

若年層の地域間移動に地域属性が与える影響に ついてのクロスセクション分析

福井 昭吾

概要

日本国内における地方から都市部への若年層の集中は地方経済を停滞させ少子化を加速する可能性があることから、重大な問題として認識されている。本研究では、その集中の要因を探るため、市区町村を対象に若年層の転入超過数および転入超過率を被説明変数、人口・労働環境・産業構造などの地域属性を説明変数とするモデルを構築しクロスセクション分析を行った。実際の分析では、被説明変数における外れ値の影響を軽減するために分位回帰を行い、また、各地域の特性を総合的に捉えるために主成分回帰を導入した。分析の結果、都市化の進んだ市区町村ほど転入超過数・転入超過率が高い傾向がある一方、一部の市区町村では都市化が進んでいないにもかかわらず転入超過率が高いことを示した。

キーワード 若年層の地域間移動、分位回帰、主成分回帰

1 はじめに

現在、地方から都市部への若年層の集中は、日本国内における大きな問題として認識されている。若年層の都市部への集中は、地方経済を衰退させるだけでなく少子化を加速させると考えられる。経済活動は人々の労働・消費といった経済的な行動により成り立っているため、人口が多い地域ほど経済活動は活発であり経済の規模は大きくなる。地方から都市部への若者の集中は地方における若者の人口を減少させるため、地方の経済活動は停滞し経済の規模は縮小することになる。また、地方から都市部への若者の集中は少子化を加速させうる。福井(2018)では、出生率の要因について総務省統計局「全国消費実態調査」の経済圏を対象に分析を行い、各経済圏の都市化の程度と出生率との間に負の相関があることを示した。この結果

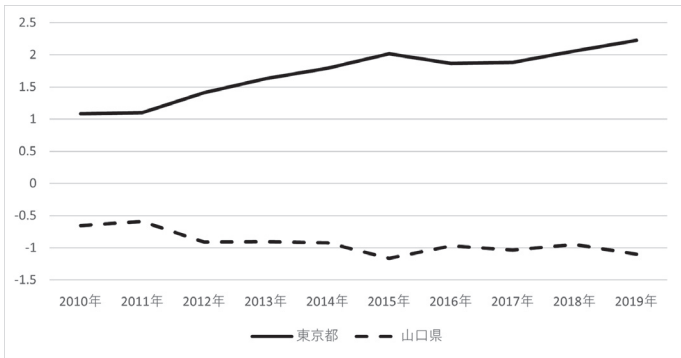


図1: 東京都および山口県における転入超過率の推移

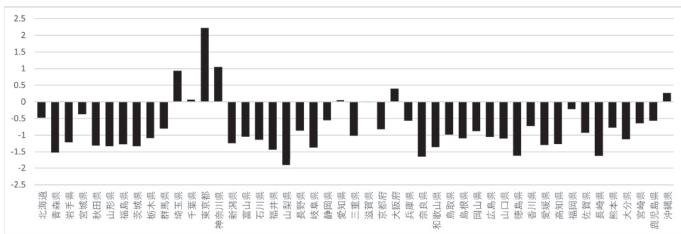


図2: 2019年時点における各都道府県の転入超過率

からただちに、都市化の進んだ地域ほど人々の出産しない傾向が強いと結論づけることはできない。例えば、出産を望まず仕事に専念したいと考える人が大企業への就職を望み都市部に移動するという関連もありうる。しかし、都市化と出生率の負の相関が、都市部に住むことによる出産への態度の変化を反映している可能性があるならば、都市部への若年層の集中は若者の出生行動を抑制し、さらなる少子化をもたらすことになるだろう。これらは、2014年の「ストップ少子化・地方元氣戦略」による報告書でも指摘されている。

図1は、東京都および山口県における20歳から39歳を対象とする転入超過率の推移を示している。また、図2には2019年時点の各都道府県の同転入超過率を示している。転入超過率は20歳から39歳までの転入超過数を20歳から39歳までの人口で割ったものである。図1より、東京都への若年人口の流入が増加傾向に

あること、一方で山口県からの若年人口の流出が増加傾向にあることが分かる。紙面の都合により省略しているが、地方にある都道府県の多くで若年人口の流出について山口県と同様に推移している。また、図2より、近年では東京都およびその周辺の都道府県に若年人口の流入が集中していることが分かる。このように、地方から東京都およびその周辺地域への若年人口の集中は、データからも明らかである。

若年層の地域間移動について考察している研究は多い。太田(2005)では地域における若者の雇用について考察する中で、都道府県を対象に新卒求人倍率・生涯賃金といった雇用の状況が純流出率に与える影響を分析し、新卒求人倍率が高い都道府県ほど若者の純流出率が低いことを示している。林他(2003)は京都府の市町村を対象に若年男性の純移動率と産業構造との関連を分析し、第3次産業就業者比率と純移動率との間に正の相関があることを示している。後河(2019)は、都道府県を対象に教育環境要因・経済的環境要因・アメニティ要因を表す各説明変数が若者の地元残留やUターンの希望にどう作用するかを分析し、雇用状況が良い都道府県ほど地元への就職を希望する若者の割合が高いことを明らかにした。

先行研究の多くは、特定の一つの地域や都道府県を対象として、地域属性と若年層の移動との関連を捉えるものが多い。そこで本研究では、全国の市区町村を対象として、地域属性と若年層の地域間移動との関連についてクロスセクション分析を試みる。都道府県を対象とする場合と比較して、市区町村の分析では利用可能なデータが制限される。例えば、後河(2019)ではモデルの説明変数の一つに消費支出を用いており、その値は総務省統計局の「家計調査」または「全国消費実態調査」の調査結果に基づいている。しかし、「家計調査」と「全国消費実態調査」のいずれも一部の市区町村のみが調査対象となることから、すべての市区町村を対象とする分析で消費支出をモデルの説明変数に含めることは難しい。一方で、すべての市区町村を対象とすることで、都道府県よりも人口・経済の規模が小さい地域も分析対象に含まれるため、都道府県単位の分析では捉えきれない事象を明らかにできる可能性がある。

本研究では、地方への移住・定住につながるような地域間移動を対象とする。大学等への進学に伴う移動は入学から卒業までの間に限られることが多いため、これらの移動を除外する。したがって、対象とする人の年齢を20歳以上から39歳以下に限定する。また、この研究では、地域属性の出生率への影響を分析した福井(2018)の手法を土台として分析を進める。

