

IV-188 パネルサンプルの初期消耗バイアスを修正した動的な買物目的地選択モデル

愛知県土木部 正会員 上田 善之
名古屋大学工学部 正会員 佐々木邦明

1.はじめに

同一の個人や世帯などを繰り返し観測したパネルデータは、選択行動の動的特性を把握するのに非常に有効であると考えられ、これまで欧米を中心として事例研究が行われており、我が国においても、データの収集とともにその事例研究が増えつつある。しかしながら、パネルデータの行動モデルへの適用に関して、いくつかの問題点が考えられており、その一つが消耗バイアスの問題である。一般にパネルサンプルは、調査を重ねるごとに減少していく。そして、回答者がパネル調査から脱落していく時に分析対象とする行動と相関を持った場合に消耗バイアスが発生すると考えられる。そこで本研究では、パネルサンプルの消耗バイアスを修正したモデルを提案し、買物目的地選択行動にその適用を行い、有効性を検討する。

2.初期消耗について

近年、調査経費の低減や調査の信頼性を確保する目的で、郵送で回収する調査が増えてきている。郵送回収に期待しうる有効回答数は一般に低く、調査票の内容にもよるが配布枚数の20～30%程度が多い。表1は国内の主なパネル調査における回収状況を示したものである。

表1 代表的なパネル調査における初期回収状況

	配布数	回収数 (回収率)	有効回収数 (回収率)	備考
甲府買物 パネル	1500	653 (43.5%)		現場配布郵送 回収
千葉モノ レールパ ネル	5279		4264 (80.8%)	訪問配布訪問 回収
堺所要時 間パネル	2290		634 (27.7%)	現場配布郵送 回収
茨木駐車 場パネル	2757	856 (31.0%)	831 (30.1%)	現場配布郵送 回収

これによると、現場配布郵送回収調査は配布数に比

べて、初期回収数あるいは初期有効回答数がかなり少ないことが分かる。このようにサンプルが調査票を受け取ってから最初のwaveにかけて消耗するものを本研究では特に「初期消耗」と定義する。これは今まであまり考慮されることが少なかったが、実際のサンプリングにおいては非常に大きなバイアスを発生させるものになりうると考えられる。

3.消耗バイアスの修正方法

本研究では甲府市の郊外に立地したSC（ショッピングコンプレックス）の来訪者の、休日における買物・交通について調査した甲府買物パネルデータを用いて、買物目的地選択モデル及びパネル調査参加モデルを構築する。両モデルとも効用最大化理論に基づき、ロジットモデルを用いる。

消耗現象が、個人属性と関連していることが既存の研究により報告されているが、それをサンプルの各サブグループが異なる割合で調査から抜け落ちる現象と考えると、パネルサンプリングは各回で抽出比率が次第に変化する層別抽出と考えることができる。この考えに基づき、あるサンプルがグループgからのものであり、かつ選択結果i、特性Xを有する尤度 $f(i, X)$ は次式で表される。

$$f'(i, X) = \frac{P(i|X, \theta)P(X)H(g)}{Q(g)} \quad (1)$$

ここに、

$P(i|X, \theta)$ ：特性X、パラメータ θ のもとで選択肢iを選択する条件付き確率

$P(X)$ ：特性Xの母集団分布

$Q(g)$ ：グループgの母集団シェア

$H(g)$ ：グループgのサンプル内シェア

今回事例研究の対象とする買物目的地選択のような場合、 $Q(g)$ は一般的には求めることはできないが、パネルデータではt時点におけるグループgの滞留人数が分かっており、各wave間での滞留確率が知られた場合、wave0すなわち調査票を配った時点の各グループの人数の期待値を求めることができ、

疑似母集団シェアを式(2)のように計算できる。

$$Q(g|\delta, Z) = \frac{H_{0g}}{\sum_{k \in G} H_k} = \frac{N_{t,g}}{\left(P(g|\delta, Z) \right)^t} \bigg/ \sum_{k \in G} \left(\frac{N_{t,k}}{\left(P(g|\delta, Z) \right)^t} \right) \quad (2)$$

