

공공빅데이터를 활용한 1인당 주거면적 추정에 관한 연구

- 서울의 단독 및 다세대 주택을 중심으로 -*

임재빈** · 이상훈***

A Study on Estimating Housing Area per capita using Public Big Data - Focusing on Detached houses and Flats in Seoul -*

Lim, Jae-Bin**, Lee, Sang-Hoon***

국문요약 본 연구는 건축물대장, 주민등록대장 등 공공빅데이터의 활용성을 탐구하기 위해, 비교적 간단한 구조를 가진 맨큐-와일(MW)모형을 활용해 1인당 주거면적 추정을 시도하였다. 그 결과 공공빅데이터를 활용하여 정기조사 방식에 버금가는 모형을 수립할 수 있고, 기존 정기조사 방식으로는 어려웠던 기초자치단체별 모형수립도 가능성을 확인할 수 있었다. 공공빅데이터로부터 일반단독주택과 다세대주택 샘플을 판별하는 과정을 설계하여, 10세 연령대별 1인당 주거면적을 추정하고, 인구주택총조사, 주거실태조사 등 기존 정기조사 자료를 활용한 결과와 비교해 일치시킨 후, 서울시 25개 자치구별 1인당 주거면적을 도출하였다. 공공빅데이터는 지식영역을 확장시켜주는 장점이 있지만 본래의 작성의도와 다른 목적으로 생성된 자료를 활용한다는 점에서 근본적 한계는 존재한다. 또 개인정보 접근이라는 어려운 과정은 분석을 보다 신중히 진행해야 하는 부담을 주고, 비식별화를 거친 자료를 분석함에 따라 연구설계가 어려워지는 문제도 있다. 향후 공공빅데이터가 기존 통계조사를 보완하거나 대체할 수도 있도록 가공하는 방법 등에 대한 꾸준한 연구가 필요할 것으로 보인다.

주제어 공공빅데이터, 맨큐-와일 모형, 주거면적

Abstract: The purpose of this study is to estimate the housing area per capita for verifying if the public Big Data, of the building ledger and resident registration ledger, can be used as well as the National Census and Housing Survey. The Mankiw and Weil (MW) model was constructed by extracting samples of general detached houses and flat houses from the public big data, and compared with the result from traditional survey method. Then, the MW models of 25 municipalities in Seoul was established. As a result, it can be confirmed that it is possible to establish MW

* 본 논문은 한국지역학회 2019 후기학술대회 발표 논문을 일부 수정, 보완한 것임

** LH 토지주택연구원 수석연구원(주저자: jb.lim@lh.or.kr)

*** LH 토지주택연구원 수석연구원(교신저자: icarus@lh.or.kr)

models comparable to regular surveys using public big data, and to establish a model for each basic localities which was difficult to use as a regular survey method. Public Big Data has the advantage of expanding the knowledge frontier, but there are some limitations because it uses data generated for other original purposes. Also, the difficult process of accessing personal information is a burden to carry out analysis. It is expected that continuing research should be needed on how public Big Data would be processed to complement or replace traditional statistical surveys.

Key Words: Public Big Data, Mankiw and Weil Model, Housing Area

1. 연구의 배경

건축물대장과 같은 대규모 행정자료에 초고속 추출·분석기술이 접목되면서, 공공빅데이터로 명명된 새로운 지식 영역이 확장하고 있다(성옥준, 2017). 최근 주목받고 있는 빅데이터 기술은 대량의 정형, 비정형(구술, 음성, 영상 등) 자료를 컴퓨터 및 네트워크를 통해 신속하게 처리·분석한다(조영임, 2013). 특히 주택시장은 수백만 명에 이르는 시민들이 매일 주거 선택과 매매, 임대, 매물에 대한 평가 등을 반복하며, 정보 수요의 크기와 성격이 변화무쌍하기 때문에 빅데이터의 활용성이 크다(경기연구원, 2015).

공공빅데이터는 거의 전 분야를 망라하기 때문에, 행정 영역에 따라 구체적인 정의와 활용 방법이 다양하다(김기환, 2013; 성옥준, 2017). 통계청은 2015년 인구주택총조사부터 행정 빅데이터를 활용한 등록센서스를 도입하였고, 기본적인 전수 조사항목은 주민등록대장, 건축물대장 등 11개 기관의 21종 행정자료를 활용하게 되었다. 덕분에 기존 5년마다 5,000만 명을 대상으로 실시하던 가구 방문 조사를 1,000만 명 대상으로 줄일 수 있었으며, 조사비용도 절반으로 감소하였다(이상훈 외, 2016). 또, 정부는 국가중점데이터 개방전략에 따라 2019년까지 64개 분야 436개 공공빅데이터를 공공데이터포털을 통해 제공하고 있다(윤상오·현지우, 2019).

만큐-와일(Mankiw-Weil) 모형(또는 'MW 모형')은 연령별 1인당 주거면적 등의 주택수요를 추정함으로써 미래 주택수요를 예측할 수 있는 모형이다(Mankiw & Weil, 1989). 가구별 구성원들의 연령과

그들의 주거면적만 알고 있어도 가격변화에 의한 자본 이득, 조달금리, 주택 관련 세제 등 변동 요인을 배제하고 기본적인 모형 구축이 가능하다. 이런 간결성 덕에 MW모형은 공공빅데이터 활용성 모색에 우선적으로 고려하기에 적합하다(장재일·안건혁, 2008).

본래 MW모형은 자료구축에 긴 시간과 큰 비용이 필요해 조사에 따라 1-2년이 소요되고, 충분한 샘플 확보도 어렵다는 문제가 있다(신미림·남진, 2011; 이경민 외, 2012). MW 모형 수립을 위해 널리 활용되는 국토교통부의 주거실태조사가 2017년부터 확대되어 매년 전국 6만여 가구를 조사하고 있고, 광역시도 차원에서 별도 조사를 실시하는 경우도 있으나, 여전히 기초자치단체 별로 분석할 경우 샘플이 1,000개도 안 되는 경우가 많다. 여기에 단독주택, 공동주택 등으로 주택의 종류를 한정하면, 샘플이 수백 개조차 확보되지 않는다.

대규모 행정자료인 건축물대장과 주민등록대장은 주거상황을 실시간에 가깝게 반영하는 전수 자료로, 기초자치단체 단위로 분석하거나 주택의 종류를 세분하였을 경우에도 통계적으로 충분한 샘플을 확보할 수 있다. 또, 본래 MW 모형이 구조적으로 주거 수요(demand)보다는 주거 수요(needs), 즉 각 개인이 주거 면적을 얼마나 차지하는지를 확인하는 모형이라는 점에서(장재일·안건혁, 2008), 실시간 자료를 활용함으로써 모형의 활용도를 크게 높일 수 있다. 국토교통부는 2017년부터 주거실태조사에 건축물대장을 참고하고 있다.

본 연구의 목적은 공공빅데이터인 건축물대장과 주민등록대장을 활용하여 MW모형 구축 일괄처리화 과

정을 수립하고, 그 결과가 기존의 정기조사(인구주택총조사, 주거실태조사) 자료를 사용한 모형과 차별화된 효용성이 있는지 확인하는 것이다. 물론 전문적인 통계 보정이 이뤄지는 정기 조사보다 높은 신뢰도를 기대하기는 어렵다. 그러나 기존 방식의 결과물에 비교해 수용 가능한 수준의 모형이 도출될 수 있다면, 또 기존 방식으로는 불가능했던 자료의 일괄처리 수시 활용과 세밀한 분석이 가능하다면, 공공빅데이터의 활용성을 확대할 수 있는 기반을 마련할 수 있을 것이다.

2. 선행연구

1) 빅데이터

빅데이터는 대용량 데이터를 분석하여 가치 있는 정보를 생성하고 이를 활용하는 정보화 기술을 포괄한다. 데이터의 자원화(mining), 가공(integration), 분석(analysing)의 3가지 분야의 전략이 필요하다. 방법적으로는 데이터 수집-정제-분석-정보화-활용의 5 단계를 거친다(장혜정, 2018). 데이터 분석 방법에 따라 상황인식, 의사결정, 그리고 근미래예측 등의 영역까지 확대하고 있다(김연우 외, 2017).

빅데이터 연구는 크게 4가지 유형으로, 활용가능 빅데이터를 탐색하고 구현 알고리즘과 플랫폼을 설계·제시하는 연구(김연우 외, 2017; 장혜정, 2018), 기술적으로 빅데이터 시스템을 구상·구현하는 연구(조정희·이강우, 2018; 임종태 외, 2019), 모형을 수립하지 않는 데이터분석·시각화 연구(김홍범·심영석, 2018; 양승혁·강준호, 2019), 그리고 빅데이터를 활용해 계량분석 모형을 수립하는 연구이다(권기동 외 2015; 이종소 외, 2019). 물론 이들 유형은 완전한 경계를 이루지 않고 일부 중첩적인 면모도 있으며, 빅데이터의 정의에 따라 그 수가 달리 카운트된다.

