

都市再生効果分析のためのヘドニック分析における 距離変数の効果最適化と標本選択最適化（概要）

研究責任者 東京大学大学院工学系研究科都市工学専攻

教授 浅見泰司

共同研究者 空間情報科学研究センター

客員研究員 蛭田有希

1. はじめに

1-1 背景

都市再生の事業や規制を検討する上で、その政策効果を的確に求めることが重要である。そのための計量手法としてしばしばヘドニック分析が用いられる。実証研究では、中心地、最寄り駅、主要施設等から価格観測地点までの距離が、属性（説明変数）として用いられることが多く（e. g. Berry, 1976, Noonan, Krupka, & Baden, 2007）、近年の GIS の普及により、距離変数の活用はさらに一般的なものとなっている。（以降、中心地、最寄り駅、主要施設等のランドマークを「施設等」とよび、各標本点から施設等までの距離や時間を示す変数を「距離変数」と呼ぶこととする。）しかし、このような距離変数を複数用いると、回帰係数が歪むことが知られている（Heikkila, 1988）。

複数の距離変数を用いることがヘドニック分析結果に与える影響として下記の2点を指摘することができる。これらは何れも距離変数間の共線関係に起因する。一つ目の問題は、近くに似た限界効用を持つ施設等が集積している場合、標本点からそれらの施設等までの距離を説明変数としたヘドニックモデルを構築しても、実際にどの施設等が物件価格に対して影響を及ぼしてい

るかを特定することが困難だという問題である。実在する市街地では、土地利用制限が行われているため、物件価格に同じような影響を与える施設同士が近くに集積しているケースが多い。そのため、このような問題は生じやすい（Deaton & Hoehn, 2004）。二つ目の問題は、例えば、迷惑施設から離れるほど好ましい施設に近くなるというケースにおいては、迷惑施設から離れることと、好ましい施設へ近づくことのどちらの影響が生じているかの判断が難しいという問題である。このようなケースにおいて、両方の施設等までの距離を説明変数として用いたヘドニックモデルを構築しても、実際にどちらが物件価格に対して影響を及ぼしているかを特定することは困難である。

これらの問題は、複数の距離変数を用いるヘドニックモデルでは正確な回帰係数を推定すること難しいという統計的な問題である。仮に、統計的に正確な回帰係数を推定することが出来るのであれば、施設までの距離を距離変数として用いてモデルを構築し、その結果に基づいて、類似施設の中からより影響の大きい施設を特定することも可能であり、迷惑施設と好ましい施設のうち、影響の大きい施設を特定することもできる。しかし、実際には、回帰係数の正確な推定そ

のものが困難となる。

統計的に正確な回帰係数の推定を行えない原因は、距離変数同士の相関にある。一つ目の問題では距離変数同士が正の相関を持ち、二つ目の問題では距離変数同士が負の相関を持つ。距離変数が相関する場合、多重共線性の影響が生じる。本来、多重共線性は、2変数以上の線形関係や、直交性の欠如等と定義される (e.g. Chatterjee & Hadi, 2006) が、現実のデータではそのような完全な多重共線性が生じることは少なく (浅見, 2015)、ここで議論されるのは複数の説明変数が互いに関連し、線形従属に近い関係をもつことによる不完全な多重共線性 (Asteriou & Hall, 2007) の影響である。多重共線性の影響としては、推定される回帰係数が安定せず、サンプルによって大きくばらつくため、信頼できないということ、また、推定される回帰係数の分散は大きくなるものの、被説明変数の推定結果、テストデータによる予測結果は、その影響をそれほど受けないこと、が知られている (e.g. Alin, 2010, Kutner, Nachtsheim, Neter, & Li, 2005 p. 283)。

用いる距離変数の数が空間の次元数よりも多い重回帰モデルでは理論的に変数同士が相関しやすく、Heikkila (1988) は、このような多重共線性を、特に、空間的多重共線性 ("spatial multicollinearity") と呼びその問題を指摘した。また、空間的多重共線性の影響が、ヘドニック分析において不適切なバイアスを生み出していることは、複数の実証研究によって指摘されている (e.g. Deaton & Hoehn, 2004 ; Ross, Farmer, & Lipscomb, 2011)。空間的多重共線性への対

応としては、標本点を新たに取得できる場合には、相関を小さくする標本点の配置をあらかじめ数理的に導いて用いるための方法が提案されている (貞広, 2015)。しかし、多くの研究では、既存の限られた観測地点のデータが用いられており、既存の観測地点を標本とした場合の対応も求められる。既存の観測点における空間的多重共線性への対応としては、主成分分析や因子分析により変数を集約する方法 (e.g. Chatterjee & Hadi, 2006) や、VIF 等により多重共線性を発見し相関する変数群から一つの変数を選択する方法 (e.g. 吉田, 1987)、サンプルサイズを増やす方法 (e.g. Alin, 2010)、が用いられる。主成分分析による方法は、「重みづけされた距離変数によって表現される被説明変数は、実際には様々な周辺環境の複合的な効果によって形成される物件価格である」との観点からは妥当な手法であり、変数を一つに絞り込む方法は、そのような複合的な効果を一つの変数の回帰係数として読み取る手法だといえる。しかし、何れの方法も、個々の距離変数に対する適切な回帰係数が得られないことから、個々の施設等がもたらす限界効用の多寡を示すことには適さない。また、サンプルサイズを増やす方法は、限られたデータしか扱うことのできない実証研究においては難しい。

ヘドニックモデルを用いた実証研究では、環境要因の経済価値の評価が目的とされることが多く、複数の距離変数を用いることにより生じる空間的多重共線性の影響は、分析の結論を覆しかねない重大な問題といえる。そのため、複数の距離変数を用いたヘドニック分析で生じる多重共線性の影響を

低減するための手法の開発が求められる。

1-2 目的

本研究では、ヘドニック分析において距離関連変数を説明変数に入れる際に生じる多重共線性の影響を、軽減するための対処方法を提示することを目的とする。具体的には下記の3点を明らかにすることが目的である。

- 1) 2つの距離変数を用いた際に生じる空間的多重共線性の影響、および、影響を計量するために適切な指標。
- 2) 多重共線性によって、推定される回帰係数がばらつくという問題に対処可能な事後処理方法。
- 3) 開発した事後処理方法を実際の事業評価に適用した際に得られる手法の効果。

2. 分析

2-1 共線性による影響

2-1-1 影響

上述の多重共線性の影響は数式によって説明することが可能である。今、被説明変数 y を n 個の観測値からなるベクトル、行列 X を、 n 個の観測値、 p 種類の説明変数からなる行列とし、重回帰式 $y = Xa + \mu$ ($\mu \sim N(0, \sigma^2 I)$) を考え、回帰係数ベクトルを $a_k = (a_1, \dots, a_k, \dots, a_p)'$ と表すとき、偏回帰係数 \hat{a}_k の分散は式(2-1)のとおりであるが、 $p = 2$ について考えると推定されるパラメータ \hat{a}_1 の分散 $V(\hat{a}_1)$ は式(2-2)のように書ける(吉田, 1987)。

$$V(\hat{a}_k) = \sigma^2 (X'X)^{-1} \quad (2-1)$$

$$V(\hat{a}_1) = \sigma^2 / n(1 - r_{12}^2) \quad (2-2)$$

ここで、 r_{12} は1番目の説明変と2番目の説明変数との相関係数である。 r_{12}^2 が1に近づくほど式(2-2)の分母は小さくなるため $V(\hat{a}_1)$ は大きな値をとることとなり、 \hat{a}_1 は安定した値とならないことが分かる。

これに対して、パラメータの和の分散 $V(\hat{a}_1 + \hat{a}_2)$ は式(2-3)のように書ける。

$$V(\hat{a}_1 + \hat{a}_2) = \frac{\sigma^2}{1 - r_{12}^2} (2 - 2r_{12}) = 2 \frac{\sigma^2/n}{1 + r_{12}} \quad (2-3)$$

式(2-2)とは異なり、(2-3)では r_{12} が1に近づいても分母は小さくならず、 $V(\hat{a}_1 + \hat{a}_2)$ も r_{12} が大きくなるほど大きな値をとることはない。このことから、多重共線性により生じた回帰係数の歪の被説明変数の推定結果やテストデータによる予測結果への影響は大きくないことが示唆される。

したがって、多重共線性による影響の中でも、「サンプリングごとに推定される回帰係数が安定しない問題(回帰係数の歪)」に対処する手法が重要となる。

2-1-2 影響を計量する指標の特定

回帰係数が真値から乖離する程度の大小は、個々の回帰係数でそれぞれ異なる。そこで、回帰係数の歪をどのような指標で表すべきかについて知る必要がある。ここでは、回帰係数の歪を表す指標の候補として下記の5指標 $C^{*a} \sim C^{*e}$ を挙げ、この中から適切な指標を特定する。

$$C^{*a} = \sqrt{(\hat{\beta}_0^* - 0)^2} \quad (2-4)$$

$$C^{*b} = \sqrt{(\hat{\beta}_1^* - 1)^2} \quad (2-5)$$

$$C^{*c} = \sqrt{(\hat{\beta}_2^* - 1)^2} \quad (2-6)$$

$$C^{*d} = \sqrt{(\hat{\beta}_1^* - 1)^2 + (\hat{\beta}_2^* - 1)^2} \quad (2-7)$$

$$C^{*e}_i = \sqrt{(\hat{\beta}_0^* - 0)^2 + (\hat{\beta}_1^* - 1)^2 + (\hat{\beta}_2^* - 1)^2} \quad (2-8)$$

