

離散型サンプルセレクションモデルによる自動車保有と都市環境の因果分析

渡邊 萌¹・円山 琢也²

¹学生会員 熊本大学 大学院自然科学教育部 (〒860-8555 熊本市中央区黒髪2-39-1)

E-mail: 197d9225@st.kumamoto-u.ac.jp

²正会員 熊本大学准教授 くまもと水循環・減災研究教育センター (〒860-8555 熊本市中央区黒髪2-39-1)

E-mail: takumaru@kumamoto-u.ac.jp

土地利用パターンや道路幅員、アクセシビリティなどの都市環境の違いが、住民の自動車保有に与える影響は大きい。それらの影響を正しく把握することは、自動車依存社会からの脱却や、コンパクトシティ施策の効果予測のために必要となる。そのためには、(a) 共変量の影響の除去、(b) 選択バイアスの補正の二つを考慮した解析が求められる。既存の手法を用いて自動車保有などの離散的な交通行動と都市環境の因果関係を明らかにする際、選択バイアスと関係する未観測の変数を考慮できない点が大きな課題であった。本研究では、未観測変数によるバイアスを明示的に補正する因果推論手法であるサンプルセレクションモデルの枠組みで、離散型の結果変数を扱うことができるモデルを提案する。また、熊本都市圏PT調査データへの適用を行い、手法の妥当性を検討する。

Key Words : causal inference, sample selection model, self-selection, built environment, car ownership

1. はじめに

自動車依存型社会からの脱却やコンパクトシティ施策の効果計測には、居住地の都市環境が自動車保有に与える影響を明らかにすることが重要である。一般的に、都市部と郊外部では住民の自動車保有割合は異なり、都市部の住民の自動車保有割合は低く、郊外部の住民の自動車保有割合は高い傾向にある¹⁾。これは、土地利用パターンや道路幅員、商業施設や公共交通機関までのアクセス等、様々な都市環境の差異が要因として考えられる。ここで郊外部居住者が都市部に住み替えを促す政策によって、自動車保有率が変化するかどうかを検討するとする。このとき、都市部と郊外部の自動車保有率の差を、予想される変化として直接的に解釈することはできない。なぜなら、都市部と郊外部における都市環境の差異と自動車保有の間の相関関係は、交絡や逆の因果性が介在する可能性があり、必ずしも因果関係を意味しないためである。より信頼性の高い結果に基づく政策的示唆を得るためには、統計的因果推論手法に基づいた因果効果の算出が必要である²⁾。しかし、自動車保有などの離散的な交通行動を扱う既存の因果推論手法は限られており、ま

た分析には人々の交通行動と居住地選択に関する多くの情報が必要となる。そのため、これまであまり知見が蓄積されておらず、離散的な交通行動をより少ない情報量で分析することのできる手法の開発が望まれる。

本研究では、サンプルセレクションモデルの枠組みで、離散的な結果変数を扱うための統計的因果推論手法を提案する。具体的には、居住地の都市環境の差異(都市部・郊外部)が与える、自動車保有率への因果効果を推定する手法を提案する。さらに、熊本都市圏居住者の自動車保有行動へ適用し、手法の妥当性を検討する。本稿では、離散的な結果変数を扱えるサンプルセレクションモデルを「離散型サンプルセレクションモデル」と呼称する。

2. 既存手法の整理と課題

本研究で提案する手法は、Rubin³⁾によって提案された反実仮想アプローチ⁴⁾に分類される。以下、一般的な反実仮想アプローチの具体的な流れと既存手法の限界を踏まえた上で、本研究で提案するモデルの意義を述べる。