본 연구는 특히 네 번째의 모형 수립 연구에 해당한다고 볼 수 있으며, 빅데이터 비활용 방식의 모형이 있을 경우 비교·검증이 중요함에도, 그 사례는 드문 편이다(권기동 외, 2015; 전해정, 2017; 김보찬 외,

2018; 이종소 외, 2019). 권기동 외(2015)는 교통안전 체험교육 정보시스템에 축적된 차량운행기록자료를 분석하여 23개 위험운전유형을 판단할 수 있는 알고리즘을 설계하고 위험판단기준(임계값)을 추출하였다. 동시에 주행실험을 통해 사람이 실제 위험하다고 느끼는 기준을 측정하고 상호비교하였다.

전해정(2017), 김보찬 외(2018)은 부동산 빅데이터와 데이터마이닝 기법을 활용해 가격결정 모형을 수립하였다. 전해정(2017)은 서울시 아파트실거래가격에 대한 각 아파트의 공간적 특성을 수집하고 이를 의사결정나무, 다중회귀, 신경망 등의 방법으로 분석하여 상호 비교하였고 신경망 분석 방법이 가장 우수한 결과를 보임을 확인하였다. 김보찬 외(2018)는 부동산실거래 애플리케이션의 대화가 주거용 부동산 임대료 샘플 12만 개를 웹크롤링 방법으로 수집하고 주변 지역 대학, CCTV 거리 등을 매칭하여 의사결정나무, 다중회귀, 신경망 등으로 분석하였으며, 의사결정나무가 가장 우수한 결과를 보인 것으로 확인하였다.

이종소 외(2019)는 통신망 접속자료, 토지피복자료 등을 활용해 국가하천 친수지구의 방문자 수를 추정하는 모형을 수립하였다. 기존에는 하천 이용 수요를 조사하기 위해 현장조사, 드론촬영 등의 직접 조사를 실시하였으나 빅데이터를 활용해 80% 내외의 설명력을 가진 회귀모형 수립이 가능했다.

2) MW 모형

MW모형은 한 지역 또는 국가의 연령별 인구 구성, 즉 인구 피라미드가 대상지역 주택의 총 주택수요를 설명할 수 있다고 전제한다. 연령별로 추정한 1인당 주택수요가 장기적으로 불변한다고 가정하면 미래의 인구 피라미드 변화에 따라 주택수요를 산술적으로 예측할 수 있다(Mankiw & Weil, 1989; 김준형 외, 2013). 기본적으로 다중회귀분석을 기반으로 개별 가구의 주택수요(주거면적, 주택가격 등)를 종속변수로, 각 가구의 연령별 가구원 수를 독립변수로 하여 그 계수를 추정하는 구조를 취한다(김경환, 1999).

$$D = \sum_{N=1} D_j \quad (\text{수식 1})$$

D_j : j 번째 가구의 주택수요면적

N : 총 가구 수

$$D_j = \alpha_0 Dummy0_j + \alpha_1 Dummy1_j + \dots + \alpha_{99} Dummy99_j \quad (\text{수식 2})$$

$Dummy_i$: j 가구 i 세 가구원 수

인구 피라미드는 출생률과 사망률을 고려해 비교적 안정적인 장기예측이 가능하므로, MW 모형은 장기 주택수요모형으로 유용하다. 특히 베이비붐, 고령화 등 인구 변화가 중요 이슈인 사회에서는 MW모형의 활용도가 높다(Mankiw & Weil, 1989; Cutler, 1990; Muellbauer & Murphy, 1997). 또 생애주기투자가설과 연계하여 주택시장은 물론 경제 전반에의 영향을 예측하는 연구에 활용되기도 하였다(Bakshi & Chen, 1994). 1인당 주거면적은 한국은 50대 전후, 캐나다는 45세, 일본은 60대가 가장 크다는 개별 연구결과들이 있었으며, OECD 15개 국가를 대상으로 한 연구는 70세라고 보고하였다(김진유·박지윤, 2017).

탄생 코호트(또는 출생 코호트)의 고유성이 나타날 경우 MW모형의 기본 가정이 깨어질 수 있다. MW모형은 현재의 30세와 20년 후의 30세가 같은 주거수요를 가질 것이라고 가정하고 있으나, 현실적으로 현재의 30대는 20년 후의 50세와 동일 집단이기 때문이다. 시민이 나이들과 함께 국가 경제도 성숙하기 때문에 연령에 따른 주거 소비 양태는 모든 세대에 걸쳐 동일하게 반복하지 않을 수 있다. 또 각 탄생 코호트는 일정한 간격을 두고 부모세대와 자식세대로 묶이기 때문에 세대간 전달된 경험에 따라 선호도에 차이가 있을 수 있다(김미경 외, 2014).

따라서 세대 효과의 유무를 판별하거나 이를 고려하여 주거소비 변화를 고려한 모형들도 제시되고 있다. 김준형·김경환(2011)은 55~65세 집단은 중대형 주택에 계속 거주하는 비율이 55세 미만 집단과 유사하지만, 65세 이상 집단은 크게 낮아짐을 토대로, 베이비붐 세대의 주거면적 감소의 경향이 당장보다는 약 10년의 시차를 두고 발생할 것으로 전망하였다. 반면 김기용·이창무(2010)는 고령화에 따른 주거면적 감소 경향은 만 79세에 이를 때까지도 발견되지 않았

으며, 오히려 1인 및 2인 가구의 경우 가구주의 연령이 65세 이상일 때 더 넓은 주거면적을 소비하는 것으로 나타난다고 보고하였다. 이정민 외(2012)는 보금자리 주택의 공급이 주택시장의 수급에 미치는 영향을 분석하면서 연령별 계수의 변화를 시계열적으로 추정하여 미래의 계수로 적용하기도 하였다.

기본적인 MW모형은 주택의 가격과 수요자의 소득을 고려하지 않기 때문에 주택 수요 모형보다는 주택 수요 모형으로 분류되기도 한다. 이를 수요모형으로 보완하기 위해 가격, 소득 등을 고려하는 수정 MW모형도 다뤄지고 있다(Swan, 1995). 이때 가격은 주거비용, 소득은 항상소득으로서 변수를 중첩하여 반영하는 경우가 많다(정의철·조성진, 2005; 황종규, 2016). 가구소득의 주 수입원이 가구주로부터 발생한다는 점에 주목하여 가구주 나이를 주요 변수로 분석하는 변형이 있으며(전성애·형남원, 2012), 1인 가구, 2인 가구, 아이 없는 부부 가구 등의 소득 특성을 감안한 연구도 발표된 바 있다(신미림·남진, 2011; 황종규, 2016; 이창무·박지영, 2009). 그 외 MW모형은 가구특성 변수를 일부 변형한 연구 등 다양하게 응용되고 있다(이영은·안정근, 2003; 정창무, 2008; 박천규 외, 2009; 장철민·강정규, 2013).

3) 선행 연구와의 차별성

본 연구의 차별성은, 첫째로 이제까지 인구주택총조사, 주거실태조사와 같은 주기적인 전문 조사에 의존해 수립되어온 MW모형에 공공빅데이터를 접목할 수 있는지 확인하는 것이다. 국토개발연구원(1997), 김경환(1999)이 MW모형에 한국가구패널을 사용한 이후, 주거실태조사(이창무·박지영, 2009; 김진유·박지윤, 2017), 가계동향조사(신미림·남진, 2011), 한국노동패널(황종규, 2016; 정보선 외, 2018), 가구소비실태조사(정의철·조성진, 2005; 조성진·조주현, 2013), 인구주택총조사(전성애·형남원, 2012) 등의 자료를 활용해 왔다. 그러나 현재까지 공공빅데이터를 접목한 MW모형 수립 사례는 확인되지 않고 있다(<표 1> 참조).

