

위계선형모형을 이용한 인구이동 흐름 분석

An Analysis of the Migration Flow Using Hierarchical Linear Model

이희연 서울대학교 환경대학원 환경계획학과 교수(제1저자)
Lee Heeyeon Professor, Dept. of Environmental Planning,
Graduate School of Environmental Studies,
Seoul National Univ.(Primary Author)
(leehyn@snu.ac.kr)

노승철 서울대학교 환경대학원 환경계획학과 박사과정
Noh Seungchul Graduate Student, Dept. of Environmental Planning,
Graduate School of Environmental Studies,
Seoul National Univ.
(goonise@snu.ac.kr)

목 차

I. 서론

1. 연구배경 및 목적
2. 연구방법 및 자료

II. 인구이동의 현황과 특징

1. 개인 속성에 따른 인구이동 특징
2. 인구이동의 공간적 특징

III. 인구이동 모형 구축과 변수 선정

1. 인구이동 흐름 분석을 위한 모형의 선정
2. 위계선형모형의 개념과 특징
3. 위계선형모형의 수준별 분석단위 및 변수 선정

IV. 인구이동 모형의 추정결과

1. 위계선형모형의 수준별 분산 비교
2. 위계선형모형의 추정결과 및 풀이

V. 결론

※ 본 논문은 2010년 국토연구원에서 수행한 “국토정책 시뮬레이션 모형의 구축 연구”의 협동연구로 수행되었음.

I. 서론

1. 연구배경 및 목적

인구이동은 개개인이나 가구가 좀 더 나은 주거환경으로 삶의 보금자리를 옮겨 생활의 질을 개선하고자 하는 기본 욕구에서 시작되지만, 그 결과는 전입지와 전출지의 인구, 사회, 경제적인 측면에 상당한 영향을 미치게 된다(Greenwood, 1985; Shaw, 1975). 인구이동으로 인한 전출지와 전입지의 인구수 및 인구구조의 변화는 주택, 경제, 토지이용에 직접적인 영향을 주지만, 더 나아가 통행유발과 교통흐름에도 간접적인 영향을 미치면서 도시 공간구조에 변화를 가져오기도 한다.

우리나라는 지난 50년 동안 괄목할 만한 경제성장과 더불어 도시화가 급속도로 진전되었으며 그에 따라 인구이동이 상당히 활발하게 이루어져 왔다. 1960년대에서 1980년대 초반까지 이루어진 인구이동은 주로 농촌에서 도시를 향한 이촌향도가 주류를 이루었으나, 농촌의 인구가 지속적으로 감소하면서 1980년대 후반 이후 인구이동의 양상은 크게 달라지고 있다. 주로 도시들 간 인구이동과 비수도권에서 수도권으로 이동이 활발하게 이루어지는 가운데 1990년대에 접어들면서 수도권 내에서의 인구이동이 주도하고 있다(이희연, 2008). 우리나라 인구이동의 특징은 특정지역으로의 방향성을 가진 집중적 흐름으로 인해 주택문제와 교통혼잡 및 주택·토지가격의 급등과 같은 문제들을 야기시킨다는 점이다.

이렇게 인구이동은 도시 및 지역의 다차원적인 변화를 가져오는 가장 핵심요소임에도 불구하고 아직까지 인구이동의 흐름을 설명하고 더 나아가 예측하기 위한 정교한 모델이 아직까지 개발되어 있지 못하다. 전통적으로 인구이동에 관한 연구는

거시적 관점과 미시적 관점에서 각각 이루어져 왔으며, 그에 따라 적용된 모형과 변수들도 차별화되었다. 거시적 관점에서의 연구는 지역 간 인구이동에 영향을 미치는 결정요인을 분석하는 데 초점을 두고, 실업률, 소득, 취업기회 등과 같은 경제적 요인이나 교육, 주택, 문화시설 및 삶의 질과 같은 요인들을 통해 지역 간 인구이동을 설명하고 있다(Michalos, 1996). 한편 미시적 관점에서의 연구는 개개인의 이동행태나 성향에 대해 초점을 두고, 개개인이 특정한 목적지를 선택하는 의사결정과정들을 통해 인구이동을 설명하고 있다(Cadwallader, 1989).

우리나라의 인구이동 관련 연구들을 보면 인구학적 측면, 경제적 측면, 공간적 측면, 행태적 측면 등 다양한 관점에서 이루어지고 있으며, 분석방법론도 결정론적 접근방법과 확률론적 접근방법을 도입하여 분석하고 있다. 인구이동 요인을 설명하기 위해 회귀분석을 이용한 연구가 가장 많이 이루어지고 있는 가운데(김성태·장정호, 1997; 유경문, 1991; 이문식, 2004; 하상근, 2005), 회귀분석에서는 설명변수들 간의 상호 연관관계 또는 환류효과를 나타내기 어렵다는 점을 고려한 다양한 접근방법도 시도되었다. 즉, 연립방정식을 이용한 연구(김갑성, 2006)나 경로분석(김미희, 1983; 이희연·박정호, 2009)을 이용한 방법도 시도되었다. 또한 이동요인을 밝히기 위해 판별분석(김정호, 1990), 요인분석(정성기·이성근·고덕균, 2008), 그랜저 인과관계(오정일·안기돈, 2007) 방법들도 활용되었다. 한편 선택확률모형을 이용하여 이동자들의 이동성향을 분석하거나(곽인숙·김순미, 1996; 신현곤, 1998), 개인 또는 가구의 목적지 선택확률에 영향을 미치는 요인을 분석한 연구도 이루어졌다(이성우, 2001; 권상철, 2005; 최유미·남진, 2008).

이상에서 살펴본 바와 같이 우리나라의 인구이동 현상을 설명하기 위한 많은 연구들이 수행되어 왔지만, 아직까지 인구이동의 흐름을 설명하고 더 나아가 예측할 수 있는 모형에 대한 연구는 다소 미미한 편이다. 이러한 배경하에서 본 논문은 인구이동의 흐름을 설명하는 모형을 구축하는 데 목적을 두었다. 특히 도시 간 인구이동의 흐름을 설명해주는 모형 구축을 통해 향후 인구이동의 흐름을 예측하는 데 활용될 수 있도록 하기 위해 궁극적인 목적을 두었다.

2. 연구방법 및 자료

본 논문에서는 인구이동의 현황과 특징을 분석한 결과를 토대로 하여 인구이동의 흐름을 설명할 수 있는 모형을 구축하고자 하였다. 먼저 최근 10년간(2000~2009년) 도시 간에 이루어진 인구이동 흐름을 분석한 결과 각 도시별 전출량 또는 전입량의 시계열적 상관계수는 0.95 이상으로 매우 높게 나타났다(<부표 1> 참조). 즉, 우리나라의 경우 각 도시를 떠나는 전출자수와 해당 도시로 들어오는 전입자수는 매년 거의 유사함을 말해주며, 따라서 특정 시점의 한 연도를 선택하여 인구이동 흐름을 분석하여도 거의 일반화된 결과를 도출할 수 있음을 시사해준다.

뿐만 아니라 어느 특정 연도를 선택하여 시·군·구 간 전입량과 전출량 간의 관계를 분석한 결과도 거의 모두 상관계수가 0.99로 나타났다(부표 2> 참조). 따라서 우리나라의 경우 연도별 차이가 없으며, 전입자가 많은 지역이 전출자도 많고,

인구이동 모형을 구축할 경우 전입 모델과 전출 모델을 별도로 구축할 필요성이 없음을 시사해준다.

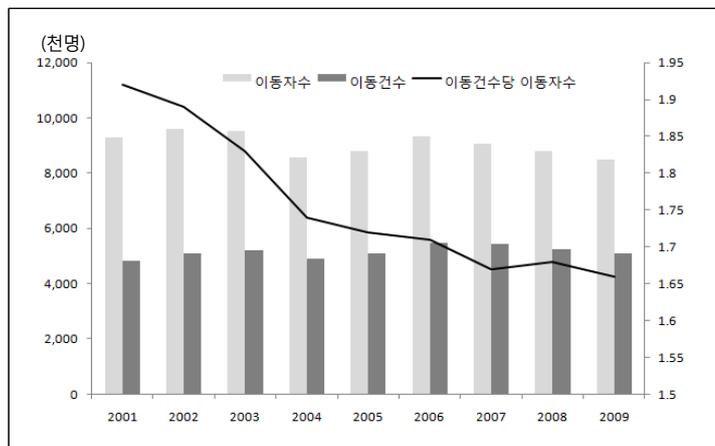
이와 같은 분석결과를 토대로 하여 본 논문에서는 2008년도 시점을 대상으로 하여 인구이동 흐름을 분석하였다. 사용한 자료는 2008년도 인구이동 통계 원시자료이며, 총 1,187만 4,714명의 이동자가운데 20세 이상의 개인 이동자를 대상으로 분석하였다. 본 논문에서는 위계선형모형을 이용하였으며 사용한 소프트웨어는 HLM 6.0이다.