まず、回帰式 $y^{*e} = \beta_0 + \beta_1 d_a + \beta_2 d_b$ を考え、Wang, Asami, & Sadahiro, 2015 に準ずる方法を用いて、微小領域において推定した回帰係数 ($\hat{\beta}_0^*$, $\hat{\beta}_1^*$, $\hat{\beta}_2^*$) に基づき 1)各指標 ($C^{*a} \sim C^{*e}$) を算出し、同じ微小領域で 2)距離変数間の相関係数 r 、3)推定結果の歪 B_i 、を算出する。そして、1)各指標 ($C^{*a} \sim C^{*e}$) と 2)距離変数間の相関係数 r 、1)各指標 ($C^{*a} \sim C^{*e}$) と 2)推定結果の歪 B_i の相関関係を把握する。これは、5指標のうち、距離変数間の相関 r との関連が最も強い指標と、推定結果の歪 B_i との関連が最も強い指標が一致した場合、その指標は、距離変数間の相関により生じた回帰係数の歪を最も的確に表現している指標だと考えられ、同時に、回帰係数の歪による被説明変数の推定結果への影響についても的確に表現している指標だと考えられるためである。

表- 1 距離変数間の相関係数 r と各指標との相関

	C^{*a}	C^{*b}	C^{*c}	C^{*d}	C^{*e}
MIC	0.193	0.186	0.186	0.202	0.199
ρ^2	0.003	0.003	0.003	0.003	0.004
$MIC - \rho^2$	0.190	0.183	0.183	0.199	0.196

表- 2 推定結果の歪 B_i と各指標との相関

	C^{*a}	C^{*b}	C^{*c}	C^{*d}	C^{*e}
MIC	0.192	0.195	0.195	0.294	0.269
ρ^2	0.026	0.029	0.029	0.034	0.031
$MIC - \rho^2$	0.166	0.166	0.166	0.260	0.238

MIC: Maximum Information Coefficient,
 ρ : Pearson product-moment correlation coefficient

表- 1 は、各指標 ($C^{*a} \sim C^{*e}$) と距離変数間の相関係数 r との相関関係を、表- 2 は、各指標 ($C^{*a} \sim C^{*e}$) と推定結果の歪 B_i との相関関係を示したものである。

表中の MIC (Maximum Information Coefficient) は、たとえ非線形な関係であっても変数間の関連性の強弱を確認できる相関係数である。MIC の値は 0 から 1 の範囲をとり、決定係数と概ね同様にとり扱うことができる。また、MIC とピアソンの積率相関係数(ρ)の二乗との差 ($MIC - \rho^2$) は非線形性の指標とされ、2 変量間が線形従属の関係ある場合 $MIC - \rho^2$ は 0 となる (Reshef et al., 2011)。

表- 1 より、距離変数間の相関係数 r と最も強い相関を示す指標は C^{*d} であり、その MIC は 0.202 であることが分かる。これはピアソンの積率相関係数 0.449 ($n=6561$) に相当し、明らかに有意な非線形の相関があるといえる。表- 2 より、推定結果の歪 B_i と最も強い相関を示す指標も C^{*d} であり、その MIC は 0.294 であることが分かる。これはピアソンの積率相関係数の 0.542、0.543 ($n=6561$) に相当し、明らかに有意な非線形の相関がある。したがって、5 指標のうち距離変数間の相関係数 r との関連が最も強い指標と、推定結果の歪 B_i との関連が最も強い指標は一致し、何れも C^{*d} である。

なお、距離変数間の相関 r と推定結果の歪 B_i との関係をみると、MIC が 0.0734 であり、その関連性は回帰係数の歪を介したときの関連性と比較して弱い (図- 1)。



図- 1 距離変数間の相関・回帰係数の歪・推定結果の歪の関係

以上より、 C^{*d} は、1)距離変数間の相関により生じた回帰係数の歪を反映すると同

時に、回帰係数の歪による2)被説明変数の推定結果への影響を表しているものと考えられ、多重共線性により生じた歪を最も的確に表現している指標だと考えられる。そのため、回帰係数の歪のサンプルごとのばらつきを最小化する手法の検討にあたっては、 C^d の最小化を目指す。

2-2 事後処理方法

距離変数間の相関による空間的多重共線性の影響を低減するための事後処理方法を検討する。本研究では、サンプルサイズの減少による精度の低下を避けるため、観測値自体に重みづけをする方法を示す。具体的には、回帰モデルに用いる観測値に距離変数間の相関に応じた重みづけをし、距離変数間の相関による空間的多重共線性の影響を低減する方法を示す。

2-2-1 提案手法

1) 対象領域と設定

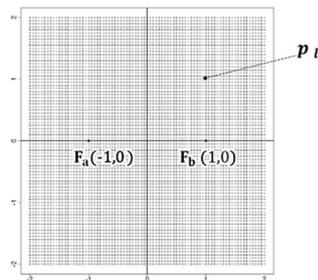


図-2 対象領域および条件設定
($\delta = 1, g = 0.05, n = 6,559$)

二次元平面上の領域 $T = \{(x,y) | -2 \leq x \leq 2, -2 \leq y \leq 2\}$ を対象とする。距離変数算出の起点となる施設等の位置を $F_a(-\delta, 0)$ 、 $F_b(\delta, 0)$ の2地点とし、対象領域内の間隔 g の格子点から、距離変数の起点となる施設等の位置2地点を除外した n 個の格子点 $p_l \quad l \in [1, n]$ を観測地点とする。起点を F_a 終点を p_l とする直線の距離を d_{a_l} と

し、起点を F_b 終点を p_l とする直線の距離を d_{b_l} とする。

$\delta = 1, g = 0.05, n = 6,559$ とした場合を例に、条件設定の概念を図-2に示す。

2) モデル

距離 $d_{a_l}, d_{b_l} \quad l \in [1, n]$ を説明変数、 $y_{e_l} \quad l \in [1, n]$ を被説明変数とした回帰モデル(2-9)を考える。

回帰モデルの被説明変数となる y^{*e}_l は、期待値 η^{*e}_l (2-11) に誤差項 ε^{*e}_l を加えて発生させた実現値であり、(2-10) で表される。

η^{*e}_l は、(2-11) で表される期待値である。誤差項 ε^{*e}_l は、平均0分散 v の正規分布する値から n 個の値をランダムサンプリングして発生させた誤差である。 Y^{*e}_l は、 y^{*e}_l を被説明変数、 d_{a_l}, d_{b_l} を説明変数とした回帰モデル(2-9)によって推定した切片 $\widehat{\beta}_0^{*e}$ 、回帰係数 $\widehat{\beta}_1^{*e}, \widehat{\beta}_2^{*e}$ により得られた被説明変数の推定値であり、(2-12) で表される。

C_e は、回帰係数 $\widehat{\beta}_1^{*e}, \widehat{\beta}_2^{*e}$ により求めた回帰係数の歪を表す指標であり、(2-13) で表される。

平均0分散 σ^2 の正規分布する値から n 個の値をランダムサンプリングして発生させる操作を t 回繰り返し、 t 種類の異なる誤差項 $\varepsilon^{*e}_l \quad e \in [1, t]$ による、 t 種類の異なる実現値 $y^{*e}_l \quad e \in [1, t]$ を被説明変数として、 t 種類の回帰モデルを構築し、 t 種類の $C_e \quad e \in [1, t]$ を得る。

$$\text{回帰式} \\ y^{*e} = \beta_0 + \beta_1 d_a + \beta_2 d_b \quad (2-9)$$

$$\text{実現値} \\ y^{*e}_l = 0 + 1d_{a_l} + 1d_{b_l} + \varepsilon^{*e}_l \quad \varepsilon^{*e}_l \sim N(0, \sigma^2) \quad (2-10)$$

$$\text{期待値} \\ \eta^{*e}_l = E[y^{*e}_l] = 0 + 1d_{a_l} + 1d_{b_l} \quad (2-11)$$

推定値

$$Y^{*e}_l = \widehat{\beta}_0^e + \widehat{\beta}_1^e d_{a_l} + \widehat{\beta}_2^e d_{b_l} \quad (2-12)$$

回帰係数の歪

$$C_e = \sqrt{(\widehat{\beta}_1^e - 1)^2 + (\widehat{\beta}_2^e - 1)^2} \quad (2-13)$$

3) 目的関数

距離変数間の相関による空間的多重共線性は、推定される回帰係数の分散を大きくする。つまり、距離変数間の相関の強い全てのサンプルで回帰係数の歪 C_e が大きくなることはなく、同程度に相関の強いデータセットを使って構築した回帰モデルであっても、回帰係数の歪のない場合も、歪の大きい場合もあり、 C_e はばらつくことになる。そのため、あるモデルによる C_e を最小化するのではなく、複数の実現値 $y^{*e}_l \quad e \in [1, t]$ により推定された回帰係数 $\widehat{\beta}_1^{*e}$ 、 $\widehat{\beta}_2^{*e} \quad e \in [1, t]$ による複数の回帰係数の歪 $C_e \quad e \in [1, t]$ の総和を最小化できるように目的関数を設定し、その値を指標値 f で表す(2-14)。

目的関数

$$f = \sum_{e=1}^t \sqrt{(\widehat{\beta}_1^{*e} - 1)^2 + (\widehat{\beta}_2^{*e} - 1)^2} \quad (2-14)$$

4) 重み関数

回帰式の観測値に重みづけすることにより、 f を最小化したい。空間的多重共線性は距離変数同士の相関が原因となり生じているため、重みベクトル算出のための関数(重み関数とよぶ)は、距離変数間の相関係数 r_l の関数とし、重み関数 $W(r_l)$ を求める。

(1) 距離変数間の相関係数

観測地点ごとの距離変数間の相関係数 r_l を求める。施設等の位置 $F_a(-\delta, 0)$ 、 $F_b(\delta, 0)$ と各標本点 p_l を結ぶベクトルをそれぞれ a_l 、 b_l とすると、2つのベクトル a_l 、 b_l がなす角 $\cos \theta_l$ は、 a_l と b_l の内積を正規化した式(2-15)で表され、 $\cos \theta_l$ は、Wang et al.,

