本日の発表内容

- “A simple method to estimate discrete-type random coefficients logit models”
  - Forthcoming in *International Journal of industrial Organization*
  - 「ランダム係数ロジットモデル」の推定方法についての論文

1. ランダム係数ロジットモデルについて
2. 論文内容説明

1. ランダム係数ロジットモデルによる需要推定

製品間の競合度合いの推定

- 製品間の競合度合いを知ることは、多くの文脈で重要になる
  - 例）市場画定、政策効果、イノベーションの影響
- データから「製品jの需要量が、ライバル製品kの価格上昇によってどれだけ増えるか（$\frac{\partial 需要量_j}{\partial 価格_k}$）」を推定したい
  - たくさん増える　→　代替性が高く、競合している
  - あまり増えない　→　あまり競合していない


- 例）プリウスとハスラーがどれほど競合しているかを知りたい
  - 「∂プリウスの需要量/∂ハスラーの価格」を推定したい

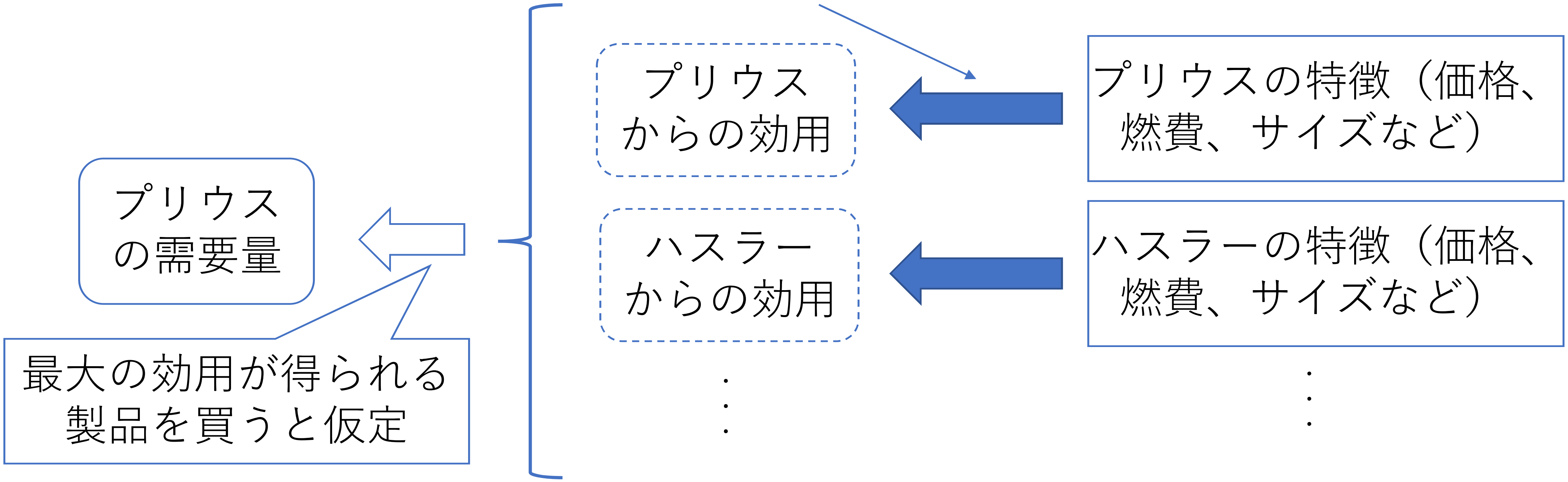
もっともシンプルな方法：<u>需要量と価格の関係（需要関数）</u>を直接推定

プリウスの需要量	←	プリウスの価格、 ハスラーの価格、 ・・・

- 例）「プリウスの需要量 $= a_1 \cdot$ プリウスの価格 $+ a_2 \cdot$ ハスラーの価格 $+ \cdots$」というようなモデルのa_1やa_2を推定する
- ライバル製品の価格の係数（例：a_2）が$\dfrac{\partial 需要量_j}{\partial 価格_k}$を表す

- しかし、製品数が多いと、推定するパラメター数が膨大になってしまう
  - 全製品ペアの両方向それぞれに対応するパラメターを推定することになる
  - 製品ペア数：$(製品数^2 - 製品数)/2$　（例：製品数100→ペア数4950）

ひとつの解決策：
製品特徴と効用の関係（効用関数）を推定



- 効用関数における製品特徴（価格も含む）の係数を推定する
    - 通常、製品ペア数＞製品特徴数
- 推定した効用関数→需要関数（需要量と価格の関係）を導出→$\dfrac{\partial 需要量_j}{\partial 価格_k}$

ランダム係数ロジットモデル (Random Coefficient Logit: RCL)

- 現在、需要推定で広く用いられているモデル
- （消費者$\boldsymbol{i}$・製品j・市場$\boldsymbol{m}$の）効用：

係数がランダムに分布

$$u_{ijm} = \boldsymbol{x}'_{jm}\boldsymbol{\beta}_i + \xi_{jm} + \epsilon_{ijm}$$

製品特徴（価格など）

観察できない製品特徴からの効用

消費者×製品に固有の好み～第1種極値分布

- 効用が最大となる製品を選ぶ

→ $\boldsymbol{i}$さんが製品jを選ぶ確率：$p_{ijm} = \frac{\exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}_i + \xi_{jm}\right)}{1+\sum_{k=1}^{Jm}\exp\left(\boldsymbol{x}'_{km}\boldsymbol{\beta}_i + \xi_{km}\right)}$

→ jの市場シェア：$s_{jm} = \int p_{ijm}\, dF(\boldsymbol{\beta}_i)$

$\boldsymbol{\beta}_i$の分布関数

→ 需要関数：需要量$_{jm}$ = 市場規模$_m \cdot s_{jm}$ → これを偏微分すれば$\frac{\partial 需要量_j}{\partial 価格_k}$が分かる