II. 인구이동의 현황과 특징

1. 개인 속성에 따른 인구이동 특징

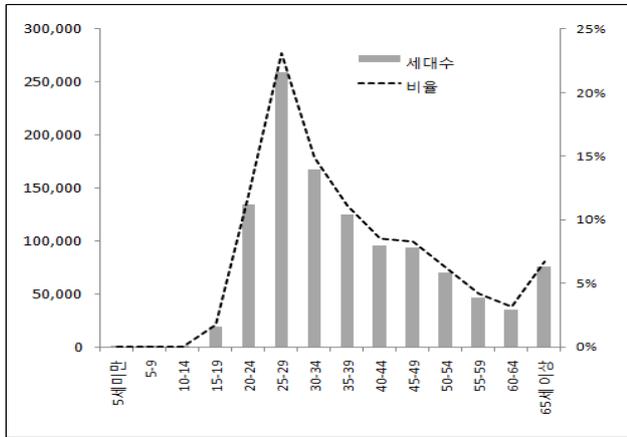
2000년대의 우리나라 인구이동 현황을 보면 이동자수는 2001년 929만 명에서 2009년에는 848.7만 명으로 매년 약간씩 줄어들고 있다. 이에 따라 인구이동률도 19.3%에서 17.1%로 감소되었지만 여전히 우리나라의 인구이동률은 선진국가들에 비하면 높은 편이다(Week, 2002). 그러나 이동건수를 보면 2001년 483.8만 건에서 2009년 509.9만 건

그림 1 _ 이동건수당 이동자수의 변화(2001~2009)



자료: 통계청, 각 연도, 국내 인구이동통계.

그림 2_ 연령층별 1인 세대 이동 비율(2008년 기준)



자료: 통계청. 2008. 국내 인구이동통계.

으로 오히려 증가하고 있다. 이에 따라 이동건수당 이동자수는 2001년 1.92명에서 2009년 1.66명으로 낮아졌다(<그림 1> 참조). 이는 점차 1인 이동자

수가 많아지고 있음을 시사해준다. 이동건수당 이동자수를 시·군 내부이동과 시·군 간 이동으로 구분하여 보면 2008년의 경우 지역 내 이동은 2.00명인데 비해 지역 외부로의 이동은 1.47명으로 나타나고 있어 지역 내 이동보다 지역 간 이동 시에 이동건수당 이동자수가 훨씬 적게 나타나고 있다.

이러한 경향은 인구이동 유형별로 1인 이동이 차지하는 비율을 비교해보면 더 잘 나타나고 있다. 2008년도 인구이동을 보면 1인 이동 비율이 전체의 75%를 차지하고 있다. 그러나 이 비율을 인구이동 흐름별로 보면 매우 대조적인 양상을 보이고 있다. 비수도권 → 수도권으로 이동하는 경우 1인 이동비율이 83.1%로 가장 높게 나타나는 반면

표 1_ 이동유형별 1인 인구이동 비율 비교(2008년 기준)

전출	전입	1인 이동건수	건수당 이동자수	1인 이동 비율(%)
수도권	수도권	664,870	1.62	68.4
	비수도권	278,606	1.36	80.1
	소계	943,476	1.55	71.5
비수도권	수도권	333,712	1.30	83.1
	비수도권	667,727	1.44	76.6
	소계	1,001,439	1.39	78.7
시	시	1,469,156	1.50	73.4
	군	214,903	1.38	78.3
	소계	1,684,059	1.49	74.4
군	시	232,539	1.36	79.3
	군	28,317	1.43	76.7
	소계	260,856	1.37	79.0
합계		1,944,915	1.47	75.0

자료: 통계청. 2008. 국내 인구이동통계.

에 수도권 내에서 이동하는 경우 1인 이동비율은 68.4%로 가장 낮게 나타나고 있다. 이는 수도권으로 진입하는 경우 가족이 함께 이동하는 것이 상당히 어려움을 시사해준다.

1인 이동에는 개인이동과 세대이동이 다 포함되므로, 본 논문에서는 전입지 기준으로 1인 세대 이동만을 추출하여 연령층별로 비교하였다. 여기서 말하는 1인 세대란 이동자수가 1명이면서 전입구분에서 세대 구성인 경우를 말한다. 2008년도의 경우 전입하면서 세대를 구성하는 1인 세대의 비율이 62.1%로 나타났으며, 다른 세대로의 편입은 약 38%로 나타났다. 1인 세대 이동의 연령층별 점유율을 보면 25~29세 연령층이 전체 1인 세대 이동자의 23.2%로 가장 높게 나타나고 있으며, 그 뒤를 이어 30~34세 연령층이 14.9%, 20~24세 연령층이 12%를 차지하고 있다. 그러나 65세 이상의 연령층이 차지하는 비율도 6.7%로 상당히 높은 편이다(<그림 2> 참조).

특히 1인 세대 이동은 지역 내 이동에 비해 지역 간 이동에서 더 두드러지게 나타나고 있어 장거리 이동 시 젊은 연령층이 1인 세대를 구성하여

이동하고 있음을 말해준다.

2. 인구이동의 공간적 특성

2008년도 우리나라의 인구이동의 흐름을 보면 시↔시 간의 이동이 전체 이동량의 78.2%를 차지하는 것으로 나타났는데, 이는 1995년도의 67%, 2000년의 72.9%에 비하면 상당히 높아진 것이다. 이는 인구이동의 후기에는 도시 간의 이동이 주류를 이루게 된다는 Zelinski(1971)의 인구이동 변천 모델과도 일치한다고 볼 수 있다. 한편 시↔군 이동은 1995년에는 전체 이동의 약 30.5%를 차지하였으나 2005년 이후 전체 이동에서 차지하는 비율은 20.4%로 다소 감소하였다. 이 가운데 군→시로의 인구이동 비율은 약 10% 수준에 머물고 있어 농촌에서 도시로의 이동도 많이 약화되고 있음을 말할 수 있다. 한편 군↔군 간의 이동은 1.4%에 불과하여 더 이상 농촌에서의 인구이동은 거의 일어나고 있지 않음을 말해준다(<표 2> 참조). 따라서 우리나라의 인구이동은 도시 간의 이동이 주도하는 가운데 도시에서 인근 주변의 농촌으로, 또는 인근

표 2_ 시-군 간 인구이동 흐름 변화(1995~2008년)

전출지	전입지	1995년	2000년	2005년	2008년
시	시	2,490 (67.0)	2,772 (72.9)	3,020 (78.9)	2,990 (78.2)
	군	560 (15.1)	456 (12.0)	357 (9.3)	379 (9.9)
군	시	574 (15.4)	508 (13.4)	403 (10.5)	400 (10.5)
	군	91 (2.5)	68 (1.8)	50 (1.3)	53 (1.4)
합계		3,715 (100.0)	3,804 (100.0)	3,830 (100.0)	3,822 (100.0)

자료: 통계청, 각 연도, 국내 인구이동통계.

표 3_ 집중형 인구이동 흐름을 보이는 상위 10개 도시의 특징

전출지 (출발지)	상대적 엔트로피	최상위 전입지가 차지하는 비중(%)	전입비중으로 본 상위 5개 도시(목적지)						누적 80%에 속하는 도시의 수
			전입 비중	1위	2위	3위	4위	5위	
구리시	2.68	35.1	73.7	남양주	서울	인천	의정부	용인	11
나주시	2.66	54.7	70.3	광주광역시	서울	목포	영암군	인천	17
양산시	2.46	45.7	70.1	부산	울산	서울	김해	대구	14
경산시	2.65	54.9	69.9	대구	서울	영천	부산	울산	12
김포시	2.34	31.5	69.0	서울	인천	고양	부천	안산	15
양주시	2.19	24.7	67.3	의정부	서울	동두천	포천	고양	14
하남시	2.46	49.9	67.0	서울	광주	남양주	성남	인천	16
경주시	2.01	21.5	67.0	울산	포항	대구	서울	부산	17
의정부시	2.27	41.4	65.2	서울	양주	포천	동두천	남양주	16
광명시	2.31	43.4	64.2	서울	인천	시흥	부천	안산	13

자료: 통계청, 2008. 국내 인구이동통계.

주변의 농촌에서 도시로 이동하는 근거리 이동이 이루어지고 있다고 풀이할 수 있다.

본 논문에서는 우리나라 도시별로 인구이동의 전출·입 패턴과 인구이동 흐름의 집중화 수준을 분석하였다. 지니계수와 상대적 엔트로피 지수¹⁾를 산출해본 결과 우리나라 도시들의 인구이동 흐름은 매우 집중화된 패턴으로 나타났다. 즉, 출발지를 떠나 164개 시·군 가운데 전입하는 목적지 수는 가장 작은 경우 11개 도시로 나타났으며, 가장 많은 경우에도 41개 도시로 집중하여 전입하는 패턴을 보이고 있다.

