

るような変化が生じていることが指摘されている(Caltabiano 2008, Castiglioni and Dalla Zuanna 2008)(図2)。もし日本における TFR 上昇も、こうしたイタリアにおける変化と共通していれば、回復傾向が全国的に広がり、出生率低下に歯止めがかかる可能性がある。また西欧・北欧に固有な現象との見方もある第二人口転換が、異なる文化的背景をもつ先進国にも今後波及する可能性を示唆することになる。日本以外に著しい出生率の低下を経験している東アジア諸国の将来にとっても、重要な知見を与えることになる。

そこで、日本においても第二の人口転換に特徴的な現象が出生率の上昇に結びついているかを検証することになるのだが、ここでは、Castiglioni らのイタリアについての先行研究(Castiglioni and Dalla Zuanna 2008)と同じアプローチを採用したい。すなわち、国内の地域差に着目し、TFR の回復が、第二人口転換の特徴をより強くもつ地域で起こっているのかを確認するというものである。

出生力の地域パターン

本研究の目的は、出生力変動と家族形成パターンおよび価値観や社会経済的諸条件との関係を明らかにすることである。出生力変動の要因を探る場合、当該変動の先行指標の特定およびその傾向の把握が有効である。一般的には時間的変化を分析対象としてその要因をさぐるアプローチが主流であるが、場合によっては、行動変化が一国内における地域の事情の違いを反映して、地域間格差として現れることがある。本研究も、この地域差を分析することで、変動要因の特定に迫ることとした。

日本における出生力の地域パターンについては、出生力変動の動向と要因に迫る目的でこれまでも多くの研究者が注目してきた(高橋 1976, 石川 1992, 河野 1992, 廣嶋・三田 1995, 高橋 1997, 清水 2004)。出生力の地域パターンは第一の出生力転換を挟んで「東高西低」の地理的パターンから、西日本や大都市圏で著しく低下するという過程を経て、さらにその後格差が縮小する局面を迎えたが(高橋 1976, Nakagawa 2003)、近年では地域間の相対的格差が再び拡大しているとの指摘もある(清水 2004)。さらに出生力低下を結婚行動および夫婦の出生行動といった要因別にとらえる研究も進んでいる。佐々井(2007)は夫婦出生力そのものに着目し、地域ブロックを単位とした分析を行っている。標本調査(出生動向基本調査)の個票を使い、結婚時の妻の年齢、妻の最終学歴、妻の就業状況、親との同別居、居住地域の人口集中地区区分といった構造的要因と地域ブロック固有の効果を検証している。完結出生児数については、妻の結婚年齢(-)、妻の学歴(-)、人口集中地区(-)、親との同居(調査時)(+)との関連が見られ、東北地方で有意に高い結果が得られている。しかし出生過程途上の夫婦の出生子ども数は、東海や九州で高いといった違いがみられた。みずほ情報総研(2005)のレポートでも、都道府県別有配偶出生力をいくつかの社会経済変数および気温といった指標で回帰した分析結果が示されているが、有配偶出生力の西高東低の傾向が、気温(+)、男性の長時間労働(-)、保育所在所児割合(+))と関

係していることが示されている。さらに山内らは都道府県単位ではなく都市圏という生活圏に着目し、1980年以降の出生力に関し、婚姻力変動の寄与とその地域差の分析を試みている(山内・西岡・小池 2005)。都市圏規模に関わらず夫婦の出生力低下が見られることや、都市圏の中心部で顕著であった出生力の落ち込みが、次第に郊外でも起きていることなどが指摘されている。

このように、出生力の地理的な違いは過去も現在も存在しているが、地域の分析単位に何を、出生力の指標として何を使うべきかについては分析の目的に依存すると言える。

まず本研究で検証したい出生力は、ある一時点における出生力のレベルそのものというよりも、第二人口転換の中で、家族形成パターンが変わり、一時的に落ち込んだ期間 TFR がキャッチアップ効果によって回復するという一連の過程の有無である。従って、期間 TFR の反転上昇の有無(度合いの違い)を表す指標を分析対象とする。ちなみに、イタリアの先行研究である Castiglioni and Dalla Zuanna (2008)においても地域別の出生率の回復を外国人出生力、婚外出生割合、一人当たり所得、基準年の出生率で回帰するモデルの推定が試みられているが、そこでは 1986-95 年の TFR に対する 1996-2000 年の TFR の比が従属変数として用いられている。

また分析する地域の単位としては 47 の都道府県に着目した。第二人口転換仮説は、個々人の社会的属性の変化のみならず、ライフスタイルに関する価値観の変化の重要性を強調している。都道府県単位というのは、個々人の生活圏としては大きいかもしれないが、文化圏としては「県民性」という言葉があるように共通する文化や慣習が作用する単位として適切な範囲と考えられる。先の論文を含むイタリアの先行研究では国内 20 の地方を分析単位としており(Caltabiano 2008, Castiglioni and Dalla Zuanna 2008)、また Lesthaeghe らが行った、米国における第二人口転換仮説の適用についての地域分析は、50 の州を単位としている(Lesthaeghe and Neidert 2006)。このような理論的な理由の他に、分析実行上の制約も関わっている。第二人口転換に関わるような家族形成形態に関するデータの中には、標本調査から集計する必要があるものがあり、信頼性を保証するためには都道府県単位の集計が限界であると考えられるからである。以上のような事情をふまえ、出生力の反転上昇に関する指標が都道府県単位でどのように異なるのか、そしてそれが他のどのような変数と関連しているのかを探ることとした。

データと方法

どのような変数に着目するかについては、近年のイタリアにおける期間出生率反転とその関連要因の分析を参考にした。Castiglioni らは、近年のイタリア北部における期間出生率の反転上昇に連動する変化として、以下の 4 点を挙げている(Castiglioni and Dalla Zuanna 2008)。(1)晩産化が若年世代で止まり、30 代以降でのキャッチアップ効果が出てきたというもの、(2)増加の一途をたどる外国人の出生率が高く、全体の出生率上昇に貢献していると

いうもの、(3)他の欧州諸国で観察されているように、同棲や婚外出生、離婚などの新しい結婚パターンが普及している地域で、出生率が回復しているというもの、(4)子育てを困難にしていた一連の要因（子どもにかかる経済力や時間の不足）が解消された地域で出生率が回復したというものである。関連のメカニズムの検証や厳密な定量化等については、より詳細な分析がまたれるところであるが、いずれの要素も近年の北部の出生率上昇に関連しているとの見解を示している。

本研究では、期間出生率の反転上昇という現象に関連する要素として、晩産化に関する指標、未婚、同棲経験、婚外出生、婚前妊娠、離婚、同棲といった家族形成の新たな形態に関する指標、その他の社会経済状況、価値観を示す変数として、女性の大学進学率、男性の失業率、農業人口割合に着目した。

出生力や結婚、離婚に関する指標は人口動態統計および国勢調査人口、国勢調査間の推計人口などを基に指標化した。都道府県別の TFR 算出にあたっては 5 歳階級別の外国人を含む人口を分母としている。2007 年の都道府県別出生率の状況については石川(2008)を参照されたい。その他、同棲や婚前妊娠結婚といった公表値で得られない指標は、国立社会保障・人口問題研究所が実施している「出生動向基本調査」や毎日新聞社人口問題調査会による「第 1 回人口・家族・世代調査」を用いる。女性の大学進学率については「学校基本調査」、男性の完全失業率、農業人口割合については、「国勢調査」の都道府県別の集計データを利用した。

