

大阪大学経済学

第66卷 第3号
2016年12月

OSAKA
ECONOMIC
PAPERS

大阪大学経済学会
大阪大学大学院経済学研究科
大阪府豊中市待兼山町

大阪大学経済学

(欧文誌名 Osaka Economic Papers)

本誌は大阪大学経済学会・大阪大学大学院経済学研究科の紀要として年4回、邦文ならびに欧文の論稿によって刊行される。

本誌の編集は、大阪大学経済学会によって選ばれた編集委員3名により行われる。編集委員は寄稿された研究成果を選定し、論文・覚書・資料および書評に類別して本誌を編集する。

大阪大学大学院経済学研究科に所属する研究者はその研究成果を本誌に寄稿することができる。なお、大阪大学大学院経済学研究科に所属しない研究者による研究成果も、大阪大学大学院経済学研究科における研究と密接な関係にあるものについては寄稿することができる。

なお、寄稿する際は「大阪大学経済学会」会員として、年会費¥4,000を納入する必要がある。

大阪大学経済学会会則

- 第1条 本会は大阪大学経済学会と称する。
- 第2条 本会は経済学、経営学の研究と発表を目的とする。
- 第3条 本会の事務所を大阪大学大学院経済学研究科資料室に置く。
- 第4条 本会は下記の事業を行う。
1. 雑誌「大阪大学経済学」の発行（年4回）
 2. 研究会及び講演会の開催（随時）
 3. その他、評議員会で適当と認めた事業
- 第5条 本会は下記の会員を以て組織する。
1. 普通会員（大阪大学大学院経済学研究科の教員、大阪大学の院生・学生・卒業生及び評議員会の承認を得た者）
 2. 賛助会員（本会の事業を賛助する者）
- 第6条 会員は本会の諸事業に参加できる。
- 第7条 本会に下記の役員を置く。役員の任期は2年とする。
1. 会長（大阪大学大学院経済学研究科長を以ってこれに充てる）
 2. 評議員（大阪大学大学院経済学研究科の教授・准教授・講師を以ってこれに充てる）
 3. 雑誌編集・庶務・会計の委員若干名（評議員中より互選する）
 4. 書記若干名（助手から互選する）
- 第8条 本会の運営はすべて評議員会の決議による。
- 第9条 会長は本会を代表する。
- 第10条
1. 普通会員は会費として年額4,000円を納入するものとする。
 2. 賛助会員は会費として年額10,000円以上を納入するものとする。
- 第11条 本会則の変更は評議員会の決議による。

大阪大学経済学会評議員

会長 堂目卓生

評議員 (ABC順)

阿部 顕三	鳩澤 歩	堂目卓生	深尾 葉子
福重 元嗣	福田 祐一	二神 孝一 (庶務)	廣田 誠
石黒 真吾	祝 迫 達郎 (会計)	勝 又 壮太郎 (会計)	葛城 政明
小林 敏男	松村 真宏	村宮 克彦	中川 功一
西原 理 (編集)	西村 幸浩	大西 匡光	太田 亘
恩地 一樹	小野 哲生	大屋 幸輔	Pierre-Yves Donzé (編集)
佐井 りさ	佐々木 勝	関口 倫紀	椎 葉 淳
高橋 慎	竹内 恵行	谷崎 久志	友部 謙一
浦井 憲	白井 正樹	渡辺 泰明	Wirawan Dony Dahana
許 衛 東	山本 千映	山本 和博 (編集)	山本 達司
安田 洋祐			

大阪大学経済学 第66巻 第3号

目 次

論文

人口移動要因としての地域アメニティ近接性 當 麻 雅 章 1

株式市場を反映するセンチメント・インデックスの構築と株価説明力の実証分析
..... 數 見 拓 朗 24

人口移動要因としての地域アメニティ近接性*

當 麻 雅 章[†]

概 要

本研究は、都道府県データを用いて、アメニティが人口移動に及ぼす影響を分析した。特に、地域間のアメニティ近接性が移住費用の低下となるかどうか検討した。移住による生活環境の変化が非金銭的な移住費用となって、居住地選択に影響する可能性がある。都道府県レベルの移住データを用いて推定した結果、アメニティは二つの性質を持つことを明らかにした。一つ目は、先行研究と同様に、その豊富さがその地域での効用を上昇させる垂直的異質性である。もう一つの性質は、地域間近接性、すなわち、移動先と移動元とでアメニティの特徴が似ているほど移住費用を低下させるという性質である。従って、アメニティの効果を分析する際に、各地域のアメニティの垂直的異質性に加えて、地域間のアメニティ近接性についても評価する必要がある。

JEL Classification : R12, R23

キーワード：アメニティ近接性、人口移動、集計ロジットモデル

1 はじめに

本研究の目的は、非金銭的要因、特にアメニティが地域間人口移動パターンに及ぼす影響に

ついて分析することである。地域間人口移動パターンは、各地域の効用と地域間移住費用によって決定される。各地の効用は、その地域の経済的豊かさや、居住環境など非金銭的要因によって決定される (Rodriguez-Pose and Ketterer (2012))。一方、移住費用には、金銭的費用に加えて、生活環境の変化などの非金銭的費用が含まれる (Sjaastad (1962))。本研究では、特に、生活環境の地域間近接性が地域間移住費用の低下となり、その結果、人口移動パターンに影響するかどうか検証する。

地域間人口移動は、各地域に居住することで得られる満足度 (効用) と、現在の居住地から他の地域への移住費用によって決定され、主な機会は、就職や転職といった経済的要因や進学や結婚といったライフサイクル的要因である (黒田ほか (2008))。特に、就業機会は人口

* 本研究に関して丁寧かつ熱心なご指導を賜りました東京大学佐藤泰裕准教授に心より感謝致します。また、有用かつ的確なコメント頂きました大阪大学佐々木勝教授、谷崎久志教授に深謝致します。並びに、東京大学 The Urban Economics Workshopにおいて貴重なコメントを頂きました金本良嗣教授、田淵隆俊教授、高橋孝明教授、太田充准教授、また日本経済学会 2014 春季大会において丁寧な助言を頂きました諸先生方、岡本亮介准教授、玉井寿樹准教授に感謝の意を表します。応用地域学会第 28 回大会において、討論者を務め有益な助言を頂きました経済産業研究所研究員近藤恵介氏に深く感謝いたします。そして、多くのご指摘を下さいました山本和博研究室の皆様にご感謝いたします。なお、本研究について有りうべき誤りは、筆者の責に帰せられる。

[†] 大阪大学大学院経済学研究科 経済学専攻 博士後期課程 連絡先：nge017tm@student.econ.osaka-u.ac.jp

流入に及ぼす影響力が大きい (Greenwood and Hunt (1989))。日本の高度経済成長期では、大都市への流入が大きく、これは就職や進学が主な理由とされている (平成 17 年度版労働経済の分析 (厚生労働省))。

しかし、一人当たり所得の増加は、アメニティに対する需要を増加させる (Graves and Linneman (1979)) ので、先進国では、経済社会要因に加えて、アメニティ (居住環境) が生活満足度 (効用) に影響し、その結果、人口移動パターンに影響を及ぼす。実際、人口移動の先行研究では、人口流入を促す要因の一つとしてアメニティに注目が集まり、その効果を分析している。まず、アメリカ国内の人口移動分析では、特に高度な能力を持つ労働者を引き付ける要因としてアメニティが注目され、温暖な気候、豊かな自然環境、芸術をはじめとする文化的活動の活発さなどが能力の高い労働者を惹き付けることが示されてきた (Graves (1979), フロリダ (2014))。次に、EU 内の人口移動パターンについても、アメニティの効果が検証され、効果を持つか否か盛んに議論がなされた。近年の EU 内における地域間人口移動移動では、賃金、失業率といった経済的要因に加えて、平均気温や降水量、また自然の豊かさなど、アメニティが人々を惹き付ける一因であることが示されている (Rodoriguez-Pose and Ketterer (2012))。さらに、日本では、人口移動におけるアメニティの分析は少ないが、伊藤 (2003) では重力モデルを用いて気候アメニティが人口移動に与える効果を分析している。その結果、気温の低い地域ほど転出が多くなっており、日本においてもアメニティが人口移動に影響していることが示された。

各地域の効用が移住パターンに影響する一方、地域間移住費用もまた、人口移動パターンに影響を及ぼし、さらに、この移住費用は、移住先によって異なる。人口移動には、通常費用が伴い、それは金銭的費用と非金銭的費用に分

けることができる (Sjaastad (1962))。Sjaastad (1962) によれば、非金銭的費用には、周囲の環境変化、例えば、家族や友人、慣れた生活環境から離れることが含まれる。この慣れた環境には、都市的生活スタイルや気候といった生活環境要因、すなわち、アメニティも含まれるであろう。つまり、現在の生活環境から離れた地域への移住には、費用がかかる。従って、移住費用は、移住先ごとに異なり、そのことが移住パターンを決定する要因となりうる。

そこで、本研究では、日本の都道府県間人口移動のデータを用いて、アメニティの地域間近接性が人々の移住先の決定に影響を及ぼすかどうかを検証する。他の都道府県へ移住する場合、金銭的費用だけでなく、生活環境の変化といった非金銭的費用についても負担する。本研究ではアメニティの変化を非金銭的移住費用と考え、3つの要因を分析する。1つ目は都市化の程度であり、都市部の住民にとって同じような都市部への移住の方が費用が小さいと想定する。2つ目は同程度の気温をもつ地域への移住費用は小さいため、移住が増加する。つまり、本研究では、例えば、温暖な地域の住民にとっては温暖な地域への移住の方が環境変化が小さいため、移住費用は小さい、また、比較的寒冷的な地域の住民にとっては寒冷的な地域へ移住する方が移住費用は小さいと予想する。最後は降水日数を用いて、気温と同様に、降水日数が似通った地域ほど移住費用が小さいと想定する。

本研究では、アメニティには二つの性質があることを発見した。まず、アメニティは垂直的異質性を持つものがある。本研究では、都市化の低い (自然が豊か) な地域ほど、移住者を惹きつける結果となった。一方、その近接性が人口移動に影響するアメニティもある。推定結果から、降水日数は、垂直的異質性ではなく、その地域間近接性の低下が移住確率を低下させた。このため、アメニティの効果を議論する際は、垂直的異質性に加えて、その近接性につい

でも検証すべきである。

本研究の貢献は、アメニティが人口移動に与える影響を、移住費用の側面から分析したことである。非金銭的費用の効果は、家族や友人と離れることによる (Sjaastad (1962), Schwart (1973))。また、社会的ネットワークがその地域への移住費用を低下させるとすると、経済的格差がないとしても移住が発生する (Carrington, Enrica, and Vishwanath (1996))。このように、ネットワークに関する費用が及ぼす影響は分析されているものの、これ以外の環境要因がもたらす効果は、我々が知る限り、分析されていない。

本論文の構成は以下の通りである。第2章では、移住行動のコンセプトを示し、アメニティ近接性と移動者数より算出される従属変数を定義、加えて、計量経済モデルを構築する。続く第3章では、分析に使用するデータセットを紹介し、また、近年の移住パターンとアメニティの関係を見る。第4章では、アメニティの効果を推定した。ここでは、アメニティの近接性だけでなく、垂直的異質性についてもその効果を推定し、さらに近接性については、異時点間の比較のために、1975年から1985年間の推定を行った。加えて、移住者の属性別にアメニティ近接性の効果を推定した。第5章では、推定結果を比較し、各アメニティの性質を検討した。そして、最後の章では、本研究の結論、政策的含意、そして今後の課題を述べ、本論文の結びとする。

2 集計ロジットモデル

本章では、移住行動のコンセプトを述べ、アメニティの近接性と“移住率” (従属変数) を定義する。本研究で使用する集計ロジットモデルとは、マイクロデータを用いた非集計ロジットモデルと同様な定式化に基づいたモデルである。しかし、非集計ロジットモデルとの違い、

選択確率をグループ (移住前地域) ごとの集計データから得ることができるため、OLSを利用できる。また、地域間移動者だけでなく、域内移動者、すなわち、地域間を移動しなかった人々の情報も分析に使用することができる (Schultz (1982))。

2.1 移住行動とアメニティ近接性

本研究におけるフレームワークは以下のとおりである。まず、人々は確率 λ で移住するための機会が与えられる。そして、移住の機会が巡ってきた人は、現在の居住地で得られる効用と各移住先で得られる効用を比較し、最も高い効用を得る地域へ移住する (域内移動も含めて)。すなわち、

$$U_{ij,t} - U_{ii,t} > 0, \quad (1)$$

ここで、 $U_{ij,t}$ は t 期に地域 i から地域 j へ移住したときの効用を表し、これは移住費用も考慮している。また、 $U_{ij,t}$ は、次式のようなランダム効用を仮定する。

$$U_{ij,t} = V_{ij,t} + \epsilon_{ij,t}, \quad (2)$$

ただし、 $V_{ij,t} = V_{j,t} - d_{ij,t} = \theta_j + X_{j,t}\beta - R_{ij}\gamma - A_{ij,t}\delta$ 。ただし、 $V_{j,t}$ を確定項、 $\epsilon_{ij,t}$ を確率項とし、その確率分布をタイプ1の極値分布と仮定する。また、 $V_{j,t}$ を t 期における地域 j の効用を表す関数、 $d_{ij,t}$ を t 期における地域 ij 間の移住費用を表す関数として、線形関数を仮定する (ただし、 $d_{ii} = 0$ とする)。さらに、 θ_j と $X_{j,t}$ はそれぞれ地域 j における固定効果と経済社会的変数を表す。一方、地域間の関係性を表す特徴のうち、時間に依存しない特徴を R_{ij} 、時間を通じて変化する特徴を $A_{ij,t}$ で表し、この $A_{ij,t}$ が地域アメニティの近接性を表す。本研究では、この近接性をアメニティの差の絶対値、すなわち、 $|A_{j,t} - A_{i,t}|$ と定義する (この定式化の問題点は第5章で議論する)。従って、 $A_{ij,t}$ の増加は、地域 ij 間の近接性の低下を意味する。各要因の係数ベクトルは β 、 γ 、 δ で表す。

地域効用の確率分布における仮定から、地域 i から地域 j への移住確率は、

$$p_{ij,t} = \frac{\exp(V_{ij,t})}{\sum_{s=1}^{47} \exp(V_{is,t})} \quad (3)$$

と書ける。ここで、 $p_{ij,t}$ と、地域 i から地域 j へ移住する確率、すなわち移住の機会が巡ってきた時、域内移動を選択する確率 $p_{ii,t}$ との比を取ると、次のように変形できる。

$$\frac{p_{ij,t}}{p_{ii,t}} = \frac{\exp(V_{ij,t}) / \sum_{s=1}^{47} \exp(V_{is,t})}{\exp(V_{ii,t}) / \sum_{s=1}^{47} \exp(V_{is,t})} = \frac{\exp(V_{ij,t})}{\exp(V_{ii,t})}$$

