

交通手段選択における 非観測要因の考慮方法

Consideration of unobservable aspects in travel mode choice modeling

大関 正博*
Masahiro OSEKI

*被災地支援・地域計画学研究室（指導教員：奥村 誠 教授）

Travel mode choice behavior is dependent not only on service level of transport modes, but also on personal attributes or preferences, which cannot easily be observed in ordinal sampled surveys. Because these unobserved aspects also affect the choice of residence location choice and the observed transport service level, then correlation with the error term violates a fundamental condition of reliable unbiased regression. Cao et al.(2009) summarized the reason to this self-selection or also called as endogeneity problem, and classified countermeasures up to 2009. This study tests three methods of them applicable to the cross-sectional survey of Tohoku University Students (2016). Instead of a propensity score matching method previously reported in Cao et al.(2009), inverse weighting method was applied and good performance was proved.

Key Words : Travel model choice, unobservable aspects, sample-selection, propensity score

1. はじめに

交通計画において、交通条件の改善に伴い個人の交通行動がどの程度変化するかは予測は不可欠であり、各種の交通手段選択モデルが提案されてきた。図-1のように個人の交通手段に選択行動には、個人の性別、職業、住んでいる地域の交通条件などの外部から観測できる観測可能要因だけでなく、態度・選好などの個人的な要因も影響を与えるが、調査方法や費用などの制約により情報が得られず、非観測要因として扱う必要に迫られる。この非観測要因はまた、自身の好む交通手段の利便性が高い地域を選択するという形で交通条件にも影響を与える。この条件下で非観測要因の影響を取り入れずに交通手段選択モデルを推定すると、交通条件からの真の影響に加えて非観測要因からの影響が混入し、パラメーター値を過大推定してしまうという問題が発生する。

1990年代以降、多くの交通行動モデル研究者がこの問題の回避方法の研究に取り組み、Mokhtarian, Cao¹⁾, Caoら²⁾が約20年間の研究レビューを行い既存の問題回避方法を、9つに分類している。

交通計画の実務上は、計画の内容やある程度具体化してきた時点で限られた費用の下で調査を行うこととなるため、1時点の横断的調査に頼らざるを得ないケースが多い。そのため、本研究は9つの方法の中で1時点の横断的調査データに適用できる可能性がある、統計的コントロール法、操作変数法、傾向スコア法の3つを比較する。その際操作変数法を改良した松島ら³⁾の方法と、傾向スコア法の中でもマッチングの恣意性問題を回避できる逆数重みづけ法⁴⁾を適用する。

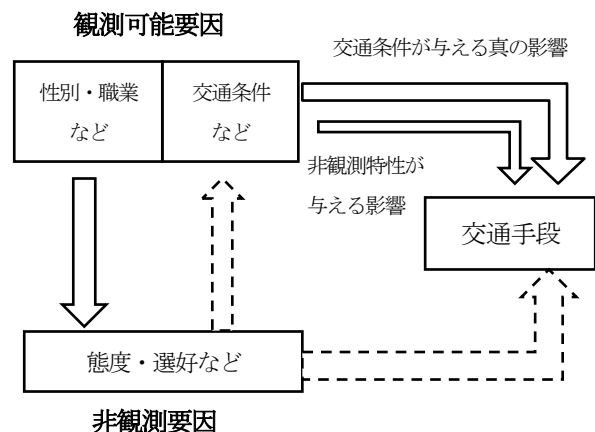


図-1 交通手段選択の概念図

2. 非観測要因の考慮方法に関する既存の研究

2.1 Caoら²⁾が分類した各方法の概要と特徴

(1)直接質問法

各調査対象者に詳細な聞き取りを行い、交通行動を決定した意思決定過程を確認する方法である。この方法は統計的処理が困難で、質問の言葉遣いや調査者の態度の影響を受けやすいため、本研究では対象としない。

(2)統計的コントロール法

通常の交通行動調査では非観測となるAT要因を合わせて調査し、BE変数と一緒に用いてその効果を統計的に分離する方法である。アンケート調査で過去の居住地の選択時点でのATを調査することが難しいという問題がある。

