

지리시간가중 회귀모형을 이용한 주택가격 영향요인 분석[†]

박세희¹ · 김민수² · 백장선³

¹²³전남대학교 통계학과

접수 2016년 12월 30일, 수정 2017년 1월 17일, 게재확정 2017년 1월 19일

요약

본 연구는 아파트 개별 실거래가격에 대한 시공간 자료를 활용하여 아파트 매매가격에 영향을 미치는 요인을 시계열적 흐름과 공간적 변화를 반영한 지리시간가중 회귀모형 (geographical temporal weighted regression; GTWR)모형을 적용하여 분석하였다. 기존 연구에서 활용되었던 일반적인 접근방법인 최소제곱 (ordinary least square; OLS) 회귀모형과 공간 데이터를 분석하기 위한 공간계량 모델 중 가장 많이 활용되고 있는 지리가중 회귀모형 (geographically weighted regression; GWR)과 달리 GTWR은 주택가격 특성을 고려함에 있어서 시간과 공간을 함께 고려함으로써 보다 정밀한 평가모형이 될 것으로 기대되었다. 본 연구에 사용된 주택가격결정 설명 요인들 중에서 건축연도 및 전용면적이 주택가격을 결정하는데 유의적인 영향을 미치는 것으로 나타났으며, 주택가격이 시간적 공간적 특성 모두에 의하여 유의적으로 설명되었다.

주요용어: 공간계량모형, 아파트 실거래가, 지리가중회귀모형, 지리시간가중회귀모형.

1. 머리말

주택 가격을 결정함에 있어 지리적 위치와 시간은 가장 중요한 요소이다. 지리적 위치의 경우 부동산이라는 재화의 비이동성 특성상 같은 공간을 공유하는 가까운 지역끼리 서로 영향을 주며 비슷한 가격을 형성하게 된다. 우리나라의 경우 행정구역 기준으로 나누어지는 공간을 생각해본다면 역세권이나 개발 지역 행정구역을 중심으로 공간적으로 인접해 있는 공간일수록 더 많은 영향을 받아 높은 가격을 형성하는 반면, 인접하지 않고 거리가 멀어질수록 주택 가격에 적은 영향을 미치며 낮은 가격을 형성하게 되는 이유이다. 시간에 따른 주택가격은 변동 폭이 크지 않지만 항상 과거에 형성된 주택가격을 기준으로 현재의 주택가격 형성에 영향을 미치므로 시간적 요소 또한 중요하게 고려해야 할 사항이다.

지금까지 국내 주택가격에 대한 선행연구는 대부분 최소제곱회귀모형 및 공간계량 모형 중 하나인 지리가중 회귀모형 (GWR)을 주로 활용한 연구들이 대부분이었다. 지리가중 회귀모형의 원리는 회귀점으로부터 인접한 지역일수록 더 많은 영향을 받아 상대적으로 큰 가중치를 주는 원리로 개별 지역마다의 회귀계수를 추정할 수 있지만 특정 시점 기준으로 개별 지역마다 회귀분석을 실시하므로 시간의 요소는 고려되지 않는 분석 방법이다.

Kang (2010)은 GWR 접근법을 서울시 아파트 실거래 자료에 적용하여 주택 특성 (층, 면적), 교통 및 입지특성 (지하철 역 직선거리 등), 어메니티 특성 (공원 직선거리), 교육특성 등의 요인이 아파트 실

[†] 본 결과물은 농림축산식품부의 재원으로 농림수산식품기술기획평가원의 첨단생산기술개발사업 (315012-03-1)의 부분 지원 및 농촌진흥청 연구사업 (PJ01182303)의 부분 지원을 받아 연구되었음.

¹ (61186) 광주광역시 북구 용봉로 77, 전남대학교 통계학과, 석사과정.

² (61186) 광주광역시 북구 용봉로 77, 전남대학교 통계학과, 교수.

³ 교신저자: (61186) 광주광역시 북구 용봉로 77, 전남대학교 통계학과, 교수. E-mail: jbaek@jnu.ac.kr

거래가격에 미치는 영향이 각기 다른 여러 주택하위시장으로 구성되어 있음을 밝혔다. Oh 등 (2014)은 부산광역시 공동주택을 대상으로 노령화지수, 경제인구수, 개별공시지가, 사용승인년도 등의 요인이 매매가와 전세가에 미치는 영향이 지역별로 다르게 나타남을 분석을 통해 보이며 부동산 시장의 차별화 현상에 대해 미시적인 접근 분석방법으로 GWR모형이 효율적이라고 판단하였다. 그 외 Lee와 Park (2015)은 공간시차모형 및 공간오차모형을 활용하여 대구, 경북 지역 단위면적당 아파트 가격을 예측하여 두 모형의 성능을 비교평가 하였다. Kim (2015)은 서울시 행정동들에 대해 아파트 전용면적별 실거래 매매가를 기준으로 군집분석을 하였다.

