

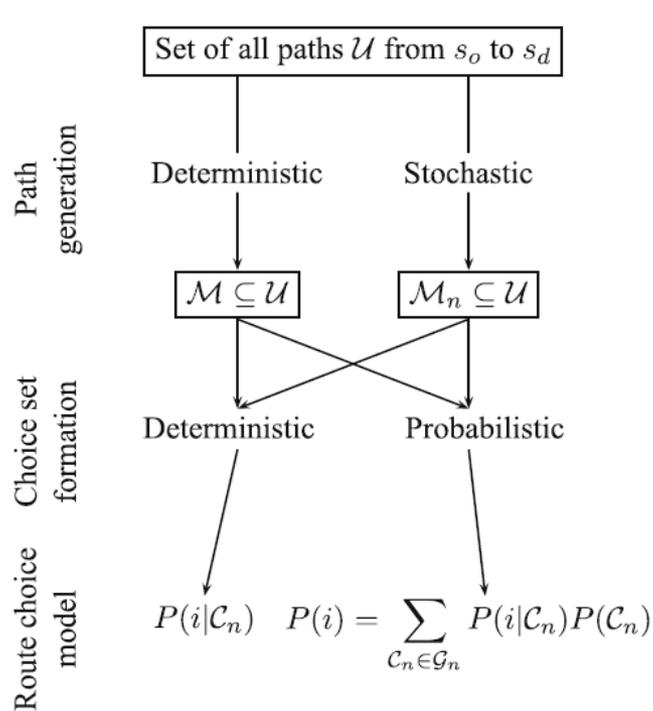
Sampling of alternatives for route choice modeling

E.Frejinger.M.Bierlaire,M.Ben-Akiva
transportation reserch partB, Vol.43,
pp. 984-994, 2009.

M1 齊藤

1. 経路選択モデル構築の手順

- 経路選択モデルは実際の推定までにいくつかのステップを踏む必要がある。



- ① 経路集合 \mathcal{M} を生成する
- ② 選択枝集合 \mathcal{C}_n を生成する
- ③ モデルを推定する

経路選択モデル推定のためには経路生成アルゴリズムが必要になる。その方法としては、決定論的な方法と確率論的な方法がある。

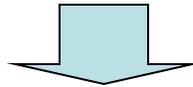
Fig. 1. Choice set generation overview.

提案手法の背景・目的

- 本論文では、選択肢集合生成の新しい手法を提案.
- 生成された選択肢集合の信頼性や現実性について評価することは難しい. なぜなら, モデラーは実際の選択肢集合を知りえないから.
- 現存の選択肢集合生成方法
 - 現実の選択肢集合生成が目標
- 今回の方法
 - 精度のよいパラメータ値を得ることが目的. 有効性の評価は, 真値との比較によって行う.
- 今回の方法では, 観測された経路はもともとサンプルに含めている

2.代替選択肢のサンプリング

- M_{nI} は代替選択肢の部分集合を用いて推定される.



- 個人 n が選択肢 i を選択する確率はモデラーによって定義される選択肢集合に依存する. 選択肢集合 \mathcal{C}_n が選ばれたとき選択肢 i が選択される条件付き確率は以下のように表される.

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in} + \ln q(\mathcal{C}_n|i)}}{\sum_{j \in \mathcal{C}_n} e^{\mu V_{jn} + \ln q(\mathcal{C}_n|j)}} \quad \dots (1)$$

ここで, μ : スケールパラメータ

V_{in} : 確定効用

$\ln q(\mathcal{C}_n|j)$: サンプリングバイアス補正項

2.代替選択肢のサンプリング

- もし全ての選択肢が同確率で選択されるならば, 式(1)のサンプリングバイアスはキャンセルアウトされ, 全ての選択肢から選択するのと同じ方法で推定できる



全選択肢集合が非常に大きいときは適切ではない

- 選択肢間の選ばれやすさに差がある(重み付けサンプリング)と考えると, サンプリングバイアス補正が重要になる.