$\dfrac{\partial 需要量_j}{\partial 価格_{\boldsymbol{k}}}$において$\boldsymbol{\beta_i}$の分布が重要な役割

$$\frac{\partial 需要量_j}{\partial 価格_k} = \int \alpha_i p_{ijm} p_{ikm}\, dF(\boldsymbol{\beta_i})$$

$\boldsymbol{\beta_i}$のなかの価格の係数

- 全員同じ係数（$\boldsymbol{\beta_i} = \boldsymbol{\beta}$）のとき：$\dfrac{\partial 需要量_j}{\partial 価格_k} = \alpha s_{jt} s_{kt}$

  →　自身のシェアとライバル製品のシェアだけで決まってしまう（製品特徴が似ているかどうかとは無関係に決まってしまう）
- $\boldsymbol{\beta_i}$のばらつき「大」のとき：

  →　製品特徴が似ている製品間でp_{ijm}とp_{ikm}が同時に大きくなり、$\dfrac{\partial 需要量_j}{\partial 価格_k}$が大きくなる

- $\boldsymbol{\beta_i}$の分布に関するパラメターを推定することで、$\dfrac{\partial 需要量_j}{\partial 価格_k}$を柔軟に推定できる

参考

製品レベルのデータで推定できる

（市場m；製品j）

- 不可欠：
  - 市場シェア$_{jm}$（←数量$_{jm}$、市場規模$_{m}$）
  - 製品特徴$_{jm}$（価格も含む）
- 基本的には必要：
  - 価格の操作変数$_{jm}$（費用変数$_{jm}$など）
- できれば欲しい：
  - 消費者特徴の分布$_{m}$（所得や年齢の分布など）

参考

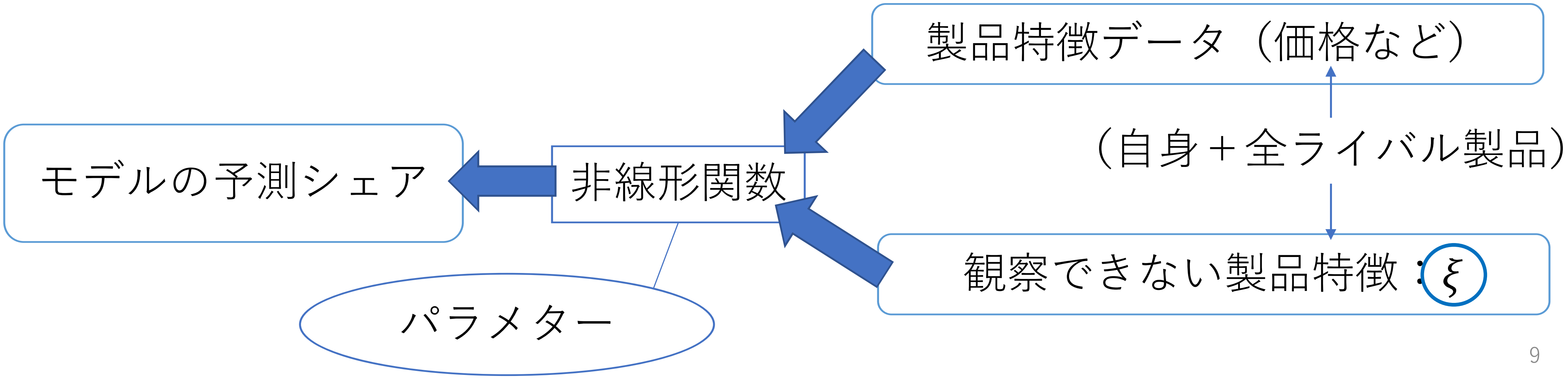
推定方法

観察できない製品特徴からの効用

- 「ξ_{jm}は外生変数と無関係（$E[\xi_{jm} * 外生変数_{jm} = 0]$）」という前提

→「$\xi_{jm} *$ 外生変数$_{jm}$」の標本平均をできるだけゼロに近づけるようにパラメター値を決める

  - ξ_{jm}の値はデータとして得られない
    →「モデルの予測シェア＝シェアデータ」となるような残差として得る（パラメター値に応じてξ_{jm}の値が決まる）



参考

ξ_{jm}の値を得るのはけっこう大変

- モデルの予測シェアは、全製品の製品特徴とξの非線形関数
  → あるパラメター値の下で「モデルの予測シェア＝シェアデータ」となるようなξ_{jm}を求めるのはけっこう大変
  → 一般には解析的に求められず、数値的に探索する（BLP: Berry, Levinsohn, & Pakes, 1995がその方法を確立）

（推定方法についての参考文献）

- Nevo (2000)：基本を分かりやすく解説（MATLABコードを公開）
- Conlon and Gortmaker (2020)：計算手順などに関するその後の研究蓄積を総合的に解説（Pythonコードを公開）

参考

参考文献

- Berry, S. , Levinsohn, J. , Pakes, A. , 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890 .

- Conlon, C. , Gortmaker, J. , 2020. Best practices for differentiated products demand estimation with pyblp. RAND J. Econ. 51 (4), 1108–1161 .

- Nevo, A. , 2000. A practitioner’s guide to estimation of random-coefficients logit models of demand. J. Econ. Manage. Strategy 9 (4), 513–548 .