방대한 자료가 필요한 MW모형의 어려움을 극복·

〈표 1〉 본 연구와 MW 모형 주요연구 비교

저자	분석대상·단위	사용 자료	MW 모형 독립 변수
본 연구	서울시 및 서울시 25개 구	건축물대장, 주민등록부	연령(10세 단위)
김경환(1999)	전국	한국가구패널, 자체설문	연령(8세 단위)
정의철·조성진(2005)	전국	가구소비실태조사	연령(5세 단위), 주거비용, 항상소득
임종현·이천기·이주형(2007)	서울시 4개 존	자체설문조사	연령(5세 단위)
이창무·박지영(2009)	전국(소득그룹별)	주거실태조사	연령(5세 단위), 가구특성(1인, 부부), 광역시도(더미)
황현정·심보람·임엽(2009)	경기도 31개 시군	경기도 생활수준 및 의식구조 조사	연령(5세 단위), 주거비용, 소득
신미림·남진(2011)	서울시	가계동향조사	연령(5세 단위)
전성애·형남원(2012)	전국	인구주택총조사	연령(5세 단위), 주거비용, 항상소득
리우취엔·황은정·박환용(2013)	중국 상하이시	상하이시 거주민조사	연령(5세 단위)
김준형·천현숙·김민철(2013)	전국	한국노동패널조사	연령(5세 단위), 주거비용, 소득
김진유·박지윤(2017)	수도권	주거실태조사	연령(5세 단위), 주거비용, 연소득
조성진·조주현(2013)	전국	가구소비실태조사	연령(5세 단위), 주거비용, 항상소득, 가구특성(노령, 자가, 1인)
김미경·이창무·송기범(2014)	전국	주거실태조사 인구주택총조사	연령(5세 단위)
황종규(2016)	전국(2인 가구)	한국노동패널	연령(5세 단위), 주거비용, 항상소득
서원석·강민성(2017)	전국(아파트)	주거실태조사	연령(5세 단위), 주거비용, 항상소득, 자가주택 여부, 가구특성(1인, 2인)
강민성·서원석(2018)	서울시	주거실태조사	연령(5세 단위), 주거비용, 항상소득, 자가주택 여부, 가구특성(1인, 2인)
정보선·조용경·이상엽(2018)	전국(임대가구)	한국노동패널	연령(5세 단위), 주거비용, 항상소득, 가구특성(1인, 신체활동 제약), 아파트 여부

보완하려는 시도는 드물지만 계속되어 왔다. 임종현 외(2007)는 서울시 공동주택수요 특성을 분석하기 위해 전수대비 0.1%를 직접 조사하였다. 공동주택만을 구분하여 분석하기에는 기존 통계자료의 샘플 수가 부족했기 때문으로 생각된다. 또 김미경 외(2014)가 인구주택총조사에 노동패널 자료를 접목해 소비변화를 반영할 수 있도록 한 사례가 있다.

두 번째 차별성은, 공공빅데이터 활용에 의한 확장성을 탐구하는 것이다. 본 연구는 샘플을 풍부하게 확보할 수 있는 건축물대장, 주민등록대장을 사용하면서, 기존 연구에서 잘 다루지 않았던 기초자치단체별 MW모형과 주택유형별 MW모형을 제시할 수 있었다. 기존 연구는 주로 전국과 광역시·도 분석을 수행했다(이주형 외, 2009). 또, 인구주택총조사를 활용한 연구라도 기초자치단체 단위로 분석하지는 않았다

(전성애·형남원, 2012; 조성진·조주현, 2013). 황현정 외(2009)가 경기도 31개 시군 MW모형을 구축한 사례가 있으나 샘플 수가 시군별로 최소 377개, 최대 999개로 10개 내외의 변수가 삽입되는 MW모형에 비해 추가될 때 샘플 수 확보에 어려움이 있었다.

주택유형별 MW모형은 더욱 드물었는데, 그간 제공된 자료들은 샘플 수가 적어 주택유형까지 구분한 분석이 어려웠을 것으로 생각된다(임종현 외, 2007). 샘플 수가 풍부한 인구주택총조사는 1% 패널자료를 무료 공개하고 2% 패널자료는 유료로 제공되어 오다가 2016년이 되어서야 무료 공개로 전환되었다. 또 주거실태조사는 2016년까지 전체 샘플이 2만여 개에 불과하였고, 2017년부터 6만여 개로 늘어났다.

세 번째 차별성은 공공빅데이터의 일괄처리 알고리즘을 고민하는 것이다. 빅데이터는 본래의 목적과 다

〈표 2〉 본 연구에 활용된 주요DB의 개요

DB	DB특성				본 연구 활용자료 특성		
	대상	제공 샘플수	자료제공 갱신주기	관리주체	활용 자료	분석대상	입수처
건축물대장	국내 건축물	전수	수시	국토교통부	주소, 연면적, 주택 수	서울 25개구, 2016년	국토교통부 공간 빅데이터 사업단
주민등록대장	국내거주 내외국인	전수(비식별화)	수시	행정안전부	주소, 가구원 연령		
인구주택 총조사	국내거주 내외국인	전수대비 2%	5년*	통계청	가구원 연령, 전용면적	서울, 2015년	통계청 MDIS
주거실태조사	선정 패널	6만 개	매년	국토교통부	가구원 연령, 가구면적	서울, 2017년	

* 2015년부터 등록센서스 활용항목 조사결과는 매년 발표하나 MDIS시스템에 DB 업로드는 하지 않음

른 용도로 사용되는 것이기에, 인구주택총조사와 같은 전수자료나 주거실태조사와 같은 뚜렷한 용도에 맞춰진 자료에 비해 활용하기 까다로운 등 상대적 열위에 있다. 따라서 공공빅데이터는 수시로 업데이트된다는 점을 강조할 필요가 있고, 간결한 일괄처리 알고리즘을 고안하여 반복 작업이나 데이터클리닝의 비효율 문제를 해결하는 방안을 마련함으로써 활용성을 높일 수 있다(조영임, 2013; 경기연구원, 2015).

3. 분석 방법과 기초 통계

1) 자료 구축 과정

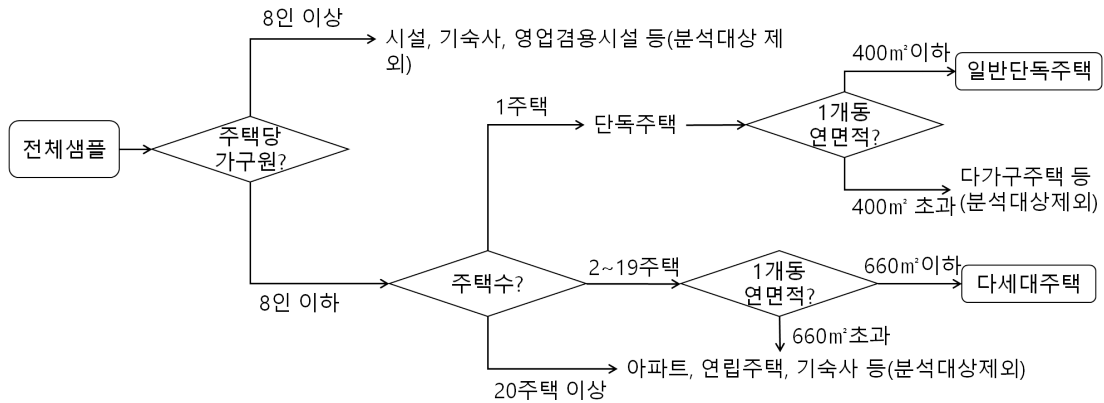
본 연구는 공공빅데이터인 건축물대장, 주민등록대장을 활용하며, 통계청 인구주택총조사, 국토교통부 주거실태조사 자료를 비교분석한다. 건축물대장과 주민등록대장은 수시로 업데이트하는 대규모 행정자료라는 점에서 공공빅데이터에 해당한다고 판단하였다. 시점은 2016년 전후, 대상은 서울의 일반단독주택 및 다세대주택이다. 주택의 종류를 한정하는 이유는 공공빅데이터 취득 자료의 샘플 단위가 도로명주소로 되어있어 수백~수천 가구가 하나의 도로명주소로 되어있는 공동주택은 분석이 어려웠기 때문이다. 인구주택총조사는 5년 단위로 제공되므로, 2015년 자료를 사용하였다. 이 조사는 2015년 등록센서스 전환 이후 매년 발표되지만, 원자료를 제공하는 통계청 MDIS

시스템은 현재 2015년 자료까지만 업로드하고 있다. 주거실태조사는 2016년 자료를 사용하고자 하였으나 샘플 수 부족으로 비교분석이 어려워 2017년 자료를 사용하였다(〈표 2〉 참조).