空間分析を行う際には、都道府県単位の shape ファイルを利用する。shape ファイルには都道府県の位置情報が含まれており、相互の空間的関係の情報を分析に取り入れることを可能にする。位置情報のみの shape ファイルに、上記の都道府県別の情報をマージすることによって分析データを構築する。

はじめにこれらの集計データがどのような地域分布をしているかを視覚的に把握するため、探索的空間分析 (Exploratory spatial data analysis(ESDA)) を行う。ここでは空間相関 spatial correlation の検証として TFR 回復指標の Moran の I 係数の算出を試みる。続いて、TFR 回復指標が、その他の変数の関数として表されるモデルを構築し推定する。その際、当該地域の構造要因だけでは説明できない近隣地域の影響（フィードバック効果、集住効果、類似反応効果、空間依存性など）といったものが存在する可能性を考え、誤差が空間的に相関している空間誤差モデルの推定を試みた。

空間回帰分析には、代表的なものとして空間ラグモデルと空間誤差モデルがある。両者は空間的依存性を異なるメカニズムでモデル化している。空間ラグモデルは一般に以下のように表記される（行列記法）(Anselin 1988)。

$$y = \rho W y + X \beta + \varepsilon,$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

W は各地域の結びつきを示すウェイト行列である。 ρ が 0 なら空間依存性は存在しないことになり、通常の OLS が適用できる。しかし 0 でない場合、OLS では正しい値が推定できないので、最尤法を用いて推定することが提案されている。なお、このモデルは、contextual effect model といわれるような、diffusion 効果を扱う研究で提案されてきたモデルと基本的には同じ構造であると言える。例えば、出生力転換の伝播理論を整理して Casterline (2001) の提示している出生力の dynamics に関する代表的なモデルは、以下のように表記されている。

$$Y_{i,t} = X_{i,t}\beta + \alpha \sum Y_{j,t-1}W_j + \varepsilon_{i,t}$$

ここで、Y は出生行動の指標、X は出生力の代表的な規定要因、W は個人 j の個人 i にとっての重要度（親密度）を示すウェイト行列、t は時点を表す。

他方、空間誤差モデルは以下のように表記できる。

$$y = \lambda W\xi + X\beta + \varepsilon,$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

ラグモデルと同様、W は各地域の結びつきを示すウェイト行列であり、空間依存性を示す λ が 0 でない場合、空間依存性を無視した OLS では係数の分散が過小に推定されてしまう。そこで最尤推定法による空間誤差モデルが必要となる。

両者のモデルは、空間依存性の解釈が異なるという意味で大きく違う。空間ラグモデルは、観察したいある地域の y 自体が、他の地域の y に影響を与えるというメカニズムに基づく。一方、空間誤差モデルは、隣接する地域の誤差項に依存性があり、それが隣接地域の y 変数に同じような影響を与えるとの考え方に基づく。モデルの選択に際しては、観察したい現象の空間依存性にどのような説明が可能かを十分検討することが重要である(Ward and Gleditsch 2008)。

一般に近隣地域の効果を検証する研究では、地域の集計データのみを使う場合と、地域の集計データが組み込まれた個人に関する調査データが使われることがある。この研究は、前者の集計データを利用した研究にあたる。また探索的空間分析および空間回帰分析には地理学者 Anselin らが開発した空間分析の専用ソフトである GeoDa を用いた。GeoDa はフリーソフトであり、以下のサイトからダウンロードできる：<http://geodacenter.asu.edu/>。分析に必要な変数が入った shape ファイルを読み込み、ウェイト行列の作成や Moran' s I の算出、空間回帰分析の推定ができる。

分析に用いる変数

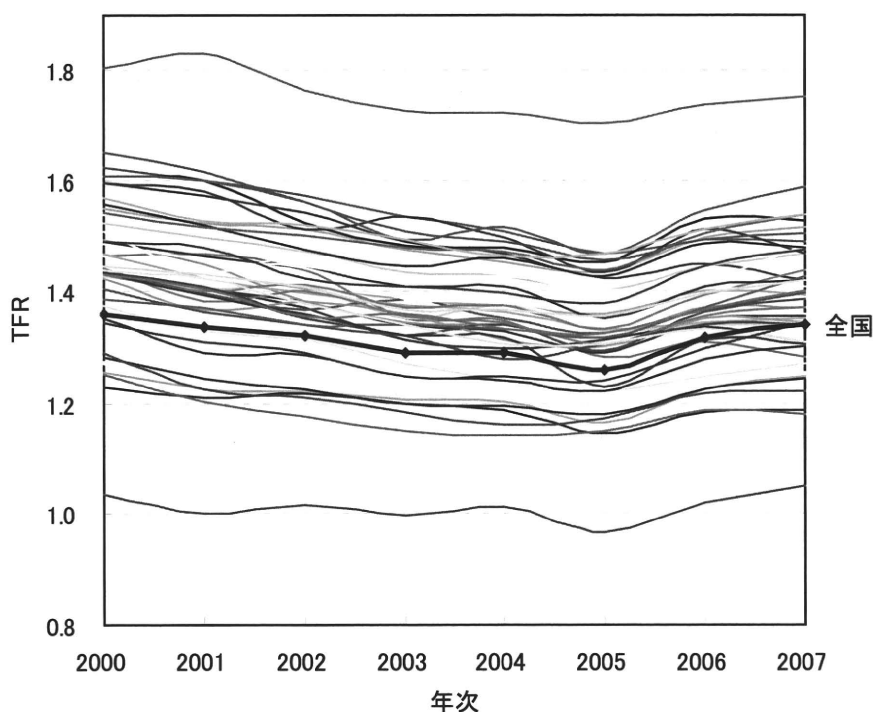
分析に用いる都道府県別の変数を、以下では目的変数、説明変数の順に説明したい。

目的変数は期間 TFR の反転傾向を象徴的にとらえるものとして、2000 年以降の最低 TFR に対する 2007 年 TFR の比(TFR 反転指標、以後変数名は RAT)を用いることにした。TFR 自体、年齢別出生率の合計という合成指標であり、その変動は大変複雑なメカニズムがかかわるので、その変動を出生行動の変化と結びつけることは容易でない。ここでは出生行動の指標というよりも、合成指標の反転が、女性の高学歴化や新しい家族形成パターンの容認傾向を含む第二人口転換が進展していると結論づけるための十分条件ではないが必要条件の一つであると考え、TFR の反転を象徴的に示す指標としてこの指標を用いる。

2000 年以降の TFR の推移を全国および都道府県別に示したのが図 3 である。多くの都道府県で 2005 年以降傾向が変わっていることが分かる。地域ごとの TFR 上昇比を地図上に描くと(十分位で表示)図 4 のようになる。西日本で上昇率が高いところが多い傾向が確認できる。東京から愛知にかけての太平洋側でも高い。

最低出生率が低かったところで反動として出生率が上昇したという可能性が考えられるが、TFR 上昇比と最低出生率との関係をみたところ、はずれ値である沖縄と東京を除けば、両者にはっきりした関係はなかった(図 5)。

図 3 全国および都道府県別 TFR の推移：2000-2007



注：5 歳階級別年齢別出生率の合計値として算出。

図4 2007年のTFR反転上昇比(十分位)