2015 の方法で算出される相関係数 r と殆ど変わらない。これを相関係数 r_l とする。図-3に、 $\delta = 1$ 、 $g = 0.05$ 、 $n = 6,559$ とした場合を例に、相関係数 r_l の算出の概念、および、相関係数 r_l の空間分布を示す。

$$r_l = \cos \theta_l = \frac{a_l \cdot b_l}{|a_l| |b_l|} \quad (2-15)$$

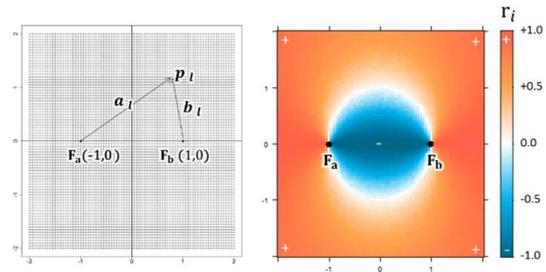


図-3 相関係数 r_l 算出の概念(左)
相関係数 r_l の空間分布(右)

(2) 距離変数間の相関係数の密度

観測地点や施設等の位置によって、距離変数間の相関係数 r_l の頻度は異なり、回帰係数の歪の生じ方は、距離変数間の相関係数 r_l だけでなく、相関係数 r_l の密度にも依存するものと考えられる。そのため、重みベクトルを特定するためには、相関係数 r_l だけでなくその密度も考慮する必要がある。距離変数間の相関係数 r_l の密度は、カーネル密度推定により推定可能である。ここでは、平均0分散1の正規分布によるカーネル関数により密度関数 $D(r_l)$ を求め密度 d を得た。図-4に、 $\delta = 1$ 、 $g = 0.05$ 、 $n = 6,559$ とした場合を例に、得られた密度関数 $D(r_l)$ を示す。

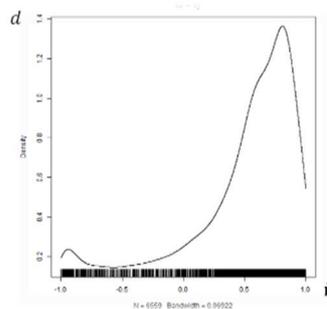


図-4 相関係数 r_l の密度 d

(3) 重み関数

最適な重みの特定にあつたては、可能な限り柔軟な重み関数を用いる必要がある。本研究では、相関係数 \hat{r} と相関係数の密度 d を変数とする適切な多項式を求めることにより、重み関数 $W(\hat{r}, d)$ を特定する。 $W(\hat{r}, d)$ は、1) 相関係数 \hat{r} 、2) 相関係数 \hat{r} の関数である密度ベクトル $d = D(\hat{r})$ 、3) 両者の交互作用項をからなる多項式による加法モデルであり、(2-16)表される。また、簡単のため、

$$(\hat{r}^1, \dots, \hat{r}^g, \dots, \hat{r}^5, d^1, \dots, d^h, \dots, d^5, \hat{r}^1 d^1, \dots, \hat{r}^g d^h, \dots, \hat{r}^5 d^5)$$

$$= (x_1, \dots, x_i, \dots, x_{35}),$$

$$g \in [1,5], h \in [1,5], i \in [1,35]$$

と置き換え、 $W(\hat{r}, d)$ を、 x の関数 $W(x)$ とし、式(2-17) で表す。

$$d = D(\hat{r})$$

$$W(\hat{r}, d) = \exp\left(\sum_{g=1}^5 a_g \hat{r}^g + \sum_{h=1}^5 a_h d^h + \sum_{h=1}^5 \sum_{g=1}^5 a_{gh} \hat{r}^g d^h\right) \quad (2-16)$$

$$W(x) = \exp\left(\sum_{i=1}^{35} \alpha_i x_i\right) = \exp(\mathbf{X}\boldsymbol{\alpha}) \quad (2-17)$$

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,35} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,35} \end{pmatrix}$$

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_i, \dots, \alpha_{35})^t$$

5) 重み関数の母数最適化 (アルゴリズム)

$W(x)$ による重みベクトル \mathbf{w} を、回帰式(2)に適用した際に得られる指標値 f を最小化するよう、重み関数 $W(x)$ の母数ベクトル $\boldsymbol{\alpha}$ を特定する。そのためのアルゴリズムを以下に示す。

Algorithm1

【入力】

- 初期値として $\boldsymbol{\alpha} = (0, \dots, 0)^t$ を与える

- インターバル interval として 0.01 を与える

Repeat

- Algorithm2 により $\boldsymbol{\alpha}^{*1}$, f_1^{\min} および, $\boldsymbol{\alpha}^{*q}$, f_q^{\min} を得る

- 初期値を $\boldsymbol{\alpha}^{*q}$ に置き換える

end if $f_1^{\min} = f_q^{\min}$

【出力】

- $\boldsymbol{\alpha}^{*q}$, f_q^{\min} を $\boldsymbol{\alpha}^{\text{best}}$, f_q^{best} として出力する

Algorithm2

【入力】

- 入力値として初期値となる母数ベクトル $\boldsymbol{\alpha}$ を与える

- Algorithm3 により $\boldsymbol{\alpha}^{*1} = (\alpha_1^{\min}, \alpha_2, \dots, \alpha_q)^t$, f_1^{\min} を得る

- Algorithm3 の初期値を $\boldsymbol{\alpha}^{*1}$ に置き換える

For $s, m = (2, \dots, s+1, \dots, q-1)$

- Algorithm3 により $\boldsymbol{\alpha}^{*m} = (\alpha_1, \dots, \alpha_{m-1}, \alpha_m^{\min}, \alpha_{m+1}, \dots, \alpha_q)^t$, f_m^{\min} を得る

- Algorithm3 の初期値となる母数ベクトル $\boldsymbol{\alpha}$ を $\boldsymbol{\alpha}^{*m}$ に置き換える

end for

- Algorithm3 により $\boldsymbol{\alpha}^{*q} = (\alpha_1, \dots, \alpha_{q-1}, \alpha_q^{\min})^t$, f_q^{\min} を得る

【出力】

- $\boldsymbol{\alpha}^{*1} = (\alpha_1^{\min}, \alpha_2, \dots, \alpha_q)^t$, f_1^{\min} および, $\boldsymbol{\alpha}^{*q} = (\alpha_1, \dots, \alpha_{q-1}, \alpha_q^{\min})^t$, f_q^{\min} を出力する

Algorithm3

【入力】

- 入力値として初期値となる母数ベクトル $\boldsymbol{\alpha}$ を与える。

- 母数ベクトル $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s, \dots, \alpha_{35})^t$ のうち、ある母数 α_s について下記の α_s^{org} , α_s^{pos} , α_s^{neg} 適用し、母数ベクトル $\boldsymbol{\alpha}^{\text{org}} = (\alpha_1, \dots, \alpha_s^{\text{org}}, \dots, \alpha_{35})^t$, $\boldsymbol{\alpha}^{\text{pos}} = (\alpha_1, \dots, \alpha_s^{\text{pos}}, \dots, \alpha_{35})^t$, $\boldsymbol{\alpha}^{\text{neg}} = (\alpha_1, \dots, \alpha_s^{\text{neg}}, \dots, \alpha_{35})^t$ を得る。

- $\alpha_s^{\text{org}} = \alpha_s$

- $\alpha_s^{\text{pos}} = \alpha_s + \text{interval}$

- $\alpha_s^{\text{neg}} = \alpha_s - \text{interval}$

- $\exp(\mathbf{X}\boldsymbol{\alpha})$ に母数ベクトル $\boldsymbol{\alpha}^{\text{org}}$, $\boldsymbol{\alpha}^{\text{pos}}$, $\boldsymbol{\alpha}^{\text{neg}}$ を適用し、重みベクトル $\mathbf{w}^{\text{org}} = \exp(\mathbf{X}\boldsymbol{\alpha}^{\text{org}})$, $\mathbf{w}^{\text{pos}} = \exp(\mathbf{X}\boldsymbol{\alpha}^{\text{pos}})$, $\mathbf{w}^{\text{neg}} = \exp(\mathbf{X}\boldsymbol{\alpha}^{\text{neg}})$ を得る。

- 回帰式(2-9)の被説明変数である実現値 y_e , 説明変数である距離変数 d_a および d_b に、重みベクトル

w^{org} , w^{pos} , w^{neg} で重みづけして、それぞれ、得られた回帰係数 $\hat{\beta}_{0e}$, $\hat{\beta}_{1e}$, $\hat{\beta}_{2e}$ によって回帰係数の歪 $c_e = \sqrt{(\hat{\beta}_{1e}-1)^2 + (\hat{\beta}_{2e}-1)^2}$ を求め、 C_e^{org} , C_e^{pos} , C_e^{neg} を得る。

- $f = \sum_{s=1}^q \sqrt{(\hat{\beta}_{1e}-1)^2 + (\hat{\beta}_{2e}-1)^2}$ により、指標値 f_s^{org} , f_s^{pos} , f_s^{neg} を得る。
- 指標値 f_s^{org} , f_s^{pos} , f_s^{neg} のうち、最小の指標値 f_s^{min} , および、 f_s^{min} を算出した重み関数の母数 α_s^{min} を特定する。

【出力】

- f_s^{min} , および、母数 α_s を α_s^{min} に置き換えた母数ベクトル $\alpha^{*s} = (\alpha_1, \dots, \alpha_{s-1}, \alpha_s^{min}, \alpha_{s+1}, \dots, \alpha_q)^t$ を出力する。

Algorithm3 では、個々の母数 α_s を増加、維持、減少させた3つのパターンのうち、最も指標値 f を小さくする α_s を選択して、より指標値を改善する母数ベクトル α^{*s} を出力する。Algorithm2 では、Algorithm3 によって α_1 から α_{35} の順に母数を順次特定し、初期値として与えられた母数ベクトル α^{*1} と、最後に得られた母数ベクトル α^{*35} を母数とする重み関数を得、これを回帰モデルに適用した際の指標値 f_1^{min} , f_{35}^{min} を出力する。Algorithm1 では、指標値 f_1^{min} , f_{35}^{min} が一致するまで、Algorithm2 による母数の特定をくり返すことを指示する。

なお、Algorithm2 では、 α_1 から α_{35} の順に母数を特定するアルゴリズムを示したが、 α_{35} から α_1 の順に母数を特定するアルゴリズムも構築した。これは、母数の特定順序が異なる両アルゴリズムによって得られた重みの一致を確認することにより、特定された重みの妥当性を確かめるためである。 α_1 から α_{35} の順に母数を特定したアルゴリズムをfrom1、 α_{35} から α_1 の順に母数を特定したアルゴリズムをfrom35と呼ぶ。

