

研究論文

小地域別地価水準のローカル回帰モデル推定
～埼玉県さいたま市を例として

植 杉 大

Local Model Estimation of Land Price Level in Each Small Area:
A Case Study in Saitama City

Dai Uesugi

要 旨

本論文では、埼玉県さいたま市の地価公示データに対して、地理的加重回帰モデル (Geographically Weighted Regression model:GWR model) および複数のグローバル回帰モデルを適用し、地価水準を推定するモデルを比較検討した。複数のグローバル回帰モデルには、ベンチマークとして通常の OLS 回帰を、また空間計量経済学分析で頻繁に用いられる空間自己回帰モデルおよび空間誤差モデルを採用した。結果的に、複数のグローバル回帰モデルと比較して、ローカル回帰モデルである GWR を適用することによって地価水準の推定精度を改善させることが確認できた。また、GWR によって求められたパラメータ推定値を丁目区分ごとに地図上で可視化することによって、グローバル回帰モデルでは不可能な、小地域に係る各説明変数の土地価格に与える影響の相違を把握することができた。

1 はじめに

一般的に、地価形成の要因分析や地価水準の変動、特定地点の地価の予測等を行う場合、ヘドニック回帰モデルが用いられる。ヘドニック・アプローチは、財・サービスの有する各品質の合成がその価格を与えると考えるものである。したがって、土地や不動産のように、市場に同質的なものが存在せず差別化や多様化の著しい財の場合、ヘドニック回帰モデルは非常に強力な分析ツールである。さらにその実証分析方法は、対象となる土地や不動産の価格を、CBDへの時間距離、最寄駅からの時間距離や面積、属する地域の様々な特性などに回帰する極めて単純なものであるため、ヘドニック回帰モデルを通常の最小二乗法 (Ordinary Least Square:OLS) で推定するといった研究が従来なされてきた。

しかし不動産は同一用途に基づいて社会的・地域的にまとまりを有している。したがって個別土地の価格形成も地域的なまとまりを示し、近隣地域に属する土地や不動産の価格は強い相関を持つことが予想される。そのためサンプルの独立性という、OLS推定の前提が満たされないという批判から、空間的相関を考慮した計量経済モデルの研究がおこなわれてきた。この空間計量経済学における代表的なモデルとして、空間自己回帰モデル (Spatial Autocorrelation Regression model:SAR model) や空間誤差モデル (Spatial Error Model:SEM) などが挙げられる。

上記のいずれも、研究の対象となる地域に属する標本データを用いて、その地域全体に係るパラメータ推定を行う。そこでは、対象地域の空間的な同質性を強調し標本データに基づく情報を集約するとともに、当該地域における一般的な法則性を見出すことに重点が置かれる。したがって、土地の価格を各説明変数に回帰した場合、例えば CBD への時間距離に係る係数パラメータの推定量は地域全体で一つしか求められない。このようなモデルは対象地域全体の推定を行うことから、グローバル回帰モデルと呼ばれる。

これに対して、従来のグローバル回帰モデルにかわり空間的非正常性を仮定したローカル回帰モデルによる研究が行われている。これらの研究は、研究の対象となる地域に属する標本データを用いて、局所的なパラメータ推定を行うことを目的としており、対象地域の一般的な法則性の発見や情報の集約という観点よりも、研究地域内での差異の発見や例外的な地点の発見などを重視する。したがって、このようなモデルは局所的に係数パラメータなどを推定することとなるため、ローカル回帰モデルと呼ばれる。本論文で用いる地理的加重回帰モデル (Geographically Weighted Regression model:GWR model) は、局所多項式回帰モデル (Local Polynomial Regression model:LPR model)¹などと並び、ローカル回帰モデルの代表的なモデルである。

ところで近年、大都市圏と地方との地価動向が二極化しており、大きな経済的問題とされている。このような二極化現象は一国全体だけではなく、非常に小さなエリア内においても見られる現象である。本論文で取り上げた埼玉県さいたま市についても、市の北部と南部では東京

¹ 局所多項式回帰モデルについては、パラメータのベイズ推定を行っているものとして Clapp, Kim and Gelfand(2002)[4]が挙げられる。

都心へのアクセスの容易さ等により、また市の中央部と東部・西部では、幹線鉄道への近接性や地域の自然環境等によりその性質を大きく異にしている。市の中心である浦和区や大宮区においては、駅周辺の再開発により業務地区の集積が進み、さらに都心通勤者をターゲットとしたマンション開発が盛んであるため、マクロ土地価格動向と比較しても、土地価格は微増している。一方市の周縁部、特に岩槻区や見沼区などにおいては、新築戸建住宅を中心に値崩れが激しく、新価格、さらには新・新価格と価格を下げたとしても買い手が現れないという現象が多く観察される。

開発業者など不動産の売り手が土地の仕入れを行う際に問題となるのは、どれくらいの値段で仕入れればいいのかはもちろんのこと、どの地点の土地を選択すればよいのかである。買い手が将来の地価動向を予測して物件購入を検討しているとすれば、資産価値が下落しそうな地域の物件購入は見送るはずである。売り手の開発業者は、土地を仕入れて建物を建て、何らかの付加価値を付けて販売することを生業としているので、地域的に多少不利であっても付加価値の付け方や営業努力等で契約締結にこぎつけようとする。すると、仕入価格に関する判断と比較して地点に関する判断を軽視する可能性が高い。この地点あるいは地域に関する特性に関してグローバル回帰モデルでは必要な情報が得られない。さらに空間的非定常性が存在するならば、当該地点あるいは地域について、ヘドニック回帰モデルの属性に係るパラメータがグローバルな推定パラメータと大きく異なる可能性がある。すると、属性値を代入して求めた土地価格の予測値も大きく異なってしまうかもしれない。このような問題は、買い手である不動産需要者にも当てはまる。