さらに両辺の対数を取ると、

$$\log\left(\frac{p_{ij,t}}{p_{ii,t}}\right) = V_{ij,t} - V_{ii,t} \\ = \theta_j - \theta_i + (X_{j,t} - X_{i,t})\beta - R_{ij}\gamma - A_{ij,t}\delta \quad (4)$$

を得る。

2.2 データとの対応と移住率

次に、移住確率と移住者のデータと対応させるため、選択確率と移住者数との関係を導く。ここで域内では同質な人々を仮定すると、地域 i の住人 $N_{i,t}$ のうち地域 j へ移住する人数 $m_{ij,t}$ は、

$$m_{ij,t} = \lambda N_{i,t} p_{ij,t}$$

で与えられると仮定する。簡単化のため、移住機会が与えられる確率 λ は全地域で共通とし、移住者数は人口に比例すると仮定する。

この式を書き換えると、地域 i から地域 j への移住確率は、

$$p_{ij,t} = \frac{m_{ij,t}}{\lambda N_{i,t}}$$

で与えられる。これを式 (3) に代入すると、

$$\log\left(\frac{p_{ij,t}}{p_{ii,t}}\right) = \log\left(\frac{m_{ij,t}/\lambda N_{i,t}}{m_{ii,t}/\lambda N_{i,t}}\right) = \log\left(\frac{m_{ij,t}}{m_{ii,t}}\right)$$

を得る。従って、式 (4) は、移住者数を使って、

$$\log\left(\frac{m_{ij,t}}{m_{ii,t}}\right) = \theta_j - \theta_i + (X_{j,t} - X_{i,t})\beta - R_{ij}\gamma - A_{ij,t}\delta$$

と書き直せる。以下では、この左辺を“移住率”と呼ぶ。

2.3 識別および推定式

本節では推定に用いる計量経済モデルを構築し、加えて、アメニティの垂直的異質性を推定するときに用いるモデルを述べる。

識別にあたり特に問題となるのは、人口移動と経済社会状況が同時に決定されることである (Treyz et al. (1993), Ozgen, Nijkamp, and Poot (2011))。つまり、居住地は移住先の経済社会的状況に基づいて決定される一方、人々の移動によって地域の居住者が変わることによって地域特性が変化する。そのため、人口移動データを含めた空間的分析の場合、推定に当たって内生性の問題が生じる。従って、被説明変数と同時点の説明変数は推定式に含めるべきではない。この問題を回避するために、すべての説明変数について一期ラグを取る。

従って、アメニティの水平的効果を推定するモデルは、次のように書き表せる。

$$\log\left(\frac{m_{ij,t}}{m_{ii,t}}\right) = \theta_j + \theta_i + (X_{j,t-1} - X_{i,t-1})\beta + R_{ij}\gamma + A_{ij,t-1}\delta + year_t + e_{ij,t} \quad (5)$$

ここで、 $year_t$ は年ダミー、 $e_{ij,t}$ は誤差項を表す。本研究において着目するのは、アメニティの地域間類似性を表す $A_{ij,t-1}$ の効果、 δ である。

一方、アメニティの垂直的異質性効果を推定するために、これを定義する必要がある。本研究では、アメニティの垂直的異質性を、居住地と比較した変数 ($A_{j,t-1} - A_{i,t-1}$) と定義する。従って、垂直的異質性の効果 (ζ) を推定する式は、

$$\log\left(\frac{m_{ij,t}}{m_{ii,t}}\right) = \theta_j + \theta_i + (X_{j,t-1} - X_{i,t-1})\beta + R_{ij}\gamma + (A_{j,t-1} - A_{i,t-1})\zeta + year_t + e_{ij,t} \quad (6)$$

となる。

アメニティの垂直的異質性の効果について、その解釈は、例えば、移住先の気温の高さが移住率を増加させるかどうかである (他の条件を所与として、気温の高さが人々を惹きつける

か)。一方、本研究で用いるアメニティの近接性 ($|A_{j,t-1} - A_{i,t-1}|$) は、現在の居住地と比較したアメニティの違いが人口移動に及ぼす影響として解釈される。つまり、近接性に関する係数 (δ) が負であれば、例えば、大きな気温の違いに直面すると、その地域間では近接性が低いため、移住確率が低下することを意味する。

3 データ

本研究では、三つの都道府県レベルデータセットを構築した。一つ目は、1997年から2006年までのデータである。このデータセットを用いて、アメニティの近接性効果と垂直的異質性効果を推定する。また、近接性の効果に関して、異時点間の比較をするために、1975年から1985年の期間についても推定した。さらに、移住者の属性別にアメニティの近接性効果を推定するために、2000年の国勢調査を使用して移住率を算出した。

各データセットでは、47都道府県すべての移動者データを用いて、移動前地域と移動先地域の組み合わせ ($47 \times 46 = 2162$) それぞれを一つの標本としてデータセットを作成した。推定に際して、識別のため、移住率と説明変数の間に1期ラグをとった。また、1997年から2006年のデータセットでは、移住者数が0人のサンプルがあり被説明変数を定義できないためにサンプルから取り除いた。結果としてこのデータセットでは、サンプルサイズは19457 ($= 2162 \times 9 - 1$) となった。また、1975年から1985年におけるデータセットのサンプルサイズは21620であり、2000年のデータセットでは、クロスセクションデータであるため、そのサンプルサイズは2162である。

被説明変数である移住率は、第3章で述べたように、地域間移動者と地域内移動者を使って求められる。すなわち、本研究では、地域*ij*間の移住率を、移動前地域*i*から移動先地域*j*へ

の移動者数と地域*i*の域内移動者数の比を対数化したものと定義する。

人口移動パターンを決定する要因は、社会経済的要因とその他の要因に分けることができる。社会経済的要因は、先行研究によって数多く分析されており、特に、経済的要因は地域の効用を決定する重要な要因であることが様々な研究で示されている（例えば、Partridge and Rickman (2003)）。本研究ではアメニティに注目するが、先行研究に従い、重要と考えられる経済社会的要因を含めて推定する。また、地域間の関係性を表す地域間距離や社会的ネットワークも含める。

なお、以下で述べる各変数の定義、予想符号条件、データソースを表1にまとめる。

3.1 移住率

被説明変数は、移動元地域*i*から移動先地域*j*への移住者数と移動元地域*i*内における移動者数に基づいて求められる。地域間の移住者数と域内移動者数の比を取り、それを対数化したものを被説明変数とし、第二章で述べたように、これを移住率と呼ぶ。地域間移動者数と域内移動者数のデータソースは住民基本台帳人口移動報告である。ただし、移住者の属性別分析のために、2000年の国勢調査を用いて、属性別の移住率を算出する。

3.2 経済社会的要因

まず、経済状況は移住先の決定に大きな影響を及ぼすと考えられる。そこで地域の経済状況として、一人当たり県民所得を用いる。一人当たり県民所得は県民経済計算より取得し、生産要素を提供した県内の居住者に帰属する所得として定義される。所得は地域の効用水準に大きな影響を与えると考えられるため、予想される符号条件は正である。

移住に際して、就業機会も重要な決定要因である。たとえ豊かな地域に移住しても就業でき

表 1 データ, 予想符号, 定義, データソース

変数名	予想符号	定義	データソース
従属変数			
移住率		地域 <i>i</i> から地域 <i>j</i> への地域間移住者数と地域 <i>i</i> における域内移動者数の比を対数化; $\log \left(\frac{\text{地域 } i \text{ から } j \text{ への移住者数}}{\text{域内 (} i \text{ から } i \text{) での移動者数}} \right)$	「住民基本台帳人口移動報告」(総務省) および国勢調査 (総務省)
説明変数			
一人当たり県民所得	+	県内純生産と県外純所得の和を都道府県人口で除したもの	「県民経済計算」(内閣府)
有効求人倍率	+	有効求人数(年度計)を有効求職数(年度計)で除したもの	「職業安定業務統計」(厚生労働省)
事業所数	+	民営、国及び地方公共団体の事業所数	「事業所・企業統計調査」(総務省)
第三次産業就業者比率	+	第三次産業就業者/労働人口	「国勢調査」(総務省)
大卒者人口比率	+	国勢調査における大卒者数を都道府県人口で除したもの	「国勢調査」(総務省)
大学数	+	都道府県内に所在する大学本部の数	「学校基本調査」(文部科学省)
若年者人口比率	?	15歳から24歳人口を都道府県人口で除したもの	「人口推計」(総務省)
地域間距離	-	都道府県庁所在地間距離	国土地理院
累積移住者数	+	1954年から1973年までの移住者数の総和	「住民基本台帳人口移動報告」(総務省)
アメニティ			
都市化	- (近接性)	都道府県総面積に占める国立公園、国定公園、および都道府県立自然公園の割合	「自然公園の面積」(環境省)
年間平均気温	- (近接性)	1日24回の観測値から算術平均により日平均気温を求め、日の値から同様にして月平均気温を求めた後、12か月分の月平均気温を算術平均した値	「気象統計」(気象庁)
年間降水日数	- (近接性)	日降水量が1ミリメートル以上であった日の年間の日数	「気象統計」(気象庁)

なければ、その地域に居住することのメリットは小さいであろう。そこで、本研究では、これを有効求人倍率と事業所数によって捉える。まず、有効求人倍率は、有効求人数を有効求職数で除したものとして定義され、「職業安定業務統計」（厚生労働省）によって計算された指標を使用した。有効求人倍率の符号条件は、移住者は職を得る機会に恵まれている地域へ流入が起こる（Partridge and Rickman (2003)）と考えられるので、正である。

次に、事業所数が多いほど、就業の機会も多いと想定する。日本では、生涯を通じた平均的移住回数が少ない（Nakajima and Tabuchi (2011)）。これは主な移住機会のひとつは転職や就職など労働に関わるためと考えられる。従って、事業所が多く就業機会が多いと予想される地域への移住が多い可能性が考えられる。本研究では、事業所数を政府・民間の区別をせず、両方を足し合わせた総数で定義し、事業所・企業統計をデータソースとする。予想される符号条件は正である。

移住パターンは産業構造によって影響を受ける可能性がある。それは、産業により移住の容易さが異なる可能性が考えられるためである。農林水産業は生産の性質上、特定の場所においてのみ生産活動が可能であることから、地域間移動は少ないと考えられる。また、工業の場合、生産地である工場は、一度立地すれば移転が少ないと予測される。一方、第三次産業の場合、生産資源は人とともに移動すると考えられるため、相対的に移住は容易と考えられる。以上のように、従事する産業によって移住行動が変わることが考えられ、比較的移住の費用が低い、第三次産業従事者が多く移住しているであろう。転勤、就職が移住の主な理由であるとする、第三次産業に従事している労働者の多い地域に人々は移住していると考えられるため、予想される符号条件は正である。産業構造を表す変数として、第三次産業従業者比率を用いる。

第三次産業従業者比率は労働人口のうち第三次産業に従事する労働者の割合として定義し、国勢調査をデータソースとして算出した。国勢調査が実施されていない期間においては、調査された時点間の幾何平均によりこれを求めた。

大卒者が多いほど、将来にわたって、地域経済の高いパフォーマンスを期待できる。能力の高い労働者が増加すると他の労働者に正の外部性が生じ、その結果、生産性が高まることが期待できる。このような外部性が移住者の意思決定に影響する可能性を考慮し、本研究では、高度なスキルを持つ労働者を大卒者によって表す。大卒者人口比率は、都道府県人口に占める大卒者人口の割合として定義される。ただし、大卒者人口は国勢調査の大調査において調べられるため、10年の一度しかデータを取得できない。そこで、国勢調査が実施されていない年度については、調査された期間の幾何平均によって計算された大卒者人口と、各年の都道府県人口によって大卒者比率を求める。

人口移動パターンは経済的要因だけでなく、それぞれの地域の社会構造によっても決定される（Rodríguez-Pose and Tselios (2010)）。まず、大学進学は移動の理由の一つと考えられる（黒田ほか (2008)）。データの制約上、移住者個人の特性を考慮することはできないが、大学数が多い地域ほど他地域からの進学者による移住が増加すると予想される。従って、予想される符号条件は正である。ただし、「学校基本調査」（文部科学省）では大学数はその都道府県に立地する大学本部の数として定義される点に、注意が必要である。このため、都道府県に分かれて複数のキャンパスをもつ大学については、大学本部を抱えていないキャンパスの影響を考慮できない。

次に、地域住民の人口構成が人口移動に及ぼす影響を考慮する。若年者は環境の変化に対応する能力が高いなど、移住を容易に行うことが可能と考えられる。そのために、15～24歳を

若年者とし、若年者人口を都道府県人口で除したものを若年者人口比として定義し、推定に含める。ただし、若年者が多い地域へ移住するか定かでないため、予想される符号条件は不明である。また、年齢別人口の調査は毎年ではないので、「人口推計」(総務省)の年齢別人口を代理として用いる。

3.3 距離および社会的ネットワーク

地域間移住行動においては、地域間距離といった、地域間の関係性に基づいた変数も考慮すべきであると考えられる。本研究では地域間の関係性を表す要因として、地域間距離と、以下で定義する社会的ネットワークを推定に含める。

地域間の距離が離れるほど移住数は減少することが長年支持されてきた(Lucas(2001))。これは重力モデルと呼ばれ、移動者数は人口やGDPなど規模の大きい地域間同市で増加すると考えている。本研究では、県庁所在地間距離を地域間距離として定義し、国土地理院よりデータを取得した。

潜在的移住者にとって、自らの知り合いがいる地域への移住は容易になる可能性がある。例えば、家族や親戚、友人などのネットワークによって、居住地を決定しているかもしれない。本研究では地域間のネットワークを過去の移住者によって構築されたものと考え、これを累積移住者数によって定義する。住民基本台帳人口移動報告をデータソースとして、高度経済成長期の1954年から1973年まで移動者数の総和によって定義する。このようなネットワークは過去の移住者数が増加するほどその効果は強まると考えられるため、予想される符号条件は正である。