2 データの設定

以下の分析では、20歳以上39歳以下の転入超過数および転入超過率を用いて分析を進める。転入超過数とは一定期間内の各地域における転入数から転出数を引いたもので、総務省統計局「住民基本台帳人口移動報告年報」で年齢階級別の数値が公表されている。若年層を対象とする転入数はその地域における若年人口の多さに左右される。したがって、若年人口の少ない地域について、転入超過数のみで転入・転出の実状を捉えることは難しい。例えば、若年人口が2万人の地域と100人の地域とで10人の転出超過の重大さは後者の地域の方がより深刻だろうし、若年人口が100人の地域における10人の転入超過はその地域にとって非常に大きな転入といえるだろう。そこで、各地域における若年層の地域間移動を詳細に把握するため、転入超過数に加えて、転入超過数を若年人口で割った転入超過率を用いることにする。転入超過率の高い地域では若年人口の規模に対して多くの若者が転入していることから、転入超過数が小さくても、その地域の経済活動に対する影響は大きいといえよう。本研究では、20歳以上39歳以下の人々を対象とするため、転入超過率を(20歳以上39歳以下の転入超過数) / (20歳以上39歳以下の人口)と計算する。

各種地域属性の中から、転入超過数および転入超過率の説明要因を設定しよう。先行研究に基づき、若年人口・人口密度・失業率・賃金・産業別就業者比率を説明要因として設定する。ただし、人口密度は全年齢を対象とする可住地人口密度を用いる。これに、各市区町村の経済的な規模、生活環境、および、家庭環境を表す要因を加える。経済的規模を表す要因として売上金額(対数)および事業所当たり年間商品販売額を、生活環境を表す要因として10万人当たり大型小売店数・人口当たり病院数・保育施設当たりの児童数を、家庭環境を表す要因として核家族世帯割合を、それぞれ説明要因として設定する。

表1は、転入超過数、転入超過率、およびその説明要因である変数について、分析で用いるデータの出所をまとめたものである。今回の分析では、指定のない限り2015年時点のデータを用いる。分析の対象は、2015年における国内の全ての市町村および東京都内の区のうち、データに非数値を含む市区町村を除いた1728市区町村とする。売上金額および事業所当たり年間商品販売額は転入超過数および転入超過率からの影響が考えられる。その場合、被説明変数から説明変数への逆因果によりパラメータ推定量の一致性が満たされない可能性がある。そこで、逆因果

表 1: モデルの変数およびデータの出所

変数名	単位	出所および計算方法
転入超過数	人	「住民基本台帳人口移動報告年報」における 20歳以上39歳以下転入超過数を導出
転入超過率	%	転入超過数 ÷ 若年人口
若年人口	万人	「住民基本台帳に基づく人口、人口動態及び世 帯数調査」より20歳以上39歳以下人口を導出
人口密度	人	「社会・人口統計体系」の 可住地面積 1km ² 当たり人口密度
失業率	%	「社会・人口統計体系」の完全失業率
所得	千円	「社会・人口統計体系」の課税対象所得 (納税義務者1人当たり)
第1次産業就業者比率	%	「社会・人口統計体系」の第1次産業就業者比率 (対就業者)
第2次産業就業者比率	%	「社会・人口統計体系」の第2次産業就業者比率 (対就業者)
第3次産業就業者比率	%	「社会・人口統計体系」の第3次産業就業者比率 (対就業者)
売上金額 (対数)	百億円 (対数)	「経済センサスー活動調査」における 売上金額 (2013年) の対数
事業所当たり年間商品販売額	百万円	「社会・人口統計体系」の商業年間商品販売額 (卸売業+小売業) (事業所当たり、2013年)
10万人当たり大型小売店数	店	「社会・人口統計体系」の大型小売店数 (人口10万人当たり、2014年)
10万人当たり病院数	施設	「社会・人口統計体系」の一般病院数 (人口10万人当たり)
保育施設当たり児童数	人	「社会福祉施設等調査」の保育所等在所見数 ÷ 「社会福祉施設等調査」の保育所等数
核家族世帯割合	%	「社会・人口統計体系」の核家族世帯割合

に伴う問題を回避するため、売上金額および事業所当たり年間商品販売額について2013年のデータを用いる。以下の分析では、産業就業者比率として第2次産業就業者比率および第3次産業就業者比率を使い、第1次産業就業者比率を除いている。これら三つの産業就業者比率をすべて説明変数とする場合、各市区町村の三つの産業就業者比率の総和がほぼ1になることから、その総和と定数項の相関係数がほぼ1に等しくなり回帰式は完全多重共線性に極めて近い状態になる。多重共線性とは、ある一つの説明変数と別の説明変数からなる線形結合との間に相関がある状態をいう。特に、ある一つの説明変数と別の説明変数の線形結合との相関が1ま

表 2: 各データの代表値

代表値	平均値	標準偏差	第 1 四分位	第 3 四分位
転入超過数	0.027	648.160	-131	-7
転入超過率	-1.375	2.342	-2.546	-0.294
若年人口	1.719	4.732	0.151	1.435
人口密度	1,390.079	2,566.447	250.075	1,236.525
失業率	4.001	1.351	3.200	4.700
所得	2,797.993	524.430	2,491.075	2,981.825
第 1 次産業就業者比率	11.018	10.294	2.900	16.317
第 2 次産業就業者比率	25.159	8.099	19.140	30.837
第 3 次産業就業者比率	61.333	8.885	55.237	67.790
売上金額 (対数)	2.564	1.871	1.284	3.761
事業所当たり年間商品販売額	189.033	345.209	82.400	220.125
10 万人当たり大型小売店数	9.855	8.945	0.000	14.760
10 万人当たり病院数	6.704	7.343	0.000	9.125
保育施設当たり児童数	82.843	37.373	60.562	103.092
核家族世帯割合	56.177	6.565	52.017	60.340

表 3: 転入超過数上位 5 市区町村

順位	市区町村名	転入超過数
1	大阪府 大阪市	11674
2	神奈川県 川崎市	9377
3	愛知県 名古屋市中区	7317
4	東京都 世田谷区	5792
5	福岡県 福岡市	5630

表 4: 転入超過数下位 5 市区町村

順位	市区町村名	転入超過数
1	東京都 八王子市	-2713
2	福岡県 北九州市	-2361
3	神奈川県 横須賀市	-1916
4	奈良県 奈良市	-1151
5	山口県 下関市	-1025

たは -1 となる状態を完全多重共線性といい、完全多重共線性のもとでは最小二乗法による推定値を計算することができない。この完全多重共線性の問題を避けるため、説明変数から第 1 次産業就業者比率を除くことにする。

