

연구논문

서울시 행정 구역과 이웃 효과 관계 탐색: 한국종합사회조사(KGSS) 및 행정 자료 연계*

김이현** · 김지범*** · 강정한****

지금까지 사회조사 자료와 지역 단위 행정 자료를 결합하는 이웃 효과 연구는 주로 시·군·구 단위에서 진행해 왔으나, 시·군·구 단위가 생활환경 단위로서의 지역을 대변하지 못하며, 시·군·구 내의 더 작은 지역 단위 간의 이질성이 서로 상쇄되어 지역 효과를 포착하기 힘들었다. 이러한 문제의식에서 출발한 본 연구는 서울시 주민에게 미치는 동 수준의 지역 효과를 탐색하고 향후 연구의 방향을 모색하고자 하였다. 사회조사 자료는 2003~2016 한국종합사회조사(KGSS) 누적 응답자 중 서울 거주자를 선택했으며, 서울시 동 수준의 행정 정보는 2003~2017 서울열린데이터광장, 2006~2017 중앙선거관리위원회 및 2006~2017 서울부동산정보광장에서 수집, 결합하였다($N=2,656$). (1) 우선 지역 수준에서는 출생률, 고령자 인구 비율, 기초생활수급자 비율, 투표율, 평균 아파트 거래가격 등을 구 단위와 동 단위에서 수집한 후 분산을 분석한 결과, 동 간의 이질성이 구 간의 이질성보다 큰 것으로 나타나 다수준 분석에서 동 수준 지역 효과의 잠재력이 있었다. (2) 다음으로 개인 수준에서는 한국종합사회조사 응답 중 가구소득 수준, 주관적 계층, 주관적 정치 성향, 타인에 대한 신뢰 등의 분산을 분석한 결과, 역시 모든 변수에 대해 동 간의 분산이 구 간의 분산보다 더 컸다. 이는 응답자의 태도 차이가 구 간의 지역 차이보다 동 간의 지역 차이로 설명될 가능성이 높음을 암시한다. (3) 실제로 다수준 분석을 통해 이러한 태도 변수들에 대해 응답자의 성별, 연령, 교육수준을 통제하고 지역 수준에서 평균 아파트 거래가격과 고령자 인구 비율이 미치는 효과를 살펴본 결과, 구 수준보다는 동 수준에서 이러한 변수들의 효과가 더 분명한 경향이 있었다. 예를 들어 고령자 인구 비율이 높은 동에 사는 응답자일수록 소득이 낮고 자신의 주관적 계층을 낮게 평가하는 경향이 통계적으로 유의했으나, 구 수준에서는 이러한 경향이 발견되지 않았다. 마지막으로 서울시민을 대상으로 한 이러한 예시 분석을 바탕으로 다수준 다출처 분석이 갖는 잠재력을 토의하였다.

주제어: 지역 효과, 한국종합사회조사, KGSS, 서울시, 행정동, 부동산, 다수준 다출처 분석

* 이 논문(저서)은 2016년도 정부재원(교육부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2016S1A3A2925033).

** 연세대학교 사회학과 석사과정(vpfh626@yonsei.ac.kr).

*** 성균관대학교 사회학과 부교수(jbk7000@skku.edu).

**** 연세대학교 사회학과 교수(jhk55@yonsei.ac.kr), 교신저자.

I. 서론

한 사람이 거주하는 지역 사회와 이웃이 그 사람과 갖는 관계는 사회학의 핵심 주제 중 하나로, 사회학의 등장 이래 수많은 연구들이 이에 기여해 왔다. 그중에서도 사회 해체 이론(social disorganization theory)은 지역의 도시성(urbanicity)과 경제적 여건이 사회적 관계에 미치는 영향에 관심을 가졌다. 구체적으로는 도시 지역의 인구 구성, 밀도, 이질성, 빈곤 수준, 주거 안정성 등의 지역 효과 혹은 이웃 효과(neighborhood effect)가 사회적 유대와 어떻게 관련되는지 관찰한 바 있다(Shaw & McKay 1969; Wirth 1938). 미국에서 먼저 촉발된 이웃 효과 논의는 한국에서 또한 많은 주목을 끌었으며, 한국에서 이웃 효과가 과연 존재하는지, 존재한다면 어떻게 측정될 수 있는지, 측정된다면 어떤 기제로 이웃 효과가 나타나는지에 대한 논의가 활발하게 진행되었다.

한국에서의 이웃 효과 논의의 핵심은 개인 수준의 특성을 넘어 개인이 속한 지역 사회의 특성, 그리고 개인과 지역 사회 사이의 상호작용이 개인에 미치는 영향력에 집중해 왔다(권오상 외 2013; 김교성 2010; 백정우 외 2013; 심재만 2017; 장동호 2011; 한보영 외 2015). 이러한 연구들에서는 공통적으로 지역의 도시화 수준과 발달 성격, 지역 사회의 인구 구성, 생활 조건 및 기반 시설 등 개인의 삶에 영향을 미치는 사회경제적 요소를 탐구해 왔다.

그러나 이웃 효과 연구는 측정과 관련된 두 가지 차원에서 우선적으로 어려움을 겪는다. 그 두 가지 차원이란 (1) 이웃의 경계를 선정하기 위한 주관적, 객관적 방식 사이의 딜레마와 (2) 개인 수준과 집단 수준을 연계할 수 있는 자료의 한계이다. 먼저 이웃의 경계 구획 차원에서 생각해보자. 사회적 자본에 대한 논의에서 영향을 받은 이웃 효과 연구는 기본적으로 이웃 사이의 상호 신뢰와 연대를 기반으로 사회적 응집력(social cohesion)이 작용하여 생활환경이 개인에게 영향을 미친다고 이해한다(Sampson et al. 1997). 하지만 사회적 응집력이 비공식적인 통제로서 나타나기 때문에 주민들의 삶에 직·간접적으로 영향을 미치는 지역 사회의 지리적 경계를 구획하는 것은 쉽지 않다(Brattbakk 2014). 이와 같은 경계의 모호함은 같은 주제를 다루면서도 상이한 연구결과로 나타나기도 한다. 이웃을 측정하는 데 있어서 어떤 단

위 혹은 대리변수를 사용하는지에 따라 연구결과에 일관성이 결여될 수 있다(Ginther et al. 2000; Mansky 1993).

한편으로는 한 도시의 동질적 지역에서 이웃 효과가 가시적으로 나타날지에 대해서도 의견이 분분하다. 한 도시 내에서 지역들이 상대적으로 동질적이라도 지역 효과가 나타난다는 연구(Jensen et al. 2008)도 있지만, 이에 대한 우려 또한 적지 않다. 관련하여 이웃 효과의 한국적 맥락을 지적하기도 하는데, 특히 수도권에서는 생활 공간이 독립적이지 않고 대부분 지역 경계를 넘나드는 생활권의 공유로 인하여 이웃 효과가 상대적으로 잘 드러나지 않을 것으로 예측하기도 한다(한보영 외 2015).

이웃 효과 연구는 이웃의 구획에서 오는 정의의 문제뿐만 아니라 개인 수준과 지역 수준 자료의 연계 문제에 직면한다. 한국에서 설문조사를 진행하면서 실제 주거 구역 등을 수집 혹은 공개하는 데이터가 드물며, 개인 수준의 자료에서 충분히 세밀한 거주지역을 공개한 경우에도 거주지역의 특성을 파악할 수 있는 자료가 부족한 경우가 많다. 이러한 한계를 극복하기 위해 몇몇 연구에서는 설문조사 데이터를 활용하여 응답자가 주관적으로 인식하는 동네의 특성이나 사회적 유대의 수준을 지역 수준의 변수로 도출하는 방식으로 연구를 진행하기도 했다(Sampson et al. 1997; 김희운 외 2010; 이화용 2016).