2. 論文内容の紹介

“A Simple Method to Estimate Discrete-type Random Coefficients Logit Models”

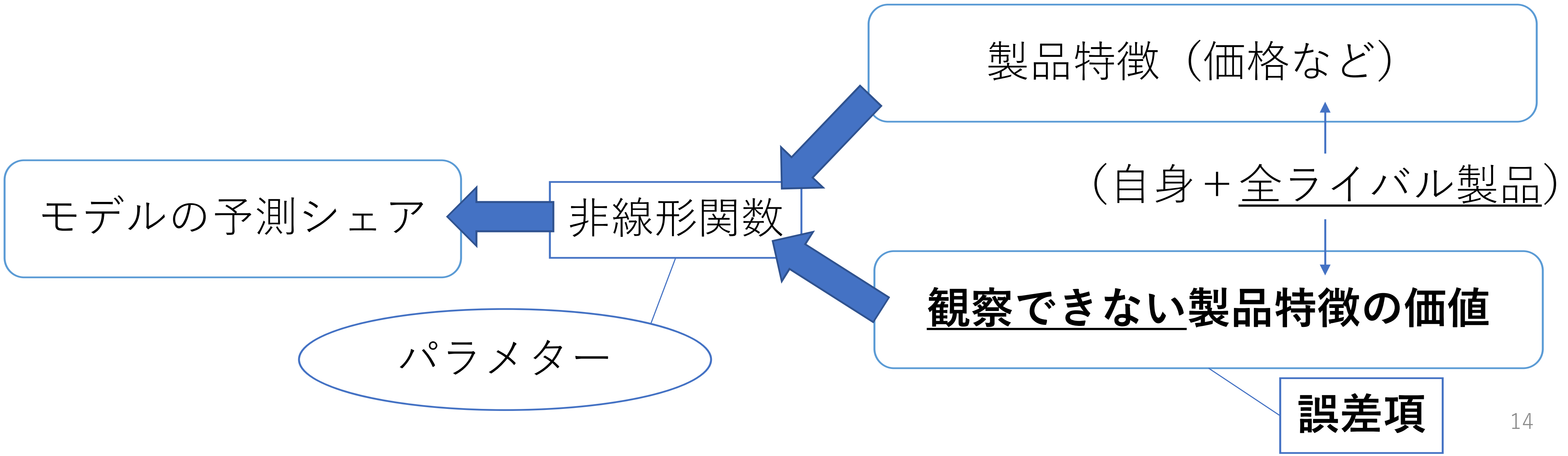
(forthcoming in *International Journal of Industrial Organization*)

集計データでのRCL推定

- ランダム係数ロジット(Random Coefficients Logit: RCL)需要モデル
  - 比較的少ないパラメター数
  - 差別化された製品間の代替性を柔軟に推定可能
- 製品レベルの集計データによる推定方法
  - Berry, Levinsohn, & Pakes (1995 ECTA: BLP）が確立
  - 以降、さまざまな改善（cf. Conlon & Gortmaker, 2020 RAND）
  - しかし、複雑なプログラムでの長時間の計算を要する
- 本論文：特定の状況なら使える、シンプルな推定方法を提案

BLPの推定方法の概略

- 市場シェアの式の「誤差項」に基づくGMM
- 「シェアデータ＝予測シェア」となる誤差項の値を求める
  - （一般には）誤差項を解析的に計算できない　→　数値計算で探す



多重ループ構造

モデルの予測シェア ＝ データ

となる誤差項の値を探索

内側ループ

GMM目的関数を最小化するパラメター値を探索

外側ループ

- それによる問題：
  - 複雑なプログラム
  - 長い計算時間
  - 内側での計算誤差によって、外側が収束しなくなったり、正しくない点に収束したりする恐れ（Dubé, Fox, & Su, 2012 ECTA）

本論文で提案する方法

- 市場シェアの式を変形し、誤差項の値を解析的に求める方法
  - ほかはBLPと同様
  - 「内側ループ」が無くなり、多重ループ構造が解消する

- その方法を使うための条件：
  - (a) 離散確率分布に従うランダム係数（「離散タイプRCL」）
    - 有限の消費者タイプ（例：ビジネス客/レジャー客）
  - (b)タイプ別販売量データがある
    - 「全製品合計」でいい（例：全航空会社の合計のビジネス客数）

離散タイプRCLの利用例

航空	Berry et al. (1996 NBER) Berry & Jia (2010 AEJMicro) Ciliberto & Williams (2014 RAND)
ケチャップ	Besanko et al. (2003 ManS)
ビデオゲーム	Nair (2007 QME)
処方薬	Iizuka (2007 RAND)
炭酸飲料	Eizenberg & Salvo (2015 AEJ:Applied)
ガソリン	MacKay & Remer (2019 SSRN)

タイプ別販売量データ

- RCL推定に必要な数量データ：製品別販売量
  - 通常、その消費者タイプ別内訳は分からない

- しかし、別データとして、総販売量における消費者タイプ別内訳が得られることはある

  （例）
  - アメリカのthe Consumer Expenditure Survey→新車購入者の所得階級別内訳（Petrin, 2002 JPE）
  - 日本の航空旅客動態調査→ビジネス客とレジャー客の割合
  - POSデータ→購入者の情報（Backus et al., 2021 NBER）

参考

タイプ別販売量データがあれば、モーメントを追加する推定も可能

- タイプ別販売量に関するモーメント条件を追加することで、推定の精度を高められる可能性がある
  - Petrin (2002 JPE) が、BLPの推定方法について、購入者特徴（所得階級など）に関するモーメント条件追加によって推定精度が高まることを示している
- 本論文の推定方法についても、モーメント条件追加によって推定精度が高まることをモンテカルロ実験で確認できた

Literatureでの位置づけ

数値積分
↓

“Simulation error”
(Berry et al, 2004 RES)

モデルの予測シェア ＝ データ

となる誤差項の値を探索

内側ループ

GMM目的関数を最小化するパラメター値を探索

外側ループ

- Simulation errorが、計算上の問題を引き起こす主因（Brunner et al., 2017）
- 離散タイプRCLではシェアの計算時に数値積分不要（e.g., Berry & Jia, 2010 AEJMicro）