(3)操作変数法

BE 変数との相関が強く AT 変数との相関が弱い変数を見出して操作変数として用いる。まず BE 変数を操作変数に回帰して AT 変数との相関が弱い推定値を得る。第 2 段階として、この推定値を BE 変数の代わりに用いた推定を行う。この方法の成否は、最初に挙げた条件を満たすような操作変数がうまく選定できるかに依存する。

(4) サンプル選択モデル

各個人の AT 要因を考慮して第一段階の居住地選択行動を離散的にモデル化したのち、第二段階として各グループごとに別々の交通行動モデルを推定する。この方法では、AT 要因の間接的な影響を定数項や他のパラメータの違いに固定化させたに過ぎず、過大推定の解消にはつながらないという問題がある。

(5) 傾向スコア法

もし計画的な実験が可能であれば、他の変数を揃えた 2 組のサンプル群を用意し、一方を固定したまま（対照群）他方に介入（処置群）すれば、介入効果と他の変数の差異は無相関となるため、群間の差異を処置の効果と見なすことができる。ところが居住地選択は BE 変数の影響を受けるためこの条件を満たされない。傾向スコアとは第一段階の選択を説明する変数を意味し、群間においてその値が均等になるようにサンプルを調整することで、着目する観測可能 BE 変数と他の変数との相関を打ち消すのが傾向スコア法である。第一段階の居住地選択モデルは無相関化する目的を満たせば十分であり、変数の有意性や実証的な意味を考えなくてもよい。居住地選択段階の社会経済的要因データを必要としない。

(6) 横断面構造方程式モデル

BE 変数、TB 変数と同時に AT 要因を調査し、これらの相互の影響関係を線形連立方程式として記述して構造パラメーターを同時推定する。異なる因果構造を柔軟に仮定し、パラメーターの有意性検定を通じて構造同定ができる。同時に観測した変数で因果関係を識別することの問題、異なる構造のモデルが同一の適合度を持ち、一意に同定ができないなどの問題がある。

(7) 同時選択モデル

各居住地における交通手段の離散的選択モデルと、その上位の居住地の離散的選択モデルを同時に作成する方法で、Nested Logit や Nested Probit モデルが使われる。仮定した因果構造について不整合の有無を確認するにとどまり、一つの構造を同定できないという問題がある。

(8) 相互依存選択モデル

(7) では 2 つのレベルの選択の説明変数は異なるものを用いるが、さらに一方の選択結果が他方の説明変数に入る構造を仮定したモデルである。(6) と同様に同時に観測した変数で因果関係を識別することの問題がある。

(9) 縦断的デザイン

以上の(6)-(9)のモデルでは意思決定のタイミングや時間スケールが異なる居住地選択段階と交通行動段階を一時点の調査データで獲得することの困難さに直面する。そのため同一個人の行動を複数の時点で追跡してパネルデータを獲得する縦断的な調査のデザインが望まれる。しかし、縦断的な調査は一時点の調査に比べて時間、費用を多く必要とすることが決定的な問題である。

2.2 1 時点の横断面交通行動調査データへの適用性

本研究では、1 時点の交通機関選択行動調査のデータを取り上げる。過去の居住地選択時の要因に関する質問を行うことが難しいことを踏まえると、適用性が高いのは、(2)統計的コントロール法、(3)操作変数法、および(5)傾向スコア法の 3 つである。

2.3 松島ら³⁾による操作変数法の拡張

松島ら³⁾は、第 2 段階が旅行回数のような連続変数の回帰ではなく、交通機関選択のような離散選択モデルの場合に適用できるよう、コントロール関数法を拡張し、通勤時の交通手段選択モデルの推定に適用した。具体的には、観測 BE 変数である最寄り駅までのアクセス時間は、非観測 AT 要因である個人の交通手段に対する嗜好の影響を受けていると考え、居住地の地価を操作変数に用いて第 1 段階の推定を行った。第 2 段階ではその推定値ではなく、残差を AT に対応する説明変数として用いて交通手段選択モデルを推定した。その結果、残差(嗜好に係る変数)が有意に推定された一方で、アクセス時間は非有意と推定された。