반면 객관적인 지표를 사용하는 경우에는, 일반적으로 행정 단위를 기준으로 동네를 구획하여 지역 특성을 측정하였다. 이러한 접근에서는 행정 지역 단위가 중앙 정부 혹은 지방자치단체가 서로 다른 정책을 시행하는 기준이며, 사람들의 인식적인 차원에서 지역의 경계를 구분하는 기본적 단위라는 점에 주목한다. 게다가 중앙 정부 혹은 지방자치단체는 행정적 필요에 의해 오랜 시간 지역에 대한 정보를 풍부하게 수집해왔다. 오늘날에 이르러서는 점차 더 많은 행정 자료를 민간에 공개하고 있는데, 이는 (1) 지방자치 권역 내 전 지역을 포괄하여 대표성이 있으며 (2) 공적 시스템에 근거하는 만큼 신뢰성이 높다는 장점이 있다. 행정 부처별로 관리 주체가 다르기 때문에 여러 출처의 자료를 연계해야 하는 복잡한 과정을 거치지만, 지역 자료를 수집할 다른 대안이 드문 상황에서는 좋은 선택지가 되고 있다. 따라서 이웃 효과를 검증하고자 하는 연구자들은 주로 중앙정부 혹은 지방자치정부에서 공개하는 행정 자료를 수집한 뒤 개인 수준에서 배경, 행동, 태도 등을 알 수 있는 설문조사 자료와 결합하여 다수준 분석을 시행한다.

이때 사용하는 지역의 단위는 다양한데, 기존에는 중앙정부에서 공개하는 자료가

주로 기초지방자치단체, 즉 시·군·구 단위로 제시되었기 때문에 한국에서의 이웃 효과 연구 역시 시·군·구 단위에서 진행되는 경우가 많았다(Han et al. 2013; 김교성 2010; 박은옥 2013; 백학영 2007; 한보영 외 2015). 이에 따르면 기초지방자치단체는 각종 보건 및 사회적 인프라 지표들이 취합되는 단위이기 때문에 비슷한 생활 조건을 가질 것으로 여겨져 왔다(김형용 외 2012). 그 외에도 지역 수준을 다양하게 제시하고 이웃 효과를 검증하는 연구들이 다수 존재하는데, 시, 군 수준에서 지역의 동질적 특성을 보거나(권오상 외 2013), 광역지방자치단체 수준에서 이웃 효과를 확인(김교성 외 2009; 장동호 2011), 혹은 한 지역 내에서 소지역 간 차이를 검증하는 연구(심정하 외 2009) 등이 있었다.

그러나 기초지방자치단체 단위의 자료가 내부적으로 동질적이지 않아 생활환경 단위로서의 지역을 대변하지 못한다는 문제의식이 꾸준히 제기되어 왔다(박종순 2006). 한 지역 내에서 더 작은 지역 간의 이질성이 서로 상쇄되기 때문에 이웃 효과를 검증하기 어렵다는 것이다. 게다가 더 작은 지역 단위를 사용할 때 변수 간 연관성이 더욱 강하게 드러난다는 것이 일부 연구에서 지적되었다(김형용 외 2014). 가령, Han et al.(2013)은 서울시 구 수준에서 총선 투표율은 주관적 행복감과 관련이 없다는 결과를 보여 주었지만, 그렇다고 해서 다른 지역 단위에서도 행복감이 총선 투표율과 연관이 없다는 것을 의미하지는 않는다고 언급하고 있다. 마침 2010년대 이후로는 지방자치단체를 중심으로 더 세밀한 단위에서 다양한 행정 자료를 공개하는 일련의 조치들이 이루어져, 읍·면·동 단위에서의 이웃 효과를 확인할 수 있는 기반이 마련되었다. 그중에서도 서울시는 다방면으로 행정 자료를 수집하고 일반 시민에게 공개하는 정책을 적극적으로 펼치면서 지역 수준 자료의 토대를 마련해 주어, 개인 수준 자료에 연결할 수 있게 해주었다.

본 연구는 시·군·구 단위에서 읍·면·동 단위로의 지역 단위 세밀화에 관심을 가지고 서울시 주민에게 미치는 동 수준의 지역 효과를 탐색하여 향후 연구의 방향을 모색하고자 한다. 이러한 시도는 우선 방법론적으로 다수준(multi-level) 분석을 활성화시키고 사회에 축적된 자료의 가치를 증진시킬 수 있다. 즉, 개인 수준의 사회조사 자료와 동 수준의 행정 자료의 결합은 자료의 활용도를 높일 뿐만 아니라 사회과학연구 발전의 밑바탕이 될 수 있다.

또한 이론적으로는 우리 사회의 이웃 효과, 혹은 사회적 자본이나 통합에 대한 논의를 활성화하고 경험적 증거를 축적시킬 수 있다. 즉, 그간 우리나라에서 지역 효과 연구가 활성화되지 못한 것이, 지역 간 동질적이고 통합적 사회라서 지역 간

사회적 자본의 차이가 없어서인지, 아니면 그러한 차이를 포착할 만큼 세밀한 지역 단위로 내려가지 못해서인지 밝히는 데 기여할 수 있을 것이다.

마지막으로 서구에서는 다수준 다출처(multi-level multi-source) 데이터베이스 구축과 관련하여 어려운 점과 주의할 점이 보고되고 있지만(Smith & Kim 2013), 한국에서는 이런 점에 관한 심층적 보고는 드문 것으로 보인다. 본 연구에서는 이용한 자료 연계와 관련하여 유의해야 할 점도 제한적이거나 짚어보려 한다.

II. 자료 구성 및 방법

1. 연구자료 구성

본 연구는 서울시 25개 자치구와 424개 행정동 수준(2017년 기준)에서 지역 효과를 각각 검증하기 위해 2003년부터 2017년까지의 서울시 자료를 여러 가지 출처로부터 수집하여 지역 수준 변수를 구성하였으며, 여기에 개인 수준의 사회조사 자료를 연계하였다. 먼저 지역을 서울시로 한정된 것은 여러 가지 이유가 있겠으나 근본적으로는 자료의 한계가 명확했기 때문이다. 다른 지역의 경우 자료의 공개 수준이 상이하여 세부 지역 단위 간 비교가능성이 낮은 반면, 서울시는 비교적 세밀한 지역 단위인 행정동 수준의 자료를 폭넓게 공개하여 지역 특성 변수의 수집이 용이하였다. 게다가 도시 지역일수록 인구가 밀집되어 있고 같은 자치구 사이에서도 행정동 사이에 이질성이 클 가능성이 높아 더욱 세밀한 지역 단위에서의 이웃 효과를 검증하는 연구에 적합하다고 판단하였다.

서울시를 구획하는 세부 지역 단위로는 행정동 외에도 법정동이 존재하나, 본 연구에서는 행정동 단위를 사용하였다. 법정동은 토지의 구획과 부동산 등 법률 행위의 단위인 데 반해 행정동은 지역의 행정 편의와 관리를 위해 사용하는 단위이므로 상대적으로 지역 인구 균형이 잘 맞추어져 있다는 장점이 있다. 게다가 서울시가 공개하는 지역 수준 자료가 기본적으로는 행정 단위인 행정동을 기준으로 수집되었기 때문에 거의 모든 자료가 행정동 단위로 취합되었다. 그러나 효율적인 행정 관리 차원에서 수시로 변경되는 행정동의 특성상 다수준 분석을 하기 위해서는 연도별 일관성을 유지하는 것이 중요한 과제가 된다. 실제로 서울시의 행정동은 2008년

한 차례 대규모 재편성을 겪은 데다 이후로도 소소한 재편성을 거쳤으며, 2015년에는 송파구 위례동이 신설되기도 하였다. 이러한 상황에서 지역적 일관성을 유지하기 위해, 2015년부터 유지되고 있는 424개 행정동 기준에 맞추어 변경된 지역을 일일이 확인하여 가장 비슷하게 연관된 동으로 코드화했다.

서울시 자치구와 행정동 수준의 지역 자료를 구성하기 위해 세 가지 출처에서 자료를 수집하였다. 첫째, 서울열린데이터광장(data.seoul.go.kr)으로부터 인구 구성 및 세대 수, 외국인 수, 65세 이상 고령자 인구, 혼인 및 이혼 건수, 출생자 및 사망자 수, 인구 이동 추이, 기초생활수급자 수와 같이 기본적인 행정 자료를 자치구와 행정동 단위 각각에서 수집하였다. 둘째, 중앙선거관리위원회(info.nec.go.kr)로부터 자치구별 및 행정동별로 전체 투표율과 정당 및 후보별 지지율을 얻을 수 있었다. 셋째, 서울부동산정보광장(land.seoul.go.kr)으로부터 해당 기간 동안 거래된 아파트 건수별로 가격과 정보를 수집하여, 아파트 위치에 따라 자치구 및 행정동 수준으로 통합하여 평균 가격과 표준편차, 거래량을 계산하였다.

