, 2011; 이상일·조대현, 2012; 통계청, 2019b). 그런데 코호트-요인법의 가장 중요한 단점은 출산력, 사망력, 인구 이동에 대한 상세하고 광대한 자료를 요구한다는 점이다. 따라서 세 컴포넌트에 대한 자료 구득이 용이하지 않은 소지역에 대한 인구 추계에 대한 적용성은 높지 않았다.

이러한 측면에서 본 연구는 우리나라 소지역 인구 추계 기법으로서 추세외삽법(trend extrapolation)의 가능성에 주목하고자 한다. 여기에는 두 가지 이유가 있다. 첫째, 앞에서 언급한 것처럼, 소지역의 경우 완전한 코호트-요인법을 적용할 수 있는 수준의 데이터가 존재하지 않을 가능성이 크다. 둘째, 소지역 인구 추계의 경우 주로 성과 연령에 의해 규정되는 인구 집단별 세부 인구 추정이 목적이 아니라 전체 인구 추정이 목적인 경우가 많은데, 이 경우 코호트-요인법 등과 같은 보다 복잡한 추계 기법은 오히려 과도한 측면이 있다. 셋째, 추세외삽법은 인구 추계의 기법임과 동시에 보다 복잡한 인구 추계를 위한 세부 모형 설정 기법이기도 하다. 예를 들어, 코호트-요인법의 실행을 위해서는 미래의 출산력 수준, 사망력 수준, 인구 이동 수준의 추정을 위한 모형을 설정해야만 한다. 이때 추세외삽법은 매우 다양한 방식으로 활용될 수 있다.

따라서 본 연구의 주요 목적은 추세외삽법이 우리나라 소지역 인구 추계에 대해 가지는 함의를 검토하는 것이다. 우선적으로, 추세외삽법이 인구 추계 기법 전체에서 차지하는 위상에 대해 소개한다. 그리고 나서 추세외삽법에 속하는 다양한 하위 기법들의 특징을 Smith *et al.*(2013)에 의거해 정리한다. 각 하위 추세외삽법의 실행성을 우리나라 서울시 구별 주민등록인구에 대한 인구 추계를 통해 제시한다. 마지막으로 추세외삽법의 가능성과 한계를 토론한다.

II. 소지역 인구 추계 기법으로서의 추세외삽법

1. 인구 추계 기법의 분류

소지역 인구 추계에 가장 적절한 방법론을 선정하기 위해 우선적으로 필요한 것은 가용한 인구 추계 기법에 어떤 것들이 있고, 그것들의 특성이 어떠한지를 파악하는 것이다. 몇몇 학자들이 다양한 인구 추계 기법을 분류하는 체계를 제안해 왔다. 우선 Klosterman(1990)은 추계 기법을 크게 외삽 기법(extrapolation techniques), 코호트-요인법(cohort-component techniques), 경제분석 기법(economic analysis techniques)의 세 가지로 구분하였다. 이에 대해 Swanson and Tayman(2012)는 외삽법(extrapolation), 비율법(ration), 징후법(symptomatic), 회귀분석법(regression), 요인법(component), 표본기반법(sample based), 기타 기법 등의 7가지로 세분하였다. 그러나 가장 널리 사용되고 있는 인구 추계 기법의 분류법은 Smith *et al.*(2013)이 제안한 것이다(그림 1 참조). 그들에 따르면 ‘객관적’ 인구 추계 기법은 크게 네 가지로 구분되는데 각각의 특성을 살펴보면 다음과 같다.

첫째, ‘추세외삽법(trend extrapolation)’은 ‘추세연장법’으로 번역되기도 하는데, 과거의 추세가 미래에도 지속될 것이라는 가정에 기반한 기법이다. 과거 수년 혹은 수십 년 관측치를 이용해 미래 시점의 인구수를 추정한다. 이 기법은 적용 방식에 따라 크게 세 가지 방식으로 세분되며 뒤에서 자세히 다룰 것이다. 둘째, ‘코호트-요인법(cohort-component method)’은 개별 인구 집단(주로 성과 연령에 의해 구분되는 인구 집단)별로 인구 변동의 세 가지 요소인 출생, 사망, 인구 이동을 개별적으로 고려하여 인구를 추정하는 기법이다. 세 가지 요소 각각에 대한 추계가 이루어지기 때문에 전 세계적으로 가장 널리 사용되고 있는 지배적인 방법론으로 자리잡고 있다. 이 코호트-요인법은 다시 세부적으로 총이동(gross migration) 코호트-요인법, 순이동(net migration) 코호트-요인법, 해밀턴-페리(Hamilton-Perry) 모형으로 나누어진다. 셋째, 구조 모형법(structural models)은

첫째, ‘추세외삽법(trend extrapolation)’은 ‘추세연장법’으로 번역되기도 하는데, 과거의 추세가 미래에도 지속될 것이라는 가정에 기반한 기법이다. 과거 수년 혹은 수십 년 관측치를 이용해 미래 시점의 인구수를 추정한다. 이 기법은 적용 방식에 따라 크게 세 가지 방식으로 세분되며 뒤에서 자세히 다룰 것이다. 둘째, ‘코호트-요인법(cohort-component method)’은 개별 인구 집단(주로 성과 연령에 의해 구분되는 인구 집단)별로 인구 변동의 세 가지 요소인 출생, 사망, 인구 이동을 개별적으로 고려하여 인구를 추정하는 기법이다. 세 가지 요소 각각에 대한 추계가 이루어지기 때문에 전 세계적으로 가장 널리 사용되고 있는 지배적인 방법론으로 자리잡고 있다. 이 코호트-요인법은 다시 세부적으로 총이동(gross migration) 코호트-요인법, 순이동(net migration) 코호트-요인법, 해밀턴-페리(Hamilton-Perry) 모형으로 나누어진다. 셋째, 구조 모형법(structural models)은

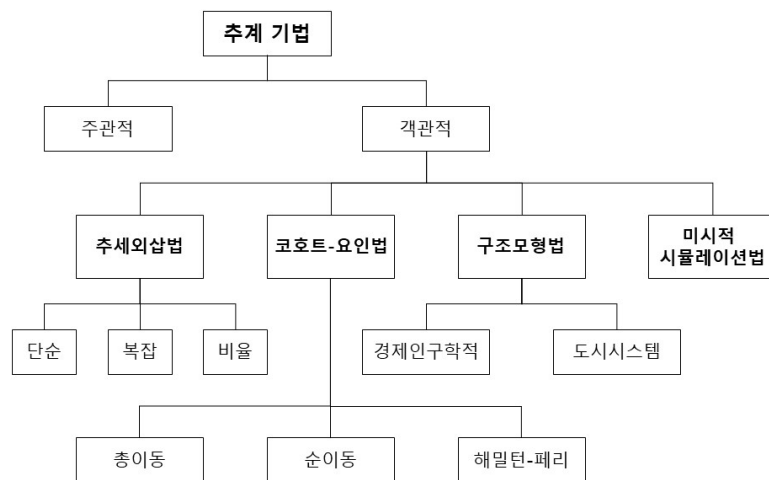


그림 1. 인구 추계 기법의 분류

* Smith *et al.*(2013)의 Fig. 1.1을 수정함.

인구 변수와 비인구 변수 간의 인과 관계를 이용해 미래 인구를 추정하는 기법이다. 이는 다시 경제인구학적(economic-demographic) 모형(우리나라에서의 적용 예로 이상일(2006) 참조)과 도시 시스템(urban system) 모형으로 나뉜다. 넷째, 미시적 시뮬레이션법(micro-simulation)은 개별 인간이나 가구를 대상으로 인구학적 의사 결정이나 행동 방식을 정식화함으로써 인구 추계를 실행하는 방식이다.

이러한 네 가지 인구 추계 기법의 장단점을 정리하면 다음과 같다(Smith *et al.*, 2013). 첫째, 추세의삽법은 적용이 단순하고 많은 데이터를 요구하지 않는다는 장점이 있지만, 주로 총인구 추계에 주로 사용된다는 단점이 있다. 즉, 인구 코호트에 대한 개별적인 추계가 이루어지지 않는다. 둘째, 코호트-요인법은 가장 과학적인 인구 추계 기법이라는 데는 의심의 여지가 없지만 많은 데이터를 요구하기 때문에 소지역의 인구 추계에 대한 적용성은 떨어지는 편이다. 셋째, 구조 모형법은 비인구학적인 요소를 도입한다는 측면에서 장점이 있지만, 고용 추계를 통한 간접적 인구 추계라는 점과 추세 외삽법과 마찬가지로 주로 총인구 추정에 한정된다는 단점이 있다. 미시적 시뮬레이션법은 인구 추계 기법으로서는 아직 실험적인 단계에 있어 적용성은 극히 낮다. 본 연구에서는 추세외삽법의 가능성과 한계에 집중하고자 한다.

2. 추세외삽법의 종류와 원리: 서울시 인구 추계에의 적용

추세외삽법은 크게 단순외삽법(simple extrapolation methods), 복잡외삽법(complex extrapolation methods), 비율외삽법(ratio extrapolation methods)으로 나뉘는데 각각의 특성을 간략히 정리하면 다음과 같다(Smith *et al.*, 2013). 단순외삽법은 과거의 두 연도(기본 연도(base year)와 추계 시작 연도(launch year))의 값을 이용해 직선 혹은 곡선 함수의 파라미터 값을 도출해 내고 이를 적용해 추계 목표 연도(target year)의 인구 수를 추정하는 것이다. 주로 선형(linear), 기하(geometric), 지수(exponential) 함수식이 이용된다. 둘째, 복잡외삽법은 과거 수년간의 데이터로부터 직선 혹은 곡선 형태로 표현되는

경향성을 도출한다. 선형추세(linear trend) 모형 외에 다항(polynomial), 지수(exponential), 로지스틱(logistic) 곡선적합(curve fitting) 기법이 있으며, 가장 복잡한 형태인 ARIMA(Autoregressive Integrated Moving Average) 모형도 사용된다. 셋째, 비율외삽법은 큰 지역 대비 작은 지역의 인구 비 개념을 이용하여 인구 추계를 실행하는 방식이다. 여기에는 고정-할당(constant-share), 변이-할당(shift-share), 성장-할당(share-of-growth) 기법이 포함된다.