提案手法：離散タイプRCLについて、タイプ別売上データを使い、誤差項を解析的に求める（内側ループ不要）

以下の流れ

- 離散タイプRCLモデル
- 提案する推定方法
- モンテカルロ実験
- 日本の航空旅客市場データを使った推定

離散タイプRCLモデル

モデル

- 市場m $(1, \dots, M)$における選択肢：
  - 製品 $j = 1, \dots, J_m$
  - 外部オプション（どれも買わない）$j = 0$

- 消費者iは条件付き間接効用u_{ijm}を最大にする選択肢をひとつ選ぶ

$$u_{ijm} = \boldsymbol{x}'_{jm}\boldsymbol{\beta}_i + \xi_{jm} + \epsilon_{ijm}$$

製品特徴（価格など）

観察できない製品特徴

消費者×製品に固有の好み
〜第1種極値分布

- 「離散タイプ」ランダム係数： $\boldsymbol{\beta}_i = \boldsymbol{\beta}^t \in \{\boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^T\}$
  - $\boldsymbol{\beta}^t$になる確率：γ^t（$\sum_{t=1}^{T} \gamma^t = 1$）

市場シェア

- タイプtの消費者が製品jを選ぶ確率：

$$p_m(j|\boldsymbol{\beta}^t) = \frac{\exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm}\right)}{1 + \sum_{k=1}^{J_m} \exp(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km})}$$

- 製品jの市場シェア：

$$s_{jm} = \sum_{t=1}^{T} \gamma^t p_m(j|\boldsymbol{\beta}^t) = \sum_{t=1}^{T} \gamma^t \frac{\exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm}\right)}{1 + \sum_{k=1}^{J_m} \exp(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km})}$$

全製品のξ

参考

離散タイプのランダム係数

- 連続的なランダム係数（例：正規分布に従う係数）に比べて、選好の多様性は限定的になる

- しかし、比較的少数のパラメーターで係数間の相関を表現できる
- また、市場シェアを解析的に計算できる（数値積分不要）

- さらに、離散タイプRCLでは本研究で提案する推定方法（「内側ループ」を省いた推定）が使える

提案する推定方法

推定方法の概略

- ξ_{jm}の求めかた以外はBLPとまったく同じ
  - ξ_{jm}と外生変数の積からなるモーメント条件に基づくGMM
  - 標本モーメントを計算するためにξ_{jm}の値を計算する
  - ξ_{jm}の値は、モデルの予測シェアとシェアデータとの乖離を説明する残差として得る

（再掲） $s_{jm} = \sum_{t=1}^{T} \gamma^t \frac{\exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm}\right)}{1+\sum_{k=1}^{Jm} \exp\left(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km}\right)}$

全製品のξ

ξ_{jm}の計算

$$s_{jm} = \sum_{t=1}^{T} \gamma^t \frac{\exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm}\right)}{1 + \sum_{k=1}^{J_m} \exp(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km})}$$

全製品のξ

- シェアは、全製品のξ_{jm}についての非線形関数
  → 一般には ξ_{jm}を解析的に計算できない

- BLP：数値計算（contraction mapping）
  - 多重ループ構造→長い計算時間・内側ループの計算誤差による問題

- 提案する方法：シェアの式にタイプ別販売量データを組み入れることで、 ξ_{jm}を解析的に求める
  → 内側ループが不要になり、多重ループ構造解消

シェアの式の変形

$$s_{jm} = \sum_{t=1}^{T} \gamma^t \frac{\exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm})}{1 + \sum_{k=1}^{J_m} \exp(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km})}$$

- タイプtがどれかは買う確率： $\tilde{s}_m^t = \sum_{j=1}^{J_m} p_m(j|\boldsymbol{\beta}^t) = \frac{\sum_{j=1}^{J_m} \exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm})}{1+\sum_{j=1}^{J_m} \exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm})}$
- タイプtが外部オプションを選ぶ確率：$1 - \tilde{s}_m^t = \frac{1}{1+\sum_{j=1}^{J_m} \exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm})}$

シェアの式の分母

- シェアの式に代入： $s_{jm} = \sum_{t=1}^{T} \gamma^t \exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t + \xi_{jm})(1 - \tilde{s}_m^t)$

$$\Leftrightarrow \ln s_{jm} - \ln\left\{\sum_{t=1}^{T} \gamma^t \exp(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t)(1 - \tilde{s}_m^t)\right\} = \xi_{jm}$$

他製品のξが消えている

→　$\tilde{s}_m^t$が分かれば、ξ_{jm}を解析的に計算できる　　$\left(\tilde{s}_m^t = \frac{\text{タイプ別販売量}_m^t}{\gamma^t \cdot \text{市場規模}_m}\right)$

ξ_{jm}を解析的に得られる理由

製品jのシェア ← 非線形関数 ← ξ_{jm} & $\xi_{km}(\forall k \neq j)$

→　一般には「予測シェア＝シェアデータ」となるξ_{jm}を解析的に求められない

- しかし、$\xi_{km}(\forall k \neq j)$は、「$\sum_{k=1}^{J_m} \exp(\boldsymbol{x}'_{km}\boldsymbol{\beta}^t + \xi_{km})$」（☆）の形でまとまって入っている
  - ☆はタイプtの消費者にとっての「全製品の価値の合計」を意味する
  - どれかは買う確率（$\tilde{s}_m^t$）によって丸ごと置き換えられる