지역 수준과 연계할 개인 수준 자료로는 한국종합사회조사(Korean General Social Survey: KGSS)를 사용하였다. 한국종합사회조사는 18세 이상 한국어를 할 수 있는 성인 대상의 면접조사로 2003년부터 2014년까지는 매년, 2016년부터는 격년으로 시행되어 온 반복횡단 조사자료이다(김지범 외 2017). 최근에는 응답자의 거주지를 행정동 수준에서 비공개로 제공하고 있다. 한국종합사회조사에서는 행정동 자료(지역코드 자료)를 제공하기 전에, 2003년부터 2016년까지 행정동을 행정안전부 홈페이지에서 제공하는 ‘2015년도 지방자치단체 행정구역 및 인구 현황’(이하 ‘현황’)을 기준으로 조화화(harmonization) 작업을 거쳤다. 이때 행정동의 변화로 인하여 기존 KGSS 자료의 행정동이 ‘현황’의 행정동으로 1:1 매칭되지 않는 경우는 가구 주소를 이용하여 행정동을 새롭게 부여하였다.

본 연구는 한국종합사회조사 비공개 데이터에 접근하기 위해 담당 기관인 성균관대학교 동아시아학술원 서베이리서치센터에 소재 대학의 IRB 승인서, 타인과의 자료 공유 금지 확인서를 제출하여 응답자의 거주지역 정보를 받을 수 있었다. 그리고 전국을 대상으로 하는 한국종합사회조사 누적 응답자 중 서울 거주자만을 선택하여, 연도와 행정동을 기준으로 서울시 지역 수준 자료를 연계하였다($N=2,656$). 결합한 자료의 연도별 수집 항목은 <표 1>과 같다. 변수별로 활용할 수 있는 연도가 달라서, 개인 수준 자료 및 연계 자료의 경우 2005년까지와 2015년을 제외하고 2006년부터 2016년까지 10개년도로 구성되었다.

<표 1> 결합된 자료의 연도별 정보

변수	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
한국종합사회조사(KGSS)															
주민등록인구															
혼인 및 이혼 건수															
인구 이동															
기초생활수급자 수															
출생자 및 사망자 수															
대통령 선거 투표/정당 지지율															
총 선거 투표/정당 지지율															
지방선거 투표/지지율															
신고된 아파트 거래가격															
다출처 연계 자료															

각각의 자료 소개에 이어 마지막으로는 다수준 다출처 연계 자료를 구축하면서 겪는 어려움과 장점 모두를 간략하게 기술하고자 한다. 여러 가지 출처의 데이터를 연결시킬 때에는 무엇보다 중요한 일은 연계의 핵심인 행정동을 고유하고 안정적으로 유지하는 것이다. 만일 두 자료에서 사용하는 행정동이 일치하지 않는다면 서로 다른 지역을 같게, 혹은 같은 지역을 다르게 취급하는 문제가 발생한다. 그런데 행정동의 특성상 시기에 따라 지역의 경계가 바뀌기 때문에, 이를 선거 투표율 자료나 아파트 가격과 연계하기 위해서는 시간이 지나도 안정적으로 유지되도록 행정동을 조화화해야 한다. 본 연구에서는 통계분류포털에서 제공하는 2017년도 한국행정구역분류 자료를 기준으로 하되, 이전 연도의 행정동이 2017년도와 연계될 수 있도록 다음의 다섯 가지 방법을 이용하였다. 첫째, A동과 B동이 C동으로 합동된 경우에는 A동과 B동의 변수를 합산하여 산출했다. 둘째, A동이 B동과 C동으로 분동된 경우에는 지도에 근거해 가장 유사한 지역으로 대체하거나, 비율이 비슷할 경우 절반씩 분할했다. 셋째, A동이 B동으로 개칭된 경우에는 A동의 변수를 그대로 입력했다. 넷째, A지역에 있던 B동이 C지역에 있는 D동이 된 경우에는 기존의 변수를 입력했다. 다섯째, 합동 및 분할 과정에서 인구가 1,000명 이하가 된 행정동은 이상수치가 될 수 있다고 판단해 결측값 처리했다. 또한 행정동의 변경 내역이 생길 때마다 저장하여, 선거 자료와 서울시 행정 자료 전체에 대해 동일하게 적용할 수 있

도록 하였다. 이러한 방법들을 이용하면 동 수준에서 누락되는 값은 최소화해도 다소 부정확할 수 있음을 유의해야 한다.

이러한 어려움에도 행정 자료를 사용하는 이점은 연계를 위한 보조 자료가 풍성하다는 점이다. 앞서 언급한 한국행정구역분류 이외에도, 행정 자료를 이용함으로써 번지, 본번, 부번 등 특정 주소를 행정동으로 변환, 거래 건별 아파트 가격으로부터 아파트 평균 가격을 산출할 수 있었다. 정리하자면 다출처 자료인 만큼 형식적으로 다른 형태로 저장될지라도, 행정 정보인 만큼 여러 가지 보조 자료를 이용하여 원하는 식별자로 변환하기 쉽다.

2. 변수 구성

본 연구는 향후 연구에 도움이 되기 위한 초기 탐색의 성격을 지니기 때문에 가능하면 대부분의 변수를 별다른 가공 없이 사용하였다. 우선 지역 단위 정보로는 출생률, 지역의 사회적 자원을 측정할 수 있는 65세 이상 고령자 인구 비율(Cagney 2006), 투표율(Han et al. 2013)이 있으며, 사회 해체 이론을 기초로 지역 효과를 살펴볼 때 이용하는 경제적 변수(Shawn & McKay 1969; Sampson et al. 1997)로는 기초생활수급자 비율과 신고된 아파트 평균 거래가격을 사용했다. 개인 단위 정보에서는 가구소득 수준, 주관적 정치 성향, 주관적 계층 수준, 타인에 대한 신뢰를 사용하였다. 지역 단위 정보에서 출생률, 65세 이상 고령자 인구 비율, 기초생활수급자 비율은 서울열린데이터광장에서 출생아 수, 고령자 수, 기초생활수급자 수를 전체 인구로 나눈 것이다.

한편 신고된 아파트 평균 거래가격은 서울부동산정보광장에서 취합한 아파트 거래 건수 자료로부터 수집하였다. 서울부동산정보광장은 거래된 아파트를 건수별로 ‘시·군·구, 번지, 본번, 부번, 건물명, 계약일, 신고일, 거래금액(만원), 건물전용면적(m²), 층’으로 정리하여 공개하고 있다. 특정 주소로부터 시·군·구와 행정동 기준으로 아파트 가격을 종합(aggregate)하기 위해 공개된 행정동 구획 자료를 이용하여 주어진 시·군·구, 번지, 본번, 부번 정보를 시·군·구와 행정동으로 변환시켰다. 그리고 같은 시·군·구 또는 행정동에 속한 거래 건수를 묶어, 아파트 평균 거래가격 변수를 얻을 수 있었다. 본래 자료는 만 원을 기준으로 가격을 표시하였으나, 본 연구에서는 가독성을 위해 억 원을 기준으로 조정하였다.

투표율의 경우에는 선거별 투표율을 통합한 일반 투표율과 정당별 지지율로, 중

양선거관리위원회에서 제공하는 행정동별 투표 정보를 수집하였다. 이때 일반 투표율은 대통령 선거(2007년, 2012년, 2017년), 국회의원 선거(2008년, 2012년, 2016년), 지방 선거(2006년, 2010년, 2014년)를 각각 치러진 연도에 따라 통합한 것이다. 2012년의 경우 대통령 선거와 국회의원 선거가 모두 진행되었으므로 편의상 대통령 선거로 통합하여 사용하였다. 또한 특정 후보자에 의한 투표율 등락을 최대한 배제할 수 있도록 국회의원 선거의 경우 비례대표국회의원 선거 자료를, 지방 선거의 경우 광역의원비례대표의원 선거 자료를 사용하였다.