개별 추세외삽법의 실행성을 검토하기 위해 서울시의 노원구와 송파구의 인구 추계를 실행하고자 한다. 기본 데이터는 지난 20년간(2000~2020년) 구별 주민등록인구 데이터이다(그림 2). 노원구와 송파구는 2000년 서울시 25개 구 중 인구가 가장 많은 두 구였다. 그러나 지난 20년간 매우 대조적인 인구 변동 패턴을 보여주었다. 노원구의 경우, 2001년 646,905명으로 정점에 이른 후 2009~2010년을 제외하고 지속적인 인구 감소를 보여주고 있다. 특히 2010년 혹은 2016년 이후는 인구 감소 경향이 심화되는 모습을 보여주고 있다. 2020년 인구는 523,037명으로 4위로 하락하였다. 이와는 대조적으로 송파구는 매우 복잡한 양상을 보여주고 있다. 전체적으로는 인구 증가를 보여주고 있지만, 2000~2004년은 감소(2004년 최소 인구 604,293명 기록), 2004~2010년은 증가(2010년에 최대 인구 685,279명 기록), 2010~2016년은 감소, 2016~2019년은 증가, 마지막으로 2019~2020년은 감소 등 감소와 증가가 반복적으로 발생하고 있다. 하지만 전 기간에 걸쳐 항상 서울시 전체에서 가장 인구가 많은 구를 기록했으며 전체적으로는 인구가 증가하는 추세를 보이고 있다. 2020년 인구 667,960명으로 여전히 서울시에서 인구 가장 많은 구를 기록하고 있다.

2000~2020년 동안 서울시 전체 인구는 10,311,314명에서 9,668,465명으로 감소하여, 연평균 인구성장률 -0.32%를 기록하였다. 이 기간 중에 노원구는 연평균 인구성장률 -0.95%를 기록하여 종로구(-1.11%)에 이어 가장 인구 감소의 속도가 빨랐던 구였다. 이에 비해 송파구는 연평균 인구성장률 0.04%를 기록하여 강서구, 서초구, 은평구에 이어 4위를 기록했으며 인구가 증가한 네 개 구에 포함되었다. 보다 최근

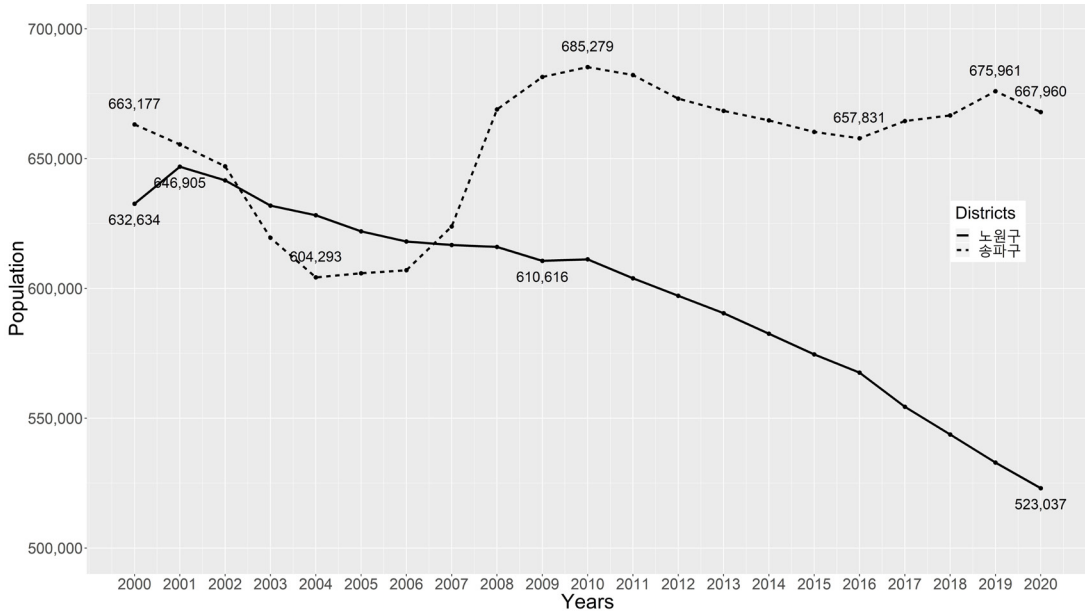


그림 2. 서울시 노원구와 송파구의 주민등록인구 변동 추이, 2000~2020년

* KOSIS 자료를 토대로 작성.

의 연도를 살펴보면, 2015~2020년 기간 동안 연평균 인구성장률 -0.72%를 기록해 감소 속도가 더 빨라 졌는데, 노원구는 -1.88%로 인구 감소 속도가 가장 빠른 구를, 송파구는 0.23%로 인구 증가 속도가 가장 빠른 구를 기록하였다. 결론적으로, 노원구는 인구가 지속적이고도 점진적으로 감소하는 간명한 인구 변동 패턴을 보이는 소지역을, 송파구는 전체적으로는 인구가 증가하지만 구간별로 인구의 증가와 감소가 반복되는 복잡한 인구 변동 패턴을 보이는 소 지역을 대변한다. 매우 상이한 인구 변동 특성을 보이는 두 구를 선정하여 인구 추계를 실행함으로써, 기법 간의 장단점이 보다 선명하게 드러날 것으로 기대한다.

1) 단순외삽법

(1) 선형외삽법

선형외삽법(linear extrapolation method)은 미래의 특정 기간 인구 수 변화는 과거의 동일 기간 인구 수 변화에 의해 결정된다는 점을 가정한다. 과거 두 시점의 인구로부터 그 기간의 AANC(average annual numeric change, 연평균변화수)를 계산해야

한다(Smith *et al.*, 2013).

$$AANC = (P_t - P_b)/y \tag{1}$$

여기서 P_t 은 추계 시작 시점(launch year)의 인구, P_b 는 기본 시점(base year)의 인구, y 는 기본 기간(base period), 즉 기본 시점과 추계 시작 시점의 연차(year difference)를 의미한다. 이 수식이 의미하는 바는 미래 인구 추계를 위해서는 과거 두 시점의 인구 값을 가지고 있어야 한다는 것이다. 보다 먼 과거의 인구가 기본 시점의 인구(P_b)가 되는 것이고, 보다 가까운 과거의 인구가 추계 시작 시점의 인구(P_t)가 되는 것인데, 보다 가까운 과거의 인구가 반드시 추계 시작 시점이 되어야 할 필요는 없지만, 통상적으로 구득 가능한 가장 가까운 시점의 인구를 사용하게 되고, 그 시점의 인구로부터 인구 추계를 시작하는 것이 합리적이기 때문에 이렇게 설정하게 된다. 예를 들어 2015년과 2020년의 인구를 활용해 2030년의 인구 수를 추정한다고 하면, 기본 시점의 인구는 2015년의 인구, 추계 시작 시점의 인구는 2020년의 인구, 기본 기간은 5년(2015~2010)이 된다. 즉,

$AANC = (P_{2020} - P_{2015})/5$ 가 된다.

AANC가 산출되면 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$P_t = P_l + (z \times AANC) \quad (2)$$

여기서 P_t 는 추계 목표 시점(target year)의 인구, P_l 은 추계 시작 시점의 인구, z 는 추계 기간(projection horizon)이다. 추계 기간은 추계 시작 시점과 추계 목표 시점의 연차를 의미한다. 위의 예로 보자면 추계 기간은 10(2030-2020)이 되어, $P_{2030} = P_{2020} + (10 \times AANC)$ 를 통해 인구 추계가 이루어진다.

서울시 데이터를 사용하여 이 과정을 설명하면 다음과 같다. 2015~2020년 인구를 바탕으로 2030년 인구를 추정하는 것이므로, 2015년이 기본 시점, 2020년이 추계 시작 시점, 2030년이 추계 목표 시점이 된다. 식 (1)과 (2)를 이용하여 2030년 노원구 인구를 추정하면 다음과 같다.

$$AANC = (523,037 - 574,583)/5 = -10,309.2$$

$$P_{2030} = 523,037 + 10 \times (-10,309.2) = 419,945$$

동일한 방식으로 2030년 송파구 인구를 추정하면 다음과 같다.

$$AANC = (667,960 - 660,302)/5 = 1,531.6$$

$$P_{2030} = 667,960 + 10 \times (1,531.6) = 683,276$$

결국 매년 얼마만큼의 인구가 증가하는지를 과거 두 시점의 인구로부터 구하고, 그것을 추계 연차에 곱하여 증가분을 구한 후, 그것과 추계 시작 시점의 인구수를 합산함으로써 특정 미래 시점의 인구를 추정하게 된다. 선형외삽법에 의한 인구 추계의 또 다른 특징은 개별 하위 지역의 인구 추계 값을 모두 합하면 전체 지역에 대한 인구 추계와 동일해진다는 점이다. 예를 들어, 서울시 전체에 이 기법을 적용하면 8,961,033명이 되는데 이는 동일한 기법을 25개 개별 구에 적용하여 얻은 추계값의 합산값과 동일해진다.

(2) 기하외삽법

기하외삽법(geometric extrapolation method)은 미래의 연평균 인구변화율은 과거의 연평균 인구변

화율과 동일할 것이라는 점을 가정한다. 우선, 기본 기간, 즉 기본 시점과 추계 시작 시점 간의 연평균기하변화율(average annual geometric rate of change)을 다음의 수식에 의거해 계산한다(Smith *et al.*, 2013).

$$r = (P_l / P_b)^{1/y} - 1 \quad (3)$$

여기서 P_b , P_l , y 는 식 (1)에서의 정의와 동일하다. 앞의 예를 적용하면, $r = (P_{2020}/P_{2015})^{1/5} - 1$ 의 공식에 의거해 구해진다. r 이 산출되면 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$P_t = P_l(1 + r)^z \quad (4)$$

여기서 P_t , P_l , z 는 식 (2)에서의 정의와 동일하다. 예에 적용하면, $P_{2030} = P_{2020}(1 + r)^{10}$ 이 된다. 기하외삽법은 선형외삽법과 달리, 개별 하위 지역의 인구 추계값의 합은 전체 지역에 대한 인구 추계값과 동일하지 않다.

서울시 데이터에 적용하여 2030년 노원구 인구를 추정하면 다음과 같다.

$$r = (523,037/574,583)^{1/5} - 1 = -0.01862$$

$$P_{2030} = 523,037 \times (1 - 0.01862)^{10} = 433,403$$

동일한 방식으로 2030년 송파구 인구를 추정하면 다음과 같다.

$$r = (667,960/660,302)^{1/5} - 1 = 0.00231$$

$$P_{2030} = 667,960 \times (1 + 0.00231)^{10} = 683,543$$

(3) 지수외삽법

지수외삽법(exponential extrapolation method)은 기하외삽법과 매우 유사한데, 평균지수변화율(average annual exponential rate of change)을 계산하여 인구 추계에 투입한다(Smith *et al.*, 2013).

$$r = \ln(P_l/P_b)/y \quad (5)$$

여기서 여기서 P_l , P_b , y 는 식 (1)에서의 정의와 동일하다. 예에 적용하면, $r = \ln(P_{2020}/P_{2015})/5$ 가 된다. r 이 산출되면 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$P_t = P_0 e^{rt} \quad (6)$$

여기서 P_t , P_0 , r , t 는 식 (2)에서의 정의와 동일하다. 예에 적용하면, $P_{2030} = P_{2020} e^{r \times 10}$ 이 된다.