(1) 反実仮想アプローチによる因果効果の推定法

a) 共変量による調整

本研究のように都市環境の差異が自動車保有に与える効果を推定する際、独立変数である都市環境の差異(都市部・郊外部)と、結果変数である自動車保有の両方に関係する共変量の影響を除去する必要がある。例えば、都市部における18歳未満の人口比率が高ければ、都市部の自動車保有割合は当然低くなる。このような影響が除去された効果を算出しなければ、都市環境の差異が自動車保有に与える因果効果を明らかにすることはできない。本研究で提案するモデルでは、都市部と郊外部の結果変数にそれぞれ共変量を説明変数とした潜在的な回帰関数を仮定し、都市部と郊外部の結果変数の期待値の差を因果効果とすることで共変量による調整を行う。

b) 選択バイアスの補正

都市環境の差異と自動車保有との間の因果関係を分析する際には、都市環境の差異が自動車保有に与える効果に介在する選択バイアスを補正する必要がある。一般的に反実仮想アプローチにおいて対処が必要となる選択バイアスとは、対象者が観測されるかどうか観測値以外の要因にも依存している状況で、単純な解析を行うことにより生じるバイアスである。具体的な例として、車を保有せず日常的に公共交通機関を利用する人は、より公共交通機関が整備された都市部を居住地として選択する可能性がある(residential self-selection: 居住地自己選択)⁹⁾。この場合、人々の居住地選択(都市部・郊外部)は彼らの自動車保有に対する嗜好や習慣と強く相関してしまい、単純な解析ではバイアスのかかった結果となる。

(2) 選択バイアスの補正方法

分析の際、選択バイアスが生じない場合は上述した共変量による調整のみで適切な解析が可能である。しかし、本研究のように選択バイアスが存在する可能性を無視できない場合は、共変量による調整と同時に選択バイアスを補正する必要がある。このとき、反実仮想アプローチでは主に以下の二つが方法として挙げられる⁹⁾。

a) 観測値による補正

一つ目は、選択バイアスを誘発すると考えられる嗜好や習慣、その他様々な要因に関する情報を調査により取得する方法である。すなわち、対象者が観測されるかどうか影響を及ぼす要因を可能な限り全て観測する方法である。これにより、統合離散選択モデル⁷⁾や傾向スコアをはじめとした様々な手法を用いた解析による推論が可能になる。しかし、これらの解析手法は選択バイアスの要因が全て観測できていることを前提としているため、観測値以外の要因が存在する場合、結果の信頼性が損なわれる。そのため、必然的に質問事項の数は多くなり、高い調査コストを必要とする。

b) サンプルセレクションモデル

二つ目は、未観測な要因による選択バイアスを明示的に補正する手法を用いて解析を行う方法である。具体的な手法として、サンプルセレクションモデルが挙げられる。この手法はヘックマンのプロビット選択モデル⁸⁾を統計的因果推論の考え方を導入して拡張したものであり、交通分野での適用研究も見られる⁹⁾。サンプルセレクションモデルでは、一般的には結果変数に重回帰モデルを仮定する。それゆえ、サンプルセレクションモデルの結果変数は連続量である必要があり、自動車保有(保有する、保有しない)のような離散的な結果変数を持つ問題にそのまま適用することができない。

(3) 既存手法の課題

現状、本研究で対象とする自動車保有のような離散的な交通行動と都市環境の差異との因果関係を明らかにするための手法は限られている。そのため、都市環境の差異が自動車保有に与える因果効果を算出するためには、新たな調査を行い解析に必要な情報を完全に取得する必要がある。その理由を、(a) 解析手法の課題と(b) データ上の課題の点からそれぞれ説明する。

a) 解析手法の課題

上述したように、未観測な要因による選択バイアスを明示的に補正するサンプルセレクションモデルは、連続量の結果変数を持つ交通行動(トリップ数、自動車保有台数など)しか扱うことができない。したがって、都市環境の差異が自動車保有のような離散的な交通行動に与える因果効果を推定する際に、未観測な選択バイアス要因が存在する場合、因果効果を正しく推定する手法は現状存在しない。

b) データ上の課題

選択バイアスと関係する嗜好や習慣等に関する情報は、パーソントリップ調査(以下PT調査)などの都市圏を対象とした伝統的な大規模調査では十分に取得しないケースがほとんどである。また、選択バイアスと関係する嗜好や習慣は、対象とする交通行動によって異なるため、一般的に分析の度に新しく調査を実施し、取得する必要がある。