3.4 アメニティ

アメニティは、各地域の垂直異質性だけでなく、地域間の近接性が移住費用となって影響す

る可能性がある。これまで、社会経済的要因は多くの分析で考慮されてきたが、近年は地域の特性を表すアメニティに注目が集まっている。例えば、気候や公園といった、地域に特有のアメニティ垂直性が近年の人口移動パターンの決定要因の一つとなっていることを示した(フロリダ(2014), Rodoriguez-Pose and Ketterer(2012))。そこで、本研究では、地域のアメニティとして代表的な要因と考えられる都市化と気候(気温、および降水日数)について、その近接性の効果を推定する。つまり、移動先とのアメニティの差異、すなわち、生活環境の変化が移住費用となって人口移動に影響するかどうか検証する。

都市化アメニティの変化は、移住費用となる可能性が考えられる。たとえば、東京の大都会の住人であれば、同じような大都市である大阪への移住を好み、一方で、都会暮らしをしたことがないような、地方都市在住の人は、同じような地方都市あるいは田舎暮らしを好むかもしれない。つまり、都市化が現在の居住地と比較して変化しない似た地域を選択する可能性がある。それは、都市化が変化することで、住人との付き合い方など、生活様式に変化が生じ、この変化が負担となるためと考えられる。本研究では、「自然公園の面積」(環境省)¹を用いて、この都市化を測定する。都道府県面積に占める自然公園の割合が高い地域は、土地が開発されておらず、都市化されていない地域と想定する。

さらに気候アメニティの地域間近接性を測り、これが人口移動に及ぼす効果を推定する。気候の変化が大きい地域間を移動すると、季節ごとの過ごし方といった、生活パターンの変更を余儀なくされるであろう。つまり、このような地域間の移動には、生活様式の変更によって

¹ 自然公園とは、国または地方自治体によって決定される自然保護区域であり、国立公園、国定公園、都道府県立自然公園の総称である。また、自然公園は、国・地方自治体によって、自然の風景地として相應しい区域が選ばれる。

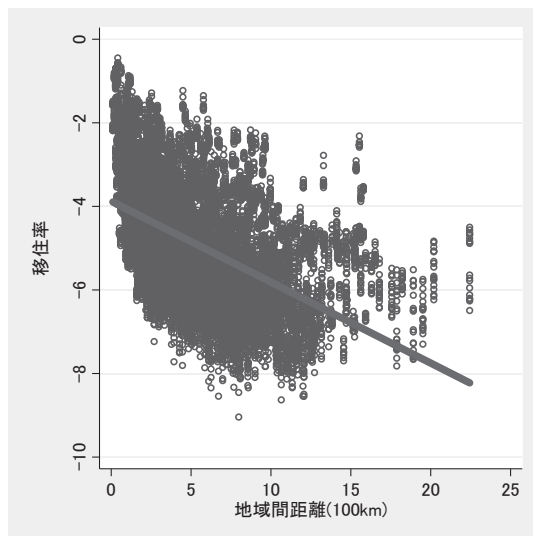
負担がかかると考えられる。

本研究では、気候アメニティとして、年間平均気温と年間降水日数を用いる。まず、年間平均気温は、次のように計算される。1日24回気温を計測し、そこから日平均気温を求める。この日平均気温から月平均気温を求め、さらに月平均気温から年間平均気温が求められる。また、年間降水日数は、日降水量が1mm以上であった日の年間の合計として定義される。気温および降水日数ともに、気象庁気象統計をデータソースとしている。観測は主に県庁所在地で行われ、各都道府県における気象の特性を把握している。気候に関するアメニティ近接性についても、都市化の場合と同様の方法で測定する。気候に関する変数の利点は、外生的に決定されるため、内生性の心配が小さいことである。

3.5 日本の移住の概観および移住率と距離及びアメニティの相関

本節では、日本の移住行動を概観し、また、移住率とアメニティとの間に見られる傾向を述べる。まず、移住の傾向を知るために、移住先としてどの都道府県が選択されているかを確認する。2000年における、移住先としてもっとも多く選択されているのは東京都である。東京を除いた46道府県中19道府県で東京への移住が最も多く観察され、それは主に東北、関東、中部からの移住である。次いで、10都道府県で大阪への移住が最も多くなっており、主に関西、四国からの移住が多い。さらに、6都道府県で福岡県への移住が最も多くなっており、すべて九州地方からである。つまり、大都市への移住が多く、それは近隣地域からの移住が多いことがわかる。

次に、移住先選択と地域間距離について確認すると、図1から明らかなように、距離と移住率との間には負の相関が観察される。つまり、地域間距離が低下するほど、移住率は低下する傾向にある。



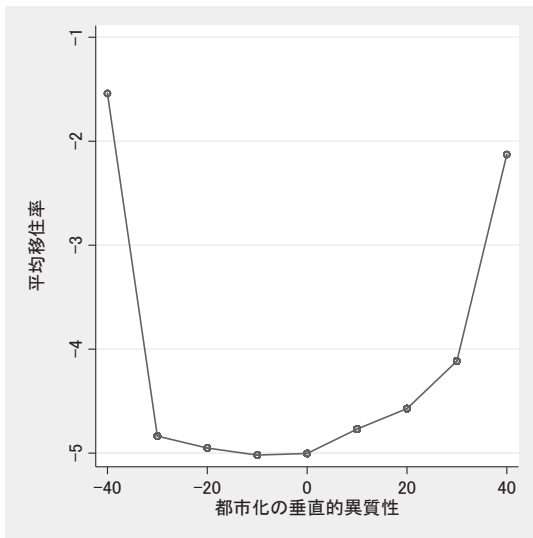
注釈：移住率データの期間は1998年から2006年の9年間。移住率は、地域*i*から*j*への移住者数を*i*域内移動者数で除し、その対数値を指す。

図1 移住率と地域間距離

さらに、移住率とアメニティの関係を観察する²。まず、図2を見ると、移住率と都市化の関係はU字型を描いている。これは、都市化の程度が大きく変化するような移住が多い、つまり、都市化の程度が低くなる移住が多いと同時に、都市化の程度が高くなる移住も多いことを表している。ただし、U字の両端には注意が必要である。この両端では移住率が高くなっているが、これらの階層に属するサンプルは“滋賀-京都”間だけである。従って、U字の両端部は、距離の近さによる移住の多さを反映している可能性がある。つまり、U字型の関係は見せかけである可能性がある（図2と推定結果との比較は、第5章で行う）。

次に、図3、図4より、気温と降水日数に関して、居住地域との大きな変化を伴う地域ほど、移住率が低下している。つまり、居住地域

² 図2、3、4では、横軸は、アメニティの近接性ではなく、垂直的異質性を表す。また、各図の縦軸では、アメニティの垂直的異質性を複数の階層に分け、その階層内の平均移住率をプロットしている。



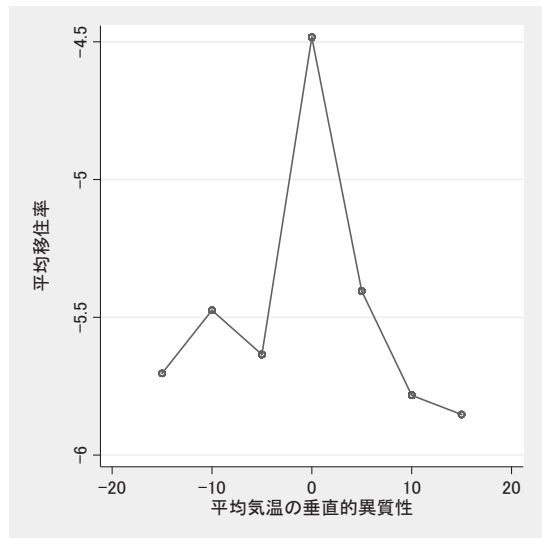
注釈：都市化の地域差とその階層内平均移住率をプロットした。横軸は、移住先と移住前地域の自然公園の差を表し、正であれば移住先の自然が豊かであることを意味する。各階層の幅は10%とした。例えば、都市化の地域差0は、-5%以上から5%未満の階層を意味する。ただし、両端の-40及び40の階層は、それぞれ、-35未満、35以上のクラスを表す。また、平均移住率は、各階層内に属するサンプルの移住率に関する平均を表す。

図2 移住率と都市化垂直的異質性

と比較してアメニティが増加あるいは減少のいずれにおいても、移住率と負の相関を持つように見える。これより、気候の変化が移住率を下げている可能性が考えられる。

4 推定結果

本章では、アメニティの地域間近接性及び垂直的異質性の効果に関する推定結果を示す。まず、近年の人口移動において、アメニティの近接性が人口移動に及ぼす効果を示し、次に、先行研究のように、アメニティの垂直的異質性効果の推定結果を示す。そして、時間を通じた効果の変化を分析するために、1975年から1985年のデータを用いて、アメニティの近接性が人口移動に及ぼす効果を推定した。最後に、国勢調査を利用して、移住者の属性別にアメニティ



注釈：平均気温の地域差とその階層内平均移住率をプロットした。横軸は、移住先と移住前地域の平均気温の差を表し、正であれば移住先の平均気温が高いことを意味する。各階層の幅は5℃とした。例えば、平均気温の地域差0は-2.5℃以上から2.5℃未満の階層を意味する。ただし、両端の-15及び15の階層は、それぞれ、-12.5未満及び12.5以上の階層を表す。また、平均移住率は各階層内に属するサンプルの移住率に関する平均を表す。

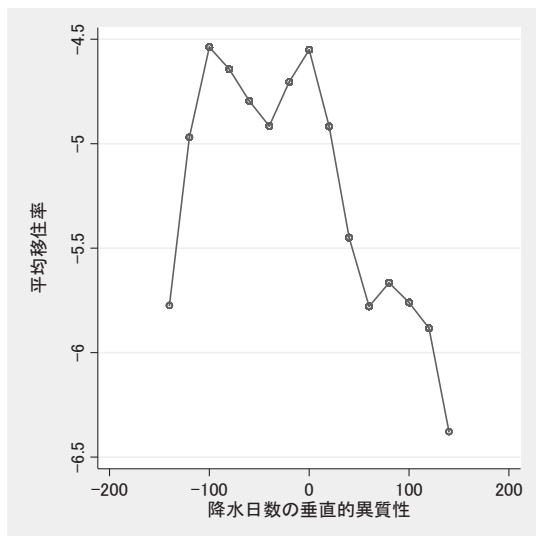
図3 移住率と平均気温垂直的異質性

の近接性効果を示す。

推定結果を示したそれぞれの表は、経済社会的要因、地域間距離及び社会的ネットワークを含めた、分析の基準となるモデルの結果と、それぞれのアメニティ変数を加えたときの推定結果から構成される。以下では、初めに、経済社会的要因の効果から示し、次いで地域間距離、社会的ネットワークの効果を述べる。最後にアメニティ要因の効果を示す。

4.1 水平的異質性の効果

本節では、本研究が着目する、アメニティの水平的性質が人口移動に及ぼす効果を見ていく。推定結果は表2の通りである。



注釈：降水日数の地域差とその階層内平均移住率をプロットした。横軸は、移住先と移住前地域の降水日数の差を表し、正であれば移住先の降水日数が多いことを意味する。各階層の幅は20日とした。例えば、平均気温の地域差0は、-10日以上から10日未満の階層を意味する。ただし、両端の-140及び140の階層は、それぞれ、-130未満及び130以上の階層である。また、平均移住率は各階層内に属するサンプルの移住率に関する平均を表す。

図4 移住率と降水日数垂直的異質性

4.1.1 経済社会的要因，地域間距離，社会的ネットワーク

はじめに、経済的要因が移住先決定の重要な要因であることが確認できる。県民所得の係数は、すべてのモデルにおいて正で有意な結果となっており、移住者はより豊かな地域へ移住する傾向を示している。推定した係数は0.0077程度であるので、その効果は、居住する地域より県民所得が10万円高い地域への移住率は0.8%ほど上昇することを意味する。また、有効求人倍率の係数は正で有意となっている。これは、移住者がより就業しやすい地域へ移住することを意味する。推定値をみると0.14程度となっているので、有効求人倍率が0.1高い地域への移住率が1.4%程度上昇することを示している。

一方、経済的要因には効果を持たない要因も

含まれている。事業所数の効果を見ると、それは統計的に有意となっていない。さらに、産業構造を表す第三次産業従業者比率も、係数は予想した通り正であるものの、統計的に有意ではない。

次に、社会構造の効果を見ると、統計的に有意な要因は大卒者比率と若年者比率であり、これらはともに負の係数を持つ。推定結果は、大卒者比率の1%の上昇が移住確率を2.3%程度減少させ、また、若年者比率の1%の上昇は移住確率を1.6%程度減少させることを示している。一方、大学数については、推定した係数は正であるが、統計的に有意ではない。

距離および社会的ネットワークについて、推定した係数はすべてのモデルで統計的に有意な結果となった。距離の係数は-0.3程度を示しており、これは、地域間距離が100km遠い地域への移住率が30%程度低下することを意味する。つまり、近隣の都道府県を選択する傾向を示した。また、過去の移住者数で測った、社会的ネットワークの係数は正である。従って、これは、移住者は社会的に結びつきが強い地域を選択することを示している。

4.1.2 アメニティの近接性

アメニティの近接性は、アメニティの地域間差に絶対値をとったものと定義している（第3章を参照）。そのため、これらの変数はアメニティに関する「距離」を表し、値が大きくなるほど地域間の近接性の低下を意味する。

まず、都市化の近接性は効果を示さなかった。推定結果から、自然公園を含めたモデルすべてについて、統計的に有意ではなかった。これは、自然公園によって表される地域の特性、すなわち都市化の類似性が移住先の選択に影響しないことを示している。ただし、この推定結果は、図2で見られる特徴と整合的ではない。つまり、図2では都市化の変化が大きいと移住率は増加する傾向にあるが、推定結果は都市化