表 2 は、これらのデータの基本的な代表値を計算したものである。表 2 より、多くの市区町村で転入超過数・転入超過率がマイナスであることが分かる。表 3 および表 4 はそれぞれ転入超過数の上位 5 市区町村と下位 5 市区町村を、表 5 および表 6 はそれぞれ転入超過率の上位 5 市区町村と下位 5 市区町村を示している。転

表 5: 転入超過率上位 5 市区町村

順位	市区町村名	転入超過率
1	鹿児島県 十島村	25.000
2	島根県 知夫村	13.846
3	鹿児島県 三島村	12.000
4	東京都 御蔵島村	11.579
5	東京都 小笠原村	10.283

表 6: 転入超過率下位 5 市区町村

順位	市区町村名	転入超過率
1	群馬県 南牧村	-13.000
2	長野県 天龍村	-12.195
3	千葉県 勝浦市	-11.504
4	秋田県 大湯村	-11.324
5	北海道 長万部町	-9.968

入超過数は、大都市ほど大きな値を取り地方都市ほどその値が小さくなる傾向がある。一方、転入超過率は地方の町村において極めて高いまたは低い値となっている。これらの町村では若年人口が小さいため、転入超過数の程度は小さくても転入超過率が大きく反応しているといえる。

3 実証分析

転入超過数または転入超過率を被説明変数とする各種モデルの推定を試みる。転入超過数と転入超過率のそれぞれについて、表 1 の説明変数をもとに重回帰分析を行った結果を表 7 に示している。なお、転入超過率の分母が若年人口であることを考慮し、転入超過率のモデルでは説明変数から若年人口を除いている。

転入超過数のモデルについて推定結果を確認しよう。以降、パラメータの統計的有意性に言及する際は、有意水準を 5% とする。若年人口および人口密度は転入超過数に対して有意に正の影響を及ぼし、若年人口の多い市区町村・人口密度の高い市区町村ほど転入超過数が多い傾向がある。失業率についてパラメータは有意に負であることから、失業率の高い地域ほど転入超過数は低い。一方、課税対象所得のパラメータは有意ではなく、所得の高さは転入超過数に影響しない。売上金額（対数）は転入超過数に負に作用しているが、事業所当たりの商品販売額は正に作用している。保育施設当たり児童数・10 万人当たり病院数・10 万人当たり大型小売店数はいずれもその影響は有意でない。核家族世帯割合が転入超過数に与える影響は有意に負となっている。産業就業者比率について、第 2 次産業就業者比率による影響はみられないが、第 3 次産業就業者比率が高い市区町村ほど転入超過数が小さい傾向がある。

転入超過率のモデルでは、人口密度が転入超過率に有意に正の影響を及ぼしてい

表 7: 重回帰分析のパラメータ推定値 (括弧内の数値は標準誤差)

	被説明変数	
	転入超過数	転入超過率
若年人口	83.164*** (2.722)	
人口密度	0.109*** (0.006)	0.0001*** (0.00003)
失業率	-24.058*** (8.527)	-0.134*** (0.043)
所得	-0.056* (0.032)	0.001*** (0.0002)
第2次産業就業者比率	3.044* (1.625)	-0.013 (0.008)
第3次産業就業者比率	-4.703*** (1.557)	0.028*** (0.008)
売上金額 (対数)	-116.053*** (8.861)	0.016 (0.040)
事業所当たり年間商品販売額	0.344*** (0.041)	0.0001 (0.0002)
10万人当たり大型小売店数	-0.978 (1.475)	0.003 (0.007)
10万人当たり病院数	2.588* (1.414)	-0.015** (0.007)
保育施設当たり児童数	0.299 (0.330)	0.006*** (0.002)
核家族世帯割合	-7.284*** (1.798)	-0.016* (0.009)
定数項	778.048*** (137.404)	-5.196*** (0.692)
サンプルサイズ	1,728	1,728
決定係数	0.592	0.196
自由度調整済み決定係数	0.589	0.191
残差標準誤差	415.668 (自由度 = 1715)	2.106 (自由度 = 1716)
F 値	207.014*** (自由度 = 12,1715)	38.110*** (自由度 = 11,1716)

*p<0.1; **p<0.05; ***p<0.01

る。また、失業率が低く課税対象所得が高い市区町村ほど転入超過率が高い傾向がある。第3次産業就業者比率は、転入超過数のモデルと異なり、転入超過率に対して有意に正の影響を及ぼしている。保育施設当たり児童数のパラメータ推定値が有意に正であり、保育施設当たりの児童数が多い地域ほど転入超過率が高いといえる。10万人当たり病院数は転入超過率に有意に負の影響を及ぼすことから、病院の充実度は転入超過率を低下させる要因となっている。

表2の転入超過数・転入超過率の代表値、および、表3-表6の値より、転入超過数と転入超過率の双方について外れ値が存在する可能性が高い。そこで、スチューデント化残差を使い外れ値を探索する。以下ではGreene (2020)に基づいて、スチューデント化残差の計算とそれによる外れ値の判定について説明する。残差のベクトルを \mathbf{e} 、観測対象 i の残差を e_i 、サンプルサイズを n 、説明変数の数を K とするとき、観測対象 i のスチューデント化残差 $e(i)$ は

$$e(i) = \frac{e_i}{\sqrt{1-h_{ii}}} / \sqrt{\frac{\mathbf{e}'\mathbf{e} - e_i^2 / (1-h_{ii})}{n-1-K}}$$

と計算される。ただし、 h_{ii} は

$$h_{ii} = \mathbf{x}'_i \left(\mathbf{X}'_{(i)} \mathbf{X}_{(i)} \right)^{-1} \mathbf{x}_i$$

と計算する。ここで、 \mathbf{x}_i は観測対象 i の説明変数の値からなるベクトル、 $\mathbf{X}_{(i)}$ は説明変数行列 \mathbf{X} から i 行目の \mathbf{x}_i を除いた行列である。理論上、スチューデント化残差は自由度 $n-1-K$ の t 分布に従う。そこで、標本が十分に大きい状況でスチューデント化残差を計算したとき、その絶対値が2を超えるような観測対象について、その被説明変数は外れ値である可能性が高いとみなすのである。

転入超過数および転入超過率について、表7の推定結果から市区町村ごとにスチューデント化残差を求めた。表8および表9は転入超過数のモデルにおけるスチューデント化残差の上位・下位5位の市区町村を、表10および表11は転入超過率のモデルにおけるスチューデント化残差の上位・下位5位の市区町村を、それぞれ示している。表8-表11より、転入超過数と転入超過率のいずれにおいても複数の地域でスチューデント化残差の絶対値が2を超えていることから、これらの地域における転入超過数または転入超過率は外れ値である可能性が高い。