以上の考えのもと、本論文ではローカル回帰モデルを用いてさいたま市内の丁目区分別の地価水準の推定方法を提案する。そこでローカル回帰モデルによる推定がどれだけグローバル回帰モデルに対して有効なのかを検討し、さいたま市の土地の価格形成要因を GIS を用いてビジュアル化することでその様子を分析し、市場に対する有益な情報を提供できるかを考察する。

本論文の構成は以下のとおりである。第 2 節では、GWR モデルの構成及びその特徴を概説し、他のグローバル回帰モデルとの比較基準について説明する。第 3 節では、回帰分析に用いるデータの説明を行う。第 4 節では、分析のベンチマークとしての OLS 推定、空間計量経済学分析で頻繁に用いられる空間自己回帰モデルおよび空間誤差モデル、および GWR による推定結果を示し、対象地域の全体的なモデルの当てはまりを見る。そこで、他のグローバル回帰モデルと比較して GWR による推定精度の優位性を確認する。また局所的なモデルの当てはまりを見るために、各地点のローカルパラメータに関する分析、および各地点の決定係数に関する分析を行う。第 5 節では、丁目区分ごとのパラメータ推定量の水準を図示することで、限定的な地域内における説明変数の影響の強弱を確認し、地域不動産情報として有益なものが得られるかを考察する。第 6 節では、論文全体のまとめと今後の研究の方向性について検討する。

2 モデル

本論文では、ローカル回帰モデルとして地理的加重回帰モデル (Geographically Weighted Regression model:GWR)を用いる。このモデルは、Brunsdon, et al.(1996)[2]や Fotheringham, et al.(2002)[5]によって考案されたものであり、対象地域内における全体的傾向を研究する一般のグローバル回帰モデルと異なって、むしろ対象地域内における例外的挙動を示す小地域の探索や地域差を強調する目的に用いられる。したがって、Fotheringham, et al(2002)で示されている例や清水・唐渡(2007)[7]のように、同一財の存在しない土地や不動産に対してはごく自然に適用されると考えられる。しかしグローバル回帰モデルに属する空間計量経済学的アプローチと比較して、GWRの当該分野への適用例はそれほど多くはない。そこで本節では、GWRのモデルに関する説明や検定方法、モデルの評価等について、GWRの教科書の文献であるFotheringham, et al.(2002)に従って紹介を行う。

2.1 GWRについて

グローバル回帰モデルとローカル回帰モデルの相違を端的に言えば、回帰式の説明変数と被説明変数との関数関係が対象地域内で同一であるか、あるいは対象地域内で局所的に変動するかである。より具体的に言えば、回帰分析により対象地域全体で各パラメータが唯一推定されるか、あるいは地点ごとに異なるかということである。ここで特に、空間的に変動する場合を空間的非定常性 (spatial nonstationarity) という。

土地価格や不動産価格に関するヘドニック回帰モデルでは、都心への時間距離や最寄駅からの時間距離、地理的特性を表すダミー変数が説明変数として用いられる。もし空間的非定常性が存在する場合、これらに係る局所的なパラメータ値の間に無視できない差異が生じる可能性が高い。さらに対象地域内に位置するある地点の予測値を求める場合、グローバル回帰モデルに基づいて予測を行うのが妥当でない可能性もある。

空間的非定常性が存在する場合、対象地域で局所的に変化する土地価格や不動産価格を推定するために、パラメトリック手法とセミパラメトリック手法が用いられる。パラメトリック手法の代表的なものとしては、多項式展開モデル (parametric polynomial expansion model) が挙げられる。このモデルは局所的な当てはまりを重視し高次元多項式を用いてパラメータ推定を行う。しかしどの程度の次元の多項式を用いるかの基準は曖昧であり、また次元が増えれば推定パラメータ数も増加するので推定の自由度が低下する。一方セミパラメトリック手法としては、局所回帰モデル、スプライン関数を用いたモデル、LOWESS 回帰モデル、カーネル回帰モデル、そして本論文で用いる GWR などが挙げられる。これらセミパラメトリックあるいはノンパラメトリック手法は、回帰モデルに先験的な関数形を与えず、空間的非定常性を各標本点の位置情報 (座標値) のみで除去することを特徴としている。中でも、GWR を用いる最大の利点は、なじみの深い通常の回帰分析の枠組みを用いているため直観的で解釈が明快であることだといえる。

いま、対象地域内の n 個の地点で収集された標本データセットがあるとする。説明変数の数は定数項を含め $K+1$ 個あるとする。被説明変数を y_i 、説明変数を X_{ik} とする。ここで、 $1 \leq i \leq n$ 、 $0 \leq k \leq K$ である。 ε_i を誤差項とし、正規分布に従うと仮定する。

すると、GWR は以下のようにあらわされる。

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^K \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad (1)$$

ここで、 u_i が緯度、 v_i が経度である²。この式からわかるように、GWR と通常のグローバル回帰モデルとの相違点は、推定パラメータが標本点の位置情報である空間座標の関数となっているため、標本点ごとのローカルパラメータが求められる点である。さらに、標本地点 i からすべての標本地点までの距離を反映した重み行列を \mathbf{w}_i とすれば、上記の式は以下のように行列表記される。

$$\mathbf{w}_i \mathbf{y} = \mathbf{w}_i \mathbf{X} \boldsymbol{\beta}_i + \varepsilon_i \quad (2)$$

ここで、各行列およびベクトルを以下のように定義する。

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nK} \end{pmatrix},$$

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{0i} \\ \vdots \\ \beta_{Ki} \end{pmatrix} = \begin{pmatrix} \beta_0(u_i, v_i) \\ \vdots \\ \beta_K(u_i, v_i) \end{pmatrix}, \varepsilon_i = \begin{pmatrix} \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

$$\mathbf{w}_i = \text{diag}\{w_{i1}, \dots, w_{in}\}$$

したがって、パラメータ推定量は以下の式で求められる。

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}' \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_i \mathbf{y} \quad (3)$$