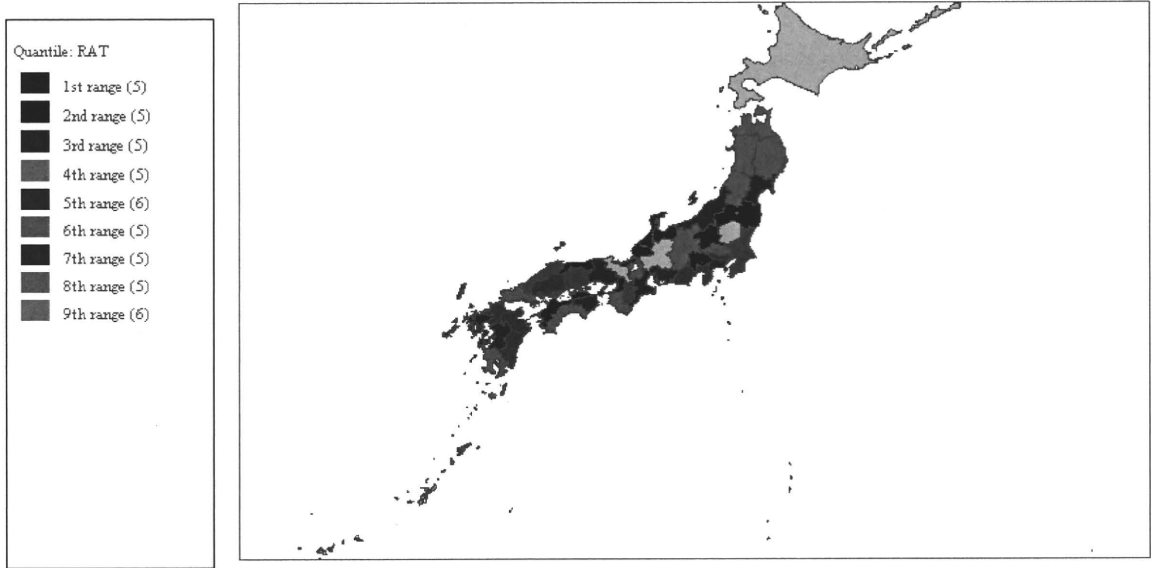
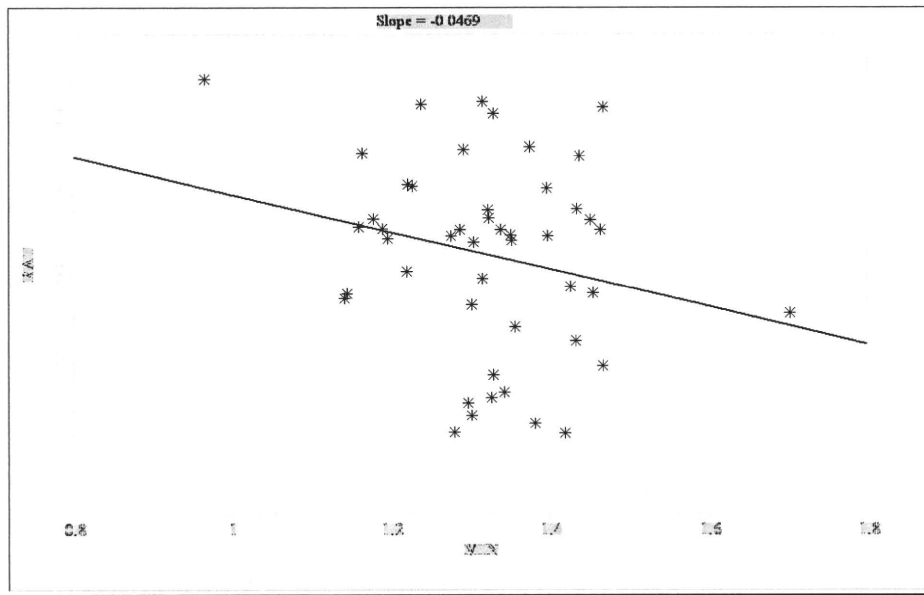


図5 TFR反転上昇比(縦軸)と最低出生率(横軸)の関係



TFR反転指標のMoran's Iの算出

TFR反転指標の空間自己相関 Spatial autocorrelation をMoran's Iで検証する。Moran's Iとは、ある変数と、その変数の周辺の加重平均値の相関関係を示す指標である。Moran's Iは以下のように表記される(Cliff and Ord 1973, Moran 1950)。

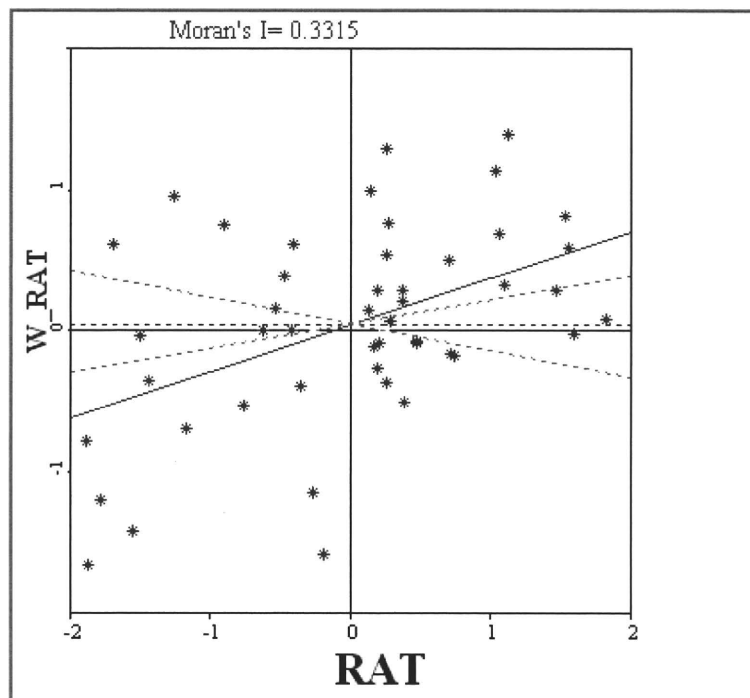
$$I = \left(\frac{n}{\sum_i \sum_j w_{ij}} \right) \left(\frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \right)$$

i, j は n ある地域の地域単位を示す。 w_{ij} は空間ウェイトで地域 i と j の結びつきを示す。 w_{ij} は $n \times n$ の空間ウェイト行列 W の要素である。

当該地域の値と周囲の平均値になんの関係性もない場合には 0 に近い値になり、値が正で大きいほど相関が高く（値の高い地域は周囲も高い）、値が負で大きいほど、高い値の地域の周囲は値が低い(社会科学ではまれ)という状況を示す。ピアソンの相関係数と概念的には近いが Moran's I は -1 と 1 の間に納まらない(Voss et al. 2004)。

図 6 は Moran 散布図で、横軸が当該地域の TFR 反転指標、縦軸がウェイト行列によって定義された周囲の TFR 反転指標の平均値である。TFR 反転指標の Moran's I を計算したところ、0.3315 で有意に高く、TFR 反転指標に関して空間的自己相関 Spatial autocorrelation が存在している、すなわち、出生率の反転が生じている地域では周囲でも反転起きている可能性が高いことを示す。

図 6 TFR 反転指標に関する Moran 散布図と Moran's I



注：横軸は当該地域の TFR 反転指標、縦軸はウェイト行列によって定義された周辺地域における TFR 反転指標の平均値。

Moran's I の高さは、TFR の反転という現象が、地理的に近い地域で近い傾向を持っているということの意味する。こうした空間的依存関係は、一部は構造的要因が近隣地域で似通っているとう事情に依っているかもしれないが、そのような構造的要因の効果を取り除いた後でも観察されるとすれば、そこには構造的要因では説明できない、地理的空間の依存性の存在が考えられる。本研究後半で、誤差が空間的に相関している空間誤差モデルを用いて検証する。

LISA (Local indicator of spatial association) クラスタ

前節で算出した全体的な空間的相関係数に対し、それぞれの地域がどのように貢献しているのかを確認する方法として、LISA 統計量を利用することができる(Anselin 1995)。

