A comparative analysis of implicit and explicit methods to model choice set generation

Michel Bierlaire
Ricardo Hurtubia
Gunnar Flötteröd

9th Swiss Transport Research Conference
September 9, 2009

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Motivation

- Standard choice models: choice set is characterized by deterministic rules
  - Explicit (un)availability of the alternative
  - Explicit restrictions
- Some choice sets are not deterministic
  - Fuzzy rules
  - Depending on unobserved attributes
  - Complex interaction between decision maker and the environment
- Methods to model choice set generation process are usually complex: solutions?

Outline

1. Motivation
2. Deterministic choice set
3. Probabilistic choice set generation
4. Constrained Multinomial Logit
5. Comparison of approaches
    - A simple example
    - Synthetic data
6. Conclusions

Deterministic choice set

- Assumption: known choice-set

$$A_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is available to individual } n \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} P_n(i|\mathcal{C}_n) &= \Pr\left(U_{in} \geq U_{jn}, \forall j \in \mathcal{C}_n\right) \\ &= \Pr\left(U_{in} + \ln A_{in} \geq U_{jn} + \ln A_{jn}, \forall j \in \mathcal{C}\right) \end{aligned}$$

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} = \frac{e^{V_{in} + \ln A_{in}}}{\sum_{j \in C} e^{V_{jn} + \ln A_{jn}}}$$

Probabilistic Choice Set (PCS)

- Manski (1977):

$$P_n(i) = \sum_{C_m \subseteq C} P_n(i|C_m) \cdot P_n(C_m)$$

Sub-set Universal choice-set

  - Choice-set is a latent construct (not observed)
  - Alternative selection and choice-set generation are separate processes
  - Computational complexity (combinatorial number of possible choice sets = $2^j - 1$)

Constrained Multinomial Logit (CMNL)

- Martínez et al. (2009):

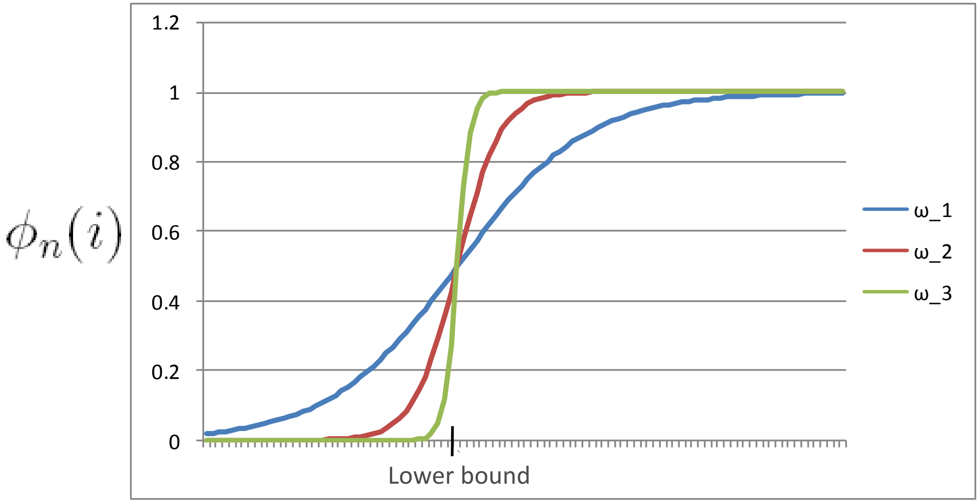
  “eliminates” unfeasible alternatives with a probabilistic rule.