外れ値を含むデータに対して最小二乗法による回帰分析を行う場合、その推定結果は外れ値に強い影響を受ける。そこで、外れ値からの影響を軽減するため、中央

表 8: 転入超過数のモデルにおけるスチューデント化残差上位 5 市町村

順位	市区町村名	残差
1	大阪府 大阪市	14.593
2	神奈川県 川崎市	14.564
3	東京都 江東区	7.020
4	東京都 世田谷区	6.832
5	愛知県 名古屋市	6.356

表 9: 転入超過数のモデルにおけるスチューデント化残差下位 5 市町村

順位	市区町村名	残差
1	東京都 八王子市	-9.226
2	福岡県 北九州市	-9.105
3	兵庫県 神戸市	-8.519
4	京都府 京都市	-8.082
5	神奈川県 横須賀市	-6.076

表 10: 転入超過率のモデルにおけるスチューデント化残差上位 5 市町村

順位	市区町村名	残差
1	鹿児島県 十島村	13.338
2	島根県 知夫村	7.434
3	鹿児島県 三島村	6.245
4	和歌山県 北山村	5.426
5	福島県 昭和村	5.364

表 11: 転入超過率のモデルにおけるスチューデント化残差下位 5 市町村

順位	市区町村名	残差
1	千葉県 勝浦市	-5.046
2	群馬県 南牧村	-4.788
3	秋田県 大潟村	-4.772
4	長野県 天龍村	-4.736
5	北海道 長万部町	-4.093

値回帰を含む分位回帰による推定を行う。分位回帰は、

$$P(y_i \leq \mathbf{x}'_i \boldsymbol{\beta}) < p \tag{1}$$

を満たすようなパラメータ $\boldsymbol{\beta}$ の値を推定値とする手法である*1。ただし、 y_i は観測対象 i の被説明変数である。この式の p を 0.5 と置く場合を中央値回帰という。分位回帰を推定する際に残差の絶対値を用いることから、最小二乗法による回帰と比較して外れ値が推定結果に与える影響は小さいとされる。

分位回帰で得られる回帰式は、被説明変数である観測値の集合をそれぞれの割合が $(1-p)$ 対 p となるように分割する。前者の集合では被説明変数の実績値が理論値よりも大きく、後者の集合では実績値が理論値よりも小さい。したがって、 p の値を変えることで、被説明変数の値が高い集団や低い集団において説明変数が与え

*1 分位回帰については、Koenker and Hallock (2001) 等を参照せよ。また、その推定に関わる計算については、Greene (2020) 等で説明されている。

表 12: 転入超過数を被説明変数とする分位回帰のパラメータ推定値 (括弧内の数値は標準誤差)

	($p = 0.2$)	($p = 0.5$)	($p = 0.8$)
若年人口	-23.456 (14.489)	32.097** (15.264)	120.746*** (21.372)
人口密度	0.022** (0.010)	0.045*** (0.011)	0.072*** (0.015)
失業率	-7.486*** (1.199)	-7.711*** (1.516)	-5.770*** (2.071)
所得	-0.009 (0.009)	-0.003 (0.010)	0.015 (0.010)
第 2 次産業就業者比率	0.176 (0.274)	-0.099 (0.314)	-0.981** (0.431)
第 3 次産業就業者比率	-1.344*** (0.271)	-1.217*** (0.265)	-1.074** (0.425)
売上金額 (対数)	-47.153*** (4.072)	-48.874*** (3.700)	-56.618*** (5.714)
事業所当たり年間商品販売額	0.160*** (0.033)	0.268*** (0.074)	0.306* (0.162)
10 万人当たり大型小売店数	-0.309 (0.193)	-0.387 (0.247)	-0.898** (0.406)
10 万人当たり病院数	0.684*** (0.192)	0.935*** (0.171)	0.336 (0.243)
保育施設当たり児童数	0.208*** (0.039)	0.142*** (0.043)	0.107** (0.043)
核家族世帯割合	-0.455 (0.342)	-1.636*** (0.412)	-3.364*** (0.630)
定数項	74.279** (37.614)	136.875*** (40.646)	220.948*** (50.459)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

る影響の違いを捉えることが可能である。

表 12 は転入超過数を被説明変数とするモデルについて、式 (1) の p を 0.2, 0.5, 0.8 とした場合の分位回帰の結果を示している。また、表 13 は転入超過率を被説明

表 13: 転入超過率を被説明変数とする分位回帰のパラメータ推定値 (括弧内の数値は標準誤差)

	($p = 0.2$)	($p = 0.5$)	($p = 0.8$)
人口密度	0.0001*** (0.00002)	0.0001*** (0.00002)	0.0001*** (0.00003)
失業率	-0.167*** (0.047)	-0.183*** (0.040)	-0.156*** (0.045)
所得	0.0003 (0.0002)	0.001*** (0.0002)	0.001*** (0.0003)
第2次産業就業者比率	-0.012 (0.008)	-0.012* (0.006)	0.005 (0.010)
第3次産業就業者比率	0.001 (0.009)	0.015** (0.007)	0.027*** (0.009)
売上金額 (対数)	0.459*** (0.031)	0.183*** (0.034)	-0.177*** (0.036)
事業所当たり年間商品販売額	0.0001 (0.0003)	0.001 (0.0004)	0.002* (0.001)
10万人当たり大型小売店数	0.014** (0.007)	0.003 (0.006)	0.004 (0.010)
10万人当たり病院数	0.001 (0.007)	-0.008 (0.006)	-0.023** (0.010)
保育施設当たり児童数	0.008*** (0.001)	0.007*** (0.001)	0.004** (0.002)
核家族世帯割合	0.017** (0.008)	0.002 (0.007)	-0.011 (0.010)
定数項	-5.731*** (0.687)	-4.518*** (0.526)	-4.012*** (0.804)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

変数とするモデルに対し、表 12 と同様に分位回帰を行った結果である。はじめに、 $p = 0.5$ とする中央値回帰の結果を考察する。転入超過数のモデルでは、10 万人当たり病院数と保育施設当たり児童数のパラメータが正で有意となったが、それ以外のパラメータの有意性は最小二乗法による結果と変わらない。また、有意となったパラメータの符号も最小二乗法の結果と同じである。ただし、パラメータの推定値