2.4 傾向スコア法に関する研究

Rosenbaum, Rubin⁴⁾は、個人 i の共変量ベクトルを x_i 、割付変数の値を z_i とするとき、2 群のうちの群 1 に割り付けられる確率 $e_i = \Pr(z_i = 1 | x_i)$ を個人 i の傾向スコアと定義した。交通行動分析では、都市部と郊外部の 2 群を考え、2 項ロジットモデルを推定してその効用関数値を傾向スコアとして用いることが行われてきた。

しかし、地域が 3 つ以上に分けられ、非観測要因が類似する人々が特定の地域を選びやすいという傾向が認められるのであれば、これらの地域を群と考え多項ロジットモデルを推定して傾向スコアを作成することもできる。この傾向スコアの値は、各個人が特定のゾーンに住居しやすいという性質の強さを表しており、その利用方法として 3 つの方法が提案されている。

(1) 傾向スコアを非観測要因を表す新たな変数と考え、説明変数に用いる。これにより観測可能変数が負担していた説明力を傾向スコアに取り戻させ、過大推定を抑えることが期待できる。

(2) ある群の中の個々のサンプルの傾向スコアの値を基準として、他の群のサンプルの中からスコアの値が近いサンプルを選びなおすマッチングを行う。その結果、す

すべての群のサンプルは傾向スコアが類似するサンプルで構成され、群間の差異はすべて観測可能要因の影響によるものと見なすことが可能となる。ただし具体的なマッチングの方法や閾値設定に恣意性があるほか、マッチングの結果サンプル数が減るという問題がある。

(3) 逆数重みづけ法 傾向スコアを作るための群選択モデルの確率値の逆数を用いて、各サンプルに重みづけするのが逆数重みづけ法である。恣意性がなく、もとのサンプルをすべて使い各群の非観測要因の分布を揃えることができる方法である。本研究ではこの方法を適用する。

3. 実証分析の対象データ

3.1 調査とデータの概要

東北大学所属の学生・教職員を対象に2016年6月上旬に実施された交通行動に対するWebアンケート調査のデータを用いて、調査時点の通学交通手段モデルの推定を行う。本研究では、居住地の選択において個人の指向の影響が強いと考えられる青葉山キャンパスに通う学生を取り上げる。居住地は、第4回仙台市PT調査の大ゾーン区分を参考に、青葉山までの交通条件が明確に異なる片平地区と八木山地区を再区分したゾーン分けを行った。また自転車や徒歩で移動する割合の多い川内・青葉山ゾーンに居住する学生は調査対象から外し、540人の学生のデータを用いる。

3.2 カテゴリカル因子分析

アンケート調査の中では、意識に関して表-1の質問を行い、5(とてもそう思う)から1(全くそう思わない)までの5件法での回答が得られている。この結果にカテゴリカル因子分析を行い、プロマックス回転を行って算出した2つの因子負荷量を表-2に示す。この結果から、因子1の値が大きいほど金銭の負担を気にせず、因子2の値が大きいほど事故のリスクを重視し、エネルギー節約に積極的であるという個人の考え方を表している。

表-1 交通に対する意識調査の内容

5(とてもそう思う)~1(全くそう思わない)で回答
・資源やエネルギー活用のため、極力公共交通や自転車を使うべきである(エネルギー)
・交通事項を防ぐため、極力公共交通を使うべきである(事故防止)
・自転車や徒歩は、金銭的負担が軽く、優れた交通手段である(金銭負担)
・季節や天候に左右されずに、通勤・通学できる交通手段が必要である(天候)

4. 非観測要因の考慮方法の比較

4.1 観測可能要因のみを考慮した比較モデル

式(1)、(2)のような公共交通 $j = p$ (鉄道・バス) と車両 $j = c$ (車・バイク・原付) の2項ロジットモデルを考え

表-2 因子分析の結果

	因子負荷量		共通性
	因子1	因子2	
エネルギー	-0.17	0.659	0.568
事故防止	0.131	0.983	0.861
金銭負担	-0.986	-0.101	0.888
天候	-0.152	0.134	0.061
因子間相関	-0.472		