3. サンプリング補正

- 経路生成におけるサンプリング手順を次のように定める
 - 全集合 \mathcal{U} から R_n 個の経路を取り出す: $\tilde{\mathcal{C}}_n$
 - $\tilde{\mathcal{C}}_n$ に選択経路を加える: $|\tilde{\mathcal{C}}_n| = R_n + 1$
- 全ての経路 $j \in \mathcal{U}$ は確率 $q(j)$ で選択される
- R_n 回でそれぞれ経路 j が $(\tilde{k}_{1n}, \tilde{k}_{2n}, \dots, \tilde{k}_{jn})$ 回取り出される確率は以下のようなになる.

$$P(\tilde{k}_{1n}, \tilde{k}_{2n}, \dots, \tilde{k}_{jn}) = \frac{R_n!}{\prod_{j \in \mathcal{U}} \tilde{k}_{jn}!} \prod_{j \in \mathcal{U}} q(j)^{\tilde{k}_{jn}} \quad \dots (2)$$

- 経路 j が表れる回数は, $k_{in} = \tilde{k}_{jn} + \delta_{jc}$ なので, 経路 i が選択されたときにその選択肢集合が \mathcal{C}_n である確率は以下のようなになる

$$q(\mathcal{C}_n|i) = q(\tilde{\mathcal{C}}_n|i) = \frac{R_n!}{(k_{in} - 1)! \prod_{\substack{j \in \mathcal{C}_n \\ j \neq i}} k_{jn}!} q(i)^{k_{in}-1} \prod_{\substack{j \in \mathcal{C}_n \\ j \neq i}} q(j)^{k_{jn}} \quad \dots (3)$$

3. サンプリング補正

- 式(3)を変形すると以下ようになる.

$$q(\mathcal{C}_n|i) = \frac{R_n!}{\frac{1}{k_{in}} \prod_{j \in \mathcal{C}_n} k_{jn}!} \frac{1}{q(i)} \prod_{j \in \mathcal{C}_n} q(j)^{k_{jn}} = K_{\mathcal{C}_n} \frac{k_{in}}{q(i)} \quad \dots (4)$$

ここで, $K_{\mathcal{C}_n} = \frac{R_n!}{\prod_{j \in \mathcal{C}_n} k_{jn}!} \prod_{j \in \mathcal{C}_n} q(j)^{k_{jn}}$

- 式(1)に代入すると $K_{\mathcal{C}_n}$ はキャンセルアウトされるので, 個人nが選択肢iを選択肢集合 \mathcal{C}_n の中から選ぶ選択確率は以下ようになる.

$$P(i|\mathcal{C}_n) = \frac{e^{\mu V_{in} + \ln \left(\frac{k_{in}}{q(i)} \right)}}{\sum_{j \in \mathcal{C}_n} e^{\mu V_{jn} + \ln \left(\frac{k_{jn}}{q(j)} \right)}} \quad \dots (5)$$

4. 確率的経路生成方法

- 各選択肢がサンプリングされる確率 $q(j)$ を定める.
- ODペア (s_o, s_d) が与えられたとき, それぞれのリンク $\ell = (v, w)$ の重みは一般化費用にもとづいた最短距離に対するリンク ℓ の距離で定められる.

$$x_\ell = \frac{SP(s_o, s_d)}{SP(s_o, v) + C(\ell) + SP(w, s_d)} \quad \dots (6)$$