한편 개인 단위 자료, 즉 한국종합사회조사에서는 매년 반복 조사되는 질문들을 선택하여, 가구소득 수준, 주관적 정치 성향, 주관적 계층 수준, 타인에 대한 신뢰 변수를 사용하였다. 가구소득 수준은 세금공제 이전 기준으로 월평균 가구소득을 만 원 단위로 묻은 뒤 50만 원씩 구간을 나눠 0부터 21까지 범주형으로 재가공한 변수로, 소득 없음을 0으로, 천만 원 이상을 21로 코드화했다. 주관적 정치 성향의 경우 5점 척도인데, 1에 가까울수록 진보적, 5에 가까울수록 보수적이라고 답했다. 주관적 계층 수준은 한국 사회의 최하층을 1로, 최상층을 10으로 나눌 때 본인이 어느 정도 수준에 속하는지 물어 1부터 10까지 코드화했다. 마지막으로 타인에 대한 신뢰에 대해서는 말 그대로 다른 사람을 얼마나 신뢰할 수 있는지 물어 4점 척도로 측정되었는데, 0부터 3까지 재코딩하여 3에 가까울수록 신뢰가 높도록 나타나게 하였다.

3. 분석모형

본 연구가 결합한 행정 자료는 동 지역 수준의 자료인 동시에 연도를 단위로 하는 패널 자료이기도 하다. 그리고 이렇게 결합된 동 단위 패널 자료는 다시 반복 횡단면 자료(repeated cross-sectional data)인 한국종합사회조사 누적 자료와 응답자의 거주 동을 기준으로 결합하였다. 이러한 최종 자료의 특성 때문에 자료의 출처별로 주의할 점은 다음과 같다.

첫째, 중앙선거관리위원회에서 수집한 변수들은 연도 간 분산이 매우 크다. 예를 들어 투표율은 그 해 실시된 선거가 대통령 선거인지 지방 선거인지 등에 영향을 많이 받고, 같은 대통령 선거라고 해도 그 해의 정치적 사건이나 이슈에 따라 투표율이 많이 달라진다. 따라서 지역 간 투표율의 차이를 살펴보기 위해서는 이러한 연도 간 차이를 통제할 필요가 있다.

둘째, 서울열린데이터광장에서 수집한 일부 변수들은 연도에 따라 일정한 경향을 보인다. 예를 들어 지역별 출생률이나 사망률은 꾸준히 감소하고 고령자 인구 비율은 꾸준히 증가한다. 이러한 변수들은 지역 간 차이를 분석할 때 연도에 따른 경향을 통제하는 것이 필요하다.

셋째, 한국종합사회조사에서 결합한 응답자는 서울시의 모든 동에 분포하지 않으며 연도별 동 분포도 다르다. 행정 자료와 결합 후, 2006년부터 2016년까지 2015년을 제외한 10개년도 누적 응답자의 거주 동은 총 183개다. 어떤 동에서는 총 5개 연도에 걸쳐 47명의 응답자가 조사되었지만, 어떤 동은 단 두 명만 조사되었다. 평균적으로는 여러 해에 걸쳐 동별로 14.5명이 조사되었다. 결과적으로 특정 변수, 예를 들어 응답자의 소득 수준이 동 내에서 보이는 분산에는 연도 간 변이가 혼합되어 있고, 동 간에 보이는 분산 역시 연도 간 차이가 혼합되어 있다. 따라서 연도 간 차이를 통제하고 지역(즉, 동) 간 차이를 분석해야 한다.

지금까지 살펴본 세 가지 문제에 모두 대응하기 위해서는 연도 간 차이를 가능한 있는 그대로 통제하는 것이 필요하다. 따라서 특정 변수에 대해 총 분산을 지역 간 분산과 지역 내 분산으로 분해할 때, 연도별 터미를 투입하여 연도 고정효과(year fixed-effect) 모형을 활용할 것이다. 이러한 연도별 고정효과는 선거 성격의 연도 간 차이(즉 대통령 선거인지 지방 선거인지 등)에 따른 투표율 차이, 사회인구학적 변수의 연도에 따른 경향성, 한국종합사회조사 변수에서 나타나는 지역 간 차이와 혼합된 연도 간 차이를 통제해 준다. 물론 연도 고정효과는 모든 지역에 동일한 정도로 적용된다고 가정하기 때문에, 지역 간 차이에 반영된 연도 간 차이를 완벽히 분리해 낸다고 볼 수는 없을 것이다.

III. 분석결과

한국종합사회조사와 결합한 자료를 분석하기 전에 우선 동 단위에서 연도별로 수집한 지역 자료의 특성을 알아보고자 한다. 이러한 지역 자료는 동 수준 패널 자료라 할 수 있는데, 다섯 개의 변수별로 구 수준 패널 자료와 비교·분석하여 <표 2>에 정리하였다. 우선 변수별로 구와 동 단위에서 기술통계를 제시한 후, 각 변수에 대해 구/동 간 표준편차, 구/동 내 표준편차를 추정하였다. 구/동 간 표준편차는 각

변수를 종속변수로 하는 회귀분석에서 구/동 단위의 무작위 효과(random effects)로부터 추정했고, 구/동 내 표준편차는 잔차에서 추정했다. 연도 고정효과의 추정치는 표에서 생략하였다.

<표 2> 구와 동 수준 변수의 분산 분해 분석 · 비교

변수	출생률		고령자 인구 비율		기초생활수급자 비율		투표율		평균 아파트 거래 가격(억 원)	
	구	동	구	동	구	동	구	동	구	동
단위 평균	.0083	.0083	.0999	.0998	.0204	.0209	.6067	.5931	4.517	4.397
표준편차	.0012	.0026	.0267	.0312	.0080	.0195	.1020	.1144	1.903	2.630
최소값	.0042	0	.0524	.0218	.0063	0	.4227	.2632	1.889	.355
최대값	.0114	.0410	.1723	.2497	.0463	.2347	.8080	.8650	12.69	42
단위 간 표준편차	.0009	.0020	.0142	.0193	.0072	.0185	.0155	.0412	1.777	2.296
단위 내 표준편차	.0004	.0015	.0034	.0087	.0021	.0064	.0119	.0209	.4209	1.122
단위 수	25	424	25	424	25	424	25	424	25	424
사례 수	325	5,077	375	6,344	375	4,518	225	3,383	300	5,028

주: 모든 모형에 연도 고정효과 포함됨.

우선 변수별로 기술통계를 살펴보자. 평균값은 구 단위와 동 단위 간에 거의 동일한 반면, 표준편차나 최소값과 최대값 사이의 간격은 동 단위에서 큰 것을 볼 수 있다. 동별 평균 출생률은 약 0.8%로 이는 전체 인구 100명당 출생아가 0.8명이라는 것을 의미한다. 출생률이 0인 동이 있었던 반면, 4.1%인 동도 있었다. 동별 65세 이상 고령자 인구 비율은 평균이 10.0%로, 100명 중 10명 정도가 고령자인 것으로 나타났다. 고령자 인구 비율은 출생률보다 편차가 훨씬 커, 2%부터 25%까지 상대적으로 큰 범위 안에 분포했다. 기초생활수급자는 한 동에 평균 2.1% 정도 있으나, 동 간 격차가 커서 특정 연도에 한 명도 없는 동이 있었던 반면, 23.5%인 동도 있었다. 동별 평균 투표율은 59.3%인데, 최고 86.5%까지 올라간 경우는 2017년 대통령 선거 시기였다. 동별 아파트 거래가격 평균은 약 4억 3,970만 원이었는데, 이러한 거래가격은 역시 동 간 편차가 매우 커서, 최소 3550만 원부터 최대 42억까지 분포했다. 서울부동산정보광장에서 수집한 자료가 아파트만을 포함하는 데다 특정

연도에 실제 신고된 거래 건수가 적을 수 있어 더욱 극단적인 값을 보이는 것으로 이해할 수 있다.