주택가격에 영향을 미치는 요인의 경우, Hong과 Lee (2015)는 서울시 아파트 가격에 영향을 미치는 요인으로 단지요인, 환경요인, 시설요인, 교육요인으로 구분하여 이들 요인을 2003년에서 2013년까지 11개년동안 검토해본 결과, 그 영향력이 시공간적으로 변화하고 있는 패턴을 분석을 통해 보였다.

본 연구에서는 지리적 위치와 시간적 요소를 포함하여 지리가중 회귀모형을 확장시킨 지리시간가중 회귀모형을 이용하여 광주광역시 82개 행정동별 아파트 실거래 가격에 영향을 미치는 요인에 대해 분석하였다.

본 논문은 제2절에서 지리가중 회귀모형과 시간적 요인을 고려하여 확장시킨 지리시간가중 회귀모형에 대해 설명하고, 제 3절에서는 분석자료 설명 및 분석결과에 대해 해석한다. 제 4절에서는 결론 및 향후 연구과제에 대하여 제안한다.

2. 분석모형

2.1. 지리가중회귀모형 (Geographically weighted regression model)

지리가중회귀모형은 전통적인 최소제곱모형 추정방법을 확장하여 지역별로 회귀계수를 추정하는 방법으로써 모형 식은 다음과 같다.

$$Y_{ij} = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) X_{ijk} + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m \quad (2.1)$$

여기서 Y_{ij} 는 i 번째 지역의 j 번째 관측 값의 주택가격, X_{ijk} 는 그 관측 값의 k 번째 설명변수, (u_i, v_i) 는 i 번째 지역의 지리적 위치를 나타내는 공간적 좌표, $\beta_0(u_i, v_i)$ 는 절편계수를 나타내며, $\beta_k(u_i, v_i)$ 는 i 번째 지역의 k 번째 회귀계수를 나타낸다.

최소제곱추정모형과 다르게 이 모형의 특징은 지역적으로 개별 회귀계수를 추정함으로써 지역적 영향력을 파악할 수 있다는 장점이 있다. 위 모형의 회귀계수를 추정하기 위해서는 지리적 위치 i 지역을 기준으로 거리가 가까울수록 거리가 먼 곳 보다 공간적으로 더 많은 영향을 받는 것을 가정으로 한다. 따라서 회귀계수 $\hat{\beta}(u_i, v_i)$ 의 추정은 다음의 식으로 계산되어 진다.

$$\hat{\beta}(u_i, v_i) = \left[X^T W(u_i, v_i) X \right]^{-1} X^T W(u_i, v_i) Y \quad (2.2)$$

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nm} \end{pmatrix}, X = \begin{pmatrix} 1 & x_{111} & \cdots & x_{p11} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{11m} & \cdots & x_{p1m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n1} & \cdots & x_{pn1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1nm} & \cdots & x_{pnm} \end{pmatrix}, \hat{\beta}(u_i, v_i) = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_{p-1} \\ \widehat{\beta}_p \end{pmatrix}$$

Y 는 n 개 지역별 m 번 관측된 주택가격을 나타내는 종속변수 관측값인 $(nm \times 1)$ 행렬, X 는 독립변수 관측값들로 이루어진 $(nm \times (p+1))$ 행렬이다.

$W(u_i, v_i) = \text{diag}(\text{diag}(\alpha_{i1}, \alpha_{i1}, \dots, \alpha_{i1}), \text{diag}(\alpha_{i2}, \alpha_{i2}, \dots, \alpha_{i2}), \dots, \text{diag}(\alpha_{in}, \alpha_{in}, \dots, \alpha_{in}))$ 는 공간가중행렬을 나타내는 $(nm \times nm)$ 대각행렬로써 각각의 부분 대각행렬 $\text{diag}(\alpha_{ij}, \alpha_{ij}, \dots, \alpha_{ij})$ 는 각 지역의 관측개수 m 개를 가지며 요소들은 지리적 위치 i 지역을 기준으로 거리의 영향력을 나타낸다. 따라서 가중행렬은 데이터 위치 i 번째 지역의 j 번째 관측 값마다 계산되어 진다.