ここで, $C(\ell)$: リンク ℓ の一般化費用

$SP(v_1, v_2)$: ノード v_1 と v_2 における最短経路の一般化費用

4. 確率的経路生成方法

- それぞれのリンク ℓ の重みを x_ℓ から定義するため、以下のような関数を用いる。

$$\omega(\ell|b_1, b_2) = 1 - (1 - x_\ell^{b_1})^{b_2} \quad \dots (7)$$

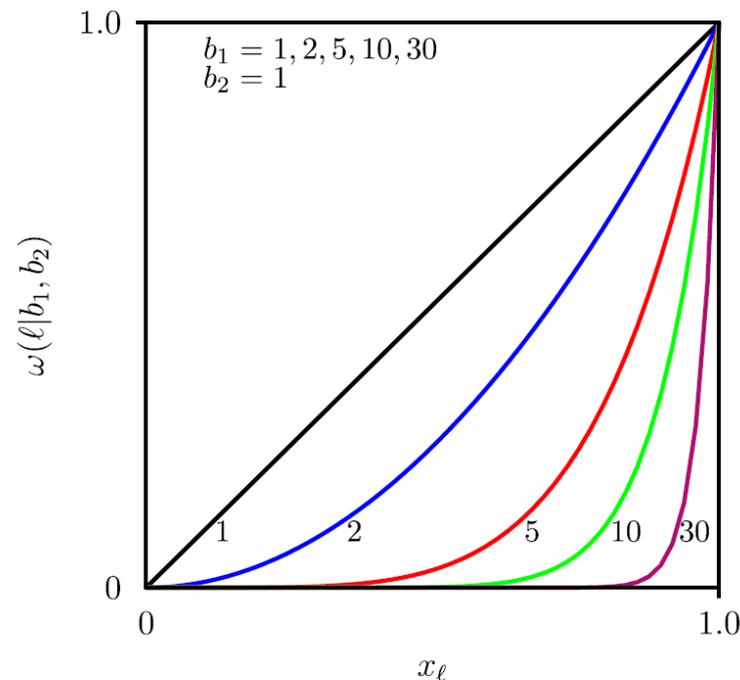


Fig. 2. Kumaraswamy distribution: cumulative distribution function.

4. 確率的経路生成方法

- それぞれのリンクの重みが決まったので、ランダムウォークを用いてそれぞれの経路の生成確率を求める.

初期化: $v = s_0, \quad \Gamma = \emptyset$

繰り返し: $v \neq s_d$ のとき

\mathcal{E}_v に含まれるそれぞれのリンク $\ell = (v, w)$ の重みを式(6)にもとづいて以下のように定める.

$$x_\ell = \frac{SP(v, s_d)}{C(\ell) + SP(w, s_d)} \quad \dots (8)$$

このときリンク $\ell = (v, w)$ の選択確率は以下のように表わされる.

$$q(\ell | \mathcal{E}_v, b_1, b_2) = \frac{\omega(\ell | b_1, b_2)}{\sum_{m \in \mathcal{E}_v} \omega(m | b_1, b_2)} \quad \dots (9)$$

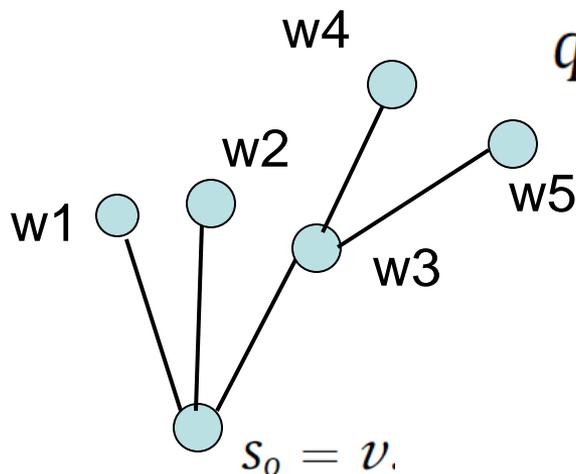
4. 確率的経路生成方法

取り出し: 式(9)で定めた確率にもとづいて, \mathcal{E}_v に含まれるリンク (v, w^*) をランダムに選択する.

経路更新: $\Gamma = \Gamma \cup (v, w^*)$

次のノード: $v = w^*$

経路 j が生成される確率 $q(j)$ は, リンク Γ_j を順番に選択する確率に等しく, 以下の式で表される.



$$q(j) = \prod_{\ell \in \Gamma_j} q(\ell | \mathcal{E}_v, b_1, b_2) \quad \cdots (10)$$

5.PS特性

- 経路同士の重複による相関を考慮するために, MNLの確定効用にPS(path size)特性を導入する.
- 従来のPSは以下の式で表される.

$$PS_{in}^{\mathcal{C}} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{M_{an}}$$

Γ_i : 経路iに含まれるリンク集合

L_a : リンクaの長さ

L_i : 経路iの長さ

$M_{an} = \sum_{j \in \mathcal{C}_n} \delta_{aj}$, $\delta_{aj} : 1$ if path j contains link a
0 otherwise

5.EPS特性

- 今回は, PS特性を補正したEPSを用いる.

$$\text{EPS}_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{M_{an}^{\text{EPS}}}$$

ここで,
$$M_{an}^{\text{EPS}} = \sum_{j \in \mathcal{C}_n} \delta_{aj} \Phi_{jn}$$

$$\Phi_{jn} = \begin{cases} 1 & \text{if } \delta_{jc} = 1 \text{ or } q(j)R_n \geq 1 \\ \frac{1}{q(j)R_n} & \text{otherwise.} \end{cases}$$

この定義は3で用いられたサンプリングプロトコルにもとづいている.