ここで、 $\mathbf{W}_i = \mathbf{w}_i' \mathbf{w}_i$ である。以降では、この \mathbf{W}_i を空間加重行列と呼ぶ。空間加重行列は、対角成分が標本地点 i からすべての標本地点までの距離を反映した重みであり、非対角成分が 0 の行列である。したがって、以下のようにあらわされる。

$$\mathbf{W}_i = \text{diag}\{w_{i1}^2, \dots, w_{in}^2\} = \begin{pmatrix} W_{i1} & 0 & \cdots & 0 \\ 0 & W_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{in} \end{pmatrix} \quad (4)$$

² 本論文で採用した緯度・経度は、世界測地系によるものである。

この空間加重行列の対角要素 $W_{ij}(1 \leq j \leq n)$ を決定するのがカーネル関数である。カーネル関数は、回帰を行う各地点から標本地点の距離が遠くなるに従い加重を低減させるという発想に基づく。カーネル関数としていくつか提案がなされているが、特に以下のようなガウス型関数がよく用いられている。Brunsdon, et al.(1996)で推奨されているガウス型のカーネル関数は以下のようにあらわされる。

$$W_{ij} = \sqrt{\exp\left(-\left(\frac{d_{ij}}{\theta}\right)^2\right)} \quad (5)$$

ここで d_{ij} は標本地点 i から他の標本地点 j までのユークリッド距離を表わす。また、 θ はカーネルのバンド幅を表わす。最適なバンド幅は、一般的にクロスバリデーション (Cross Validation:CV) に基づき、ヘドニック回帰モデルの理論値と実績値の平方和 (CV 値) を最小化することで求められる。

また、カーネルのバンド幅の設定方法は、固定型カーネルと適応型カーネルとがある。固定型カーネルは、一般化交差確認法 (Generalized Cross Validation:GCV) において GCV 値を最小にするバンド幅を求めた後にこれを固定して用いる方法である。一方適応型カーネルは、初めに回帰を行う地点の近傍にある標本地点の個数を決定し、回帰地点から最も遠い標本地点までの距離をバンド幅とする方法である。したがってバンド幅について、前者は固定的だが、後者は回帰地点ごとに可変的である。

2.2 グローバル回帰モデル

本論文で GWR と比較を行うグローバル回帰モデルとしては、ベンチマークとして OLS、空間計量経済学で頻りに用いられるモデルとして空間自己回帰モデル (Spatial Autocorrelation Regression model:SAR model) と空間誤差モデル (Spatial Error Model:SEM) を取り上げる。OLS については説明を省略し、ここでは SAR と SEM のモデルの構造をごく簡単に概説するにとどめる。

一般化した空間自己回帰モデルは以下のようにあらわされる。

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{\Omega}_1 \mathbf{y} + \mathbf{X}\beta + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{\Omega}_2 \mathbf{u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \quad (6)$$

ここで、 $\mathbf{\Omega}_1, \mathbf{\Omega}_2$ は空間重み行列である³。 ρ は空間ラグのパラメータである。これは時系列分析でもなじみのある自己相関パラメータと同様のものである。さらに λ は誤差項に空間自己

³ SAR や SEM を推定する場合に用いられる空間重み行列は、GWR における空間重み行列とは、その作成方法が異なる。本研究のように標本地点の緯度・経度が判明している場合、一般的に 2 地点間の距離の逆数を近接性の指標として採用し、行和が 1 となるように標準化した空間重み行列を用いることができる。

相関があることを示すパラメータである。したがって、もし $\Omega_2=0$ ならば、次のようになる。

$$\begin{aligned} \mathbf{y} &= \rho\Omega_1\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0},\sigma^2\mathbf{I}_n) \end{aligned} \quad (7)$$

また、 $\Omega_1=0$ ならば、次のようになる。

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta}+\mathbf{u} \\ \mathbf{u} &= \lambda\Omega_2\mathbf{u}+\boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0},\sigma^2\mathbf{I}_n) \end{aligned} \quad (8)$$

前者のモデルが SAR であり、後者のモデルが SEM である。

2.3 GWR と他のグローバルモデルとの比較の方法

はじめに、各モデルの推定に先立ち、対象地域内における空間従属性の有無を検定するために、Anselin(1988)[1]の LM 検定を行う。これは、一般化した空間自己回帰モデル(6)式における $\rho=0$ 、あるいは $\lambda=0$ を検定するものである。

次に、GWR モデルの検定として、2種類の検定を行う。第一の検定は、GWR と OLS の分散分析を行い、GWR の適用によって OLS に対して有意な改良がなされたのを見るものである。その際には、Brunsdon, et al.(1999)[3]による擬似 F 統計量を用いる。第二の検定は、GWR で求められた各パラメータ推定量は OLS 推定量と有意に異なるかを見るものである。つまり、各パラメータ推定量が対象地域内で空間的に変動するかを確認する。そのために、Leung, et al.(2000)[6]による検定を用いる。

さらに、OLS やその他のグローバル回帰モデルと GWR について、グローバル決定係数や赤池情報量基準 (Akaike Information Criteria:AIC) 等の評価基準となる統計量を用いて各モデルの優劣を評価する。グローバル決定係数が大きければ適合度が高いと評価される。AIC は小さければ小さいほど適合度が高いと評価される⁴。

3 データ

本論文では、比較的狭い地域的範囲においても土地価格の二極化が発生しており、それが住宅購入・住宅投資を行う場合に深刻な差異を引き起こす可能性があるとの問題意識のもと、対象地域として埼玉県さいたま市を選択した。さいたま市は埼玉県南部に属し、多くの JR 線等が東西南北に走っているため都心へのアクセスは至便である。市の中央に位置する浦和区や大