2-2-2 シミュレーションの条件設定

提案手法において、任意に設定する条件として下記7種類がある。

【設定条件】

T: 対象領域

δ : 施設等位置 $F_a(-\delta, 0)$, $F_b(-\delta, 0)$

n: サンプルサイズ (グリッドサイズ)

σ : 実現値の生成にあたり期待値に加える誤差の標準偏差

t: 指標 w 算出にあたり構築するモデルの数

interval: アルゴリズムによる重み関数の最適化にあたり母数を変化させる程度 (インターバル)

from: 重み関数の母数 α の特定順序

ここでは手法が適用できる条件設定の範囲について検討する。計算資源には限りがあり、すべての組み合わせを検討することはできないが、可能な範囲で妥当な条件設定を示す。対象領域T、施設等位置 δ は、所与のものとする。まず、interval、fromについて示し、次に、サンプルサイズn、構築モデル数tについて検討する。最後に、誤差の分散 σ^2 、回帰係数が異なる条件への適用に触れる。なお、実在市街地で得られる多様な観測値を想定すれば、同じ対象領域Tで施設等位置 δ を変化させた際、手法を適用できる δ の範囲を把握することも必要だが、この点に関しては今後の課題とする。

1) 重み関数の母数特定順序・インターバル

提案手法のアルゴリズムでは、重み関数の母数を特定する際に、母数を少しずつ変化させることにより、より歪を小さくできる母数を求める。アルゴリズムに入力値として与えるintervalは、母数を少しずつ変化させる際にどの程度変化させるかを定める値である。intervalの値が小さいほど精緻な重みが得られる。他方、from1とfrom35では重み関数の母数の特定順序が異なり、最も精緻な重みでは、母数の特定順序にかかわらず得られる母数は一致する。得られた重み関数の形状が from1とfrom35とで大

大きく異なることは、有効な重みが得られていないことを意味する。from1とfrom35で特定される重みベクトルが一致する精緻な重み関数を得るためには、intervalを可能な限り小さくする必要がある。intervalが十分に小さければ、母数を特定する順序（from1とfrom35との違い）にかかわらず一意に重みを特定することが可能だが、計算資源が膨大となる。そのため、実現可能な範囲で有効な重みを特定できるintervalを選択する必要がある。図-5は、intervalを変化させ特定した重み関数を示したものである。母数は完全に一致しないものの、interval0.01まで小さくすると、from1とfrom35で重み関数の形状は大きく異なることが分かる。ここでは、interval0.01を妥当なインターバルとみなす。

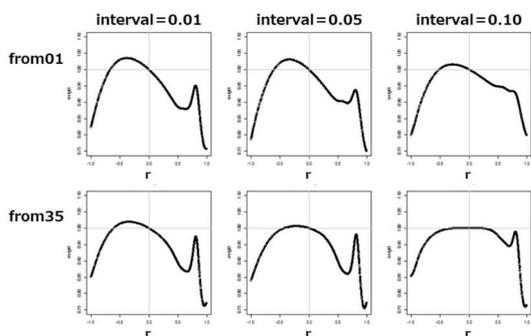


図-5 母数特定順序とインターバルの違いによる重み関数の比較 (n=6559, t=300 を例に)

2) 構築モデル数 t とサンプルサイズ n

アルゴリズムにおける構築モデル数 t や、サンプルサイズ n が小さければ、有効な重みベクトルの特定は難しい。ここでは、t と n の下限値について検討する。具体的には、1) 指標 f 算出にあたって t がいくつ以上となると安定した重みが得られるか、2) サンプルサイズ n は最低どの程度必要か、について示す。なお、ここで用いるアルゴリズムは、

from01、interval = 0.01である。まず、下記のパターンを組み合わせ、全部で 49 種類の重みを特定し、その有効性を検討した。

- ・ アルゴリズムにおける構築モデル数 t を変化させた 8 つの条件
t ∈ [100,200,300,400,500,1000,1500,2000].
- ・ 図-6のとおり、格子間隔gによりサンプルサイズ n を変化させた 7 つの条件
g ∈ [0.05,0.10, ..., 0.35].

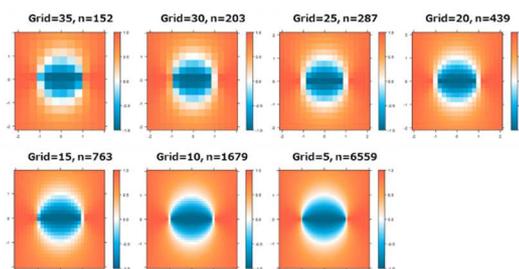


図-6 格子間隔の違いによるサンプルサイズの設定

(1) 構築モデル数 t

重みづけをしていない OLS に生じる回帰係数の歪 (f*OLS) に対して、重みづけにより低減された歪 (f*OLS - f*WLS) が占める割合 (以下、歪の低減割合) と呼ぶ) を下式で表す。

$$\text{「歪の低減割合」} = \frac{f^{*OLS} - f^{*WLS}}{f^{*OLS}} \quad (2-18)$$

図-7は、tの違いによる「歪の低減割合」を示したものである。

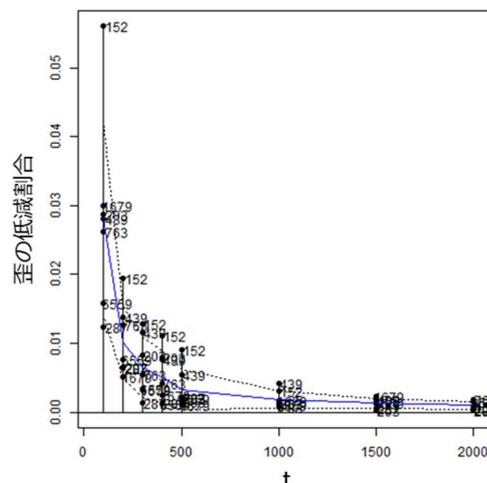


図-7 tごとの歪の低減割合
(実線：平均、点線：標準偏差1の範囲、
図中の数値：サンプルサイズn)

図- 7 より、 t が大きいほどサンプルサイズによる「歪の低減割合」のばらつきは減少し、 t が 1500 程度になれば「歪の低減割合」約 0.0015 で安定する。つまり、 t が十分に大きいとサンプルに依存しない重みが特定されやすくなる。また、 t が大きいほど「歪の低減割合」の値は小さくなる。

(2) サンプルサイズ n

図- 8 は、特定される重みが特に不安定な $t = 100$ 、 $t = 200$ を除いて、サンプルサイズごとの「歪の低減割合」を示したものである。(1)で示した通り、 t が大きいほど平らなグラフとなっており、サンプルサイズに依存しない結果が得られることが分かる。加えて、サンプルサイズが大きいと、 t によるばらつきも減少し一定の値に近づくことが分かる。

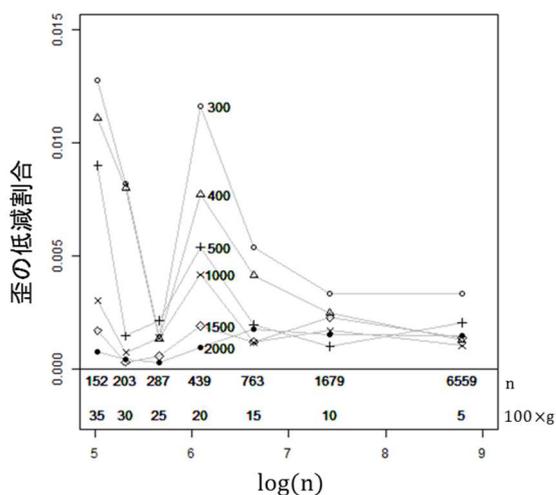


図- 8 サンプルサイズごとの「歪の低減割合」

(3) 重みの形状

図- 9 は、 t 、 n ごとの重み関数を図示したものである。 $t = 100$ では、 n の違いによって重みの形状は全く異なるが、 $t = 2000$ では、 n にかかわらず、枠内 ($t = 2000$, $n = 6559$)の重みの形状に近づく。これは、 t を大きくすることにより、サンプルごとの

誤差項の違いによる影響よりも、距離変数間の相関の影響を反映した重み関数が特定されやすくなるためだと考えられる。しかし、 $t = 2000$ であっても、依然として、 n の違いによって重みの形状に違いが認められ、安定した重みは得られていない。より安定した重みを得るためには、一定以上のサンプルサイズの確保と、十分に大きな t の確保が必要だと考えられる。

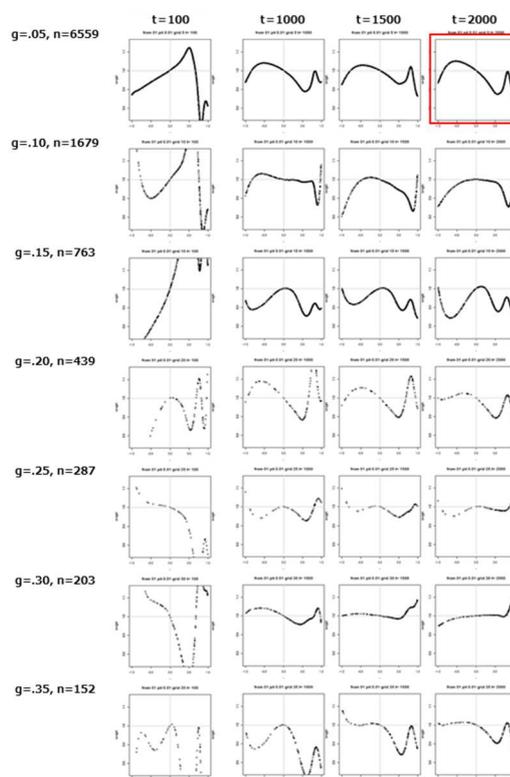


図- 9 重み関数

(横軸：距離変数間の相関係数 r 縦軸：重み)