구 단위보다 동 단위에서 변수의 표준편차가 큰 것은 연도 고정효과를 통제한 후 지역 간 편차와 지역 내 편차로 나누어 추정한 결과에서도 나타난다. 모든 변수에 대해 지역 간 편차와 지역 내 편차 모두에서 동 수준 자료가 큰 값을 보인다. 이러한 결과는 구보다는 동 단위에서 지역 변수를 수집했을 때, 사례 수도 17배 가량 증가(424개 동 대 25개 구)하고 변수의 변이(variation)가 커서 통계적 추론에 활용할 때 더 큰 검정력을 보일 가능성이 높음을 뜻한다. 이제 이러한 동 수준 변수들의 효과를 추정해볼 개인 수준의 종속변수들을 한국종합사회조사 자료에서 검토해 보자.

한국종합사회조사에서 선택한 네 개의 응답변수들에 대해 기술통계와, 연도 고정효과 및 구/동 무작위 효과를 투입한 회귀모형을 수행한 결과를 <표 3>에 정리하였다. 각 종속변수에 대해 구 단위와 동 단위에서 각각 회귀모형을 추정하였다. 우선 기술통계에서 가구소득과 주관적 계층은 값이 클수록 높은 소득이나 상위 계층을 뜻한다. 주관적 정치 성향은 값이 클수록 보수적임을 뜻하며, 타인에 대한 신뢰는 값이 클수록 신뢰함을 뜻한다. 기술통계를 보면 가구소득이나 주관적 계층에 비해 주관적 정치 성향과 타인에 대한 신뢰는 각각 5점, 4점 척도로 측정되어 회귀분석을 수행하기에 적절하다고 보기 힘들다. 그러나 앞서 언급했듯이 본 연구는 특정 종속변수에 대해 정교한 설명적 모형을 구성하는 것이 목적이 아니라, 동 수준 분석의 유용성을 변수 간에 비교하며 탐색하는 것이 목적이기에, 4개의 변수에 대해 일괄적으로 회귀모형을 적용하였다. 타인에 대한 신뢰는 다른 반복 문항에 비해 조사한 연도 수가 적어 응답자 수 및 응답자의 주거 동 수가 적다.

구 수준과 동 수준의 분산을 살펴보면, 주관적 정치 성향을 제외한 세 변수 모두에 대해 구 간 분산보다 동 간 분산이 뚜렷이 큰 것을 알 수 있다. 주관적 정치 성향의 경우는 동 간 분산이 구 간 분산의 약 1.5배, 주관적 계층 수준의 경우는 약 2배 가량 된다. 즉 한국종합사회조사 응답자의 태도나 의식이 구 간 차이보다 동 간 차이가 큰 경향이 있음을 알 수 있다. 이러한 결과는 응답자의 태도나 의식이 구 수준보다는 동 수준 지역 변수로 설명할 여지가 크다는 것을 암시한다. 전체 분산에서 지역 간 분산이 차지하는 비율(ICC)을 비교해 보더라도 모든 변수에 대해 구 간 분산보다는 동 간 분산이 차지하는 비율이 높다. 동 수준 ICC 값이 많게는 약 4.8배(타인에 대한 신뢰), 적게는 약 1.5배(주관적 정치 성향)가 된다. 그러나 주관적 정치 성향의 경우 동 간 분산은 통계적으로 유의하지 않고, 구 간 분산도 5% 수

준에서만 유의하여, 지역 간 분산을 가정하지 않는 OLS 모형으로 예측해도 충분할 것으로 보인다.

모형의 적합도를 구 수준 모형과 동 수준 모형 간에 비교하기 위해, 로그 우도(log likelihood), AIC, BIC 세 지표를 표에 제시하였다. 주관적 정치 성향을 제외한 세 변수 모두에 대해서, 동 수준의 모형에서 적합도 지표들이 구 수준 모형에서보다 절댓값이 작았다. 즉, 동 수준 지역 효과 모형 적합도가 구 수준의 모형보다 좋다고 할 수 있다(Burnham & Anderson 2002; Kutner et al. 2005). 정리하자면, KGSS에서 추출한 네 종속변수 중 지역 간 차이가 불분명한 주관적 정치 성향을 제외한 세 변수 모두에서 구 단위보다는 동 단위에서 지역 간 차이를 더 잘 드러낸다고 할 수 있다.

<표 3> 한국종합사회조사 변수의 분산 분해, 구 수준과 동 수준 분해

변수	가구소득 수준	주관적 정치 성향	주관적 계층 수준	타인에 대한 신뢰
평균	9.1587	3.0129	4.8695	1.3684
표준편차	5.6438	.9882	1.7137	.7080
최소값	0	1	1	0
최대값	21	5	10	3
구 간 분산 추정치	2.8568***	.0063*	.2786***	.0028
구 간 분산의 비율(ICC)	.0896	.0066	.0937	.0056
Log likelihood	-7847.1581	-3598.0503	-5089.7628	-1887.4748
AIC	15718.32	7220.101	10203.53	3792.95
BIC	15788.30	7290.311	10274.09	3842.24
구 수	24	24	24	24
동 간 분산 추정치	5.0732***	.0091	.5543***	.0134**
동 간 분산의 비율(ICC)	.1603	.0094	.1871	.0270
Log likelihood	-7818.4123	-3600.008	-5038.8887	-1884.8922
AIC	15660.82	7224.016	10101.78	3787.784
BIC	15730.81	7294.227	10172.34	3837.078
동 수	183	183	183	160
사례 수	2,520	2,568	2,644	1,767

주: 1. 모든 모형에 연도 고정효과 포함됨.

2: + $p < .10$, * $p < .05$; ** $p < .01$; *** $p < .001$.

그렇다면 네 응답변수 중 어떤 변수들이 동 단위 지역 효과를 드러낼 가능성이 높을까? 가구소득 수준이나 주관적 계층 수준 모두 전체 분산에서 동 간 분산이 차지하는 비율이 상당히 높아(약 16~19% 사이), 이러한 종속변수는 동 간 차이에 의해 설명될 가능성도 높으며 다수준 분석을 할 만한 좋은 후보들이라 할 수 있다. 반면에 주관적 정치 성향이나 타인에 대한 신뢰 변수는 동 간 분산의 비율이 작아(각 0.94%와 2.70%) 서울 주민의 주관적 정치 성향은 동 간 이질성이 크지 않다고 할 수 있다.

이상의 추론을 검증해보기 위해 <표 3>과 동일한 네 개의 종속변수에 대해 개인 수준과 지역 수준 변수를 추가한 다수준 분석모형의 결과를 <표 4>에 정리하였다. 응답자 수준에서는 성별, 연령, 교육수준을 통제하였다. 교육수준은 7점 척도이며 최저 수준은 무학, 최고는 대학원 졸업이다.

지역 수준에서는 평균 아파트 거래가격과 고령자 인구 비율을 독립변수로 선택하였는데, 이 변수들은 <표 2>에서 살펴본 지역 수준 변수들 중 네 개의 종속변수에 대해 독립변수로 투입했을 때 가장 강한 효과를 보인 평균 아파트 거래가격을 우선 선택하고, 이와 함께 투입했을 때도 통계적으로 유의한 효과를 보인 고령자 인구 비율을 추가하였다. <표 4>는 이 변수들의 효과를 구 수준과 동 수준 각각에서 추정된 모형(모형 1과 모형 3)을 비교하였으며, 더불어 동 수준에서는 지역 수준 변수가 없는 모형(모형 2)도 제공하여, 지역 수준 변수를 추가함에 따라(모형 3) 증가한 설명력이 어느 정도인지 평가할 수 있도록 하였다.

응답자 수준과 지역 수준 변수의 모든 회귀계수는 고정효과로 추정하였고, 이러한 고정효과를 추정하고 남은 구와 동 수준의 지역 효과는 무작위 효과로 추정하였다. 이러한 무작위 효과의 분산으로 구와 동 간 분산을 추정하였다.