LISA 統計量に基づき、当該地域の値と、周辺地域の値が、ともに高い(High,High)、ともに低い(Low-Low)、その他(High, Low)、(Low,High)地域に分類し図示してみた(図7)。

東北地方は当該地域も周囲も低いという特徴、九州の一部で、当該地域も周囲も高いという地域が見られ、山梨については、当該地域は低い周囲が高いという特徴が見られる。

図7 LISA クラスタ図



空間依存性とは何か

出生力反転上昇指標のMoran's Iを算出することによって、この指標が空間的依存性をもっている可能性が示唆された。はたして空間依存性とは何なのであろうか。近年、さまざまな社会科学研究分野でこうした空間依存性の存在が認識され、空間依存性がなぜ生じるのかについても議論が深まっている。これまで議論されているメカニズムをVossら(2004)の整理を参考にしながらいくつか紹介したい。

1) フィードバック

社会的プロセスの多くで、個人や世帯は他者と交流し、相互に影響を受ける。こうした影響は接触頻度が多いほど強いと考えられ、一般に居住地域が近いほどそうした接触の機会が多いと考えられる。こうした考えは、採用／伝播理論(adoption/diffusion theory)(Rogers 1995)あるいはエージェント・インタラクション理論(Irwin and Bockstael 2004)に基づいている。

出生行動については、以前から伝播のメカニズムが指摘されていた(Rogers 1995, Retherford and Palmore 1983, Montgomery and Casterline 1996, Kohler 2000, Casterline 2001)。日本については河野(1992)において出生力低下の地理的連続的広がり指摘され、近年では池(2009)において伝播拡散理論に関する詳細な概念整理が行われている。空間効果を検証するモデルはこうした伝播理論における実証に対しても貢献できる可能性をもっていると言えよう。

2) 集住化 grouping forces

同じような属性や特徴をもつ人々は、自ら選択して、あるいは社会的、経済的、政治的に強要されて、近い場所に集住する傾向がある。こうした場合の空間依存性は、観察される属性をモデルに入れることである程度は説明される。しかしそれでも依存性が消えない場合は、空間誤差モデルの導入が必要であり、観察されない共通の特徴の存在を示唆することになる。

3) 反応の類似性 grouping responses

同じような属性や特徴をもつ人々は、外部の刺激に対して、同じような反応を示す可能性が高い。我々の社会には、しばしば文脈効果contextual effectが存在している。たとえば、地方特有の産業構造や労働環境、長期にわたって蓄積してきた文化の影響、自然地理的環境は、その周辺一帯の人に影響を与える。どのような地域に住むかによって、外部の刺激に対するキャパシティ(人的、社会的資本)が異なるかもしれない。そのような場合、地域特有の事情を選別し、空間的レジームspatial regimes(検証したい関係において構造的な違いを示す小区域)として、他の地域と弁別することが有効である。それを怠ると、空間依存性の要因は誤差の中に残ってしまうことになる。

4) 厄介な自己相関 nuisance autocorrelation

検証したい空間的プロセスが、分析者が指定した分析単位よりもより大きい範囲を単位として生じているときは、否応なく空間的自己相関が観察されてしまう。実質的な自己相関と同時に、このような厄介な自己相関が分析に入り込む可能性がある。Anselin(1988)は両者を「実質的な伝染(real contagion, spatial dependence)」と「見掛け上の伝染(apparent contagion, spatial heterogeneity)」という言葉を用いて区別している。分析者は空間的に観察された関係をみるとき、まずその関係が反応 reaction なのか相互作用 interaction なのか、すなわち異質性なのか空間依存なのかを問わなければならない。もし、関係が単なる当該地域の地理的、文化的、経済的要因(これら要因

自体が空間依存している状況が多い) に対する反応に過ぎないのなら、それらの要因を統制した通常の回帰分析を行うだけでよい。独立変数が、従属変数の自己相関のほとんどを説明することになる。そうした属性でコントロールしてもなお、誤差に自己相関が残る場合は、空間依存性が実質的に存在し、メカニズムの解明が必要ということになる。

出生力反転上昇指標の地域的自己相関の検出だけでは、それがどの程度関連要因の構造自体の近似性によるものなのか、あるいは出生力反転現象そのものが、近隣地域で類似する特別なメカニズムが存在しているのかわからない。そこでまず、関連要因の構造だけで説明するモデルを推定し、構造をコントロールした上で、さらに空間依存性が残っているかを確認する作業に移っていく。まずは本研究で着目する関連要因の地域パターンを確認しておこう。

説明変数の地域パターン

第二の人口転換において注目されている変数について地域パターンを確認する。ここでは晩産化指標、20代後半未婚率、第1子婚外出生率割合、第1子婚前妊娠出生率割合、再生産年齢有配偶離婚率、35歳未満未婚女性の同棲経験割合、全女性の同棲経験割合を見ていく(図8～図14)。

晩産化指標には、Lesthaegheらが米国の州別の第二人口転換度合いを検証した際に用いている指標を利用した(Lesthaeghe and Neidert 2006)。20代の年齢別出生率の合計値に対する30代以降の年齢別出生率の合計値の比である。大きいほど、晩産化が進んでいることを意味する。晩産化指標については、首都圏および阪神地域の都市部で高い傾向が見られる。一方九州南部、中国地方西部、宮城県を除く東北地方で低い。20代の未婚率も、晩産化指標とほぼ同様の傾向を示すが、愛知、岐阜、長野、富山、新潟といった中部・北陸地方で、晩産化の水準と比べると未婚率が相対的に低い傾向が見られる。第1子婚外出生率割合は、大分を除く九州、四国東部、福島県、北海道で高い。未婚率が相対的に低かった中部・北陸は低い傾向がある。第1子婚前妊娠出生率割合は、人口動態統計に基づいて算出した。東北地方および九州で高く、首都圏、中部、関西地域では低い。離婚については、通常の離婚率(人口に対する離婚発生数)ではなく、離婚するリスクのある人からの発生率をとらえるため、年齢別有配偶女子人口に対する当該年齢の離婚数を15歳から49歳まで足し上げた数値を用いた。この再生産年齢有配偶離婚率は、北海道、高知、岡山、鳥取で高く、その他九州西部でも高い傾向がある。