MW모형 수립을 위해 주소지별 연령별 가구원 수, 주거면적의 관계를 분석한다. 주민등록대장과 건축물대장의 주소 체계는 동일하기 때문에 이들을 상호 연동하면 MW모형 구축에 필요한 연령별 가구원 수와 전체 주택면적 자료를 획득할 수 있다는 것이 본 연구의 착안점이다. 행정안전부가 관리하는 주민등록대장은 세대(가구)별 주소, 세대원의 정보(이름, 주민등록번호, 성별, 연령) 세대원 간의 관계 등을 기록하고 있다. 또 국토교통부가 관리하는 건축물대장은 주소지별 건축물의 종류, 구조, 소유자의 주소 및 성명, 대지면적, 건축면적, 건축연면적 등을 제공하고 있다.

주민등록대장은 개인정보보호를 위해 원자료를 직접 열람할 수 없었기 때문에 국토교통부 공간빅데이터사업단으로부터 가공된 자료를 제공받았다. 분석대상지인 서울의 건축물대장(주택)을 기준으로 주민등록정보 연계를 신청하였다. 구득 자료는 주소가 비식별 처리되어(김경열·권현영, 2014; 이성엽, 2018), 도로명주소에서 본번과 부번이 제거되었으며, 연령별 가구원 수도 10세 단위로만 제공되었다. 건축물대장의 정보는 건축물연면적과 주택수, 주택가격만 제공되었다(〈그림 1〉 참조).

주거비용과 소득은 고려하지 않았다. 주택수요 추



〈그림 2〉 연구대상 샘플 분류 방법

아파트단지는 거주용 주택 공간뿐 아니라 관리소, 노 인정, 복도 등 공용면적과 상가까지 연면적에 포함되어 왜곡이 심하게 발생할 것으로 우려되었다. 반면, 다세대주택은 일반적으로 공용면적의 비율이 크지 않아 왜곡의 우려가 적다고 판단하였다.

3) 주택 분류

본 연구는 아파트, 시설 등을 제외하고 일반단독주택과 다세대주택 모형만을 구축하였다. 주택의 분류는 건축법을 참고하였다. 건축법상 단독주택은 1주택에 1가구가 거주하는 일반단독주택 외에, 다가구주택, 즉 1개동 연면적 660m² 이하이면서 19가구 이하가 거주하는 주택(주택사용부가 3개 층 이하일 것)을 포함한다. 그 외 영업겸용주택, 공관, 복지시설 등도 단독주택의 일종이다. 또 건축법상 다세대주택은 공동주택의 한 종류로서, 1개동 연면적 660m² 이하, 2호 이상 19호 이하 주택(4층 이하)의 요건을 가진다. 그 외 연립주택, 아파트, 기숙사 등이 공동주택으로 분류된다.

분석대상 단독주택은 일반단독주택으로 한정하되, 가구원 8인 이하로 구성된 1주택이면서 1개동 연면적 400m² 이하로 구성된 샘플로 분류하였다. 다만 가구원 수 기준의 차이를 확인하기 위해 별도로 4인 이하의 경우도 분석하였다. 이런 제한은 1주택 샘플 중에서 다가구주택을 최대한 배제하기 위해 필요하다. 다세대주택은 주택당 가구원 8인 이하의 2호 내지 19호

로 구성된 주택이면서, 1개동 연면적 660m² 이하로 구성된 샘플로 분류하였다.

일반단독주택 분류기준으로 주택당 가구원을 8인 이하로 제한한 것은 함께 분석한 인구주택총조사와 주거실태조사 자료상 서울시 일반단독 및 다세대 주택당 가구원 최댓값이 6~9인으로 나타나고 있었기 때문이다(〈표 5〉 참조). 또 본래 일반단독주택은 다가구주택과 구분하는 연면적 기준이 없으나, 인구주택총조사와 주거실태조사 자료상 일반단독주택 연면적은 400m²를 넘지 않았기 때문에 이를 적용하였다(〈그림 2〉 참조).

4) 기초 통계

공공빅데이터 자료의 총 샘플 수는 412,972개로 이에 포함된 총 주택 수는 3,030,760호이고, 주민등록인구는 8,889,388명이다. 서울시 열린데이터광장에 따르면, 실제 집계된 2016년 서울시 총 주택 수는 3,644,101호, 총 가구는 378만 가구이고, 주민등록 인구는 1,002만 명이다. 따라서 분석자료의 주택 수는 실제 집계의 83.4%, 주민등록 인구는 88.7%로 일부 누락되어 있으나, 본 연구는 사실상 전수 자료에 해당한다고 간주하였다(〈표 4〉 참조).

공공빅데이터 자료에서 일반단독주택 가구원 8인 이하 기준으로 분류된 샘플의 총 주민등록인구는 591,199명이고, 샘플 수(주택 수와 같음)는 135,755개이다(〈표 4〉). 4인 이하 기준으로 분류된 샘플 수

〈표 4〉 사용된 공공빅데이터 자료의 기초통계

구분	전체 샘플 (서울지역)	일반단독주택 (서울지역)		다세대주택 (서울지역)	
		1개동 연면적 400㎡ 이하 1주택 가구원 4인 이하	1개동 연면적 400㎡ 이하 1주택 가구원 8인 이하		
정의	최초취득 샘플			1개동 연면적 660㎡ 이하 2~19주택	
총 샘플 수(개) (전체샘플대비 비율)	412,792 (100.0%)	74,491 (18.0%)	135,755 (32.9%)	189,440 (45.9%)	
총 인구 수(명) (전체인구대비 비율)	8,889,388 (100.0%)	203,999 (2.3%)	591,199 (6.7%)	2,552,430 (28.7%)	
총 주택 수(호) (전체주택대비 비율)	3,030,760 (100.0%)	74,491 (2.5%)	135,755 (4.5%)	1,101,516 (36.3%)	
주소당 주택수	최솟값	1.0	1	1	2.0
	중간값	2.0	1	1	5.0
	평균값	7.3	1	1	5.8
	최댓값	6,864.0	1	1	19.0
	표준편차	62.9	0	0	2.9
주택당 가구원 (명)	최솟값	0.008	1.0	1.0	0.1
	중간값	3.0	3.0	4.0	2.3
	평균값	4.4	2.7	4.4	2.4
	최댓값	142.3	4.0	8.0	56.3
	표준편차	3.6	1.0	2.1	1.2
주택당 평균연령 (세)	최솟값	5.0	5.0	5.0	5.0
	중간값	62.9	58.3	60.7	63.8
	평균값	62.2	58.6	59.9	63.5
	최댓값	90	90.0	90.0	90.0
	표준편차	10.1	13.5	12.2	8.6
주택당 연면적 (㎡)	최솟값	9.5	9.5	9.5	9.9
	중간값	77.7	88.7	129.3	51.9
	평균값	125.9	117.5	139.7	60.5
	최댓값	661.5	399.9	400.0	329.9
	표준편차	114.0	83.8	85.7	33.4

는 74,491개이다. 실제 집계된 서울시 2016년 단독주택은 영업겸용을 포함해 135,936호이므로 양자 모두 실제 집계와 큰 차이는 없다고 볼 수 있다. 주택당 평균 가구원은 8인 이하 기준시 4.4인, 4인 이하 기준시 2.7인으로 인구주택총조사의 평균 2.4인보다는 다소 높은 편이다. 평균연령은 8인 이하 기준시 59.9세, 4인 이하 58.6세로 총조사 평균 51.2세보다 다소 높다. 주택당 연면적 평균도 8인 이하 139.7㎡, 4인 이하 117.5㎡로 총조사 평균 90.1㎡보다 크다(〈표 5〉).

다세대주택으로 분류된 샘플 수는 189,428개이고,

여기에 포함된 주택 수는 1,101,516호, 주민등록인구는 2,552,430명이다. 실제 집계된 서울시 2016년 다세대주택은 699,446호이므로 차이가 큰 편이다. 이는 다세대주택으로 분류된 주택 수에 아파트가 일부 유입되었을 가능성을 생각하게 하는데, 19호 이하 주택으로 구성된 소규모 아파트의 경우 5층 이상의 높이를 가진다는 점 외에는 다세대주택과 큰 차이가 없다. 주민공동시설이 없거나 드물고, 단지로 개발되지도 않기 때문이다. 따라서 연구 결과에의 영향은 적을 것으로 판단하였다.