다른 도시들에 비해 보다 더 집중형의 인구이동 흐름을 보이고 있는 상위 10위 도시들의 특징을 보면 상위 5개 전입지가 차지하는 비중이 전체 전입자의 66% 이상을 차지하고 각 도시로 전입해오는

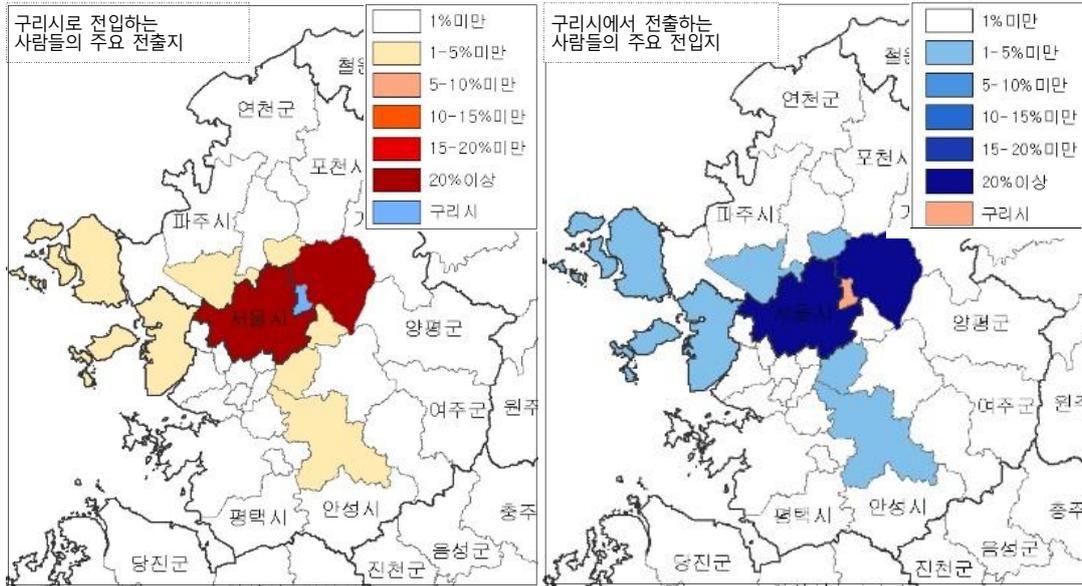
누적 이동자의 80%를 차지하는 전입지수가 20개 미만이며, 상대적 엔트로피 지수도 2 이상으로 나타났다(<표 3> 참조). 이러한 집중형 패턴을 보이는 도시들의 인구이동 흐름을 보면 서울시로 집중하거나 광역시 또는 도내 수위 도시 및 인근 도시로 집중하는 현상이 두드러지게 나타나고 있다.

개별 도시별로 인구이동의 흐름을 분석한 결과 주로 광역시·도 내에서 인근 도시로 이동하는 비중이 가장 높게 나타났으며, 사·도를 넘어선 이동의 경우 수도권으로 이동하거나 광역시로 이동하는 경우가 가장 보편적으로 나타나고 있었다. 일례로 가장 집중형 인구이동의 흐름을 보이고 있는 구리시의 전출 패턴과 전입 패턴을 보면 거의 유사하게 나타나고 있다. 164개 시·군 가운데 구리시와 전출입하고 있는 도시의 수는 11개로 나타났으며,

1) 엔트로피 지수를 산출하여 비교하는 경우 전입도시의 수에 따라 지수값이 달라지므로 집중도를 비교하기 어려움. 따라서 최대 엔트로피값(Hmax)에서 엔트로피값(H)을 뺀 상대적 엔트로피 지수(I)를 산출하여 비교하였음.

$$I = H_{\max} - H, \text{ 여기서 } H_{\max} = -\sum_{i=1}^n \frac{1}{n} \ln \frac{1}{n} = \ln n, \quad H = -\sum_{i=1}^n P_i \ln P_i$$

그림 3_ 구리시의 전출입 패턴



부산을 제외한 나머지 도시들은 모두 수도권 내 도시들로, 특히 구리시와 인접한 서울과 남양주시로 집중하여 전출·입하고 있다(<그림 3> 참조).

한국교통연구원에서 제공하는 각 도시 간 시간거리를 이용하여 다음과 같은 중력모형을 구축하여 추정하였다.

III. 인구이동 모형 구축과 변수 선정

$$M_{ij} = k \frac{P_i^\alpha \cdot P_j^\beta}{d_{ij}^\gamma}$$

$$\ln(M_{ij}) = \alpha \ln P_i + \beta \ln P_j + \gamma \ln(d_{ij}) + \ln k$$

1. 인구이동 흐름 분석을 위한 모형의 선정

앞에서 인구이동의 특징을 살펴본 결과 도시 간 이동이 전체의 약 80%를 차지하고 있었다. 이에 따라 본 논문에서는 도시 간에 이루어지고 있는 인구이동 흐름에 초점을 두었으며, 제주시와 서귀포시의 경우 이동을 제약하는 거리에 대한 인지와 비용이 상당히 다르다는 점에서 제외시켜 82개 도시 간에 이루어지는 인구이동의 흐름을 설명하기 위한 모형을 구축하는 데 초점을 두었다.

인구이동의 흐름을 설명하는 모형을 구축하기에 앞서 본 논문에서는 2008년도 82개 도시 간에 이루어진 인구이동에 대한 O-D 행렬을 구축하고,

그 결과 전입지와 전출지 인구, 그리고 두 지역 간의 거리 변수(시간 거리)로 도시 간 인구이동 흐름의 약 84%가 설명되었다. 특히 전출지와 전입지의 계수를 비교해보면 전출지 인구 계수가 전입지 인구 계수보다 아주 미미하게 더 크게 나타나고 있으나, 거의 차이가 없다고 볼 수 있다. 따라서 인구이동의 흐름에서 전입지와 전출지의 영향력은 거의 같다고 풀이할 수 있다. 이와 같은 결과는 우리나라의 경우 각 도시별 전출량과 전입량 간의 상관계수가 0.99로 높게 나타난 결과를 통해서도 이해된다. 한편 거리조락계수도 -1.167로 나타나면서 거리 영향력 계수도 중요함을 말해준다(<표 4> 참조).

표 4_ 82개 도시 간 중력모델 추정결과

구분	회귀계수		t값	유의수준
	비표준화	표준화		
상수	-13.628		-82.511	.000
ln전입지 인구	.944	.547	111.403	.000
ln전출지 인구	.949	.550	112.100	.000
ln이동시간	-1.167	-.469	-95.594	.000

그러나 이렇게 추정된 회귀모형은 고정효과(fixed effect)만을 고려하기 때문에 인구이동의 흐름을 예측하고자 하는 경우 모든 도시에 동일한 회귀계수를 적용하게 된다. 즉, 회귀모형에서 추정된 전출지와 전입지의 인구수와 이동거리에 대한 회귀계수가 모든 도시에 동일하게 적용되기 때문에 전입지와 전출지의 특성을 충분히 반영하지 못하게 된다. 이는 더 나아가 우리나라 인구이동 흐름에서 두드러지게 나타나는 도내 이동이나 권역별 특성을 반영할 수도 없다. 뿐만 아니라 전입자가 많은 도시가 전출자도 많은 우리나라 도시 간 인구이동의 흐름을 설명하고자 하는 경우 인구이동 흐름에 영향을 주리라 예상되는 요인들(예: 일자리 기회, 주택 기회 등)을 모형에 투입하는 경우 변수 간에 높은 공선성으로 인해 모델의 추정결과를 신뢰할 수 없게 된다. 따라서 전통적인 회귀모형을 활용하여 우리나라 인구이동의 흐름을 설명하고 더 나아가 인구이동의 흐름을 예측하는 데 상당한 한계를 지니게 됨을 시사해준다.

이러한 문제점을 고려하여 본 논문에서는 위계를 나누어 분산을 분석하는 위계선형모형(Hierarchical Linear Model: HLM)을 선정하였다. 전통적 회귀모형은 표본들이 단일 수준에서 독립적이라는 가정하에서 분석이 이루어지는 반면에 위계선형모형은 표본들이 각기 다른 특성을 지닌 상위

수준에 속하는 위계구조를 지니고 있다는 전제하에서 분석이 이루어진다(강상진, 1998).

본 논문에서는 위계를 달리하는 위계선형모형에서 1수준은 82개 도시 간에 이루어지는 6,642개(82×82) 이

동흐름을 분석단위로 삼았다. 그러나 두 도시 간에 이동흐름이 없는 36개 이동흐름을 제외하여 총 6,606개 이동흐름이 분석단위이며, 종속변수는 각 도시로 유입되는 전입자수다. 즉, 1수준 모형은 6,606개 흐름에서의 전입자수 차이(종속변수의 분산)에 영향을 미치는 요인들을 추정하는 모형이다.

한편 위계선형모형의 2수준 분석단위는 82개 전입도시로 설정하였다. 2수준 모형은 다른 도시들로부터 해당 도시로 전입한 82개 집계된 흐름(aggregated flow)에서의 전입자수의 차이(종속변수의 분산)를 설명하는 것으로, 전입지의 특성을 설명하는 모형이라고 볼 수 있다. 이는 전입지의 특성이 1수준의 설명변수들에 미치는 영향이 다르며, 그 결과 특정도시로의 전입자수가 다를 것이라고 전제하였기 때문이다.