(4) 離散型サンプルセレクションモデルの意義

上述したように、既存手法の限界と大きな調査コストにより、離散的な交通行動と都市環境の差異との因果関係を明らかにした既存研究は限られている。そのため、これまで知見があまり蓄積されていない。しかし、離散型サンプルセレクションモデルの開発により、様々な分析が既存のデータを用いて可能になる。日本においても、PT調査をはじめとした様々な交通行動調査が長年実施されている。一般的にこれらの調査では選択バイアス

の要因は十分に観測されていない。しかし、上述の通りサンプルセレクションモデルは観測値以外の要因によるバイアスを明示的に補正することができるため、過去に取得されたそれらのデータを用いて推論を行うことができる。しかしながら、筆者の知る限り、離散的な結果変数を扱えるサンプルセレクションモデルへの拡張を試みた研究は存在しない。次章では、一般的なサンプルセレクションモデルの拡張である離散型サンプルセレクションモデルの提案を行う。

3. 離散型サンプルセレクションモデル

(1) サンプルセレクションモデル

まず、一般的なサンプルセレクションモデルを説明する。ここでは、一般的なサンプルセレクションモデルにおける、(a) 共変量による影響の除去、(b) 選択バイアスの補正方法の2点に焦点を当てる。基本的にこの2点は、次節で説明する離散型サンプルセレクションモデルでも共通である。

$$y_{1i}^* = x_{1i}\beta_1 + u_1, \quad (1)$$

$$y_{1i} = \begin{cases} 1 & (\text{if } y_{1i}^* > 0) \\ 0 & (\text{if } y_{1i}^* \leq 0) \end{cases}$$

$$y_{2i} = x_{2i}\beta_2 + \varepsilon_2 \quad (\text{if } y_{1i} = 1), \quad (2)$$

$$y_{3i} = x_{3i}\beta_3 + \varepsilon_3 \quad (\text{if } y_{1i} = 0),$$

i は対象とする個人、 x は共変量、 β は係数ベクトル、 $u_1, \varepsilon_2, \varepsilon_3$ は誤差項、 y_1^* は潜在変数である。式(1)は一般的に選択方程式と呼ばれる。連続量の観測値である y_2 と y_3 が同時に観測されることはなく、 $y_1^* > 0$ の時に y_2 が観測され、 $y_1^* \leq 0$ の時に y_3 が観測されるとする。

サンプルセレクションモデルの枠組みでは、共変量を条件付きとした y_2 と y_3 の期待値の差の標本平均を、共変量の影響が除去された因果効果とするのが一般的である。ただし、このとき誤差項間で相関が生じている場合、 y_2 と y_3 は不偏推定量でも一致推定量でもなくなる。これにより生じるバイアスが、サンプルセレクションモデルの枠組みにおける選択バイアスとなる。よって誤差項間の相関を明示的に考慮することにより、選択バイアスの補正を行う。誤差項の同時分布は、多変量正規分布による分散共分散行列を仮定するのが一般的である。

$$\begin{pmatrix} u_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_2 & \sigma_3 \\ \sigma_2 & \sigma_2^2 & 0 \\ \sigma_3 & 0 & \sigma_3^2 \end{pmatrix} \right], \quad (3)$$

識別性の問題から u_1 の分散は1とする。 σ_2^2, σ_3^2 は結果変数である y_2 と y_3 の分散、 σ_2, σ_3 は共分散である。一般的なサンプルセレクションモデルのさらなる詳細と推定法については、既存研究⁹⁾を参照されたい。

(2) 離散型サンプルセレクションモデル

本節では離散型サンプルセレクションモデルを提案する。上で示した一般的なサンプルセレクションモデルとの違いは、(a) 結果変数が潜在変数、(b) 誤差項の同時分布が相関行列の2点である。順を追って説明する。