서울시 데이터에 적용하여 2030년 노원구 인구를 추정하면 다음과 같다.

$$r = \ln(523,037/574,583)/5 = -0.01880$$

$$P_{2030} = 523,037 \times e^{(-0.01880 \times 10)} = 433,403$$

동일한 방식으로 2030년 송파구 인구를 추정하면 다음과 같다.

$$r = \ln(667,960/660,302)/5 = 0.00231$$

$$P_{2030} = 667,960 \times e^{(0.00231 \times 10)} = 683,543$$

서울의 예에서도 알 수 있듯이, 기하외삽법과 지수외삽법은 거의 동일한 값을 산출하며, 전자가 후자보다 항상 큰 값을 갖는다. 보통 둘 중에서 지수외삽법이 사용된다. 두 기법 모두 과거에 매우 빠르게 성장한 지역이 미래에 급속히 증가하는 결과를 보이게 된다.

2) 복잡외삽법

복잡외삽법(complex extrapolation method)은 단순외삽법에 비해 다음의 세가지 점에서 더 복잡하며, 따라서 더 정교한 방법이다(Smith *et al.*, 2013). 첫째, 복잡외삽법은 기본 기간에 대한 더 많은 과거 자료를 요구한다. 단순외삽법의 경우는 기본 시점과 추계 시작 시점이라는 두 시점의 데이터를 요구하지만 복잡외삽법은 더 많은 시점의 과거 데이터를 필요로 한다. 둘째, 미래 인구 추정의 수식이 훨씬 더 복잡하다. 이는 미래 인구 변동에 대한 보다 폭넓은 가정을 적용할 수 있게 해준다. 셋째, 통계적 절차를 통해 도출된 파라미터를 통해 미래 인구 추계의 간격을 설정할 수 있다. 그러나 기법의 복잡성은 항상 구현의 어려움을 동반하며, 복잡한 기법이 항상 더 좋은 예측력을 보여주는 것은 아니다.

복잡외삽법의 적용은 다음의 3단계로 구성된다(Smith *et al.*, 2013). 첫 번째 단계에서는 기본 기간 내에서 동일 간격의 다수의 시점을 정하고 그 시점에 대한 인구를 수집한다. 두 번째 단계에서는 특정

한 수리 모형(mathematical models)을 선택하고, 곡선적합(curve fitting)이라고 불리는 과정을 통해 해당 파라미터를 추정한다. 수리 모형의 선택이 복잡외삽법의 핵심인데, 수리 모형의 선택은 인구 변동의 본질에 대한 개념화와 미래 인구 변동의 경향성에 대한 프레임을 반영하기 때문이다. 세 번째 단계에서는 수리 모형과 추정된 파라미터를 이용해 인구 추계를 실행한다. 즉, 단순외삽법이 과거 데이터로부터 도출된 AANC나 r 을 미래 인구 추계에 그대로 적용하는 것과 마찬가지로 복잡외삽법은 과거 데이터로부터 도출된 모형 파라미터를 미래 인구 추계에 그대로 적용한다.

(1) 선형추세모형법

과거 시점의 데이터로부터 다음과 같은 선형추세 모형(linear trend model)을 수립한다. 일반적으로 OLS(ordinary least square) 회귀분석을 통해 적합된다(Smith *et al.*, 2013).

$$Y = a + bX \quad (7)$$

여기서 Y 는 인구이고, X 는 시점(연도)이다. 두 파라미터 혹은 회귀계수 중 b 가 보다 중요한 의미를 갖는데, 추정된 연평균 인구증가분이다. 즉, 매년 b 만큼의 인구가 지속적으로 증가한다는 것을 의미한다. 이는 개념적으로 선형외삽법의 AANC와 동일한 것으로, 그 값의 추정 방식이 다를 뿐이다.

파라미터가 추정되면, 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$P_t = a + bX_t + cf \quad (8)$$

여기서 P_t 는 추계 목표 시점의 인구, X_t 는 추계 목표 시점에 상응하는 시점값(time value), cf 는 보정계수(calibration factor)이다. X_t 는 추계 목표 시점에서 기본 시점(과거 데이터의 가장 이른 시점)을 빼고 거기에 1을 더한 값이다. 예를 들어 2000~2020년의 연 단위 과거 데이터를 통해 수리 모형을 적합한다고 하면, 추계 목표 시점 2030년의 시점값은 $31(2030 - 2000 + 1)$ 이 된다. 보정계수는 추계 시작 시점의 실측값에서 예측값을 뺀 값이다. 이는 적어도 추계 시작 시점에 대한 모형 추정값이 실측값과 동일하게

만들기 위한 조치로, 모형이 과거 모든 시점에 대해 적용된 것이므로 그 모형에 의거해 추계 시작 시점(데이터의 가장 늦은 시점일 가능성이 높음)을 정확히 추정할 가능성은 희박하기 때문에 이를 미리 조정하는 것이다. 추계 시작 시점의 실측값이 예측값보다 크다면, 그것은 모형이 과소추정을 한다는 것이므로 그만큼 값을 높여 주어야 하고(양의 보정계수 값), 반대로 추계 시작 시점의 실측값이 예측값보다 작다면 모형이 과대추정을 한다는 것이므로 그만큼 값을 낮추어 주어야 한다(음의 보정계수 값).

서울시의 사례를 통해 선형추세모형을 설명하면 다음과 같다. 우선 앞의 단순외삽법과 달리 2000~2020년의 21개 연도 모두를 사용한다. 즉, 21개의 데이터 포인트를 가지고 OLS 회귀분석을 실행하고 그 결과 도출된 회귀식을 바탕으로 인구 추계를 시행하는 것이다. 노원구의 경우 회귀계수는 각각 659,073.6과 -5,602.4이다. 이는 노원구의 인구가 매년 5,602.4명 감소한다는 것을 의미한다. 회귀식의 결정계수(R^2)는 0.9279로 매우 높다. 2020년의 실측값으로부터 예측값을 빼서 조정계수값을 구하고, 식 (8)에 의거해 2030년의 노원구 인구를 추정하면 다음과 같다. 위에서 설명한 것처럼, 2030년의 시점값은 31이다.

$$cf = 523,037 - 541,423.4 = -18,386.4$$

$$P_{2030} = 659,073.6 - 5,602.4 \times 31 - 18,386.4 = 467,013$$

송파구의 경우 회귀계수는 각각 630,587.1과 2,170.0이다. 이는 송파구의 인구는 노원구와 달리 매년 2,170명 증가한다는 것을 의미한다. 그런데, 회귀식의 결정계수(R^2)는 0.2653으로 노원구에 비해 매우 낮다. 이는 그림 2에도 나타나 있는 것처럼, 2000~2020년간의 송파구의 인구 변화 추이가 매우 복잡하게 나타나기 때문이다. 노원구와 동일한 방식으로 조정계수값과 2030년의 인구를 추정하면 다음과 같다.

$$cf = 667,960 - 676,157.5 = -8,197.5$$

$$P_{2030} = 630,587.1 + 2,170.0 \times 31 - 8,197.5 = 689,660$$

(2) 다항곡선적합법

단순외삽법에서 선형외삽법이 직선 패턴을, 기하

외삽법과 지수외삽법이 곡선 패턴을 가정했던 것처럼, 복잡외삽법에서도 앞의 선형추세모형법이 직선 패턴을 가정한다면, 다항곡선적합법(polynomial curve fitting)과 이후로 소개되는 기법은 모두 곡선 패턴을 가정한다. 시작은 선형추세모형법과 유사하다. 우선, 과거 시점의 데이터로부터 다음과 같은 다항곡선모형을 수립한다. 선형경향모형과 마찬가지로 일반적으로 OLS(ordinary least square) 회귀분석을 통해 적합된다(Smith *et al.*, 2013).

$$Y = a + b_1X + b_2X^2 + b_3X^3 + \dots + b_nX^n \quad (9)$$

선형추세모형법과 마찬가지로 Y 는 인구이고, X 는 시점(연도)이다. 이러한 모형을 다항회귀모형(polynomial regression model)이라고 하는데, 앞의 선형추세모형은 이러한 모형의 특수한 형태인 일차다항(first degree) 모형이다. X 와 X^2 가 사용되면 2차다항 모형, X , X^2 , X^3 이 사용되면 3차다항 모형이 되는 식이다.

보통 이차 다항 모형 혹은 이차 함수(quadratic function) 모형이 사용되는데, 이차 함수 모형이 수립되면 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$P_t = a + b_1 X_t + b_2 X_t^2 + cf \quad (10)$$

식 (8)에서와 마찬가지로 P_t 는 추계 목표 시점의 인구, X_t 는 추계 목표 시점에 상응하는 시점값, cf 는 보정계수이다. 파라미터 b_1 은 선형 요소(linear term)을 담당하고 b_2 는 곡선 요소(parabolic term)을 담당한다. 이때 중요한 것은 두 파라미터의 부호 간 조합이다(Smith *et al.*, 2013). 두 파라미터가 모두 양의 값을 보인다면, 추계 기간 동안 인구가 가속 증가한다는 것을 의미하는 반면, 두 파라미터가 모두 음의 값을 보인다면, 추계 기간 동안 인구가 가속감소한다는 것을 의미한다. 선형 요소가 양의 값을, 곡선 요소가 음의 값을 보인다면, 추계 기간 동안 인구가 감소 증가를 한다는 것을 의미하며, 중국에는 증가를 멈추고 인구 감소로 바뀐다는 것을 의미한다. 선형 요소가 음의 값을, 곡선 요소가 양의 값을 보인다면, 추계 기간 동안 인구가 감속 감소를 한다는 것을 의미하며, 중국에는 감소를 멈추고 인구 증가로 바뀐

다는 것을 의미한다.

