

Correcting for endogeneity in behavioral choice models with social influence variables

Walker, J., Ehlers, E., Banerjee, I., Dugundji, E.

Transportation Research Part A, Vol.45, pp.362-374, 2011.

社会的影響変数を用いた行動選択モデルにおける内生性の修正

2011/10/17(月)

秋期集中論文ゼミ

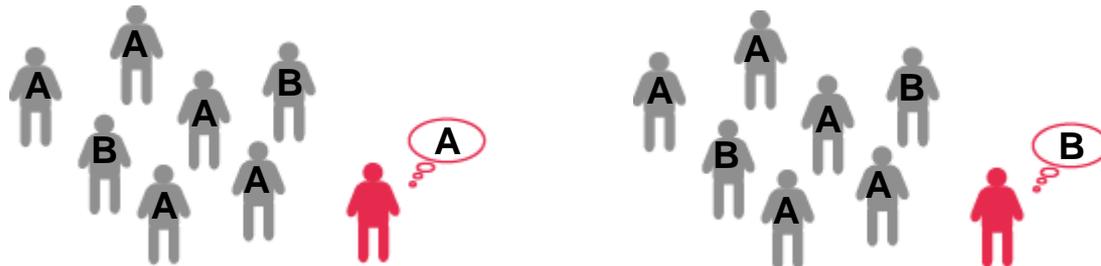
B4 福士薫

# 発表の流れ

- はじめに
  - 社会的影響とは
- 問題点
  - 内生性とは
- 研究の目的
- 提案する手法
  - BLP法, 操作変数法
- 手法の適用
- 推定結果
- まとめ

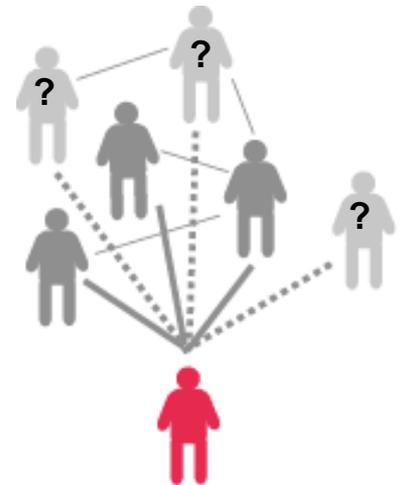
# はじめに

- 心理学や行動経済学において、行動選択における社会的影響の重要性が注目されている。
- Manski (1995)の社会集団相互作用に関する3つの仮説
  - 内生効果:個人がある行動をとる傾向は、集団の中でどれだけその行動が選択されているかによって異なる。
  - 文脈効果:個人がある行動をとる傾向は、集団の構成員の(平均的な)個人属性に応じて異なる。
  - 相関効果:同じ集団の中の個人は、同様の社会環境のもとにある、あるいは同様の属性を有しているので、同じような振る舞いをする傾向がある。



# はじめに

- McFadden(2009)
  - 個人の行動選択は
    - 他者からの情報, 集団の中で蓄積された情報
    - 他者の行動からの類推
    - 他者により課せられる制約 ...により影響を受ける.
- 離散選択モデルは個人の行動選択に基づいているため, 多様な意思決定者間の相互作用を扱うのは難しい.
  - 意思決定者に関係する社会的ネットワークの範囲の把握.
  - 意思決定者それぞれで, 他者から受ける影響の度合いも異なる.



# はじめに

- このようなネットワークの厳密な把握をせず，総体としての社会的影響を示すfield effect変数を用いて，行動選択における社会的影響を離散選択モデルに組み込みたい。
  - field effect変数: peer groupにおける，ある選択肢を選んだ人の割合
  - peer group: 社会経済的に近い属性を有する，あるいは居住地などが空間的に近接している意思決定者の集団

- field effect変数 ( $F_{in}$ ) を取り入れた効用関数

$$U_{in} = V(x_{in}, S_n; \beta) + \gamma F_{in} + \varepsilon$$

意思決定者  $n$  の属する peer group における，  
選択肢  $i$  を選んだ人の割合

→ 社会的影響の度合いを示すパラメータ  $\gamma$  を推定したい。

# 問題点

- しかし, このような社会的影響を説明するための変数をモデルに取り入れると複雑化しやすい. なぜか?

→endogeneity(内生性)の問題.

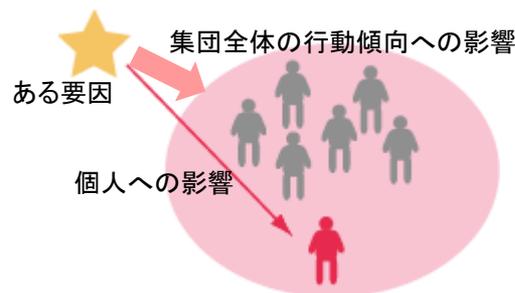
- Manski(1993)の‘reflection problem’

- ある集団の中での平均的な行動が個人の行動にどの程度影響を及ぼしているのかを特定するのは困難.
- データで観測しきれない要因が, 個々の意思決定にも( $\varepsilon$ に反映)集団全体の行動傾向( $F_{in}$ に反映)にも影響を及ぼす.

→モデルの誤差項と説明変数が相関を持つ. (=内生性)

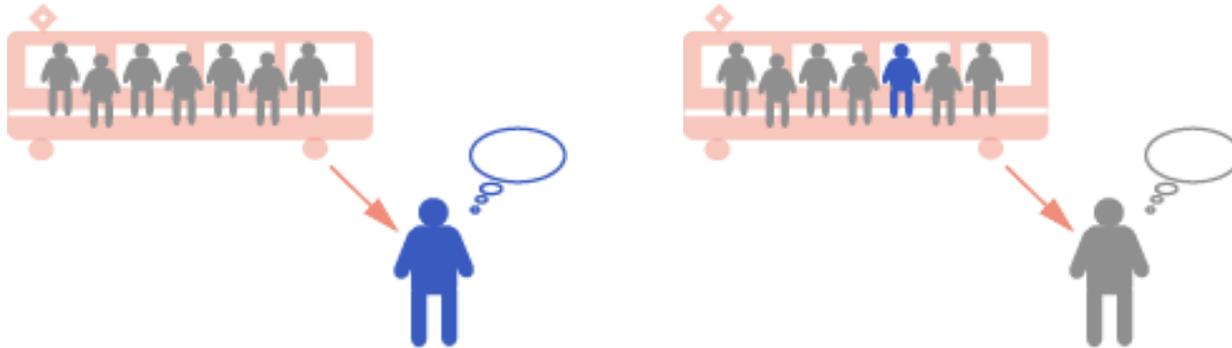
$$U_{in} = V(x_{in}, s_n; \beta) + \gamma F_{in} + \varepsilon$$

内生変数



# 内生性

- 例えば,
  - ある人が徒歩か電車かを選択する場合, 所要時間や費用などの他に, 混雑の度合い(=他の人がどれくらい電車を使うか)も重要な要因になりうる.
  - この場合, ある人の選択は他者に影響を受けるが, 同時に他者の選択にも影響を及ぼしていることになる.



- このような相互作用をfield effect変数としてモデルに取り入れたいたが, 内生性のためにパラメータ $\gamma$ の推定値が過大になったり, 一致性を持たなかったりする.

→内生性の修正が必要.

# 研究の目的

- 離散選択モデルにおける内生性の修正法を提案し、社会的影響を示すfield effect変数の、内生性バイアスが修正されたパラメータを求めたい。

$$U_{in} = V(x_{in}, S_n; \beta) + \gamma F_{in} + \varepsilon$$

- 線形回帰モデルでは、内生変数は比較的扱いやすい。
  - 操作変数法: 線形回帰モデルにおける内生変数に対処する推定法。
- 離散選択モデルの場合、内生性の修正を行うと効用関数の誤差項が変化することがあり、モデルのscaleが変わってしまう。  
→より扱いやすい修正法を提案したい。

# 提案する手法

- BLP法(Berry, Levinsohn and Pakes, 1995 and 2004. )  
離散選択モデルにおける内生性の修正方法. 内生変数を定数で置き換えることで内生性をなくす.
  - この方法では先に述べたような修正によるスケールの変化がなく, 推定がしやすい.
  - しかし定数で置き換えてしまうため, これだけでは内生変数のパラメータは推定できない.→求めたかったfield effect変数のパラメータがわからない.
- そこで以下のような2段階の手法を提案する.

step1. BLP法を用いて選択モデルから内生変数(= field effect変数)を取り除く.

step2. 操作変数法により, field effect変数のパラメータを推定する.

# BLP法

- market levelで生じる内生性を修正する推定法(ここではmarketはpeer group.)
- 個人単位の相互作用ではなく, 本研究で扱おうとしているpeer group全体による社会的影響の内生性を修正できる.

## BLP法の流れ

- 誤差項の分割
- 内生変数 → market specific定数への置き換え
- market specific定数の推定

# BLP法

- 誤差項の分割 :  $\varepsilon = \dot{\varepsilon}_{in} + \ddot{\varepsilon}_{in}$  とする.

$$U_{in} = V(x_{in}, S_n; \beta) + \gamma F_{in} + \varepsilon$$

$\dot{\varepsilon}_{in}$  :  $F_{in}$  と相関なし  
 $\ddot{\varepsilon}_{in}$  :  $F_{in}$  と相関あり

$$\rightarrow U_{in} = [\gamma F_{in} + \ddot{\varepsilon}_{in}] + V(x_{in}, S_n; \beta) + \dot{\varepsilon}_{in}$$

内生性の問題を  
生じる部分.

- 意思決定者  $i$  の属する peer group を  $m$  とすると,

$$U_{in_m} = [\gamma F_{im} + \ddot{\varepsilon}_{im}] + V(x_{in_m}, S_{n_m}; \beta) + \dot{\varepsilon}_{in_m}$$

どの説明変数とも  
相関を持たない誤差項.

peer group  $m$  にとっての、選択肢  $i$  の平均的な  
望ましさ. peer group ごとで異なるが、 $m$  内の  
意思決定者間では変わらない.

意思決定者個人によって異なる  
効用を示す部分.

# BLP法

- 内生変数→market specific定数への置き換え  
market specific定数:  $\alpha_{im} = [\gamma F_{im} + \dot{\epsilon}_{im}]$  として,

$$U_{in_m} = \alpha_{im} + V(x_{in_m}, S_{n_m}; \beta) + \dot{\epsilon}_{in_m}$$

各選択肢×各peer groupの固有定数.  
peer group内での平均的な社会的影響を示す.

ここで,  $\alpha_{im}, x_{in_m}, S_{n_m}, \dot{\epsilon}_{in_m}$  はそれぞれ相関を持たない.  
→内生性が取り除かれた.

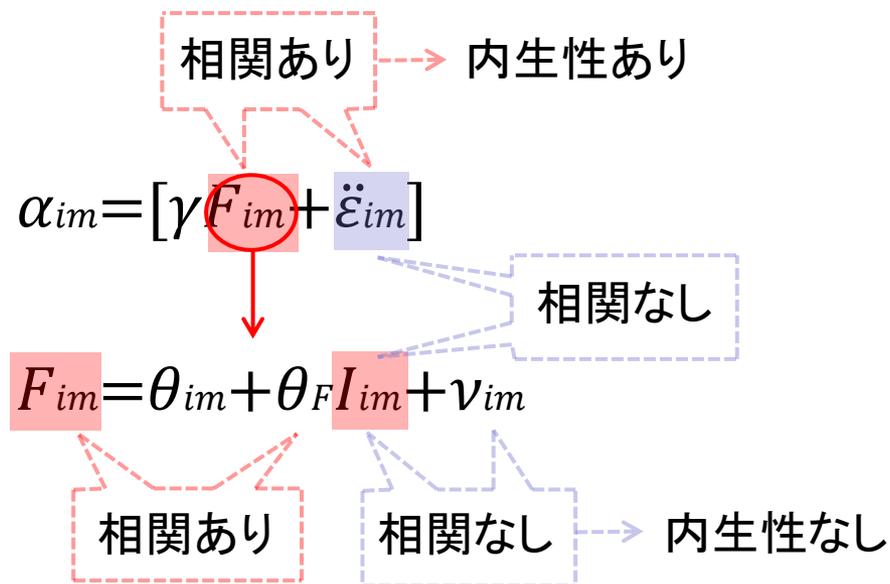
- market specific定数の推定
  - $x_{in_m}, S_{n_m}$  のデータを用いて,  $\alpha_{im}, \beta$  を推定できる.

# BLP法

- BLP法により、社会的影響を示す  $\alpha_{im}$  を推定することはできる。  
しかし、もともと知りたかったのは  $F_{im}$  の内生性バイアスが修正されたパラメータ  $\gamma$ 。
- もう一度  $\alpha_{im} = [\gamma F_{im} + \varepsilon_{im}]$  に戻る。  
 $\alpha_{im}$  の推定値  $\hat{\alpha}_{im}$  と  $F_{im}$  が分かっているため、回帰分析(最小二乗法)により  $\gamma$  を推定することはできる。
- しかし、 $F_{im}$  は内生性を持ったままなので、推定される  $\gamma$  はバイアスがかかったまま。  
→操作変数法。

# 操作変数法

- 内生変数( $F_{im}$ )を操作変数( $I_{im}$ )とパラメータ( $\theta_{im}, \theta_F$ ), 誤差項( $v_{im}$ )で示す. ただし,  $F_{im}$  と  $I_{im}$  は強い相関を持つが,  $\varepsilon_{im}$  と  $I_{im}$  は相関を持たない.



- $F_{im}$  と  $I_{im}$  が分かっているなので,  $\theta_{im}$  と  $\theta_F$  を推定できる.

# 操作変数法

- $\theta_{im}$ と $\theta_F$ の推定値 $\hat{\theta}_{im}$ ,  $\hat{\theta}_F$ によって,  $\hat{F}_{im}$ を求める.

$$\hat{F}_{im} = \hat{\theta}_{im} + \hat{\theta}_F I_{im}$$

$$\begin{aligned}\alpha_{im} &= \gamma F_{im} + \ddot{\varepsilon}_{im} \\ &= \gamma(\hat{F}_{im} + C + \varepsilon) + \ddot{\varepsilon}_{im} \\ &= \gamma C + \gamma \hat{F}_{im} + \gamma \varepsilon + \ddot{\varepsilon}_{im} \\ &= \gamma_i + \gamma_F \hat{F}_{im} + \ddot{\varepsilon}_{im}\end{aligned}$$

$F_{im}$ と $\hat{F}_{im}$ との誤差

相関なし → 内生性なし

- $\alpha_{im}$ と $\hat{F}_{im}$ が分かっているので,  $\gamma_i$ と $\gamma_F$ を推定できる.

→ この $\gamma_F$ が, 求めようとしていた $F_{im}$ の内生性バイアスが修正されたパラメータである.

# 手法の適用

- **通勤手段選択モデル**(車, 公共交通, 自転車)
  - 1992~1997年にアムステルダム, アムステルベンで行われたアンケートのデータから, 自宅→職場へのトリップにのみ注目.

仮説① ある個人の通勤手段選択は, 類似した属性を有する他者(peer group)の行動, 信条, 嗜好などに依存している.

仮説② 意思決定者をpeer groupに分け, 選択モデルにfield effect 変数を取り入れると, 内生性によりfield effect変数のパラメータにバイアスがかかり, 推定値が過大になる.

→提案した2段階の手法(BLP法, 操作変数法)により実際のデータを用いた推定を行い, これらの仮説を検証する.

# 手法の適用

## 推定の流れ

- peer groupの定義 ( $\alpha_{im}$ ,  $F_{im}$  の  $m$ )
- [model1] field effect変数の導入
- [model2] BLP法の適用
- 操作変数の定義 ( $I_{im}$ )
- [model3] 操作変数法の適用

# peer groupの定義

- **spatial reference group**: 空間的近接性に基づくpeer group.
  - 近くに住んでいる人からの方がより大きく影響を受け, 類似した行動をとりやすいはず.
  - 67の郵便番号によるグループ分け.
- **social reference group**: 社会経済的属性に基づくpeer group.
  - 友人や同僚は同じくらいの収入の人が多くいはずであり, 居住地にかかわらず類似した行動をとりやすいはず.
  - 6段階の収入の階層によるグループ分け.
- 意思決定者は, 自身の属するpeer groupによってのみ影響を受け, 他のpeer groupからは影響を受けない(影響を無視できる)と仮定している.

# model 1 . field effect変数の導入

- peer groupの定義より, 1人の意思決定者はspatial/socialという2通りのpeer groupに属することになる.  
→1人につき2通りのfield effect変数. ( $F_{im(spatial)}$ ,  $F_{im(social)}$ )

$$U_{in} = \gamma F_{im(spatial)} + \gamma F_{im(social)} + V(x_{in}, s_n; \beta) + \varepsilon \quad \dots (\text{model 1})$$

$m(spatial)$ ,  $m(social)$ , の中での,  
各選択肢の平均的な好ましき.

- $\gamma$ は本当は  $\gamma_{spatial}$  と  $\gamma_{social}$ . ただし推定の結果, 両者を区別してもモデルの精度に有意な差が見られないため,  $\gamma_{spatial} = \gamma_{social} = \gamma$  とする.

## model2. BLP法の適用

- 67+6のspatial/social reference groupsの中での、車・公共交通・自転車の各選択肢の望ましさ $F_{im}$ を、それぞれ market specific定数 ( $\alpha_{im} = [\gamma F_{im} + \epsilon_{im}]$ ) で置き換える.
- ただし、各選択肢、各peer groupにおいて、定数項はどれか1つは固定 (=0) とする. このため、  
(66+6のspatial/social reference groups) × (車・公共交通の2つの選択肢)  
計144の market specific定数 ( $\alpha_{im}$ ) を他のパラメータと共に推定.

$$U_{inm} = \alpha_{im(spatial)} + \alpha_{im(social)} + V(x_{inm}, s_{nm}; \beta) + \epsilon \quad \dots (\text{model2})$$

# 操作変数の定義

- $\alpha_{im} = [\gamma F_{im} + \ddot{\epsilon}_{im}]$  の  $\gamma$  を求めるために、操作変数法を適用する.

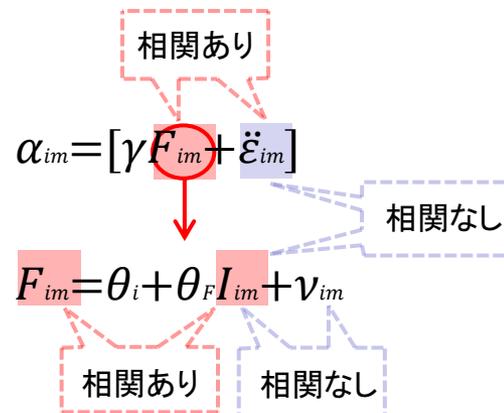
$$F_{im} = \theta_i + \theta_F I_{im} + v_{im} \quad \dots \hat{\theta}_i, \hat{\theta}_F \text{ を求める}$$

$$\hat{F}_{im} = \hat{\theta}_i + \hat{\theta}_F I_{im} \quad \dots \hat{F}_{im} \text{ を求める}$$

$$\alpha_{im} = \gamma_i + \gamma_F \hat{F}_{im} + \ddot{\epsilon}_{im} \quad \dots \hat{\gamma}_i, \hat{\gamma}_F \text{ を求める}$$

- それぞれのpeer groupにおける  $F_{im}$  について、以下のような性質を満たす  $I_{im}$  の定義が必要.

- $F_{im}$  と強い相関を持つ
- $\ddot{\epsilon}_{im}$  とは相関を持たない.



# 操作変数の定義

- spatial reference groupの場合:

$I_{im} = (\text{隣接した郵便番号区の } F_{im} \text{ の平均値})$

- 交通ネットワークは空間的に連続しているため、隣接する地区での交通手段の選択傾向は類似しているはずである。(  $F_{im}$  との強い相関)
- 個人の選択に強く影響を及ぼすのは同じ地区内に住む人々で、他の地区の人々からの影響は無視できるはずである。(  $\ddot{e}_{im}$  と相関を持たない)

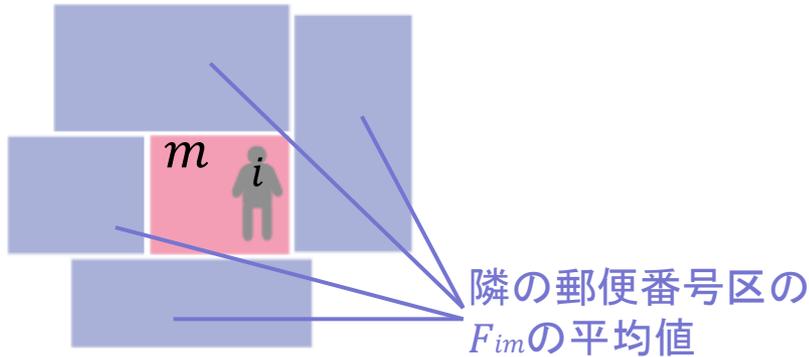
- social reference groupの場合:

$I_{im} = (1つ下の収入階層の } F_{im})$

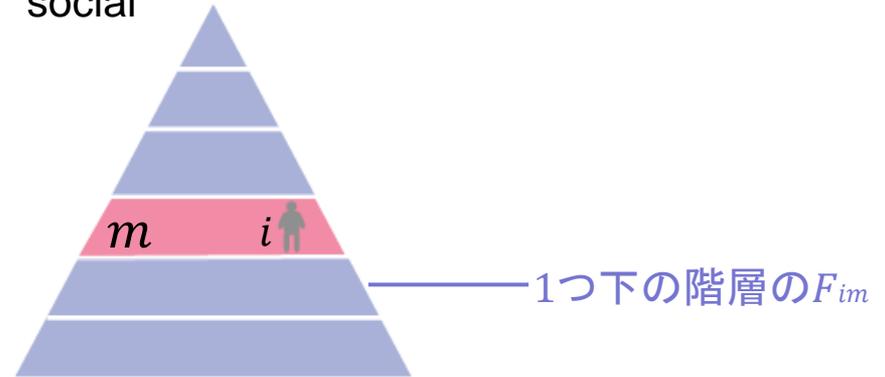
- 収入階級が近い人々の選択傾向は類似しているはずである。(  $F_{im}$  との強い相関)
- ただし、ステータスを求める傾向:個人の選択傾向は自分より高い収入階級の選択傾向に影響を受ける可能性がある。(  $\ddot{e}_{im}$  との相関あり. 操作変数として適切でない)

# 操作変数の定義

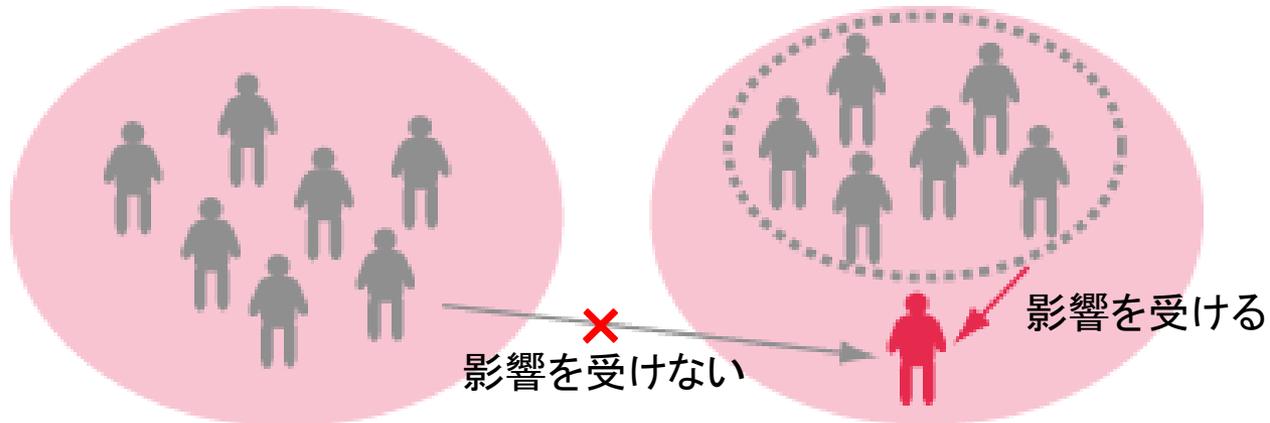
spatial



social



- 空間的/社会経済的に近い属性を持つpeer group同士の選択傾向は類似するが( $I_{im}$ と $F_{im}$ との強い相関), 個人レベルでは他のpeer groupからの影響は受けない( $I_{im}$ と $\varepsilon_i$ は相関を持たない)という仮定.



# model3. 操作変数法の適用

- 定義した操作変数を用いて、操作変数法を適用する.

$$\alpha_{im} = \gamma_i + \gamma_F \hat{F}_{im} + \dot{\epsilon}_{im} \quad \dots(\text{model3})$$

- これによって内生性バイアスの修正された推定値  $\gamma_F$  が求められ, field effect変数が選択モデルの中でどの程度影響を及ぼしているのかがわかる.
- 内生性バイアスによってfield effect変数の影響は過大に推定されると考えられるので, BLP法で置き換えたときそのまま ( $\alpha_{im} = [\gamma F_{im} + \dot{\epsilon}_{im}]$ ) 推定した  $\gamma$  の値よりも,  $\gamma_F$  の推定値の方が小さくなるはずである. (仮説②)

# 推定結果

Mode choice to Work (Logit) Alternatives: Auto, Transit, Bike	Base Model			Choice Models with Field Effect Variable (Endogenous)			First stage of BLP: Choice Models with Market-specific Constants *		
	Estimate	StdErr	t-test	Estimate	StdErr	t-test	Estimate	StdErr	t-test
Parameter (& relevant utility)									
Sociality Field Effect (auto, transit, bike)				2.49	0.20	12.3			
Number of observations	2888			2888			2888		
Number of parameters	20			19			160		
Log-likelihood at convergence	-2406			-2326			-2288		
Adjusted rho-square	0.230			0.256			0.223		

統計的に有意.

尤度は上がったが、修正済み尤度比は下がっている。

精度が上がった。

$$U_{in} = V(x_{in}, S_n; \beta) + \varepsilon$$

$$U_{inm} = \alpha_{im(spatial)} + \alpha_{im(social)} + V(x_{inm}, S_{nm}; \beta) + \varepsilon$$

$$U_{in} = \gamma F_{im(spatial)} + \gamma F_{im(social)} + V(x_{in}, S_n; \beta) + \varepsilon$$

[model2] BLP法適用後

[model1] field effect変数導入後

# 推定結果

- field effect変数を導入したことで有意な結果が得られた。

→ 仮説① ある個人の通勤手段選択は、類似した属性を有する他者(peer group)の行動、信条、嗜好などに依存している。

が証明されたといえる。

- BLP法を適用しただけでは、内生性は取り除けるが必ずしもモデルの精度は上がらない。
- 各選択肢の性質(所要時間、費用など)や意思決定者の社会経済的属性などのパラメータの推定値は、field effect変数導入前、導入後、BLP法導入後で有意な差が見られなかった。

# 推定結果

$$F_{im} = \theta_i + \theta_F I_{im} + \nu_{im}$$

	Estimate	StdErr	t Stat
Intercept Spatial (transit)	-0.00897	0.02055	-0.4
Intercept Spatial (auto)	0.0582	0.0281	2.1
Intercept Social (transit)	-0.00704	0.06815	-0.1
Intercept Social (auto)	0.0846	0.0679	1.2
Field effect instrument	0.756	0.094	8.0
Number of observations	144		
Adjusted R Square	0.540		

$I_{im}$  の定義は妥当.

内生性修正後は、パラメータの推定値・t値ともに下がっている。  
値が下がってもなお統計的には有意である。

[model2] BLP法で置き換えただけの段階.

Dependent variable: Market specific constants  
from choice model

$$\alpha_{im} = \gamma F_{im} + \varepsilon_{im}$$

	Estimate	StdErr	t Stat
Intercept Spatial (transit)	1.21	0.06	21.5
Intercept Spatial (auto)	0.250	0.070	3.6
Intercept Social (transit)	-3.45	0.19	-18.5
Intercept Social (auto)	-2.09	0.19	-11.2
Field effect	2.93	0.19	15.2
Number of observations	144		
Adjusted R Square	0.897		

[model3] 操作変数法適用後.

Dependent variable: Market specific constants from  
choice model

$$\alpha_{im} = \gamma_i + \gamma_F \hat{F}_{im} + \ddot{\varepsilon}_{im}$$

	Estimate	StdErr	t Stat
Intercept Spatial (transit)	1.18	0.09	13.1
Intercept Spatial (auto)	0.559	0.146	3.8
Intercept Social (transit)	-3.53	0.30	-11.8
Intercept Social (auto)	-1.96	0.30	-6.5
Fitted field effect	1.48	0.54	2.7
Number of observations	144		
Adjusted R Square	0.751		

# 推定結果

- さらに, model1の段階, model2の段階, model3の段階で field effect変数のパラメータとt値を比較すると以下のようなになる.
- Hausman検定により, 内生性によって本当にパラメータにバイアスがかかっていたのかどうかを検証.

Estimator	Estimate	StdErr	t-test	Hausman test Corrected vs Uncorrected	
				t-test	conclusion
[model1] $U_{im} = \gamma F_{im(spatial)} + \gamma F_{im(social)} + V(x_{in}, S_n; \beta) + \varepsilon$ Choice - uncorrected	2.49	0.20	12.3	2.0	reject null of no endogeneity (i.e., 2.49 is significantly higher than 1.48)
[model2] $\alpha_{im} = \gamma F_{im} + \varepsilon_{im}$ Regression (OLS) - uncorrected	2.93	0.19	15.2	2.9	reject null of no endogeneity (i.e., 2.93 is significantly higher than 1.48)
[model3] $\alpha_{im} = \gamma_i + \gamma_F \hat{F}_{im} + \ddot{\varepsilon}_{im}$ Regression (2 stage IV) - CORRECTED	1.48	0.54	2.7		内生性によりパラメータが過大に推定されていた.

内生性修正前は同程度.

修正後は下がっている.

# 推定結果

- BLP法→操作変数法という2段階の内生性の修正により, field effect変数のパラメータ, t値は修正前より小さくなった. **しかし, 値が下がっても統計的には有意.**
- Hausman検定の結果, 修正前後での推定値の差は有意であり, 確かに内生性によってパラメータの推定値が過大になっていたことがわかった.

→ 仮説② 意思決定者をpeer groupに分け, 選択モデルにfield effect 変数を取り入れると, 内生性によりfield effect変数のパラメータにバイアスがかかり, 推定値が過大になる.

が証明されたといえる.

# まとめ

- 空間的・社会経済的に近い属性をもった意思決定者集団の中では、集団全体の選択傾向が個人の選択に影響を及ぼしている(=field effect).
- このような影響を説明変数として、内生性を考慮せずに推定すると、その影響の大きさが過大に評価されてしまう.
- BLP法により内生性を取り除くだけでなく、操作変数法を用いた2段階の修正を行うことで(修正された)field effect変数のパラメータの推定値を求め、集団の選択傾向が個人の選択に及ぼす影響の度合いまで示し、モデルの精度向上を図った.