이와 같이 위계선형모형을 이용하는 경우 종속변수에 직접 영향을 미치는 1수준의 독립변수들과 2수준의 종속변수에 영향을 미치는 2수준의 독립변수들은 서로 다르며, 2수준에서 투입된 전입지의 지역특성 변수는 1수준의 상수항에 영향을 미치게 되는 모형을 구축할 수 있어 전입지와 전출지 간의 이동 흐름이 매우 높은 상관성을 지니고 있는 경우 단일 수준에서 분석하는 회귀모형을 활용하는 경우에 나타나는 문제점을 해결할 수 있다는 장점이 있다.

2. 위계선형모형의 개념과 특징

다층모형(multi-level model)이라고 불리기도 하는 위계선형모형은 하위 수준과 상위 수준 간의 분산(variance)을 분리하여 분석한다. 특히 위계선형모형을 사용함으로써 종속변수의 분산을 보다 더 많이 설명할 수 있으며, 각 수준에서 종속변수의 분산을 설명하는 기여 수준을 파악할 수 있다. 더 나아가 위계선형모형을 이용하는 경우 추정된 회귀계수들도 실제에 더 근접한 값을 제시하는 것으로 알려져 있다(강상진, 1995; Hofmann, 1997; Luke, 2004; Pollack, 1998). 전통적 회귀모형은 표본들이 한 수준에서 독립적이라는 가정하에서 분석이 이루어지는 반면 위계선형모형은 표본들이 각기 다른 특성을 지닌 상위 수준에 속하는 위계구조를 지니고 있다고 전제한다. 만약 위계구조를 지닌 자료를 한 수준으로만 분석하는 경우 미시적 분석결과가 거시적 해석에 대한 설명력을 결여하는 알커의 원자학적 오류(Alker's Atomistic Fallacy)나 또는 거시적 분석결과가 미시적 결과에 대한 설명력을 결여하는 로빈슨의 생태학적 오류(Robinson's Ecological Fallacy)를 범할 가능성이 높아진다(Raudenbush & Bryk, 2002). 또한 고정효과(fixed effect)만을 고려한 전통적인 회귀모형들은 변인 간의 인과관계에 초점을 두기 때문에 개개인의 특성, 그룹과 개인 간의 상호작용, 그리고 그룹의 특성을 동시에 고려하지 못하고 있으며, 이는 회귀계수의 표준오차를 실제보다 작게 추정하게 함으로써 통계적 유의성이 과대 평가될 위험을 지니고 있다(유정진, 2006). 위계선형모형의 경우 설명변수들이 포함되지 않고 상수항만을 갖는 무제약모형(unconditional model)과 각 수준별 설명변수들이 투입된 최종모형의 비교를 통해 분산의 변화를 명확하게 파악할 수

있다. 무제약모형은 상수항의 임의효과만을 가정한 모형으로 1수준의 분산과 2수준의 분산을 구분하고 비교할 수 있는 정보를 제공해준다. 즉, 무제약모형을 통해 도출된 두 수준의 분산이 통계적으로 유의하다면 각 수준이 종분산에 미치는 영향력을 비교할 수 있게 된다. 이러한 의미에서 무제약모형은 연구를 진행하며 단계별로 투입되는 독립변수의 설명력을 판단하는 기준을 제시해준다는 의미에서 기초모형이라고 불리기도 한다(강상진, 1995).

일반적으로 위계선형모형은 무제약모형과 상수항은 임의효과, 1수준 변수의 계수는 고정효과를 갖는 임의절편모형(random intercept model), 그리고 상수항과 1수준 변수의 계수가 모두 임의효과를 갖는 임의계수모형(random coefficient model)을 통해 각 수준별로 설명변수가 투입됨에 따라 줄어드는(설명되는) 분산의 양을 통해 수준별 요인의 설명력과 모델의 타당성을 검증하게 된다.

이러한 세 가지 위계선형모형을 수식으로 나타내면 <표 5>와 같다. 여기서 i 는 1수준 표본, j 는 2수준 집단을 나타내며, X_{ij} 는 j 집단에 속한 i 의 특성, W_j 는 집단 j 의 특성이다. 또한 β 는 1수준 독립변수의 추정계수, γ 는 2수준 독립변수의 추정계수이며 ϵ 과 v 는 각각 1수준과 2수준의 확률적 오차항을 나타낸다.

위계선형모형에서 2개의 수준을 갖는 가장 복잡한 임의계수모형의 일반적인 형태는 아래와 같다. 2수준 모형은 1수준 독립변수가 종속변수에 미

표 5_ 세 가지 위계선형모형의 기본 식

구분	1수준	2수준
무제약모형	$Y_{ij} = \beta_{0j} + \epsilon_{ij}$	$\beta_{0j} = \gamma_{00} + v_{0j}$
임의절편모형	$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + v_{0j}$
임의계수모형	$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + v_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + v_{1j}$

치는 영향, 2수준 독립변수가 종속변수에 미치는 영향, 그리고 1수준과 2수준의 상호작용 효과를 모두 포함하는 모형이다.

Level-1 Model: 미시 수준 모형

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + Y_{ij}$$

Level-2 Model: 거시 수준 모형

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + \mu_{1j}$$

Mixed Model: 통합 모형

$$Y_{ij} = \frac{\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}X_{ij}W_j}{\text{고정효과(Fixed Effect)}} + \frac{\mu_{0j} + \mu_{1j}X_{ij} + Y_{ij}}{\text{임의효과(Random Effect)}}$$

여기서 Y_{ij} 는 i 번째 그룹에 속해 있는 j 의 종속변수를 의미한다. β_{0j} 와 β_{1j} 는 상수항과 1수준 독립변수들의 추정계수(level-1 coefficient)들로 더 정확히 표현하면 β_{0j} 는 i 번째 그룹의 절편, β_{1j} 는 j 그룹의 X_{ij} 의 회귀계수다. γ_{00} , γ_{10} , γ_{01} , γ_{11} 는 2수준 독립변수들의 추정계수(level-2 coefficient)들로서 2수준 독립변수들이 1수준 회귀계수에 미치는 영향력을 나타낸다. 독립변수들 간의 관계를 보면 일반 선형회귀 모형과는 달리 1수준의 추정계수들은 1수준 모형에 의해 직접 추정되는 것이 아니라 2수준 추정계수들을 통해 추정된다. 2수준 모형의 상수항인 γ_{00} , γ_{10} 은 1수준 추정계수의 집단 간 평균이며 γ_{01} , γ_{11} 는 1수준 독립변수와 2수준 독립

변수 간 상호작용(cross-level interaction)효과를 보여준다. 즉, 집단 간 평균적인 효과에 2수준 단위 특성들의 영향 그리고 임의 효과인 μ_{0j} , μ_{1j} 가 합해져서 1수준 독립변수가 종속변수에 영향을 미치는 정도(기울기)가 추정되므로 2수준 집단별로 1수준의 추정계수가 달라지게 된다.

3. 위계선형모형의 수준별 분석단위 및 변수 선정

위계선형모형에서 1수준의 분석 단위는 82개 도시 간에 이루어지는 6,606개 이동 흐름이다. 따라서 1수준 모형은 6,606개 이동흐름에서 전입자수 차이(종속변수의 분산)를 설명하는 것으로, 이동흐름 자체의 특성을 설명하는 모형이라고 볼 수 있다.

1수준 모형에서 이동흐름 자체의 특성을 설명하기 위해 본 논문에서 선정한 변수들은 <표 6>과 같다. 첫 번째 독립변수로는 이동흐름에서 가장 대표

표 6_ 인구이동 모형에서 1수준의 독립변수

독립변수	설명	변수형태
지역 내 이동	전출입지가 같은 지역 내에 있는 경우 <지역구분> • 수도권: 서울특별시, 인천, 경기도 • 강원권: 강원도 • 충청권: 충청북도 • 충남권: 대전, 충청남도 • 전북권: 전라북도 • 전남권: 광주, 전라남도 • 경북권: 대구, 울산, 경상북도 • 경남권: 부산, 경상남도	더미 변수
인지 거리	$\frac{\text{도시 간 시간거리(자동차 이용)}}{\text{특정도시로의 전입자/해당도시에서의 총 전출자}}$	연속 변수 (자연 로그)
전출지 인구수	2007년 전출지의 총인구수	연속 변수 (자연 로그)