6. 数値計算

- 今回は、以下の影響を評価する目的で、合成データを用いて推定を行う。
 - サンプリング補正
 - PS特性の定義
 - 重みづけランダムウォークアルゴリズムパラメータ
 - 仮定したモデルの定義

6.1. 合成データ

- 以下に示すネットワークを用いる. このネットワークは, 実在するネットワークの一部であり, 38ノードと64リンクで成り立つ.
- ループがないので全集合を列挙することが可能で, $|\mathcal{U}| = 170$ となる.

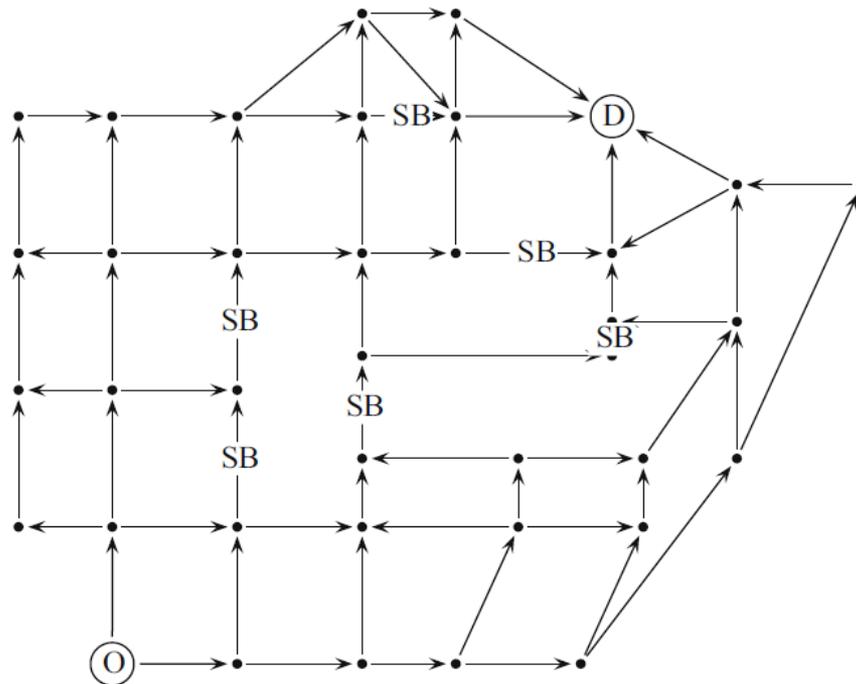


Fig. 3. Example network.

6.1. 合成データ

- それぞれ個人 n の選択肢 i に対する効用関数を以下のよう
に定める.

$$U_{in} = \beta_{PS} \ln PS_i^u + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i + \varepsilon_{in}$$

ここで, $\beta_{PS} = 1$, $\beta_{SB} = -0.1$

$$PS_i^u = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in u} \delta_{aj}}$$

6.2.モデルの特定

- サンプル補正とPS特性の異なる定式化を評価するため、以下のように5つのモデルを考慮

$$M_{PS(\mathcal{C})}^{\text{NoCorr}} V_{in} = \mu(\beta_{PS} \ln PS_{in}^{\mathcal{C}} + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i)$$

$$M_{PS(\mathcal{C})}^{\text{Corr}} V_{in} = \mu(\beta_{PS} \ln PS_{in}^{\mathcal{C}} + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i) + \ln \left(\frac{k_{in}}{q(i)} \right)$$

$$M_{PS(\mathcal{U})}^{\text{NoCorr}} V_i = \mu(\beta_{PS} \ln PS_i^{\mathcal{U}} + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i)$$

$$M_{PS(\mathcal{U})}^{\text{Corr}} V_{in} = \mu(\beta_{PS} \ln PS_i^{\mathcal{U}} + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i) + \ln \left(\frac{k_{in}}{q(i)} \right)$$

$$M_{EPS}^{\text{Corr}} V_{in} = \mu(\beta_{PS} \ln EPS_{in} + \beta_L \text{Length}_i + \beta_{SB} \text{NbSB}_i) + \ln \left(\frac{k_{in}}{q(i)} \right)$$

β_L は固定して、(データセットに応じて、-1か-0.3に固定)
 $\mu, \beta_{PS}, \beta_{SB}$ を推定

6.3. 推定結果

$\beta_L = -1$ dataset using 40 draws and $b_1 = 1$

Table 1

Detailed estimation results of PSL models.