表2 推定結果 アメニティ近接性, 1997年から2006年

	(1) アメニティ無し	(2) 都市化	(3) 平均気温	(4) 降水日数	(5) すべて
県民所得 (10万円)	0.0077*** (0.0018)	0.0077*** (0.0018)	0.0077*** (0.0018)	0.0077*** (0.0018)	0.0077*** (0.0018)
有効求人 倍率	0.1413*** (0.0121)	0.1413*** (0.0121)	0.1413*** (0.0121)	0.1413*** (0.0124)	0.1413*** (0.0123)
事業所数 (10万所)	-0.0411 (0.0254)	-0.0411 (0.0254)	-0.0411 (0.0259)	-0.0411 (0.0260)	-0.0411 (0.0264)
第三次産業 就業者比率	0.0018 (0.0016)	0.0018 (0.0016)	0.0018 (0.0016)	0.0018 (0.0016)	0.0018 (0.0016)
大卒者人口比率	-0.0225*** (0.0046)	-0.0225*** (0.0046)	-0.0225*** (0.0047)	-0.0225*** (0.0047)	-0.0225*** (0.0047)
大学数	0.0016 (0.0011)	0.0016 (0.0011)	0.0016 (0.0011)	0.0016 (0.0011)	0.0016 (0.0011)
若年者人口比率	-0.0155*** (0.0036)	-0.0155*** (0.0036)	-0.0155*** (0.0036)	-0.0155*** (0.0036)	-0.0155*** (0.0036)
地域間距離 (100km)	-0.3174*** (0.0046)	-0.3170*** (0.0046)	-0.2933*** (0.0066)	-0.3140*** (0.0047)	-0.2941*** (0.0066)
社会的ネットワーク (10万人)	0.0711** (0.0276)	0.0701** (0.0277)	0.0819*** (0.0298)	0.0698** (0.0275)	0.0782*** (0.0296)
都市化		-0.0024 (0.0021)			-0.0025 (0.0021)
平均気温			-0.0533*** (0.0100)		-0.0449*** (0.0105)
降水日数				-0.0026*** (0.0008)	-0.0020** (0.0008)
サンプルサイズ	19457	19457	19457	19457	19457
修正済み決定係数	0.988	0.988	0.988	0.988	0.988

括弧内は移動元と移動先の組み合わせ(2162通り)でクラスタリングした頑健な標準誤差。

すべてのモデルについて、年ダミー、移動前地域ダミー、移動先地域ダミーをコントロールして推定。

* p<0.1, ** p<0.05, *** p<0.01

表3 推定結果 アメニティ垂直的異質性, 1997年から2006年

	(1) アメニティ無し	(2) 都市化	(3) 平均気温	(4) 降水日数	(5) すべて
県民所得 (10万円)	0.0077*** (0.0018)	0.0075*** (0.0018)	0.0074*** (0.0018)	0.0077*** (0.0018)	0.0072*** (0.0018)
有効求人 倍率	0.1413*** (0.0121)	0.1446*** (0.0121)	0.1429*** (0.0121)	0.1411*** (0.0121)	0.1458*** (0.0121)
事業所数 (10万所)	-0.0411 (0.0254)	-0.0222 (0.0270)	-0.0438* (0.0253)	-0.0409 (0.0254)	-0.0234 (0.0269)
第三次産業 就業者比率	0.0018 (0.0016)	0.0021 (0.0016)	0.0019 (0.0016)	0.0018 (0.0016)	0.0022 (0.0016)
大卒者人口比率	-0.0225*** (0.0046)	-0.0191*** (0.0047)	-0.0219*** (0.0046)	-0.0225*** (0.0046)	-0.0181*** (0.0047)
大学数	0.0016 (0.0011)	0.0015 (0.0011)	0.0016 (0.0011)	0.0016 (0.0011)	0.0015 (0.0010)
若年者人口比率	-0.0155*** (0.0036)	-0.0171*** (0.0038)	-0.0154*** (0.0036)	-0.0155*** (0.0036)	-0.0171*** (0.0038)
地域間距離 (100km)	-0.3174*** (0.0046)	-0.3174*** (0.0046)	-0.3174*** (0.0046)	-0.3174*** (0.0046)	-0.3174*** (0.0046)
社会的ネットワーク (10万人)	0.0711** (0.0276)	0.0711** (0.0276)	0.0711** (0.0276)	0.0711** (0.0276)	0.0711** (0.0276)
都市化		0.0093** (0.0036)			0.0099*** (0.0036)
平均気温			-0.0136*** (0.0046)		-0.0146*** (0.0048)
降水日数				-0.0000 (0.0001)	-0.0001 (0.0001)
サンプルサイズ	19457	19457	19457	19457	19457
修正済み決定係数	0.988	0.988	0.988	0.988	0.988

括弧内は移動元と移動先の組み合わせ(2162通り)でクラスタリングした頑健な標準誤差。
すべてのモデルについて、年ダミー、移動前地域ダミー、移動先地域ダミーをコントロールして推定。

* p<0.1, ** p<0.05, *** p<0.01

の変化が影響しないことを示している。この点については第5章で考察する。

次に、平均気温については、すべてのモデルにおいて負の符号を持ち、1%水準で統計的に有意である。これは、移動先の平均気温が現在居住している地域と大きく異なると、移住率が低下することを意味する。推定値は-0.053から-0.045程度であり、従って、これは、平均気温が1℃異なる地域への移住は5%程度低下することを示している。

最後に、降水日数は、気温と同様に、降水日数の水平性低下が移住率を低下させることを表す結果となった。推定結果は、すべてのモデルで負で統計的に有意となっている。推定した係数は-0.0026から-0.0020であるので、これは、降水日数が10日異なると、移住率を2%程度低下させることを示している。

4.2 垂直的異質性の効果

ここでは、アメニティの垂直的異質性が人口移動に及ぼす効果を示す。本研究の目的はアメニティ近接性の分析にあるが、すべてのアメニティが近接性効果を持つかどうか明らかではない。従って、各アメニティがどちらの性質を持つのか、あるいは持たないのかが調べる必要がある。そのために、ここでは、アメニティの垂直的異質性が人口移動に及ぼす効果を推定した。そのため、前節とは異なり、各アメニティは地域間の差 ($A_{j,t} - A_{i,t}$) として定義されていることに注意が必要である。推定結果を表3に示す。

まず、経済社会的要因、地域間距離および社会的ネットワークの効果の推定結果を見ると、表2と比較して大きな違いは当然見られない。なぜなら、これらは、第2章で見たように、垂直的異質性のみを考慮しているためである。ただし、平均気温の水平的効果を推定した結果(表2モデル(3))と違って、事業所数に関しては、平均気温の垂直的性質をモデルに含めた

場合、10%水準で統計的に有意となっている。

アメニティの垂直的異質性効果に関する推定結果は以下の通りである。はじめに、都市化の効果に関しては正の係数が統計的に有意に得られている。つまり、これは、都市化されていない自然が豊かな地域ほど、移住者を惹きつけていることを示している。ただし、移住率と都市化の関係は、図2より、U字型であった。つまり、都市化が上昇しても、低下しても移住率は増加する関係を示した。この結果については第5章で考察する。

次に、気温の係数は負で有意となっており、気温の低い地域を選択することを示す結果となった。先行研究では、温暖な気候が人口流入を促す要因の一つであった(Rodriguez-Pose and Ketterer (2012))。一方、本研究の推定結果は先行研究とは逆の結果である。この結果に対して、気温に関する人々の選好は逆U字型を持っており、加えて、日本各地の平均気温は最適な気温よりも高いなどの可能性が考えられる。しかし、平均気温の結果に関して、近接性と垂直的異質性の両方で統計的に有意となっていることには注意が必要である。これについては、第5章で考察する。

最後に、降水日数の垂直的異質性が推定結果から移住先の選択に影響しているとは言えない。表4より、降水日数の垂直的異質性効果は統計的に有意ではなく、さらに係数の大きさも非常に小さい。従って、降水日数は垂直的異質性効果を持たないと結論付けられる。

4.3 異時点のアメニティ近接性効果

本節では、アメニティの近接性に関して、本章で時間を通じた効果の変化を調べるために、データセットを1975年から1985年の期間に変更して推定した結果を示す。これは、アメニティの効果は経済発展に応じて変化する可能性を考慮するためである。推定の方法、つまり、推定式および近接性の定義はこれまでと同様で

ある。推定結果を表4に示す。

4.3.1 経済社会的要因，地域間距離，社会的ネットワーク

この時代において各要因の人口移動に及ぼす効果は，近年の効果（表2）と類似した結果を示した。まず，経済社会的要因は，事業所数を除いて，定性的には同じ傾向を示したものの，推定値には時代間で差が見られる。また，事業所の推定した係数は，負で有意となっており，これは事業所の多い地域ほど移住確率が低下することを意味する。また，距離や社会的ネットワークの効果についても，ほぼ同様な結果となった。地域間距離の広がり移住確率を低下させ，社会的ネットワークが強いほど移住確率を上昇させる。

4.3.2 近代におけるアメニティの近接性が与える影響

アメニティについても，他の要因と同様に，時代が変化しても定性的に同じ結果を示した。まず，都市化の近接性は統計的に有意ではない。従って，都市化に関しては，近接性の効果を持たないと言える。

次に，気候に関しては，近年における水平的性質と同様な結果を示した。まず，平均気温の効果は，負の符号で統計的に有意となっている。従って，この結果は，移住先の決定において現在の居住地と平均気温が近い地域を選択する傾向を示している。また，降水日数についても，負の符号で統計的に有意な推定結果となっている。近年における効果と同様に，降水日数の異なる地域への移住率が低下する結果となった。

4.4 移住者の属性

本節では，国勢調査を用いて移住者の属性別にアメニティの近接性効果を推定した。まず，移住者を労働者と非労働者，また年齢に応じて

若年，中年，老年の3つに分類した。若年，中年，老年はそれぞれ，20代と30代，40代と50代，60代と70代と定義した。また，これまでと同様に，同時決定による内生性を回避するため，移住者数には2000年のデータを，説明変数には1995年のデータを用いて，従属変数と説明変数との間にラグを取った。推定結果を表5に示す。

本節のデータは二つの点で注意が必要である。まず，クロスセクションデータであるため，年ダミーは推定に含めていない。このため，推定にバイアスが生じる可能性がある。

二つ目は，時間の単位である。本節の移住率は1995年，2000年のそれぞれの時点での住所に基づいて計算された移住者数（都道府県間移動，都道府県内移動の両方）から算出されている。しかし，例えば，1995年に大阪にいた人が96年に東京へ移住し，さらにもう一度移住して2000年には神奈川に居住していた場合，大阪から神奈川への移住者として計算される。このとき，大阪から東京，東京から神奈川への移住は移住者数に含まれない。したがって，移住者データには測定誤差が生じる可能性があり，分析には注意が必要である。また，移住者の属性は2000年時点の状態を表しており，1995年から2000年にかけて属性が変わった場合を考慮できない。

アメニティ変数の中で，降水日数のみが統計的に有意な結果となった。降水日数の推定値をみると，表2と表4と同程度となっているので，頑健な推定結果と言える。一方，気温に関して，係数は負の符号であるがすべての属性で統計的に有意ではなかった。また，都市化近接性は，中年世代の人口移動に対してのみ統計的に有意な結果となり，移住を妨げる要因となっている。

表4 推定結果 アメニティ近接性, 1975年から1985年

	(1) アメニティ無し	(2) 都市化	(3) 平均気温	(4) 降水日数	(5) すべて
県民所得 (10万円)	0.0095*** (0.0025)	0.0095*** (0.0025)	0.0095*** (0.0025)	0.0095*** (0.0025)	0.0095*** (0.0025)
有効求人 倍率	0.0924*** (0.0135)	0.0924*** (0.0135)	0.0924*** (0.0135)	0.0924*** (0.0134)	0.0924*** (0.0134)
事業所数 (10万所)	-0.1025*** (0.0180)	-0.1025*** (0.0180)	-0.1025*** (0.0181)	-0.1025*** (0.0178)	-0.1025*** (0.0179)
第三次産業 就業者比率	-0.0005 (0.0040)	-0.0005 (0.0040)	-0.0005 (0.0040)	-0.0005 (0.0039)	-0.0005 (0.0039)
大卒者人口比率	0.1073*** (0.0124)	0.1073*** (0.0124)	0.1073*** (0.0124)	0.1073*** (0.0122)	0.1073*** (0.0122)
大学数	-0.0001 (0.0022)	-0.0001 (0.0022)	-0.0001 (0.0022)	-0.0001 (0.0022)	-0.0001 (0.0022)
若年者人口比率	-0.0112*** (0.0025)	-0.0112*** (0.0025)	-0.0112*** (0.0025)	-0.0112*** (0.0025)	-0.0112*** (0.0024)
地域間距離 (100km)	-0.3173*** (0.0066)	-0.3173*** (0.0066)	-0.2982*** (0.0073)	-0.3158*** (0.0066)	-0.2989*** (0.0073)
社会的ネットワーク (10万人)	0.0838** (0.0340)	0.0837** (0.0340)	0.0905*** (0.0351)	0.0803** (0.0335)	0.0863** (0.0346)
都市化		-0.0004 (0.0022)			-0.0008 (0.0022)
平均気温			-0.0581*** (0.0144)		-0.0515*** (0.0145)
降水日数				-0.0031*** (0.0008)	-0.0028*** (0.0008)
サンプルサイズ	21620	21620	21620	21620	21620
修正済み決定係数	0.985	0.985	0.985	0.985	0.985

括弧内は移動元と移動先の組み合わせ(2162通り)でクラスタリングした頑健な標準誤差。

すべてのモデルについて、年ダミー、移動前地域ダミー、移動先地域ダミーをコントロールして推定。

* p<0.1, ** p<0.05, *** p<0.01

表5 2000年の移住者の属性別、アメニティ近接性

	(1) 労働者	(2) 非労働者	(3) 若年	(4) 中年	(5) 老年
県民所得 (10万円)	0.2182*** (0.0415)	0.2134*** (0.0409)	0.1746*** (0.0412)	0.2866*** (0.0418)	0.0084 (0.0305)
有効求人 倍率	4.9064*** (0.2120)	4.3220*** (0.2016)	4.7226*** (0.2090)	4.6393*** (0.2076)	6.9325*** (0.2600)
事業所数 (10万所)	1.1303*** (0.0517)	1.1679*** (0.0516)	1.1089*** (0.0505)	1.3196*** (0.0506)	1.6426*** (0.0757)
第三次産業 就業者比率	0.3978*** (0.0201)	0.3827*** (0.0199)	0.3688*** (0.0200)	0.4356*** (0.0209)	0.2858*** (0.0163)
大卒者人口比率	-0.0153 (0.0238)	-0.0135 (0.0236)	0.0070 (0.0236)	-0.0423* (0.0238)	0.1579*** (0.0219)
大学数	-0.1340*** (0.0066)	-0.1334*** (0.0065)	-0.1239*** (0.0065)	-0.1621*** (0.0062)	-0.1467*** (0.0056)
若年者人口比率	-0.1389** (0.0652)	-0.2028*** (0.0659)	-0.0683 (0.0654)	-0.3270*** (0.0634)	0.1034* (0.0548)
地域間距離 (100km)	-0.3074*** (0.0068)	-0.3231*** (0.0067)	-0.3174*** (0.0068)	-0.2893*** (0.0065)	-0.2913*** (0.0081)
社会的ネットワーク (10万人)	0.0824*** (0.0313)	0.0702** (0.0300)	0.0764** (0.0307)	0.0798*** (0.0275)	0.1102*** (0.0382)
都市化	-0.0035 (0.0021)	-0.0028 (0.0021)	-0.0025 (0.0021)	-0.0056*** (0.0021)	-0.0026 (0.0027)
平均気温	-0.0190 (0.0155)	-0.0139 (0.0154)	-0.0166 (0.0154)	-0.0158 (0.0151)	-0.0070 (0.0167)
降水日数	-0.0028*** (0.0008)	-0.0028*** (0.0008)	-0.0026*** (0.0008)	-0.0030*** (0.0007)	-0.0040*** (0.0009)
サンプルサイズ	2162	2162	2162	2162	2130
修正済み決定係数	0.991	0.991	0.991	0.992	0.990