る。効用関数 U_{ij} には観測可能要因のみを取り入れる。すなわち個人の居住地から勤務地までの交通条件を表す変数を用いる。未知パラメーターの推定は式(3)の尤度関数を最大化する最尤法を用いて行う。

$$P_{ip} = \frac{1}{1 + \exp(V_{ic} - V_{ip})} \quad (1)$$

$$V_{ij} = \alpha x_{ij} + \gamma_j \quad (2)$$

$$\max \sum_i k_i \log(P_{ip}) + (1 - k_i) \log(1 - P_{ip}) \quad (3)$$

ただし、 P_{ij} :個人 i の交通手段 j の選択確率、 V_{ij} :個人 i が交通手段 j を選択したときの効用の確定効用部分、 x_{ij} :個人 i が交通手段 j を選択したときの交通条件、 α :交通条件パラメーター、 γ_j :手段 j の定数項、 k_i :個人 i の実際の選択を表す0-1変数。

交通条件としてゾーン代表点から青葉山キャンパスまでの所要時間、金銭費用、個人による居住地の交通条件の5段階評価値を取り入れた。符号条件と有意性を満足した推定結果は表-3に示すようにゾーン間所要時間のみを用いたモデルだけであった。内生性問題のために、この所要時間パラメーターが過大推定になっている危険性がある。

表-3 所要時間のみを考慮したモデルの推定結果

	推定値	標準誤差	P値
ゾーン間所要時間 α_T	-0.1404	0.0246	<0.0001
定数項 γ	-4.1363	0.6999	<0.0001

4.2 統計的コントロール法に基づく選好データ併用モデル

4.1のモデルに対して、式(4)のように効用関数が各個人の非観測要因も含むケースを考える。個人の非観測要因の値として3.2のカテゴリカル因子分析によって算出した2つの因子得点の値を用いる。

$$V_{ij} = \alpha x_{ij} + \beta_j y_i + \gamma_j \quad (4)$$

ただし、 y_i :個人 i の非観測要因に関連する変数、 β_j :手段 j に対する非観測要因関連変数のパラメーター。

最尤推定法による推定結果を表-4に示す。ゾーン間所要時間のパラメーター推定値は表-3の比較モデルに比べて絶対値が小さく、過大推定が補正される方向の結果が表れている。しかしその差違は小さく、標準誤差の範囲に収まっているため、統計的に意味のある補正がで

表4 選好データ併用モデル推定結果(因子変数あり)

	推定値	標準誤差	P値
ゾーン間所要時間 α_T	-0.1179	0.0257	<0.0001
因子1の得点 β_{1c}	-0.464	0.1146	0.0001
因子2の得点 β_{2c}	-0.6159	0.1074	<0.0001
性別	-0.6512	0.2401	0.0067
定数項 γ	-3.4516	0.7347	<0.0001

きたとは言えない。

4.3 操作変数法に基づく二段階推定モデル

操作変数法を参考にした二段階推定モデルを作成する。本調査では個人ごとの通勤所要時間は調査していないが、居住地の交通条件の5段階評価値を調査している。この評価値は居住地ゾーンの平均的な利便性に加えて、各個人が居住地ゾーンの中で交通条件が良い場所を選択したかどうかを反映している。そこでまず個人の評価値を各ゾーンの公共交通本数に基づく公共交通の利便性指標で線形回帰する。第2段階として、その残差を式(4)における個人*i*の非観測要因の変数 y_i として用いる。

最尤推定法による推定結果を表-5に示す。ゾーン間所要時間のパラメーター推定値は、表-3と比較して絶対値が小さくなっているものの、その差異は表-4よりもさらに小さく、標準誤差の範囲に収まっている。このため、統計的に意味のある補正ができたとは言えない。