서울시 데이터에 이차 다항 모형을 적용하여 인구 추계를 시행하였다. 노원구의 경우 회귀계수는 각각 637,262.8, 87.37, -258.6이다. 파라미터 b_1 은 양수 b_2 는 음수이므로, 미래로 갈수록 인구가 감소 증가하다가 특정 시점이 되면 증가를 멈추고 감소를 시작할 것이라는 점을 함축하고 있다. 회귀식의 결정계수(R^2)는 0.9855로 매우 높다. 2030년의 노원구 인구를 추정하면 다음과 같다.

$$\begin{aligned} cf &= 523,037 - 525,043.8 = -2,006.8 \\ P_{2030} &= 637,262.84 + 87.37 \times 31 - 258.63 \times 31^2 - \\ & 2,006.8 = 389,426 \end{aligned}$$

송파구의 경우도 노원구와 동일하게 2차 다항 모형을 적합하였다. 회귀계수는 각각 629,912.4, 2,346.0, -8.0이다. 파라미터 b_1 은 양수, b_2 는 음수이므로, 미래로 갈수록 인구가 감소 증가하다가 특정 시점이 되면 증가를 멈추고 감소를 시작할 것이라는 점을 함축하고 있다. 그러나 결정계수는 여전히 0.2654에 불과해 선형추세모형의 적합도와 거의 비슷한 수준으로 나타났다. 이 모형을 이용해 송파구의 2030년의 인구를 추정하면 다음과 같다.

$$\begin{aligned} cf &= 667,960 - 675,650.8 = -7,690.8 \\ P_{2030} &= 629,912.4 + 2,346.0 \times 31 - 8.0 \times 31^2 - \\ & 7,690.8 = 687,260 \end{aligned}$$

그런데 곡선성을 보다 잘 묘사해 줄 수 있는 고차의 다항 모형을 적합해 보았다. 결정계수는 3차의 경우 0.4673, 4차의 경우 0.7369, 5차의 경우 0.7409로 나타나 4차 다항 모형이 가장 적합한 것으로 드러났다. 그러나 4차 다항 모형에 의거한 2030년의 송파구 인구는 2,070,174로 나타나 상당히 비합리적인 결과가 도출되었다. 이는 과거 데이터의 적합도가 반드시 미래 추계의 합리성을 담보하는 것은 아니라는 점을 선명하게 보여주는 예라고 할 수 있다.

(3) 지수곡선적합법

곡선 패턴을 가정하는 또 다른 방식이 지수곡선적합법(exponential curve fitting)이다. 과거 시점의 데이터로부터 인구에 자연대수를 취한 뒤, 다음과

같은 선형추세모형을 수립한다. 다항곡선적합법과 마찬가지로 일반적으로 OLS(ordinary least square) 회귀분석을 통해 적합된다(Smith *et al.*, 2013). 그러나 자연대수를 취하는 대신, 원 인구를 종속변수로 두고 포아송 회귀분석(Poisson regression)을 통해 적합할 수도 있다. 그러나 여기서는 OLS 모형을 염두에 두고 설명하고자 한다.

$$\ln Y = a + bX \quad (11)$$

앞의 수식들에서와 마찬가지로 Y 는 인구이고, X 는 시점값이다.

이때 파라미터 b 는 매우 중요한 의미를 갖는다. 이것은 시점값이 한 단위 증가할 때마다 Y 가 $b \times 100\%$ 증가한다고 해석할 수 있기 때문인데, 이때 X 가 시점값, 즉 연단위이기 때문에 추계 기간 동안 인구가 매년 $b \times 100\%$ 씩 증가한다고 해석할 수 있다. 모형이 수립되고 파라미터가 추정되면, 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다.

$$\ln P_t = a + bX_t + cf \quad (12)$$

식 (10)에서와 마찬가지로 P_t 는 추계 목표 시점의 인구, X_t 는 추계 목표 시점에 상응하는 시점값, cf 는 보정계수이다. 보정계수에서 주의할 점은 이제 실제 값이 아니라 추정된 자연대수값 간의 차이로 정의된다는 점이다(Smith *et al.*, 2013).

추정값이 인구값 그 자체가 아니라 자연대수가 취해진 인구값이므로 다음의 식을 통해 최종적인 예측값을 획득하게 된다.

$$P_t = e^{a + bX_t + cf} \quad (13)$$

그런데 이러한 과정에서는 재전환 편향(retransformation bias)이 발생하기 때문에, 스미어링 추정량(smearing estimator)을 통해 보정해야 한다는 논리도 있다(Smith *et al.*, 2013). 스미어링 추정량은 잔차의 역로그 값의 평균을 의미하는 것으로, 그 값을 식 (13)의 지수 부분에 곱해 준다.

$$P_t = e^{(a + bX_t + cf) \times sme} \quad (14)$$

여기서 sme 는 스미어링 추정량을 의미한다. 보통 지수곡선적합법을 통한 인구 추계에서는 스미어링 추정

량은 1에 매우 가까운 값을 보이기 때문에(Smith *et al.*, 2013:196), 여기서는 식 (13)을 사용하기로 한다.

서울시 데이터에 지수곡선적합법을 적용하였다. 노원구의 경우 회귀계수는 각각 13.4034, -0.0095이다. 이는 노원구 인구가 미래에 매년 0.95%씩 감소할 것이라는 점을 함축하고 있다. 회귀식의 결정계수(R^2)는 0.9156으로 매우 높다. 2030년의 노원구 인구를 추정하면 다음과 같다.

$$cf = 13.16741 - 13.20339 = -0.03599$$

$$P_{2030} = e^{13.4034 - 0.0095 \times 31 - 0.03599} = 475,522$$

송파구의 경우 회귀계수는 각각 13.3536, 0.0034이다. 이는 송파구 인구가 미래에 매년 0.34%씩 증가할 것이라는 점을 함축하고 있다. 회귀식의 결정계수(R^2)는 0.2651로 노원구에 비해 훨씬 낮다. 2030년의 송파구 인구를 추정하면 다음과 같다.

$$cf = 13.41198 - 13.42457 = -0.01259$$

$$P_{2030} = e^{13.3536 + 0.0034 \times 31 - 0.01259} = 690,914$$

참고로 노원구와 송파구의 시미어링 추정량은 각각 1.000153과 1.000578로 이를 적용할 경우 2030년의 인구 추계값은 각각 476,472와 696,304이 된다.

(4) 로지스틱곡선적합법

지금까지의 모든 추계 기법은 인구 추계값의 한계(특히 상한)(explicit ceiling, upper limit)를 부과하지 않는다. 그러나 현실적으로 인구는 무한정 증가할 수 없고, 특정한 시점에 특정한 수준의 인구로 수렴하는 경향이 있다고 가정하는 것이 현실적이다. 이러한 측면을 반영하는 기법이 로지스틱곡선적합법(logistic curve fitting)이다. 물론 로지스틱곡선적합법 외에도 이러한 속성을 공유하는 모형들이 가능하다. 예를 들어 수정지수모형(modified exponential models)과 고펜르츠 모형(Gompertz models)도 상한을 설정할 수 있는 기법이며, 수정지수곡선(modified exponential curve)과 쌍곡선(hyperbolic curve)은 특히 하한 설정이 가능하기 때문에 빠르게 인구가 감소하고 있는 지역의 추정에 적합하다(Smith *et al.*, 2013). 로지스틱 곡선은 몇 개의 파라미터로 구성되느냐에 따라 3-파라미터, 4-파라미터, 5-파라미터 모

형이 가능하지만 여기서는 3-파라미터 로지스틱 곡선에 집중하고자 한다.

통상적으로 인구 추계에서는 Keyfitz(1968)의 모형이 주로 사용된다(Smith *et al.*, 2013).

$$Y = a/(1 + be^{-cX}) \tag{15}$$

여기서 a 는 상한점근값(upper asymptote), b 와 c 는 로지스틱 곡선의 형태를 결정하는 파라미터이다. 이 파라미터들은 총체적으로 한 인구의 상한이 얼마이며 그 상한에 언제쯤 도달하는지를 결정한다. 이 파라미터 추정을 위해서는 OLS가 아닌 일종의 반복최소제곱법(iterative least square technique)이 사용된다. 파라미터가 추정이 되면, 다음의 공식에 의거해 미래 시점에 대한 인구 추계가 이루어진다(Smith *et al.*, 2013).

$$P_t = a/(1 + be^{-cX_t}) + cf \tag{16}$$

그런데 미래 인구의 상한을 결정하는 파라미터인 a 에 대해서는 두 가지 경우가 존재한다(Smith *et al.*, 2013). 첫 번째는 통계 패키지가 자동으로 추정된 파라미터를 그대로 사용하는 경우이다. 그런데 이렇게 자동으로 추정된 파라미터는 현실적이지 않은 경우가 빈번하다. 두 번째는 연구자가 그 파라미터를 우선적으로 결정하여 모형에 투입하는 것이다. 이 두 가지 옵션은 서로 다른 장단점을 가진다. 첫 번째 옵션의 경우, 모형 결정이 용이하고 미래 인구의 성장 한계가 매우 불확실한 경우에는 장점이 있지만, 비현실적으로 큰 값은 과대 추정의 결과를 낳기 쉬운 단점이 있다. 두 번째 옵션의 경우, 모형의 현실성을 제고하고 미래 인구의 성장 한계가 매우 확실한 경우에는 장점이 있지만, 자의적 상한값 설정을 정당화하는 데 항상 어려운 점이 있는 단점이 있다.

본 연구에서는 R 의 drc 패키지(Ritz, *et al.*, 2015)를 활용하여 3-파라미터 로지스틱곡선적합법을 실행하도록 한다. 식 (15)와는 다소 다른 형식으로 다음과 같이 주어진다.

$$Y = a/(1 + e^{b(X-c)}) \tag{17}$$

여기서, a, b, c 가 곡선의 형태를 결정하는 세 파라미터이다. c 는 변곡점(inflexion point)을 나타내는데,

그 점을 중심으로 좌우대칭이 이루어진다. 앞에서와 마찬가지로 상한값을 미리 설정한 후 모형을 적합할 수도 있다. 파라미터가 추정되면 다음의 공식에 의거해 인구 추정이 이루어진다.

$$P_t = a/(1 + e^{b(x_i - c)}) + cf \quad (18)$$

서울시 데이터에 로지스틱곡선적합법을 적용하였다. 노원구에 대한 세 파라미터는 각각 653,255.6, 0.119836, 32.530129이다. 변곡점 파라미터 값이 의미하는 바는 노원구의 인구가 가속감소하고 있지만 대략 2031~2032년에 발생할 변곡점을 중심으로 감속감소가 이루어질 것이라는 점이다. 2030년의 노원구 인구를 추정하면 다음과 같다.

$$cf = 523,037 - 522,126.2 = 910.8$$

$$P_{2030} = 653,255.6 / (1 + e^{0.119836 \times (31 - 32.530129)}) + 910.8 = 357,401$$

송파구의 경우 세 파라미터 값은 각각 747,872.1, -0.026631, -62.464734이다. 변곡점 파라미터가 매우 큰 음의 값을 보이고 있다는 것은 수십년 전에 변곡이 이루어졌다는 것을 의미하는 것으로 사실상 선형 모형과 큰 차이가 없음을 시사한다. 2030년의 송파구 인구를 추정하면 다음과 같다.

$$cf = 667,960 - 674,784.5 = -6,824.5$$

$$P_{2030} = 747,872.1 / (1 + e^{0.026631 \times (31 + 62.464734)}) - 6,824.5 = 683,738$$

로지스틱곡선적합법은 많은 장점에도 불구하고, 파라미터(특히 상한값 파라미터)의 작은 차이가 추계 결과에서 큰 차이를 야기한다는 심대한 단점이 있다(Smith *et al.*, 2013).