括弧内は頑健な標準誤差。

すべてのモデルについて、移動前地域ダミー、移動先地域ダミーをコントロールして推定。

* p<0.1, ** p<0.05, *** p<0.01

5 考察

本章では、4章の結果を考察する。ここでは、アメニティのみに着目して、近接性と垂直的異質性、異時点間、さらに移住者の属性間で比較する。

5.1 近接性と垂直的異質性

推定結果の比較から、アメニティには性質の違いがあることが見て取れる。アメニティが持つ特徴としては、垂直的異質性効果、近接性効果、そしてその両方をもつ、三つの可能性がある。ここでは、各アメニティがいずれの性質にあるか検討する。

はじめに、都市化は垂直的異質性が効果をもつアメニティである。近接性と垂直的異質性の効果を比較すると、都市化は、垂直的異質性の効果に関しては統計的に有意である一方、近接性についてはそうではない。つまり、都市化が上昇するほど、移住確率が低下することを示している。従って、移住者は、都市化されていない、豊かな自然を有する地域を選んで移住する。

ここで、図2で示された、移住率と都市化におけるU字型の関係は、見せかけといえる。図2から、都市化の変化が大きいほど、移住率が高いように見える。しかし、データを観察すると、U字の両端は、“滋賀-京都”間のサンプルしか含んでいない。そのため、都市化の変化による移住率への影響を反映しているのではなく、地域間距離の近さによる移住率の高さを示している可能性があった。もし、U字型の関係が正しいとすると、近接性の低下、すなわち、都市化の変化が移住率を上昇させるはずである（つまり、都市化水平的異質性の係数は正で有意を示す）。しかし、推定結果（表2）はこれを支持していない。従って、U字の両端の移住率の高さは、距離の近さを反映した結果である。

次に、平均気温は、両方の性質を備えている。推定結果は、近接性効果、垂直的異質性効

果ともに統計的に有意を示している。従って、これは、近接性が低下するほど移住確率は低下する一方で、垂直的異質性効果の係数は負なので、より気温の低い地域を好んで移住していることを示している。この結果に対する解釈の一つは、より気温の低い地域を好むのは、洋服などの防寒機能が向上し、暑さ対策よりもそれが容易であるためかもしれない。また、平均気温が異なることが季節ごとの気温の違いを表しており、その変化に対して負担を感じていることが移住確率を下げていると考えられる。

最後に、降水日数については近接性効果をもつアメニティと言える。推定結果より、垂直的異質性の効果は統計的に有意ではなく、近接性のみが負で有意である。この結果は、移住によって降水日数が変化すると、環境変化により最適な生活パターンが変化し、この変化をコストとして捉えていると考えられる。

5.2 近接性の異時点間比較

本節では、二つの期間の推定結果を比較し、アメニティの近接性効果が、時間を通じて変化するかどうかを考察する。推定結果は、両期間で似ている。しかし、効果の大きさには差が見られ、時間を通じて外部環境に変化が起きたことがその原因と考えられる。

まず、都市化は、これまで同様に、近接性効果を持たないアメニティである。なぜなら、二つの期間の推定結果（表2と表4）は、統計的に有意ではないからである。つまり、時間を通じて外部環境が変化しても都市化アメニティの効果に変化は見られない。

一方、気候に関しては、時間を通じてその効果は弱まっている。両期間に共通して、平均気温と降水日数のそれぞれの近接性効果は負で統計的に有意である。つまり、たとえ時代が異なっても、近接性の低下は移住率を下げている。しかし、近年（97-06年）の効果は、より弱まっている。これは、技術進歩によって、

気候変化の負担が低下したためと考えられる。例えば、エアコンや空気清浄機などによって、室内に関しては場所を選ばずに調整が可能となった。このため、近年では環境変化の効果が弱まったと考えられる。

5.3 移住者の属性間比較

本節では、アメニティの近接性効果を、移住者の属性間で比較する。ここでは、データの利用可能性のため、労働状態間、そして年齢間で比較した。推定結果からは、移住者の属性により、効果に差がみられるアメニティがあることが確認できる。労働状態では推定結果に違いは見られない一方、年齢間比較では、年齢の上昇に応じて推定値に違いが見られた。

まず、労働者と非労働者間で大きな違いは見られず、アメニティの効果は労働状態、つまり、社会的状況に依存しない性質と考えられる。各アメニティについて結果を比較すると、統計的有意性や定性的な違いは見られない。つまり、降水日数に関しては社会的状態に依存しないような、非金銭的要因が影響しているものと考えられる。

一方、年齢別推定結果を見ると、大きな違いが見て取れる。まず、降水日数の水平性は、年齢が上がると、移住確率を下げる効果が強くなっている。これは、加齢によって適応能力が低下し、新たな環境をより大きく負担に感じるためと考えられる。

ただし、都市化と降水日数の近接性はこれまでの推定結果と同様な傾向を示している一方、平均気温に関しては結果が全く異なる。属性を考慮した推定結果(表4)では、すべての属性で統計的に有意ではないが、属性を考慮しない結果(表2)は統計的に有意である。この結果の違いの原因については明らかではないが、第4章で述べたように、5年間を通じた移動データであるために計測誤差が生じ、それが原因かもしれない。

5.4 問題点

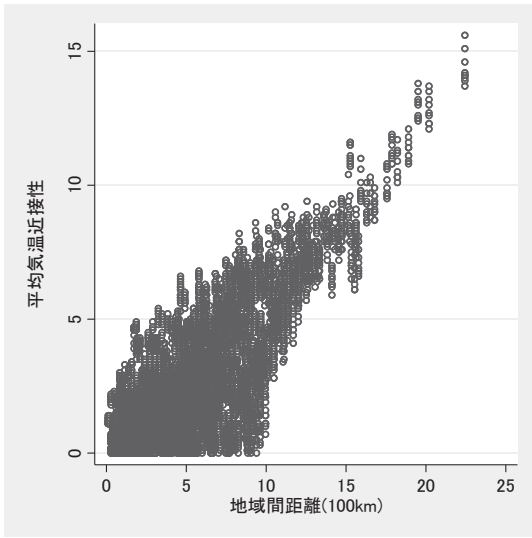
ここでは、いくつかの問題点を考察する。問題の一つ目は、近接性の定義あるいはその測定方法にある。本研究で用いた近接性の場合、垂直的異質性を含んだ定義となっている。そのため、水平性と垂直性の両方で統計的に有意であった場合、近接性による効果かどうかを識別することは困難である。

二つ目は、それぞれのアメニティ近接性と地域間距離との相関である。そこで、地域間距離と各アメニティ近接性を図5、6、7にプロットした。まず、図5に地域間距離と平均気温水平性をプロットすると、気温に関しては距離との間に高い相関があるように見える。これは、日本列島が南北に伸びているために、距離と平均気温が似た動きをするためと考えられる。したがって、平均気温については、その解釈をするとき、特に、多重共線性に注意して結論付ける必要がある。

一方、都市化と降水日数については、明らかな相関は見られない。図6に地域間距離と都市化の近接性をプロットすると、明らかな相関は見られない。これは、都市化には、地域間距離とは違った、別のメカニズムが働いているためと考えられる。また、図7に地域間距離と降水日数近接性をプロットすると、都市化と同様に、明らかな相関は見られない。したがって、これら二つの近接性については、距離との相関の問題は小さいと考えられる。

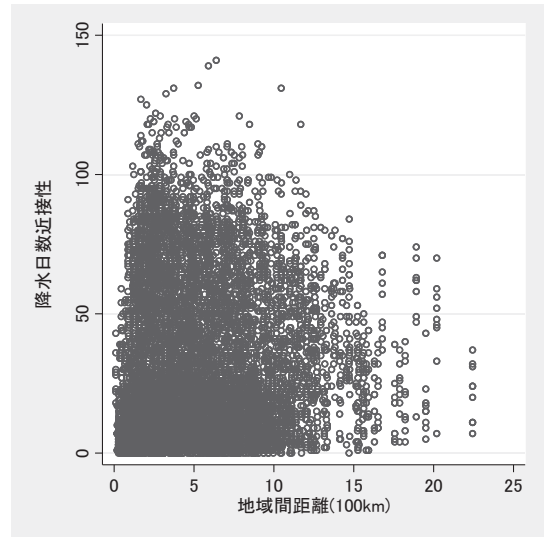
6 結論

本研究は、都道府県データを用いて、地域間のアメニティの近接性が人口移動パターンに及ぼす効果を分析した。つまり、アメニティの変化を居住環境の変化と捉え、これが移住コストとなる可能性を検証した。通常、人口移動パターンは、移住先の経済社会的要因や社会的ネットワークから影響を受けると考えられる。



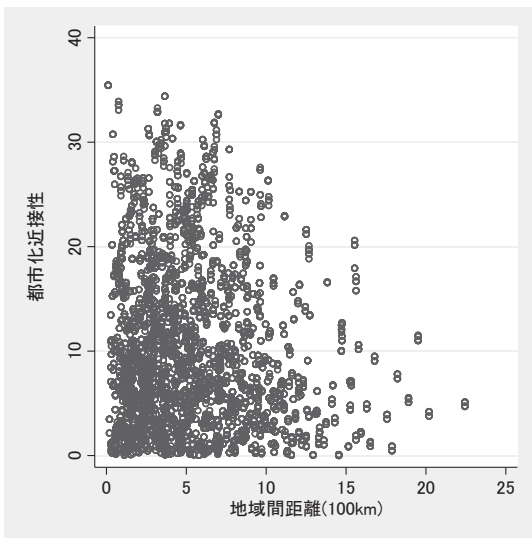
注釈; データ期間は1997年から2005年の9年間。

図5 平均気温近接性と地域間距離



注釈; データ期間は1997年から2005年の9年間。

図7 降水日数近接性と地域間距離



注釈; データ期間は1997年から2005年の9年間。

図6 都市化近接性と地域間距離

研究では、代表的なアメニティについて、近接性効果と垂直的異質性効果を推定し、結果を比較した。本研究は、地域間近接性が低下するほど、移住確率を低下させるアメニティがあることを発見した。具体的には、気候の近接性に大きな隔たりがある地域間ほど、環境変化が大きいため、その地域間での移住は低下することを示した。

先行研究では、アメニティの垂直的異質性が人口移動に与える影響を分析してきたが、本研究では地域間近接性に焦点を当てた。アメニティの垂直的異質性と近接性の両方の性質を分析した結果、垂直的異質性が人口移動パターンに影響するアメニティがある一方で、近接性が影響を及ぼすアメニティがあることが示された。従って、アメニティの効果を分析するには、垂直的異質性に加えて、地域間近接性も調べる必要がある。

政策的インプリケーションとして、移住を促すような補助金を与える場合、移住前の居住地に応じた金額を設定することが考えられる。遠く離れた地域からの移住者は、気候変化による

また、これらの要因に加え、移住者は移住によるコストを考慮して移住先を決定する。そのコストには、移住による生活環境の変化に対するものも含まれている可能性がある。このためのコストは、移動前地域と移動先地域の環境が似通っているほど、低下するであろう。そこで本

身体的、心理的費用を負担する必要がある。ある地域が他の遠い地域からの移住者を必要とすると、これらの費用に応じた補助金を与えることで移住を促すことができる可能性がある。つまり、補助金額の設定に対して、移住元地域と移住先地域の組み合わせに応じた設計が必要と考えられる。

しかし、いくつかの問題点については、今後の課題である。まず、近接性の測定方法には、垂直的異質性とは独立になるような指標が必要である。なぜなら、本研究で用いた近接性では、垂直的異質性の効果と近接性効果を識別することが困難な状況が生まれる。つまり、今回使用した近接性の定義では垂直的異質性の情報が含まれるため、近接性と垂直的異質性の両方で統計的に有意な結果が出た場合、その識別は困難となる。また、同様の理由で多重共線性が発生するため、推定式に近接性と垂直的異質性の両方を同時に推定することができない。

また、集計データのため、移住者に特有な要因について考慮できない。例えば、移住者の移動履歴は、移住の決定に影響すると考えられるが、これを考慮していない。つまり個人の移動履歴や、出身地など、個人特有の要因が移住先の決定に影響すると予想される。これにより、欠落変数バイアスによる誤った結果をもたらさうるので、より詳細な移住者データを用いて検証すべきと考える。

さらに、本研究で分析しなかったが、文化的要因の人口移動への効果を検証すべきと考えられる。文化的な近接性は、生活パターンの変化について、負担を下げると思えるので、この点は重要になる可能性がある。例えば、江戸時代など歴史において交通の便が良かったことや、街道でつながっていることが地域間交流に影響して、地域間で文化が似る可能性がある。そのような文化的な近接性に基づいた移住行動も十分考えられるが、行政区分と文化の地理的区分が完全には一致していないため、今回の分

析には取り入れなかった。今後、良い指標があれば分析に組み込んでいきたい。

参考文献

- 伊藤薫 (2003) 「バブル経済期の男女・年齢別人口移動 —1990年国勢調査人口移動集計結果を利用して—」『地域学研究』33巻, 3号, 85-102頁.
- 黒田達朗・田淵隆俊・中村良平 (2008) 『都市と地域の経済学(新版)』有斐閣, 第11章, 227-228頁.
- フロリダ, リチャード (2014) 井口典夫訳『新クリエイティブ資本論』ダイヤモンド社, 第10章.
- 厚生労働省編 (2005) 『平成17年版 労働経済の分析』, 第2部 第1章 人口減少社会における労働政策の課題, 108-109頁, www.mhlw.go.jp/wp/hakusyo/roudou/05/dl/02-02c.pdf.
- Cameron, A. Colin and Pravin K. Trivedi (2010) *Microeconometrics Using Stata*, Revised edition, College Station TX: Stata Press.
- Carrington, William J., Detragiache, Enrica and Vishwanath, Tara (1996) "Migration with Endogenous Moving Cost," *American Economic Review*, 86(4), 909-930.
- Cheire, Paul and Stefano Magrini (2006) "Population Growth in European Cities; Weather Matters — But Only Nationally," *Regional Studies*, 40(1), 23-37.
- Graves, Philip E. (1979) "A Life-Cycle Empirical Analysis of Migration and Climate, by race," *Journal of Urban Economics*, 6, 135-147.
- Greenwood, Michael J. and Hunt, Gary L. (1989) "Jobs versus Amenities in the Analysis of Metropolitan Migration," *Journal of Urban Economics*, 25, 1-16.
- Lucas, Robert E. B. (2001) "The effects of proximity and transportation on developing country

- population migrations,” *Journal of Economic Geography*, 1(3), 323-339.
- Nakajima, Kentaro and Tabuchi, Takatoshi (2011) “Estimating Interregional Utility Differentials,” *Journal of Regional Science*, 51(1), 31-46.
- Ozgen, Ceren, Peter Nijkamp, and Jacques Poot (2011) “Immigration and Innovation in European Regions,” *IZA Discussion paper*, 5676, Institute for the Study of Labor (IZA).
- Partridge, Mark D. and Rickman, Dan S. (2003) “The waxing and waning of regional economies; the chicken-egg question of jobs versus people,” *Journal of Urban Economics*, 53, 79-97.
- Rodriguez-Pose, Andrés and Ketterer D. Tobias (2012) “Do Local Amenities Affect the Appeal of Regions in Europe for Migrants?” *Journal of Regional Science*, 52(4), 535-561.
- Schultz, T. Paul (1982) “Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model,” *Economic Development and Cultural Change*, 30(3), Third World Migration and Urbanization: A Symposium, 559-593.
- Schwartz, Aba (1973) “Interpreting the Effect of Distance on Migration,” *Journal of Political Economy*, 81(5), 1153-1169.
- Sjaastad, Larry A. (1962) “The Costs and Returns of Human Migration,” *Journal of Political Economy*, 70(5), Part 2: Investment in Human Beings, 80-93.
- Treyz, George I., Dan Rickman, Gary L. Hunt and Michael J. Greenwood (1993) “The Dynamics of U.S. internal Migration,” *Review of Economics and Statistics*, 75(2), 209-214.