の絶対値について、最小二乗法で得られる値は中央値回帰よりも総じて大きい。これは、最小二乗法の推定結果が外れ値の影響を受けていることを反映していると考えられる。転入超過率のモデルでは、10万人当たり病院数のパラメータが有意でなくなっているが、それ以外のパラメータについての有意性・符号・値とも、中央値回帰と最小二乗法による回帰とで大きな差はない。転入超過率のモデルは、転入超過数のモデルほど外れ値の影響を受けていないが、一部のパラメータが有意でなくなる結果となった。次に、 $p = 0.2$ とする分位回帰の結果を見てみよう。転入超過数のモデルにおいて、中央値回帰と比較すると若年人口の影響が有意でなくなっている。すなわち、転入超過数の少ない市区町村では、若年人口は転入超過数に影響しないといえる。転入超過率のモデルでは、中央値回帰と異なり課税対象所得のパラメータが有意でなくなっており、転入超過率の低い市区町村では所得の大きさが転入超過率に影響しないことがわかる。最後に、 $p = 0.8$ とする分位回帰の結果を確認する。転入超過数のモデルでは、10万人当たり大型小売店数が有意に負の影響を及ぼしており、若年人口のパラメータの推定値が中央値回帰と比較して非常に大きくなっている。結果として、転入超過数の多い市区町村では、人口当たりの大型小売店数が多いほど転入超過数が小さく、若年人口が多いほどより転入超過数が増える傾向がある。転入超過率のモデルでは、人口密度・失業率・所得の影響は中央値回帰に近いが、売上金額(対数)のパラメータが負で有意となり、10万人当たり病院数もまた有意に負の影響を及ぼしている。したがって、転入超過率が高い市区町村では、売上金額が大きく病院数が充実するほど転入超過率が下降する傾向が新たに見られる。

地域属性と転入超過数・転入超過率の関連をさらに詳しく分析するために、主成分回帰を導入する*2。主成分回帰とは、説明変数に対して主成分分析を行い、その結果得た主成分を新たな説明変数として回帰分析を行う方法である。主成分は互いに無相関であることから、主成分回帰では多重共線性による問題が発生しない。そのため、説明変数における多重共線性による問題への対策として主成分回帰を用いることが多い。また、Jolliffe (2002) はたとえ多重共線性が問題とならない状況でも、本来の説明変数による回帰分析と比較して、主成分回帰は計算と解釈における優位性を持ちうると指摘している。主成分回帰の説明変数となる主成分は本来のモデルにおける説明変数を総合化したものであり、今回の分析では各市区町村における都市化の程度や労働環境といった地域的な特性を表す。したがって、どのような

*2 主成分回帰については、Jolliffe (2002) が詳しい。

特徴を持つ市区町村で転入超過数が多いか・転入超過率が高いかを解釈しやすくする。さらに、主成分間の相関がないため、被説明変数に対する各主成分の寄与度を容易に計算することができる。

以下の主成分回帰では、表 1 の説明変数に第 1 次産業就業者比率を加えて主成分得点を求めた後、中央値回帰および分位回帰を行う。なお、転入超過率のモデルでは引き続き若年人口を説明変数から除く。第 1 次・第 2 次・第 3 次産業就業者比率については、Aitchison (1983) の変換を適用する。第 1 次・第 2 次・第 3 次産業就業者比率の和がほぼ 1 に等しく主成分分析の固有値の一つが 0 となるため、主成分回帰ではこの固有値に対応する主成分を除く。

表 14 は転入超過数のモデルの説明変数から得た主成分得点の主成分負荷量と主成分寄与率である。また、表 15 は転入超過率のモデルについて同様に求めた主成分負荷量と主成分寄与率である。表 16 は転入超過数を被説明変数、主成分得点を説明変数とするモデルについて、式 (1) の p を 0.2, 0.5, 0.8 とした場合の分位回帰の結果を示している。表 17 は転入超過率を被説明変数、主成分得点を説明変数とするモデルについて、表 16 と同様の分位回帰を行った結果である。表 16 および表 17 には、主成分得点とパラメータの推定値から求めた寄与度を示している。この寄与度は、被説明変数の分散のうち各主成分得点の説明項の分散が占める割合であり、以下のように計算している。主成分回帰の式を

$$y_i = \alpha + \beta_1 a_{1,i} + \cdots + \beta_K a_{K,i}$$

とする。ただし、 y_i は被説明変数、 $a_{j,i}$ 、 $j = 1, \dots, K$ は第 j 主成分得点、 $\alpha, \beta_1, \dots, \beta_K$ はパラメータである。主成分得点間の相関が 0 であることからそれらの共分散もまた 0 である。これにより、

$$\begin{aligned} \text{Var}(y_i) &= \text{Var}(\beta_1 a_{1,i}) + \cdots + \text{Var}(\beta_K a_{K,i}) \\ &= \beta_1^2 \text{Var}(a_{1,i}) + \cdots + \beta_K^2 \text{Var}(a_{K,i}) \end{aligned}$$

となる。さらに、 λ_j を第 j 主成分の固有値とすると、 $\text{Var}(a_{j,i}) = \lambda_j$ となる。この結果、第 j 主成分得点による寄与度を $\beta_j^2 \lambda_j / \sum_k \beta_k^2 \lambda_k$ と定義し、分位回帰によるパラメータの推定値を β_j に、主成分分析より得た固有値を λ_j にそれぞれ代入することで実際の寄与度を計算している。

転入超過数と転入超過率に対する分位回帰で求めた寄与度に基づいて各被説明変数への影響が大きいと考えられる主成分を選び、それらの主成分を対象に分析結果を考察する。転入超過数の場合、 $p = 0.2$ の分位回帰では第 5・6・8 主成分の影響が

表 14: 転入超過数モデルの説明変数の主成分負荷量・主成分寄与率

	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分	第 5 主成分	第 6 主成分
若年人口	-0.548	-0.243	0.274	0.322	-0.290	0.452
人口密度	-0.772	-0.217	0.327	0.161	0.033	-0.185
失業率	-0.235	0.607	-0.012	0.428	0.527	0.203
所得	-0.747	-0.389	-0.050	-0.160	-0.018	-0.288
第 1 次産業就業者比率	0.929	-0.135	-0.177	-0.042	-0.053	0.143
第 2 次産業就業者比率	-0.801	0.243	0.164	-0.098	0.002	-0.025
第 3 次産業就業者比率	-0.871	-0.001	0.154	0.176	0.094	-0.234
売上金額 (対数)	-0.797	-0.003	-0.137	0.036	-0.230	0.309
事業所当たり	-0.539	-0.483	-0.431	-0.153	0.291	0.013
年間商品販売額 10 万人当たり	-0.575	-0.080	-0.575	-0.107	0.239	0.247
大型小売店数 10 万人当たり病院数	0.131	-0.097	-0.534	0.706	-0.294	-0.266
保育施設当たり児童数	-0.531	0.407	-0.217	-0.265	-0.389	0.178
核家族世帯割合	-0.322	0.655	-0.238	-0.147	-0.166	-0.323
寄与率	0.418	0.119	0.091	0.077	0.064	0.062