2.1.1. 공간가중행렬

공간가중행렬 $W(u_i, v_i)$ 는 지리적 위치 i 지역의 개별적 공간적 영향력을 나타내는 행렬로써 i 지역을 기준으로 거리가 가까운 지역일수록 더 많은 가중치를 받는 행렬로 구성되어 진다. 가중행렬의 요소인 α_{ij} 가중치를 결정하는 커널함수로는 정규분포를 이용한 Gaussian 커널함수, Cleveland와 Devlin (1988)가 제안한 Tricube 커널함수, exponential 커널함수 등 다양하다 (Lee 등, 2006). 그 중 가장 보편적으로 사용되는 exponential 커널함수는 다음의 식과 같다.

$$\alpha_{ij} = \sqrt{\exp(-d_{ij}/h)} \quad (2.3)$$

식 (2.3)에서 d_{ij} 는 i 지역에서 다른 지역 j 까지의 유클리드 거리를 나타내며, h 는 양수으로써 대역폭 (bandwidth)를 나타내는 모수이다. 대역폭은 i 지역과 j 지역 간의 거리 d_{ij} 에 따른 가중치의 민감도를 나타내며, 대역폭 h 값이 커질수록 회귀점 i 지역을 기준으로 더 넓은 범위의 데이터가 가중이 되며, h 값이 작아질수록 회귀점 i 지역을 기준으로 더 좁은 범위의 데이터가 가중이 되는 구조를 갖고 있다.

2.1.2. 최적의 대역폭 (bandwidth) 결정

GWR모형을 통해 회귀계수를 추정함에 있어 공간가중행렬 $W(u_i, v_i)$ 를 정의하기 위해 대역폭 결정이 가장 우선적으로 실시되어야 한다. 최적의 대역폭을 찾기 위한 과정은 Cross-validation을 통해 계산되어지며, GWR 모형에서 y_i 의 실제값과 예측값과의 차이의 제곱 함수를 대역폭 h 에 관한 함수로 표현하며 이 값을 최소화하는 최적의 대역폭 h 를 찾는다. 이때에 i 번째 지점에서의 본인 자신과의 거리는 0으로써 가중치가 1에 가깝게 나타나므로 i 번째 지점에서의 본인 자신의 가중치를 제외하는 방법으로 구하게 된다.

$$CVRSS(h) = \sum_i (y_i - \hat{y}_{-i}(h))^2 \quad (2.4)$$

2.2. 지리시간가중회귀모형 (Geographical and temporal weighted regression model)

앞에서 설명한 GWR모형은 회귀점 i 지점을 기준으로 다른 지점마다의 거리를 기준으로 가중행렬 $W(u_i, v_i)$ 를 구축하여 회귀계수를 추정하는 공간적 영향이 중요한 모형이지만, 지리시간가중 회귀모형 (GTWR)은 GWR 모형에 시간 변수를 포함함으로써 지점간의 공간적 거리뿐만 아니라 시간적 차이를 모형에 포함시킨 형태로서 모형식은 다음과 같다.

$$Y_{ij}^{(t)} = \beta_0(u_i, v_i, t) + \sum_{k=1}^p \beta_k(u_i, v_i, t) X_{ijtk} + \epsilon_{ijt} \quad (i=1, \dots, n, j=1, \dots, m) \quad (2.5)$$

$Y_{ij}^{(t)}$ 는 i 번째 지역의 j 번째 관측 값의 t 시점의 주택가격, X_{ijtk} 는 그 관측 값의 k 번째 설명변수, (u_i, v_i, t) 는 i 번째 지역의 지리적 시간적 위치를 나타내는 공간적 좌표, $\beta_0(u_i, v_i, t)$ 는 절편계수를 나타내며, $\beta_k(u_i, v_i, t)$ 는 i 번째 지리적 시간적 위치에 대한 k 번째 회귀계수를 나타낸다.

여기서 시간이 포함된 (u_i, v_i, t) 는 i 번째 지점의 공간 및 시간을 나타내는 좌표로 나타내며, 특정 t 시점의 회귀계수 추정치는 식 (2.6)과 같다. t 시점에서 i 지역의 회귀계수를 추정하기 위해 t 시점 자료와 t 시점 이전 최근 q 개 시점 자료를 활용한다고 할 때, Y 는 n 개 지역별 $q+1$ 개 시점별 m 번째 관측된 주택 가격을 나타내는 종속변수 관측값인 $(nm(q+1) \times 1)$ 행렬, X 는 독립변수 관측값인 $(nm(q+1) \times (p+1))$ 행렬이다.