적으로 사용되는 변수인 전출지의 인구 규모다. 전출지의 인구가 많을수록 각 전입지로의 인구수가 많아질 것이기 때문이다. 두 번째 변수는 도시 간 이동 흐름을 설명하는 ‘거리’ 변수다. 일반적으로 인구이동은 도시 간 거리가 멀수록 이동량이 줄어드는 거리조락 효과를 가지고 있다. 그러나 물리적인 거리보다는 전출자들이 전입지에 대한 심리적 거리, 이른바 인지거리가 더 중요할 수 있다. 거리가 멀더라도 특정 도시에 대한 정보나 접촉이 많은 경우 보다 가깝게 인지될 수 있으며, 아무리 가까운 도시라고 하더라도 그 도시에 대한 접촉이 전혀 없는 경우 오히려 멀게 느껴질 수도 있다. 즉, 비교적 자신이 잘 알고 있는 도시에 대해서는 물리적 거리보다 가깝게 느끼게 되고, 가까이 있더라도 그 도시에 대한 정보가 부족하다면 사람들은 더 멀게 느끼기 때문에 이동을 꺼리게 된다. 이와 같은 점을 고려하여 본 논문에서는 ‘인지거리’ 개념을 도입하였다. 인지거리는 두 도시 간 자동차로 이동하는데 소요되는 시간 거리를 ‘인지도’로 나누어서 산출하였다. 여기서 인지도 변수는 해당도시에서 이미 특정 도시로 전출해 나간 사람들이 많을수록 특정 도시에 대한 인지도가 높을 것이라는 전제하에서 2007년도 인구이동 자료를 토대로 하여 해당도시로부터 전출해 나간 사람들이 어느 도시들로 전입하였는가를 고려하였다. 즉, 해당 도시에서 전출한 총전출자수에 대한 특정 도시로의 전입자수가 차지하는 비중으로 인지도 변수를 측정하였다. 따라서 해당 전출지에서 특정 도시로 많은 사람들이 전입한 도시의 경우 인지도가 높게 되므로 물리적 거

리보다 인지거리가 짧아지게 되며, 반대로 해당 전출지에서 특정 도시로의 전입한 비중이 낮아서 인지도가 낮은 도시는 물리적 거리보다 인지거리가 더 길어지게 된다.

또한 시·군 간 이동흐름에서 나타난 특징으로 사람들은 자신이 거주하고 있는 도(道) 내로의 이동 비율이 매우 높다는 점을 감안하여 본 논문에서는 이동흐름을 설명하는 세 번째 독립변수로 ‘지역 내 이동’ 변수를 선정하였다. 지역 내 이동에서 지역은 8개 지역으로 구분하여 더미변수화하였다(<표 6> 참조). ‘지역 내 이동’ 변수의 추정계수값이 높을수록 동일지역 내 도시에서 유입하는 전입자수가 많다는 것을 의미하게 된다. 이는 전출·입지가 동일 지역 내에서 많이 이루어지고 있는 인구이동의 흐름을 통제해줄 수 있는 변수라고 볼 수 있다.

위계선형모형의 2수준의 분석 단위는 82개 도시로의 전입자 흐름의 차이를 설명하는 모형이므로, 종속변수에 영향을 주는 독립변수는 전입지의 지역특성을 나타내는 변수들을 고려하였다. 2수준 모형의 전입지의 지역특성 변수는 1수준의 상수항에 영향을 미치게 되는 전입지의 매력요인이라고 볼 수 있다. 본 논문에서는 전입지의 매력도를 나타내는 전입지의 지역특성 변수로 인구이동에 영향을 미치는 것으로 잘 알려진 경제적 요인과 주택 요인을 고려하였다. 경제적 요인으로는 취업 기회²⁾를 반영하는 2007년도 각 도시의 ‘중사자수’를 변수로 선정하였고, 주택 요인으로 신규 주택기회를 반영하기 위해 2005~2008년까지 ‘허가된 신규 주거용 건축면적’을 변수로 선정하였다³⁾. 또한 전

2) 인구이동을 유발하는 전입지의 매력도로 선정한 취업기회는 중사자수/인구수(또는 중사자수/경제활동인구수)로 상대적 취업기회를 측정하여야 하지만, 우리나라 중소도시의 경우 사회서비스업과 공공서비스업에서의 중사자수가 상대적으로 많아서 상대화시킬 경우 중소도시들의 상대적 취업기회가 더 높게 나타나는 문제점을 안고 있음. 또한 이주자들은 상대적 일자리수를 인지한다기보다는 총체적 일자리수(절대량)를 인식하여 이동한다고도 볼 수 있으므로 본 논문에서는 절대적인 일자리수로 취업기회를 측정하였음.

3) 전입지의 주거 매력도를 나타내는 주택기회는 주택수/가구수로 나누어 주택보급 수준을 측정하는 것이 바람직하지만, 우리나라

입지가 지니고 있는 지역특성 변수 이외에 ‘지명도’ 변수를 고려하였다. 이는 우리나라의 경우 인구가 동에서 가장 지명도가 높고 선호하는 지역이 수도권이라는 점을 고려한 것이다. 즉, 수도권은 다른 지역에 비해 사회·경제, 문화, 교육적인 면에서 양호한 환경을 지니고 있어 수도권으로의 인구이동 성향이 매우 높기 때문에 지명도 변수를 통제변수로 투입하기 위해 비수도권을 참조 집단으로 하여 수도권을 나타내는 더미변수를 포함하였다.

또한 ‘지역 내 이동’이 도시의 전입자수에 미치는 영향은 전입지가 속한 광역권에 따라 다르다고 전제하고 다른 광역권에 비해 자신이 거주하는 권역 외부로의 이동비중이 높은 수도권으로 참조집단으로 하여 각 광역권(충청권, 호남권, 대경권, 동남권, 강원권)을 더미변수화하였다. 따라서 각 광역권의 추정계수가 클수록 수도권에 비해 해당 권역내의 도시는 같은 권역 내의 다른 도시로부터 오는 전입자수의 비중이 크다는 것을 의미한다. 이는 곧 각 권역 내에서만 주로 이동이 이루어지는 인구이동의 폐쇄성을 의미한다고 볼 수 있다. 실제로 우리나라의 인구이동 흐름을 분석한 결과 호남권과 동남권 간에 인구이동은

거의 이루어지지 않고 있었으며, 광역권 내부에서의 이동이 주류를 이루고 있었다.

한편 도시 간 이동흐름에 영향을 주는 인지거리의 거리조락 효과도 전입지의 매력도가 클수록 작아진다고 예상할 수 있다. 즉, 인지거리 자체도 전입지가 대도시인지 또는 중도시인지에 따라서 인지거리의 조락효과가 다를 수 있다. 이에 따라 본 논문에서는 인지거리 변수도 인구 50만 명 미만의

표 7_ 위계선형모형 구축에 선정된 1수준과 2수준 변수들

1수준	2수준	변수 형태
상수항	2007년 종사자수	연속변수 (자연로그)
	2005~2008년 누적 주거건축허가 면적	연속변수 (자연로그)
	지역 지명도(수도권 더미변수: 수도권 = 1)	더미변수
지역 내 이동	충청권, 호남권, 대경권, 동남권, 강원권, 수도권(참조집단)	더미변수
인지 거리	인구 50만 명 미만(참조집단) 인구 50만 명 이상~100만 명 미만 인구 100만 명 이상	더미변수
전출지 인구수	-	-

1 수준 위계선형모델

$$LN(\text{전입자수}_{ij}) = \beta_{0j} + \beta_{1j}(\text{지역 내 이동}_{ij}) + \beta_{2j}LN(\text{인지거리}_{ij}) + \beta_{3j}LN(\text{전출지 인구수}_{ij}) + \epsilon_{ij}$$

2 수준 위계선형모델

$$\beta_{0j} = \gamma_{00} + \gamma_{01}LN(\text{종사자수}_j) + \gamma_{02}LN(\text{신축주거면적}_j) + \gamma_{03}(\text{지명도}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{충청}_j) + \gamma_{12}(\text{호남}_j) + \gamma_{13}(\text{대경}_j) + \gamma_{14}(\text{동남}_j) + \gamma_{15}(\text{강원}_j) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(50 - 100\text{만 미만}_j) + \gamma_{22}(100\text{만 이상}_j) + u_{2j}$$

의 경우 각 도시의 총주택수는 인구센서스가 실시된 2005년 데이터가 가장 최신의 자료임. 그러나 본 논문은 2008년에 이루어진 인구이동 흐름이기 때문에 2005년도 시점의 데이터를 사용할 수 없다는 자료의 한계점으로 인해 2005~2008년 동안 신규로 허가된 주택면적 변수를 사용하였음. 따라서 주택기회라기보다는 신규 주택기회 변수를 측정할 것이라고 할 수 있음.

도시를 참조집단으로 하여 인구 50~100만 명 도시, 인구 100만 명 이상 도시를 더미변수화하였다. 최종적으로 1수준과 2수준의 독립변수를 포함한 위계선형모형을 보면 <표 7>과 같다.

이와 같이 본 논문에서 구축한 인구이동모형은 1수준 독립변수들과 2수준 독립변수들을 포함한 2수준의 임의계수모형으로, 82개 도시로의 전입자 수는 경제적 요인, 주택요인 그리고 지명도에 의해 영향을 받으며(상수항), 전입지가 같은 도내에 있는 지역(지역 내 이동)인가의 여부, 인지거리, 전출지의 인구수에 영향을 받는다. 그러나 지역 내 이동은 각 광역권별로 다르게 나타나며, 인지거리에 따른 거리조각 효과도 전입지의 인구규모에 따라서도 달라지게 된다. 본 논문에서는 6,606개 1수준 표본이 82개 2수준 전입지에 속하는 위계적 자료구조로 데이터베이스를 구축하였으며, 더미변수를 제외한 모든 변수들은 자연로그값으로 변환하였다.