	第 7 主成分	第 8 主成分	第 9 主成分	第 10 主成分	第 11 主成分	第 12 主成分
若年人口	0.361	-0.035	0.019	0.168	-0.118	-0.006
人口密度	-0.018	0.210	-0.223	-0.085	0.083	-0.282
失業率	0.066	0.202	0.113	-0.105	-0.073	0.009
所得	0.134	0.118	0.089	-0.229	-0.277	0.099
第 1 次産業就業者比率	0.164	0.131	-0.028	-0.113	-0.010	-0.070
第 2 次産業就業者比率	-0.329	-0.256	0.215	0.101	-0.134	-0.118
第 3 次産業就業者比率	0.036	0.022	-0.165	0.102	0.154	0.246
売上金額 (対数)	-0.040	-0.170	0.130	-0.324	0.210	0.025
事業所当たり	0.119	0.148	0.304	0.178	0.138	-0.064
年間商品販売額 10 万人当たり	-0.037	-0.197	-0.390	0.020	-0.100	-0.029
大型小売店数 10 万人当たり病院数	-0.166	-0.008	0.047	0.030	-0.039	-0.024
保育施設当たり児童数	-0.191	0.453	-0.048	0.081	-0.022	0.032
核家族世帯割合	0.478	-0.137	0.009	0.022	0.033	-0.090
寄与率	0.046	0.038	0.032	0.021	0.017	0.014

表 15: 転入超過率モデルの説明変数の主成分負荷量・主成分寄与率

	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分	第 5 主成分	第 6 主成分
人口密度	-0.760	-0.202	0.336	0.279	0.087	0.010
失業率	-0.243	0.616	-0.121	0.423	-0.542	0.033
所得	-0.749	-0.430	0.050	-0.058	0.153	0.251
第 1 次産業就業者比率	0.935	-0.119	-0.195	-0.135	-0.040	0.043
第 2 次産業就業者比率	-0.813	0.221	0.193	-0.033	0.021	-0.205
第 3 次産業就業者比率	-0.870	-0.009	0.158	0.278	0.051	0.130
売上金額 (対数)	-0.782	-0.001	-0.179	-0.102	0.047	-0.308
事業所当たり	-0.543	-0.549	-0.327	-0.119	-0.272	0.199
年間商品販売額						
10 万人当たり	-0.588	-0.151	-0.524	-0.178	-0.344	-0.042
大型小売店数						
10 万人当たり病院数	0.134	-0.106	-0.643	0.600	0.417	-0.116
保育施設当たり児童数	-0.545	0.373	-0.215	-0.402	0.234	-0.298
核家族世帯割合	-0.346	0.611	-0.240	-0.187	0.269	0.560
寄与率	0.431	0.126	0.097	0.081	0.069	0.056

	第 7 主成分	第 8 主成分	第 9 主成分	第 10 主成分	第 11 主成分
人口密度	-0.212	-0.226	-0.073	-0.033	-0.284
失業率	-0.208	0.108	-0.081	0.108	0.010
所得	-0.122	0.086	-0.084	0.350	0.102
第 1 次産業就業者比率	-0.149	-0.044	-0.147	0.047	-0.070
第 2 次産業就業者比率	0.288	0.239	0.216	0.097	-0.115
第 3 次産業就業者比率	-0.022	-0.162	0.047	-0.183	0.243
売上金額 (対数)	0.148	0.091	-0.461	-0.075	0.019
事業所当たり	-0.146	0.310	0.090	-0.198	-0.066
年間商品販売額					
10 万人当たり	0.199	-0.384	0.099	0.089	-0.027
大型小売店数					
10 万人当たり病院数	0.021	0.056	0.075	0.029	-0.023
保育施設当たり児童数	-0.440	-0.032	0.131	-0.006	0.033
核家族世帯割合	0.111	-0.007	-0.070	-0.050	-0.090
寄与率	0.042	0.034	0.029	0.020	0.015

表 16: 転入超過数を被説明変数、主成分得点を説明変数とする分位回帰の推定値・寄与度（括弧内の値は標準誤差）

	(p = 0.2)		(p = 0.5)		(p = 0.8)	
	推定値	寄与度	推定値	寄与度	推定値	寄与度
第 1 主成分	26.063** (11.817)	0.170	-59.102*** (12.923)	0.353	-179.801*** (18.801)	0.396
第 2 主成分	-18.622 (13.030)	0.025	-100.795*** (14.844)	0.292	-207.217*** (23.466)	0.150
第 3 主成分	-19.587 (15.263)	0.021	47.174** (19.561)	0.049	181.034*** (32.084)	0.088
第 4 主成分	-39.222** (19.404)	0.071	48.597** (22.034)	0.044	194.972*** (31.520)	0.086
第 5 主成分	63.482*** (20.782)	0.155	-0.923 (25.377)	0.000	-119.708*** (38.554)	0.027
第 6 主成分	-93.931*** (35.592)	0.326	26.690 (39.738)	0.011	215.311*** (54.108)	0.084
第 7 主成分	-44.987 (32.711)	0.056	79.379** (34.607)	0.070	274.960*** (48.603)	0.102
第 8 主成分	66.341*** (10.152)	0.101	80.572*** (12.038)	0.060	93.427*** (20.207)	0.010
第 9 主成分	-15.572 (12.575)	0.005	-7.088 (16.474)	0.000	-24.143 (30.363)	0.001
第 10 主成分	19.513 (29.941)	0.005	110.441*** (30.616)	0.063	240.394*** (45.374)	0.036
第 11 主成分	8.841 (22.553)	0.001	-39.759 (26.074)	0.006	-128.922*** (39.689)	0.008
第 12 主成分	-88.013*** (18.483)	0.066	-122.114*** (21.446)	0.051	-167.089*** (32.083)	0.012
定数項	-163.729*** (13.326)		-16.045 (15.541)		184.016*** (22.039)	

*p<0.1; **p<0.05; ***p<0.01

大きく、 $p = 0.5$ および $p = 0.8$ の分位回帰では、第 1・2 主成分の影響が大きい。表 14 の主成分負荷量より、第 1 主成分の値が小さい市区町村は都市化が進んでいると解釈できる。第 2 主成分は、失業率が高く、所得が低く、さらに事業所当たりの年間商品販売額が大きい地域ほどその値が高い傾向があることから、労働環境を表すといえる。また、第 5 主成分は失業率の高さ、第 6 主成分は若年人口の多さ、第 8 主成分は保育施設の逼迫の度合いをそれぞれ表すと考えられる。以上の解釈に