탐색적 분석 과정을 거쳐 선택한 지역 변수이지만 평균 아파트 거래가격과 고령자 인구 비율은 이론적으로도 주민 의식에 다양하게 영향을 미칠 수 있다. 우선 서울에서 지역 부동산 가격이 주민의 생활과 의식에 미치는 영향은 상당히 클 것으로 예상된다. 김항기 외(2017)는 자산 가격이 높을수록 재분배를 선호하는 태도를 보인다고 분석하였으며, 지역별 지가 변동, 주택 규모, 아파트 비율, 자가 소유 여부가 정치적 참여나 특정 정당 지지로 이어진다는 일련의 연구 또한 발표되었다(박원호 2009; 백두진 외 2014; 한석준 2012). 둘째로, 현재 지역의 평균 소득 수준이나 소득 지니계수 등은 공개되어 있지 않은 상황에서 아파트 거래가격은 그 지역의 경제적 수준에 대한 합리적인 대리변수 역할을 해줄 수 있다.

한편 지역의 고령자 인구 비율은 행정 자료에서 추출할 수 있는 대표적 인구학적 변수일 뿐만 아니라, 생산 활동과 유동 인구, 경제적 수준 등 지역의 특성을 복합적으로 반영할 수 있는 변수다. 또한 고령자 인구 비율은 동별로 최소 2.2%에서 최대 25.0%까지 분포하여, 다른 지역 수준 변수인 출생아 수 비율과 기초생활수급자 비율 등에 비해 이질성이 크고 비대칭도(skewness)가 작았다. 한편 다른 동별 지역 변수와의 상관성을 검토했을 때, 높은 고령자 인구 비율은 높은 기초생활수급자 비율(0.621), 높은 장애인 비율(.605), 낮은 세대당 인구수(-.390), 낮은 출생아 수 비율(-.479)과 상관성을 보여 여타 인구통계학적 특성도 두루 반영하는 듯하다. 참고로 고령자 인구 비율과 평균 아파트 거래가격의 상관계수는 -.141로 나타났다.

<표 4> 한국종합사회조사 변수에 대한 구와 동 단위 탐색적 다수준 모형 비교

변수	가구소득 수준			주관적 정치 성향		
	모형 1 (구 단위)	모형 2 (동 단위)	모형 3 (동 단위)	모형 1 (구 단위)	모형 2 (동 단위)	모형 3 (동 단위)
여성(=1)	.7300*** (.2013)	.7035*** (.2001)	.6371** (.1995)	.0217 (.0389)	.0243 (.0121)	.0159 (.0390)
연령	-.0225** (.0073)	-.0253** (.0073)	-.0265*** (.0073)	.0118*** (.0014)	.0121*** (.0014)	.0117*** (.0014)
교육수준	1.4910*** (.0841)	1.4296*** (.0844)	1.3627*** (.0851)	-.0046 (.0162)	.0033 (.0159)	-.0071 (.0164)
평균 아파트 거래가격(억원)	.3352** (.1086)		.3378*** (.0738)	.0297+ (.0153)		.0278** (.0113)
고령자 인구 비율(%)	-.0624 (.1575)		-.2374** (.0813)	-.0064 (.0197)		-.0005 (.0121)
구 간 분산	.4692***			.0012		
동 간 분산		2.1526***	1.2790***		.0042	.0029
ICC	.0190	.0856	.0527	.0012	.0045	.0031
Log likelihood	-7458.88	-7455.33	-7434.39	-3480.03	-3483.33	-3479.70
AIC	14951.77	14940.65	14902.79	6994.05	6996.66	6993.41
BIC	15050.59	15027.85	15001.61	7093.19	7084.13	7092.55
구/동 수	19	183	183	19	183	183
사례 수	2,473	2,473	2,473	2,519	2,519	2,519

변수	주관적 계층 수준			타인에 대한 신뢰		
	모형 1 (구 단위)	모형 2 (동 단위)	모형 3 (동 단위)	모형 1 (구 단위)	모형 2 (동 단위)	모형 3 (동 단위)
여성(=1)	.1273* (.0612)	.1030+ (.0604)	.0849 (.0602)	-.0757* (.0341)	-.0753* (.0342)	-.0804* (.0342)
연령	.0001 (.0022)	-.0006 (.0022)	-.0011 (.0022)	-.0004 (.0012)	-.0005 (.0012)	-.0006 (.0012)
교육수준	.4479*** (.0254)	.4174*** (.0255)	.3970*** (.0256)	.0560*** (.0140)	.0553*** (.0139)	.0477*** (.0143)
평균 아파트 거래가격(억원)	.1353*** (.0315)		.1465*** (.0239)	.0123 (.0135)		.0118 (.0097)
고령자 인구 비율(%)	-.0580 (.0456)		-.0543* (.0267)	.0044 (.0174)		-.0169 (.0106)
구 간 분산	.0368**			.0020		
동 간 분산		.2811***	.1717***		.0092*	.0070
ICC	.0154	.1140	.0730	.0041	.0188	.0144
Log likelihood	-4795.09	-4783.46	-4756.48	-1838.05	-1837.15	-1834.11
AIC	9624.18	9596.91	9546.96	3704.09	3698.31	3696.23
BIC	9723.82	9684.83	9646.59	3780.47	3763.77	3772.60
구/동 수	19	183	183	19	160	160
사례 수	2,594	2,594	2,594	1,729	1,729	1,729

주: 1. 모든 모형에 연도 고정효과 포함됨.

2. 괄호 안은 표준오차.

3. + $p < .10$, * $p < .05$; ** $p < .01$; *** $p < .001$.

<표 4>에서 우선 응답자 수준 변수들의 회귀계수를 비교해 보면 (주관적 계층 수준에 대한 여성 더미를 제외하고는) 모형 간 추정치나 유의수준에 차이가 거의 없다. 즉 지역 수준을 구와 동 중 무엇으로 설정하는지(모형 1과 모형 3), 어떤 지역 변수를 투입하는지(모형 2와 모형 3)는 개인 수준 효과 추정에 큰 영향을 미치지 않는다.

다음으로 지역 수준 변수 중 평균 아파트 거래가격의 회귀계수를 모형 1과 모형 3 간에 비교해 보면, 동 수준에서 좀 더 유의하거나(가구소득 수준), 동 수준에서

회귀계수가 좀 더 크거나(주관적 계층 수준), 두 모형 간 차이가 없다(주관적 정치 성향과 타인에 대한 신뢰). 크게 보아 평균 아파트 거래가격은 구 수준에서 효과를 추정하건 동 수준에서 추정하건 큰 차이가 없거나 동 수준에서 미세한 우위를 보인다. 동 수준 효과(모형 3)에 근거해 해석하면, 평균 거래가격이 높은 동에 사는 응답자일수록 가구소득이 높고($p<.000$), 스스로 정치적으로 보수적이라 생각하며($p=.014$), 스스로의 주관적 계층을 높게 평가한다($p<.000$). 그러나 주관적 정치 성향은 <표 3>에서 살펴보았듯이 지역 간 차이가 거의 없기에, 지역의 평균 아파트 거래가격의 효과를 해석하는 데 주의해야 할 것이다. 지역의 아파트 거래가격과 상관관계가 높은 응답자 수준의 설명 변수가 존재할 수 있다.

지역 변수 중 고령자 인구 비율의 경우는 구 수준보다는 동 수준에서 뚜렷한 설명력을 가진다. 주관적 정치 성향과 타인에 대한 신뢰에 대해서는 구 수준과 동 수준 모두에서 유의한 영향력이 없으나, 가구소득 수준과 주관적 계층 수준은 구 수준이 아닌 동 수준에서 유의한 효과를 갖는다. 고령자 인구 비율이 높은 동에 사는 응답자일수록 가구소득이 낮고($p<.004$), 스스로의 주관적 계층을 낮게 평가한다($p<.042$).

지금까지의 분석을 정리해 보면, <표 3>에서 살펴보았듯이 한국종합사회조사 응답변수들의 변이는 구 수준보다는 동 수준의 요인으로 더 잘 설명이 될 것으로 기대하였으며, <표 4>는 이러한 기대를 뒷받침한다. 평균 아파트 거래가격과 같이 어느 수준에서건 응답변수를 잘 설명하는 경우도 있으나, 고령자 인구 비율처럼 동 수준에서만 응답변수를 잘 설명하는 경우도 있다. 구 수준에서만 유의하게 설명하는 지역 변수는 없었다. AIC나 BIC 값을 비교해봐도 구 단위 모형 1보다 동 단위 모형 3이 더 작은 값을 보여, 모든 종속변수에서 동 단위 모형의 적합도가 더 좋은 것을 볼 수 있다.