⁴ これらの推定及び検定を R で行うコマンド等については、谷村 (2010) [8]を参照のこと。

宮区などでは業務施設が集中し、さらに都心通勤者の居住する駅近のタワーマンションなどが立ち並んでいる。しかしその一方で、交通幹線から離れた周辺部においては市街化調整区域も多く、交通の利便性も低い「陸の孤島」のような地域も存在する。このようにさいたま市という非常に限定された地域においても、さらに細分した小地域ごとに様相は大きく異なっている。したがって空間的非正常性を勘案して、GWR のようなローカル回帰モデルを用いる方が適切な結果をもたらすかもしれない。

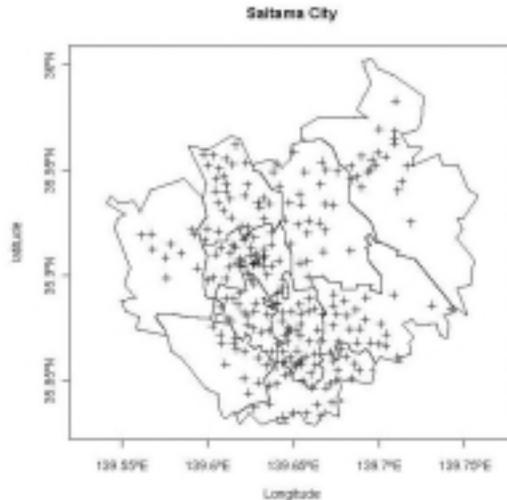


図 1: さいたま市内の標本地点

GWR に必要なデータとしては、被説明変数や説明変数に加えて、空間変数が必要になる。そこで本節では、当該地域に係る各変数の特徴、データの出所などについて説明する。なお、本論文で推定に用いる標本地点は 224 地点であり、その位置は図 1 で示されている。

3.1 被説明変数

土地・不動産にかかるヘドニック回帰モデルの場合、被説明変数はその価格（あるいは賃料）である。本論文では、国土交通省による 2008 年の『地価公示』の $1m^2$ 当たり価格を使用した。地価公示の標本地点はその地域の代表点として選択され、不動産鑑定士による鑑定価格として求められる。標本地点の価格は、最小値が 30,400 円、最大値が 2,550,000 円、平均が 271,400 円であった。その分布は図 2 に示されている。

なお、『地価公示』のデータは、国土交通省国土政策局国土情報課 GIS ホームページの国土数値情報ダウンロードサービスを利用して取得した⁵。

⁵ <http://www.mlit.go.jp/kokudokeikaku/gis/index.html>

小地域別地価水準のローカル回帰モデル推定～埼玉県さいたま市を例として

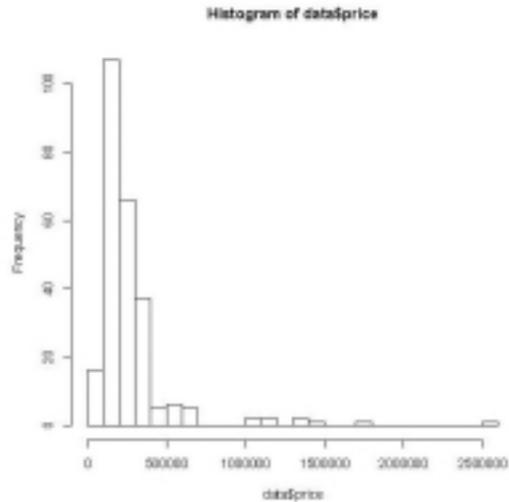


図 2: 標本地点の $1 m^2$ 当たり地価の分布

3.2 説明変数

土地・不動産にかかるヘドニック回帰モデルの場合、説明変数はその標本に係る地理的、社会的属性である。そこで説明変数は、被説明変数と同様、『地価公示』に記載されているデータから採用した。具体的には、(1)最寄駅から都心までの時間距離、(2)最寄駅からの時間距離、(3)地積（対数値）、(4)用途地域ダミー（住宅地域系、商業地域系、工業地域系）、(5)用途ダミー（住居系、店舗系、事務所系）、(6)防火地域ダミー、(7)容積率を用いた。

3.3 空間変数および地図等のデータ

『地価公示』には、標本地点の位置情報として緯度・経度が公表されているため、空間変数としてこれらを採用した。

本論文中で用いられる地図等のデータファイルは、独立行政法人統計センターのホームページ、政府統計の総合窓口『e-Stat』の「地図で見る統計（統計 GIS）」から取得した⁶。

モデル	LM	p-value
SAR	45.231	0.000
SEM	18.350	0.000

表 1: LM 検定による空間従属性の有無

⁶ <http://www.e-stat.go.jp/SG1/estat/toukeiChiri.do?method=init>

4 各モデルの推定結果と比較

4.1 グローバル回帰モデルの推定

GWR の適用に先立って、グローバル回帰モデルの推定を行う。グローバル回帰モデルとしては、ベンチマークとしてOLS、空間計量経済モデルとして標準的に用いられるSAR及びSEMを推定する。はじめにOLS推定を行い、その後(6)式の $\rho=0$ 、及び $\lambda=0$ を検定することで空間従属性の有無を確認する。もし空間従属性が確認されれば、SARやSEMの妥当性が示されることになる。

空間従属性は、Anselin(1988)によるラグランジュ乗数 (Lagrange Multiplier:LM) 診断関数を用いて検定する。空間重み行列 Ω を用いて、LMは以下のように定義されている。

$$\begin{aligned} LM &= (1/T)((\mathbf{e}'\Omega\mathbf{e})/(\mathbf{e}'\mathbf{e}/n))^2 \sim \chi^2(1) \\ T &= \text{tr}(\Omega^2 + \Omega'\Omega) \end{aligned} \quad (9)$$

ここで、 \mathbf{e} は最小2乗法による残差である。LM検定の結果を示した表1を見ると、 ρ 及び λ はともに0ではないことがわかる。従って空間従属性の存在が確認されたので、SARやSEMによる推定を実行する。