$$\hat{\beta}^t(u_i, v_i, t) = [X^T W(u_i, v_i, t) X]^{-1} X^T W(u_i, v_i, t) Y \quad (2.6)$$

$$Y = \begin{pmatrix} y_{11}^{(t)} \\ \vdots \\ y_{1m}^{(t)} \\ y_{21}^{(t-1)} \\ \vdots \\ y_{2m}^{(t-1)} \\ \vdots \\ y_{n1}^{(t-q)} \\ \vdots \\ y_{nm}^{(t-q)} \end{pmatrix}, X = \begin{pmatrix} 1 & x_{111}^{(t)} & \cdots & x_{p11}^{(t)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{11m}^{(t)} & \cdots & x_{p1m}^{(t)} \\ 1 & x_{121}^{(t-1)} & \cdots & x_{p21}^{(t-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{12m}^{(t-1)} & \cdots & x_{p2m}^{(t-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n1}^{(t-q)} & \cdots & x_{p1n1}^{(t-q)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1nm}^{(t-q)} & \cdots & x_{pnm}^{(t-q)} \end{pmatrix}, \hat{\beta}^t(u_i, v_i, t) = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_{p-1} \\ \widehat{\beta}_p \end{pmatrix}$$

지리시간가중 회귀모형의 가중행렬은 시간을 고려한 모형이기 때문에 지리가중회귀모형과 달리 i 지점 시점을 기준으로 다른 모든 시간을 가중하는 형태가 아닌, t 시점이전 과거 $t - q$ 시점까지의 데이터들은 가중시키고 t 시점보다 더 미래의 시점의 데이터 들은 가중시키지 않는 모형으로 지역 및 시점을 고려한 가중행렬을 나타내면 다음과 같은 대각행렬로 표현할 수 있다.

$$W(u_i, v_i, t) = \begin{pmatrix} daig(\alpha_{i1(\cdot)}^t) & & \cdots & & 0 \\ & daig(\alpha_{i2(\cdot)}^t) & & & \\ & & \ddots & & \\ \vdots & & & daig(\alpha_{in(\cdot)}^t) & \vdots \\ & & & & \ddots \\ 0 & & & & daig(\alpha_{i1(\cdot)}^{t-q}) \\ & & & & \ddots \\ & & & & daig(\alpha_{in(\cdot)}^{t-q}) \end{pmatrix} \quad (2.7)$$

이때 $daig(\alpha_{i1(\cdot)}^t) = diag(\alpha_{i1(1)}^t, \alpha_{i1(2)}^t, \dots, \alpha_{i1(m)}^t)$ 이다. $\alpha_{ii'(j)}^t$ 는 i 지역과 i' 지역의 t 시점에서의 j 번째 관측값과의 가중값을 나타낸다. 그런데 i' 지역의 모든 관측값들은 동일한 지역에 위치하고 있으므로 모든 j 에 대하여 동일한 가중값을 갖게된다. 따라서 우리는 $\alpha_{ii'(j)}^t$ 를 $\alpha_{ii'}^t$ 로 동일하게 표현한다.

위 가중행렬은 특정 t 시점을 기준으로 회귀계수를 추정함에 있어 가중행렬의 요소들은 자신이 포함된 t 시점의 다른 지역별 가중값 부터 $t - q$ 시점까지의 지역별 가중값을 포함한 대각행렬로 표현되며, GTWR모형 또한 GWR모형과 동일하게 exponential 커널함수를 사용하였으며 각 요소들은 다음의 식

으로 계산되어 진다.

$$\alpha_{ij} = \sqrt{\exp(-d_{ij}/h)} \tag{2.8}$$

$$d_{ij} = \sqrt{\lambda [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu (t_i - t_j)^2} \tag{2.9}$$

$$\frac{d_{ij}}{\sqrt{\lambda}} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2 + \tau (t_i - t_j)^2} \tag{2.10}$$

위의 식 (2.10)에서 λ 와 μ 는 공간단위 및 시간단위를 조정해주는 모수로써 시간적 변동성이 없을 경우, $\mu = 0$ 으로써 공간적 변동성만 고려한 GWR모형으로 판단하며, 반대의 경우 $\lambda = 0$ 으로써 시간적 변동성만 고려한 TWR모형이라고 판단한다 (Bo Huang, 2010).

따라서 시간과 공간을 고려한 GTWR모형을 구축할 때 모수 추정 개수를 줄이기 위해 $\tau = \mu/\lambda$ 로 놓으면 이는 공간과 시간과의 단위 차이를 보정해주는 모수가 된다. 예를 들어 이 논문에서 사용된 자료의 경우 지리적 좌표계는 광주광역시 지리정보 좌표계의 경우 x 축 및 y 축은 미터 (m)단위이며 시간단위는 년도이다. 따라서 시공간 단위의 차이 보정을 τ 라는 모수로 변화시킴으로써 시공간 단위차이를 보정할 수 있다.