〈표 5〉 사용된 인구주택총조사, 주거실태조사 자료의 기초통계

구분	인구주택총조사(서울지역)			주거실태조사(서울지역)		
	전체	일반단독주택	다세대주택	전체	일반단독주택	다세대주택
정의	전수의 2%	일반단독주택	다세대주택	유효 샘플	일반단독주택	다세대주택
총 샘플 수(개)	61,676	1,513	9,855	8,077	323	1,333
총 인구 수(명)	153,821	3,621	24,932	21,440	823	3,569
총 주택 수(호)	61,676	1,513	9,855	8,077	323	1,333
주택당 가구원 (명)	최솟값	1.0	1.0	1.0	1.0	1.0
	중간값	2.0	2.0	2.0	3.0	2.0
	평균값	2.5	2.4	2.5	2.7	2.5
	최댓값	10.0	8.0	9.0	9.0	7.0
	표준편차	1.3	1.3	1.3	1.2	1.2
주택당 평균연령 (세)※	최솟값	10.0	15.0	11.7	10.0	15.0
	중간값	39.3	51.7	38.3	45.0	55.0
	평균값	42.5	51.2	41.8	47.2	56.2
	최댓값	85.0	85.0	85.0	95.0	85.0
	표준편차	16.1	17.1	15.3	17.9	17.1
주택당 연면적 (㎡)	최솟값	10.0	10.0	10.0	5.0	6.6
	중간값	85.0	70.0	50.0	59.4	66.0
	평균값	120.5	90.1	51.7	65.1	70.6
	최댓값	660.0	372.0	250.0	396.0	396.0
	표준편차	105.2	65.4	22.0	31.1	37.3

※ 연령은 변수를 공공빅데이터와 통일하기 위해 10세별 단위로 재분류하여 적용

다세대주택으로 분류된 샘플의 샘플 당 주택 수는 평균 5.8호이다. 주택당 인구는 평균 2.4인으로, 총조사 평균 2.5인과 유사하다. 평균연령은 63.5세로, 총조사 평균 41.8세에 비해 크게 높다. 주택당 연면적은 평균 55.4㎡으로 총조사 평균 51.7㎡와 큰 차이는 없다(〈표 5〉 참조).

가구원 구성에 있어, 공공빅데이터 자료는 정기조사 자료와 주택당 인구수는 비슷하나 평균연령은 크게 작다. 차이가 있는 이유는 본 연구가 사용한 공공빅데이터가 주민등록대장이라는 것을 생각할 때, 젊은층이 실거주 현황에 비해 주민등록을 적극적으로 하지 않았기 때문으로 보인다. 따라서 당 공공빅데이터를 사용한 MW모형 수립시 젊은층의 1인당 주거면적은 과대추정될 소지가 있다.

정기조사 자료의 주택당 연면적은 평균값이 공공빅데이터 자료에 비해 작다. 다만, 중간값으로 비교할 때는 유사하다. 일반단독주택 분류에 가구원 수 8인

이하 기준 적용시 차이가 크게 벌어지는데, 여전히 다가구주택, 영업점용주택 등이 썩어 있기 때문에 판단된다. 가구원 수를 4인 이하로 하면 차이가 줄어든 것은 다가구주택을 배제한 효과로 생각된다. 한편, 영업점용공간이 포함될 가능성이 낮은 다세대주택의 주택당 연면적 평균은 차이가 크지않다.

4. 모형 구조별 추정 결과

1) 빅데이터-정기조사 MW모형 비교

일반단독주택의 경우, 공공빅데이터 모형과 정기조사 모형이 부분적인 유사성을 보였다. 특히 당 공공빅데이터는 정기조사에 비해 실제보다 젊은층이 적게 반영되어 있을 수 있다는 우려와 같이, 공공빅데이터의 MW모형인 모형 1(가구원 수 4인 이하), 모형 2(가구원 수 8인 이하) 모두 정기조사의 MW모형인 모형

〈표 6〉 일반단독주택 MW모형

연령 Age	〈모형 1〉 공공빅데이터(4인 이하)		〈모형 2〉 공공빅데이터(8인 이하)		〈모형 3〉 인구주택총조사		〈모형 4〉 주거실태조사	
	계수(m ² /인)	t	계수(m ² /인)	t	계수(m ² /인)	t	계수(m ² /인)	t
0~9세	25.1	23.2***	14.1	30.3***	5.7	1.0.	13.1	2.3*
10~19세	29.7	17.7***	18.4	24.6***	9.0	2.0*	11.6	2.3*
20~29세	54.2	37.9***	37.5	39.3***	28.3	8.7***	13.5	3.2***
30~39세	54.3	92.5***	40.1	100.5***	34.7	10.0***	15.1	3.3***
40~49세	48.6	120.8***	38.3	141.6***	29.2	8.0***	25.0	6.6***
50~59세	38.7	95.2***	29.8	119.2***	31.6	10.3***	22.9	5.7***
60~69세	34.7	67.1***	26.4	91.9***	33.9	11.9***	35.3	11.8***
70~79세	35.5	66.5***	28.9	95.9***	42.6	13.4***	33.7	11.8***
80~89세	35.8	71.9***	29.5	110.1***	40.8	7.9***	34.8	6.6***
90~99세	25.6	37.8***	19.8	59.1***	-※	-※	28.0	1.4.
F-value	1.369e+04(***)		3.445e+04(***)		197.1(***)		95.7(***)	
R ² (adj_R ²)	0.63(0.63)		0.72(0.72)		0.54(0.54)		0.80(0.80)	
표본크기	74,491		135,755		1,513		323	

*** : p < 0.001 ** : p < 0.01 * : p < 0.05

※모형 3은 원자료에 인구 90~99세가 포함된 가구가 전혀 없어 해당 변수를 제외하였음.

3. 모형 4에 비해 젊은 층의 1인당 주거면적(계수)이 크게 추정되었다(〈표 6〉).

주목할 점은 모형 2가 대체적으로 모형 3, 모형 4와 유사하면서도, 60세 이상의 구간에서는 모형 1이 모형 3, 모형 4와 좀 더 유사하다는 점이다. 또 모형 2의

설명력은 0.72로 모형 1의 설명력 0.63, 모형 3의 설명력 0.54보다 높다. 비록 기초통계량은 가구원 4인 이하인 샘플이 가구원 8인 이하인 샘플보다 정기조사 샘플과 유사해 보이지만, 모형 결과는 모형 2가 더 우수함을 확인할 수 있었으며, 고령자에 한해서는 모형 1

〈표 7〉 다세대주택 MW모형

연령 Age	〈모형 5〉 공공빅데이터		〈모형 6〉 인구주택총조사		〈모형 7〉 주거실태조사	
	계수(m ² /인)	t	계수(m ² /인)	t	계수(m ² /인)	t
0~9세	5.8	23.4***	5.0	9.2***	7.7	5.8***
10~19세	3.6	10.2***	7.4	14.2***	7.6	5.9***
20~29세	22.2	24.5***	15.5	39.3***	11.2	10.5***
30~39세	26.4	72.1***	21.6	60.8***	20.7	23.2***
40~49세	27.7	123.2***	21.9	50.0***	21.6	21.0***
50~59세	18.8	109.0***	19.9	50.3***	21.6	22.3***
60~69세	17.9	98.8***	23.4	52.9***	24.3	26.7***
70~79세	28.8	169.0***	24.8	39.6***	26.2	25.1***
80~89세	30.6	256.4***	22.8	18.8***	23.5	11.1***
90~99세	16.0	83.8***	-※	-※	35.0	3.8***
F-value	1.521e+05(***)		4,350(***)		763(***)	
R ² (adj_R ²)	0.89(0.89)		0.80(0.80)		0.85(0.85)	
표본크기	180,440		9,855		1,333	

*** : p < 0.001 ** : p < 0.01 * : p < 0.05

※모형 6은 원자료에 인구 90~99세가 포함된 가구가 전혀 없어 해당 변수를 제외하였음

〈표 8〉 서울시 자치구별 공공빅데이터 일반단독 주택(가구원 수 8인 이하) MW모형(25개구)