OLS、SARおよびSEMの最尤法による推定結果は、表2にまとめた。表2を見ると、各説明変数の符号条件は予想通りの結果となっている。ただし各モデルでは推定量に大きな差が生じている。例えば、定数項、最寄駅から都心までの時間距離、最寄駅からの時間距離など、ヘドニックアプローチの性質から考えると重要な説明変数項に相当のずれが生じているのが確認できる。他の説明変数項に関しては、一部を除き概ね同様な推定量が得られている。AICに基づいてモデルの当てはまりについてみると、SARが最も当てはまりがよく、SEM、OLSの順となっている。空間従属性を考慮したモデルを選択することで当てはまりが改善されたともいえる。しかし、SARモデルの説明変数の中でも、最寄駅から都心までの時間距離や最寄駅からの時間距離については、(疑似) t 値が低く有意ではない。SARは他のモデルと比較してAICが小さく適合度が最も高いとはいえども、必ずしもよいモデルとはいえない面もある。

小地域別地価水準のローカル回帰モデル推定～埼玉県さいたま市を例として

モデル	OLS	SAR	SEM
定数項	-283184 (-1.39)	-417720 (-2.36)	-81160 (-0.37)
最寄駅から都心までの時間距離	-25 (-1.95)	-6 (-0.52)	-44 (-2.70)
最寄駅からの時間距離	-3027 (-2.15)	-170 (-0.13)	-4568 (-2.00)
地積 (対数値)	52001 (2.46)	39593 (2.16)	47750 (2.50)
住宅地域	188809 (2.23)	185150 (2.53)	162125 (2.04)
商業地域	259271 (2.52)	237150 (2.66)	148777 (1.55)
工業地域	107958 (1.20)	100940 (1.30)	77755 (0.93)
住居	-69981 (-1.50)	-88514 (-2.20)	-83374 (-2.02)
店舗	-155351 (-3.58)	-124190 (-3.31)	-114367 (-2.91)
事務所	-93648 (-1.98)	-98245 (-2.41)	-75354 (-1.74)
防火地域	166590 (3.79)	57164 (1.44)	101290 (2.25)
容積率	1544 (8.39)	1298 (7.92)	1430 (8.10)
ρ		0.50 (8.57)	
λ			0.56 (7.74)
決定係数	0.66		
自由度修正済み決定係数	0.65		
AIC	6735.2	6688.1	6711.7

表 2: OLS, SAR, SEM の推定結果 (カッコ内は t 値)

4.2 GWR モデルの推定

次に GWR の推定を行う⁷。GWR の推定結果は各標本地点ごとに異なるパラメータ推定量を出力するため、その数は膨大なものになる。そこで表3では、最小値、四分位点の値、最大値

⁷ なお推定にあたってのカーネルのバンド幅の設定方法であるが、標本地点の範囲内で粗密に大きな差がないことから固定型カーネルを選択して推定を行った。

および比較として先に推定されている OLS によるグローバルパラメータ推定量をまとめた。

表3を見ると、各標本地点によってパラメータ推定量が相当に変化することが分かる。特に注目すべきは、定数項と最寄駅からの時間距離のレンジが非常に大きい点である。一般にヘドニック回帰モデルにおける定数項の推定量は、その予測値の水準に大きな影響を持つものであるため、定数項が大きく変化するという事は、各標本地点の土地価格の水準が大きく異なる可能性を示唆していると思われる。また、最寄駅から当該標本地点までの時間距離が大きく変化するという事は、同じ時間距離であったとしても、地域によっては価格に対して大きなマイナスの影響を与えることを意味しているの、土地の価格付けをする場合に対象土地がどこに位置するか十分に注意を払わなければならないことを示唆している。

	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
定数項	-737100	-603300	-415700	-154100	289100	-283184
最寄駅から都心までの時間距離	-55	-40	-30	-24	-15	-25
最寄駅からの時間距離	-10980	-5010	-2964	-660	703	-3027
地積 (対数値)	12460	57670	70800	80930	85680	52001
住宅地域	67080	176100	225100	261000	286400	1888098
商業地域	139200	195500	225600	244000	273000	259271
工業地域	-6756	90830	152900	202200	226300	107958
住居	-234900	-91370	-79020	-63140	-15180	-69981
店舗	-193600	-152500	-132400	-100000	-51560	-155351
事務所	-213300	-98680	-88870	-78470	-51650	-93648
防火地域	103000	133300	165300	185000	300400	166590
容積率	261	1438	1808	1982	2094	1544
GCV によるバンド幅	0.0381					
グローバル決定係数	0.73					
AIC	6682.9					

表 3: GWR の推定結果

	(有効)自由度	平方和	平均平方和	F-value	p-value
OLS Residuals	12.00	5.97e+12			
GWR Improvement	25.30	1.18e+12	4.67e+10		
GWR Residuals	213.70	4.79e+12	2.24e+10	2.09	0.000

表 4: GWR による改良の有無の検定 (分散分析)

それでは、GWR が OLS と比較して有意に改良されたのかを見るために、Brunsdon, et al.(1999)に従って分散分析を行う。これは OLS 残差を、GWR で説明された部分 (GWR improvement) と説明されなかった部分 (GWR residuals) に分解して両者の比率について F

検定を行うものである。検定結果は表 4 に示した⁸。

表 4 を見ると OLS と比較して、GWR による推定精度の向上が認められる結果となった。また、表 2 および表 3 のグローバル決定係数を比較すると⁹、OLS から GWR にすることで説明力の向上がみられる。さらに AIC については、SAR よりも当てはまりの向上がみられる¹⁰。したがって、GWR の適用によって推定精度の向上が実現したと考えてよいと思われる。