図8 晩産化指標（第1子の30代出生率/20代出生率） 図9 女性20代後半未婚率

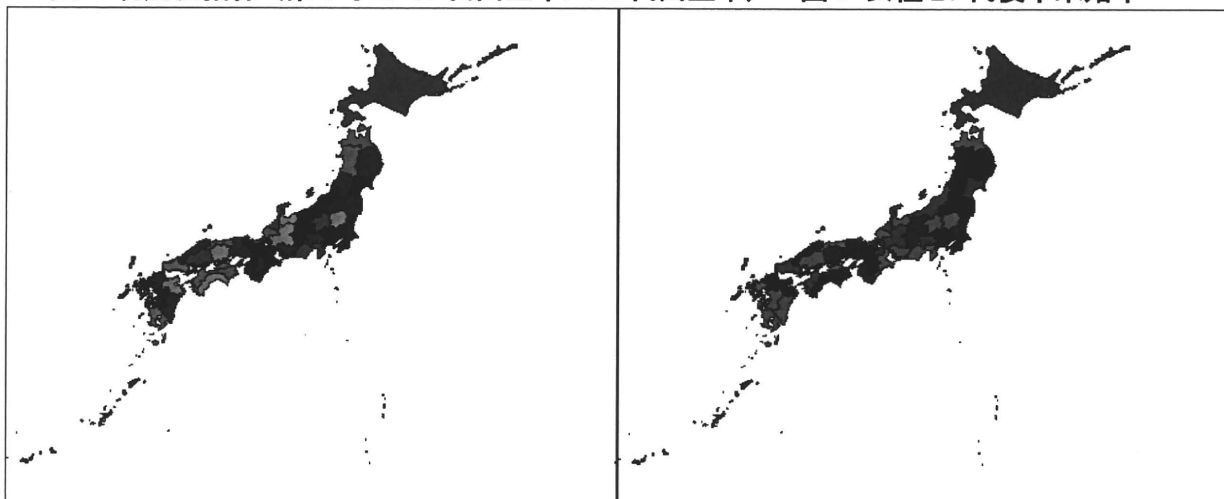


図10 第1子婚外出生率割合

図11 第1子婚前妊娠出生率割合

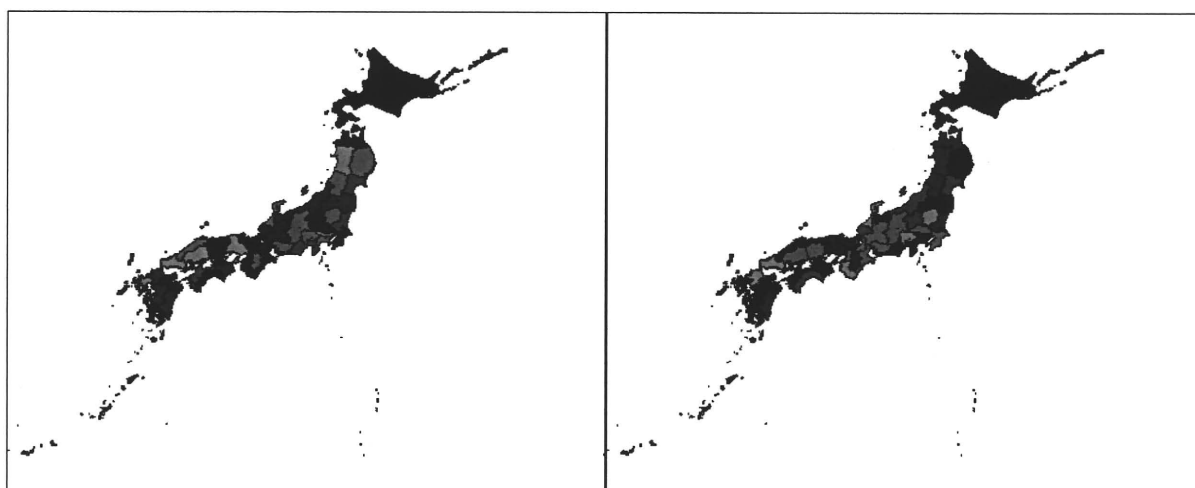
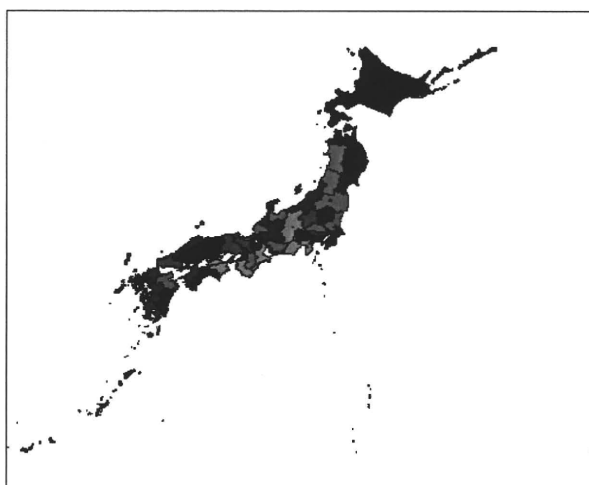


図12 再生産年齢有配偶離婚率（年齢別有配偶離婚率合計値）



35 歳未満未婚女性の同棲経験割合については、出生動向基本調査の個票を都道府県別に集計した。2002 年の第 12 回と 2005 年の第 13 回データをプールして集計している。しかし、都道府県別に集計すると標本数が少ない県も多く信頼性に欠ける。そこで、GeoDa のファンクションを利用し、いくつかの方法でスムージングを試みた結果も示した。図の左上は実数に基づいた同棲経験割合である。右上は Empirical Bayes 法でスムージングした結果、左下は Spatial rate 法でスムージングした結果、右下は Spatial empirical Bayes 法でスムージングした結果である。いずれのスムージングも、分子となる同棲経験者数と分母となる 35 歳未満女性人口を用いて算出している。Spatial rate 法と Spatial empirical Bayes 法は、空間を特定するウェイト行列を使用する。同棲経験者の多くはその後婚姻に至っている。従っ

図 13 35 歳未満未婚女性の同棲経験割合（出生動向基本調査）

調査データの実数ベース(左上) Empirical Bayes でスムージングした結果(右上)

Spatial rate でスムージングした結果(左下) Spatial empirical Bayes でスムージングした結果
(右下)

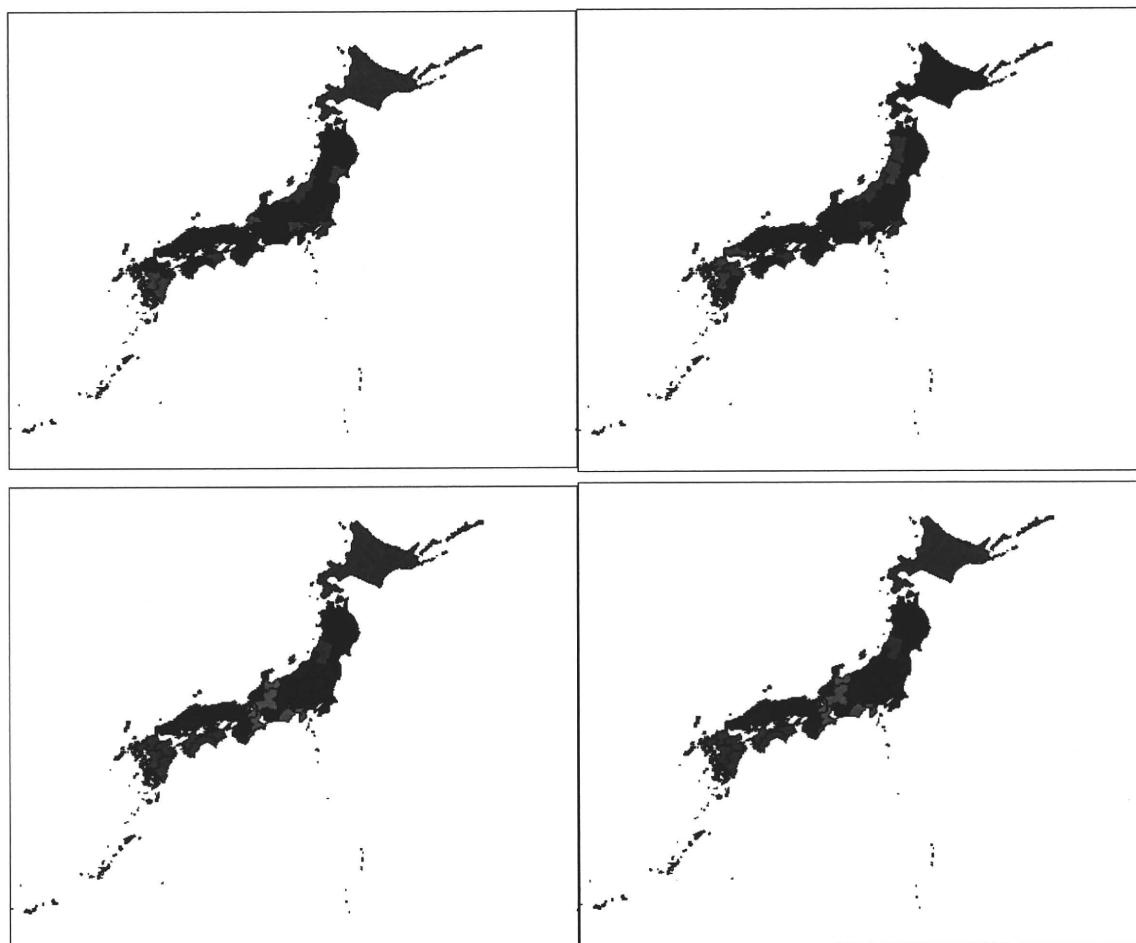
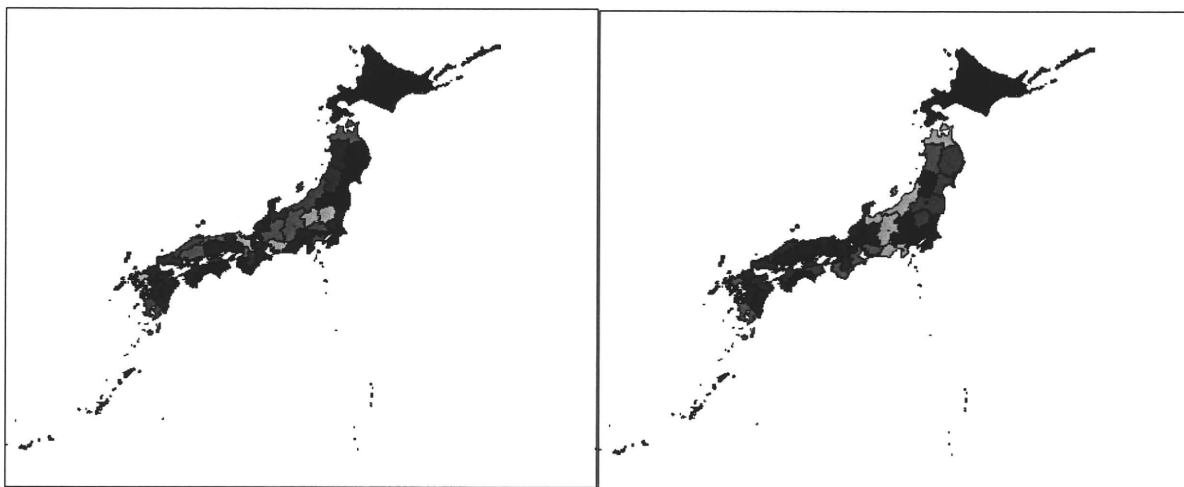