製品jのシェア ← 非線形関数 ← ξ_{jm} & $\tilde{s}_m^t$

データ

ロジットの回帰式導出と同じ理屈

- 単純なロジット（$\boldsymbol{\beta}_i = \boldsymbol{\beta}$）の回帰式（Berry, 1994 RAND）：

$$\ln s_{jm} - \ln s_{0m} = \boldsymbol{x}'_{jm}\boldsymbol{\beta} + \xi_{jm}$$

- 外部オプションのシェアs_{0m}が「全製品の価値の合計」（の小ささ）の指標となっている
- それで☆を置き換えることによって、シェアの式から他製品のξ_{km}が消える

→　変形して線形の回帰式　（$T = 1$のときのξ_{jm}の式）

$$\boxed{\ln s_{jm} - \ln\left\{\sum_{t=1}^{T} \gamma^t \exp\left(\boldsymbol{x}'_{jm}\boldsymbol{\beta}^t\right)\left(1 - \tilde{s}_m^t\right)\right\} = \xi_{jm}}$$

参考

モーメント条件の追加

- ξ_{jm}を解析的に得る点のほかは、BLPと同じ方法で推定できる
  - ξ_{jm}と外生変数の積のモーメント条件に基づくGMM
- しかし、タイプ別販売量のデータがあるなら、それに関するモーメント条件を追加することで推定精度を上げられる可能性
  - 「タイプ別の購入確率（$\tilde{s}_m^t$）」の平均について、モデルの予測とデータをマッチさせる
- Petrin (2002 JPE)と同様のアイデア
  - アメリカ新車市場のRCL需要モデルを、集計データで推定
  - 通常のモーメント条件に加えて、購入者の特徴（所得階別ごとの新車購入確率など）に関するモーメント条件を追加

モンテカルロ実験

設定：Kalouptsidi (2012 JAE)と同様

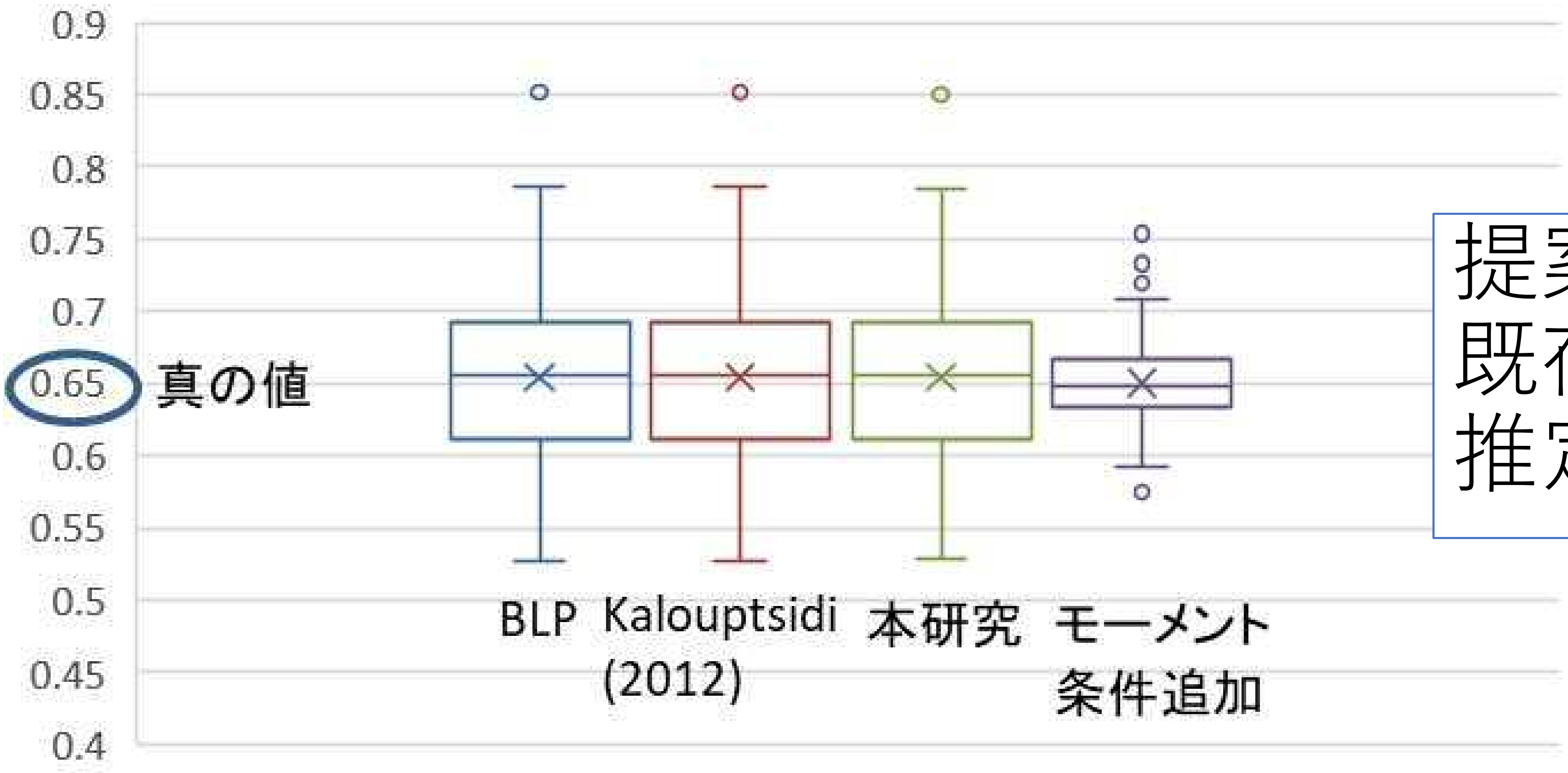
- 製品特徴：x_{jm}^{exo}とp_{jm}
- 2タイプの消費者
  - $\beta^1 = (2, -1)', \beta^2 = (3, -0.5)'$
  - 65%がタイプ1（$\gamma^1 = 0.65$）
- 50市場、100製品についてデータを発生：
  - $x_{jm}^{exo} \sim N(-3.5, 1)$
  - $\xi_{jm} \sim N(0, 0.4)$
  - $p_{jm} = 6 + 0.5x_{jm}^{exo} + 0.5\xi_{jm} + e_{jm}$，$e_{jm} \sim N(0,1)$
- 操作変数： $z_{jm}^d \sim U(0,1) + 0.25\left(1.1x_{jm}^{exo} + e_{jm}\right)$, $d = 1, \dots, 6$