	<i>F</i> -value	Pr(>)
定数項	2.15	0.000
最寄駅から都心までの時間距離	0.95	0.589
最寄駅からの時間距離	3.34	0.000
地積 (対数値)	0.85	0.787
住宅地域	0.54	0.773
商業地域	0.12	0.999
工業地域	0.65	0.711
住居	0.67	0.818
店舗	0.75	0.711
事務所	0.41	0.980
防火地域	1.81	0.014
容積率	8.39	0.000

表 5: 各パラメータに関する空間的非定常性の検定

⁸ 表 4 の自由度に関して、OLS 残差については通常自由度 (標本数 - 推定パラメータ数) が記載されているが、GWR については有効自由度 (The Effective Degrees of Freedom) が記載されている。有効自由度は

$$n - 2\text{tr}(H) + \text{tr}(H'H) \quad (10)$$

であり、ここで、 $\text{tr}(\cdot)$ は行列のトレースを表し、 H はノンパラメトリック回帰におけるハット行列である。有効自由度は、必ずしも整数値ではないのが一般的である。

⁹ グローバル決定係数は、回帰を行う地点を i および標本点を j とすると、以下の式で求められる。

$$\text{global}R^2 = 1 - \frac{\sum_i \sum_j w_{ij} (y_j - \bar{y})^2}{\sum_i \sum_j w_{ij} (y_j - \hat{y}_j)^2} \quad (11)$$

¹⁰ 一般的に AIC は、各標本の誤差項が独立かつ正規分布に従い、さらに標準誤差が等しい場合、パラメータ数を k とすると、

$$AIC = 2n \ln \sigma + n \ln 2\pi + 2k \quad (12)$$

によって求められるが、GWR における AIC は、ハット行列を H とすると、

$$AIC_{GWR} = 2n \ln \sigma + n \ln 2\pi + n + \text{tr}(H) \quad (13)$$

で求められる。したがって正確に言えば、両者の直接の比較はできないが、近似的にモデルの比較選択に用いられる。詳細については Fotheringham, et al.(2002)を参照のこと。

4.3 各ローカル推定量に関する分析

ここでは、GWR で得られたローカルパラメータ推定量に注目して、その有意性について検討する。

はじめに、各推定パラメータが空間的に変動するか、いいかえれば空間的非定常性が認められるかどうかを、Leung, et al.(2000)の方法で検定する。その結果は表5にまとめた。

表5を見ると、各係数パラメータ推定量のうち、定数項、最寄駅までの時間距離、防火地域、容積率について、標本地点によって空間的非定常性が認められるとの結果が出た。定数項と最寄駅までの時間距離については、グローバル回帰モデルの推定結果の中でもその有意性に疑問が残る結果となっていたが、この検定結果から、グローバルな唯一の推定量を求める困難性が確認される。また、その一方で最寄駅から都心への時間距離については空間的変動が認められなかった。これは、都心通勤者が市内にまんべんなく多く居住するさいたま市の全体的な特徴であり、当該地域の都心との近接関係によるものだと推察される。ただし、最寄駅からの時間については、最寄駅でのバス便への乗り換えの不便さが地域ごとにかなり異なることなどを反映しているものと考えられる。

次に、各標本地点における最寄駅から都心までの時間距離と最寄駅からの時間距離についての(疑似) t 値を検討する。これらは、グローバル回帰モデルの中で最もモデル適合度が高かった SAR において t 値が非常に小さく有意な結果が得られなかったものである。GWR による推定の結果図3のような t 値の分布となった。平均値が両方とも 5% 有意水準を超えてはいない(平均はそれぞれ -1.892、-1.92) が、SAR の結果に比べれば格段に改善している。

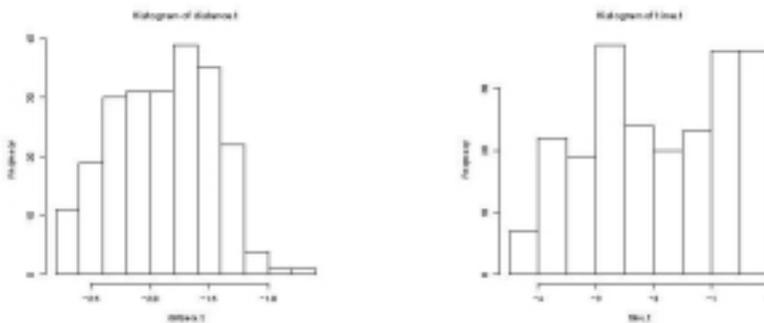


図3: 最寄駅から都心までの時間距離(左)および最寄駅からの距離(右)に係る t 値の分布

最後に各標本地点におけるモデル適合の程度を表すローカル決定係数についてみる¹¹。最小値が 0.5502、最大値が 0.7448 である。図 4 を見てわかるようにほとんどの標本地点で OLS の決定係数を上回っていることが分かる。したがって、各地点の予測値が実績値に近付いていることが期待できる。

5 小地域におけるパラメータ推定量の分布

本節では、これまで使用してきたデータに基づいて、対象地域内の小地域におけるパラメータ推定量の水準がどのように分布しているのかをビジュアル化する。これまで、地域別の価格形成要因の把握は必ずしも実証的なものではなく、ある意味不動産に携わる人間の「勘」に支えられていた。本論文での GWR 推定に基づく実証的な結果を通じて、小地域内の土地の価格形成要因についてさらなる理解が期待できる。

ところで、GWR が他の推定モデルと比較してその推定精度を改善することは既にみたが、その他の利用可能性として興味深いのは、標本地点に限らず、対象地域のあらゆる地点でのパラメータ推定が可能な点である。もちろんその場合には、その点に関するデータが存在しないので、推定誤差等に係る分析や決定係数の計算などはできない。しかし、近似として十分な情報を提供することは可能である。