Proximity of Local Amenities as a Factor in Migration

Masaaki Toma

This study uses prefectural-level data to analyze the impact of amenities on migration. In particular, we examine whether the local proximity of amenities reduces moving costs. Changes in living conditions resulting from migration are non-monetary costs that can affect the choice of residential area. Estimations using prefectural-level data on migration revealed two aspects wherein amenities affect the cost of moving. The first aspect, as indicated in prior studies, is a vertical heterogeneity wherein an area's utility increases with the availability of amenities. The second aspect is the local proximity, whereby moving costs are lower to the extent that the amenities in the origin and destination are similar. Consequently, any analysis of the utility of amenities requires assessment of their local proximity, as well as their vertical heterogeneity in each area.

JEL Classification: R12, R23

Keywords: amenity proximity, population movement, aggregate logit model

株式市場を反映するセンチメント・インデックスの構築と 株価説明力の実証分析*

數見 拓朗[†]

要 約

本論文では、日本経済新聞（日経）から、日本の経済、特に日本の株式市場を反映するセンチメントを計量化する指数（インデックス）について提案する。具体的には、2つのセンチメント指数（SI: sentiment index）を構築し、その日本株価への説明力を実証する。市場における参加者の心理や雰囲気を表すセンチメントに関する分析は、近年、学術研究のみならず、あらゆる産業実務においても、ますます多くの関心を集めている。センチメント分析には多くのアプローチが考えられるが、本研究は、Ishijima, et al. (2014) を議論の起点とする。彼らは、日経に掲載されているすべての記事を利用して、ポジティブあるいはネガティブな心理を、SIとして計量化する一手法を提案した。さらに、過去5年間の日次データを分析対象として、そのSIが3日後の日本株価を有意に予測し得ることを実証した。本研究の目的は、Ishijima, et al. (2014) の拡張として、日本の雰囲気全体ではなく、日本の経済活動のセンチメントに焦点を当てたSIを提案する。具体的には、日本の経済活動について言及している記事に限定してSIを作成する。本研究で得られた結論は以下の3点に集約される。(1) 本研究で提案するSIは、先行研究の結果とは異なり、翌営業日の対数収益率と出来高に有意な説明力を有する。(2) 提案するSIは、株価を説明する上で、リバウンドが観測されない。(3) SIは、株式市場に対して後追いで反応する、ということである。

JEL Classification : C88, E37, G17

キーワード：センチメント分析、日本経済新聞、テキスト・マイニング、株価予測可能性

1 はじめに

近年、学术界と産業界の両方でセンチメント分析が注目されている。センチメントとは、景気全般や社会心理を漠然と表すもので、これを分析することは、経済やマーケットをより深く

理解する助けとなる。

資本市場・証券市場などを念頭においたセンチメントを、市場センチメントと呼ぶことにする。その背景となる議論を以下に概観してみよう。近年の市場センチメントについての関心は、市場の合理性の仮定についての賛否を起点にしている。効率性市場仮説によれば、情報は市場全体に極めて効率的に平等に広がり、それゆえ、リスクを勘案して合理的に予想される投資リターン以上の超過リターンを、何人も得ることができないとされる (Fama, 1965, 1991)。

* 本論文を執筆するにあたり、指導教員である大阪大学大西匡光教授に感謝致します。また、多くの知識や示唆を頂いた中央大学石島博教授にも感謝致します。

[†] 大阪大学大学院経済学研究科博士後期課程、株式会社サイバーエージェント

これまで多くの実証研究がこの仮説を支持する結果を示してきたが、その一方で、近年、これを否定する研究も増えている。こうした研究の多くは、超過リターンの存在（市場アノマリー）を立証し、それを通して証券価格の予測可能性を立証するものとなっている。

一方、市場アノマリーの実証研究を支える理論的根拠についても、いくつかの理論が提示されている。特に、Kahneman and Tversky (1979) に始まるとされる行動経済学は、経済主体の行動を、その心理的側面に力点をおいて、解釈するものとなっている。その脈絡で、行動ファイナンスについて Ritter (2003) は文献を整理し、認知心理学と市場裁定機会の可能性を中心にまとめている。それによると、典型的な分析結果は、市場参加者は必ずしも合理的に取引を行うわけではない、というものである。むしろ、彼らの中で広がっている心理状態に従い、非合理的に取引を行う。そして、そうした取引が市場に超過リターンの機会を生じさせる。そのほかにも、経済主体は当該の経済活動に関係の無い情報に影響されて取引を行う、といった理論も提案されている。社会の雰囲気、世論、社会トレンドなどと言った株価に直接関連しそうな情報が株価を動かすと考えられる。「センチメント」という言葉は、そうした情報に対する認識全般を表すものであるが、これまでは、実体の無いものとして顧みられることも少なかった。

センチメント分析は、これまではっきりとした実体として捉えられることが少なかったもの、主に文章に現れる心理などを数値として定義し、計量しようとする近年の試みである。これはテキスト・マイニング技術の発達によるところが大きい。合理的市場仮説に対する疑問、行動経済学と関連する新しい理論を背景として、こうしたセンチメント分析が社会経済分析の分野で注目され始めたと言える。

近年急速に増えているセンチメント分析の文献のなかで、株式市場に関連しているものと

して特記に値するものがいくつか挙げられる。Tetlock (2007) はマスメディアと株式市場との相関を検証した。その延長線上でさらに Tetlock et al. (2008) は言語を計量化したものが個々の企業の会計利益と株式リターンを予測しえるか否か分析を行った。Bollen et al. (2011) は、Twitter に現れる文章についてインデックスを作成することを考案し、センチメント・インデックスとしていくつかのタイプを提示した。Boudoukh et al. (2012) は、新聞などのニュースが株価の動きをどのようにリードするか検証した。

日本語のデータを使った研究例も数多く存在する。例えば、沖本 et al. (2014) や五島 (2016) は、Tetlock (2007) で提唱された仮説が日本の株式市場でも成り立つかどうかを検証している。彼らの分析では、日経 QUICK やロイターニュースのデータから作成したニュース指標が、翌日のリターンや出来高に対して有意な説明力を持つことが示されている。また、インターネット上のデータを利用して、株価を予測する研究例として、坪内 et al. (2015) が挙げられる。彼らの研究では、感情辞書を準備して、Yahoo! ファイナンスの株価掲示板からインデックスを作成し、株価の予測性能を検証している。新聞記事を利用した研究として、Ishijima et al. (2014) は、毎日の新聞ニュースに現れる日本経済のセンチメントを計量的に分析した。彼らは、日々の経済状況を肯定的あるいは否定的に説明する単語の出現頻度をカウントし、指数（インデックス）とした。新聞ニュースとしては日本経済新聞（以下、日経）が使われている。その上で、彼らはそうした指数をセンチメント・インデックス（以下、SI）とよび、株価指標である日経 225 との相互関係を統計分析した。興味深いことに、彼らの結論は、日経に掲載された全ての記事から作成した SI が 3 日先の株価を予測しえるとのことであった。

以上のように、テキスト情報からセンチメントを抽出し、株式市場を予測する研究は多く存

在するが、株式市場を説明するために、どのようなソースからセンチメントを作成すべきかについて、統一的な見解はない。例えば、Twitterのデータを利用したBollen et al. (2012) や、全ての日経記事を利用したIshijima et al. (2014) のように、経済活動に関するニュースを対象にしていない記事を含めてSIを構築する研究事例がある。彼らの研究は、特に世の中の「全体」の雰囲気に着目をして、株式市場の関係を明らかにする研究と言えよう。一方で、沖本 et al. (2014) や五島 et al. (2016) のように、マーケットニュースなど株式市場に影響を与える可能性のあるニュースからSIを構築する研究も多い。特に、沖本 et al. (2014) では、どのような種類のニュースが株価に対して大きな影響を有するかを実証している。彼らの分析では、コーポレートアクションとマーケットコメントに関する記事が株式リターンに有意な説明力を有していることを示している。

以上を踏まえて、本研究の目的は、2007年1月1日から2014年12月31日までの日経の記事を分析対象とし、そのSIを計量化するにあたっては、市場に関連すると思われる記事に限定して作成したSIと、記事を限定せずに全ての記事を用いて作成したSIが、株価の説明力という点において差異があるかどうかを実証することである。具体的には、以下の手順を踏んで実証を行う。第一に、日経記事の内容ごとに分類を行う。日経記事の元データには、どのような内容が書かれているかについて分類がなされていないため、トピックモデルを利用して各記事の分類を行い、経済活動に関連したニュースを特定する。第二に、経済活動に関連したニュースに分類された記事から作成するSI、株式収益率、出来高からなる3変量VARモデルと、全ての記事を利用するSI、株式収益率、出来高からなる3変量VARモデルの、2つのVARモデルを構築する。第三に、VARモデルの推定結果の符号条件を確認する。Tetlock

(2007) で指摘された、ニュースの株価への影響に関する3つの理論（情報理論、センチメント理論、無情報理論）のうち、どの理論が成立する可能性が高いかを調査する。第四に、統計的なモデルの当てはまり度合いを確認する。以上の手順を踏むことで、本研究で提案するSIと、Ishijima et al. (2014) のSIのどちらが、株式市場をより反映するSIであるかを検証する。Tetlock (2007) の3つの理論に関しては、以下で説明する。

第一に、情報理論とは、SIの株価に対する影響は、恒久的に消滅しないという理論である。情報理論が成立する場合、SIの株価に与える影響の方向は、一貫して同じ方向を向いているはずである。第二に、センチメント理論とは、SIは短期的に株価に影響を与えるが、その影響は長期的には消滅するという理論である。第三に、無情報理論とは、SIは株価に対して影響を全く与えないという理論である。

本研究の貢献として、以下の3点が挙げられる。第一に、経済活動に関する記事を抽出する際の、記事の分類方法の提案である。日本の株式市場を説明しうるSIの構築に関して、LDAが有効であることを示す。第二に、提案するSIは、翌営業日の対数収益率と出来高を有意に説明し、株式市場に対して、後追いで反応するということである。第三に、提案するSIの株価への推定係数が、一貫して同じ方向であることを示す。これはTetlock (2007) とは異なる結果である。

本研究は次のように構成される。第2節では、分析対象とする日経記事について説明する。第3節では、それぞれの日経記事を書かれている内容をもとにクラスタリングを行い、経済活動に関連したニュースの記事を特定する。第4節では、SIの構築方法について述べる。第5節では、株式市場との相関分析のためのモデルを記す。第6節でまとめとする。

2 日経記事の分析

本節では、SIを構築する上で、対象となるテキスト情報がどのようなニュースソースなのかを理解するために、日経記事について概観する。まず、データの前処理について説明する。次に、日ごとに現れる特徴語の推移を確認しながら、日経記事と社会的イベントの関係について明らかにする。尚、本研究で利用する記事は、2007年1月1日から2014年12月31日の間に掲載された、718,743記事の日経の朝刊である。

2.1. データの前処理

本研究の分析では、以下の記事を分析の対象外とする。第一に、決算数字のみが掲載されている記事である。第二に、会社人事情報のみが掲載されている記事である。上記の記事を除くと、最終的に分析に利用する記事数は、603,063記事になる。

また、ストップワードを設定する。新聞記事の性質上、「～月」や「～日」など日付に関する言葉が多く出現する。こうした言葉は、センチメント分析で有用でないため、分析に利用しないことにする。

2.2. 社会的イベントと日経記事に現れる特徴語

サブプライムローン問題を発端とした世界金融危機や東北大震災などの社会的イベントに対して、日経でどのような話題が掲載されていたのかを確認するために、日ごとに特徴的に現れた単語を抽出する。抽出対象とした単語は全期間を通じて、100回以上出現した名詞に限定する。ただし、数字が特徴的な語として現れることを防ぐために、文書全体で30%以上使用されている名詞は対象外とする。名詞の抽出には、形態素解析ライブラリであるMeCab¹を使

用し、最新の言葉を抽出するために辞書としてmecab-ipadic-neologdを利用する²。

次に、特徴語を定義する。ここで、文書 d は、ある1日に掲載された日経記事とし、 $|D|$ は日経が発行される全日数、 $|d: d \ni w_i|$ は、単語 w_i が出現する日数であるとする。ここで、特徴語は、ある日に単語 w_i が出現した回数を $N_{w_i,d}$ としたときに、

$$tf_{w_i,d} = \frac{N_{w_i,d}}{\sum_k N_{w_k,d}} \quad (1)$$

$$idf_{w_i,d} = \log \frac{|D|}{|d: d \ni w_i|} \quad (2)$$

$$tfidf_{w_i,d} = tf_{w_i,d} \cdot idf_{w_i,d} \quad (3)$$

によって求められる $tfidf_{w_i,d}$ が高い単語とする。

こうして得られた特徴語を、リーマンブラザーズの破綻などの社会的イベント前後に分けて、表1に示す。表1から、社会的イベントごとにどのような言葉が特徴的であったかを概観する。尚、表1の網掛けになっている日付けは、社会的イベントが発生した日を表している。