式(2)で示されるパラメータベクトル θ を含んだ母集団シェアを式(1)に代入し、買物目的地選択モデルとパネル調査参加モデルに適用する。

買物目的地選択とパネル調査参加モデルの効用関数の誤差項間には何らかの相関があると考え、個人で共通の誤差項を系列相関項として導入する。

$$U_{i,n}^t = \lambda X_{i,n}^t + v_n + \xi_{i,n}^t \quad (3)$$

$$A_n^t = \vartheta Z_n^{t-1} + v_n + \zeta_n^t \quad (4)$$

ここに、

$U_{i,n}^t$: t 時点における個人 n の代替案 i に対する総効用

A_n^t : t 時点における個人 n のパネルへの参加意向を表す潜在変数

$X_{i,n}^t$: t 時点における個人 n の代替案 i に対する説明変数ベクトル

Z_n^{t-1} : $t-1$ 時点における個人 n の属性ベクトル

λ, ϑ : 未知パラメータベクトル

v_n : 個人 n の両モデルに共通の誤差項

$\xi_{i,n}^t, \zeta_n^t$: 個人 n の真にランダムな誤差項（無相関）

4. モデルの推定

まず本研究で提案したモデルの有用性を検討するために甲府パネルデータを使って消耗バイアスを修正しない買物目的地モデルを推定した結果を表2に示した。選択肢はSC、中心街、最寄り店の3つである。この結果定数項の符号がすべて正となり、パネルサンプルがSCにおけるチョイスベースドサンプルであることによると考えられる。次に消耗バイアスを修正したモデルの推定結果を表3に示す。また消耗モデルの説明変数として性別ダミーを用いた。これは、性別によるグループ分けが、ほぼ一定の滞留確率をもつからである。表4はwave3の来訪者のその後のパネル調査への滞留率を示したものである。表3の結果より定数項の符号がすべて負に変化していることから、初期消耗バイアスを修正することにより、SCを選択した人が調査に残りやすいというチョイスベースドサンプルの影響を緩和することがで

きたと考えられる。

5. おわりに

本研究ではパネル分析の際に問題となる消耗バイアスの除去を目的とし、一貫性および漸近的有効性を持つ推定値が得られるようなモデルを提案した。具体的にはパネル調査参加モデルを用いて母集団シェアを表現し、消耗現象を層別抽出としてとらえることで、サンプルにウェイトをかけるモデルである。その結果として通常の消耗だけでなく初期消耗によるバイアスも取り除かれ、より有効性の高い推定値が得られたと考えられる。今後は本研究のモデルをSPデータを用いたパネル分析などに適用していき、有用性を検討することが必要とされよう。

表2 買物目的地選択モデルのパラメータ推定結果

	推定値	t 値
SC3定数項	1.266	6.8
SC4定数項	1.241	8.4
SC5定数項	1.001	7.9
SC6定数項	1.045	9.3
距離	-0.383	-6.1
交通利便性満足度	0.205	5.6
買物利便性満足度	0.073	2.8
N=796 $\bar{\rho}^2 = 0.28$		

注) SC3はwave3におけるSCの定数項を表す、以下同様

表3 初期消耗バイアスを修正したモデルの推定結果

	推定値	t 値
SC3定数項	-1.530	-8.0
SC4定数項	-1.083	-2.4
SC5定数項	-2.208	-3.5
SC6定数項	-2.236	-3.6
距離	-0.382	-6.0
交通利便性満足度	0.204	5.5
買物利便性満足度	0.073	2.7
N=796 $\bar{\rho}^2 = 0.58$		

表4 男女別滞留人数

	wave3	wave4	wave5	wave6
男	39	19	10	5
女	122	52	36	24