図 14 全女性の同棲経験割合(毎日新聞調査)

調査データの実数ベース(左) Spatial empirical Bayes でスムージングした結果(右)



て、未婚の同棲経験者だけでは全体の同棲経験の動向を把握しきれない可能性がある。そこで、既婚者にも同棲経験を尋ねている第1回人口・家族・世代調査(毎日新聞社人口問題調査会)の個票をつかって、全女性の同棲経験割合も算出した。同じく実数ベースと Spatial empirical Bayes 法でスムージングした結果を示す。未婚者に限定した結果と違い、関東地域で高い割合を示し、未婚者では高かった九州では低くなっている。ここから考えられることは、両地域で同棲の意味づけが違うという可能性である。関東地域では同棲は結婚の前段階として経験されているのに対し、九州では必ずしも結婚に結びつかない同棲が多いということかもしれない。いずれのサンプルでも関西近辺では同棲経験率が低いという傾向がみられる。

平行等尺度プロット (Parallel coordinate plot) による変数間の関連性

TFR 反転指標 (RAT) が、他の変数とどのように関連しているかを、平行等尺度プロット (Parallel coordinate plot) で確認しておく(図 15)。隣り合う変数同士の相関関係をおおまかに見ることができると同時に、変数内部の分布についても視覚的に把握することができる。

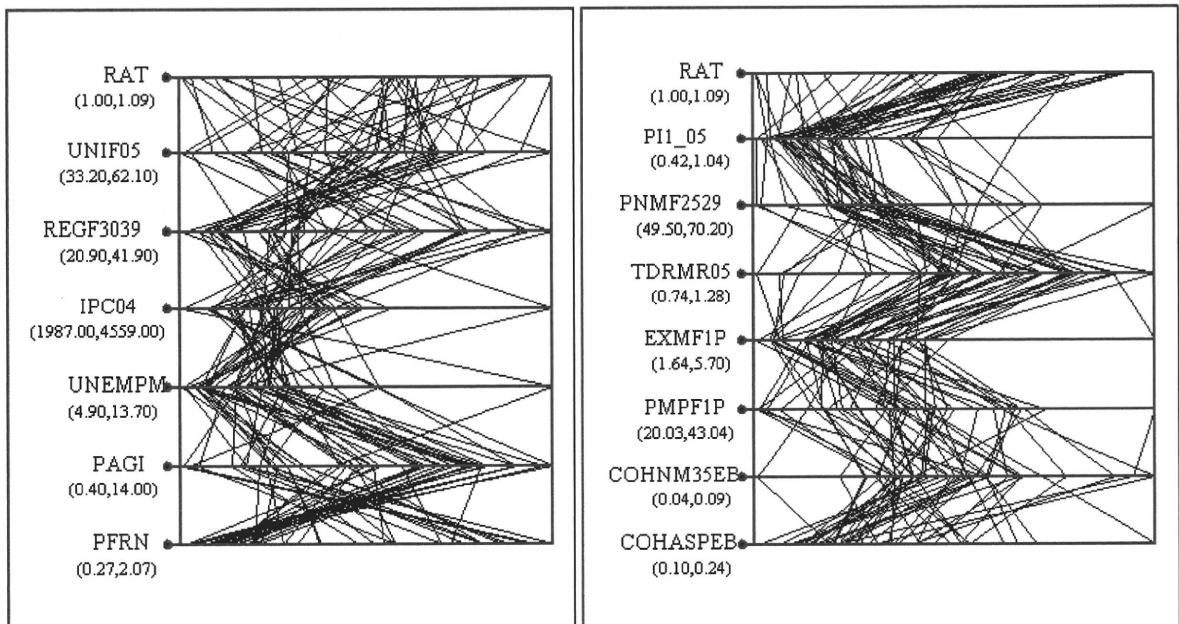
左の図は、TFR 反転指標と社会経済状況に関する変数(女性の大学進学率(UNIF05)、30代女性の正規雇用者割合(REGF3039)、2004年1人あたり所得(IPC04)、男性の失業率(UNEMPM)、農業人口割合(PAGI)、外国人割合(PFRN))の関係を見たものである。変数名の下のカッコ内の数字は指標の範囲を示す。

TFR 反転指標と進学率の関係ははっきりしないが、進学率と女性の正規雇用者割合、女性の正規雇用者割合と所得、所得と失業率には逆相関の関係があるように見える。

右の図は、TFR 反転指標 (RAT) と第二の人口転換において注目されている変数 (晩産化指標(PI1_05)、20 代後半未婚率(PNMF2529)、再生産年齢有配偶離婚率(TDRMR05)、第 1 子婚外出生率割合(EXMF1P)、第 1 子婚前妊娠出生率割合(PMPF1P)、35 歳未満未婚女性の同棲経験割合(COHNM35EB)、全女性の同棲経験割合(COHASPEB)) の関係を見たものである。

図 15 平行等尺度プロット

TFR 反転指標と社会経済状況変数(左)および TFR 反転指標と第二人口転換関連変数(右)