IV. 인구이동 모형의 추정결과

1. 위계선형모형의 수준별 분산 비교

위계선형모형은 각 수준별로 종속변수의 총분산을 설명할 수 있고 고정효과만이 아니라 임의계수모형을 통해 각 도시별로 계수가 추정된다는 점에서 일반선형회귀모형보다 더 활용성이 높다고 볼 수 있다. 그러나 이러한 활용성은 수준별로 추정하는 것이 보다 더 타당하기에 대한 통계적 검정을 거쳐

야만 한다. 일반적으로 위계선형모형의 적합성에 대한 검정은 독립변수들이 포함되지 않고 상수항만을 갖는 무제약모형과 최종분석모형에서의 분산비교를 통해 가능하다. 무제약모형은 상수항의 임의효과만을 가정한 모형으로 1수준의 분산과 2수준의 분산을 구분하고 비교할 수 있는 기준이 된다. 즉, 이 모형을 통해 도출된 두 수준의 분산이 통계적으로 유의하다면 각 수준이 총분산에 미치는 영향력을 비교할 수 있다. 또한 무제약모형은 연구를 진행하며 단계별로 투입되는 독립변수의 설명력을 판단하는 기준이 되기도 한다⁴⁾.

본 논문에서는 무제약모형과 1수준 변수만 포함된 모형, 2수준 변수까지 모두 포함된 모형을 비교하여 각 수준별 분산의 차이를 비교하고 각 수준의 변수가 포함됨에 따라 줄어드는(설명되는) 분산의 양을 통해 수준별 모형 설정의 타당성을 검증하였다. 먼저 무제약모형에서 산출된 분산량 검정 결과를 통해 위계선형모형이 수준을 구분하지 않고 하나의 모형으로 분석하는 일반회귀모형에 비해 더 나은 결과를 도출할 수 있는가를 판정하였다. 그 결과 2수준의 분산이 1.027로 통계적으로 유의미하게 나타났다. 각 수준별로 산출된 분산량을 보면 종속변수에 대한 총분산량(2.987)의 약 34%(1.027)가 2수준의 전입지의 지역특성 차이에 의해 설명되고 있으며, 나머지 약 66%(1.959)는 1수준의 독립변수들에 의해 설명되는 것으로 나타났다(<표 8> 참조).

또한 각 수준별 분산감소비율 통계치를 산출하

4) 수준별로 분산을 추정하는 것에 대한 통계적 타당성은 무제약모형에서 나타난 분산의 유의성뿐만 아니라 분산감소비율 통계(propotion reduction in variance statistics)를 산출해서도 알 수 있다(이성우 외, 2006). 분산감소비율 통계(θ)란 OLS모형의 결정계수와 마찬가지로 추가된 독립변수들이 무조건부 모형에서 존재하고 있던 분산을 설명함으로써 어느 정도로 분산을 감소시키는가를 보여주는 것이다.

$$\theta = \frac{\text{무제약모형의 } L \text{수준의 분산성분} - \text{임의계수모형의 } L \text{수준의 분산성분}}{\text{무제약모형의 } L \text{수준의 분산성분}}$$

표 8_ 위계선형모형의 수준별 분산량

분산		무계약 모형	임의계수모형	
			2수준 변수 미포함	2수준 변수 포함
1수준		1.959***	0.428***	0.427***
2수준	상수항(U0)	1.027***	1.823***	1.269***
	도내이동(U1)		0.744***	0.053***
	인지거리(U2)		0.021***	0.021***
분산 합		2.987	3.015	1.771
1수준 사례수		6,606		
2수준 사례수		82		

주: ***, **, *: 유의수준 0.01, 0.05, 0.1에서 유의함.

였다. 특히 상수항, 지역 내 이동, 인지거리에 임의 효과를 가정하고 2수준 변수를 모두 포함시킨 임의계수모형의 추정결과를 보면 총분산량이 3.015에서 1.771로 줄어들어 약 41% 분산이 감소된 것으로 나타났다⁵⁾. 이와 같은 검증결과를 통해 1수준과 2수준을 분리하여 추정하는 위계선형모형이 일반 회귀모형에 비해 수준별로 분산을 비교할 수 있으며, 종속변수의 총분산량을 더 많이 설명할 수 있는 적합한 모형이라고 판정할 수 있다.

2. 위계선형모형의 추정결과 및 풀이

2수준 변수를 포함한 임의계수모형의 추정결과를 보면 상수항, 즉 전입지의 특성을 나타낸 변수들인 종사자수와 누적 주거 건축면적, 수도권 더미변수 모두 양(+의 부호를 보이면서 통계적으로 유의미하게 나타났다. 따라서 일자리 기회와 신규주택기회가 증가할수록 전입자수가 증가하며, 수도권 도

시일수록 다른 지역의 도시들에 비해 전입자수가 더 많을 것이라고 풀이할 수 있다. 즉, 전입지의 총 종사자수가 1% 증가하는 경우 해당 전입지로의 전입자수가 약 0.595% 더 늘어나며, 신규 주거 건축면적이 1% 증가할수록 해당 전입지로의 전입자수가 약 0.171% 더 늘어나는 것으로 추정되었다 (<표 9> 참조).

또한 2수준에 수도권이 참조집단으로 되어 있는 ‘지역 내 이동’ 변수의 추정결과를 보면 다음과 같다. 수도권이 참조집단인 2수준의 상수항의 계수가 양(+)로 유의하게 나타나고 있어 수도권 도시들의 경우 지역 외부 도시로부터의 전입자수에 비해 지역 내 도시에서 유입되는 전입자수가 많음을 말해준다. 다른 권역의 더미 변수 추정계수들도 모두 양(+의 부호를 나타내면서 통계적으로 유의미하게 나타났다. 따라서 모든 지역 도시들이 지역 외부 도시에서 오는 전입자수에 비해 지역 내 도시에서 오는 전입자수가 수도권에 비해 더 많음을 말해준다. 지역별로 보면 강원권의 추정계수가 2.034로 가장 크고 이어 호남권, 동남권 순으로 나타나고 있다. 따라서 강원권이나 호남권, 동남권은 다른 지역에 비해 권역 내 이동이 주류를 이루고 있으며 따라서 인구이동의 폐쇄성이 상대적으로 높다고 풀이할 수 있다.

본 논문에서는 위계선형모형을 통해 추정된 계수들을 토대로 하여 각 도시별로 예측된 전입자수

5) 1수준의 분산감소비율 통계=(2.987-1.959)/2.987=0.344(34.4%), 2수준의 분산감소비율 통계=(3.015-1.771)/3.015=0.412(41.2%)로 산출되어, 각 수준별로 분산감소효과가 상당히 크지만, 특히 2수준에서의 분산감소효과가 더 크게 나타나고 있어 수준별로 분산을 추정하는 것이 타당하다고 판정할 수 있음.

와 실제 전입자수가 얼마나 차이가 있는가를 비교하였다. 이는 구축된 인구이동 모형의 정확성 및 신뢰성을 평가하는 것이라고도 볼 수 있다. 먼저 자연로그를 취한 예측 전입자수와 실제 전입자수 간의 상관관계를 분석한 결과 상관계수가 0.93으로 매우 높게 나타나고 있어 인구이동모형의 신뢰성이 상당히 높다고 볼 수 있다. 한편 자연로그값을 변환시켜 실제 수치로 환산하여 각 도시별 예측 전입자수와 실제 전입자수 간의 상관관계를 산출한 결과 상관계수는 0.77로 나타나 (자연로그값을 취한 예측치와 실측치 간의 상관계수에 비해 다소 낮음), 여전히 인구이동모형의 신뢰성이 비교적 높다고 볼 수 있다.

그러나 이러한 상관계수는 82개 도시 전체를 대상으로 한 것이며, 위계선형모형에서 추정된 예측치와 실측치의 차이가 도시마다 상당히 다르게 나타날 수 있다. 어떤 도시의 경우 예측치와 실측치 간의 차이가 거의 없어서 모형의 정확성이 매우 높은 도시도 있을 수 있는 반면 도시의 특이성으로 인해 모형을 통해 산출된 예측치와 실측치 간의 차이가 상당히 크게 나타나는 도시도 있을 수 있다. 본 논문에서는 각 도시별로 실제 전입자수와 예측 전입자수 간 차이인 잔차를 비교하였다. 전입자수가 도시 규모에 따라 다르기 때문에 잔차를 산출하여 비교하였다. 그 결과 82개 도시 가운데 잔차가

표 9_ 위계선형모형의 추정결과

독립 변수		무제약 모형	임의계수모형	
1수준	2수준		2수준 변수 미 포함	2수준 변수 포함
상수항	상수항	4.265***	-4.810***	-4.909***
	종사자수			0.595***
	주거용 건축면적			0.171***
	수도권			0.355***
지역 내 이동	상수항		1.421***	0.462***
	충청권			1.260***
	호남권			1.740***
	대경권			1.229***
	동남권			1.479***
인지거리	상수항		-0.422***	-0.440***
	50~100만 명 미만			0.043**
	100만 명 이상			0.059**
전출지 인구수			0.941***	0.943***

주: ***, **, *는 유의수준 0.01, 0.05, 0.1에서 유의함.