	$M_{PS(\mathcal{U})}^{\text{NoCorr}}$	$M_{PS(\mathcal{U})}^{\text{Corr}}$	$M_{PS(\mathcal{C})}^{\text{NoCorr}}$	$M_{PS(\mathcal{C})}^{\text{Corr}}$	$M_{\text{EPS}}^{\text{Corr}}$
$\hat{\beta}_{\text{PS}}$	-0.108	0.969	0.285	0.397	1.09
Rob. std	0.045	0.0541	0.0785	0.074	0.052
Rob. t -test 1	-24.62	-0.57	-9.11	-8.15	1.73
$\hat{\beta}_{\text{SB}}$	-0.547	-0.0849	-0.52	-0.00941	-0.109
Rob. std	0.0322	0.0262	0.0331	0.0261	0.0281
Rob. t -test -0.1	-13.88	0.58	-12.69	3.47	-0.32
$\hat{\mu}$	1.04	0.983	1.05	0.945	1.05
Rob. std	0.0314	0.028	0.0316	0.0264	0.0314
Rob. t -test 1	1.27	-0.61	1.58	-2.08	1.59
Final L-L	-7284.711	-6966.668	-7281.035	-7160.154	-6704.515
Adj. rho bar sq.	0.291	0.322	0.292	0.303	0.348

Null log likelihood: -10283.7, 3000 observations.

$\beta_L = -1$. Algorithm parameters: 40 draws, $b_1 = 1$, $b_2 = 1$, $C(\ell) = L_\ell$. Average size of sampled choice sets: 30.92.

BIOGEME (Bierlaire, 2007, 2003) has been used for all model estimations.

6.3. 推定結果

$\beta_L = -1$ dataset using 40 draws and $b_1 = 1$

Table 1
Detailed estimation results of PSL models.

	$M_{PS(\psi)}^{NoCorr}$	$M_{PS(\psi)}^{Corr}$	$M_{PS(\psi)}^{NoCorr}$	$M_{PS(\psi)}^{Corr}$	M_{EPS}^{Corr}
$\hat{\beta}_{PS}$	-0.108	0.969	0.285	0.397	1.09
Rob. std	0.045	0.0541	0.0785	0.074	0.052
Rob. t-test 1	-24.62	-0.57	-9.11	-8.15	1.73
$\hat{\beta}_{SB}$	-0.547	-0.0849	-0.52	-0.00941	-0.109
Rob. std	0.0322	0.0262	0.0331	0.0261	0.0281
Rob. t-test -0.1	-13.88	0.58	-12.69	3.47	-0.32
$\hat{\mu}$	1.04	0.983	1.05	0.945	1.05
Rob. std	0.0314	0.028	0.0316	0.0264	0.0314
Rob. t-test 1	1.27	-0.61	1.58	-2.08	1.59
Final L-L	-7284.711	-6966.668	-7281.035	-7160.154	-6704.515
Adj. rho bar sq.	0.291	0.322	0.292	0.303	0.348

Null log likelihood: -10285.7, 3000 observations.

$\beta_L = -1$. Algorithm parameters: 40 draws, $b_1 = 1$, $b_2 = 1$, $C(\ell) = L_\ell$. Average size of sampled choice sets: 30.92.

BIOGEME (Bierlaire, 2007, 2003) has been used for all model estimations.

- モデルの適合度は、明らかにサンプリング補正をしたほうが良くなっていることがわかる。

6.3. 推定結果

$\beta_L = -1$ dataset using 40 draws and $b_1 = 1$

Table 1
Detailed estimation results of PSL models.