GTWR 모형에서의 대역폭은 GWR모형에서 사용되어진 Cross-Validation과 동일하게 계산하여 구하며, 추가로 시공간 단위 보정 모수인 τ 값 또한, $\tau = 1$ 을 기준으로 최적의 대역폭 h 를 찾은 후, τ 값을 변화시키면서 다음의 값을 최소화하는 τ 를 찾는 방식으로 하였다.

$$CVRSS(h, \tau) = \sum_i (y_i - \hat{y}_{-i}(h, \tau))^2 \tag{2.11}$$

3. 실증 분석

3.1. 분석 자료

GTWR 모형에서 반응변수로 사용한 자료는 2006년 1월부터 2014년 12월까지 국토교통부에서 제공하는 광주광역시의 아파트 개별 실거래가 자료이다. 그 중 총 3635개 자료를 분석에 사용된 표본 자료로서 활용하여 GTWR모형을 구축하였다. 복잡한 모형의 계산을 용이하기 위하여 연도별 추정모형은 그 연도 이하 5개년 자료로부터 2000개 내외를 랜덤 추출하여 추정하였다. 지역별 자료로는 아파트가 있는 행정구역별로 총 82개 행정동으로 구분하였으며, 시간단위로는 1월부터 12월까지를 묶어 년단위 자료로 구축하였다 (Table 3.1)

Table 3.1 Sample size and average house price in Gwangju, 2006-2014

Year	Number of regions	Sample size	Average house price (₩10,000)
2006	77	385	8,128
2007	78	390	8,745
2008	81	405	9,970
2009	81	405	9,915
2010	82	410	11,329
2011	82	410	12,407
2012	82	410	12,909
2013	82	410	13,671
2014	82	410	14,962

모형에 사용된 설명변수로는 국내 선행 연구된 설명변수들을 참고하여 선정하였다 (Hong과 Lee, 2010). 아파트 건축년도, 아파트 전용면적, 행정구역 내 역세권의 영향력 을 나타내는 역세권 아파트

개수, 아파트 평균 세대수, 보육시설 개수, 1인당 평균 공원 면적을 고려하였다. 지역별 자료인 보육시설 개수, 1인당 평균 공원 면적은 행정구별 통계연보 자료를 활용하였다.

Table 3.2 Variables of GTWR regression model, 2006-2014

Variables	
y	House price (₩10,000)
x_1	Year of construction (year)
x_2	Floor area (m^2)
x_3	Number of apartment complexes near subway station
x_4	Average number of households
x_5	Number of day care centers
x_6	Average park area per person (m^2)

분석에 사용된 데이터의 시간은 회귀시점 년도를 포함한 과거 4년까지 총 5개 년도를 데이터 구축에 활용하였고 그에 따른 관측 자료 개수는 아래의 표와 같다 (Table3.3)

Table 3.3 Number of data at time point t

Data	Time point (year)	Number of observations (time period)
apt2010	2010	1995 (2006-2010)
apt2011	2011	2020 (2007-2011)
apt2012	2012	2040 (2008-2012)
apt2013	2013	2045 (2009-2013)
apt2014	2014	2050 (2010-2014)

3.2. GTWR 모형 조정을 위한 대역폭 및 τ 값 추정

다음은 GTWR모형의 최적의 대역폭 및 τ 값을 찾는 과정을 나타낸 표이다. 대역폭을 찾기 위해 $\tau = 1$ 로 초기값을 설정한 후 위의 Cross-Validation (2.11)식을 최소로 하는 최적의 대역폭을 찾은 후, τ 값을 변화시키면서 (2.11)식을 최소로 하는 τ 값을 찾는 과정을 아래의 표에 나타내었다. 그 결과 GTWR2010 모델과 GTWR2013 모델은 $\tau = 8$, 그 외 모형은 $\tau = 10$ 으로 최적의 모수가 선택 되었다 (Table 3.4)

Table 3.4 $CVRSS(h, \tau)$

Model	Bandwidth	$\tau = 1$	$\tau = 3$	$\tau = 8$	$\tau = 10$	$\tau = 12$	$\tau = 15$
GTWR2010	53	2701.0	2663.0	2552.5	2555.4	2566.6	2581.2
GTWR2011	52	3168.8	3094.3	2743.1	2700.3	2700.6	2710.5
GTWR2012	56	3327.4	3210.7	2686.8	2637.5	2644.6	2661.3
GTWR2013	54	2852.6	2716.8	2455.8	2496.3	2518.1	2536.1
GTWR2014	55	3112.4	2992.1	2584.5	2544.3	2544.1	2550.4

3.3. 모형 비교 및 평가

GTWR모형을 다른 모형과 비교하기 위하여 각 회귀시점별 데이터를 GWR, TWR모형을 적용하여 각 모형별 대역폭 h 와 RMSE값을 비교해 보았다. 그 결과 TWR모형의 경우 대역폭이 2로써 동일한 값을 나타내며 RMSE값이 GWR모델보다 높게 나타났다, GWR모형은 GTWR 모형과 비슷한 대역폭을 나타냈지만 GTWR모델보다는 RMSE값이 더 높게 나타났다.