まず、離散型サンプルセレクションモデルでは三種類の潜在変数 y_1^*, y_2^*, y_3^* を以下のように仮定する。 i は対象とする個人、 x は共変量、 β は係数ベクトル、 $u_1, \varepsilon_2, \varepsilon_3$ は誤差項である。

$$y_{1i}^* = x_{1i}\beta_1 + u_1, \quad (4)$$

$$y_{1i} = \begin{cases} 1 & (\text{if } y_{1i}^* > 0) \\ 0 & (\text{if } y_{1i}^* \leq 0) \end{cases}$$

$$y_{2i}^* = x_{2i}\beta_2 + \varepsilon_2 \quad (\text{if } y_{1i} = 1), \quad (5)$$

$$y_{3i}^* = x_{3i}\beta_3 + \varepsilon_3 \quad (\text{if } y_{1i} = 0),$$

式(4)は選択方程式であり、 $y_1^* > 0$ の時に y_2^* が観測され、 $y_1^* \leq 0$ の時に y_3^* が観測されるとする。また、 y_2^* と y_3^* が同時に観測されることはない。 y_2^* と y_3^* は結果変数であり、離散型サンプルセレクションモデルでは潜在変数を仮定する。この点において、結果変数が連続量の観測値である一般的なサンプルセレクションモデルと異なる。そのため離散型サンプルセレクションモデルでは、潜在的な結果変数 y_2^*, y_3^* の期待値の差の標本平均を因果効果と定義する。また、離散型サンプルセレクションモデルでは、結果変数の観測値は離散変数 y_2, y_3 である。

$$y_{2i} = \begin{cases} 1 & (\text{if } y_{2i}^* > 0) \\ 0 & (\text{if } y_{2i}^* \leq 0) \end{cases}, \quad (6)$$

$$y_{3i} = \begin{cases} 1 & (\text{if } y_{3i}^* > 0) \\ 0 & (\text{if } y_{3i}^* \leq 0) \end{cases}$$

式(6)に示すように、それぞれ $y^* > 0$ の時に1が観測され、 $y^* \leq 0$ の時に0が観測されると仮定する。また、誤差項の同時分布は以下のような多変量正規分布を仮定する。このとき、 σ_2, σ_3 は相関係数である。

$$\begin{pmatrix} u_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_2 & \sigma_3 \\ \sigma_2 & 1 & 0 \\ \sigma_3 & 0 & 1 \end{pmatrix} \right], \quad (7)$$

式(7)が示すように、離散型サンプルセレクションモデルではパラメータ推定時の識別性の問題から、誤差項の同時分布は相関行列となっている¹⁰⁾。一方、一般的なサンプルセレクションモデルでは連続値である結果変数の分散を表現するため、分散共分散行列を仮定するのが一般的である。この点が二つ目の違いである。

次に、尤度関数を定義する。まず、観測された結果変数が $y_{2i} = 1$ の個人を N_1 、 $y_{2i} = 0$ の個人を N_2 、 $y_{3i} = 1$ の個人を N_3 、 $y_{3i} = 0$ の個人を N_4 と分類する。このとき、尤度関数 L は式(8)のように定義される。

$$\begin{aligned}
& L(\beta, \sigma_2, \sigma_3 | y_1, y_2, y_3) \\
& \propto \prod_{i \in N_1} f(y_{1i}^* > 0, y_{2i}^* > 0 | \beta, \sigma_2) \\
& \times \prod_{i \in N_2} f(y_{1i}^* > 0, y_{2i}^* \leq 0 | \beta, \sigma_2) \\
& \times \prod_{i \in N_3} f(y_{1i}^* \leq 0, y_{3i}^* > 0 | \beta, \sigma_3) \\
& \times \prod_{i \in N_4} f(y_{1i}^* \leq 0, y_{3i}^* \leq 0 | \beta, \sigma_3),
\end{aligned} \tag{8}$$