TFR 反転指標の空間回帰分析

TFR 反転指標が、他の家族形成変数および社会経済変数とどのように関連しているのかを検証するために、従属変数を都道府県別の TFR 反転指標とした空間回帰分析を試みた。

説明変数については、社会経済変数のみを入れたモデル (ベースモデル) と、第二人口転換関連変数を加えたモデルを推定する。

さらに、それぞれについて、残差が独立であることを仮定する通常の回帰分析と残差に空間的従属性(spatial dependence)が存在していることを仮定する、空間ラグモデルと空間誤差モデルを推定し、いずれのモデルがより適合的かを検証する。以下では結果の要約を示す。

1. ベースモデル

(1) 通常の重回帰 (OLS)

出力結果 1 の SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION に結果を示す。まず推定された係数を見てみよう。農業地域 PAGI の係数が -0.0026 で有意である。農業地域で上昇率が少ないことが分かる。男性失業率は有意ではなかった。モデルの適合度を示す Akaike info criterion は -223.115 である。

回帰診断の結果を見てみると、MULTICOLLINEARITY CONDITION NUMBER は 10.56 であった。通常 20 以上が注意する目安になっているので、今回は問題ないと考えられる。

誤差項の正規性 (尖度と歪度) は Jarque-Bera 検定によって検証できる。ここでは帰無仮説が棄却されないため、正規性を仮定してもよいという判断になる。以下の分散と空間依存性に関する検定はこの正規性が保持されていることが条件となる。

GeoDa では、誤差項の等分散性の検定として、Breusch-Pagan 検定、Koenker-Bassett 検定、White 検定の結果を出力する。仮説の棄却は分散不均一を意味する (BLUE ではない)。ここでは、Koenker-Bassett 検定の確率が 0.1 を切っているため棄却され、分散不均一の可能性を示す。空間依存性が存在する場合は、等分散性が影響をうけるため、それを示唆している可能性がある。

続いて、空間依存性の診断結果を見てみよう。GeoDa は空間依存性を評価する 6 つの検定結果を出力する。 0.421 という Moran's I は有意であり、強い空間依存性があることを示している。すなわち OLS の仮定 (誤差項が $iid : e \sim iid(0, \sigma^2)$, "Independent Identically Distributed") が破れているため、OLS が使えないことを意味する。

線形モデルにおける空間依存性を検証するその他の検定としては、以下の 5 つの検定がある。空間的ラグが存在しないことをテストする Lagrange Multiplier 検定 (Lagrange Multiplier (lag))、誤差の空間的依存性を検定する Lagrange Multiplier 検定 (Lagrange Multiplier (error))、他方の存在に対する頑健性を示す統計量として、Robust LM (lag) と Robust LM (error) である。前者は空間誤差依存性がないことを想定して空間ラグ依存性を検定し、後者は逆に空間ラグ依存性がないことを想定して空間誤差依存性を検定する。最後の SARMA (spatial autoregressive moving average) は、Lagrange Multiplier (error) + Robust LM (lag) である (Portmanteau 検定)。ここでは Lagrange Multiplier (error) が棄却されているため、空間依存性の存在が示唆される。従って通常の OLS は使えない。

SPATIAL DEPENDENCE の診断

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.421923	4.4728596	0.0000077 (iidではない(dependenceが存在))
Lagrange Multiplier (lag)	1	0.8901202	0.3454447 (有意でない)
Robust LM (lag)	1	0.3733867	0.5411640 (有意でない)
Lagrange Multiplier (error)	1	15.1608237	0.0000987 (有意)
Robust LM (error)	1	14.6440902	0.0001298 (有意)
Lagrange Multiplier (SARMA)	2	15.5342104	0.0004234 (有意)

以上の診断から、空間分析を試みる必要性が示唆される。この場合は空間ラグモデルよりも空間誤差モデルが適格的であるということになる。ちなみに、もし、空間ラグモデルのラグランジュ乗数も有意だった場合、ラグランジュ乗数がより大きいほうのモデルが適格的と考えられる。しかしこのような適合性を基にした判断は、通常、空間依存性の存在に関する理論的メカニズムが不明瞭な際に行われるものであり、今回のように、従属変数が出生力である場合は、出生力そのものが空間的に依存していると考えerよりも（空間ラグモデル）、出生力の規定要因に依存メカニズムが存在すると考えた方が現実的である。従って、空間誤差モデルを採用することが適切であると思われる。

(2) 空間ラグモデル

前節の空間依存性の検定で、ラグモデルが適格的ではないことが示唆されていたが、念のため空間ラグモデルを推定し、推定結果からそれを確認してみたい。出力結果1の SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION に結果を示す。

今回の結果では推定に用いられたウェイト行列として、**q1.GAL**というファイル名が示されている。これは近隣地域を特定する行列であり、あらかじめ分析の前に準備しておく必要がある。今回はqueen方式（前後左右、および対角で接する地域、将棋で言えば「角」）を採用した。他にもrook方式（前後左右に接する地域、将棋で言えば「飛車」）や、地域の中心点からの距離を決め、その範囲内を近隣地域と定義する方法などがある。

空間ラグモデルのAkaike info criterion は-222.001で、OLSより適合度が落ちている。回帰分析の推定結果をみると係数に新たにW_RATが追加されている。この変数の係数(Rho)が0でない場合は空間依存性が存在していることを意味し、周辺地域から当該地域への平均的な影響力を示す。ここではW_RAT の係数は有意ではない ($p=0.34$)。Likelihood Ratio Test も有意でない。従って、上記の診断通り、空間ラグモデルはこのデータに対しては有効ではないことが分かる。

(3) 空間誤差モデル

続いて空間誤差モデルの結果を見てみよう。出力結果1のSUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATIONに結果を示す。ウェイト行列には同じくqueen方式の行列を使用している。Akaike info criterionは-241.064でOLSよりよくなっている。

空間誤差モデルでは、空間的に相関している誤差項として LAMBDA の係数が推定されている。

LAMBDA の係数は 0.88370 で有意性も高い ($p<0.001$)。これは周囲が1単位上がると、当該地域も0.883押上げられることを意味している。Likelihood Ratio Test も有意である

($p=0.000023$)。Breusch-Pagan testによれば、等分散性も棄却されない。空間誤差モデルは使用しているデータにある程度適合的であると言える。その結果、男性の失業率の係数 (-0.00439)、農業者割合 (-0.00393) がともに有意となった。男性の失業率 (UNEMPM) が高く、農業者割合 (PAGI) が高いほど出生率の回復が弱い、そのような構造的要因以外に地域およびその周辺に特有な地域効果の存在が考えられる (誤差項の相関として検出)。仮に隣接する地域で社会経済構造が違っていても、双方をカバーする共通する効果の影響を受けている可能性がある。価値観や子育ての方法に関する伝播効果の存在を示唆する結果である。

出力結果 1 : 地域の経済特性 (男性失業率と農業割合) のみを入れた base モデル

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : **anadata2**
 Dependent Variable : **RAT** Number of Observations: 47
 Mean dependent var : 1.04481 Number of Variables : 3
 S.D. dependent var : 0.0236169 Degrees of Freedom : 44

R-squared : 0.198344 F-statistic : 5.44318
 Adjusted R-squared : 0.161905 Prob(F-statistic) : 0.00772222
 Sum squared residual: 0.0210151 Log likelihood : 114.557
 Sigma-square : 0.000477616 Akaike info criterion : -223.115
 S.E. of regression : 0.0218544 Schwarz criterion : -217.564
 Sigma-square ML : 0.00044713
 S.E of regression ML: 0.0211455