本論文では、対象地域であるさいたま市を丁目区分にまで分割し、その小地域における価格形成要因の影響の程度、つまり回帰パラメータの水準の高低がどのように分布しているのかを図示することで、不動産の専門家だけでなく広く一般に理解可能な情報を提供できるかを検討する。丁目区分ごとの推定地点については、各地域の重心を求め、その緯度・経度を位置情報として採用することにする。

本論文では、特に定数項、最寄駅から都心への時間距離、最寄駅からの時間距離および地積について取り上げて各パラメータの水準についての考察を行うことにする。結果を図 5 から図 8 にまとめた。

¹¹ ローカル決定係数は、回帰を行う地点を i および標本点を j とすると、以下の式で求められる。

$$\text{local}R^2 = 1 - \frac{\sum_j w_{ij} (y_j - \bar{y})^2}{\sum_j w_{ij} (y_j - \hat{y}_j)^2} \quad (14)$$

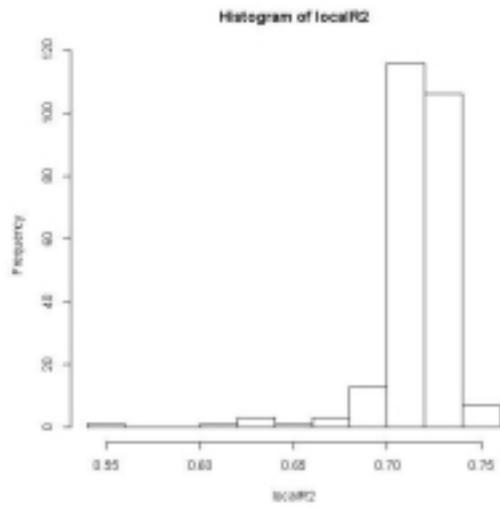


図 4: ローカル決定係数の分布

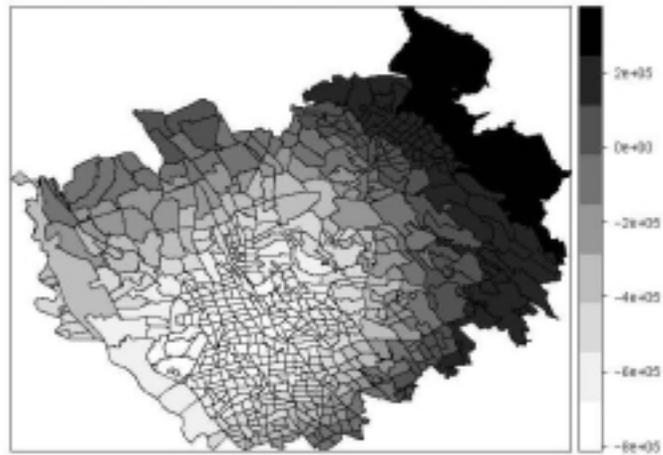


図 5: 定数項

小地域別地価水準のローカル回帰モデル推定～埼玉県さいたま市を例として

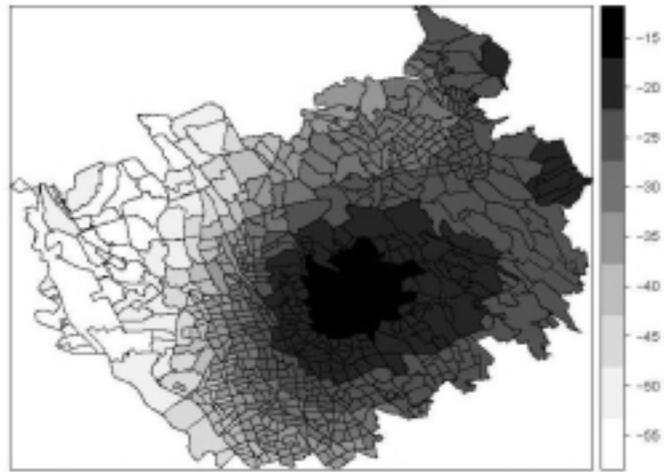


図 6: 最寄駅から都心への時間距離

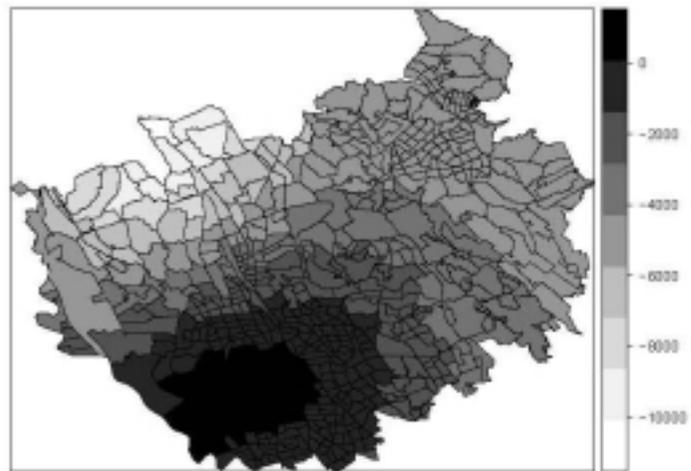


図 7: 最寄駅からの時間距離

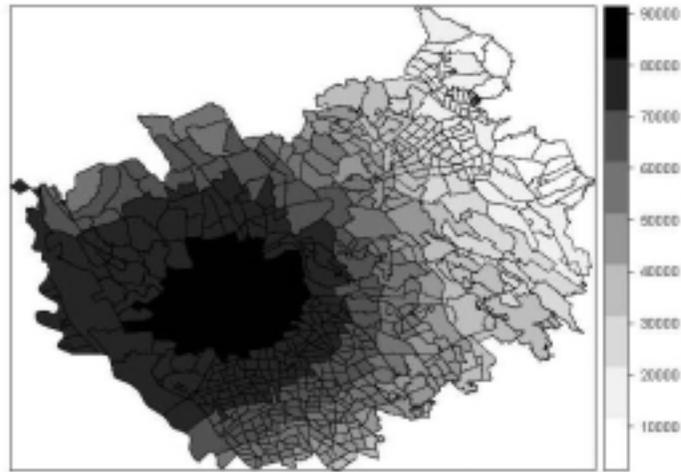


図8: 地積 (対数値)