3) 誤差の分散

提案手法では期待値 η に正規分布に従うランダムな誤差 $\epsilon_e \sim N(0, \sigma^2)$ を加えることにより、実現値 y_e を t 種類生成した。設定する分散 σ^2 が、特定される重みに及ぼす影響を確認したところ、 $\sigma = 1$ 、 $\sigma = 0.2$ 、 $\sigma = 0.1$ において、特定される重み関数の形状に大きな違いはなかった。また、分散 σ が小さい場合、重みの特定に要する時間が短くなる

ことから、計算資源節約の観点から分散 σ を期待値の分散の 1/20 程度まで小さくすることが有効だと考えられる。

4) 異なる回帰係数

ここまでの分析では、回帰係数の真値を $\beta_0^e=0$ 、 $\beta_1^e=1$ 、 $\beta_2^e=1$ としたが、異なる真値の回帰モデルによるテストでも、同様の重みを特定できる。

2-2-3 重み適用の効果

1) 特定された重み関数

表- 3 に示す条件で、重みの特定を行った。得られた母数ベクトル α^{best} を以下に、得られた重み w^{best} と r との関係を図- 10(上段)に示す。

表- 3 条件の設定

説明	設定
対象領域	$T = \{(x,y) -2 \leq x \leq 2, -2 \leq y \leq 2\}$
施設等位置 $F_a(-\delta, 0)$, $F_b(-\delta, 0)$	$\delta = 1$
アルゴリズムによる重み関数の最適化にあたり、母数を変化させる程度	interval = 0.01
重み関数の母数 α の特定順序	from01 from035
サンプルサイズ(グリッドサイズ)	$n = 6,559$ ($g = 0.05$)
指標 w 算出にあたり構築するモデルの数	$t = 2,000$
実現値の生成にあたり期待値に加える誤差の標準偏差	$\sigma = 0.1$

$$w^{\text{best}} = \exp(\mathbf{X}\alpha^{\text{best}}) \quad (2-19)$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nq} \end{pmatrix}$$

$\alpha_{\text{from1}}^{\text{best}}$

$$= \begin{pmatrix} -0.11, -0.07, -0.04, -0.07, 0.10, -0.08, -0.04, \\ 0.00, 0.01, 0.00, -0.03, 0.03, 0.00, 0.00, \\ 0.00, 0.02, 0.05, 0.00, 0.01, 0.00, 0.01, \\ 0.03, 0.00, 0.00, 0.00, 0.02, 0.06, 0.02, \\ 0.01, 0.00, -0.01, 0.03, 0.00, 0.00, 0.00 \end{pmatrix}^t$$

$\alpha_{\text{from35}}^{\text{best}}$

$$= \begin{pmatrix} -0.06, 0.01, -0.01, -0.05, 0.07, -0.05, -0.01, \\ 0.00, 0.00, 0.00, -0.04, -0.01, 0.00, 0.00, \\ 0.00, -0.06, 0.01, -0.01, 0.00, 0.00, -0.02, \\ 0.00, 0.01, -0.02, 0.01, -0.06, -0.01, -0.02, \\ 0.00, 0.02, 0.00, 0.05, 0.05, 0.05, 0.08 \end{pmatrix}^t$$

from1により特定された $\alpha_{\text{from1}}^{\text{best}}$ とfrom35により特定された $\alpha_{\text{from35}}^{\text{best}}$ とは完全に一致しないが、図- 10(上段)より、重み関数の形状は類似していることが分かる。指標値 f は、from01では4.00827、from35では4.00789であり、大きな差はみられない。

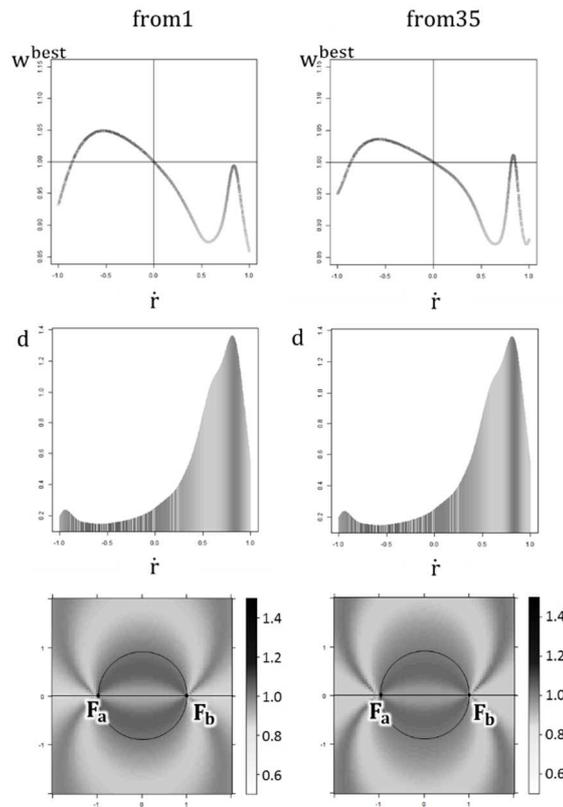


図- 10 重みの特徴

(上段：距離変数間の相関係数 r と重み、 中段： r と密度 d 、重みの関係、 下段：重みの空間分布)

* 分かりやすさのため、距離変数間の相関係数 r が0のとき重みが1となるように、得られた重みを距離変数間の相関係数 r が0のときの重みで除した値を示した。

図- 10(中段)は、距離変数間の相関係数 r の密度 d と重みの関係を示し、ウェイトが大きいほど濃い着色をしたものであり、図- 10(下段)は、同じ着色により、重みの空間分布を示したものである。密度 d が特に低い相関係数 r ではウェイトが高く(濃)、密

度dが低い場合でも、距離変数間の相関が強い相関係数rが-1 付近ではウェイトが低い(淡)重みとなっている。また、空間分布では、距離変数間の相関が強い2つの施設等(F_a、F_b)を結ぶ直線上の重みが軽い(淡)ことが確認でき、重みの形状は妥当である。

2) 重みづけの効果

表-4は、得られた重みw^{best}を、回帰式(2-9)に適用し、重み適用の効果を示したものである。

表-4 重み適用の効果

from01

	回帰係数の歪C _e $\sqrt{(\hat{\beta}_{1e}-1)^2 + (\hat{\beta}_{2e}-1)^2}$		推定結果の歪B _e $\sum_{i=1}^n (y_{e i} - \hat{\eta}_{e i})^2$		予測結果の歪B _i $\sum_{i=1}^n (\hat{y}_{e i} - \hat{\eta}_{e i})^2$	
	OLS	WLS	OLS	WLS	OLS	WLS
Min.	0.00005	0.00003	0.00006	0.00004	0.00006	0.00004
Qu.1st	0.00121	0.00121	0.01194	0.01181	0.01200	0.01185
Median	0.00190	0.00187	0.02271	0.02276	0.02272	0.02282
Mean	0.00201	0.00200	0.02933	0.02929	0.02945	0.02941
Qu.3rd	0.00267	0.00268	0.03956	0.03992	0.03980	0.04007
Max.	0.00614	0.00612	0.14584	0.15233	0.14545	0.15193
sd.	0.00107	0.00106	0.02384	0.02378	0.02392	0.02385
sum	4.01421	4.00827	58.66885	58.58170	58.90089	58.81038

From35

	回帰係数の歪C _e $\sqrt{(\hat{\beta}_{1e}-1)^2 + (\hat{\beta}_{2e}-1)^2}$		推定結果の歪B _e $\sum_{i=1}^n (y_{e i} - \hat{\eta}_{e i})^2$		予測結果の歪B _i $\sum_{i=1}^n (\hat{y}_{e i} - \hat{\eta}_{e i})^2$	
	OLS	WLS	OLS	WLS	OLS	WLS
Min.	0.00005	0.00002	0.00006	0.00003	0.00005	0.00003
Qu.1st	0.00121	0.00123	0.01194	0.01187	0.01175	0.01167
Median	0.00190	0.00187	0.02271	0.02270	0.02249	0.02265
Mean	0.00201	0.00200	0.02933	0.02928	0.02905	0.02900
Qu.3rd	0.00267	0.00268	0.03956	0.04012	0.03965	0.03972
Max.	0.00614	0.00612	0.14584	0.15249	0.14614	0.15094
sd.	0.00107	0.00106	0.02384	0.02376	0.02366	0.02358
sum	4.01421	4.00789	58.66885	58.56332	58.10684	58.00145

*WLSの太字は改善をグレー文字は悪化を示す。

t種類の異なる誤差項 ε_l^e 、 $e \in [1, t]$ による実現値 y_l^e を被説明変数、 $d_{a l}$ 、 $d_{b l}$ を説明変数とした回帰モデル(2-9)について、重みづけを行っていないOLSと、w^{best}により重みづけを行ったWLSを比較している。重みづけをしない場合のf*^{OLS}は4.01421であるのに対して、重みw^{best}により重みづけをしたf*^{WLS}は4.00827 (from01)、4.00789 (from35)あり、回帰係数の歪を表す指標fは、重みの適用により0.0059~0.0063ポイント低減された。これに伴い、被説明変数の推

定結果B_iの歪や予測結果の歪 \hat{B}_i の総和も低減されている。したがって、重みの適用により、多重共線性により回帰係数がばらつく問題が低減されているものと考えられる。しかし、f*^{OLS}に対する歪の低減割合は0.15~0.16%と僅かであり、個々の回帰モデルに重みを適用した際の歪の低減効果は必ずあるとは限らない。

2-2-4 まとめ

本節では、1)空間的多重共線性によって回帰係数の分散が大きくなる問題への対応手法を示し、2)手法が有効なシミュレーションの条件の範囲、3)手法の適用効果を示した。提案した手法の適用により、fで表される回帰係数の歪を低減することができることが分かった。しかし、その効果は大きくないということも同時に示された。次節(2-3)では、実際の観測値を用いたヘドニック分析に、提案手法を適用する。