연령	서울시 전체	중로구	중구	용산구	성동구	광진구	동대문구	중랑구	성북구	강북구	도봉구	노원구	은평구
0~9세	14.1	20.5	4.6	14.9	13.1	16.8	7.5	14.2	13.1	11.0	17.1	15.2	13.0
10~19세	18.4	27.7	22.2	11.2	12.3	11.5	22.0	18.5	17.2	15.5	21.7	22.4	12.2
20~29세	37.5	56.2	42.8	34.5	32.7	42.9	27.9	26.8	39.7	31.6	28.1	15.4	27.6
30~39세	40.1	40.1	35.4	38.0	34.8	40.4	33.4	36.2	26.2	42.8	34.0	35.9	43.4
40~49세	38.3	35.9	31.3	29.7	33.1	42.9	32.7	37.6	35.1	38.5	39.3	39.9	40.7
50~59세	29.8	28.7	25.8	23.7	23.5	32.6	23.8	30.9	28.3	30.3	32.7	30.6	33.9
60~69세	26.4	21.3	25.4	24.9	23.4	27.5	23.1	29.4	25.9	36.6	29.1	21.8	27.7
70~79세	28.9	24.5	29.2	26.9	26.1	29.7	26.7	29.4	27.4	25.7	26.8	24.6	29.4
80~89세	29.5	28.1	26.3	28.4	26.6	30.3	28.9	26.4	26.8	22.5	24.8	26.5	26.5
90~99세	19.8	23.1	18.7	15.4	19.4	21.2	20.6	16.9	16.4	15.4	19.2	24.7	19.3
R2	0.72	0.56	0.61	0.67	0.72	0.78	0.65	0.77	0.66	0.72	0.74	0.69	0.74
샘플수	생략	6,978	2,755	6,321	4,015	5,836	9,256	7,672	13,523	9,469	3,733	2,557	8,864
연령	서대문구	마포구	양천구	강서구	구로구	금천구	영등포구	동작구	관악구	서초구	강남구	송파구	강동구
0~9세	19.3	16.4	17.1	13.9	8.9	8.2	17.0	12.1	9.7	25.2	25.8	16.7	23.2
10~19세	21.2	25.0	21.3	17.3	16.2	15.0	14.9	17.6	14.1	23.7	32.7	23.9	36.3
20~29세	45.8	52.7	24.7	36.1	32.0	60.0	34.3	43.9	35.4	55.0	63.4	29.7	33.0
30~39세	47.1	46.8	36.6	39.5	38.8	62.7	39.1	40.4	45.0	54.7	65.5	33.4	40.2
40~49세	39.8	38.9	45.3	44.1	37.2	56.1	36.6	36.8	38.1	50.0	63.3	40.6	36.9
50~59세	29.1	27.5	33.3	34.9	32.1	33.1	29.6	27.7	28.0	36.0	46.9	35.0	31.3
60~69세	24.7	32.4	27.8	32.3	30.2	38.2	27.7	24.0	25.4	29.9	41.6	31.2	31.6
70~79세	27.9	32.0	28.2	35.3	31.0	34.6	31.6	26.0	31.3	32.1	34.3	33.0	25.8
80~89세	31.0	32.9	31.4	30.0	27.2	36.0	28.2	32.1	34.2	37.6	35.0	29.6	30.2
90~99세	23.4	22.6	24.2	20.0	14.8	20.5	16.6	19.7	21.3	28.0	28.2	26.1	21.1
R2	0.70	0.67	0.78	0.78	0.76	0.69	0.75	0.75	0.79	0.77	0.76	0.79	0.79
샘플수	7,077	5,403	3,349	4,491	4,769	505	6,513	6,726	7,211	2,390	1,468	1,912	2,952

*1. 연령별 셀의 값은 각 연령별 계수로 1인당 주거면적에 해당함

*2. 지면상 t값 및 p값은 별도 표기하지 않았음. 음영(밑줄) 표시된 셀을 제외한 모든 셀은 유의수준 0.001(p=0.001)에서 유의함

도 참고할 수 있음을 알 수 있다.

다세대주택의 경우, 공공빅데이터 MW모형인 모형 5가 인구주택총조사의 MW모형인 모형 6, 주거실태조사의 MW모형인 모형 7과 유사한 결과를 보였으며, 설명력도 모두 모형 5가 0.89로 모형 6의 0.80, 모형 7의 0.85보다 높았다. 여전히 젊은 층의 1인당 주거면적에 있어 모형 5가 모형 6, 모형 7보다는 큰 경향이 있으나 일반단독주택 MW 모형들인 모형1-모형 4에서의 차이보다는 적었다(〈표 7〉 참조).

종합하면, 상술한 분류 알고리즘을 활용하고 있을 때, 공공빅데이터는 기존 정기간사에 기반을 둔 MW

모형에 근사한 결과를 보여주어 활용성이 있을 것으로 판단된다. 특히 다세대주택 MW모형은 일반단독주택 MW모형보다 기존 정기간사에 의한 모형과 유사한 결과를 보였다. 또 일반단독주택은 가구원 수 8인 이하인 샘플로 한정하는 것이 가구원 수 4인 이하인 샘플로 한정하는 것보다 전체적으로 기존 정기간사자료 분석결과와 유사했다.

2) 기초자치단체별 공공빅데이터 MW모형

공공빅데이터를 활용하여 서울시 기초자치체, 즉 25개 자치구에 대한 일반단독주택, 다세대주택 MW

〈표 9〉 서울시 자치구별 공공빅데이터 다세대주택 MW모형(25개구)

연령	서울시 전체	중로구	중구	용산구	성동구	광진구	동대문구	중랑구	성북구	강북구	도봉구	노원구	은평구
0~9세	5.8	8.6	5.7	4.1	6.8	6.3	7.8	8.0	7.6	17.6	11.3	9.7	9.5
10~19세	3.6	13.1	5.2	10.8	10.3	8.4	2.3	10.3	6.1	18.0	9.5	10.1	8.8
20~29세	22.2	34.1	37.8	34.3	20.7	21.6	21.5	17.5	18.9	22.5	18.2	19.7	18.0
30~39세	26.4	33.2	27.3	32.7	25.7	26.5	26.0	24.3	25.9	29.2	22.8	20.9	23.0
40~49세	27.7	28.1	24.7	30.0	26.6	27.7	31.0	27.3	29.0	29.5	28.7	28.0	28.4
50~59세	18.8	19.6	17.4	17.3	18.3	20.7	19.8	22.3	19.5	23.1	24.2	21.1	24.0
60~69세	17.9	18.6	16.2	21.2	15.3	18.5	13.3	18.6	15.7	9.3	22.7	13.1	22.3
70~79세	28.8	31.5	28.8	33.0	24.2	29.4	24.8	24.0	25.8	16.5	24.5	22.2	25.4
80~89세	30.6	34.1	35.3	33.8	28.0	27.2	32.6	24.5	33.1	23.8	19.3	26.5	24.5
90~99세	16.0	22.1	21.0	14.6	16.0	15.5	22.2	13.4	19.8	23.5	12.9	21.5	13.5
R ²	0.89	0.83	0.85	0.81	0.87	0.92	0.84	0.90	0.87	0.89	0.89	0.90	0.82
샘플수	생략	3,748	1,822	6,181	4,430	8,747	8,256	10,339	9,813	10,445	5,941	3,313	12,479
연령	서대문구	마포구	양천구	강서구	구로구	금천구	영등포구	동작구	관악구	서초구	강남구	송파구	강동구
0~9세	8.4	7.0	7.2	4.5	5.7	6.6	6.7	5.0	2.3	5.4	-2.8	3.5	8.1
10~19세	5.5	4.8	7.0	4.3	8.5	10.0	11.0	8.2	5.6	2.5	-6.1	3.6	6.8
20~29세	27.0	24.0	20.3	25.2	16.0	18.4	12.9	19.1	22.8	36.0	20.3	18.2	11.3
30~39세	28.3	27.5	17.6	28.2	29.1	27.6	32.1	27.6	24.7	32.6	45.4	23.7	20.9
40~49세	29.3	21.9	24.5	29.3	26.3	30.8	33.6	29.1	29.1	27.3	32.5	24.9	29.0
50~59세	16.5	14.8	19.7	24.4	23.8	27.0	20.8	19.6	18.5	17.0	10.9	17.3	20.2
60~69세	15.1	15.6	17.7	22.3	24.7	23.4	21.0	18.6	20.6	20.5	19.8	21.8	20.9
70~79세	28.9	31.3	26.7	28.3	25.0	24.9	25.3	28.1	29.9	34.3	41.4	30.3	26.3
80~89세	35.5	34.0	25.6	25.8	23.3	22.6	27.9	29.4	29.8	31.8	40.6	30.2	26.5
90~99세	21.9	16.6	14.6	12.6	10.2	9.3	12.3	18.1	12.8	15.9	15.8	11.6	12.4
R ²	0.88	0.88	0.91	0.92	0.90	0.88	0.87	0.90	0.91	0.89	0.87	0.93	0.91
샘플수	6,658	8,939	7,456	9,693	7,230	10,012	7,125	8,492	11,392	4,492	5,993	10,028	6,396

※1. 연령별 셀의 값은 각 연령별 계수로 1인당 주거면적에 해당함

※2. 지면상 t값 및 p값은 별도 표기하지 않았음. 음영(밑줄) 표시된 셀을 제외한 모든 셀은 유의수준 0.001(p=0.001)에서 유의함

모형을 수립하였다. 일반단독주택은 가구원 수 8인 이하인 샘플을 대상으로 하였다. 결과적으로 총 50개 모형(25개구×2개 주택유형)을 수립하였으며, 각 모형의 계수 총 500개 중 유의수준 0.001에서 유의하지 못한 계수는 18개뿐이었다(〈표 8〉, 〈표 9〉 참조). 한편, 인구주택총조사를 활용한 모형도 시도하였으나 샘플들이 구 단위로 쪼개지면서 부족해짐에 따라 계수들이 유의하지 못하거나 계수가 음수를 보이는 등 MW모형의 형태를 갖추지 못했다.

