離散選択モデルに関する研究ノート

上智大学経済学部経済学科　竹内ゼミ
藤田光明

要旨

この研究ノートでは，伝統的な需要関数の推定の問題点と，その問題点を解消しうる離散選択モデルの理論的背景やその実証分析の例をまとめたものである．また，その分析において起こった問題に関しては，今後の課題として，解決策のヒントとともにまとめている．

目次

1　はじめに

産業組織論(Industrial Organization :IO)の理論研究において，最も注視されるトピックは企業がいかに行動するかである．しかし，その実証研究では，企業の行動を決める重要な要因の一つである費用関数を実際に観察/推定できないことが問題となる．そこで，新しい実証 IO では，実際の購買活動から観察可能な消費者の行動に注目して需要関数を推定し，そのあとで間接的に費用関数を求めようとしている．

需要関数の推定にあたって，伝統的なアプローチは，例えば二財モデルなどで消費者の効用最大化問題を解き，需要関数を求めるというものである．しかしこれは以下の二点の理由で，実証分析において便利でない．一点目は，実際の分析対象の市場では，たくさんの代替財が存在するため，推定するパラメータが膨大になり，計算が煩雑になること．二点目は，人々の選択は，「複数のものの中から，一つを選ぶ」 (離散選択)ことが多いこと．例えば，ラーメンを食べるとき，二郎，大勝軒，中本，日高屋などから一つ行く店を決めるだろう．離散選択における問題点は，離散的な選択肢があると微分ができないので，需要関数を効用最大化問題の一階の条件から導出することが不可能になるということである．以上の二点の問題を解決するために，実証 IO では，離散選択モデルの推定に関する研究を進めてきた．

2 二項ロジットモデル

2.1 二項ロジットモデルとは

二項ロジットモデル（Binomial Logit :BL）とは，離散選択モデルの中で特に，「結婚するかどうか」や「自動車を保有するかどうか」のような二者択一の問題を扱うモデルである．BL においては，例えば「結婚する」を Y = 1，「結婚しない」を Y = 0 として，説明変数に回帰するという手法がとられる．ここで，潜在変数と呼ばれる Y*を考える．Y*は，Y*の値が臨界値（通常は 0）を越えれば Y=1，超えなければ Y=0 になるというような変数で，$Y^* = \alpha + \beta X + \varepsilon$ と説明変数 X についての線形の関数として表すことを仮定している．また，誤差項の累積密度関数はロジスティクス分布を使う．ここで$P_i = E(Y_i = 1 | X_i)$とすると，

$$log\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta X_i \qquad (1)$$

と表せる．左辺の対数オッズ（これをロジットと呼ぶ）を X に回帰するとパラメータ(α，β)が得られる．

2.2 回帰の例：所得と自動車の保有割合

BL 回帰の一例を挙げる．以下の分析では統計ソフト R を用いた．表 1 は所得と自動車の保有割合を表している擬似データで，例えば一行目は，「所得 200 万の人は 50 人いて，その中で 20 人が自動車を保有している」というように読む．X を所得，P を自動車の保有確率として，これを BL 回帰した結果は表 2 のようになる．

表1　所得と自動車の保有割合

所得	表本数	保有者
200	50	20
400	100	40
600	200	70
800	300	150
1000	200	140
1200	120	100
1400	80	60
1600	40	37

所得の単位は万

引用：浅野・中村(2009)

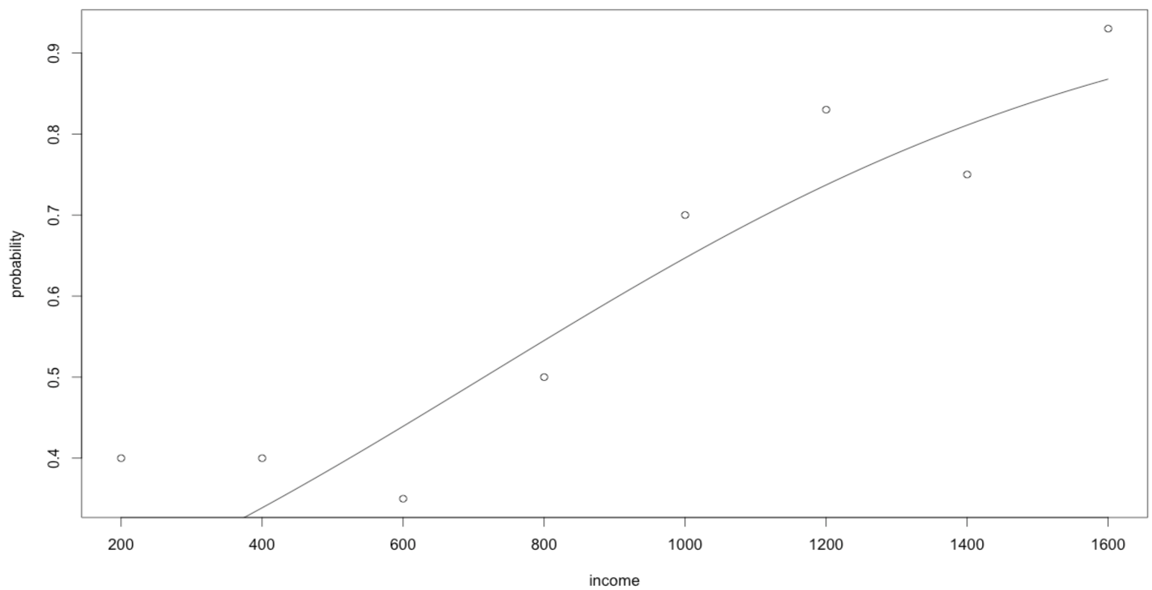
表2　回帰結果

説明変数	係数	有意性
定数	−1.519	***
所得	0.002125	***
R^2	0.794	
Adj R^2	0.725	
F値	11.56	
サンプルサイズ	1090	

*** ・・・ 1%有意. ** ・・・ 5%有意, * ・・・ 10%有意

これを見ると，所得の係数は有意にプラスであることが言える．また，図 1 に，所得の変化に応じた保有比率の予測値を示す．図 1 からも，所得が上昇するにつれ，保有確率も上昇することがわかる．

図 1　所得の変化に応じた保有比率の予測値



線が予測値，点が実現値を表す

3 複数（3 つ以上）の選択肢がある離散選択モデル

前章の二項ロジットモデルでは二者択一の問題を扱っていた．しかし，冒頭の 1 章でもラーメンの選択の例を述べたように，現実では「複数の選択肢から一つを選ぶ」場合が多い．それらを BL で記述するのは不可能であるので，新たにランダム効用と呼ばれるフレームワークを導入する．

Berry(1994)では以下のようにランダム効用を定式化した．

$$U_{ij} = X_j\beta + \alpha P_j + \xi_j + \varepsilon_{ij} \qquad (2)$$
$$\delta_j \equiv X_j\beta + \alpha P_j + \xi_j \qquad (3)$$

i は市場に参加している消費者で i = 1, …. , N, j は市場にある製品の種類を表し，j = 0, 1, …, J とする．ここで j = 0 は外部財（j=1, 2, …. , J 以外の財）としている．X_jは分析者にとって観察可能である製品 j の品質，P_jは製品 j の価格であり，(2)式は製品の品質と価格の線形和でランダム効用を表している．ξ は誤差項 ε の P と相関する部分を取り出したもので，「観察不可能な品質」と解釈することができる．(3)式は，mean utility と呼ばれ，製品 j から得られる消費者共通の効用である．また一般的に，価格と製品の品質には正の相関があるので，観察不可能な品質はモデルの

内生性の問題を生むことになる．
消費者 i が製品 j を買う確率を以下のように記述する．

$$D_{ij} = Prob\{ \varepsilon_{i0}, \ldots, \varepsilon_{iJ}: U^*_{ij} > U^*_{ij'}\ for\ j \neq j' \} \qquad (4)$$

誤差項ベクトルが i.i.d.(independent identically distributed)なら，D_{ij}は全体のマーケットシェアにもなりうる．

4. 多項ロジットモデル

4.1 多項ロジットモデルとは

(4)式の誤差項ベクトルが i.i.d.でガンベル分布に従うと仮定すると, それは多項ロジットモデル(Multinomial Logit : MNL)となり, 以下のように全体のマーケットシェアを記述できる．

$$\tilde{S}_j = \frac{exp(\delta_j)}{1 + \sum exp(\delta j\prime)} \qquad (5)$$

これにより，$\tilde{S}$ (: predicted share)が導出できる．ここで，Berry(1994)では，$\tilde{S}$がデータとしてある $\hat{S}$ (: observed share)とマッチングするようにパラメータを決めるという手法をとっている．これによって，

$$log\tilde{S}_j - log\tilde{S}_0 = X_j\beta + \alpha P_j + \xi_j \qquad (6)$$

が得られる．

4.2 操作変数の導入

Berry(1994)では価格の内生性の問題に対処するために操作変数を導入した．これによって，二段階最小二乗法(Two-stage Least Squares: 2SLS)で，(6)式を推定することができるようになった．具体的には，まず価格と相関するが，誤差項（需要ショック）と相関しない Z を操作変数とする．一段階目で価格を Z に回帰して，価格の推定値$\hat{P}(Z)$を作る．そして二段階目で，(6)の左辺を X と$\hat{P}(Z)$に回帰する．そうすることで，(6)式のパラメータを推定することが可能になる．

4.3　回帰の例：シリアル市場

MNL 回帰の一例を挙げる．表 3 は 1992 年のアメリカでの，50 種類のシリアルとその他のシリアル(=外部財)それぞれについて，市場シェア，広告費，カロリー，価格などのデータをまとめたものである．ここで, 離散選択モデルの条件として,「選択肢が網羅的である」というのがあるが，ここでは全ての（対象マーケット内の）人が，1 年の間に何らかのシリアルを買っていることを仮定している．実際，97.1%の人が買っているので，現実味のない仮定ではない．

これをまず通常の OLS で回帰してみる．(6)式において, P を transaction price, X ＝ [広告費 ｜カロリー ｜脂質 ｜砂糖]として，左辺を X と P に回帰すると表 4 の結果が得られる．

表4 OLS回帰結果

説明変数	係数	有意性
広告費	0.1196	***
カロリー	−0.0069	***
脂質	0.1048	*
砂糖	−0.0216	
Transaction Price	−0.07296	***
R^2	0.976	
Adj R^2	0.974	
F値	368.2	
サンプルサイズ	50	

*** ･･･ 1%有意, ** ･･･ 5%有意, * ･･･ 10%有意

表3 シリアル市場

	Name		Transaction Price	Shelf Price	広告費	シェア	セグメント	カロリー	脂質	砂糖
1	KG	CornFlakes	1.81	1.95	7.109	5.67	Fam	100	0	2
2	GM	Cheerios	3.16	3.47	7.287	4.38	Fam	110	2	1
3	KG	RiceKrispies	2.96	3.2	6.034	4.04	Fam	120	0	3
4	KG	FrostedFlakes	2.52	2.68	7.867	3.82	Fam	120	0	13
5	KG	RaisinBran	2.34	2.5	5.591	2.73	Fam	200	1.5	18
6	GM	Total	3.61	4.04	3.926	2.36	Adult	110	1	5
7	GM	HoneyNutChee	3.14	3.41	4.03	2.26	Fam	120	1.5	11
8	KG	SpecialK	3.48	3.78	3.531	2.16	Adult	110	0	3
9	PT	GrapeNuts	2.14	2.29	6.74	2.12	Adult	200	1	7
10	NB	SpoonSizeShd	2.81	3.05	0.025	2.08	Adult	170	0.5	0
11	QK	100%Natural	2.24	2.55	1.612	1.96	Adult	220	8	13.5
12	KG	FrostedMiniWh	2.62	2.75	6.106	1.84	Adult	170	1	10
13	KG	NutriGrain	2.87	3.1	2.508	1.55	Adult	100	1	0
14	KG	Mueslix	3.31	3.58	1.975	1.53	Adult	200	4	13
15	GM	Wheaties	2.55	2.86	2.257	1.52	Fam	110	1	4
16	PT	RaisinBran	2.23	2.57	4.361	1.46	Fam	190	1	20
17	RL	Muesli	3.34	3.93	0.215	1.26	Adult	210	2.7	14
18	KG	CornPops	3.51	3.69	3.198	1.46	Fam	120	0	14
19	GM	RaisinNutBran	2.98	3.22	1.659	1.35	Adult	210	4.5	16
20	GM	Basic4	3.27	3.63	2.51	1.31	Adult	210	3	12
21	GM	CocoaPus	3.46	3.67	2.097	1.28	Kids	120	1	14
22	GM	GoldenGraham	3.24	3.54	2.953	1.24	Kids	120	1	11
23	GM	Cinn.ToastCru	3.36	3.56	2.963	1.23	Kids	130	3	10
24	KG	FrootLoops	3.53	3.76	3.11	1.2	Kids	120	1	15
25	KG	LowFatGranola	2.68	3.1	2.327	1.17	Adult	190	3	12
26	GM	Trix	3.96	4.22	3.236	1.13	Kids	120	1.5	13
27	GM	Triples	2.33	2.8	3.036	1.12	Adult	120	1	6
28	KG	Crispix	3.28	3.49	3.225	1.12	Adult	110	0	3
29	GM	Kix	3.67	3.93	3.801	1.08	Kids	120	0.8	6
30	GM	LuckyCharms	3.45	3.72	3.079	1.08	Kids	120	1	13
31	GM	AppleCinn.Che	3.02	3.35	3.12	1.06	Fam	120	2	13
32	KG	CracklinOatBra	3.19	3.51	2.279	1.06	Adult	190	6	15
33	NB	BigBiscuitShd	2.79	3.05	0	0.99	Adult	156	1.2	9.6
34	PT	HoneyBunches	2.85	3.18	3.749	0.95	Adult	125	2.2	6
35	PT	GreatGraines	2.9	3.43	2.648	0.89	Adult	215	5.5	10.5
36	GM	OtmlRaisinCris	2.71	3.04	1.641	0.97	Adult	210	2.5	19
37	QK	OatSquares	2.43	2.71	1.472	0.94	Adult	220	3	9
38	RL	RiceChex	3.4	3.53	0.875	0.89	Adult	120	0	2
39	GM	TotalRaisinBra	3	3.5	1.874	0.89	Adult	180	1	19
40	KG	Product19	3.38	3.7	1.408	0.89	Adult	100	0	4
41	KG	AppleJacks	3.64	3.91	1.465	0.84	Kids	120	1	16
42	QK	CaptCrunch	2.55	2.86	1.714	0.83	Kids	105	1.5	11.5
43	NB	ShreddedWhea	2.82	3	2.925	0.8	Adult	160	0.5	0
44	PT	FruityPebbles	3.21	3.48	1.71	0.83	Kids	110	1	12
45	GM	Clusters	3.14	3.52	1.425	0.78	Fam	210	3.5	14
46	KG	CinnamonMini	2.75	3.14	0.002	0.76	Fam	120	0.5	12
47	KG	DoubleDipCrun	3.01	3.52	1.454	0.73	Adult	110	0	11
48	GM	MultiGrainChee	3.34	3.74	2.52	0.75	Fam	110	1	15
49	PT	Honeycomb	3.4	3.67	2.567	0.74	Kids	110	0	11
50	QK	Popeye	1.77	1.77	0	0.67	Kids	120	1	13.3
51	basketof	他のブランド	2.68		0.645	24.29				

データソース：Leading National Advertisers(1990-1993)

Matthew Shumのレクチャーノートから引用

ここで，モデルの内生性，つまり価格と残差の相関をチェックしてみる．P と残差をプロットしたものが図 2 である．

図 2 価格(P)と残差(.resid)のプロット

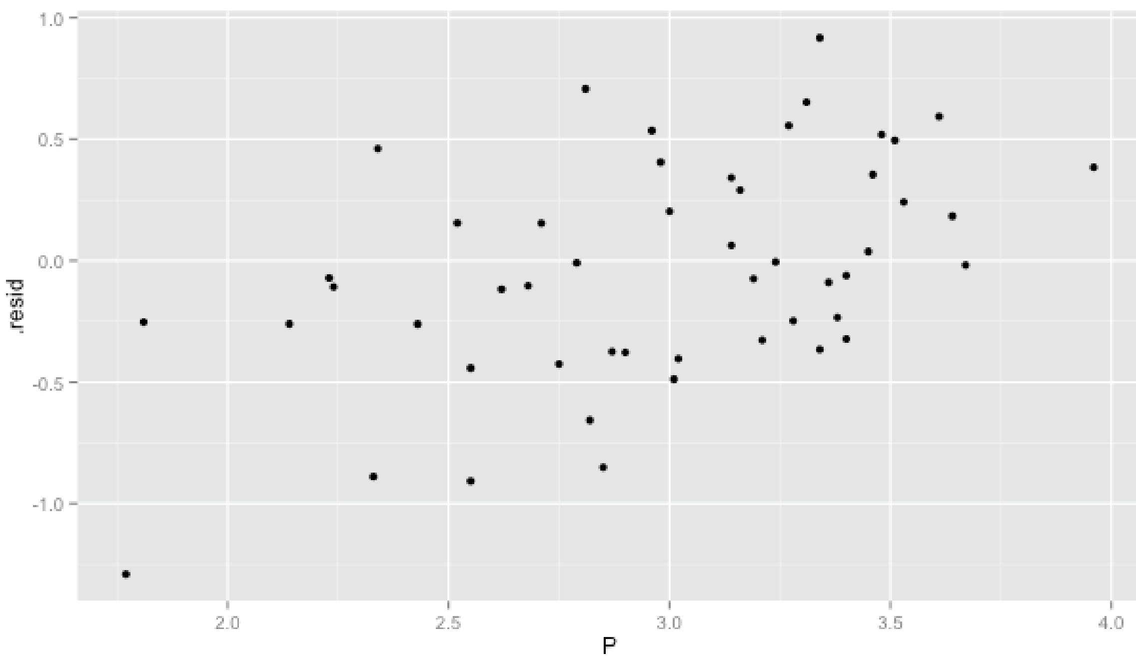


図 2 から明らかな正の相関がわかる．相関係数はおよそ 0.48 となっている．

そこで，Shelf price を操作変数として 2SLS で回帰をする．操作変数の選択に際して，Chintangunta, Dude and Singh(2003)では取引費用の操作変数として卸売価格を利用していた．これは製品の販売店の市場支配力が取引費用には反映しているが，卸売価格には反映されていないため，誤差項である需要ショックと卸売価格が無相関だと考えられるためである．これにならって，shelf price を transaction price の操作変数として推定する．しかし，shelf price と残差の相関係数はおよそ 0.47 で transaction price と残差の相関係数とほとんど変わらない．この原因は shelf price も transaction price と同様に市場支配力が反映されているためだと考えられる．したがって，shelf price は操作変数として不十分であるが，shelf price と残差の相関係数は transaction price のそれよりわずかに小さいので，ここではエクササイズ的に，shelf price を操作変数として 2SLS を行う．推定結果は表 5 である．