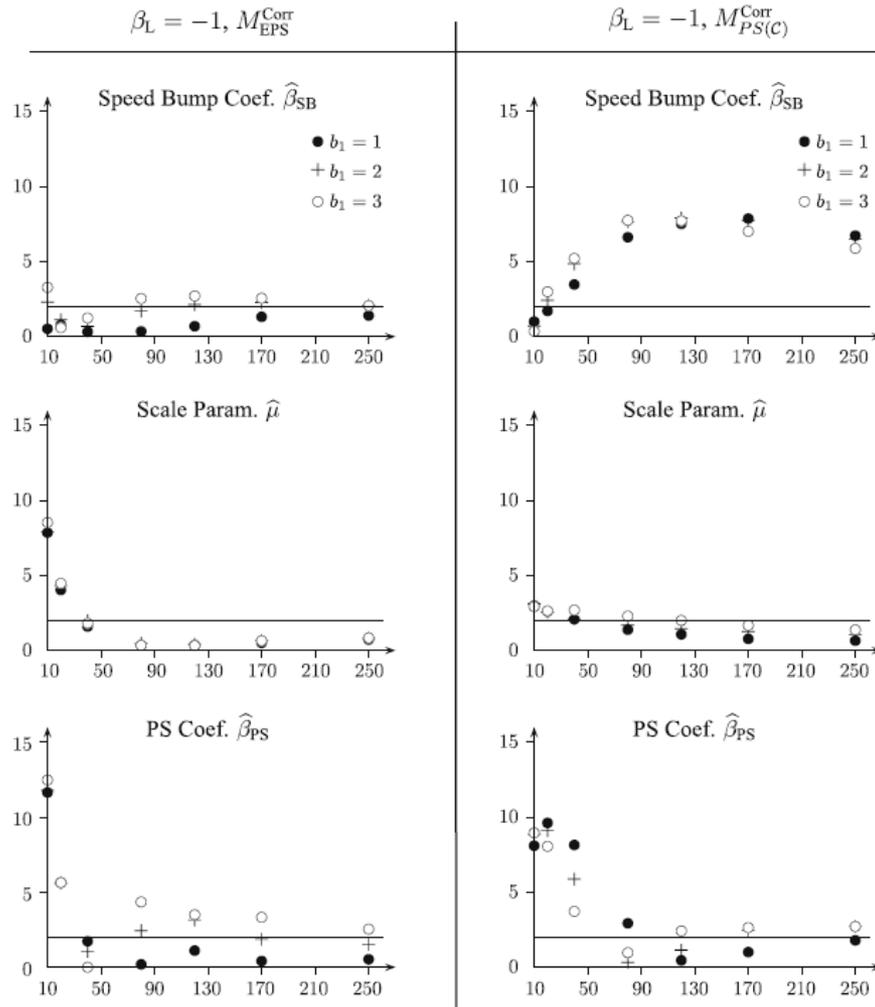
	$M_{PS(\psi)}^{\text{NoCorr}}$	$M_{PS(\psi)}^{\text{Corr}}$	$M_{PS(\phi)}^{\text{NoCorr}}$	$M_{PS(\phi)}^{\text{Corr}}$	$M_{\text{EPS}}^{\text{Corr}}$
$\hat{\beta}_{\text{PS}}$	-0.108	0.969	0.285	0.397	1.09
Rob. std	0.045	0.0541	0.0785	0.074	0.052
Rob. <i>t</i> -test 1	-24.62	-0.57	-9.11	-8.15	1.73
$\hat{\beta}_{\text{SB}}$	-0.547	-0.0849	-0.52	-0.00941	-0.109
Rob. std	0.0322	0.0262	0.0331	0.0261	0.0281
Rob. <i>t</i> -test -0.1	-13.88	0.58	-12.69	3.47	-0.32
$\hat{\mu}$	1.04	0.983	1.05	0.945	1.05
Rob. std	0.0314	0.028	0.0316	0.0264	0.0314
Rob. <i>t</i> -test 1	1.27	-0.61	1.58	-2.08	1.59
Final L-L	-7284.711	-6966.668	-7281.035	-7160.154	-6704.515
Adj. rho bar sq.	0.291	0.322	0.292	0.303	0.348

Null log likelihood: -10283.7, 3000 observations.

$\beta_L = -1$. Algorithm parameters: 40 draws, $b_1 = 1$, $b_2 = 1$, $C(\ell) = L_\ell$. Average size of sampled choice sets: 30.92.
BIOGEME (Bierlaire, 2007, 2003) has been used for all model estimations.

- PSとEPSでは、EPSのほうが適合度が高くなっている

推定結果



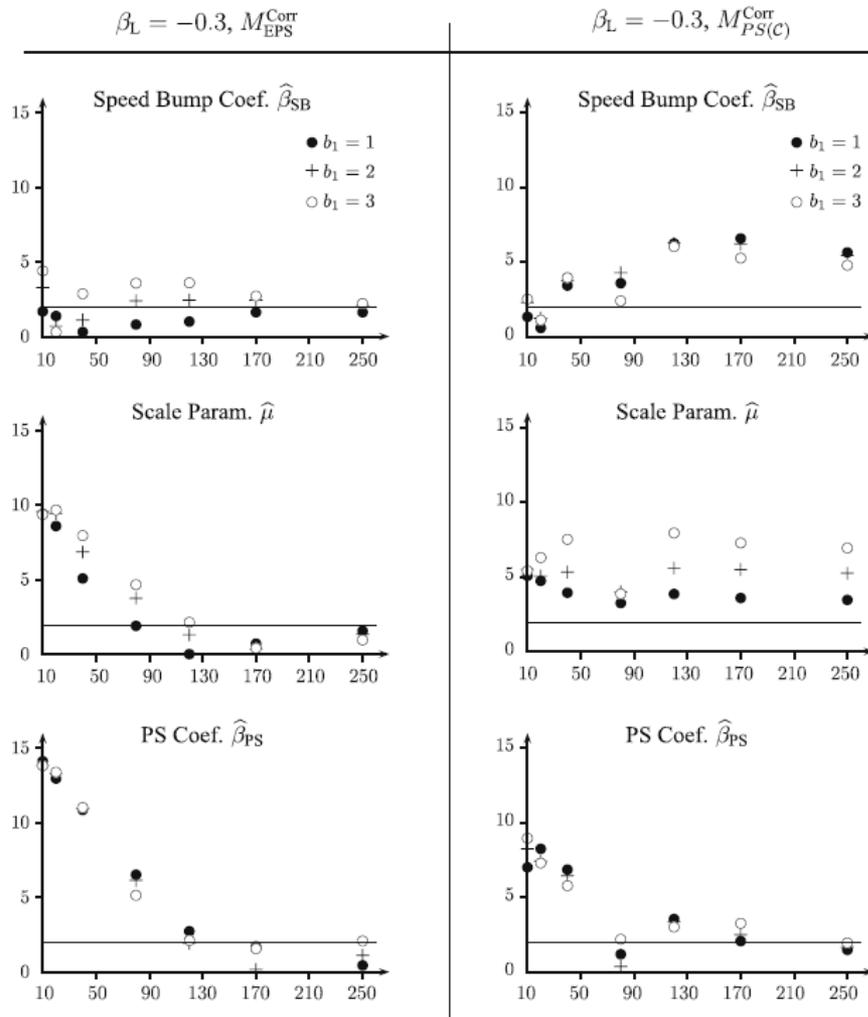
モデル M_{EPS}^{Corr} と $M_{PS(\epsilon)}^{Corr}$ の
T値の絶対値の比較

モデル M_{EPS}^{Corr} は、サンプリングの回数が増えるとともに結果が向上している。

モデル $M_{PS(\epsilon)}^{Corr}$ は、サンプリングの回数が増えても少なくとも1つはパラメータ値が改善されない結果となっている

Fig. 6. $\beta_L = -1$ dataset: t -test values for M_{EPS}^{Corr} and $M_{PS(\epsilon)}^{Corr}$, x-axis: number of draws, y-axis: absolute value of t -test w.r.t. true value.

推定結果



$\beta_L = -0.3$ としたときの
T値の比較.

$\beta_L = -1$ のときとほぼ同様の
結果が得られている.

Fig. 7. $\beta_L = -0.3$ dataset: t-test values for M_{EPS}^{Corr} and $M_{PS(C)}^{Corr}$, x-axis: number of draws, y-axis: absolute value of t-test w.r.t. true value.

結論と今後の課題

- サンプルング補正を行うと、行わない場合よりも良い結果が得られる.
- 生成されたチョイスセットだけでなく、すべての選択肢について相関があると考えてPSを設定したほうがよい結果が得られる.
- EPSによって補正を行うと、従前のモデルよりも結果が向上する.
- 今回は、合成データに対して行ったので、実際のデータについても検証していく必要がある.