(1) イチロー4000本安打達成(2013年8月21日)

アメリカ時間の2013年8月21日(日本時間22日)のブルージェイズ戦で、イチローは日米通算で4000本安打を記録した。時差の関係で、23日からイチローに関する話題が特徴的になっていることがわかる。

(2) リーマンブラザーズの破綻(2008年9月15日)

2008年9月15日は、リーマンブラザーズが破綻し、バンク・オブ・アメリカがメリルリンチの買収に合意した日である。また翌日に

開発された形態素解析を行うためのアプリケーションである。詳細は次のサイトを参照のこと。URL <http://taku910.github.io/mecab/> (アクセス日: 2016年8月21日)

¹ MeCabとは京都大学大学院情報学研究所とNTTコミュニケーション科学基礎研究所によって2013年に

² Ishijima et al. (2014) では、標準のシステム辞書を利用している。

表1 イベント前後の特徴語の推移

	2008/8/20	2008/8/21	2008/8/22	2008/8/23	2008/8/24
1	カジノ	競争政策	標本	イチロ	イチロ
2	Cluster	疎開	汚染水	汚染水	水中り
3	新電力	暫定政権	カリウム	安井	卵子
4	チャットアプリ	国有企業	lte	日米通算	イチ子
5	うちわ	山本さん	円借款	mrj	待機児童
6	LIXIL	派遣会社	評議員	鮎物	考古学
7	ビッグデータ	汚染水	暫定政権	三菱航空機	同胞
8	代謝	丸太	国保	保釈	長谷工
9	騒乱	同胞	line	中尾	野党再編
10	もんじゅ	his	化学兵器	飼い主	開催都市
	2008/9/13	2008/9/14	2008/9/15	2008/9/17	2008/9/18
1	麻生氏	モネ	事故米	リーマン	aig
2	事故米	庵	有料道路	事故米	リーマン
3	アウトレット	麻生氏	捕虜	製菓	事故米
4	アーバン	ユダヤ人	分配金	aig	リーマン・ブラザーズ
5	献金	産科	リーマン	リーマン・ブラザーズ	落雷
6	与謝野	検針	共青团	メリル	つゆ
7	小池	ネスレ	内部統制	バンカメ	裁定
8	公開討論会	古い	艦	堂	サムライ債
9	石破	貧血	総合病院	ベアー・スターンズ	MBO
10	資金管理団体	リーマン	秋田県	特別養護老人ホーム	シダックス
	2009/9/14	2009/9/15	2009/9/16	2009/9/17	2009/9/18
1	デルタ	車検	入閣	鳩山首相	商業地
2	減反	デルタ	亀井	鳩山内閣	アウトレット
3	宇宙ステーション	カシオ	郵政	衆	鳩山内閣
4	蚊	閣僚人事	環境税	鳩山政権	住宅地
5	正答	イチロ	飲酒運転	藤井裕久	地価
6	インサイダー取引	雅山	三好	政治主導	チーズ
7	有人	スプレッド	増派	水量	hiv
8	ドッキング	水族館	モリブデン	脱・官僚	MD
9	植物工場	新型インフルエンザ	郵政事業	横顔	サンマ
10	西川	ネットスーパー	カブ	節水	用船
	2011/3/9	2011/3/10	2011/3/11	2011/3/12	2011/3/13
1	カバン	日立物流	SWF	津波警報	炉心
2	ソバ	活字	古紙	震源	炉心溶融
3	林業	リビア	リビア	大津波	格納
4	リビア	監察	租	余震	冷却水
5	世界観	sap	信組	帰宅困難者	カレンダー
6	未納	音質	中国産	災害対策本部	灯火
7	フック	録音	証取	水没	避難所
8	緒方	上場会社	土木	巨大地震	海水
9	法大	紙おむつ	明細書	仙台空港	チェルノブイリ
10	ダイバーシティ	農業委員会	ニンジン	ブレート	停電
	2013/4/2	2013/4/3	2013/4/4	2013/4/5	2013/4/6
1	長嶋	isa	完全試合	マネタリーベース	サーベラス
2	モリブデン	サーベラス	alsok	黒田	以南
3	松井氏	黒鉛	月齢	物価目標	普天間
4	産業政策	roe	地熱発電	蔡	文化財
5	民主派	教育長	イー・アクセス	黒田東彦	イチ子
6	以南	九電	ダルビッシュ	量的・質的金融緩和	ダイエ
7	紀州	教育行政	lseg	当座預金	辺野古
8	北越	発送電分離	八百長	roe	roe
9	焼き鳥	貸し切り	ヒューストン	大王	n
10	再選挙	返戻	選抜高校野球	サーベラス	嘉手納基地

はAIG株が急落した。そのため、17日以降は、リーマンブラザーズ・AIG・バンカメ・メリルなど、一連の金融危機によって、損害を受けた金融機関が特徴的な言葉として現れている。

(3) 民主党政権の発足 (2009年9月16日)

民主党政権が発足した2009年9月16日以降のニュースは、「脱・官僚」や「政治主導」政権のスローガンや、政権のキーマンに関する言葉が特徴的になっている。

(4) 東北大震災 (2011年3月11日)

地震が起こった次の日の記事は、前日の地震が大きな地震であったことや、津波により大きな被害を与えたこと、東京などで帰宅困難者が発生したことなどを報道する記事が多いことがわかる。しかし、地震発生から2日後には、炉心溶融が起こったことを伝える記事が増え、原発の停止により計画停電が実施されることが伝えられる記事が多くなっていることがわかる。

(5) 異次元の金融緩和 (2013年4月4日)

就任直後の日銀総裁の黒田総裁は、「2年程度で、2%の物価上昇」という目標をアナウンスした。そのため、翌日の記事では、「マネタリーベース」や「物価目標」など、金融政策でよく利用される言葉が特徴的な語として現れている。

上記のように、他の新聞同様、日経でも様々な記事が記載されるので、新聞記事と株式市場の関係を分析する際は、記事の分類を行い、企業の経済活動に関するニュースなど、株式市場に影響を与える可能性のある記事を特定することが重要である。

3 日経記事の分類

Ishijima et al. (2014) では、分析対象時期の全ての日経記事を利用して、SIを作成し、株価との関係について考察しているが、スポーツ情報など、金融市場には関係のないニュースが含まれている。そこで、本研究では、LDAと呼ばれる文書クラスタリング技術を利用して、ある日経記事が、経済、政治、スポーツなど、どのようなことを話題にしているか特定する。文書クラスタリングの結果、経済活動について言及している日経記事を特定し、SIを作成する。

3.1. 潜在的ディリクレ配分法 (LDA)

LDA (Latent Dirichlet Allocation) はトピックモデルの一種で、ある文書に含まれる単語がどのようなトピックによって生成されたかを知ることができる。LDAでは、各文書は K 個の観測できないトピックから発生した単語で構成されている、と仮定³する。単語 v を含んだ文書集合 D と、トピック数 K を入力として、以下の2つの確率分布を推定することになる。第一に、トピック k における単語 v の出現する確率の推定である。第二に、文書 d におけるトピック k の出現する確率を推定することである。トピックごとに出現確率の高い単語を確認すれば、そのトピックがどのようなトピックであるかを分析することができる。

3.2. クラスタリング結果

前節の特徴語抽出で使用した603,063記事を用いて、クラスタリングを行う。ここで、トピック数は30個とする。表2には、トピック番号と、その各トピックで出現する確率の高い上

³ 例えば、「東京オリンピックによる経済効果は〇〇億円と推定される」という記事は「スポーツ」と「経済」という2つのトピックから構成されていると考えられる。

位5個の単語を示す。表2の単語の出現情報から、トピック1は生産量や輸出量などが変化したことについて言及しており、トピック21は、金融政策に関する話題である可能性がある。一方、トピック16・22・30はスポーツに関する話題である。単語の出現確率から、経済や市場に関連すると考えられるトピック番号と単語に影響を付けている。

4 SIの作成と社会的イベントとの関連

本節では、SIの作成について述べる。本研究では、2種類のSIを作成する。第一に、経済について話題にしている記事から作成するSIである。具体的には、経済や市場に関連するトピック $k=\{1,2,3,11,18,19,21,23,25,26\}$ の出現する確率が、正の値をとる記事、461,055件を抽出し、SIを作成する。第二に、比較対象として、全ての記事から作成するSIを作成する。

表2 トピック番号と、各トピックに出現する確率が高い上位5個の語

トピック番号	単語				
1	前年	比	増	減	回復
2	億	期	純利益	増	売上高
3	工場	生産	中国	インド	事業
4	店	店舗	商品	出店	販売
5	氏	大統領	選挙	政権	投票
6	北朝鮮	中国	米	ロシア	会談
7	研究	細胞	薬	治療	患者
8	感染	台湾	卸売市場	元代表	ウイルス
9	容疑者	逮捕	容疑	人	捜査
10	首相	民主党	自民党	氏	党
11	株	東証	貸し借り	銘柄	指定
12	原発	東電	福島	稼働	電力
13	配	予	発売	newface	販売
14	スマホ	通信	装置	半導体	パネル
15	さん	私	人	著	作品
16	位	ゴルフ	アング	番	打
17	空売り	人事	課長	比率	維新
18	価格	トン	値上げ	原料	輸入
19	数表	先物	原油	相場	残高
20	社長	氏	取締役	就任	入社
21	日銀	金利	国債	応札	物
22	試合	監督	回	戦	チーム
23	発行	格付け	銀	運用	融資
24	サービス	ネット	サイト	情報	企業
25	企業	経済	政府	制度	改革
26	ドル	安	高	銭	上昇
27	空港	日航	便	ホテル	航空
28	死去	歳	さん	人	氏
29	週	馬	特許	gi	公社債
30	位	女子	男子	五輪	大会

4.1. SIの作成

各記事に現れる単語からセンチメントを計量化する方法について述べる。Ishijima et al. (2014)と同様に、高村 (2007) による「単語感情極性対応表」というセンチメント辞書を利用する。この辞書を $D := \{(D_k, S(D_k)) | k = 1 \dots K\}$ と書く。つまり本辞書は、単語 D_k とそのセンチメント・スコア $S(D_k)$ の組より構成される。このセンチメント・スコアは、登録された単語が、どれだけポジティブ、あるいはネガティブな心理を日本人に想起させるか、という度合いを計量化したものであり、 -1 から $+1$ までの定義域を持つ。すなわち、 $-1(+1)$ に近づくほど、日本人に、よりネガティブ (ポジティブ) な心理を想起させる度合いが強くなる。ポジティブなスコア ($S(D_k) > 0$) を持つ登録単語数は 5,122 である一方で、ネガティブなスコア ($S(D_k) < 0$) を持つ登録単語数はその約 10 倍の 49,983 である。

単語感情極性対応表に記載されている単語で、スコアが 0 よりも小さい語をネガティブな単語とし、0 よりも大きい語をポジティブな単語とする。発行日 t において配信された日経新聞記事における、ネガティブな単語数を N_t とし、ポジティブな単語数を P_t とする。これを以下のように集計したものを日次の SI とする。

$$SI_t = \frac{-N_t}{P_t + N_t + 1} \quad (4)$$

但し、分母に 1 を足すのは、ポジティブな単語とネガティブな単語が出現しないときに、定義できなくなることを防ぐためである。

4.2. SIの作成

作成した SI と、社会的イベントの関係を明らかにするために、SI の 5 日移動平均線を図 1 に示す。期間中で、SI が最も小さくなるのは、2011 年 3 月 16 日である。東日本大震災から 5 日後であり、津波による被害、福島原発の炉心溶解、東京電力の計画停電など、被害の全体像が明らかになった時期である。一方、SI が最も大きくなるのは 2014 年 1 月 6 日である。この日は大発会であり、日経平均株価の終値が前年末比 382 円 43 銭安の 1 万 5908 円 88 銭の大幅安となった。しかし、この日で SI が最高値となっているのは、昨年の大納会までの 9 日間で日経平均株価は 9 日連続で上昇しており、大発会への期待が高かったためであると考えられる。

5 日本株式市場における実証分析

本節では、前節で作成した SI をもとに分析を行う。具体的には、センチメント・インデックス (SI_t)、日経 225 の対数収益率 ($N225_t$)、日経 225 の構成銘柄の出来高の合計の対数値 (Vol_t) からなる 3 変量 VAR モデルを考えて、Tetlock (2007) と沖本 et al. (2014) で検証され

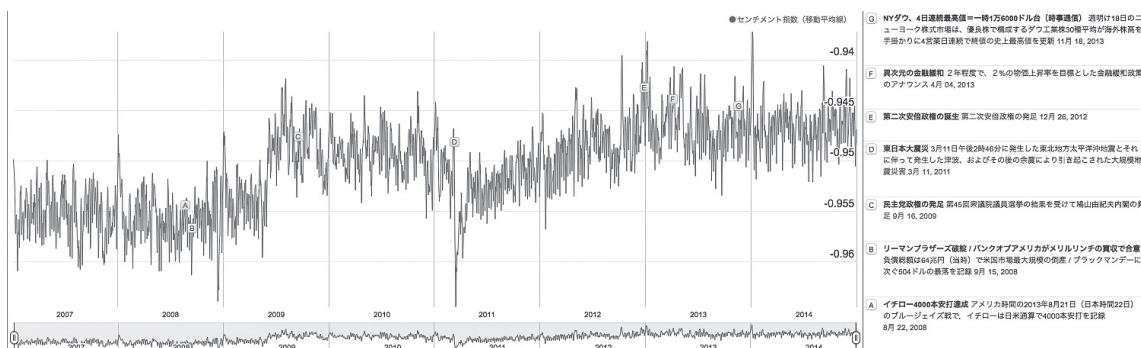


図 1 SIの5日移動平均線と社会的イベント

ている、SIの株価に与える影響について考察する。Ishijima et al. (2014) では、出来高について考慮していないが、本研究では、Tetlock (2007) と沖本 et al. (2014) に準じて、出来高を考察対象に含める。また、SIは、全期間で平均0、分散1になるように標準化を行う。

5.1. SIの対数収益率への影響

SIの対数収益率への影響を調査するために、3変量VARモデルのうち、日経225の対数収益率(N225)に関する回帰モデル：

$$N225_t = \alpha_1 + \sum_{j=1}^p \beta_{1j} N225_{t-j} + \sum_{j=1}^p \gamma_{1j} SI_{t-j} + \sum_{j=1}^p \delta_{1j} Vol_{t-j} + \varepsilon_{1t} \quad (5)$$