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	1.073068	0.01458379	73.57952	0.0000000
UNEMPM	-0.001369736	0.002091753	-0.6548268	0.5159879
PAGI	-0.002679528	0.0008895013	-3.012394	0.0042858

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 10.56096

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	0.4690646	0.7909407

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	3.793629	0.1500458
Koenker-Bassett test	2	4.798237	0.0907979

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	5.174311	0.3949797

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : **q1.GAL** (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.421923	4.4728596	0.0000077
Lagrange Multiplier (lag)	1	0.8901202	0.3454447
Robust LM (lag)	1	0.3733867	0.5411640
Lagrange Multiplier (error)	1	15.1608237	0.0000987
Robust LM (error)	1	14.6440902	0.0001298
Lagrange Multiplier (SARMA)	2	15.5342104	0.0004234

=====
END OF REPORT
=====

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : **anadata2**
Spatial Weight : **q1.GAL**
Dependent Variable : **RAT** Number of Observations: 47
Mean dependent var : 1.04481 Number of Variables : 4
S.D. dependent var : 0.0236169 Degrees of Freedom : 43
Lag coeff. (Rho) : 0.0157545

R-squared : 0.213316 Log likelihood : 115
Sq. Correlation :- Akaike info criterion : -222.001
Sigma-square : 0.000438779 Schwarz criterion : -214.6
S.E of regression : 0.0209471

Variable	Coefficient	Std.Error	z-value	Probability
W_RAT	0.01575452	0.01669354	0.9437495	0.3452976
CONSTANT	1.050691	0.02769823	37.9335	0.0000000
UNEMPM	-0.0002870923	0.002319444	-0.1237763	0.9014923
PAGI	-0.002786256	0.0008587034	-3.244725	0.0011758

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	4.213834	0.1216123

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : q1.GAL

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	0.8860615	0.3465469

=====
END OF REPORT
=====

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : **anadata2**
Spatial Weight : **q1.GAL**
Dependent Variable : **RAT** Number of Observations: 47
Mean dependent var : 1.044815 Number of Variables : 3
S.D. dependent var : 0.023617 Degree of Freedom : 44
Lag coeff. (Lambda) : 0.883703

R-squared : 0.440486 R-squared (BUSE) :-
Sq. Correlation :- Log likelihood : 123.531756
Sigma-square : 0.000312 Akaike info criterion : -241.064
S.E of regression : 0.0176656 Schwarz criterion : -235.513069

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	1.103703	0.02700625	40.86843	0.0000000
UNEMPM	-0.00438454	0.002382717	-1.840143	0.0657471
PAGI	-0.003929645	0.0008949694	-4.390815	0.0000113
LAMBDA	0.8837032	0.0444231	19.89288	0.0000000

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	1.224374	0.5421638

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : q1.GAL

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	17.94867	0.0000227

----- END OF REPORT -----

2. ベースモデルに第二人口転換変数を加えたモデル

(1) 第二人口転換関連変数投入後の通常の重回帰 (OLS)

続いて、ベースモデルに第二人口転換に関連する変数を投入し、通常の重回帰モデルを推定した。ベースモデルにあった男性失業率と農業比率に加え、晩産化指標、婚外出生、有配偶離婚率、同棲経験率、女性進学率を投入した。出力結果2の SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION に結果を示す。

ベースモデルと同様に空間依存性を診断すると、OLS の仮定が破られており、空間分析を行う必要があることが分かる。空間ラグ、空間誤差ともに乗数検定が有意であり、第二人口転換に関わるような構造的要因を入れてもなお、空間的な依存性をすべて排除できないといえることができる。いずれのモデルもデータに適合しそうだが、乗数の値は誤差モデルのほうが高いので、誤差モデルがより適合的と考えられる。このようなモデル自体のパフォーマンスを比較する以前に、理論的にも空間誤差モデルが望ましいことは先ほども述べた。出生力が相関するのは、出生力そのものが近接地域に影響をあたえるというよりも、出生行動に影響をあたえる要素を近接地域で共有していると考えの方が現実的であろう。そこで、以下では空間誤差モデルの推定結果を示す。

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : q1.GAL (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.270334	3.3295434	0.0008700(有意)
Lagrange Multiplier (lag)	1	3.4629108	0.0627598(有意)
Robust LM (lag)	1	2.6234537	0.1052950(有意でない)
Lagrange Multiplier (error)	1	6.2238573	0.0126040(有意)
Robust LM (error)	1	5.3844003	0.0203176(有意)
Lagrange Multiplier (SARMA)	2	8.8473111	0.0119903(有意)

(2) 第二人口転換関連変数投入後の空間誤差モデル

出力結果2の SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION に結果を示す。係数の推定結果をみると、空間誤差の係数 LAMBDA は有意である。隣接地域が1単位上昇すると、当該地域が0.96上昇することを意味する。男性失業率(-)、農業比率(-)も有意である。

その他、第二人口転換関連で、出生率上昇に対して有意だったのは、婚外出生 (+)、離婚 (-) であった。晩産化指標、同棲経験率、女性の大学進学率は有意でなかった。農業地域や経済が弱い地域で回復も弱いというのはイタリアとも共通する。しかしイタリアで

見られたような、新しい家族形成パターンや女性の高学歴化がTFRの反転上昇に結びついて
いるとは断定するには不十分な結果である。

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	1.117885	0.04476563	24.97196	0.0000000
UNEMPM (男性失業率)	-0.01089179	0.003280203	-3.320462	0.0008988
PAGI (農業比率)	-0.00295942	0.001114604	-2.655131	0.0079279
PII_05 (晩産化指標)	0.03780074	0.04128096	0.9156944	0.3598272
EXMF1P (婚外出生)	0.02348232	0.007793949	3.012891	0.0025879
TDRMR05 (有配偶離婚率)	-0.07485026	0.02550311	-2.934946	0.0033362
COHA (同棲経験率)	-7.150284e-006	0.0003157021	-0.02264884	0.9819303
UNIF05 (女性進学率)	0.0002869915	0.0007038415	0.4077502	0.6834572
LAMBDA	0.9630577	0.0165635	58.14335	0.0000000

出力結果 2 : 地域の経済特性 (男性失業率と農業割合) を入れたベースモデルに第二人口
転換に関連する変数を追加

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : **anadata2**
 Dependent Variable : **RAT** Number of Observations: 47
 Mean dependent var : 1.04481 Number of Variables : 8
 S.D. dependent var : 0.0236169 Degrees of Freedom : 39

R-squared : 0.350765 F-statistic : 3.01009
 Adjusted R-squared : 0.234235 Prob(F-statistic) : 0.0125799
 Sum squared residual : 0.0170195 Log likelihood : 119.513
 Sigma-square : 0.000436396 Akaike info criterion : -223.026
 S.E. of regression : 0.0208901 Schwarz criterion : -208.225
 Sigma-square ML : 0.000362116
 S.E of regression ML : 0.0190293

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.9920189	0.04766514	20.81225	0.0000000
UNEMPM	-0.008736775	0.003634978	-2.403529	0.0210902
PAGI	-0.0007712955	0.001542927	-0.499891	0.6199607
PII_05	0.03314493	0.04418017	0.7502219	0.4576232
EXMF1P	0.0246977	0.009886302	2.498174	0.0168080
TDRMR05	-0.01300876	0.03606944	-0.3606589	0.7203012
COHA	-8.898346e-005	0.0004454101	-0.1997787	0.8426908
UNIF05	0.001066818	0.0007007929	1.522301	0.1360014

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 49.85271
 TEST ON NORMALITY OF ERRORS
 TEST DF VALUE PROB
 Jarque-Bera 2 0.2034907 0.9032596

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS