

빅데이터 기반 주택수요 분석을 위한 DB 구축 방안 연구

양동석, 이상훈, 임재빈
한국토지주택공사 토지주택연구원
e-mail:blue@lh.or.kr

A Study on the DB Construction Method for Analyzing Housing Demand Analysis Based on Big-Data

Dong-Suk Yang, Sang-Hoon Lee, Jae-Bin Lim
Land & Housing Institute, Korea National Land & Housing Corporation

요 약

적절한 주택공급 및 주택정책을 위해서는 인구 및 가구 구조의 변화에 따른 주택수요의 예측의 정확성이 요구되고 있다. 본 연구에서는 기존 주택수요 예측에 있어서의 DB의 문제점들을 살펴보고 개선 방안 및 빅데이터를 활용할 수 있는 DB 구축방안을 제시하였다. 향후, 기존에 활용되지 않고 있는 주택공시가격, 건축물대장, 가계동향조사, 인구주택 총조사 등을 활용하여 주택수요를 분석할 수 있도록 파일럿시스템을 개발하여 타당성을 검토할 예정이다.

1. 서론

장기주택종합계획에 활용되는 맨큐-와일(M/W) 모형은 대표적인 주택수요 예측 모형으로 한 지역 또는 국가의 인구 피라미드 구조가 주택 총수요(면적)으로 설명할 수 있으며, 미래의 주거수요를 인구피라미드의 변화로 추정하고 있다. 그러나 소득수준이 증가할 경우 주거면적이 증가하거나 주거비용이 높을 경우 면적이 감소하는 경우 등을 고려하여야 하며, 가구 구조의 변화 및 특정 인구수에 따라 주택수요가 변화를 수용할 수 있어야 한다. 본 연구에서는 기존의 주택수요 예측 모형들에서 요구되는 변수들을 기준으로 DB 구성의 문제점들을 살펴보고 개선방안 및 주택수요관련 빅데이터 활용을 위한 데이터베이스 구축 방안을 제시한다.

2. 주택수요 분석을 위한 데이터 활용 사례

현재 주택정책에 요구되는 주택소요 (Housing needs)의 경우 거시적 관점의 외재적 변수(Extrinsic Variables)에 해당하는 데이터들을 활용하여 분석함에 따라 인구 가구 수, 주택공급량, 미분량 물량, 거래량, 매매가변동률, 전세가율, 소득 등의 데이터들이 구축되어야 한다. 그러나 이러한 외재적 변수외에 특정 사업지에 따른 주택 수요를 분석하기 위해서는 사업지의 거주환경과 관련된 미시적 내재적 변수(Intrinsic Variables)가 요구된다.[1]

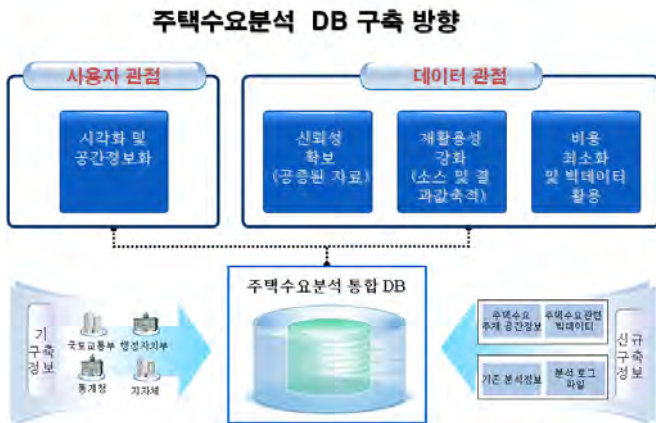
주택수요분석에 활용된 데이터 사례로는 지역별 수급 추정 및 사업지별 수요 추정 방법론 개발과 적용에서

(2014)¹⁾ 사업지 신규주택 공급량, 사업지 중위가격 대비 분양주택 가격, 사업지와 경쟁 사업지 간 거리, 사업지별 공급세대수, 사업지와 경쟁사업지 간 고용접근성 지수의 차, 대규모 개발여부 등 여섯 가지 변수 외에 종속변수로는 이주가구수 등이 사용되었고 현금화 가능자산, 주택자급대출 가능성과 같은 개인금융상태, 가구원수, 통근용이성, 학생 수, 분가여부(주택마련 필요성 변수), 주변 환경 쾌적성, 생활편의시설, 교육환경(주거환경 변수), 내부 환경 및 규모, 장래교통 편의성 (투자가치변수) [이우중, 2003]²⁾, 공급전용면적, 현관구조, 방수, 층수(단지편리성), 경과연수, 단지규모, 세대 당 주차대수 (단지쾌적성), 버스정류장 및 전철과의 거리(교통환경변수), 공원, 강, 산(조망권 변수), 학교와의 거리(교육여건 변수), 쇼핑시설 및 관공서(편의특성 변수) [김광석 등 2010]³⁾, 투기적 요인, 가구원수, 주거지 만족도, 연령, 단지 내 시설 만족도 등 투자 심리 관련 변수들의 통합 시도[이재영, 2003]⁴⁾ 등의 사례가 있다.

1) 한국토지주택공사(2014). "지역별 수급추정 및 사업지별 수요추정 방법론 개발과 적용용역"
2) 이우중, 이재영, 신동민(2003), "수도권 주택구매의 소비자 선호특성 연구", 국토계획, 38(3):135~146
3) 김광석, 박원갑(2010), "실거래 가격을 이용한 아파트 분양가 및 계약률 측정에 관한 연구", 한국전자통신학회 2010추계종합학술대회지, 4(2):205~208.
4) 이재영(2003), "수도권 주택 수요 특성 및 구매 결정요인 분석", 한국지역개발학회지, 15(2):165~182.

3. 주택수요 분석 DB 문제점 및 구축 방향

기존에 주택 수요 분석에 활용된 데이터를 살펴보았을 때, 외재적 변수의 경우 데이터 수집이 일정 부분 용이한 반면에 내재적 변수는 실제 설문 및 면접 등 데이터 구득에 많은 비용을 초래함에 따라 주택 수요 분석을 위한 DB구축은 우선적으로 비용 및 구득 시간을 최소화하는 방향으로 추진할 수 있어야 한다. 대부분의 주택 수요 분석의 경우 엑셀이나 통계 프로그램을 활용하여 그 결과 값을 제공하는 관계로 관련 데이터들이 데이터베이스로 축적되지 못하고 일회성 활용에 그치고 있는 실정이다. 그에 따라 분석결과 값뿐만 아니라 사용된 데이터를 데이터베이스에 축적하여 추후 검증할 수 있는 자료로 활용 가능할 수 있도록 DB가 구축되어야 한다. 또한, 정확한 주택 수요 분석을 위해서는 무엇보다도 기초 자료의 신뢰성을 확보할 수 있어야 함에 따라 기존 주택 수요 관련 국가 정보시스템들에서 제공하는 데이터 중심으로 DB 구축이 필요하다. 직관적인 주택 수요 분석을 위한 결과의 시각화 및 GIS 기반 시각화와 분석이 가능하도록 하는 기능 필요함에 따라 관련 데이터들이 위치정보를 포함할 수 있도록 데이터 구축할 수 있어야 한다. 마지막으로 데이터 및 분석 결과의 활용성을 높일 수 있는 부가기능을 제공할 수 있도록 확장성을 확보할 수 있어야 한다.

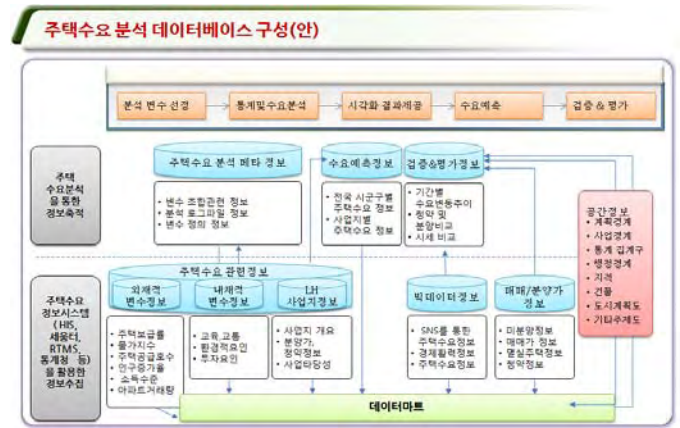


(그림 1) 주택수요분석 DB 구축방향

4. 주택수요 분석을 위한 DB 구축 방안

주택 수요 분석을 위한 데이터베이스는 주택 수요 분석에 요구되는 기초자료(외재적 변수정보, 내재적 변수정보, 기타정보) 축적과 분석과정을 통한 결과 값을 구축하도록 하는 크게 두 가지로 구성할 수 있다. 주택 수요분석에 요구되는 기초자료를 중심으로 DB를 구성하는 정보수집에서는 관련된 정보시스템들에서 해당 자료들을 연계하여 구축하도록 하고(예: 주택보급률, 물가지수, 주택공급호수, 신규가구수, 인구증가율, 소득수준, 아파트 거래량 등) 분석과정 및 분석결과를 데이터베이스화하여 추후 활용 및 검증을 할 수 있도록 데이터베이스 구축하여야 한다. 각 단계별 DB 구축과정에서 생성된 정보들은 데이터마트로

종합적으로 취합되어서 사용자가 선택하여 분석할 수 있도록 DB 구성하고 분석결과 값을 데이터베이스에 축적하여 추후 검증할 수 있는 자료로 활용 가능하도록 해야 한다. 또한 주택시장 관련 빅데이터 정보를 활용하여 1차적으로 검증 및 평가 단계에서 활용한 후 추후 수요분석 시에도 점차적으로 활용할 수 있도록 빅데이터 정보들을 단계적으로 구축할 필요가 있다. 기존에 주택 수요 분석에서는 통계프로그램을 통하여 수요분석을 수행 후 그 결과 값을 지도상에 표현하는 방식을 취한 반면에 본 연구에서 제안하는 주택 수요 DB는 위치정보를 함께 구축하여 GIS 상에서 직관적으로 분석이 가능하도록 할 수 있다. 주택 수요 관련 자료를 위치정보화하고 기존 공간정보들(주제도, 도시계획도, 인공사진 등)과 중첩하여 비교 분석할 수 있도록 DB 구축이 요구된다. 그러나 공간데이터베이스 구축비용의 어려움에 따라 단계적으로 DB를 구축할 수 있어야 한다.



(그림 2) 주택 수요분석 데이터베이스 구성(안)

외재적 변수정보는 중앙정부 차원에서 수행하는 신규주택 수요 추정에 요구되는 변수들을 중심으로 시계열별로 DB를 구축한다. 사용된 변수는 가구증가율, 소득요인, 멸실 요인 등을 활용할 수 있는데 이를 나타낼 수 있는 주택보급률, 소비자물가지수, 연간주택공급호수, 인구증가율, 소득수준 등의 변수들을 공통으로 사용한다. 보완하기 위한 외재적 변수로 신규가구증가, 연평균주택거래실적, 연평균가구증가율, 주택미분양호수, 아파트거래량 등을 사용하며, 추가적으로 분양호수, 수요권별 공급량, 가구수, 수요권별 미분양물량, 수요권별 거래 증가율, 수요권별 청약경쟁률, 수요권별 세대수 증가율, 아파트 평균가, 아파트 최고가, 분양가, 매매가 변동률, 전세가율, 부동산심리지수, GRDP(Gross Regional Domestic Product) 등을 구축하도록 한다. 내재적 변수정보는 설문조사 및 면접조사 등 직접적인 데이터 수집 비용이 발생하는 정보로 기존에 수행되었던 용역들의 성과물 중심으로 DB를 구축하는데 내재적 변수는 '교육', '교통', '생활편익환경', '자연환경', '이웃수준', '투자가치'의 6개 헤도닉변수의 평균값과

‘PSM(Price Sensitivity Meters)종합가치’ 두 가지 변수를 구하여 사용한다. 사업지정보는 사업지의 유효수요 분석을 할 경우 해당 사업지에 대한 정보들을 의미한다. 수요예측 정보에서는 주택 수요분석 결과를 분기별, 연도별로 축적하여 시계열 분석이 가능하도록 DB가 구성된다. 사용된 외재적 변수 및 가중치 등의 내용에 따른 결과 값을 시군구 단위로 나타낼 수 있다. 분석결과 값을 데이터베이스에 축적하여 추후 검증할 수 있는 자료로 활용 가능하도록 해야 함에 따라 검증 및 평가정보는 실제 분양량과 청약률을 해당 연도에 DB화하여 예측한 결과 값과 비교할 수 있도록 한다. 매매/분양가 정보는 주택 수요예측의 결과를 검증하거나 평가하고자 할 때 비교할 수 있는 데이터들로 구성되어 분양가, 매매가, 주택 채고 등을 구축한다. 빅데이터 정보는 주택 수요예측을 하기 위한 보조적인 데이터 역할을 하거나 검증/평가지 활용될 수 있는 데이터들로 구성되어 주택 수요 SNS, 경제활력정도(카드 사용액, 통화량, 에너지 사용량 등), 텍스트마이닝을 통한 구매의사량 등을 의미한다. DB 구축을 위한 데이터 구성에 따른 외부 유관 정보시스템과의 연계는 그림 3.에 나타내었다.

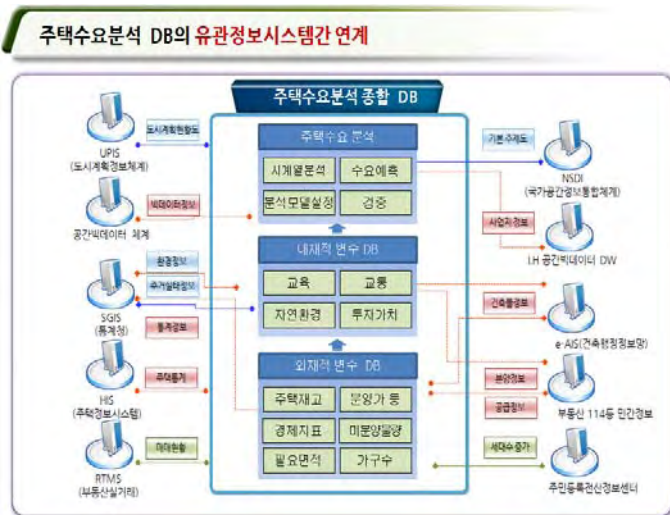
과를 DB화시키는 것도 중요하다. 그리고 기존 연구의 경우 GIS 연계분석이 이루어지지 않고 있으며 우수한 시각화 기능을 갖추지 않고 있어 직관적인 예측 및 활용성이 낮음에 따라 GIS 연계 및 시각화 기법을 제시하여 활용성 높일 수 있도록 DB를 구축할 수 있어야 한다. 전국 시군구 단위의 주택수요 분석뿐만 아니라 추후 지구단위까지 확대할 수 있는 DB 구축도 필요하다. 데이터마트 구성에 있어 단계적으로 DB 발전 방안이 요구되며 사업지구별 주택 소요량을 파악하기 위해서는 내재적 변수에 대한 DB 구축이 필요하나, 현재 비용적·시간적인 부분에서 어려움이 있으므로 단계적으로 DB 구축 방안 마련하여 발전시켜 나갈 수 있어야 한다. 앞으로 본 연구에서 제시한 DB 구축 방안에 따라 일차적으로 파일럿시스템을 구축하고 그에 따른 타당성을 검증할 계획이다.

사사

본 연구는 토지주택연구원에서 수행한 “주택분야 수요 분석 지원을 위한 빅데이터 적용방안 연구”과제의 일부 및 국책 R&D “빅데이터를 활용한 주택시장 분석 및 예측 모형 개발”과제의 일부를 발췌하여 수정·보완한 것입니다.

참고문헌

- [1] 김광석, 박원갑(2010), “실거래 가격을 이용한 아파트 분양가 및 계약률 측정에 관한 연구”, 한국전자통신학회 2010춘계종합학술대회지, 4(2):205~208.
- [2] 이상훈, 양동석, 임재빈(2016), “주택분야 수요분석 지원을 위한 빅데이터 적용방안 연구”
- [3] 이우중, 이재영, 신동민(2003), “수도권 주택구매의 소비자 선호특성 연구”, 국토계획, 38(3):135~146
- [4] 이재영(2003), “수도권 주택 수요 특성 및 구매 결정요인 분석”, 한국지역개발학회지, 15(2):165~182.
- [5] 한국토지주택공사(2014), “지역별 수급추정 및 사업지별 수요추정 방법론 개발과 적용용역”



(그림 3) 주택 수요분석 DB의 유관정보시스템간 연계

5. 결론

주택수요 분석을 위한 DB 구축은 다양하고 복잡한 데이터들을 찾아야하며, 각 데이터들의 연계도 어렵다. 또한, 정확한 주택수요 분석을 위해서는 기초 자료의 신뢰성을 확보할 수 있어야 하는데 국가정보시스템들에서 제공하는 공개 데이터들의 경우 각자의 목적에 맞게 구성되어 있어서 연계하기가 어렵고 분석에 필요한 데이터를 구할 수 없을 때는 대체 데이터들을 확보하고 그에 따라 검증모형이 요구된다. 주택수요 분석의 경우, 일반적으로 일회성으로 엑셀이나 통계 패키지를 활용하여 결과를 얻은 후 재 활용되지 않고 있음에 따라 관련 데이터들을 축적하여 정기적으로 주택수요 분석이 자동화될 수 있도록 DB 구축이 필요하며 시계열적인 분석이 가능하도록 기존 분석결