(5) ARIMA 모형법

ARIMA 모형법(Autoregressive Integrated Moving Average model, 자기회귀누적이동평균 모형)은 시계열 데이터를 통계적으로 모형화하는 여러 기법들 중 하나이다. 보통 자기회귀 모형(Autoregressive mode, AR), 이동평균 모형(Moving average model, MA), ARIMA 모형이 이 범주에 포함되는데, ARIMA 모형이 가장 진보된 형태인 것으로 인정받

고 있다. 기본적으로 본 연구에서 초점을 두고 있는 추계외삽법은 과거의 데이터를 통해 미래를 예측하는 것이기 때문에 시계열 데이터의 특성을 그대로 보유하고 있고, 따라서 시계열 통계 모델링에서 널리 사용되고 있는 ARIMA 기법의 적용은 매우 자연스러운 것이다. 더구나 ARIMA 모형은 앞에서 살펴본 다양한 형태의 회귀분석-기반 기법에 비해 보다 정확한 추계값을 산출하는 것으로 알려져 있다(Smith *et al.*, 2013). 그러나 상대적으로 기법이 복잡하여 이해하기 어렵고, 보다 많은 수의 과거 관측값이 요구된다는 의미에서 실행성이 다소 낮은 것도 사실이다.

ARIMA 모형은 AR 모형과 MA 모형을 결합한 것으로 이해할 수 있는데, AR 모형은 특정 시점의 값을 예측하기 위해 시계열상의 과거 관측값 p 개를 이용해 예측 모형을 생성하는 것이고, MA 모형은 과거 예측 오차값 q 개를 이용해 예측 모형을 생성하는 것이다(곽기영, 2019:376). 결국 세 개의 파라미터로 구성되는 ARIMA(p, d, q) 모형은 주어진 데이터를 d 차 차분(differencing)하고(차분 차수), 과거 d 개 관측값(AR 모형 차수)과 q 개 오차값(MA 모형 차수)을 통해 예측값을 산출하는 모형이다(Hyndman and Athanaspoulos, 2018). 보통 세 파라미터는 0, 1, 2 셋 중의 한 값을 갖는다. 데이터의 차분은 추세를 갖는 시계열 데이터에 정상성(stationarity)을 부여하기 위해 시행하는 작업으로 특정 시점의 관측값을 그 값에서 직전 시점의 관측값으로 차감함으로써 이루어진다. 이러한 1차 차분값(first-order differences)을 활용해도 정상성이 확보되지 않으면 2차 차분값을 활용할 수도 있다(Smith *et al.*, 2013:200; 곽기영, 2019:372). 최적의 p, q 결정은 ACF(autocorrelation function), PACF(partial autocorrelation function), AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion)과 같은 함수나 통계량을 통해 이루어진다(Smith *et al.*, 2013).

본 연구에서는 R forecast 패키지(Hyndman and Athanaspoulos, 2018; Hyndman *et al.*, 2020)를 활용하여 ARIMA 모형에 기반한 인구 추계를 실행하고자 한다. 특히 내장되어 있는 auto.arima를 실행하면 최적의 파라미터 값을 산출해 주기 때문에 적용이

용이한 장점이 있다. 노원구의 경우 최적의 모형은 ARIMA(0, 2, 0)인 것으로 드러났다. 이것은 노원구의 과거 인구 변화는 2차 차분을 통해서 정상성을 획득할 수 있는데, 특정한 자기회귀 프로세스나 이동평균 프로세스를 보여주지는 않는다는 것을 의미한다. 송파구의 경우는 ARIMA(2, 0, 0)이 최적인 것으로 드러났다. 이는 정상성을 위한 차분이 필요 없으며, 2차 자기회귀 프로세스가 지배적이라는 것을 의미한다. 노원구와 송파구의 ARIMA 모형에 의한 2030년의 추계값은 각각 424,357과 656,521이다. ARIMA 모형법이 특히 송파구 인구 추계에서 갖는 의미는 뒤에서 보다 자세하게 다루도록 한다.

3) 비율외삽법

비율외삽법(ratio extrapolation methods)은 상위 공간단위와 하위 공간단위가 계층-포섭적으로 구성되어 있고, 상위 공간단위에 대한 미래 추계값이 외부에서 주어진다 가정하에서 하위 공간단위의 인구를 추정하는 상황에 적절하다. 요구되는 데이터의 양이 적으며, 적용하기 용이하다는 특징이 있다. 그런데 비율외삽법은 상위 공간단위에 대한 미래 추계치가 외부에서 주어져야만 한다. 따라서 서울시 노원구와 송파구에 대한 비율외삽법에 기반한 인구 추계를 위해서는 서울시 전체 인구에 대한 미래 연도의 추계값이 필수적이다. 이를 위해 통계청이 2017년 7월 1일 시점의 인구를 추계 기준인구(base population)로 삼아 2047년까지 향후 30년간의 인구를 추정할 시도별 장래인구 추계(중위 추계) 결과를 활용하였다(통계청, 2019a; 2019b). 2000~2020년 자료는 12월 31일자 기준인 주민등록인구이고, 미래의 인구는 7월 1일자 기준의 장래인구 추계값이어서 직접 비교

에 다소간의 난점이 있다. 예를 들어 2020년 서울시 전체의 주민등록인구는 9,668,465이지만, 통계청의 2020년 시도별 추계값은 9,601,693이다. 그러나 추계 기법을 이해한다는 측면에서는 큰 문제가 없을 것으로 판단하여 특별한 조정 과정을 거치지 않고 통계청의 시도별 장래인구 추계값을 그대로 사용하였다.

(1) 고정-할당법

고정-할당법(constant-share method)은 하위 공간단위의 상위 공간단위에 대한 할당(비중)이 미래에도 지속될 것이라는 단순한 가정에 기반한다. 고정-할당법은 과거 한 시점의 데이터만을 요구하며 다음의 공식에 의거해 인구 추계가 이루어진다(Smith *et al.*, 2013).

$$P_{it} = \frac{P_{it}}{P_t} \times P_t \quad (19)$$

여기에서 P_{it} 는 추계 목표 시점의 하위 공간단위 i 의 인구, P_t 는 추계 목표 시점의 상위 공간단위의 인구, P_{it} 은 추계 시작 시점의 하위 공간단위 i 의 인구, 그리고 P_t 은 추계 시작 시점의 상위 공간단위의 인구이다.

표 1은 고정-할당법에 의한 서울시 노원구와 송파구의 인구 추계 과정을 보여주고 있는데, 2020년을 추계 시작 시점으로 하여 2030년의 인구를 추정하고 있다. 추계 과정을 설명하면 다음과 같다. 우선 식 (19)의 우변의 첫 번째 요소는 각 구별 할당값(2020년 서울 전체 인구에서 개별 구가 차지하는 비중)을 산출하는 것인데, 노원구와 송파구는 각각 0.05410과 0.06909이다. 2030년 추계값을 구하기 위해 이 할당값에 추계 목표시점인 2030년의 서울시 전체에 대

표 1. 고정-할당법에 의한 서울시 구별 인구 추계

	2020년		2030년 추계 ^a	2020~2030년 변화	
	인구	할당		규모	비율(%)
노원구	523,037	0.05410	495,287	-27,750	-5.31
송파구	667,960	0.06909	632,521	-35,439	-5.31
나머지구	8,477,468	0.87682	8,027,694	-449,774	-5.31
서울시 전체	9,668,465	1.00000	9,155,502 ^b	-512,963	-5.31

^a 2020 할당 × 2030 서울 전체 인구; ^b 통계청 시도별 장래 인구 추계 결과.

한 외부 추계값(9,155,502)을 곱하면, 해당 연도의 구별 추계값(495,287과 632,521)을 구할 수 있다. 이 기법은 과거 한 시점의 데이터만을 요구한다는 측면에서 간단하다는 뚜렷한 장점이 있지만, 하위 공간단위의 인구 성장 추이가 상위 공간단위의 인구성장 추이와 항상 동일할 수밖에 없다는 치명적인 단점이 있다. 표 1에서도 보듯이 2020~2030년에 서울시 전체 인구가 감소하기 때문에 모든 하위 구의 인구도 감소해야만 하고, 인구성장률도 서울시 전체와 모든 구에 걸쳐 동일하게 드러난다(-5.31%).

(2) 변이-할당법

변이-할당법(shift-share method)은 하위 공간단위의 상위 공간단위에 대한 비중(할당)이 고정되어 있는 것이 아니라 경향성을 가지면서 변화한다는 점을 가정한다. 다음의 공식에 의거해 인구 추계가 이루어진다. 고정-할당법과 달리 두 시점(기본 시점과 추계 시작 시점)의 데이터가 필요하다(Smith *et al.*, 2013).

$$P_{it} = \left[\frac{P_{il}}{P_{.l}} + \frac{z}{y} \left(\frac{P_{il}}{P_{.l}} - \frac{P_{ib}}{P_{.b}} \right) \right] \times P_{.t} \quad (20)$$

여기에서 P_{ib} 는 기본 시점의 하위 공간단위 i 의 인구, $P_{.b}$ 는 기본 시점의 상위 공간단위의 인구, z 는 추계 기간(추계 시작 시점과 추계 목표 시점의 연차), 그

리고 y 는 기본 기간(기본 시점과 추계 시작 시점의 연차)이다. 기본 기간 동안 발생한 각 하위 공간단위들의 할당 변화 규모가 미래에도 지속될 것이라는 가정 하에, 추계 시작 시점의 각 하위 공간단위 비중에 그 비중 변화값을 합산하여 특정 추계 목표 시점에서 각 하위 공간단위가 전체에서 어느 정도의 할당을 갖게 되는지를 추정하는 방식으로 추계가 이루어진다.