推定

- データセットを100回発生
- 4種類の方法で推定
  - BLP
  - Kalouptsidi (2012 JAE): “K”
    - 内側ループの計算方法を工夫（多重ループ構造は残る）
  - 本論文：“P”、“P+”（モーメント条件追加）
- two-step efficient GMM

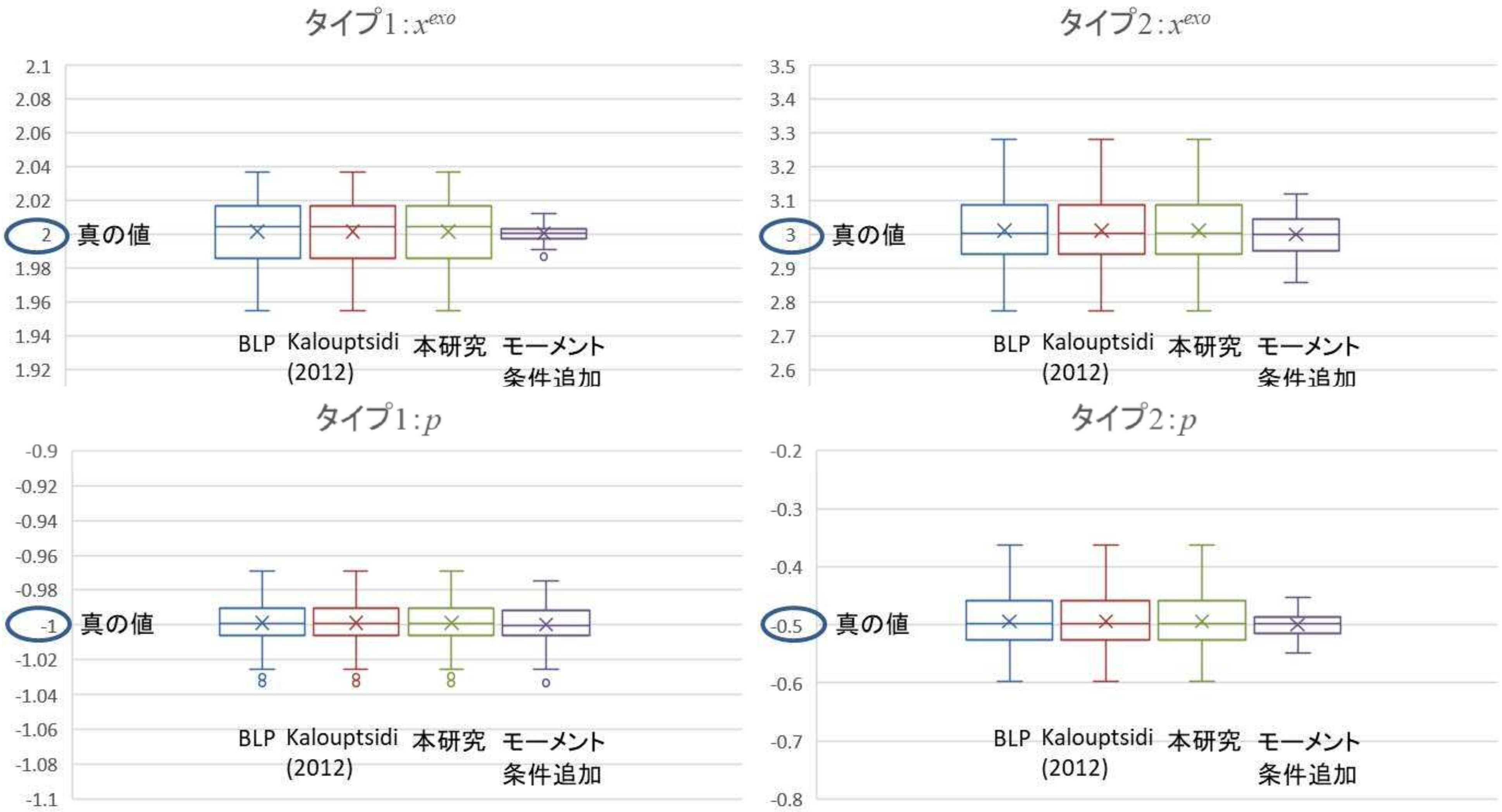
※個人のウェブサイトでコードを公開している

推定値の分布（γ^1）

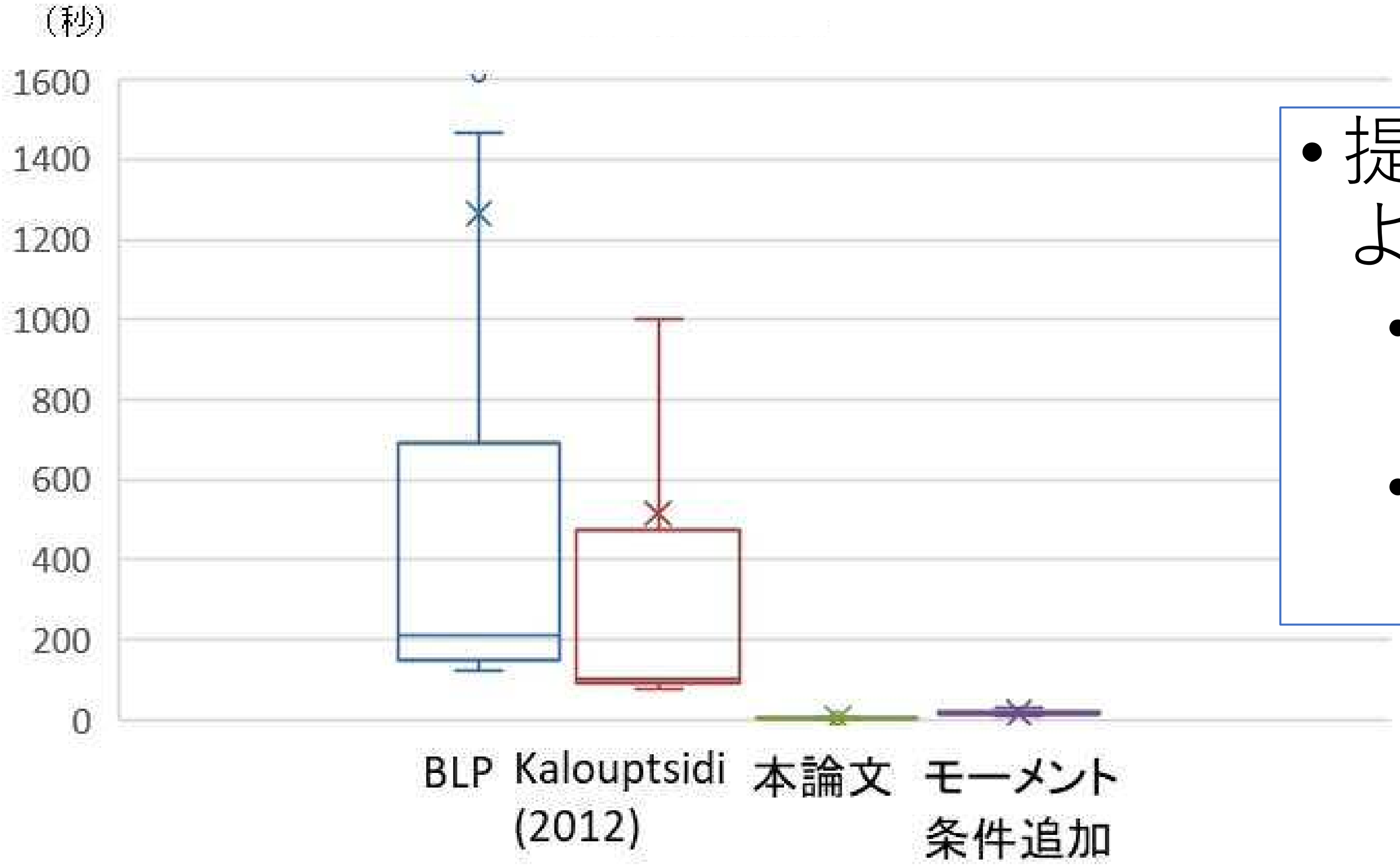


提案方法で、
既存方法と同様の
推定値が得られる

推定値の分布（その他）


タイプ1 : x^{exo}
2.1
2.08
2.06
2.04
2.02
2
真の値
1.98
1.96
1.94
1.92
BLP
Kalouptsidi
(2012)
本研究
モーメント
条件追加
タイプ2 : x^{exo}
3.5
3.4
3.3
3.2
3.1
3
真の値
2.9
2.8
2.7
2.6
BLP
Kalouptsidi
(2012)
本研究
モーメント
条件追加
タイプ1 : p
-0.9
-0.92
-0.94
-0.96
-0.98
-1
真の値
-1.02
-1.04
-1.06
-1.08
-1.1
BLP
Kalouptsidi
(2012)
本研究
モーメント
条件追加
タイプ2 : p
-0.2
-0.3
-0.4
-0.5
真の値
-0.6
-0.7
-0.8
BLP
Kalouptsidi
(2012)
本研究
モーメント
条件追加