このとき、

$$\begin{aligned}
& f(y_{1i}^* > 0, y_{2i}^* > 0 | \beta, \sigma_2) = \Phi_2(x_{1i}\beta_1, x_{2i}\beta_2, \sigma_2), \\
& f(y_{1i}^* > 0, y_{2i}^* \leq 0 | \beta, \sigma_2) = \Phi_2(x_{1i}\beta_1, -x_{2i}\beta_2, -\sigma_2), \\
& f(y_{1i}^* \leq 0, y_{3i}^* > 0 | \beta, \sigma_3) = \Phi_2(-x_{1i}\beta_1, x_{3i}\beta_3, -\sigma_3), \\
& f(y_{1i}^* \leq 0, y_{3i}^* \leq 0 | \beta, \sigma_3) = \Phi_2(-x_{1i}\beta_1, -x_{3i}\beta_3, \sigma_3).
\end{aligned} \tag{9}$$

Φ_2 は二変量正規分布の累積分布関数(CDF)である。

本研究では、居住地（都市部、郊外部）と自動車保有（保有する、保有しない）を観測値とする。具体的には、都市部に居住している住民は $y_{1i} = 1$ 、郊外部に居住している住民は $y_{1i} = 0$ が観測されたとし、自動車を保有している住民はそれぞれ $y_{2i} = 1$ 、 $y_{3i} = 1$ 、自動車を保有していない住民はそれぞれ $y_{2i} = 0$ 、 $y_{3i} = 0$ が観測されたと定義する。

離散型サンプルセレクションモデルは結果変数に潜在変数を仮定しているため、一般的なサンプルセレクションモデルと異なり、最尤推定法によりパラメータを求めることが困難である。よって、本研究ではMCMCによる解法アルゴリズムを構築し実装を行う予定である¹⁰⁾。

4. 今後の流れ

本稿では、離散型サンプルセレクションモデルの提案を行った。今後は、2012年熊本都市圏PT調査データへの適用を行い、提案手法の妥当性を検討する。具体的には、自動車依存の高い熊本都市圏内の都市部と郊外部で、自

動車保有確率の差異を明らかにし、得られた知見を基に議論を行う。解析結果は発表会当日に発表予定である。

参考文献

- 1) 自動車検査登録情報協会：都市別の自家用乗用車の普及状況, 2019.
- 2) 織田澤利守, 大平悠季：交通インフラ整備効果の因果推論：論点整理と展望, 土木学会論文集D3（土木計画学）, Vol.75, No.5, pp.L1-L15, 2019.
- 3) Rubin, D. B.: Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, Vol.66, pp.688-701, 1974.
- 4) 星野崇宏：調査観察データの統計科学：因果推論・選択バイアス・データ融合, 岩波書店, 2009.
- 5) Cao, X., Mokhtarian, P. L. and Susan L. Handy.: Examining the impacts of residential self-selection on travel behavior: a focus on empirical findings, *Transport reviews*, Vol.29, No.3, pp.359-395, 2009.
- 6) Mokhtarian, P. L. and Cao, X.: Examining the impacts of residential self-selection on travel behavior: A focus on methodologies, *Transportation Research Part B: Methodological*, Vol.42, No.3, pp.204-228, 2008.
- 7) Bhat, C. R. and Guo, J. Y.: A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels, *Transportation Research Part B: Methodological*, Vol.41, No.5, pp.506-526, 2007.
- 8) Heckman, J. J.: Sample selection as a specification error, *Econometrica*, Vol.47, pp.153-161, 1979.
- 9) Mokhtarian, P. L. and van Herick, D.: Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches, *Journal of Transport and Land Use*, Vol.9, No.1, pp.9-28, 2016.
- 10) Chib, S. and Greenberg, E.: Analysis of multivariate probit models, *Biometrika*, Vol.85, No.2, pp.347-361, 1998.

DEVELOPING A DISCRETE CHOICE MODEL WITH SAMPLE SELECTION: EXAMINING THE BUILT ENVIRONMENT IMPACTS ON CAR OWNERSHIP

Hajime WATANABE and Takuya MARUYAMA