2-3 事業評価のバイアスの測定

2-2で示した手法を、実際の不動産価格を用いたヘドニックモデルに適用しその効果を検証する。

2-3-1 対象範囲・観測値

1) 対象範囲

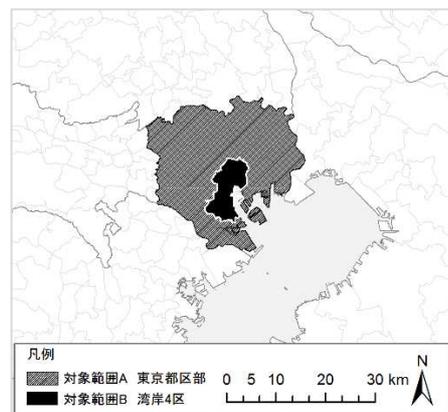


図-11 対象範囲

考案した事後処理方法を2つの対象範囲に適用する。対象範囲Aは東京都区部である。対象範囲Bは中央区、千代田区、港区、品川区の4区である。対象範囲を図-11に示す。

2) 観測値

アットホーム株式会社から提供された不動産取引データのうち、2011年から2014年に登録された、中古分譲マンションのデータを用いる。提供されたデータに対して下記のようなデータの精査を行い、最終的に得られたデータのサンプルサイズは、対象範囲A（東京都区部）では149,657、対象範囲B（湾岸4区）では24,134であった。これらのデータセットから、それぞれ20000をランダムサンプリングした。なお、全ての観測値を用いず、ランダムサンプリングを行ったのは、2つの対象範囲におけるサンプルサイズの統一をはかり、同一の条件で比較を可能にするためである。

【データクレンジング】

- ・ 築年数が古すぎる物件（片側の標準偏差2以上）を除外
- ・ 面積が狭すぎる物件、広すぎる物件（両側の標準偏差2以上）を除外
- ・ 建物階数がエラーと考えられる0階と60階以上を除外
- ・ 所在階が文字列で入力されているものは除外し、数値のみを採用

2-3-2 重みの特定

1) 条件

実際の物件位置に基づき、2-2と同様の方法で重みの特定を行う。2-2における F_a は東京駅、 F_b は横浜駅とし、東京駅から各物件の位置までの距離を距離変数 d_a 、横浜駅から各物件の位置までの距離を距離変数 d_b とする。条件の設定は表-6に示す通りである。

表-5 条件の設定

対象領域	対象範囲A(東京都区部), 対照範囲B(湾岸4区)	
施設等	東京駅 横浜駅	
"interval"	アルゴリズムによる重み関数の最適化にあたり、母数を変化させる程度	interval = 0.01
"from"	重み関数の母数 α の特定順序	from01
n	サンプルサイズ(グリッドサイズ)	n = 20,000
t	指標w算出にあたり構築するモデルの数	t = 1,500
σ	実現値の生成にあたり期待値に加える誤差の標準偏差	σ : 期待値の標準偏差の1/20

2) 距離変数間の相関とその密度

各対象範囲における相関係数 r の空間分布は図-12に示すとおりであり、その密度 d は図-13に示すとおりである。また、距離変数 d_a と d_b との相関係数 r は、対象範囲A（東京都区部）では-0.02128、対象範囲B（湾岸4区）では-0.96466である。

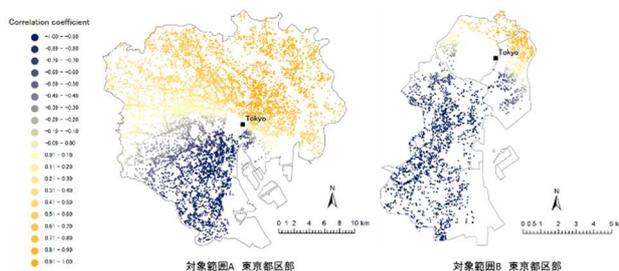


図-12 距離変数間の相関

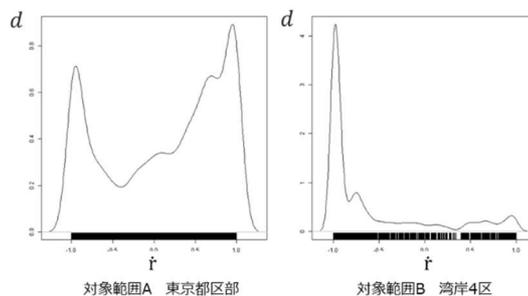


図-13 相関係数の密度

3) 特定された重み関数

得られた母数ベクトル α^{best} を以下に、得られた重み w^{best} と \hat{r} との関係を図-14(上段)に示す。

$$w^{\text{best}} = \exp(X\alpha^{\text{best}}) \quad (2-20)$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nq} \end{pmatrix}$$

$$\alpha_A^{\text{best}} = \begin{pmatrix} -0.01, -0.02, 0.00, -0.01, -0.02, -0.01, -0.01, \\ -0.01, 0.00, -0.01, -0.01, 0.00, 0.00, 0.00, \\ 0.00, -0.01, -0.01, -0.01, 0.00, 0.00, 0.00, \\ 0.00, 0.00, 0.00, 0.00, -0.01, 0.00, 0.00, \\ 0.00, 0.00, -0.01, 0.00, 0.00, 0.00, 0.00 \end{pmatrix}^t$$

$$\alpha_B^{\text{best}} = \begin{pmatrix} -0.02, -0.01, 0.00, -0.02, 0.01, -0.01, 0.00, \\ 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, \\ 0.00, -0.01, 0.00, 0.00, 0.00, 0.00, 0.00, \\ 0.00, 0.00, 0.00, 0.00, -0.01, 0.00, 0.00, \\ 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00 \end{pmatrix}^t$$

図-14(中段)は、距離変数間の相関係数 \hat{r} の密度 d と重みの関係を示し、ウェイトが大きいほど濃い着色をしたものであり、図-14(下段)は、同じ着色により、重みの空間分布を示したものである。密度 d が特に高い相関係数 \hat{r} ではウェイトが低く(淡)、密度 d が低い場合でも、距離変数間の相関が強い相関係数 \hat{r} が|1|付近ではウェイトがやや低い(淡)重みとなっている。

4) 重みづけの効果

表-6は、表-4と同様に得られた重み w^{best} を、回帰式(2-9)に適用し、重み適用の効果を示したものであり、重みづけを行っていないOLSと、 w^{best} により重みづけを行ったWLSを比較している。

表-6 重み適用の効果

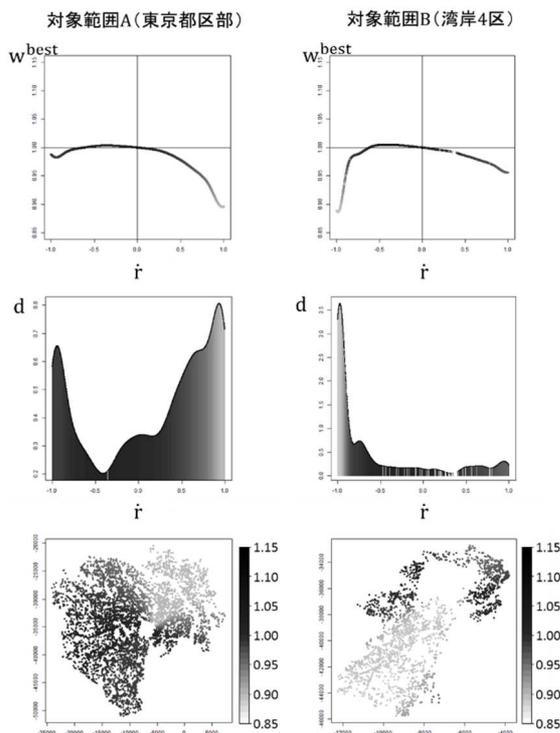


図-14 重みの特徴
(上段：距離変数間の相関係数 \hat{r} と重み， 中段： \hat{r} と密度 d ，重みの関係， 下段：重みの空間分布)

* 分かりやすさのため、距離変数間の相関係数 \hat{r} が0のとき重みが1となるように、得られた重みを距離変数間の相関係数 \hat{r} が0のときの重みで除した値を示した。

	回帰係数の歪 C_e		推定結果の歪 B_e		予測結果の歪 B_i	
	OLS	WLS	OLS	WLS	OLS	WLS
Min.	0.00005	0.00004	3033	4319	3033	4319
Qu.1st	0.00059	0.00059	165084	161815	165084	161815
Median	0.00089	0.00089	320282	321282	320282	321282
Mean	0.00098	0.00098	408514	407699	408514	407699
Qu.3rd	0.00129	0.00129	550420	550287	550420	550287
Max.	0.00352	0.00355	2150784	2120857	2150784	2120857
sd.	0.00054	0.00054	340586	339176	340586	339176
sum	1.46388	1.46309	612771062	611548357	612771062	611548357

	回帰係数の歪 C_e		推定結果の歪 B_e		予測結果の歪 B_i	
	OLS	WLS	OLS	WLS	OLS	WLS
Min.	0.00001	0.00000	94	52	94	52
Qu.1st	0.00035	0.00035	3713	3667	3713	3667
Median	0.00068	0.00068	7013	7075	7013	7075
Mean	0.00081	0.00081	8985	8981	8985	8981
Qu.3rd	0.00117	0.00116	11978	11881	11978	11881
Max.	0.00373	0.00373	58725	59046	58725	59046
sd.	0.00059	0.00059	7691	7686	7691	7686
sum	1.22007	1.21902	13477682	13471536	13477682	13471536

* WLSの太字は改善をグレー文字は悪化を示す。

対象範囲A(東京都区部)では、重みづけをしない場合の f^{OLS} は1.46388であるのに対して、重みづけをした f^{WLS} は1.46309であり、回帰係数の歪を表す指標 f は重みの適用により0.0010ポイント低減された。これに伴い、被説明変数の推定結果 B_i の歪や予

測結果の歪 \hat{B}_i の総和も低減されている。
 対象範囲 B (湾岸 4 区) では、重みづけをしない場合の f^{*OLS} は1.22007であるのに対して、重みづけをした f^{*WLS} は1.21902であり、回帰係数の歪を表す指標 f は重みの適用により0.0008ポイント低減された。これに伴い、被説明変数の推定結果 B_i の歪や予測結果の歪 \hat{B}_i の総和も低減されている。