表 17: 転入超過率を被説明変数、主成分得点を説明変数とする分位回帰の推定値・寄与度（括弧内の値は標準誤差）

	(p = 0.2)		(p = 0.5)		(p = 0.8)	
	推定値	寄与度	推定値	寄与度	推定値	寄与度
第 1 主成分	-0.557*** (0.017)	0.767	-0.444*** (0.028)	0.780	-0.349*** (0.047)	0.504
第 2 主成分	-0.115*** (0.030)	0.010	-0.301*** (0.082)	0.105	-0.499*** (0.141)	0.301
第 3 主成分	-0.191*** (0.036)	0.020	-0.051 (0.059)	0.002	0.135 (0.106)	0.017
第 4 主成分	-0.268*** (0.038)	0.033	-0.188*** (0.038)	0.026	-0.185*** (0.062)	0.027
第 5 主成分	0.303*** (0.048)	0.036	0.215*** (0.068)	0.029	0.04 (0.115)	0.001
第 6 主成分	-0.254*** (0.044)	0.021	0.024 (0.064)	0.000	0.367*** (0.115)	0.073
第 7 主成分	-0.143** (0.061)	0.005	-0.280*** (0.065)	0.030	-0.367*** (0.100)	0.054
第 8 主成分	-0.195*** (0.056)	0.007	-0.132 (0.101)	0.005	-0.033 (0.182)	0.000
第 9 主成分	-0.755*** (0.061)	0.095	-0.244*** (0.073)	0.016	0.242* (0.131)	0.016
第 10 主成分	-0.199*** (0.071)	0.004	-0.130 (0.097)	0.003	0.043 (0.166)	0.000
第 11 主成分	-0.149** (0.074)	0.002	0.157** (0.075)	0.003	0.223* (0.129)	0.007
定数項	-1.318*** (0.054)		-0.119** (0.047)		1.202*** (0.072)	

*p<0.1; **p<0.05; ***p<0.01

より、転入超過数の小さい市区町村では、若年人口が多いほど転入超過数が大きくなる傾向が特に強い。一方、これらの市区町村では、失業率が高く保育施設の逼迫度が高いほど転入超過数が大きくなっており、失業率の低い地域や保育施設に余裕がある地域が好まれるだろうという一般的な想定とは異なる結果になっている。その原因の一つとして、転入超過数の少ない地域では、本研究で設定した説明変数以外の要因による影響が大きく、失業率の高さ・保育施設の逼迫度はそれらモデルに含まれない要因を反映している可能性がある。中央値回帰の結果から、市区町村全体の傾向として都市化の度合いと労働環境の状況により転入超過数が決定し、都市

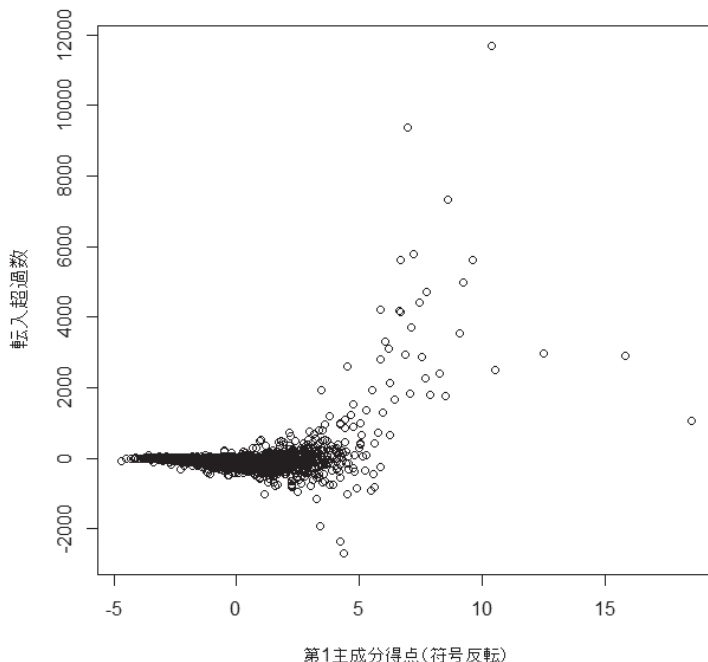


図 3: 転入超過数と都市化の度合いの散布図

化の進んだ市区町村および労働環境が良好な市区町村ほど転入超過数が多い。これは、転入超過数の大きな市区町村の場合も同様である。

転入超過率の場合、 p が 0.2, 0.5, 0.8 のすべての状況で第 1・2 主成分による影響が大きい。表 15 の主成分負荷量から、転入超過数のモデルと同様、第 1 主成分は都市化の程度を、第 2 主成分は労働環境の状況をそれぞれ表し、第 1 主成分の値が小さい市区町村ほど都市化が進んでおり、第 2 主成分の値が小さいほど労働環境の状況が比較的良好であるといえる。これらの解釈より、都市化が進み労働環境が良好な市区町村ほど転入超過率が高い。ただし、転入超過率が高い市区町村では、労働環境の影響が転入超過率に対してより強く影響している。

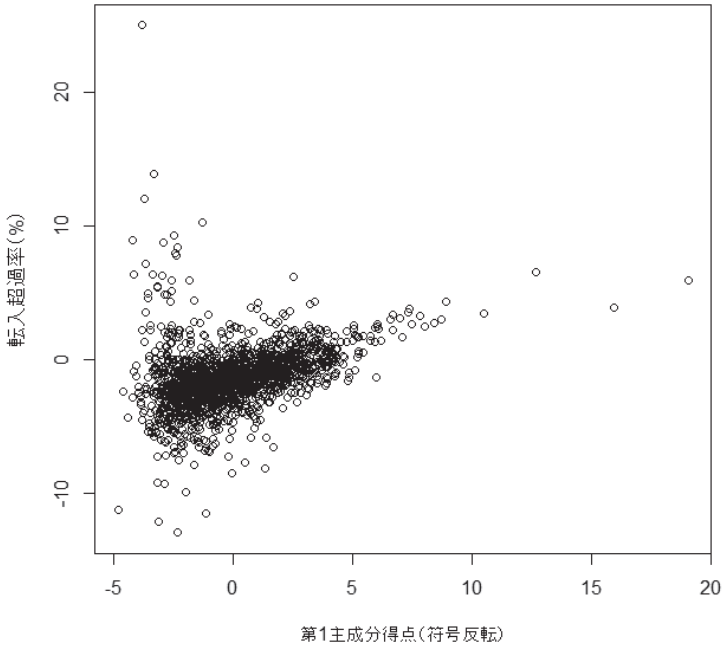


図 4: 転入超過率と都市化の度合いの散布図

以上の結果から、転入超過数・転入超過率とも都市化の程度に強く影響を受けていることが分かった。そこで、転入超過数および転入超過率と都市化の進行度との関連を詳しく見てみよう。図 3 は都市化の度合いを示す第 1 主成分得点と転入超過数の散布図である。なお図 3 では第 1 主成分得点の符号を反転しており、その得点が高い市区町村ほど都市化が進んでいることを表すように修正している。この散布図より、都市化の進行度と転入超過数との間には正の相関があることが分かる。実際には、都市化の進んでいない大部分の市区町村で転入超過数は 0 に近い値となり、都市化の進んだ一部の市区町村で転入超過数が大きくなっている。これは、地方にある多くの市区町村から一部の都市へ若者が集中しているという現状を反映し