Table 3.5 Comparison of RMSEs of TWR, GWR and GTWR Models

Data	TWR		GWR		GTWR		
	h_t	RMSE	h_s	RMSE	h_{st}	τ	RMSE
apt2010	2	3486.65	51	2225.32	53	8	1813.12
apt2011	2	3560.87	52	2700.18	52	10	1666.67
apt2012	2	3573.88	54	2908.54	56	10	1822.41
apt2013	2	2832.90	53	2630.36	54	8	1786.57
apt2014	2	3501.49	53	2858.00	55	10	1908.31

GTWR모형에 대하여 다른 모형들 보다 유의한 모형인지 평가하기 위해 Xuan 등 (2015)이 제안한 모형검정 방법을 통해 모형을 비교 검정하였다.

첫 번째로 검정방법으로 OLS모형과 GTWR모형을 비교하고, 두 번째로는 공간적 영향을 배제한 TWR 모형과 비교, 마지막으로 시간적 요인을 배제한 GWR모형과 비교하였다. 검정 가설을 정리하면 다음과 같다.

1. Test for Global stationarity

$$\begin{cases} H_0^{(1)} : y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i \\ H_1 : y_i = \sum_k \beta_k(u_i, v_i, t_i) x_{ik} + \epsilon_i \end{cases}$$

$$F_1 = [RSS(H_0^{(1)}) - RSS(H_1)] / RSS(H_1) \tag{3.1}$$

2. Test for Spatial non-stationarity

$$\begin{cases} H_0^{(2)} : y_i = \sum_k \beta_k(t_i) x_{ik} + \epsilon_i \\ H_1 : y_i = \sum_k \beta_k(u_i, v_i, t_i) x_{ik} + \epsilon_i \end{cases}$$

$$F_2 = [RSS(H_0^{(2)}) - RSS(H_1)] / RSS(H_1) \tag{3.2}$$

3. Test for Temporal non-stationarity

$$\begin{cases} H_0^{(3)} : y_i = \sum_k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \\ H_1 : y_i = \sum_k \beta_k(u_i, v_i, t_i) x_{ik} + \epsilon_i \end{cases}$$

$$F_3 = [RSS(H_0^{(3)}) - RSS(H_1)] / RSS(H_1) \tag{3.3}$$

세 가지 가설들에 대한 검정결과, GTWR2010 모형을 제외하고 모든 모형에서 GTWR모형에 대한 유의한 결과가 나타났으며 (유의수준 0.05), 따라서 GTWR 모형이 다른 모형보다 자료에 더 적합함을 알 수 있다. 다만 GTWR2010모형의 경우 시간적 영향을 배제한 GWR모형과의 비교결과인 F_3 의 검정이 유의하지 않은 결과를 나타냈다. 이는 2006년-2010년 자료의 경우 시간적 영향력이 유의하지 않음을 의미한다.

Table 3.6 Tests for GTWR regression Model

F -test	F_1	p -value	F_2	p -value	F_3	p -value
GTWR2010	3.670	<0.001	3.182	<0.001	0.506	1
GTWR2011	4.927	<0.001	3.565	<0.001	1.625	<0.001
GTWR2012	3.793	<0.001	2.846	<0.001	1.547	<0.001
GTWR2013	3.034	<0.001	1.995	<0.001	1.168	<0.001
GTWR2014	3.038	<0.001	2.367	<0.001	1.243	<0.001

다음으로는 설명변수들에 대한 유의성 검정으로써 주택가격에 영향을 미치는 요인이 어떤 것인지 분석하였다. 연도별 설명변수 유의성 검정결과 건축년도, 전용면적이 유의하게 나타났으며, 건축년도의 경우 2010년 및 2011년을 제외한 2012년부터 2014년까지 모형에서 유의하게 나타났다.