まず図5の定数項についてであるが、市の南西部（浦和区等が位置する）と特に北東部（岩槻区の方）を比べるとその差が非常に大きくなっていることが分かる。したがって、両地区にある同じ価格の土地については、価格形成要因たる各説明変数の影響度を示す推定パラメータもその差が非常に大きくなっていることが予想される。

次に最寄駅から都心への時間距離については、検定段階では大きな空間的変動はないという結果であった。しかし細かく図示してみると様相がずいぶん変わって見える。市の中央付近部はいわゆる「見沼田んぼ」が広がっている芝川周辺地域であり、通勤に関する利便性は低い。このような地域の土地は、他地域に比較してそもそも通勤に向いていないので、都心への時間距離に敏感に反応しないと考えられる。逆に、このような属性に敏感に反応する（1分通勤時間が長くなった場合、価格の下落が大きい）のであれば、その土地は幹線鉄道等の近くに所在するはずである。図6で見ると、確かに市の北西部は大宮駅を代表とするさいたま市の交通の集積地であり、分析結果と符合する。

最寄駅から当該土地への時間距離についてであるが、時間距離が大きくなってもそれほど価格が下がらない地域は図7によると市の南西部である。この地域は浦和区などが所在しており、幹線鉄道としては、東北本線、高崎線、京浜東北線が南北に縦断しており、またこれらと並行して埼京線が走っている。さらに武蔵野線もあり、近い駅がいくつもあるという地域である。したがってそのような地域では、最寄駅までの時間距離はそれほど重要な価格形成要因ではないと考えられる。一方、交通手段やターミナルとなる駅が非常に限定的であり、またそういう交通事情であるからこそ乗り継ぎのバス便が少ないなど不利な交通環境の地域である場合、最寄駅から当該土地までの時間距離が大きいと予想される。このような地域は市の北部から東部にかけて存在しており、図7の様子とやはり符合する。

最後に図 8 の地積に関してであるが、まさに市の中心市街地に位置する地域でパラメータ推定値が大きくなっている。この地域は業務集積地であり、また都心通勤者をターゲットとしたマンション開発が盛んな地域である。特に最近の駅近マンションのブームもあり、一団の大きな土地は価値が高いものとみられる。それに対して市の東部（岩槻区や見沼区）は、そのような土地需要もなく密集地区でもないため、地積の大きさには敏感に反応はしないと思われる。

以上の考察は地域全体を対象としていたが、細分されたすべての丁目区分に対して、GWR によって適合度の高いパラメータ推定がなされているので、実証的に土地の価格形成要因を把握することができる。またそれをビジュアルを通じて認識することで、小地域をさらに細分した地域の特性も容易に確認することができる。

6 まとめと今後の課題

本論文では、さいたま市を例にとりローカル回帰モデルの GWR を用いることで土地価格およびその価格形成要因について相当の地域差が存在することが確認された。他のグローバル回帰モデルと比較して GWR はモデルの適合度が高く、土地価格あるいは不動産価格の推定においては、グローバルなパラメータ推定よりもむしろローカルなパラメータ推定を行う意義は大きいことが分かった。また、さいたま市という限定的な研究対象地域でありながら、細分された丁目区分におけるパラメータの空間的相違の大きさは無視できないものであったため、土地の価格形成要因の地域差を分析するために、GWR は有益な情報をもたらすことが確認された。さらに興味深いのは、GWR を利用することによって、対象地域に属するあらゆる地点について、その地点独自のパラメータで予測価格を求めることができることである。本論文での結果をさらに発展させてより精度の高い土地・不動産価格の推定方法を模索することは、地域不動産情報の充実の実現を通じて、情報の非対称性の大きい不動産市場の効率化に資するものと考えられる。

問題点としては、ヘドニック回帰モデルの説明変数の選定や推定手順の洗練化に加えて、時系列的な価格動向を考慮したモデルの設計が挙げられる。不動産データは入手が非常に困難であり、地点的にも時点的にもデータの密度が疎である。精度の高い推定を行うためには、今回の研究のようなクロスセクション・データのみならず、過去時点のデータも活用する必要がある。これらは今後の研究課題としたい。

参考文献

- [1] Anselin, L.(1988) “Spatial Econometrics: Methods and Models”, Dordrecht: Kluwer.
- [2] Brunson, C., Fotheringham, A.S. and Charlton, M.E.(1996), “Geographically weighted regression : A method for exploring spatial nonstationarity”, *Geographical Analysis* 28, pp.281-298.
- [3] Brunson, C., Fotheringham, A.S. and Charlton, M.(1999), “Some notes on parametric significance tests for geographically weighted regression”, *Journal of Regional Science* 39, pp.497-524.
- [4] Clapp, J.M., Kim H.J. and Gelfand A.E.(2002), “Spatial Prediction of House Price Using LPR and Bayesian Smoothing”, *Real Estate Economics*, 30:4, pp.505-532.
- [5] Fotheringham, A.S., Brunson, C. and Charlton, M.(2002), “Geographically Weighted Regression: The Analysis of Spatially Varying Relationships”, Chichester, UK: Wiley.
- [6] Leung, Y., Mei, C.-L. and Zhang, W.-X.(2000), “Statistical tests for spatial nonstationarity based on the geographically weighted regression model”, *Environment and Planning A* 32, pp.9-32.
- [7] 清水千尋・唐渡広志, 『不動産市場の計量経済分析』, 朝倉書店.
- [8] 谷村晋 (2010), 『地理空間データ分析』, Rで学ぶデータサイエンス7 (金明哲編), 共立出版.