ている。図4は第1主成分と転入超過率の散布図である。図4もまた第1主成分得点の符号を反転させている。こちらも都市化の進行度と転入超過率との間に正の相関がみられるが、都市化の進んでいない市区町村ほど、転入超過率のばらつきが大きいことが分かる。すなわち、転入超過率については都市化以外の要因も大きく、都市化が進んでいなくても高い転入超過率を達成している地域が存在していることを示している。

4 まとめ

市区町村の各種データに基づいて、様々な地域属性が若年層の転入超過数および転入超過率に与える影響を分析した。以下では、その分析結果について総括する。

スチューデント化残差により、若年層の転入超過数・転入超過率の双方について外れ値が存在する可能性が高い。これにより、最小二乗法による回帰分析は、これらの外れ値から強い影響を受けることになる。外れ値による影響を軽減するために本研究では中央値回帰および分位回帰を導入し、特に転入超過数の分析で、最小二乗法による回帰分析の推定結果が外れ値から強い影響を受けていることが分かった。したがって、以下では中央値回帰・分位回帰の結果に基づいて総括を進めていく。

転入超過数と転入超過率について、若年人口・人口密度といった人口的要因が正の影響を、失業率が負の影響をもたらすことが明らかとなった。保育施設の余裕の度合いは負の影響を及ぼしており、我々の想定や経験とは逆の結果となっている。しかし、後の主成分回帰により、転入超過数・転入超過率の変動の大部分は、都市化の程度により説明されることが分かった。一般に、都市化が進んだ地域ほど若年人口は多く人口密度が高いが、保育施設の余裕の度合いは低くなる。つまり、都市化の進んだ地域への若者の転入超過が多いことが、保育施設の余裕度の低い地域への若者の転入超過という結果をもたらしたといえる。したがって、地域における転入超過の要因を明らかにするならば、人口・失業率・保育施設の余裕度などの個別の要因だけでなく、各市区町村の都市化といった総合的な要因についても考察する必要があると結論付けられる。

以上の結果は、地域における転入超過の要因について追加的な分析の必要性を示している。第一に、転入超過数の少ない市区町村における転入超過の要因についてのさらなる分析が必要である。これらの市区町村では、都市化による影響は有意ではない。また、失業率が高く保育施設の逼迫度が高いほど転入超過数が多い傾向

があり、一般的な想定とは異なるものである。先に考察したように、転入超過数の少ない市区町村では、本研究で設定した説明変数以外の要因が転入超過数に影響しており、その要因が失業率や保育施設の逼迫度との相関を通じて、転入超過数に影響している可能性がある。この場合、欠落変数バイアスによりパラメータの推定量が一致性を持たないという問題が発生する。したがって、今回設定したモデルに対してより多くの説明変数を追加して分析を行うべきである。ただし、市区町村の地域属性に関するデータについて、利用可能なデータは必ずしも多くない。そこで、代理変数の導入等によってデータの利用に関わる制限に対処し、可能な限り多くの説明変数を用いて分析を進める必要がある。

第二に、都市化の進んでいない地域における転入超過率のばらつきについての分析が必要である。表 17 より、転入超過率の変動の多くは都市化により説明できることを明らかにしたが、図 4 より、都市化の進んでいない市区町村では転入超過率に大きなばらつきがあることを示した。これは、都市化が進んでいない地域において都市化以外の要因が転入超過率に影響しており、その要因を明らかにすることで転入超過率を上昇させる施策を提示できる可能性を示唆している。例えば、鹿児島県十島村の転入超過率は 25% であり、これは今回利用したデータにおいて最も高い転入超過率である。一方、十島村の第 1 主成分得点は 3.78 でその都市化の度合いは低い。つまり、十島村は都市化がそれほど進んでいないにも関わらず極めて高い転入超過率を実現しているのである。十島村に限らず、鹿児島県の島嶼部ではその人口・経済的規模に対して移住希望者が多いことが知られている。今回の分析で用いたモデルでは、その要因を十分に説明することができていない。例えば、自然観光資源の数・出身地に対する愛着の程度・各市区町村による移住支援の種類や度合いといった新たな属性を説明変数に加えた分析により、転入超過率に影響する要因をさらに詳しく捉える必要があるだろう。都市化がそれほど進んでいないにもかかわらず転入超過率の高い地域について、さらなる分析を通じて転入超過率が高い原因を明らかにすることができれば、地方への若者の移住に対して効果的な施策を提示できる可能性がある。

参考文献

- Aitchison, J. (1983) "Principal Component Analysis of Compositional Data," *Biometrika*, Vol. 70, No. 1, pp. 57-65.
- Greene, W. H. (2020) *Econometric Analysis(Global Edition)*: Pearson Educa-

tion, 8th edition.

Jolliffe, I. T. (2002) *Principal Component Analysis*: Springer-Verlag, 2nd edition.

Koenker, R. and K. F. Hallock (2001) “Quantile Regression,” *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143-156.

福井昭吾 (2018) 「出生率の地域差に関するクロスセクション分析」, 『九州大学経済学研究』, 第 85 巻, 第 2・3 号, 43-66 頁.

太田聰一 (2005) 「地域の中の若年雇用問題」, 『日本労働研究雑誌』, 第 539 号, 17-33 頁.

後河正浩 (2019) 「若者の地域間移動の傾向と要因—都道府県データでみる大学進学・初職就職時の地域間移動—」, 『京都産業大学経済学レビュー』, 第 6 号, 1-42 頁.

林直樹・齋藤晋・高橋強 (2003) 「農村地域における若年層男性の人口移動と産業構造」, 『農村計画学会誌』, 第 22 巻, 31-36 頁.