4. Test for significance of variables

$$\begin{cases} H_0^{(4)} : y_i = \sum_{k \neq j} \beta_k(u_i, v_i, t_i)x_{ik} + \epsilon_i \text{ (except } x_j \text{ variable)} \\ H_1 : y_i = \sum_k \beta_k(u_i, v_i, t_i)x_{ik} + \epsilon_i \text{ (full model)} \end{cases}$$

$$F_4 = [RSS(H_0^{(4)}) - RSS(H_1)] / RSS(H_1) \tag{3.4}$$

Table 3.7 Test for the significance of the explanatory variable

<i>F</i> -test		x_1	x_2	x_3	x_4	x_5	x_6
GTWR2010	F_4	1.005	3.149	0.019	0.026	0.024	0.019
	<i>p</i> -value	0.468	<0.001	1.000	1.000	1.000	1.000
GTWR2011	F_4	1.109	3.598	0.018	0.051	0.061	0.035
	<i>p</i> -value	0.064	<0.001	1.000	1.000	1.000	1.000
GTWR2012	F_4	1.122	2.591	0.016	0.034	0.038	0.050
	<i>p</i> -value	0.043	<0.001	1.000	1.000	1.000	1.000
GTWR2013	F_4	1.277	2.612	0.023	0.036	0.049	0.041
	<i>p</i> -value	<0.001	<0.001	1.000	1.000	1.000	1.000
GTWR2014	F_4	1.175	3.314	0.035	0.032	0.054	0.038
	<i>p</i> -value	0.007	<0.001	1.000	1.000	1.000	1.000

따라서 최종 선택된 설명변수들로 구축된 모형의 회귀계수를 추정한 결과는 다음의 표와 같이 요약하였다 (Table 3.8)

Table 3.8 GTWR parameter estimate summaries for 2010-2014 models

Model	GTWR2010 (bandwidth=53)				
	Min	LQ	Med	UQ	Max
Intercept	-2097851	-934865	-726400	-602740	-196223
Floor area	82.72	122.39	154.22	196.79	306.60
Model	GTWR2011 (bandwidth=52)				
	Min	LQ	Med	UQ	Max
Intercept	-1337780	-970584	-728006	-594009	-1320
Floor area	68.52	132.55	158.90	227.23	418.80
Model	GTWR2012 (bandwidth=56)				
	Min	LQ	Med	UQ	Max
Intercept	-1302338	-863452	-710583	-535318	-199253
Year of construction	94.06	266.57	354.38	430.76	656.27
Floor area	77.1	133.7	166.1	219.3	318.4
Model	GTWR2013 (bandwidth=54)				
	Min	LQ	Med	UQ	Max
Intercept	-2063911	-860546	-759674	-629973	-296690
Year of construction	145.1	316.5	380.7	431.2	1031.8
Floor area	104.1	147.2	176.1	200.3	298.0
Model	GTWR2014 (bandwidth=55)				
	Min	LQ	Med	UQ	Max
Intercept	-2031234	-916243	-779590	-635487	-300758
Year of construction	149.3	317.5	390.4	458.1	1015.3
Floor area	99.08	169.45	189.67	205.50	310.55

주택가격 영향요인의 시간적 변화를 살펴보기 위하여 2010년과 2014년의 아파트 전용면적의 추정 회귀계수를 지역별로 차트를 그려본 결과가 Figure 3.1에 나타나 있다. 2010년 서남권의 전용면적이 아파트 가격에 큰 영향을 미치는 반면, 2014년의 경우 신지구가 개발된 북부권을 중심으로 전용면적의 아파트 가격에 미치는 영향력이 매우 크게 변화하였음을 알 수 있다.