표 2는 변이-할당법에 의한 서울시 노원구와 송파구의 인구 추계 과정을 보여주고 있다. 우선적으로 고려되어야 할 것은 개별 공간단위의 할당값이 2020~2030년 사이에 얼마만큼 변화할 것인가를 추정하는 것이다. 이 값이 추정되면 그것을 2020년의 할당값과 합하여 2030년의 할당값을 구하고, 그것에 2030년 서울시 전체 인구를 곱하면 해당 구의 인구 추계값을 획득하게 되는 것이다. 노원구를 예로 들어 이 과정을 설명하면 다음과 같다. 과거 두 시점인 2000년과 2020년의 노원구의 할당값은 각각 0.06135와 0.05410이다. 이 두 값의 차인 -0.00726이 2000~2020년에 발생한 할당에서의 변화값인데, 이 변화의 경향이 추계 기간 동안에도 지속될 것이라고 가정한다. 그런데 이 할당 변화값은 20년간에 걸쳐 발생한 것이고 추계 기간은 10년이므로, 2020~2030년에 발생할 할당 변화값은 2000~2020년에 발생한 할당 변화값의 절반일 것으로 추정하는 것이 합리적이다

표 2. 변이-할당법에 의한 서울시 구별 인구 추계

	2000년		2020년		할당에서의 변화	
	인구	할당	인구	할당	2000~2020	2020~2030 ^a
노원구	632,634	0.06135	523,037	0.05410	-0.00726	-0.00363
송파구	663,177	0.06432	667,960	0.06909	0.00477	0.00239
나머지구	9,015,503	0.87433	8,477,468	0.87682	0.00249	0.00124
합계	10,311,314	1.00000	9,668,465	1.00000	0.00000	0.00000
	2030년		2020~2030년 변화			
	할당 ^b	추계 ^c	규모	비율(%)		
노원구	0.05047	462,070	-60,967	-11.66		
송파구	0.07147	654,362	-13,598	-2.04		
나머지구	0.87806	8,039,070	-438,398	-5.17		
합계	1.00000	9,155,502 ^d	-512,963	-5.31		

* ^a 2000~2020 할당에서의 변화 × 0.5; ^b 2020 할당 + 2020~2030 할당에서의 변화; ^c 2030 할당 × 2030 서울시 전체 인구; ^d 통계청 시도별 장래 인구 추계 결과.

(-0.00363). 결국 2030년 노원구의 추정 할당값은 이 할당 변화값에 2020년의 할당값을 더하면 구할 수 있게 되고(0.05047), 그 추정 할당값에 2030년 서울시 전체의 인구 추계값(9,155,502)을 곱하면, 2030년 노원구 인구 추계값 462,070을 구할 수 있게 된다.

표 2의 2020~2030년 비율(%) 변화를 보면, 앞의 고정-할당법과 달리 구와 서울시 전체의 인구성장률이 각기 다르게 계산된 것을 볼 수 있다. 이는 변이-할당법의 고정-할당법에 대한 상대적인 장점이다. 그러나 이 기법의 가장 큰 단점은 기본 기간 중 인구가 감소했거나 미세하게 증가한 하위 공간단위의 경우, 상당히 먼 미래의 추계값이 급격히 줄어들 수 있고 심지어 음수가 될 수도 있다는 점이다(Smith *et al.*, 2013). 노원구의 경우 그러한 경향이 잘 드러나는데, 앞의 고정-할당법과 비교했을 때 2020~2030년의 인구변화율이 훨씬 낮게 추정되었다(-5.31에서 -11.66). 물론 이러한 극단적인 경우는 잘 발생하지 않을 뿐만 아니라 할당 변화값에 특정한 제약을 부과함으로써 해결될 수도 있지만, 장기 추계에 적용하는 경우에는 큰 한계점이 되는 것은 분명한 사실이다.

(3) 성장-할당법

성장-할당법(share-of-growth method)은 하위 공간단위의 성장이 상위 공간단위의 성장에서 차지하는 비중이 미래에도 동일하다고 가정하며, 다음의 공식에 의거해 인구 추계가 이루어진다(Smith *et al.*, 2013).

$$P_{it} = P_{il} + \frac{P_{iu} - P_{ib}}{P_{l} - P_{b}} \times (P_t - P_l) \quad (21)$$

기본 기간 중에 상위 공간단위에서 발생한 인구 증감 중 특정 하위 공간단위에서 발생한 인구 증감이 차지하는 비중(할당)을 계산하고, 그것을 추계 기간에서 발생할 상위 공간단위에서의 인구 증감에 곱함으로써 해당 하위 공간단위에서 얼마만큼의 인구 증감이 추계 기간 중 발생할 것인가를 계산하여 미래 인구를 예측하는 방식이다.

표 3은 성장-할당법에 의한 서울시 노원구와 송파구의 인구 추계 과정을 보여주고 있다. 기본 기간

(2000~2020년)에 노원구와 송파구는 각각 -109,597명 증가와 4,783명 증가를 기록하였다. 동일 기간, 서울시 전체의 인구 증감은 -642,849명이었고, 이 값으로 노원구와 송파구 각각의 인구 증감값을 나누면 인구 변화에서의 할당값을 구할 수 있다(각각 0.17049와 -0.00744). 일단 이 할당값이 단순한 비중이 아니라 부호가 부여된 비중이라는 사실을 인식할 필요가 있다. 서울시 예의 경우, 분모가 음수이므로 인구가 감소한 노원구의 할당값은 양수, 인구가 증가한 송파구의 할당값은 음수라는 점을 염두에 둘 필요가 있다. 이 부호가 부여된 할당값이 미래에도, 즉 추계 기간 중에도 지속될 것이라는 것이 성장-할당법의 가장 중요한 가정이다. 2030년의 서울시 전체에 대한 추계값을 적용하면, 2020~2030년 서울시 전체의 인구증감을 계산할 수 있고(-512,963), 그 값을 각 구의 2000~2020년의 변화 할당값에 곱하면 추계 기간 중 얼마만큼의 인구증감이 발생하는지를 추정할 수 있다. 그 값을 2020년의 구 인가와 합산하면 최종적으로 2030년의 인구 추계값을 구할 수 있게 된다. 노원구의 경우, 나머지 두 기법에 비해 가장 감소의 속도가 가파른 것으로 추계되지만(-11.66%), 송파구의 경우는 나머지 두 기법과 달리 증가할 것으로 추계되었다(0.57%).

이 기법의 문제점은 기본 기간 중에 발생한 하위 공간단위에서의 인구 증감($P_{il} - P_{ib}$)과 상위 공간단위의 인구 증감($P_l - P_b$)의 방향이 서로 다른 경우에 발생한다(Smith *et al.*, 2013). 만일 기본 기간 동안 상위 공간단위는 증가했고, 특정 하위 공간단위는 감소했다고 하자. 이런 경우 해당 하위 공간단위에 대한 증분비(수식의 마지막 분수)는 음수가 될 것이다. 그런데, 추계 기간 동안 상위 공간단위가 증가했다고 하면, 위의 수식에 의거해 해당 하위 공간단위의 인구는 감소하게 된다. 비록 기본 기간 동안 해당 하위 공간단위의 인구가 감소했다 하더라도, 추계 기간 동안 상위 공간단위의 인구가 증가했다면, 해당 하위 공간단위의 감소폭은 줄거나 어떤 경우에는 증가해야 한다. 그러나 현 수식에서는 기본 기간 동안 상위 공간단위가 증가한 상황에서 하위 공간단위가 감소하면, 추계 기간 동안 해당 하위 공간단위의 인구는 무조건 감소해야 한다. 반대로 추계 기간

표 3. 성장-할당법에 의한 서울시 구별 인구 추계

	2000~2020년 변화		인구		2020~2030년 변화	
	증감	할당	2020년	2030년 ^a	규모 ^b	비율(%)
노원구	-109,597	0.17049	523,037	435,584	-87,453	-16.72
송파구	4,783	-0.00744	667,960	671,777	3,817	0.57
나머지구	-538,035	0.83695	8,477,468	8,048,142	-429,326	-5.06
합계	-642,849	1.00000	9,668,465	9,155,502 ^c	-512,963	-5.31

* ^a 2020 인구 + 2020~2030 인구 증감; ^b 2000~2020 변화 할당 × 2020~2030 서울시 인구 증감; ^c 통계청 시도별 장래 인구 추계 결과.

동안 상위 공간단위가 감소하면 해당 하위 공간단위는 증가하게 된다.

서울시의 경우를 보면, 노원구와 송파구의 인구 추계가 합리적으로 이루어진 것을 볼 수 있는데, 이는 2020~2030년 서울시 전체 인구의 감소라는 공공로운 전제 때문에, 각 구의 할당 부호가 반전하면서 발생한 일이라는 점을 알 수 있다. 즉, 2000~2020년에 인구 감소를 보인 노원구는 해당 기간 중에 서울시 인구가 감소하기 때문에 할당값이 양수로 전환했고, 다시 2020~2030년 서울시 전체 인구가 감소하는 것으로 전제되었기 때문에 해당 기간 동안 인구 증감의 부호가 다시 음수로 전환하여, 최종적으로 추계 기간 중 인구가 감소하는 것으로 나타나게 된 것이다. 송파구는 정반대의 상황을 보여준다. 그러나 2020~2030년에 서울시 전체 인구의 증가가 전제된다면, 부호의 반전이 발생하지 않기 때문에, 노원구의 인구는 증가하고 송파구의 인구는 감소하는 비합리적인 추계 결과가 산출되었을 것이다. 따라서 성장-할당법은 부호와 관련된 이러한 치명적인 한계가 있으며, 이러한 문제를 해결하기 위해서는 일종의 부가적인 조정 과정이 요구되는데, 이는 본 논문의 범위를 넘어서는 것으로 다른 독립적인 논문에서 다루어져야 한다.

3. 추계 결과 비교

표 4는 11개의 서로 다른 추계 외삽 기법의 결과와 그 결과들 간의 편차를 보여주고 있다. 추계 목표 연도는 2030년, 2040년, 2047년인데, 2047년이 선택된 것은 비율외삽법이 서울시 전체에 대한 외부 추계값을 필요로 하고, 통계청의 최신 시도 단위 추

계는 2047년까지이기 때문이다. 표의 하단에 있는 추계 편차에서 규모는 최대 추계값과 최소 추계값 간의 차이를 의미하고, 비율(%)은 그 값의 최소 추계값에 대한 비중을 백분율로 표현한 것이다. 예를 들어 서울시 송파구의 2047년의 추계에서, 추계 편차 규모 156,977은 최대 추계값인 731,760(지수곡선)에서 최소 추계값인 574,784(고정-할당)을 뺀 값이고, 추계 편차 비율(%) 27.3은 추계 편차 규모를 최소 추계값으로 나누고 거기에 100을 곱한 값이다.

이 표로부터 몇 가지 결론이 도출되었다. 첫째, 미래로 갈수록 추계 편차가 증가한다. 서울시 노원구와 송파구 모두 2030년, 2040년, 2047년으로 갈수록 추계 편차의 규모와 비율(%)이 모두 증가한다. 이는 매우 당연한 것인데, 먼 미래일수록 추계의 합리성이 하락할 수밖에 없고 이는 추계 기법 간의 추계값에서의 편차가 더 커질 수밖에 없는 이유가 된다.

둘째, 서울시 노원구의 경우, 다항곡선적합법과 로지스틱곡선적합법은 2047년에 대해 극단적으로 작은 값(43,577과 89,376)을 산출하여, 가장 불합리한 추계 기법인 것으로 드러났다. 서울시 노원구의 경우, 특히 2010년 이후로 가속감소하는 경향을 보여주고 있는데, 두 복잡외삽법은 이러한 경향을 극단적으로 대변하는 모형을 구성함으로써 먼 미래에 대해 극단적으로 작은 추계값을 산출하게 되었다. 다른 두 복잡외삽법(선형추세모형법과 지수곡선적합법)이 상대적으로 합리적인 추계값을 산출한 것은 21년 간의 과거 인구에 대한 평균 경향선을 적합하였기 때문이다. 결론적으로, 매우 빠르게 감소하거나 성장하는 지역에 대해서는 다항곡선적합법이나 로지스틱곡선적합법은 적절하지 않은 것으로 보인다.

표 4. 서울시 구별 인구 추계 결과 비교

추계 기법	서울시 노원구			서울시 송파구		
	2030년	2040년	2047년	2030년	2040년	2047년
단순외삽법						
선형	419,945	316,853	244,689	683,276	698,592	709,313
기하	433,403	359,129	314,849	683,543	699,491	710,874
지수	433,403	359,129	314,849	683,543	699,491	710,874
복잡외삽법						
선형추세	467,013	410,989	371,772	689,660	711,360	726,551
다항곡선	389,426	204,089	43,577	687,260	704,960	716,398
지수곡선	475,522	432,323	404,441	690,914	714,657	731,760
로지스틱	357,401	174,679	89,376	683,738	696,336	703,559
ARIMA	424,357	325,677	256,601	656,521	654,893	654,846
비율외삽법						
고정-할당	495,287	472,344	450,076	632,521	603,221	574,784
변이-할당	462,070	408,988	368,577	654,362	644,879	628,370
성장-할당	435,584	363,279	293,103	671,777	674,932	677,995
추계 편차						
규모	137,886	297,665	406,499	58,393	111,435	156,977
비율(%)	38.58	170.41	932.83	9.23	18.47	27.31

셋째, 서울시 송파구의 경우는 상대적으로 비율외삽법이 다른 추계 기법과는 다른 결과를 보여주고 있다. 특히 고정-할당법과 변이-할당법이 가장 작은 추계값을 산출하고 있는데, 이 두 기법이 상대적으로 빠른 인구 성장을 보여주고 있는 송파구의 상황을 잘 반영하지 못하기 때문이다. 예를 들어, 고정-할당법은 오로지 2020년의 서울시 전체 인구 중 송파구가 차지하고 있는 비중만큼을 미래의 모든 시점에 그릴 것이라고 가정함으로써, 송파구가 2000~2020년에 가장 빠른 인구 성장을 보여주었다는 점을 전혀 반영하지 못하는 방법론적 한계를 지니고 있음을 보여주는 것이다. 또한 변이-할당법의 경우도, 2000년과 2020년 두 시점 사이에서의 송파구의 비중 변화는 송파구의 인구 성장 추이를 제대로 반영하지 못하기 때문에 일종의 과소평가가 발생하는 것으로 보인다. 결론적으로, 상위 공간단위의 인구는 감소하는 경우 점진적으로 증가하는 하위지역의 인구 추계 기법으로서 비율외삽법은 상대적으로 열등한 것으로 평가된다.

넷째, 11개 추계 기법 중 8개는 서울시 송파구의 인구 증가 경향을 추정했고, 나머지 3개(ARIMA 기

법, 고정-할당법, 변이-할당법)는 인구 감소 경향을 추정하였다. 두 가지 비율외삽법은 앞에서 언급한 과소추정의 단점이 있어 제외하면 ARIMA 기법만 남는다. 송파구의 2020년 인구가 2000년의 인구보다 크기 때문에 두 시점 인구의 비교에만 전적으로 의존하는 단순외삽법에 속하는 세 기법이 모두 증가 경향을 추정한 것은 너무나 당연한 것이다. 또한 2000~2020년을 전체적으로 볼 때 증가하는 경향이 있다고 보는 것이 타당하기 때문에 복잡외삽법에 속하는 대부분의 기법이 증가 경향을 추정한 것 역시 매우 당연한 것이다. 그러나 실질적으로 송파구 인구의 변동 추이는 매우 복잡하며, 이러한 복잡한 과거 패턴으로부터 미래의 경향성을 추정하는 것은 쉬운 일이 아니다. 이런 상황 속에서 ARIMA 기법이 인구 감소를 추정했다는 사실은 주목할 필요가 있다.

이를 좀 더 자세히 살펴보고자 한다. ARIMA 기법은 기본적으로 시간적 자기상관을 반영하는 기법이다. 따라서 가까운 시점에 발생한 사건, 즉 2019~2020년의 인구 감소를 기본적으로 반영할 것이고, 전체적으로 송파구가 보여주었던 일종의 반복성도 반영할 것이다. 결과적으로 이 기법은 향후 대

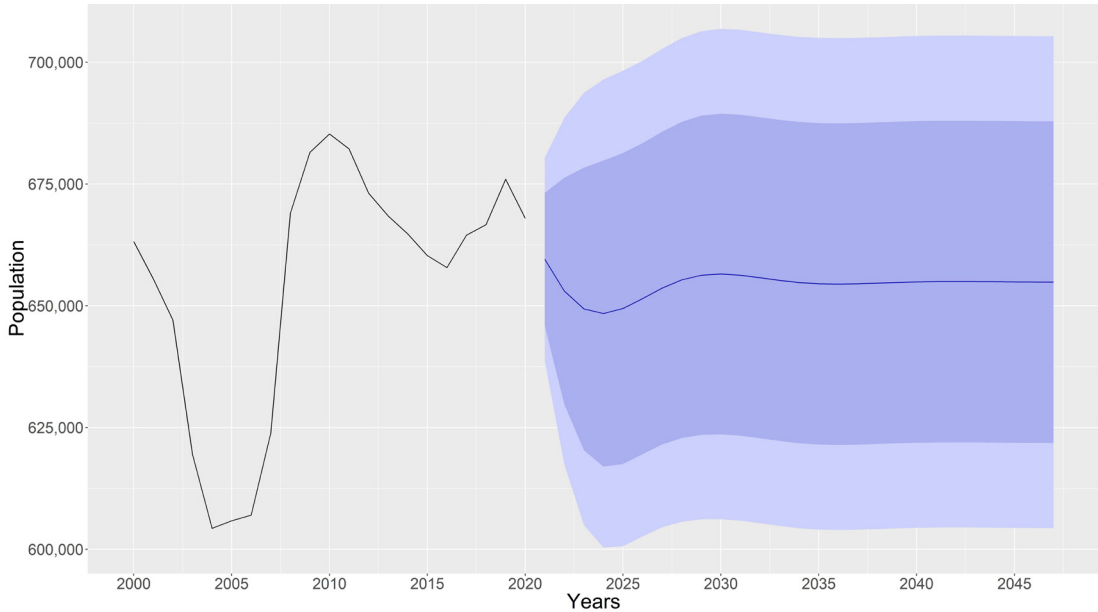


그림 3. ARIMA 모형법에 의한 서울시 송파구 인구 추정

약 5년간의 감소(2024년경에 최저점에 이름), 뒤 이
은 5년간의 증가, 그 이후의 평탄하지만 미세한 감
소 경향을 추정하고 있다(그림 3 참조). 결국 2030년
인구는 감소 이후 회복하여 정점에 이른 인구 수를,
2040년은 2030년 이후 감소하던 인구가 2036년에
저점을 통과하고 미세하게 상승하는 경향을 반영하
고 있으며, 2047년은 2042년 이후 미세하게 감소하
는 경향을 반영하고 있다. 결론적으로 과거 인구 성
장의 패턴이 복잡한 경우 ARIMA 모형의 장점이 가
장 잘 드러나는 것으로 해석된다.

III. 결론

본 연구의 주된 목적은 우리나라 소지역의 인구
추계를 위한 방법론으로서 추세외삽법이 가지는 가
능성과 한계를 살펴보는 것이었다. 추세외삽법은 크
게 단순외삽법, 복잡외삽법, 비율외삽법으로 범주화
되며, 모두 11개의 기법으로 세분화된다(Smith *et al.*,
2013). 단순외삽법은 과거의 두 시점에 대한 정보만
을 참조하여 추계를 실행하는 기법으로, 미래 경향
성의 형태를 어떻게 정식화하느냐에 따라 선형외삽
법, 기하외삽법, 지수외삽법으로 세분화된다. 복잡외

삽법은 단순외삽법보다 더 많은 과거 정보를 활용
하여 추계를 실행하는 기법으로, 미래 경향성을 어
떤 회귀식으로 정식화하느냐에 따라 선형추세모형
법, 다항곡선적합법, 지수곡선적합법, 로지스틱곡선
적합법, 그리고 ARIMA 모형법으로 세분된다. 비율
외삽법은 독립적으로 주어지는 상위 공간단위에 대
한 미래 추계값을 바탕으로, 과거의 하나 혹은 두 시
점에 대한 정보가 보여주는 비율 관계에 기반하여
추계를 실행하는 기법으로, 어떤 비율 관계를 활용
하느냐에 따라 고정-할당법, 변이-할당법, 그리고 성
장-할당법으로 세분된다.