について、 SI_{t-1} の係数 γ_{1j} を確認する。 p はVARモデルのラグ次数を表し、最大のラグ次数を5として、AICが最も低くなるような次数を選択する。AICが最も低くなった次数は5であった。推定結果をまとめたものが、表3である。ここでカッコ内はNewey-West標準誤差を用いて計算した t 値を報告している。また、*、**、***は、それぞれ10%、5%、1%の有意水準で有意であることを表す。これらのことは、以下の表4と表5でも同じとする。

LDAを用いて、経済活動に関連する記事のみを利用する場合、 SI_{t-1} の係数と SI_{t-3} の係数が、有意水準10%で有意に正になっており、有意に負になっている他の係数はない。この結果は、沖本 et al. (2014) の結果と整合的である一方、Tetlock (2007) とは異なる結果である。

沖本 et al. (2014) が指摘するように、経済活動に関するニュースが株価に対して本源的な情報を保有しているという情報理論が成立している可能性がある。

一方、全ての記事を利用する場合、4営業日前の SI_{t-4} の係数が負で有意になっており、Tetlock (2007) で指摘されているリバウンドが観測される。この結果は、Tetlock (2007) の結果と整合的であり、SIの株価への影響は一時的なものである、というセンチメント理論が成立している可能性があることを示している。

記事を限定して作成したSIの推定結果に注目すると、有意な係数は、全て正である。一方、 SI_{t-4} の係数と SI_{t-5} の係数は負になっているが、有意ではない。つまり、どの時点においても、ポジティブ度合いが増せば、対数収益率が上昇することを意味している。これは、記事を経済ニュースのみに限定しているため、株式市場に影響のあるSIを作成できているからだと考えられる。

一方、全ての記事を利用して作成したSIでは、対数収益率への影響が正と負で逆転してしまう。これは、SIが、社会面で掲載される殺人事件のように、株式市場とは関係のない記事を含んでしまっているからだと考えられる。

さらに、経済活動に関連する記事から作成したSIのモデルの方が、全ての記事を利用して作成したSIのモデルよりも、調整済み決定係数がわずかであるが、高くなっている。このこ

表3 SIの対数収益率への影響に関する推定結果

SI _{t-j}	LDAによって記事を限定した場合		全ての記事を利用した場合	
SI _{t-1}	0.00091*	(1.76056)	0.00088*	(1.70551)
SI _{t-2}	0.00029	(0.59390)	0.00005	(0.10246)
SI _{t-3}	0.00083*	(1.65662)	0.00091*	(1.81535)
SI _{t-4}	-0.00073	(-1.50278)	-0.00081*	(-1.64917)
SI _{t-5}	-0.00019	(-0.40743)	-0.00017	(-0.35028)
Adj. R	0.00531		0.00490	

とから、株価を説明するSIの作成には、経済活動には関係のない記事を除いて作成した方がよいと考えられる。

5.2. 対数収益率のSIへの影響

次に、対数収益率のSIへの影響を考察する。このことを考察するために、3変量VARモデルのうち、センチメント・インデックス (SI_t) の回帰式 (6) に注目する。上記の分析同様、 p はVARモデルのラグ次数を表し、最大のラグ次数を5として、AICが最も低くなるような次数を選択する。AICが最も低くなった次数は5であった。

$$SI_t = \alpha_2 + \sum_{j=1}^p \beta_{2j} N225_{t-j} + \sum_{j=1}^p \gamma_{2j} SI_{t-j} + \sum_{j=1}^p \delta_{2j} Vol_{t-j} + \varepsilon_{2t} \quad (6)$$

もし、新聞記事の中に、株価に関する記述があれば、 $N225_{t-j}$ の係数が正か負かのいずれかで有意になるはずである。推定結果を表4に示す。

記事を限定して作成したSIと、全記事から作成したSIのどちらの場合も、 $N225_{t-2}$ の係数と $N225_{t-3}$ の係数が正で有意である。これは、作成したSIに、2・3営業日前の、日本の株式市場の状況を表現していることを示している。つまり、作成した2種類のSIは、株式市場に後追いで反応している、ということである。

また、SIとして、記事を限定して作成したSIを採用した方が、調整済み決定係数が高くなっており、モデルの当てはまりが良くなる。

しかし、なぜSIが前営業日の対数収益率に反応していないのかは、今後の研究課題である。

5.3. SIの出来高への影響

最後に、SIの出来高への影響を考察する。Delong et al. (1990) やCampbell et al. (1993) に基づき、Tetlock (2007) は、対数収益率・SI・出来高の3変量VARモデルだけではなく、SIの絶対値を含めた4変量VARモデルを考えている。本研究も、それに倣い、SIと出来高の関係に注目するために、4変量VARモデルの下記の回帰式 (7) を考える。ここで、 SI_{t-j} の係数と $|SI|_{t-j}$ の係数が重要である。

$$Vol_t = \alpha_3 + \sum_{j=1}^p \beta_{3j} N225_{t-j} + \sum_{j=1}^p \gamma_{3j} SI_{t-j} + \sum_{j=1}^p \psi_{3j} |SI|_{t-j} + \sum_{j=1}^p \delta_{3j} Vol_{t-j} + \varepsilon_{3t} \quad (7)$$

上記までの分析同様、 p はVARモデルのラグ次数を表す。ラグ次数を1から5まで変化させてみて、AICが最も低くなった次数5を選択した。推定結果を表5に示す。

前営業日のSIの出来高への影響を表わす SI_{t-1} の係数に注目する。記事を限定して作成したSIのモデルでは、 SI_{t-1} の係数が正で有意になっているが、全ての記事を利用して作成したSIのモデルでは、有意になっていない。つまり、この結果は、前営業日の経済活動に関するポジティブなニュースは出来高を増加させ、経済活動に関するネガティブなニュースは出来高を減少させることを示している。一方、前営業日のSIの絶対値の出来高への影響を表す $|SI|_{t-1}$

表4 対数収益率のSIへの影響に関する推定結果

N225 _{t-j}	LDAによって記事を限定した場合		全ての記事を利用した場合	
	係数	(t)	係数	(t)
N225 _{t-1}	1.57024	(1.36680)	1.43639	(1.70551)
N225 _{t-2}	2.49047**	(2.12557)	2.60836**	(0.10246)
N225 _{t-3}	2.20995**	(2.19445)	2.60992***	(1.81535)
N225 _{t-4}	0.47296	(0.42204)	0.96105	(-1.64917)
N225 _{t-5}	0.63010	(0.57654)	0.26785	(-0.35028)
Adj. R	0.36680		0.36570	

表5 SIの出来高への影響に関する推定結果

SI_{t-j}	LDAによって記事を限定した場合		全ての記事を利用した場合	
SI_{t-1}	0.01010**	(2.00535)	0.00730	(1.45662)
SI_{t-2}	0.01287***	(2.64606)	0.01057**	(2.12214)
SI_{t-3}	0.00385	(0.80720)	0.00416	(0.81910)
SI_{t-4}	0.02001***	(3.62202)	0.02411***	(4.39540)
SI_{t-5}	-0.02834***	(-5.54388)	-0.02459***	(-4.85821)

$ SI _{t-j}$	LDAによって記事を限定した場合		全ての記事を利用した場合	
$ SI _{t-1}$	0.00493	(0.69298)	0.00454	(0.61831)
$ SI _{t-2}$	0.00123	(0.17111)	0.00819	(1.08125)
$ SI _{t-3}$	0.01067	(1.54344)	0.01885**	(2.55177)
$ SI _{t-4}$	0.00695	(0.83503)	0.00438	(0.52270)
$ SI _{t-5}$	-0.01180	(-1.52057)	-0.01462*	(-1.89555)

Adj. R	0.63690		0.63680	
---------------	---------	--	---------	--

の係数は、 SI_{t-1} の結果と異なり、両方のSIのモデルで、有意な結果とはならなかった。

また、記事を限定したSIを変数として採用する方が調整済み決定係数は高くなり、モデルの当てはまりが良くなる。

6 結論

本研究では、日経記事を書かれている内容ごとに分類し、経済活動に関する記事より計量された、ポジティブさとネガティブさを反映した日次SIを作成した。その日次SIと株式市場との関連についてVARモデルにより実証分析を行ったところ、次のような知見を得た。日経記事の、経済や市場を話題にした記事から作成したSIは、(1)翌日の対数収益率と出来高を有意に説明し、(2)株式市場に対して、後追いで反応するということである。

7 参考文献

Asur, S. and Huberman, B.A. (2010) "Predicting the

Future with Social Media," <http://www.arxiv.org> arXiv:1003.5699v1.

Bollen, J., Mao, H. and Zeng, X. (2011) "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2(1), 1-8.

Boudoukh, J., Feldman, R., Kogan, S. and Richardson, M. (2012) "Which News Moves Stock Prices? A Textual Analysis," *NBER Working Paper*, No. w18725.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1992) "Trading volume and serial correlation in stock returns," *National Bureau of Economic Research*, No. w4193.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990) "Noise trader risk in financial markets," *Journal of political Economy*, 703-738.

Fama, E.F. (1965) "The Behavior of Stock-market Prices," *The Journal of Business*, 38(1), 34-105.

Fama, E.F. (1991) "Efficient Capital Markets: II," *Journal of Finance*, 46(5), 1575-1617.

Gruhl, D., Guha, R., Kumar, R., Novak, J. and

- Tomkins, A. (2005) "The Predictive Power of Online Chatter," *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, New York: ACM Press, 78-87.
- Ishijima, H., Kazumi, T. and Maeda, A. (2015) "Sentiment Analysis for the Japanese Stock Market," *Global Business and Economics Review*, 17(3), 237-255.
- Kahneman, D. and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263-291.
- Liu, Y., Huang, X., An, A. and Yu, X. (2007) "ARSA: A Sentiment-aware Model for Predicting Sales Performance using Blogs," *SIGIR '07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York: ACM, 607-614.
- Loughran, T. and McDonald, B. (2011) "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, 66(1), 35-65
- Mishne, G. and Glance, N. (2006) "Predicting Movie Sales from Blogger Sentiment," *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*.
- Ritter, J.R. (2003) "Behavioral Finance," *Pacific-Basin Finance Journal*, 11(4), 429-437.
- Schumaker, R.P. and Chen, H. (2009) "Textual Analysis of Stock Market Prediction using Breaking Financial News: The AZFin text system," *ACM Transactions on Information Systems*, 27(2), 12:1-12:19.
- Tetlock, P.C. (2007) "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, 62(3), 1139-1168.
- Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. (2008) "More than Words: Quantifying Language to Measure Firms' Fundamentals," *Journal of Finance*, 63(3), 1437-1467.
- 石島博・数見拓朗・前田章 (2015) 「市場センチメント・インデックスの構築と株価説明力の分析：日次データによる検証」『経済政策ジャーナル』11巻2号, 7-10頁.
- 沖本竜義・平澤英司 (2014) 「ニュース指標による株式市場の予測可能性」『証券アナリストジャーナル』52巻4号, 67-75頁.
- 五島圭一・高橋大志 (2016) 「ニュースと株価に関する実証分析：ディープラーニングによるニュース記事の評判分析」『証券アナリストジャーナル』54巻3号, 76-86頁.
- 高村大也 (2007), 単語感情極性対応表, http://www.lr.pi.titech.ac.jp/~takamura/pubs/pn_ja.dic (accessed 2015-08-01).
- 坪内孝太・山下達雄 (2015) 「株価掲示板情報の感情解析と株価との相関の研究」『2015年度人工知能学会全国大会講演集』1J5-OS-13b-2in.

Quantifying Market Sentiment for the Empirical Analysis in the Japanese Stock Market

Takuro Kazumi

The purpose of this paper is to quantify the market sentiment as two indexes and examine whether they can help predict stock prices in the Japanese market. Sentiment analysis is gaining increasing interest in both academia and business. Along these lines, Ishijima et al. (2014) created a sentiment index that quantifies the positive or negative emotion that might appear in entire articles of Nikkei which is the most popular business newspaper in Japan. They concluded that the sentiment index significantly predicts stock prices three days in advance. We re-examine their results by suggesting a new sentiment index quantified from the articles limited to the economic-activity-related news and explore the implication on how the sentiment index can help explain Japanese stock price. Our findings are three-fold: (i) Sentiment index created from the articles limited to the economic-activity-related news significantly allows us to explain Nikkei 225 and the market trading volume of the next business day. (ii) We cannot observe the return reversal referred in the literature. (iii) SI will follow the stock price.

JEL Classification: C88, E37, G17

Keywords: Sentiment Analysis, Nikkei, Text Mining, Predictability of Stock Price

Editorial Policy

The Osaka Daigaku Keizaigaku (English title, Osaka Economic Papers) is published quarterly by the Economic Society of Osaka University and the Graduate School of Economics, Osaka University. The articles may be either in Japanese or in Western languages.

The Journal shall be under the editorial direction of an editorial board of three persons chosen from members of the Graduate School of Economics of Osaka University. The editorial board shall select papers for publication from submissions and classify them into the following categories : articles, notes, data, and book reviews.

Researchers who belong to the Graduate School of Economics of Osaka University may submit their studies for publication to this journal. Those who do not belong to the Graduate School may also publish their papers in this journal, if their contribution is closely related to research being undertaken in the Graduate School of Economics of Osaka University.

In the case of contributed manuscripts, the author should be a member of the Economic Society of Osaka University, who has paid the yearly membership fee of 4,000 yen.

大阪大学経済学 第66巻 第3号 (通巻215号)

平成28年12月発行

編集兼発行人 〒560-0043 豊中市待兼山町1番7号

印刷所 〒920-0855 金沢市武蔵町7番10号

発行所 〒560-0043 豊中市待兼山町1番7号

堂目卓生

能登印刷株式会社

大阪大学経済学会・大阪大学大学院経済学研究科

tel/fax 06-6850-5270

振替 00940-2-19842

OSAKA ECONOMIC PAPERS

Vol.66

No.3

December 2016

Articles

- | | | |
|--|---------------|----|
| Proximity of Local Amenities as a Factor in Migration | Masaaki Toma | 1 |
| Quantifying Market Sentiment for the Empirical Analysis in the Japanese Stock Market | Takuro Kazumi | 24 |

THE ECONOMIC SOCIETY OF OSAKA UNIVERSITY
GRADUATE SCHOOL OF ECONOMICS, OSAKA UNIVERSITY
TOYONAKA, OSAKA, JAPAN