計算時間

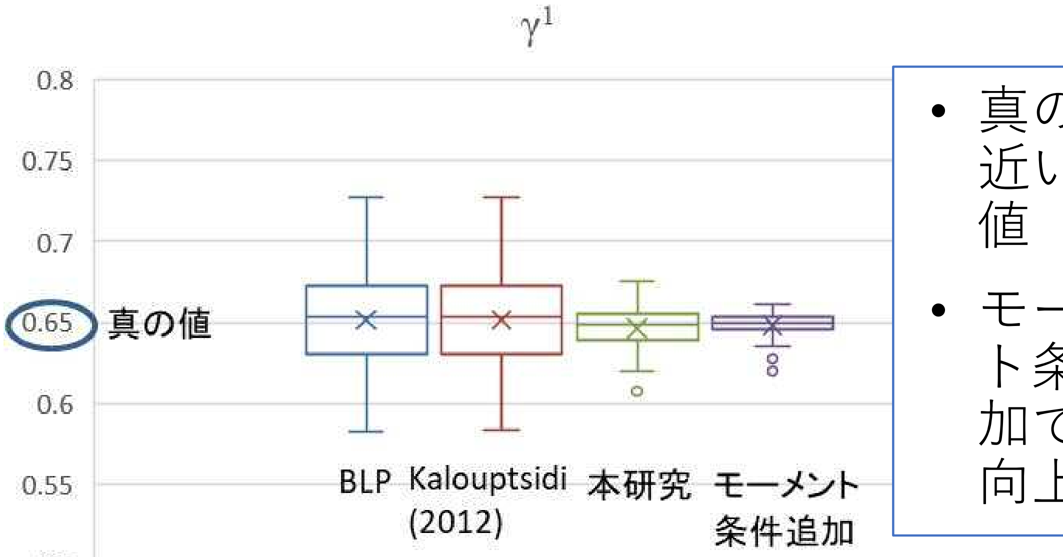


- 提案方法は、既存方法より早い
  - 中央値でBLPの1/47、Kの1/22
  - データセットによっては1/9000

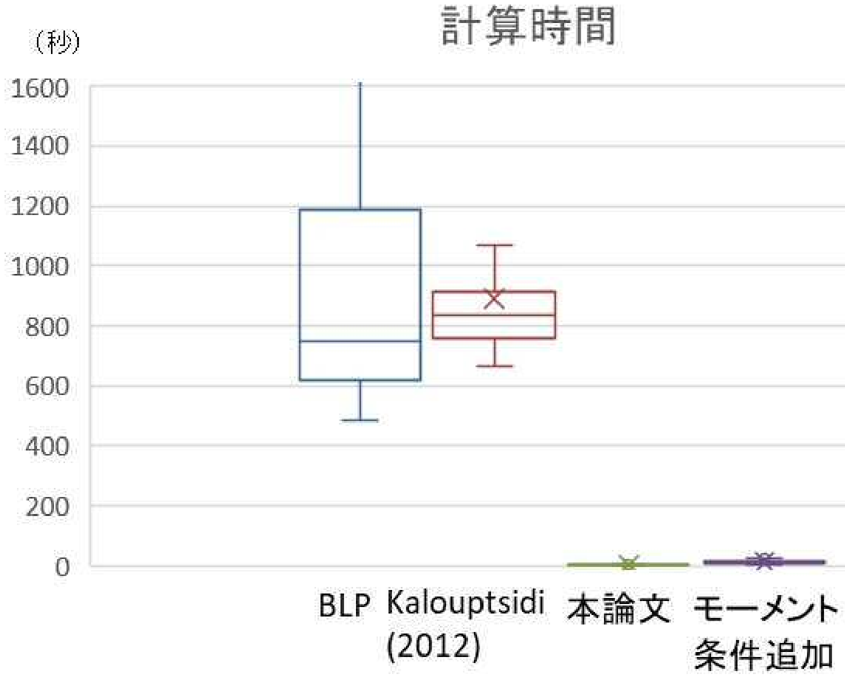
参考

頑健性確認：外部オプションのシェア（s_{0m}）「小」

- $x_{jm}^{exo} \sim N(-1.5, 1)$（※元は$N(-3.5, 1)$）
- s_{0m}「小」だと、BLPのcontraction mappingの収束スピードが遅くなる（Kalouptsidi, 2012 JAE）



- 真の値に近い推定値
- モーメント条件追加で精度向上



- BLPやKは長くなる
- 提案方法はほぼ同じ

→　提案方法の優位さが際立つ（中央値で、BLPの1/87、Kの1/143倍）

日本の航空旅客市場データを使った推定

タイプ別販売量（旅客数）データ

- 2タイプ（$T = 2$）：
  - ビジネス
  - レジャー

- 製品j：航空会社
- 市場m：路線

	(A) 総旅客数	(B) タイプ別割合		(C) タイプ別旅客数	
ソース	航空輸送統計年報	航空旅客動態調査		(A) × (B)	
		ビジネス	レジャー	ビジネス	レジャー
平均	47,598	0.49	0.51	26,622	20,976
標準偏差	103,862	0.20	0.20	67,998	41,827
最小値	781	0.05	0.08	78	120
第1四分位	5,338	0.33	0.35	1,776	2,934
中央値	12,840	0.51	0.49	6,188	6,219
第3四分位	48,713	0.65	0.67	23,219	21,601
最大値	821,611	0.92	0.95	528,833	317,313

注： 2003年11月のデータ。路線数は169。総旅客数は当該路線の全航空会社の旅客数の合計。

参考

推定結果

		(1) 離散タイプRCL 提案手法			(2) ロジット 2SLS		
ビジネスタイプの割合（γ）		0.75	(0.27)	***	-		
運賃	ビジネス	-0.20	(0.13)		-0.15	(0.04)	***
	レジャー	-0.05	(0.04)				
ln(フライト頻度)	ビジネス	1.88	(0.67)	***	0.94	(0.12)	***
	レジャー	0.60	(0.38)				
定数項		-4.57	(0.72)	***	-4.28	(0.64)	***
距離		0.03	(1.74)		0.62	(1.58)	
距離^2		2.30	(1.40)		2.33	(1.35)	*
距離^3		-0.72	(0.39)	*	-0.63	(0.35)	*
相関係数		0.36			0.27		
自己価格弾力性		-2.10			-3.36		

注：サンプルサイズは240（観測単位：航空会社×路線）。相関係数は、モデルの予測シェアとデータについてのもの。自己価格弾力性は航空会社×路線 ペアごとに計算したものの平均。カッコ内の数字は標準誤差。***と*はそれぞれ1%と10%有意を表す。

まとめ

- 離散タイプRCLについて、「内側ループ」（誤差項の数値計算）なしで推定する方法を提案
  - BLPなどと同様の推定値
  - 計算時間はほぼゼロになる

- 追加的に必要なものはタイプ別販売量データだけ
  - 製品別ではなく全製品合計で良い

参考

今後の課題

- 航空旅客動態調査を使った分析
  - ビジネス客vsレジャー客
  - 航空アライアンスの影響：国内旅客vs国際旅客

- 推定方法の改良
  - タイプ別販売量データ（T×M個）から、直接、各市場における各タイプの割合（$\gamma^t \rightarrow \gamma_m^t$：T×M個）を求められるのではないか？

ありがとうございました！