비록 탐색적 모형이기는 하지만 본 연구에서 채택한 두 개의 동 수준 변수는 종속변수를 비교적 잘 설명하는 것으로 보인다. <표 4>의 모든 종속변수에 대해 모형 2에서 모형 3으로 오면서 동 간 분산이 상당히 감소하였다. 즉, 모형 2의 독립변수들에 의해 설명되지 않고 남아 있는 종속변수의 지역 간 차이가 모형 3에서 지역 변수를 투입한 후에 많이 감소하였다. 물론 종속변수가 타인에 대한 신뢰일 경우는 두 개의 동 단위 변수 모두 유의한 효과를 보이지 못했으며, 구 간 분산(모형 1)이나 동 간 분산(모형 3)의 추정치가 통계적으로 유의하지 않았다. 본 연구에서 탐색적으로 선택한 지역 변수들이 한국종합사회조사의 다양한 응답변수들을 모두 잘 설명하지는 못할 것이다. <표 2>에 제시하지는 않았지만 몇몇 종속변수에 유의한 효

과를 보이는 동 수준 변수들이 있었다. 세대당 인구수, 장애인 비율, 여성 비율, 아파트 거래가격의 표준편차 등이 이러한 동 수준 지역 변수들이었다. 연구주제에 따라 종속변수가 정해진다면 이러한 지역 변수들을 포함하여 다양한 동 수준 변수들의 효과를 좀 더 세밀히 추정하고 해석해 볼 수 있을 것이다.

IV. 토의 및 결어

국내에서는 왜 사회조사 자료를 기반으로 한 지역 효과 분석이 기대만큼 활성화 되지 못하는가? 대한민국은 인구 밀도가 높고 동질적인 사회라 개인에게 미치는 지역 효과가 크지 않기 때문일 수 있다. 반면 거주지역이 실제로 대한민국 국민에게 매우 중요한 의미를 가짐에도, 그동안 주로 활용되어 온 시·군·구 수준의 변수로는 이러한 지역 맥락을 포착하지 못하기 때문일 수도 있다. 본 연구는 후자의 가능성을 탐색하기 위하여, 세 개의 행정 자료와 한국종합사회조사 연계 자료를 결합한 뒤 서울시의 구와 동 수준에서의 지역 변수 효과의 차이를 살펴보았다.

분석의 결과, 공개 자료를 통해 모은 서울시 동 수준 변수들은 구 수준에서보다 큰 지역 간 차이를 보여 다수준 분석에서 지역 간 차이를 반영할 잠재력이 있었다. 또한 한국종합사회조사 내 서울시 거주자들의 응답은 구 간 차이보다 동 간 차이가 더 큰 변화폭을 보여 구보다는 동 수준 지역 효과를 포착할 가능성이 높았다. 지역의 평균 아파트 거래가격과 고령자 인구 비율을 예시로 그 효과를 추정해본 결과, 구 수준보다는 동 수준에서 응답자의 응답변수들을 더 잘 설명함을 관찰하였다.

본 연구가 분석을 위해 지역 수준 변수나 응답자 수준 변수를 선택할 때 구 간 차이보다 동 간 차이가 두드러진 변수들을 의도적으로 선택한 것은 아니다. 그러나 본 연구가 지역 수준의 모든 변수에 대해 구 간 변이와 동 간 변이를 비교하지 않았기에, 구 간 변이가 동 간 변이보다 큰 지역 변수가 존재할 수 있다. 마찬가지로 본 연구는 한국종합사회조사의 모든 응답변수에 대해 지역 간 변이를 분석하지 않았다. 본 연구에서 살펴본 응답자의 응답은 뚜렷한 지역 간 차이를 보이지 않거나, 보인다면 동 간 차이가 구 간 차이보다 컸다. 일부 응답자의 태도는 구 간 차이가 더 클 수도 있다.

이러한 일련의 분석과 제한점은 향후 연구를 위한 이론적, 방법론적 함의를 갖는

다. 우선 이론적으로 본 연구는 사회 해체 이론의 적합성 검증에 필요한 경험적 연구의 발판을 제시한다고 할 수 있다. 사회 해체 이론은 거주지역의 중요성을 강조하였고, 이 이론에 바탕을 둔 서구의 많은 경험적 연구는 지역 특성이 다양한 영역에서 지역에 사는 개인에게 영향을 미친다는 점을 밝혀왔다. 한국에서도 아직 이론적으로 정교화되지는 않았지만 주택 문제로 인한 빈번한 주거지역 이동, 신도시, 아파트 공화국, 젠트리피케이션, 지방 도시의 소멸 등으로 인한 공동체의 붕괴가 지속적으로 논의되고 있다. 지역 효과의 연구가 쌓이고 지역 효과의 기제를 파악한다면, 마을 만들기 사업 등 정책적으로 장려되는 공동체 복원에도 기여할 수 있으리라 예상한다.

이러한 기여를 위해서는 본 연구에서 효과를 탐색해 본 지역 부동산 거래가격이 반영하는 개념이 무엇인지, 지역의 고령자 인구 비율이 무엇인지에 대한 후속 연구와 토론이 필요하다. 본 연구에서 추정한 지역 ‘효과’는 탐색적이고 통제변수들도 간단한 모형에서 추정한 통계적 연관관계를 뜻하며, 이러한 효과가 실질적 효과로 해석되기 위해서는 이론적 초점에 따른 명확한 종속변수와 더 풍부한 자료 및 통제변수 등이 갖추어져야 할 것이다. 거주지역의 부동산 거래가격은 그 지역의 부의 수준을 반영할 수도 있으며, 거주자의 부동산 투기 성향을 반영하거나, 혹은 지역의 치안 정도를 반영할 수도 있다. 한 지역의 고령자 인구 비율은 고령화 정도를 뜻하는데, 서울의 경우 고령화가 심한 지역일수록 평균 소득 수준이 낮거나 주민 간 소득 불평등이 클 수도 있으며, 지역 경제 활동이 침체될 가능성을 생각해볼 수 있다. 지역 단위에서 소득 정보가 잘 공개되지 않는 우리나라에서 지역의 부의 평균 수준이나 불평등 수준을 포착할 수 있는 대리변수(proxy)의 탐색도 의미 있는 이론적 활동이 될 수 있을 것이다.

이러한 이론적 관심은 방법론적 발전도 도모하게 유도할 것이다. 향후 연구에서는 이웃의 경계와 관련하여, 거주자의 주관적 인식에 기초하여 이웃의 경계를 정의할 수도 있고(Coulton et al. 2001), 비슷한 동네 특성을 보이는 몇 개의 동을 묶는 시도도 할 수 있을 것이다(Sampson et al. 1997). 또한 본 연구에서는 서울시 자료와 중앙선거관리위원회 자료만 결합했지만, 구글 스트리트 뷰어(Odgers et al. 2012) 등을 통하여 지역의 물리적 환경을 파악하거나, 공공데이터 포털에서 제공하는 지역의 노인 복지시설, 파출소, 보건소 등 다양한 기관 및 시설을 이웃의 지표로 활용할 수도 있다. 디지털 기기와 센서의 발달에 따라 위치 기반으로 자동 축적될 데이터(passive data)를 고려해 본다면, 지역 수준 정보의 활용 가능성은 무궁무진하다. 특

히 이러한 정보는 행정동 단위보다 더욱 정교한 단위에서 수집되기 때문에, 본 연구에서 아파트 거래가격의 동 수준 평균을 구했듯이, 행정동 수준으로 합산하기 용이하다. 향후 위치 기반 개인정보의 프라이버시 문제를 해결해 나간다면, 다수준 다출처 분석을 통해 동보다도 세밀한 지역이나 주거환경 수준의 효과를 탐색할 가능성은 급격히 증가할 수 있다. 이러한 이웃 경계와 지표에 대한 활발한 연구는 궁극적으로 지역적 특성이 어떻게 해당 지역에 사는 사람들에게 영향을 미치는지에 대한 설명을 가능하게 할 수 있다. 가령 왜 지역에 따라 지역 주민들의 행동 및 태도에 차이가 나는지를 설명하기 위해 지역의 공공 기관 및 복지시설 자원의 영향을 파악할 수 있을 것이다(Leventhal & Brooks-Gunn 2000).