11개 세부 기법의 실행성을 평가하기 위해 서울시
노원구와 송파구에 대한 미래 인구 추계를 실행하였
다. 중요한 결과는 다음과 같다. 첫째, 미래로 갈수
록 추계 기법 간 편차가 증가한다. 미래로 갈수록 인
구 추계의 불확실성이 높아지고 이것이 추계 기법의
결과간 차이를 확대하는 것으로 해석된다. 둘째, 매
우 빠르게 인구가 증가하거나 감소하는 지역의 경우
다항곡선적합법이나 로지스틱곡선적합법은 상대적
으로 그 적절성이 떨어진다. 셋째, 상위 공간단위의
인구 증감과 그 방향이나 정도에서 다른 특성을 보
이는 하위지역의 경우는 비율외삽법에 속하는 기법

의 적절성이 상대적으로 낮다. 넷째, 과거 인구 변화 패턴이 복잡한 지역의 경우, ARIMA 모형의 장점이 가장 잘 드러난다.

본 연구에서 다루어진 추세외삽법은 적은 양의 과거 데이터를 필요로 하고, 추세 논리가 간명하며, 적용이 손쉽다는 특성으로 인해 소지역 인구 추계를 위한 기법으로서 큰 가능성을 지니는 것으로 평가할 수 있다. 그러나 그러한 특성은 동시에 추세외삽법의 치명적인 한계를 규정하기도 한다. 우선, 어떤 '과거'를 참조했느냐에 따라 추세 결과가 완전히 달라진다. 서울시 송파구의 경우 기본 기간을 2000~2020년 혹은 2015~2020년이 아니라 2010~2020년을 두었다면 대부분 추세 기법은 인구 감소 경향을 보여 주었을 것이다. 그러나 추세외삽법의 보다 치명적인 한계는 인구 변동의 구성 요소(출산력, 사망력, 인구 이동)를 정교하게 다루지 못하고, 총량에 대한 직접적인 추정에만 집중한다는 것이다. 개별 구성 요소 별로 정교한 추산 기법이 개발되어 있으며(우해봉 등, 2016 참조), 그것의 합산으로서의 인구 총량 추세 개념이 훨씬 더 진보된 인구 추세 방법론이다.

소지역 인구 추계를 위한 추세외삽법의 가능성을 극대화하기 위해서는 다음의 두 가지가 반드시 고려되어야 할 것으로 보인다. 첫째, 추세값의 활용에 있어서 일종의 절충적 관점을 채택할 필요가 있다. 즉, 특정한 인구 추세 기법에 전적으로 기댈 것이 아니라 다양한 추세 기법을 적용하고 그 결과를 바탕으로 이차적인 처치를 하는 것이다. 특히 본 연구에서 다루어진 추세외삽법의 경우 상대적으로 실행이 간단한 장점이 있기 때문에 11개의 모든 기법을 적용해 11개의 서로 다른 추세값을 얻고, 그 11개의 추세값으로부터 최종적인 추세값을 추출하는 방식이 합리적인 대안일 수 있다. 예를 들어 11개 중 최대값 몇 개와 최소값 몇 개를 제외한, 중간에 위치한 추세값의 평균을 구하는 방식이 고려될 수 있다(Smith *et al.*, 2013:209).

둘째, 추세외삽법에 의한 소지역 인구 추세값은, 독립적인 차상위 공간단위의 추세값을 통한 통계과정(controlling)을 반드시 거쳐야 한다(Smith *et al.*, 2013:258). 일반적으로 국가 전체나 넓은 지역에 대한 인구 추계의 결과는 과거 인구 변동 패턴이 보여

주는 안정성이나 적용되는 추세 기법(예, 코호트-요인법)의 정교성 등으로 인해 타당성이나 신뢰성이 높은 경향이 있다. 그러므로 개별 소지역에 대한 인구 추세 결과를 그 지역이 포함된 상위 공간단위 전체에 대한 독립적인 추세 결과를 준거로 삼아 조정하는 과정을 거치는 것은 매우 합리적인 대안이 아닐 수 없다. 이것은 단순히 하위 공간단위의 인구 추세값의 합을 독립적으로 주어지는 차상위 공간단위 추세값과 일치시키는 정도를 넘어서 소지역 인구 추계의 타당성을 고양시키는 다양한 노력의 일환으로 이루어져야 한다. 다양한 통계 기법이 존재하며(Smith *et al.*, 2013:258-272), 그 기법들이 구체적으로 어떻게 작동하며, 우리나라 소지역 인구 추계에 어떻게 사용될 수 있는냐는 매우 중요한 사안이며 다른 연구 논문의 주제로 남겨두고자 한다.

참고문헌

- 박기영, 2019, 「R을 이용한 통계데이터분석」, 청람.
- 김정희, 2008, “고령화시대의 수도권 인구구성 변화 예측,” 한국도시지리학회지, 11(1), 31-42.
- 김계국·안수민·박성구·좌승희, 2006, “도시기본계획 인구지표의 설정실태와 관리방안에 관한 연구: 경기도 시군 도시기본 계획을 중심으로,” 경기연구원 기본연구, 3-6.
- 김태현·김동희·정구현, 2006, “코호트 요인법을 이용한 시군구별 장래인구추계,” 통계연구, 11(2), 1-39.
- 우해봉·양지윤·조성호·안형석, 2016, 「인구추계 방법론의 현황과 평가」, 한국보건사회연구원.
- 이상림·조영태, 2005, “H-P 기법을 이용한 기초자치단체의 장래인구추계,” 한국인구학, 28(1), 149-172.
- 이상복, 2007, “Population projections for local governments in Korea: Based on Hamilton-Perry & auto regression,” 한국데이터정보과학회지, 18(4), 955-961.
- 이상일, 2006, “고용 변동에 따른 인구 규모와 산업 구조 상의 장기적 변화 추정: 2020년 서울을 사례로,” 지리교육논집, 50, 27-39.
- 이상일, 2014, 「시나리오 기반의 시군구 단위 인구추계 모형 개발」, 국토연구원 최종보고서.
- 이상일·조대현, 2012, “지역간 인구이동의 예측을 통한 우리나라 시도별 장래 인구 추세: 다지역 코호트-요인법의 적용,” 대한지리학회지, 47(1), 98-120.
- 전광희·이상일·박민규·계봉호, 2014, 「등록센서스 기반 주요 인구통계의 일관성과 시의성 확보방안 연구」, 통계청.
- 조대현·이상일, 2011, “이지역 코호트-요인법을 이용한 부산광역시 장래 인구 추계,” 대한지리학회지, 46(2), 212-

- 232).
- 통계청, 2017, 보도자료: 시군구 장래인구추계 프로그램 개편 및 지자체 보급.
- 통계청, 2019a, 보도자료: 장래인구특별추계(시도편): 2017~2047.
- 통계청, 2019b, 「장래인구추계」 통계정보보고서.
- Baker, J., Alcántara, A., Ruan, X., Wakins, K., and Vasan, S., 2013, A comparative evaluation of error and bias in census tract-level age/sex-specific population estimates: Component I (net-migration) vs Component III (Hamilton-Perry), *Population Research and Policy Review*, 32(6), 919-942.
- Baker, J., Alcántara, A., Ruan, X., Watkins, K., and Vasan, S., 2014, Spatial weighting improves accuracy in small-area demographic forecasts of urban census tract populations, *Journal of Population Research*, 31(4), 345-359.
- Baker, J., Swanson, D., and Tayman, J., 2020, The accuracy of Hamilton-Perry population projections for census tracts in the United States, *Population Research and Policy Review*, 1-14.
- Davis, H.C., 1995, *Demographic Projection Techniques for Regions and Smaller Areas: A Primer*, Vancouver, BC: UBC Press.
- Hauer, M.E., 2019, Population projections for US counties by age, sex, and race controlled to shared socioeconomic pathway, *Scientific Data*, 6, 190005.
- Hyndman, R. and Athanasopoulos, G., 2018, *Forecasting: Principles and Practice*, 2nd Edition, Otexts.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., and Yasmeeen, F., 2020, *forecast: Forecasting Functions for Time Series and Linear Models*, R package version 8.13, <https://pkg.robjhyndman.com/forecast/>.
- Inoue, T., 2017, A new method for estimating small area demographics and its application to long-term population projection, in Swanson, ed., *The Frontiers of Applied Demography*, Cham: Springer, 473-489.
- Keyfitz, N., 1968, *An Introduction to the Mathematics of Population*, Reading: Addison Wesley.
- Klosterman, R.E., 1990, *Community Analysis and Planning Techniques*, Savage: Rowman & Littlefield Publishers.
- Rayer, S. and Smith, S.K., 2014, Population projections by age for Florida and its counties: Assessing accuracy and the impact of adjustments, *Population Research and Policy Review*, 33(5), 747-770.
- Ritz, C., Baty, F., Streibig, J.C., and Gerhard, D., 2015, Dose-Response Analysis Using R, *PLOS ONE*, 10(12), e0146021.
- Smith, S.K. and Rayer, S., 2012, Projections of Florida Population by County, 2011-2040, *Florida Population Studies*, Bulletin 162, Gainesville, FL: Bureau of Economic and Business Research, University of Florida.
- Smith, S.K., Tayman, J., and Swanson, D.A., 2013, *A Practitioner's Guide to State and Local Population Projections*, Dordrecht: Springer.
- Swanson, D.A., Scholttmann, A., and Schmidt, B., 2010, Forecasting the population of census tracts by age and sex: An example of the Hamilton-Perry method in action, *Population Research and Policy Review*, 29(1), 47-63.
- Swanson, D.A. and Tayman, J., 2012, *Subnational Population Estimates*, New York: Springer.
- Swanson, D.A. and Tayman, J., 2014, Measuring uncertainty in population forecasts: A new approach employing the Hamilton-Perry method, in *Annual Meeting of British Society for Population Studies*, Winchester, UK, 8-10.
- Wilson, T., 2011, Modelling with NEWDSS: Producing state, regional and local area population projections for New South Wales, in Stillwell, J. and Clarke, M. eds., *Population Dynamics and Projection Methods*, Dordrecht: Springer, 61-97.
- Wilson, T., 2015, New evaluations of simple models for small area population forecasts, *Population, Space and Place*, 21(4), 335-353.
- KOSIS 홈페이지, <https://kosis.kr/>

교신: 이상일

08826, 서울시 관악구 관악로 1,
서울대학교 사범대학 지리교육과
(이메일: si_lee@snu.ac.kr)

Correspondence: Sang-Il Lee

08826, 1 Gwanak-ro, Gwanak-gu, Seoul, Korea,
Department of Geography Education,
Seoul National University (Email: si_lee@snu.ac.kr)