일반단독주택 MW모형의 계수들은 대체적으로 유의한 결과를 보이고 있고, 설명력은 0.70 전후로 형성

되어 있다. 하나라도 유의하지 못한 계수를 가진 구는 중구, 노원구, 송파구, 금천구뿐인데, 특히 금천구는 다른 구와 달리 유의하지 못한 계수가 3개 포함되었다. 금천구는 샘플이 505개로 25개 일반단독주택 MW 모형 중 가장 적은데, 다른 구는 적어도 수천 개의 샘플이 확보되었다.

25개 일반단독주택 MW모형들은 낮은 연령대의 1인당 주거면적이 높은 연령대보다 크게 나타나는 경향을 보인다. 이는 정기조사에 기반을 둔 모형 3이나 모형 4가 연령이 증가할수록 1인당 주거면적이 증가하는 결과를 보인 것과는 차이가 있는 것으로, 기초통

계로 확인하였듯이 저 연령대의 주민등록대장 등재가 적었던 것이 한계점으로 작용한 것으로 생각된다. 추후 좀 더 실용적인 방향으로 개선하기 위해서는 연령별 계수의 보정방법 등을 고안해야할 것으로 생각된다. 한편, 연령별 계수를 시계열적으로 확보하여 추적하는 경우에는 상대적인 변화를 중심으로 분석하면서 문제를 완화할 수 있을 것으로 생각된다(이영은·안정근, 2003; 진미운·김경선, 2010; 이정민 외, 2012).

다세대주택 MW모형의 계수들은 일반단독주택 MW모형에 비해 유의하지 못한 계수가 많으나, 설명력은 0.85 전후로 더 높게 형성되어 있다. 특히 소위 강남4구라 불리는 서초구, 강남구, 송파구, 강동구에 유의하지 못한 계수가 몰려있으며, 중구, 동대문구, 서대문구, 관악구도 1-2개의 변수가 유의하지 않다. 서초구와 강남구는 샘플수가 5천 개 전후로 많은 편인데도 유의하지 않은 것은 주목할 만하다.

25개 다세대주택 MW모형들은 대체적으로 정기조사에 기반한 모형 6이나 모형 7과 유사한 형태를 보인다. 공공빅데이터의 다세대주택 자료는 가구 단위로 구성되어 있지 않고, 다세대주택 1개 동(정확히는 1개 주소) 단위로 구성되어 있음에도 유사한 결과가 나타나는 것은 고무적이라 할 수 있다. 본 연구의 방법론이 지속적인 1인당 주거면적 모니터링에 유용할 것으로 기대할 수 있다. 다만 강남4구의 결과가 미진한 것은 보완이 필요할 것으로 생각된다.

5. 요약 및 결론

본 연구는 맨큐-와일 모형을 활용하여 1인당 주거면적을 추정함에 있어 건축물대장과 주민등록대장을 연계한 공공빅데이터가 기존 정기조사(인구주택총조사, 주거실태조사)에 못지않은 수준에 이르는지, 또 공공빅데이터를 활용할 경우 누릴 수 있는 장점이 있을지 확인하고자 하였다. 이를 위해 건축물대장과 주민등록대장을 연계하여 전수자료를 구축하였으며, 일반단독주택과 다세대주택 샘플을 일괄판별 추출하는 조건을 설정하여 MW모형을 구축하였다. 추가적으

로 인구주택총조사와 주거실태조사 자료를 처리하여 MW모형을 구축함으로써 상호비교하였다. 이를 통해 작성한 모형수립 방법을 서울시 25개 자치구별로 적용하여 자치구별 MW모형을 수립하였다. 그 결과, 본 연구는 다음과 같은 결론을 얻을 수 있었다.

첫째, 건축물대장과 주민등록대장을 도로명주소로 연결하고, 주택유형을 구분하는 일괄조건을 정함으로써, 정기조사 자료에 기반한 MW모형과 유사한 수준의 MW를 수립할 수 있다. 비록 일반단독주택(분류기준: 가구원 8인 또는 4인 이하, 1주택, 연면적 400m² 이하)은 기존 정기조사 방식과 일부 차이를 보였지만 다세대주택(가구원 8인 이하, 2~19주택, 연면적 660m² 이하)은 비교적 유사한 결과를 얻을 수 있었다. 기존 정기조사 결과를 활용한 연구들은 주택유형별 MW모형을 수립한 사례가 드물다는 점에서 중요한 개선점이다. 다만 젊은 층의 주민등록이 실제보다 적게 반영되어 이들의 1인당 거주면적이 과다추정되는 문제가 있었다. 추후 장기적인 추정을 통해 인구주택총조사 결과와 비교하여 보정해 나감으로써 결과를 개선할 필요가 있다. 또, 본 연구는 자료구득의 한계로 일반단독과 다세대주택 MW모형만 수립할 수 있었으므로, 아파트 MW모형은 후속 연구가 필요하다.

둘째, 공공빅데이터는 기초지자체 단위의 모형 수립시 정기조사보다 유리한 장점이 있다. 공공빅데이터는 충분한 샘플 확보가 가능하였으며, 유의수준 0.001에서도 500개 계수 중 482개 계수가 유의한 결과를 보였다. 반면 정기조사는 샘플 수가 부족하여 기초자치단체 단위의 모형 수립이 불가능했다. 물론 인구주택총조사 자료가 2% 샘플이 아닌 10% 또는 전수 샘플일 경우 더 우수한 모형 수립도 가능할 것으로 기대된다(윤주현 외, 2006; 은기수, 2007; 한국인구학회, 2012). 그러나 본 연구의 방법론을 활용하면 비교적 적은 비용으로 수시 업데이트된 결과를 얻을 수 있는 장점이 있다.

본 연구의 공공빅데이터와 이를 활용해 구축하는 MW모형을 실시간으로 업데이트하여 시계열 자료로 축적할 경우 활용성이 더 클 것으로 기대된다. 본 연구는 비교적 간단한 알고리즘만으로 MW모형을 수립

할 수 있었으며, 이를 일괄처리함으로써 지속적인 모니터링과 연구 확장도 기대할 수 있다. 또 현재 서울시 25개 구에만 적용한 방법론을 전국 기초지자체로 확대하기도 간단할 것으로 기대된다.

공공빅데이터를 활용함으로써 연구영역을 확장할 수 있다는 것은 장점이지만, 별도의 목적으로 생성된 자료를 활용한다는 점에서 공공빅데이터의 근본적 한계는 존재한다. 본 연구는 이를 극복하기 위해 분석대상 개념들을 재정의하고, 범위를 한정하는 등의 작업이 필요했다.

또 개인정보의 접근이라는 어려운 과정은 분석을 보다 신중히 진행해야 하는 부담이 될 수 있고, 비식별화를 거친 자료를 분석함에 따라 연구설계가 어려워지는 문제도 있었다. 제공 받은 공공빅데이터 자료가 동호가 분리되지 않고 도로명주소 단위로 제공되었기 때문에 대규모 아파트단지는 처음부터 제외하고 분석해야 했으며, 연립주택도 단지로 구성되는 경우가 많아 일괄 제외해야 했다. 그럼에도 한정된 정보라는 불리함을 역이용해 간결하고 안정적인 알고리즘을 제시하고자 노력하였다. 또, 비식별화 등은 연구자 개인의 한계로, 자료접근이 자유로운 기관의 경우 비교적 쉽게 해결할 수 있을 것이다.

마지막으로, 공공빅데이터를 분석해 새롭게 도출한 정보들은, 기존의 방법의 결과물이 없어 비교가 곤란하므로 그 신뢰성을 직접 확인할 수 없는 것도 한계점이었다. 본 연구는 이와 같은 확장성을 제시하는 데 초점을 두어야 했지만, 추후에는 반대로 공공빅데이터가 제공하는 새로운 정보가 맞는지 정기조사 방식을 통해 확인하고 신뢰도를 높이는 작업도 이어져야 할 것으로 생각된다. 앞으로도 공공빅데이터가 기존의 통계조사를 보완하거나 대체할 수 있을 정도로 의미 있는 결과를 보여줄 수 있도록 가공하는 방법 등에 대한 꾸준한 연구가 필요할 것이다.