본 연구가 시도한 방법은 다른 지역이나 사회조사 자료로 확대될 수 있다. 한국종합사회조사 응답자를 서울 외에도 잘 디지털화되어 있는 다른 도시의 행정 자료와 결합한다면, 분석의 지역적 범위를 넓혀 외적 타당성을 높일 수 있을 것이다. 또는 서울서베이와 같이 매년 약 2만 가구의 서울 시민을 조사한 자료를 본 연구가 통합한 서울시 행정 자료와 결합한다면, 서울시 주민에게 미치는 동 수준의 지역 효과에 대해 훨씬 풍부한 연구가 가능할 것이다. 이러한 자료의 결합은 개인 정보 보호에 대한 우려를 일으킬 수 있으나, 성균관대학교 서베이리서치센터는 누적 한국종합사회조사에 대해 적절한 비공개 원칙을 세우고 심사를 통해 이미 이러한 자료를 제공하여 본 연구를 가능하게 하였다. 이러한 선구적 시도를 잘 보완하고 확대한다면 축적되어 온 사회조사 자료의 가치를 극대화하고 향후 조사연구의 발전을 위해 일조할 수 있을 것이다.

참고문헌

- 권오상·강혜정. 2013. “농가 소득에 있어 ‘이웃 효과’ 추정.” 《농촌경제》 36(2): 97-113.
- 김교성. 2010. “서울시 가구 빈곤의 원인에 관한 다층 분석.” 《한국사회정책》 17(1): 215-240.
- 김지범·강정환·김석호·김창환·박원호·이윤석·최슬기·김솔이. 2017. 《한국종합사회조사 2003-2016》. 성균관대학교 출판부.
- 김항기·권혁용. 2017. “부동산과 복지국가-자산, 부채, 그리고 복지태도.” 《한국정치학회 한국정치학회보》 51(1): 261-285.

- 김형용·최진무. 2014. “서울시 소지역 건강불평등에 관한 연구.” 《한국지역지리학회지》 20(2): 217-229.
- 김희윤·백학영. 2010. “아동의 학업성취도에 대한 동네효과 분석: 부모의 교육적 관여의 매개효과를 중심으로.” 《사회복지정책》 37(3): 53-84.
- 박원호. 2009. “부동산 가격 변동과 2000년대의 한국 선거: 지역주의 "이후"의 경제투표에 대한 방법론적 탐색.” 《한국정치연구》 18(3): 1-28.
- 백두진·김재태. 2014. “서울시 하위지역 주거용 부동산 특성과 2012년 대선투표 결과에 관한 탐색적 공간분석.” 《부동산학연구》 20(4): 113-130.
- 백정우·최석현. 2013. “지역사회 정치 참여의 구조적 성격 고찰.” 《한국사회학회 사회학대회 논문집》 597-612.
- 심재만. 2017. “서울 어느 동네 누가 건강할까?.” 서우석·변미리·김백영·김지영(편). 《서울사회학》 71-98. 나남.
- 심정하·손미아. 2009. “강원도의 뇌혈관질환 사망과 지역사회 건강 지표와 관련성.” 《농촌의학·지역보건》 34(1): 1-12.
- 이화용. 2016. “이웃효과가 독거노인의 우울에 미치는 영향-이웃유대의 매개 효과를 중심으로-.” 서남대학교박사학위논문.
- 장동호. 2011. “거주지역의 경제적 특성에 따른 대중의 불평등 인식수준과 빈곤관의 차이 분석.” 《한국지역사회복지학》 37: 259-284.
- 탁장한. 2016. “소속동네의 빈곤이 주민들의 이웃관계에 미치는 맥락효과: 동네의 거주불안정성 및 무질서의 영향을 중심으로.” 서울대학교석사학위논문.
- 한보영·강정한. 2015. “한국사회의 고밀도 생활권과 이웃관계가 우울감에 미치는 영향.” 《한국인구학》 38(3): 75-111.
- 한석준. 2012. “지역경제와 투표참여: 기초지자체에 대한 경험적 분석.” 《사회과학연구》 20(1): 210-238.
- Brattbakk, I. 2014. “Block, Neighbourhood or District? The Importance of Geographical Scale for Area Effects on Educational Attainment.” *Geografiska Annaler: Series B, Human Geography* 96(2): 109-125.
- Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach (2nd ed.)*, Springer-Verlag.
- Cagney, K.A. 2006. “Neighborhood Age Structure and Its Implications for Health.” *Journal of Urban Health* 83(5): 827-834.
- Coulton, C.J., J. Korbin, T. Chan, and M. Su. 2001. “Mapping Residents’ Perceptions of Neighborhood Boundaries: A Methodological Note.” *American Journal of Community Psychology* 29(2): 371-383.

- Ginther, D., R. Haveman, and B. Wolfe. 2000. "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust Are the Relationships?." *The Journal of Human Resources* 35(4): 603-642.
- Han, S., H. Kim, E.S. Lee, and H.S. Lee. 2013. "The Contextual and Compositional Associations of Social Capital and Subjective Happiness: A Multilevel Analysis from Seoul, South Korea." *Journal of Happiness Studies* 14(4), 1183-1200.
- Jensen, B. and M. Harris. 2008. "Neighbourhood Measures: Quantifying the Effects of Neighbourhood Externalities." *The Economic Record* 84(264): 68-81.
- Kutner, M.H., C.J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied Linear Statistical Models* (5th edition). Boston, Massachusetts: McGraw-Hill.
- Leventhal, T. and J. Brooks-Gunn. 2000. "The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes." *Psychological Bulletin* 126(2): 309-337.
- Manski, C. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3): 531-542.
- Odgers, C.L., A. Caspi, C.J. Bates, R.J. Sampson, and T.E. Moffitt. 2012. "Systematic Social Observation of Children's Neighborhoods Using Google Street View: A Reliable and Cost-effective Method." *Journal of Child Psychology and Psychiatry* 53(10): 1009-1017.
- Sampson R., S. Raudenbush, and F. Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science New Series* 277(5328): 918-924.
- Shaw, C.R. and H.D. McKay. 1969. *Juvenile Delinquency and Urban Areas*. The University of Chicago Press.
- Tom W. Smith and Jibum Kim. 2013. "An Assessment of the Multi-level Integrated Database Approach." *The ANNALS of the American Academy of Political and Social Science* 645:185-221.
- Wirth, L. 1938. "Urbanism as a Way of Life." *American Journal of Sociology* 44: 3-24.

Exploring the Relationship between Administrative Units and Neighborhood Effects in Seoul: Linking Korean General Social Survey with Three Administrative Data Sources

Eehyun Kim
Jibum Kim
Jeong-han Kang

Studies on neighborhood-level effects in South Korea have been focusing on the level of gu, which may fail to represent regions as living environments because goon/gu have large within-region variation. This study zooms into the level of dong, which is much finer than gu and may revive research on neighborhood-level effects in South Korea.

We selected Seoul residents from a dataset of Korean General Social Survey (KGSS, 2003-2016) and matched them to dong-level information from three administrative data sources: Seoul Open Data Plaza (2003-2017), National Election Commission (2006-2017), and Seoul Real Estate Information Plaza (2006-2017) ($n=2,656$).

Variance analysis of selected variables from KGSS and administrative data confirmed that between-dong variations are larger than between-gu variations and thus suggests that dong-level estimations are more promising than gu-level estimations of neighborhood effects on residents. We estimated how much Seoul citizens' various attitudes were affected by the average price of apartment transactions and the percentage of the older population at the dong level and compared them with those at the gu level. Estimates at the dong level tended to be larger and more significant than those at the gu level. Controlling for gender, age and education levels, respondents from dongs with higher percentage of older people reported lower household income and lower self-appraised class, which were not observed in the same analysis with gu-level variables. By conducting multi-level, multi-source analysis, we showed the analytic potential of combining dong-level information with social surveys from large cities in South Korea.

Key words: KGSS, Seoul, neighborhood effects, dong as administrative district, real estate, multi-level analysis, multi-source linkage, between-dong variance