しかし、 f^{*OLS} に対する歪の低減割合は、対象範囲 A (東京都区部) で0.054%、対象範囲 B (湾岸 4 区) で0.086%と僅かであり、個々の回帰モデルに重みを適用した際の歪の低減効果は必ずあるとは限らない。

2-3-3 ヘドニックモデルへの適用

得られた重みをヘドニックモデルに適用することにより、重みづけが、モデルの説明力、回帰係数の推定結果、被説明変数の推定結果に及ぼす影響を把握する。

1) 変数

アットホーム株式会社から提供された不動産取引データのうち、2011年から2014年に登録された、中古分譲マンションのデータを用いる。

(1) 被説明変数

被説明変数は物件価格であり、AICに基づく検討により物件価格の対数を採用した。各対象範囲の被説明変数の統計量を表-7に示す。

表-7 被説明変数

対象範囲A (東京都区部)	Min.	Median	Mean	Max.
物件価格	900,000	24,800,000	28,276,298	418,000,000
log(物件価格)	13.710	17.026	16.961	19.851
対象範囲B (湾岸4区)	Min.	Median	Mean	Max.
物件価格	900,000	28,800,000	37,761,885	500,000,000
log(物件価格)	13.710	17.176	17.176	20.030

(2) 説明変数

量的な説明変数の一覧を表-8に、ダミー変数の一覧を表-9に示す。距離変数は、各物件から東京駅までの距離、および、新宿駅までの距離である。その他の変数は物件価格に影響を及ぼすと考えられる物件の状態や立地に関する情報である。

表-8 説明変数一覧 (量的変数)

対象範囲A (東京都区部)	Min.	Median	Mean	Max.
東京駅までの距離	527	8,224	8,561	20,042
横浜駅までの距離	11,202	28,269	27,505	42,517
築年数	3	28	26	47
専有面積	4	53	52	481
建物階数	1	8	9	58
最寄り駅までの距離	14	433	484	2,680
対象範囲B (湾岸4区)	Min.	Median	Mean	Max.
東京駅までの距離	527	4,207	4,694	10,932
横浜駅までの距離	16,662	23,356	23,424	29,776
築年数	2	32	27	47
専有面積	10	47	52	491
建物階数	1	10	12	58
最寄り駅までの距離	17	333	360	984

表-9 説明変数一覧 (ダミー変数)

対象範囲A (東京都区部)			
ダミー変数の種類	所在階ダミー	路線ダミー	用途地域ダミー
ベース	(F1.01)	(JR山手線)	(商業地域)
ダミー変数	FLB6_B1	JR横須賀線 京成金町線	東急目黒線 第一種低層住居専用地域
	FL02_05	JR京浜東北根岸線 京成本線	東京メトロ丸ノ内線 第一種中高層住居専用地域
	FL06_10	JR京葉線 小田急小田原線	東京メトロ丸ノ内方南支線 第一種住居地域
	FL11_20	JR高崎線 湘南新宿ライン 宇都宮線	東京メトロ銀座線 第二種低層住居専用地域
	FL21_59	JR埼京線 成田スカイアクセス	東京メトロ千代田線 第二種中高層住居専用地域
		JR常磐線 西武新宿線	東京メトロ東西線 第二種住居地域
		JR千代田 常磐緩行線 西武池袋 豊島線	東京メトロ南北線 準住居地域
		JR総武 中央緩行線 西武有楽町線	東京メトロ日比谷線 準工業地域
		JR総武本線 都営三田線	東京メトロ半蔵門線 工業地域
		JR中央本線 都営新宿線	東京メトロ副都心線 工業専用地域
		JR東海道本線 都営浅草線	東京メトロ有楽町線
		JR東北本線 都営大江戸線	東京メトロ丸の内線
		つくばエクスプレス 都電荒川線	東京臨海高速鉄道りんかい線
		ゆりかもめ 東急世田谷線	東武伊勢崎 大師線
		京王井の頭線 東急多摩川線	東武亀戸線
		京王線 東急大井町線	東武東上線
		京急空港線 東急池上線	日暮里舎人ライナー
		京急本線 東急田園都市線	北総鉄道
		京成押上線 東急東横線	

対象範囲B (湾岸4区)			
ダミー変数の種類	所在階ダミー	路線ダミー	用途地域ダミー
ベース	(F1.01)	(JR山手線)	(近隣商業地域)
ダミー変数	FLB6_B1	JR横須賀線 京王線	東京メトロ丸ノ内線 第一種低層住居専用地域
	FL02_05	JR京浜東北根岸線 京急本線	東京メトロ銀座線 第一種中高層住居専用地域
	FL06_10	JR京葉線 湘南新宿ライン 宇都宮線	東京メトロ千代田線 第一種住居地域
	FL11_20	JR埼京線 都営三田線	東京メトロ東西線 第二種中高層住居専用地域
	FL21_59	JR総武 中央緩行線 都営新宿線	東京メトロ南北線 第二種住居地域
		JR総武本線 都営浅草線	東京メトロ日比谷線 商業地域
		JR中央本線 都営大江戸線	東京メトロ半蔵門線 準工業地域
		JR東海道新幹線 東急大井町線	東京メトロ有楽町線
		JR東海道本線 東急池上線	東京メトロ丸の内線
		ゆりかもめ 東急目黒線	東京臨海高速鉄道りんかい線

2) 説明力への影響

対象範囲 A (東京都区部) では、OLS の自由度調整済決定係数は0.8187、重みを適用した WLS の自由度調整済決定係数は0.8188であった。対象範囲 B (湾岸 4 区) では、OLS

の自由度調整済決定係数は0.8381、重みを適用した WLS の自由度調整済決定係数は0.8389であった。何れの対象範囲においても、重みの適用により説明力はやや向上した。

3) 推定回帰係数への影響

重みを適用していない OLS と、重みを適用した WLS による回帰係数の推定結果を比較したが、何れの対象範囲においても、OLS と WLS で、推定された回帰係数とその有意性に大きな差はみられなかった（報告書 表-10、表-11 参照）。対象範囲 A（東京都区部）で、OLS と WLS との回帰係数の差が最も大きかったのは、路線ダミー“北総鉄道”であり、OLS が-0.288、WLS が-0.293、差は0.005であった。対象範囲 B（湾岸4区）で、OLS と WLS との回帰係数の差が最も大きかったのは、路線ダミー“JR 東海道新幹線”であり、OLS が-0.085、WLS が-0.077、差は0.008であった。

差はみられない。

次に、重みを適用していない OLS と、重みを適用した WLS との、個々の観測地点における推定価格の差を示す。図- 16 は、推定価格の差の空間分布、および、その頻度を示したものである。何れの対象範囲においても、OLS と WLS の推定価格の差は、ほとんどの物件で50万円未満であり、大きな差はみられない。

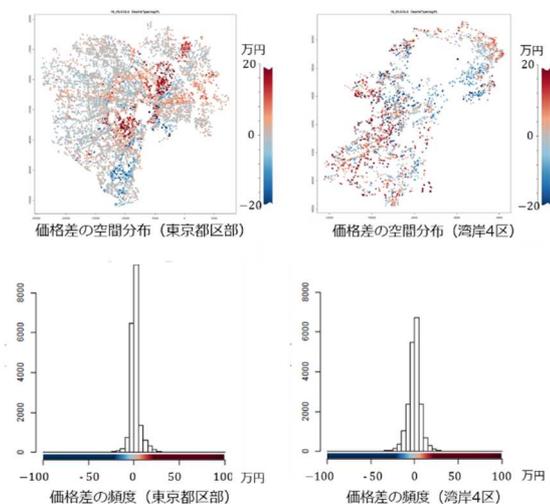


図- 16 価格推定結果の差 (OLS-WLS) とその頻度

4) 推定価格への影響

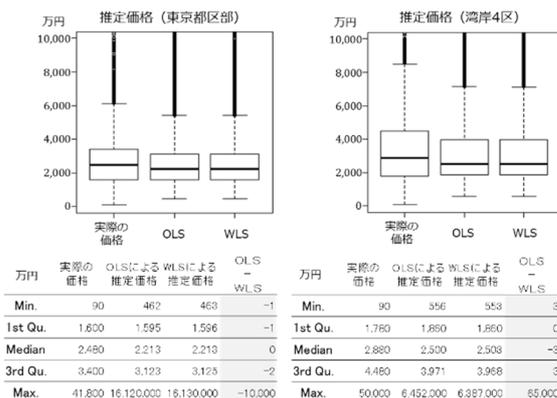


図- 15 価格推定結果

図- 15 は、実際の価格、重みを適用していない OLS による推定価格、重みを適用した WLS による推定価格を示したものである。重みを適用していない OLS と重みを適用した WLS で、推定価格の記述統計量に大きな

2-3-4 まとめ

2-2 で提案した手法を実際の不動産物件の位置に適用し、重みを得た。回帰係数の歪を表す指標 f は重みの適用により低減され、被説明変数の推定結果 B_i の歪や予測結果の歪 \hat{B}_i の総和も低減された。しかし、 f^{*OLS} に対する歪の低減割合は、対象範囲 A（東京都区部）で0.054%、対象範囲 B（湾岸4区）で0.086%と僅かであり、個々の回帰モデルに重みを適用した際の歪の低減効果は必ずあるとは限らないと考えられた。

次に、得られた重みをヘドニックモデルに適用し、重みづけが、モデルの説明力、回帰係数の推定結果、被説明変数の推定結果

に及ぼす影響を把握した。重みの適用により、モデルの説明力は向上することが示されたが、回帰係数の推定結果、被説明変数の推定結果への影響は僅かであった。

なお、重み適用の影響は僅かであるものの、モデルの説明力、回帰係数の推定結果、被説明変数の推定結果の何れにおいても、対象範囲 A（東京都区部）より、対象範囲 B（湾岸 4 区）の方が重み適用の影響が大きい。このことは、距離変数 d_a と d_b との相関係数 r が、対象範囲 A（東京都区部）では -0.02128 であるのに対して、対象範囲 B（湾岸 4 区）では -0.96466 と相関が強いため、生じている空間的多重共線性の影響が大きいためだと考えられ、距離変数間の相関係数 $|r|$ がより 1 に近い場合、手法適用の効果が、より顕著となる可能性を指摘できる。