表-5 二段階推定モデルの推定結果

	推定値	標準誤差	P値
ゾーン間所要時間 α_T	-0.1243	0.0249	<0.0001
利便性の残差 β_{dc}	-0.361	0.0672	<0.0001
定数項	-3.6993	0.7092	<0.0001

4.4 傾向スコア逆数重み付けモデル

第一段階として各個人が居住地を選択する多項ロジットモデルを作成し、モデルを用いて個々のサンプルの現在の居住地の選択確率値を求めて傾向スコアとする。

個人*i*の傾向スコア PS_i は各個人が居住地*j*に住む確率 $P_i(j)$ から式(5)-(7)のように求められる。

$$PS_i = P_i(j) * D(t_i) \quad (5)$$

$$D(t_i) = \begin{cases} 1 & (t_i = j) \\ 0 & (t_i \neq j) \end{cases} \quad j = 1, \dots, J \quad (6)$$

$$P_i(j) = \frac{\exp(a_j s_i + b_j f_{1i} + c_j f_{2i} + d_j)}{\sum_{j=1}^J \exp(a_j s_i + b_j f_{1i} + c_j f_{2i} + d_j)} \quad (7)$$

ただし、 t_i :個人*i*の居住地割り当てを表す変数 ($t_i = 1, \dots, J$), s_i :個人*i*の性別ダミー(男性0, 女性1), f_{1i}, f_{2i} :個人*i*の因子1, 2の値, a_j :性別ダミーの居住地*j*に対するパラメーター, b_j, c_j :居住地*j*に対する因子1, 2得点のパラメーター, d_j :居住地*j*に対する定数項のパラメーター。

式(8)のように PS_i の逆数をかけた対数尤度関数を用いて最尤推定法によってパラメーターを推定する。

表-6 傾向スコア逆数重み付けモデルの推定結果

	推定値	標準誤差	T値
ゾーン間所要時間 α_T	-0.1065	0.0041	-26.0097
定数項 β	-2.908	0.1101	-26.4093

$$\max \sum_i \frac{1}{PS_i} (k_i \log(P_{ip}) + (1 - k_i) \log(1 - P_{ip})) \quad (8)$$

この推定結果を表-6に示す。これを表-3と比較すると、所要時間パラメーター推定値の絶対値は0.035と、表-6の標準誤差の約8倍小さく、有意に異なっている。これより過大推定を補正できたと考えられる。

4.5 推定結果の考察

4.2の選好データ併用モデルは、因子得点と所要時間の間の相関が残存していたため、大きな差異を生まなかったと考えられる。4.3の二段階推定モデルでは、考慮すべき個人間の差異が利便性評価値だけでは十分に表現できなかったことが、有意な補正ができなかった理由であると考えられる。4.4の傾向スコア逆数重み付けモデルは所要時間パラメーターの絶対値が最も小さく、さらに標準誤差が小さい推定ができていることから、他の3つのモデルとの差異が統計的に有意となった。

5. おわりに

本研究では、交通手段選択に交通条件などの観測可能特性と、態度や選好などの非観測特性が与える影響の分離方法の既存研究を確認し、東北大学学生の通学行動に関する横断面調査データに適用できる3つの方法の有効性を実証的に確認した。その結果、傾向スコア法の中でも最近に開発された逆数重み付け方法によって、非観測要因を考慮しない交通機関選択モデルの過大推定を補正できることを確認した。本研究では学生の間で公共交通と車両に次いで多い自転車の利用を対象としていない。3手段以上の選択モデルへの適用が今後の課題である。

参考文献

- 1) Mokhtarian P.L., Cao X.: Examining the impacts of residential self-selection on travel behavior: A focus on methodologies, Transportation Research Part B, 42 (2008) 204-228.
- 2) Cao X., Mokhtarian P. L., Handy S. L.: Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings, Transport Reviews, 29,3 (2009) 359-395.
- 3) 松島格也・小林潔司・福井浩: データ内生性を考慮した交通機関選択モデル土木学会論文集 D3, 69,5 (2013) 1_511-1_521.
- 4) Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

(2017年2月2日提出)