表5 2SLSの回帰結果

説明変数	係数	有意性
広告費	0.113	***
カロリー	−0.0065	***
脂質	0.1111	*
砂糖	−0.0209	
Transaction price	−0.7461	***
R^2	0.977	
Adj R^2	0.974	
F値	373.4	
サンプルサイズ	50	

*** ・・・ 1%有意. ** ・・・ 5%有意, * ・・・ 10%有意

表 4 の結果と比較すると，価格の係数の絶対値が少し大きくなっていることがわかる．これを解釈するために，回帰係数の確率極限の話を紹介する．回帰係数 b の確率極限は $\mathrm{plim}(b) = \beta + \mathrm{Cov}(X, \varepsilon) / \mathrm{Var}(X)$ と書くことができる．つまり，回帰係数と誤差項に正の相関がある今回の場合は，通常の OLS だと回帰係数の絶対値を過小評価してしまう．2SLS では，価格の回帰係数の絶対値は大きくなっているので，内生性の問題はわずかながら軽減されていることがわかる．

5. 今後の課題

5.1 操作変数の選択

4 章の MNL の 2SLS において，transaction price の操作変数として shelf price を使うことは適切でないことがわかった．したがって，別の操作変数を考える必要がある．操作変数の例として，Genesove and Mullin (1998)では米国の製糖市場の需要関数の推定にキューバからの原材料の輸入費を用いたり，Hausman and Leonard (2002)では，地理的に異なる他の市場に投入されている同一商品の価格を用いたりしている．また, Berry, Levinsohn and Pakes (1995)では特性値と呼ばれる，財の特性を表す変数を操作変数として推定している．

5.2 IIA 特性の解消

MNL で, 任意の 2 つの選択肢の選択確率を現した時に, それがその選択肢の mean utility のみから決定されている，つまり選択肢集合に含まれるほかの選択肢の影響を受けなくなる.

$$\frac{D_j}{D_j'} = \frac{exp(\delta_j)}{exp(\delta_j')} \qquad (7)$$

これが IIA（Independence of Irrelevant Alternative）特性と呼ばれるもので，邦訳すると「無関係な選択肢からの独立」となる．IIA 特性の問題点は以下で説明する，赤バス/青バス問題のような事態を引き起こす点である.

赤バス/青バス問題とは，例えばある市に 2 つの交通手段：歩く(W)，赤いバスに乗る(RB)があるとする．そのシェアは W:50% RB:50%でその比:W/RB = 1 と表すことできる．ここで新たに 3 つめの選択肢として青いバス(BB)が参入してきたとする．IIA 特性は以前と同様に W/RB = 1 を示す．しかし青いバスは赤いバスと完全代替財なので，実際には W:50% RB:25% BB:25%, W/RB = 2 となるはずである.

この IIA 特性を解消するために，Berry, Levinsohn and Pakes (1995)はランダム係数と呼ばれる，それぞれの消費者固有の係数を導入した．これによってマーケットシェアにおいて，IIA 特性が解消された.

6. 参考文献

浅野・中村(2009):『計量経済学』有斐閣.

Berry, S. (1994): “Estimating Discrete Choice Models of Product Differentiation,” RAND Journal of Economics, 25, 242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995): “Automobile Prices in Market Equilibrium,” Econometrica, 63, 841–890.

Chintagunta, P., Dube, J., and Singh, V. (2003): “Balancing Profitability and Customer Welfare in a Supermarket Chain,” Quantitative Marketing and Economics, 1, 111-147.

Hausman, J. and Leonard, G. (2002): “The competitive effects of a new product introduction: A case study,” Journal of Industrial Economics, 50, 237–263.

Genesove, D. and Mullin, P. (1998): “Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914,” RAND Journal of Economics, 29, 355-377.

Matthew Shum のレクチャーノート,
< http://people.hss.caltech.edu/~mshum/gradio/ioclass.html>.