율은 0.77%로 매우 낮았다. 즉, 서울시로 실제 전입한 사람의 수는 53만 9,806명인데 비해 모형을 통해 추정된 예측 전입자수는 54만 3,952명으로 예측값이 4,146명 더 크게 추정되었다. 그러나 이는 실제 전입자수 약 54만 명의 0.77%에 불과한 것으로 예측값의 정확도가 매우 높다고 볼 수 있다. 서울에 이어 경기도 이천시, 경남 거제시, 경남 통영시, 부산광역시 순으로 잔차율이 낮게 나타났다 (<표 10> 참조).

전반적으로 볼 때 잔차율이 낮은 도시들은 광역시를 비롯한 지방 수위도시들, 그리고 비교적 대도시와 거리가 먼 도시들이다. 이에 반해 잔차율이 가장 높은 도시로는 경기도 김포시로 나타났다. 김포

표 10_ 인구이동 예측치와 실측치 간의 잔차율이 작은 상위도시들

순위	도시명	실제 전입자수	예측 전입자수	잔차	잔차율
1	서울	539,806	543,952	-4,146	-0.77
2	이천시	14,204	13,845	359	2.53
3	거제시	19,104	18,440	664	3.47
4	통영시	9,738	9,083	655	6.72
5	부산	109,723	102,096	7,627	6.95
6	진주시	15,537	16,640	-1,103	-7.10
7	대전	67,905	59,609	8,296	12.22
8	창원시	32,384	28,125	4,259	13.15
9	사천시	8,869	7,688	1,208	13.31
10	안동시	8,140	6,985	1,155	14.19

주: 잔차율 = 잔차 / 실제 전입자수 × 100

시의 경우 실제 전입자수는 3만 4,767명인데 비해 예측 전입자수는 1만 4,003명으로 과소추정하였으며, 잔차율은 약 60%에 달하고 있다. 김포시 외에도 잔차율이 높은 도시들을 보면 의정부시, 하남시, 구리시, 과천시 등 경기도 내 도시들이 많고 특히 서울과 인접한 도시들의 잔차율이 높게 나타났다.

V. 결론

인구이동은 개개인이나 가구가 좀 더 나은 환경을 지닌 지역으로 이동하려는 기본 욕구에서 시작되지만, 그 결과는 전입지와 전출지의 인구, 사회, 경제적인 면에 상당한 영향을 미치게 된다. 우리나라 인구이동의 특징을 보면 점점 더 1인 이동의 비율이 늘어나고 있으며, 특히 지역 내 이동보다 지역 간 이동 시에 이동건수당 이동자수는 1.46명으로 훨씬 더 작게 나타나고 있다. 또한 비수도권에서 수도권으로 이동하는 경우 1인 세대의 이동 비율이 전체 이동가구의 83%를 차지할 정도로 매우 높게 나타났

다. 이는 우리나라의 경우 점점 더 인구이동을 통해 1인 세대수가 늘어나고 있으며, 향후 가족 구성에 상당한 문제를 야기시킬 것임을 시사해준다.

82개 도시 간에 이루어지는 인구이동 흐름의 패턴을 분석한 결과 전입지와 전출지 간의 상관관계는 0.95 이상으로 매우 높게 나타나고 있어, 전출자가 많은 도시는 전입자수도 많음을 말해준다. 또한 인구이동의 공간적 패턴을 보면 전출지를 떠나 전입하는 도시의 수는 선택적이며 소수의 도시들로 집중하여 유입하는 패턴을 보이고 있다.

본 논문에서는 이와 같은 인구이동 흐름의 특징을 토대로 하여 지금까지 동일한 수준에서 분석되어오던 지역 간 인구이동 모형을 위계선형모형을 이용하여 개별도시 간에 이루어지는 이동흐름 특성과 전입지의 지역특성으로 수준을 달리하여 분석하였다. 도시 간에 이루어지는 전입량의 차이를 설명하기 위해 1수준에서는 전출지의 인구수, 인지거리, 지역 내 이동 변수를 선정하였다. 그러나 2수준에서는 도시 간 전입량의 차이를 설명하는 전입

지의 특성 변수로 일자리 기회와 주택 기회를 고려하였으며, 전입지 선택에서의 수도권을 선호하는 성향을 반영한 모형을 구축하였다.

위계선형모형을 추정한 결과 도시 간 전입량의 차이(중속변수)를 나타내는 충분산의 약 66%는 1수준에서 설명되었으며, 나머지 34%는 2수준에서 설명되었다. 즉, 도시 간 인구이동 요인들의 영향력을 수준을 달리하여 추정함으로써 인구이동 흐름을 보다 더 정확하게 설명할 수 있었다. 본 논문에서는 인구이동 흐름에서 나타난 특징을 반영하여 인접 도시의 영향력과 이동거리에 따른 거리조락 효과가 도시가 속한 지역 또는 인구규모에 따라 다르다는 점을 모형에 반영하였다. 그 결과 인구이동 모형의 예측력은 상당히 높게 나타났다. 위계선형모형을 통해 예측된 전입자수와 실제 전입자수 간의 상관계수는 상당히 높게 나타났으며, 잔차율도 서울의 경우 1% 미만으로 상당히 정확하게 추정되었다. 그러나 개별도시별로 잔차율을 보면 전반적으로 대도시 인접 도시들이 실제 전입자수와 예측된 전입자수 간에 많은 차이를 보이고 있다. 인구이동모형에서 거리조락 효과를 고려하였음에도 불구하고 대도시와 인접해 있는 도시들에게 미치는 대도시의 영향력이 상대적으로 매우 큰 것으로 나타났다.

두 도시 간에 이루어지는 이동량은 전출지의 배출요인과 전입지의 흡인요인이 동시에 작용하여 나타난 결과다. 그럼에도 불구하고 본 논문에서 구축한 인구이동모형의 경우 전입지의 특성은 모든 전출지에 동일한 영향력을 미치는 것으로 구축되어 있으며, 전출지의 배출요인이 포함되지 못하였다. 또한 한 시점만을 대상으로 하였기 때문에 일자리 매력도가 시차를 두고 인구이동에 영향을 줄 수 있는 점을 고려하지 못하였다는 한계점도 있다.

향후 인구이동의 흐름을 보다 더 정확하게 설명

하고 더 나아가 예측하기 위해서는 개별도시의 특성뿐만 아니라 그 도시가 속한 광역권 또는 생활권의 특성을 포함할 수 있는 모형으로 확장할 필요성이 있다. 본 논문에서 사용된 위계선형모형은 2수준 모형이지만 앞으로 3수준 모형으로 확장하여 각 전입지들이 속한 권역의 특성을 나타내는 변수들을 더 포함시킨다면 인구이동의 흐름을 유발하는 권역별 영향력이 추가되어 인구이동의 흐름을 보다 더 잘 설명할 수 있을 것이다.

한편 본 논문의 모형에서 예측치와 실측치의 차이가 가장 크게 나타난 도시들은 주로 수도권에 속해 있는 도시들, 특히 서울과 매우 인접해 있는 도시들이었다. 이는 수도권 내에서 일어나고 있는 인구이동의 흐름과 이동 요인은 전국적 차원에서 도시들 간에 이루어지는 인구이동과는 다르다는 것을 시사해준다. 따라서 수도권을 대상으로 한 인구이동 모형이 별도로 구축되어야 하며, 더 나아가 각 권역별 인구이동 모형도 개발될 필요성이 있을 것이다.

참고문헌

- 강상진. 1995. "다층통계모형의 방법론적 특성과 활용방법". 교육평가연구 제8권 제2호. 서울 : 한국교육평가학회. pp63-94.
- . 1998. "교육 및 사회연구를 위한 연구방법으로서 다층모형과 전통적 선형모형과의 비교 분석 연구". 교육평가연구, 제 11권 2호. 서울 : 한국교육평가학회. pp207-258.
- 곽인숙·김순미. 1996. "가족생활주기별 주거소비 및 주거이동지향성에 관한 연구". 한국가정관리학회지 제14권 제4호. 경기 : 한국가정관리학회. pp233-248.
- 권상철. 2005. "우리나라 수도권으로의 인구이동: 시기별 유출지역 특성과 이주자 선별성의 상대적 중요도 평가". 한국지