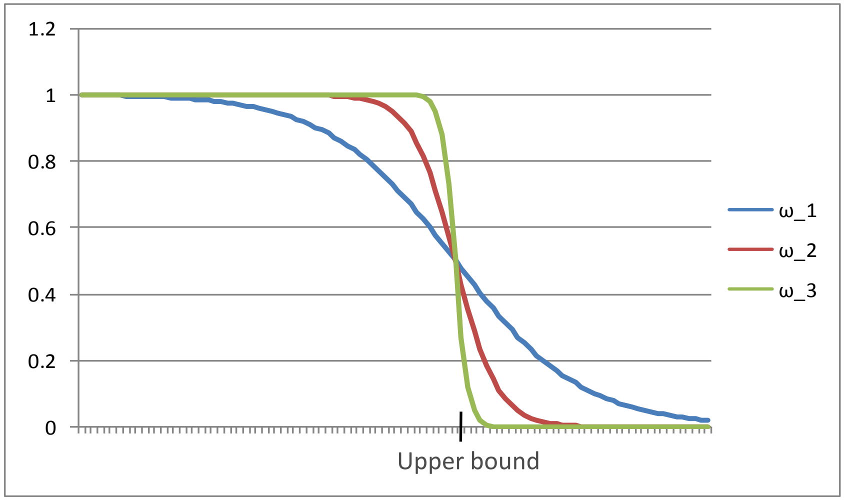
$$P_n(i) = \frac{e^{V_{in} + \ln \phi_n(i)}}{\sum_{j \in C} e^{V_{jn} + \ln \phi_n(j)}}$$

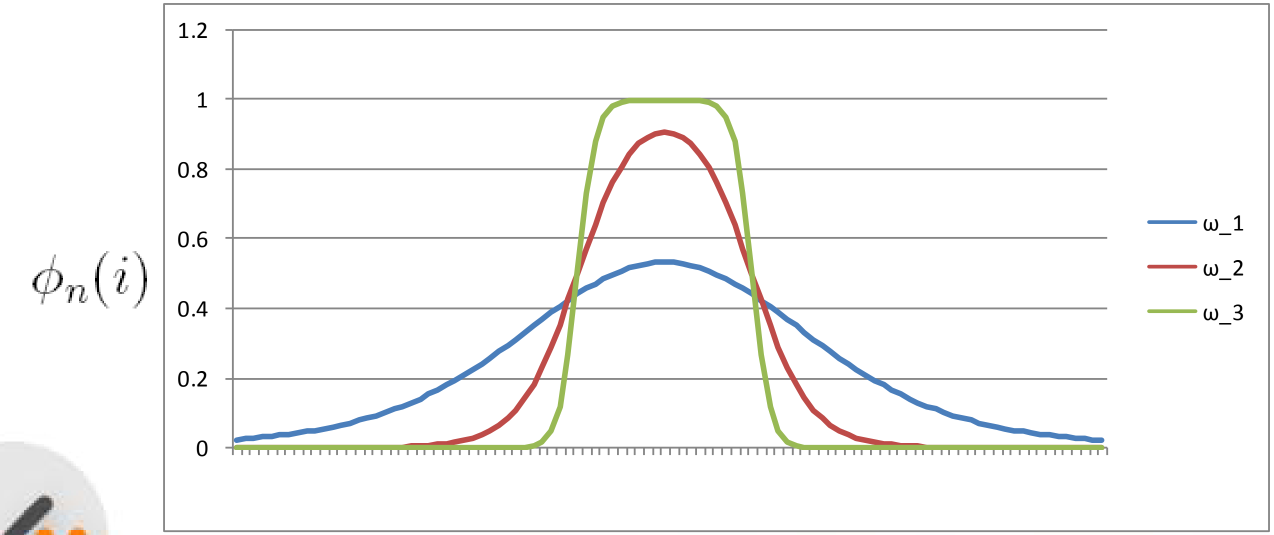
Probability of alternative *i* being available or considered by user *n* :

$$\phi_n(i) = \frac{1}{1 + \exp(\omega(\underset{\substack{\uparrow \\ \text{Attribute}}}{y_i} - \underset{\substack{\uparrow \\ \text{Constraint}}}{a_n}))} = \begin{cases} 1 \text{ if } \; y_i << a_n \\ 0 \text{ if } \; y_i >> a_n \end{cases}$$

Constrained Multinomial Logit (CMNL)







ω_1 < ω_2 < ω_3

Constrained Multinomial Logit (CMNL)