참고문헌

강민성·서원석, 2018, 가구생애주기 및 점유형태별 주거격차 분석: 거주면적과 주택수요 비교를 중심으로, 『한국지역개발학회지』, 30(2), pp.117-135.

- 경기연구원, 2015, 경기도 주택·부동산 빅데이터 분석 및 활용체계 연구.
- 국토개발연구원, 1997, 주택수요구조 변화전망에 관한 연구.
- 권기동·주재홍·진장원, 2015, 『정보화연구』, 12(4), pp.515-523. (??논문제목이 없어요)
- 김경렬·권현영, 2014, 공공데이터 활용을 위한 개인정보보호 제도의 개선 과제, 『경제규제와 법』, 7(2), pp.22-37.
- 김경환, 1999, 인구의 연령구조 변화와 주택 수요 및 주택 가격, 『대한부동산학회지』, 17, pp.69-84.
- 김기용·이창무, 2010, 고령화 사회와 주택수요, 『대한국토·도시계획학회 춘계산학협동학술대회 논문집』.
- 김기환, 2013, 공공부문 빅데이터의 활용성과 위험성, 『정책분석평가학회보』, 23(2), pp.1-27.
- 김미경·이창무·송기범, 2014, 세대(탄생코호트) 효과를 고려한 실용적 장기 주택수요 모형, 『국토계획』, 49(8), pp.35-48.
- 김연우·김병훈·고건식·최민웅·송희섭·김기훈·유승훈·임종태·복경수·유재수, 2017, 실시간 가상 빅데이터를 활용한 홍수 재난안전 시스템 설계 및 구현, 『한국콘텐츠학회논문지』, 17(1), pp.351-362.
- 김준형·김경환, 2011, 고령화와 주택시장: 은퇴전후 주택소비 변화를 중심으로, 『부동산학연구』, 17(4), pp.59-71.
- 김준형·천현숙·김민철, 2013, 주택수요의 규모별 분포 예측: 맨큐-와일 모형에서 추계가구자료의 활용, 『국토계획』, 48(2), pp.263-279.
- 김진유·박지윤, 2017, 가구원 수별 가구수를 활용한 주택수요 추정 모형 연구, 『부동산학연구』, 23(4), pp.65-76.
- 김홍범·심영석, 2018, 연관규칙 마이닝 기법을 활용한 관광지 이미지 빅데이터 분석에 관한 연구, 『관광레저연구』, 30(12), pp.57-76.
- 리우취엔·황은정·박환용, 2013, M-W모형을 이용한 상하이시 주택수요 추정에 관한 연구, 『주거환경』, 11(2), pp.205-218.
- 박천규·이수욱·손경환, 2009, 가구생애주기를 감안한 주택수요특성 분석 연구, 『국토연구』, 2009.3, pp.171-187.
- 서원석·강민성, 2017, 수정 M-W모형을 이용한 가구의 장기주택수요에 영향을 미치는 주거환경특성 분석, 『대한부동산학회지』, 35(3), pp.5-23.
- 성육준, 2017, 데이터 생애주기 관점에서 본 공공부문 빅데

- 이더 활성화 방안, 『한국지역정보학회지』, 20(2), pp.25-41.
- 신미림·남진, 2011, 서울시 1인가구의 주택수요 예측, 『국토계획』, 46(4), pp.131-145.
- 양승혁·강준호, 2019, 프로스포츠클ubs의 위키커뮤니케이션 전략 빅데이터 분석기법을 활용한 프로야구 심판 금전거래 사건 사례분석을 중심으로, 『한국스포츠키업경영학회지』, 24(2), pp.15-31.
- 윤상오·현지우, 2019, 공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구, 『한국공공관리학보』, 33(1), pp.219-247.
- 윤주현·강미나·박천규, 2006, 인구주택총조사로 본 주거특성 변화 및 정책 시사점, 『국토정책 Brief』, 101, pp.1-8.
- 은기수, 2007, 『인구주택총조사 종합분석』.
- 이경민·김호기·서범준·정창무, 2012, 보금자리주택이 규모별 주택시장에 미치는 영향 예측, 『국토계획』, 47(2), pp.83-98.
- 이상훈·양동석·임재빈, 2016, 주택분야 수요분석 지원을 위한 빅데이터 적용방안 연구, 토지주택연구원.
- 이성엽, 2018, 한국의 데이터주도 혁신에 대한 법의 대응과 진화, 『경제규제와 법』, 11(2), pp.147-167.
- 이영은·안정근, 2003, 수도권 권역별 주택수요 분석, 『국토계획』, 38(6), pp.61-74.
- 이종소·이상은·최진영, 2019, 국가하천 친수지구 공간관리를 위한 통신 빅데이터 활용성 검토: 자료검증과 이용지표 선정, 『국토연구』, 101, pp.3-18.
- 이주형·임종현·이천기, 2009, 가구특성에 따른 주택의 점 유형태 및 유형 선택에 관한 연구, 『국토계획』, 44(3), pp.79-93.
- 이창무·박지영, 2009, 가구특성을 고려한 장기주택 수요 예측모형, 『국토계획』, 44(5), pp.149-161.
- 임종태·복경수·유재수, 2019, 빅데이터 분석기법을 이용한 실시간 대중교통 경로 안내 시스템의 설계 및 구현, 『한국콘텐츠학회논문지』, 19(2), pp.460-468.
- 임종현·이천기·이주형, 2007, 서울시 인구구조에 따른 공동주택수요 특성 및 전망, 『국토연구』, 2007,6, pp.147-162.
- 장재일·안건혁, 2008, 경제변수와 확률을 이용한 주택 수요 예측, 『국토계획』, 43(7), pp.31-42.
- 장철민·강정규, 2013, 가격-면적 근집에 의한 주택 수요에 관한 연구: 부산시를 중심으로, 『주거환경』, 11(1), pp.205-219.
- 장혜정, 2019, 안전도시 구현을 위한 빅데이터 적용 알고리즘 연구, 『도시설계』, 19(1), pp.37-51.
- 전성애·형남원, 2012, 인구구조 및 가구특성 변화를 고려한 주택수요의 예측, 『국토계획』, 47(3), pp.191-208.
- 정보선·조용경·이상엽, 2018, 고령자 맞춤형 공공임대주택 수요추정에 관한 연구, 『주택도시연구』, 8(1), pp.37-52.
- 정의철·조성진, 2005, 인구구조 변화에 따른 장기주택수요 전망에 관한 연구, 『국토계획』, 40(3), pp.37-46.
- 정창무, 2008, 맨큐-웨일 장기주택수요모형의 적정성 평가-서울지역을 중심으로, 『계획』, 24(2), pp.245-252.
- 조성진·조주현, 2013, 주택 수요변화 분석을 통한 장기주택 수요 전망에 관한 연구: 2000년, 2010년 자료를 중심으로, 『국토계획』, 48(5), pp.251-268.
- 조영임, 2013, 빅데이터의 이해와 주요 이슈들, 『한국지역정보학회지』, 16(3), pp.43-65.
- 조정희·이강우, 2018, Marmot: 공간 빅데이터 처리를 위한 하둡 기반의 고성능 데이터 저장 관리 시스템, 『한국지형공간정보학회지』, 26(1), pp.3-10.
- 진미운·김경선, 2010, 주택수요 평가지표의 개발과 적용 연구, 『국토연구』, 2010.12, pp.3-23.
- 한국인구학회, 2012, 『인구주택총조사 표본결과 심층분석』.
- 황종규, 2016, 2인 가구 장기주택수요 전망에 관한 연구, 『주거환경』, 14(2), pp.21-32.
- 황형정·심보람·임업, 2009, 경기도 인구구조에 따른 주택 수요 추정, 『지역연구』, 25(4), pp.107-121.
- Bakshi, G. & Chen, Z., 1994, Baby boom, population aging, and capital markets, 『Journal of business』, pp.165-202.
- Cutler, D., 1990, 『An aging society: Opportunity of challenge』.
- Mankiw, N. & Weil, D., 1989, The baby boom, the baby bust, and the housing market, 『Regional science and urban economics』, 19(2), pp.235-258.
- Muellbauer, J., & Murphy, A., 1997, Booms and busts in the UK housing market, 『The Economic Journal』, 107(445), pp.1701-1727.
- Swan, C., 1995, Demography and the demand for housing A reinterpretation of the Mankiw-Weil demand variable, 『Regional Science and Urban Economics』, 25(1), pp.41-58.

계재신청 2020.02.03.
심사일자 2020.02.14.
계재확정 2020.02.26.
주저자: 임재빈, 교신저자: 이상훈