3. おわりに

3-1 結果概要

本稿では、中心地、最寄り駅、主要施設等から価格観測点までの距離が説明変数として用いられるヘドニックモデルにおいて、複数の距離変数を用いると多重共線性の影響で回帰係数が歪む(Heikkila, 1988)という点に対処する方法を検討した。

最初に、2次元空間における空間的多重共線性の生じ方を把握し、回帰係数の歪を表すのに最適な指標を特定した。次に、回帰係数の歪を最小化できるよう回帰モデルに重みづけを行う手法を示した。最後に、この手法を実際の観測地点に適用し、得られた重みをヘドニックモデルに適用することで、手法の有効性について検討した。

空間的多重共線性の影響については、距

離変数間の相関が強いほど回帰係数の分散が大きくなるという問題が確認された。また、被説明変数の推定結果への影響は、回帰係数の歪への影響と比較して大きくないことから、多重共線性の影響を低減するためには、推定される回帰係数の真値からのずれを最小化する必要があることを示した。

回帰係数の歪を表す指標の検討では、複数の指標の候補の中から、距離変数間の相関 r や被説明変数の推定結果の歪 B との間の関連性が最も強いことから、 $C^{*d} = \sqrt{(\widehat{\beta}_1 - 1)^2 + (\widehat{\beta}_2 - 1)^2}$ を選択した。また、距離変数間の相関 r と C^{*d} 、被説明変数の推定結果の歪 B と C^{*d} の関連性が強いものに対して、距離変数間の相関 r と被説明変数の推定結果の歪 B の関連性は弱いことを示した。

回帰係数の歪を最小化するための手法の提案では、誤差項の異なる多様な実現値を用いた複数の回帰モデルにより C^{*d} を算出し、その総和を最小化できるような重みを特定する手法を示した。具体的には、観測地点ごとに得られる距離変数間の相関係数 r とその密度 d を変数とする多項式による重み関数を考え、 C^{*d} の総和を最小化する重み関数となるように母数を特定するアルゴリズムを示した。手法の適用の結果、 f で表される回帰係数の歪を低減できることが分かった。しかし、その効果は、提案手法を適用していない場合に生じる回帰係数の歪の約 0.15% と僅かであった。

得られた重みをヘドニックモデルに適用することにより、重みづけが、モデルの説明力、回帰係数の推定結果、被説明変数の推定

結果に及ぼす影響を把握した。重みの適用により、モデルの説明力が向上することが示されたが、回帰係数の推定結果、被説明変数の推定結果への影響は僅かであった。

1-1 結論

本研究において、本提案手法の適用効果は殆どみられなかった。その理由について考察する。

まず、本手法では、用いるt種類の実現値が同じであれば、パラメータの特定順序にかかわらず類似した重みが特定された。このことから、intervalを細かくする等アルゴリズムを精緻化すれば、あるサンプル群に対して特定される重みは、一意に決まるものと考えられ、アルゴリズム自体に問題はない。

そこで、回帰係数の歪が生じる原因に着目すると、大きく2つの原因が挙げられる。1) 距離変数間の相関による多重共線性の影響、2) 実現値のばらつき方による影響、の2種類である。サンプルサイズが小さい場合や、tが小さい場合には2) 実現値のばらつき方の影響が大きく、用いたサンプル群の実現値のばらつき方の違いに依存した重み(サンプル群固有の重み)が特定される。このような条件では、たとえ観測地点が同じ(空間的多重共線性の影響が同じ)であっ

ても、実現値の誤差項の異なるサンプル群を用いれば異なる重みが特定される。これに対して、tやサンプルサイズが十分に大きければ、実現値のばらつき方の影響は小さくなり、距離変数間の相関による多重共線性の影響を最小化する重み(共線性の影響を減らす重み)が特定されやすくなる。ところが、tを十分に大きくして得られた重みを適用した場合、低減できる歪は、重みを適用していない場合に生じる回帰係数の歪の約0.15%と僅かであった。このことから、回帰係数に生じる歪のうち、多重共線性の影響によって生じている歪は大きくないことが示唆される。

本研究におけるシミュレーションでは、距離変数 d_a と d_b との相関係数 \bar{r} は-0.298であり、相関が強いとは言えない。そのため、多重共線性の影響が生じにくい条件設定だったといえる。また、実際の不動産物件位置に基づく対象範囲B(湾岸4区)における検討では、距離変数 d_a と d_b との相関係数 \bar{r} は-0.965であり、相関が強いものの、地点の偏り等、その他の影響を排除しきれていないものと考えられる。今後、図-17のような多様な施設位置を設定したシミュレーションを行うことにより、相関係数 \bar{r} の異なる条件下での手法の有効性を検討することが課題である。

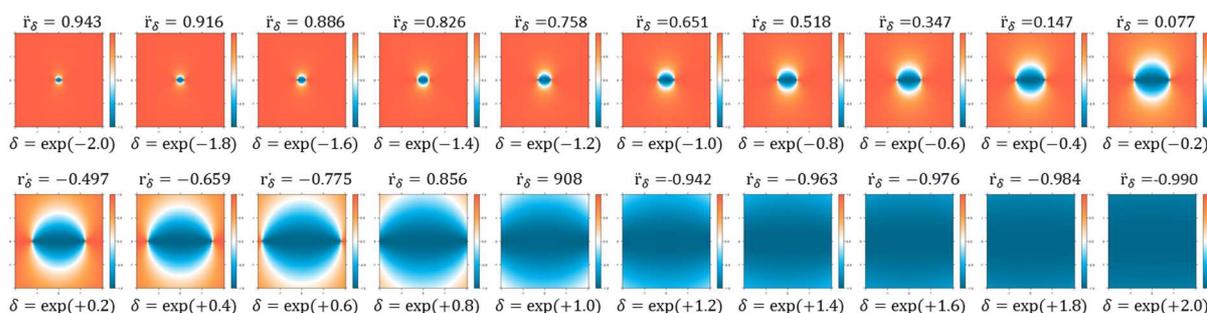


図-17 異なる δ における距離変数間の相関係数 \bar{r} の空間分布

なお、本手法では、距離変数間の共線性の影響を低減する手法を示したが、実際のヘドニックモデルでは他の変数にも共線関係が生じている可能性も高い。そのような場合には、距離変数間の共線性の影響の低減のみを考慮して得た重みによって重みづけを行うと、かえって回帰係数にバイアスを生じさせる可能性がある。そのため、ヘドニックモデルに本手法を適用する際には、慎重な検討を要する。

また、本研究で提案したアルゴリズムは、多大な計算資源を要するものであり、計算容量や計算時間による制約が生じた。今後、より効率的なアルゴリズムを考案することにより、精緻な重みを特定することも課題である。

なお、本研究では重み関数の母数の特定方法として、本提案手法の他、次の方法も検討したが、何れの方法でも、適切な重みの特定には至らなかった。

- ・ ランダムに母数を与え指標値が最も小さくなる母数ベクトルを特定する方法。
- ・ 次数の少ない多項式の母数を総当たりし、指標値が最も小さくなる母数ベクトルを特定する方法（グリッドサーチ）。
- ・ 区分線形関数により段階的に区分点を増していく方法。
- ・ 既存のアルゴリズム（Nelder-Mead 法、SANN 法）。

【参考文献】

- ・ Alin, A. (2010). Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics, 2(3), 370–374. <http://doi.org/10.1002/wics.84>
- ・ Asteriou, D., & Hall, S. G. (2007). Applied econometrics : a modern approach using EViews and Microfit (Rev. ed). Palgrave Macmillan.
- ・ Berry, B. J. L. (1976). Ghetto expansion and single-family housing prices: Chicago, 1968–1972. *Journal of Urban Economics*, 3(4), 397–423. [http://doi.org/10.1016/0094-1190\(76\)90038-3](http://doi.org/10.1016/0094-1190(76)90038-3)
- ・ Chatterjee, S., & Hadi, A. S. (2006). Regression analysis by example (4th ed). Wiley-Interscience.
- ・ Deaton, B. J., & Hoehn, J. P. (2004). Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. *Environmental Science & Policy*, 7(6), 499–508. <http://doi.org/10.1016/j.envsci.2004.08.003>
- ・ Heikkila, E. (1988). Multicollinearity in Regression Models with Multiple Distance Measures. *Journal of Regional Science*. Heightsstown: Wiley Subscription Services, Inc. <http://doi.org/10.1111/j.1467-9787.1988.tb01087.x>
- ・ Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). Applied Linear Statistical Models. McGraw-Hill Irwin.
- ・ Noonan, D. S., Krupka, D. J., & Baden, B. M. (2007). Neighborhood dynamics and price effects of superfund site clean-up. *Journal of Regional Science*, 47(4), 665–692. <http://doi.org/10.1111/j.1467-9787.2007.00525.x>
- ・ Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., ... Sabeti, P. C. (2011). Detecting Novel Associations in Large Data Sets. *Science*, 334(6062), 1518–1524.
- ・ Ross, J., Farmer, M., & Lipscomb, C. (2011). Inconsistency in Welfare Inferences from Distance Variables in Hedonic Regressions. *The Journal of Real Estate Finance and Economics*, 43(3), 385–400. <http://doi.org/10.1007/s11146-009-9221-z>
- ・ Wang, Y., Asami, Y., & Sadahiro, Y. (2015). A Study on Inference from Distance Variables in Hedonic Regression. *International Journal of Social, Education, Economics and Management Engineering*.
- ・ 吉田光雄. (1987). 重回帰分析における多重共線性と Ridge 回帰について. *大阪大学人間科学部紀要*, 13, 227–242.
- ・ 浅見泰司. (2015). 都市工学の数理. 日本評論社.
- ・ 貞広幸雄. (2015). 回帰分析において距離変数間の相関が小さい標本点配置. *都市計画論文集*, 50(3), 279–285. <http://doi.org/10.11361/journalcpj.50.279>