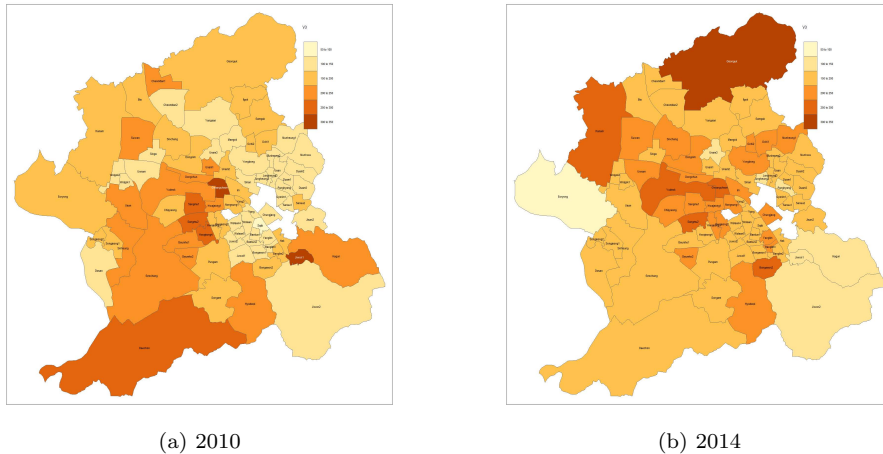


Figure 3.1 The regional plot for the regression coefficients of floor area

4. 결론 및 향후 연구과제

본 연구는 2010년부터 2014년까지의 광주광역시 행정동 대상으로 건축년도, 전용면적, 역세권 아파트 개수, 아파트 평균 세대수, 보육시설 개수, 1인당 평균 공원 면적이 아파트 가격에 지역별로 어떠한 영향을 미치는지 공간적 시간적 특성을 고려하여 분석하였다. 분석을 위해 지리시간가중 회귀모형을 이용하여 공간 및 시간적 특성을 분석하였으며, 그 결과 아파트 가격에는 공간적 및 시간적 이질성이 존재하여 지역 및 연도별로 국지적인 회귀계수를 추정할 필요가 있다고 판단되었다. 지리시간 가중 회귀모형인 GTWR은 기존의 주택가격 평가모형이었던 GWR 및 OLS 모형보다 더 낮은 RMSE값을 나타냄으로써 보다 더 정밀한 주택가격 평가모형을 추정할 수 있었으며, 비교 모형에 대한 검정 평가 결과 또한 2010년 GTWR모형의 시간적 이질성이 유의하지 않음을 제외하고는 2011년부터 2014년까지의 GTWR모형은 시간적 공간적으로 유의한 영향력을 나타냈다. 설명변수에 대한 유의성 검정 결과에서는 건축년도 및 전용면적의 경우에만 의미있는 것으로 나타났다. 따라서 향후 연구에서는 아파트 가격에 영향을 미치는 시간적 공간적 설명변수들을 추가하면 더욱 신뢰도 있는 모형을 구축할 것으로 생각된다.

References

- Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596-610.
- Fotheringham, A. S., Crespo, R. and Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, **47**, 431-452.
- Hong, H. Y. and Lee, J. H. (2015). A time series and spatial analysis of factors affecting housing prices in Seoul. *Seoul City Research*, **16**, 87-108.

- Huang, B, Wu, B. and Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, **24**, 383-401.
- Kang, C. D. (2010). GWR approach for real estate appraisal : The case of Seoul apartment. *Korean Appraisal Review*, **20**, 107-132.
- Kim J. J. (2015). Cluster analysis for Seoul apartment price using symbolic data. *Journal of the Korean Data & Information Science Society*, **26** 1239-1247.
- Lee, S. W., Yoon, S. D., Park, J. Y. and Min, S. H. (2006). *The practice on spatial econometric model*, Pakyoungsa, Seoul
- Lee, W. J. and Park, C. Y. (2015). Prediction of apartment prices per unit in Daegu-Gyeongbuk areas by spatial regression models. *Journal of the Korean Data & Information Science Society*, **26**, 561-568.
- Oh, Y. K., Kang, J. G. and Kim, J. M. (2014). Analysis of regional characteristics that affect housing prices using a GWR model - Focused on Busan metropolitan city. *Tax Accounting Research*, **40**, 1-17.
- Xuan, H., Li, S. and Amin, M. (2015). Statistical inference of geographically and temporally weighted regression model. *Pakistan Journal of Statistics*, **31**, 307-325.

Application of geographical and temporal weighted regression model to the determination of house price[†]

Saehee Park¹ · Minsoo Kim² · Jangsun Baek³

^{1,2,3}Department of Statistics, Chonnam National University

Received 30 December 2016, revised 17 January 2017, accepted 19 January 2017

Abstract

We investigate the factors affecting the price of apartments using the spatial and temporal data of private real estate prices. The factors affecting the price of apartment were analyzed using geographical and temporal weighted regression (GTWR) model which incorporates the temporal and spatial variation. In contrast to the OLS, a general approach used in previous studies, and GWR method which is most widely used for analyzing spatial data, GTWR considers both temporal and spatial characteristics of the house price, and leads to better description of the house price determination. Year of construction and floor area are selected as the significant factors from the analysis, and the house price are affected by them temporally and geographically.

Keywords: Geographical and temporal weighted regression, geographically weighted regression, spatial nonstationarity, temporal nonstationarity.

[†] This research is partially supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) (315012-03-1) and Development of Science and Technology for the Korean Agriculture and Beyond research (PJ01182303) fund.

¹ Master student, Department of Statistics, Chonnam National University, Gwangju 61186, Korea.

² Professor, Department of Statistics, Chonnam National University, Gwangju 61186, Korea.

³ Corresponding author: Professor, Department of Statistics, Chonnam National University, Gwangju 61186, Korea. E-mail: jbaek@jnu.ac.kr