- 역지리학회지 제11권 제6호. 서울 : 한국지역지리학회. pp571-585.
- 김갑성. 2006. “수도권주택건설과 인구유입의 인과관계”. 수도권 택지개발사업의 수요와 파급효과에 대한 토론회 자료집. 서울 : 대한국토·도시계획학회. pp25-49.
- 김성태·장정호. 1997. “한국 지역 간 인구이동의 경제적 결정요인 : 1970-1991”. 국제경제연구 제3권 제2호. 서울 : 한국국제경제학회. pp175-197.
- 김미희. 1983. “주거이동 지향성에 관한 연구: 서울시를 중심으로”. 대한가정학회지 제21권 제4호. 서울 : 대한가정학회. pp103-114.
- 김경호. 1990. “도시기구의 주거이동에 관한 사회학적 접근”. 도시행정학보 제3권. 서울 : 한국도시행정학회. pp79-100.
- 신현곤. 1998. “이주의사결정에 있어 경제적 요인의 역할: 확률적 효용모형의 적용을 통한 인구특성별 차별적 형태파악을 중심으로”. 연세경제연구 제5권 제1호. 서울 : 연세대학교 경제연구소. pp69-97.
- 오정일·안기돈. 2007. “우리나라 15개 시도의 일자리수 변화와 인구이동 간 인과성 검증”. 국토연구 제53권. 경기 : 국토연구원. pp57-76.
- 유경문. 1991. “인구이동의 결정요인에 관한 실증분석: 한국의 경우 1966-1985를 중심으로”. 경제학연구 제39권 제1호. 서울 : 한국경제학회. pp157-209.
- 유정진. 2006. “위계적 선형모형의 이해와 활용”. 아동학회지, 제27권 제3호. 서울 : 한국아동학회. pp169-187.
- 이문식. 2004. “Economic Determinants of Recent Internal Migration in Korea Industrial Economics Association”. 산업경제연구 제17권 제4호. 경남 : 한국산업경제학회. pp.1155-1170.
- 이성우. 2001. “지역특성이 인구이동에 미치는 영향 : 독립이동과 연계이동”. 지역연구, 한국지역개발학회, 제18권 3호, 경북 : 한국지역개발학회. pp.19-44.
- 이성우·윤성도·박지영·민성희. 2006. 공간계량경제모형응용. 서울 : 박영사.
- 이희연. 2008. 인구이동 확장모형 개발 및 실증 분석. 경기 : 국토연구원.
- 이희연·박정호. 2009. “경로분석을 이용한 인구이동 결정요인들간의 인과구조”. 한국경제지리학회지 제12권 제2호. 강원 : 한국경제지리학회. pp123-141.
- 정성기·이성근·고덕균. 2008. “아파트 주거만족도에 따른 주거이동 결정요인”. 부동산학보 제33권 제13호. 서울 : 한국부동산학회. pp236-247.
- 최유미·남진. “서울이 가국특성별 주거선택 요인에 관한 연구”. 국토계획 제43권 제3호. 서울 : 대한국토·도시계획학회. pp195-211.
- 하상근. 2005. “지역 간 인구이동의 실태 및 요인에 관한 연구: 경상남도의 기초자치단체를 중심으로”. 지방정부연구 제9권 제3호. 부산 : 한국지방정부학회. pp309-332.
- Cadwallader, M. 1989. “A Conceptual Framework for Analysing Migration Behaviour in the Developed World”. *Progress in Human Geography*, vol.13, no.4. London : Sage. pp494-511.
- Greenwood, M. J. 1985. “Human Migration: Theory, Models, and Empirical Studies”. *Journal of Regional Science* vol.25. NJ : Blackwell Publishing. pp521-544.
- Hofmann, D. A. 1997. “An Overview of the Logic and Rationale of Hierarchical Linear Models”. *Journal of Management* vol.23, no.6. London : Sage. pp723-744.
- Luke, D. A. 2004. *Multilevel Modeling*. Thousand Oaks, CA : Sage.
- Michalos, A. C., 1996. “Migration and the Quality of Life: a Review Essay”. *Social Indicator Research* vol.39, no.2. Netherlands : Springer. pp121-166.
- Pollack, B. N. 1998. “Hierarchical Linear Modeling and the Unit of Analysis Problem: a Solution for Analyzing Responses of Intact Group Members”. *Group dynamics* vol.2, no.4. Washington : American Psychological Association. pp299-312.
- Raudenbush, S. W. and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousands Oaks, California : Sage.
- Shaw, R. P. 1975. *Migration, Theory and Fact*. Philadelphia: Regional Research Institute.
- Week, J. R. 2002. *Population: An Introduction to Concepts and Issues*. 8th ed. Belmont : Wadsworth Pub.
- Zelinski, W. 1971. “The Hypothesis of the Mobility Transition, *Geographical Review*”. *Geographical Review* vol.61, NY : American Geographical Society. pp219-249.

- 논문 접수일: 2010.10.25
- 심사 시작일: 2010.11. 4
- 심사 완료일: 2010.11.25

ABSTRACT

An Analysis of the Migration Flow Using Hierarchical Linear Model

Keywords: Population Movement, Migration Flow, Hierarchical Linear Model, Origin-destination

Migration is an individual or household behavior in order to live a better residential environment, but the results of the population movement have considerable impacts on population, social and economic aspects in the origin and destination. The purpose of this paper is to develop the model of migration flows among 82 cities, to identify the determinants of migration flows. This paper is composed of two parts. First, the migration pattern and migration flows among regions in Korea was analyzed in terms of the personal characteristic and the spatial perspective. Second, the migration flow was analyzed using hierarchical linear model decomposing determinants of migration flow at the different levels. According to the estimation results of HLM, 66% of the total variance of the migration flow among 82 cities(the dependent variable)was explained by 1-level determinants such as perceived distance, population of the origin, and intra-migration and the remaining 34% was explained by 2-level determinants such as employment opportunities, new housing opportunities, and regional attraction. In the light of the high correlation coefficient between predicted immigrants and actual immigrants of each city, hierarchical linear model could be explained more precisely the migration flow in Korea.

위계선형모형을 이용한 인구이동 흐름 분석

주제어: 인구이동, 이동 흐름, 위계선형모형, 전출지-전입지

인구이동은 개개인이나 가구가 좀 더 나은 주거환경으로 삶의 보금자리를 옮기려는 기본 욕구에서 시작되지만, 그 결과는 전입지의 전출지의 인구, 사회, 경제적인 면에 상당한 영향을 미치게 된다. 본 논문은 우리나라 82개 도시 간의 인구이동의 흐름을 분석하고 인구이동에 영향을 미치는 결정요인을 밝히려는 데 목적을 두었다. 본 논문은 크게 두 부분으로 구성되었다. 첫째, 개인적 특성에 따른 인구이동 현상과 공간적 측면에서의 우리나라의 인구이동 흐름을 분석하였다. 둘째, 위계선형모형을 이용하여 개별도시 간 전입자수를 이동흐름 특성과 전입지의 지역특성으로 수준을 달리하여 분석하였다. 위계선형모형의 추정결과 도시 간 전입자 수(종속변수) 차이의 약 66%는 1수준의 이동흐름을 나타내는 변수들(거리, 전출지 인구수, 지역 내 이동)에 의해 설명되었고, 나머지 34%는 2수준의 전입지 특성 변수들(고용기회와 주택기회, 지역 매력도)에 의해 설명되었다. 본 연구모형을 통해 예측된 각 도시의 전입자수와 실제 전입자수 간의 상관계수를 산출한 결과 매우 높게 나타났다. 본 논문의 의의는 도시 간 이동흐름에 영향을 미치는 요인들을 수준을 달리하여 추정하는 경우 인구이동 흐름을 보다 더 정확하게 설명할 수 있음을 보여주었다는 점에 있다.

[부록]

부표 1_ 82개 도시별 전출량과 전입량 간의 시계열적 상관관계(2000~2009년)

전입량	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	합계	평균
2000	1											
2001	.986	1										
2002	.976	.986	1									
2003	.977	.983	.988	1								
2004	.957	.971	.978	.979	1							
2005	.964	.977	.981	.979	.989	1						
2006	.954	.976	.978	.973	.979	.990	1					
2007	.954	.965	.965	.960	.971	.980	.978	1				
2008	.939	.948	.950	.943	.957	.964	.958	.988	1			
2009	.943	.958	.959	.955	.967	.975	.971	.984	.981	1		
합계	.979	.989	.990	.988	.988	.993	.989	.988	.976	.982	1	
평균	.979	.989	.990	.988	.988	.993	.989	.988	.976	.982	1.000	1
전출량	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	합계	평균
2000	1											
2001	.996	1										
2002	.993	.997	1									
2003	.987	.990	.993	1								
2004	.980	.984	.988	.992	1							
2005	.979	.982	.986	.989	.993	1						
2006	.977	.982	.985	.982	.985	.993	1					
2007	.969	.975	.979	.980	.985	.992	.994	1				
2008	.964	.970	.975	.975	.979	.989	.990	.995	1			
2009	.960	.967	.972	.976	.980	.988	.989	.994	.993	1		
합계	.988	.992	.994	.994	.994	.996	.995	.994	.990	.989	1	
평균	.988	.992	.994	.994	.994	.996	.995	.994	.990	.989	1.000	1

자료: 통계청. 각 연도. 국내 인구이동통계.

부표 2_ 전입률(량)과 전출률(량) 간의 상관관계

구분		1997	2000	2005	2008
전입률과 전출률	시도	0.961**	0.972**	0.950**	0.949**
	시·군·구	0.855**	0.867**	0.879**	0.889**
전입량과 전출량	시도	0.992**	0.997**	0.997**	0.997**
	시·군·구	0.978**	0.990**	0.990**	0.989**

주: **유의수준 0.05에서 유의함.