- CMNL:
  - Does not require enumeration of choice sets
  - Simulates the construction of the individual’s choice set
  - Heuristic based on assumptions over the utility’s functional form:

$$V_{in}^{'} = \underbrace{V_{in}}_{\text{Compensatory part}} + \underbrace{\phi_n(i)}_{\text{Non-Compensatory part (penalty)}}$$

  - ➔ CMNL is an approximation to the choice-set generation procedure

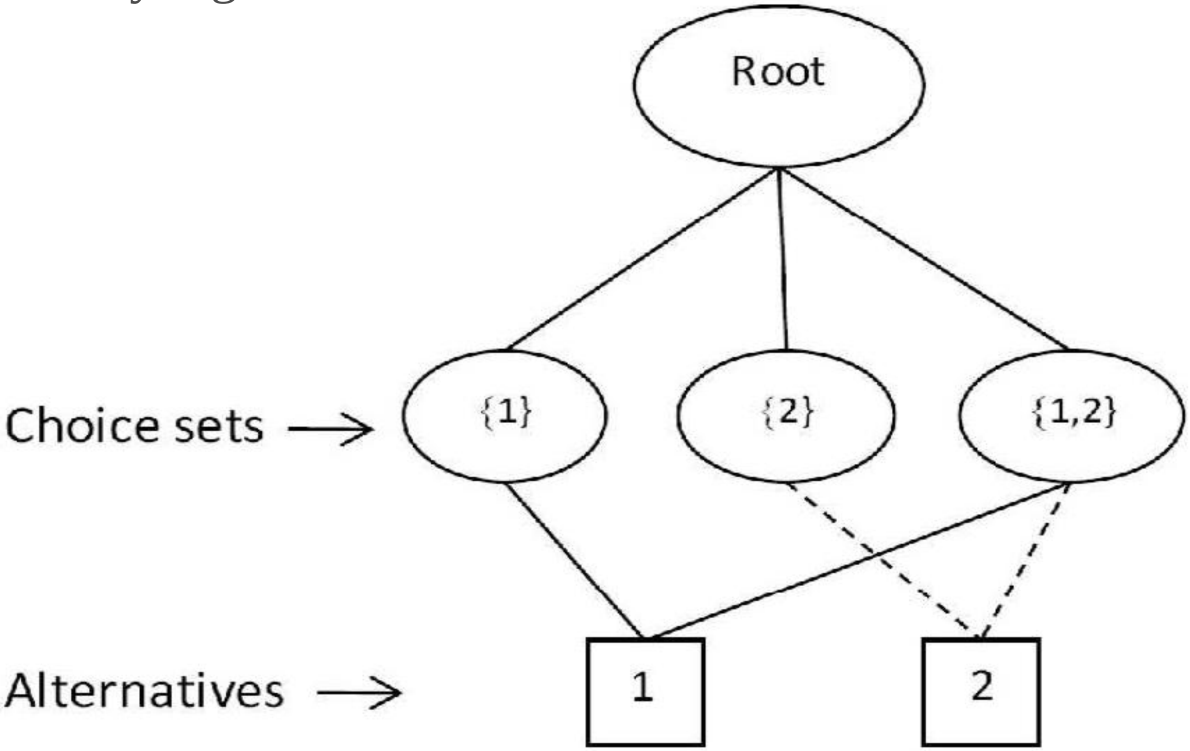
How good is this approximation?

Comparison of approaches

$$P_n(i) = \sum_{C_m \in C} \left(\frac{e^{V_{in}}}{\sum_{j \in C_m} e^{V_{jn}}} \cdot P_n(C_m) \right) \overset{?}{=} \frac{e^{V_{in} + \ln \phi_n(i)}}{\sum_{j \in C} e^{V_{jn} + \ln \phi_n(j)}}$$

A simple example

Binary logit



$$P(\{1\}) = 1 - \phi(2)$$

$$P(\{2\}) = 0$$

$$P(\{1,2\}) = \phi(2)$$

A simple example

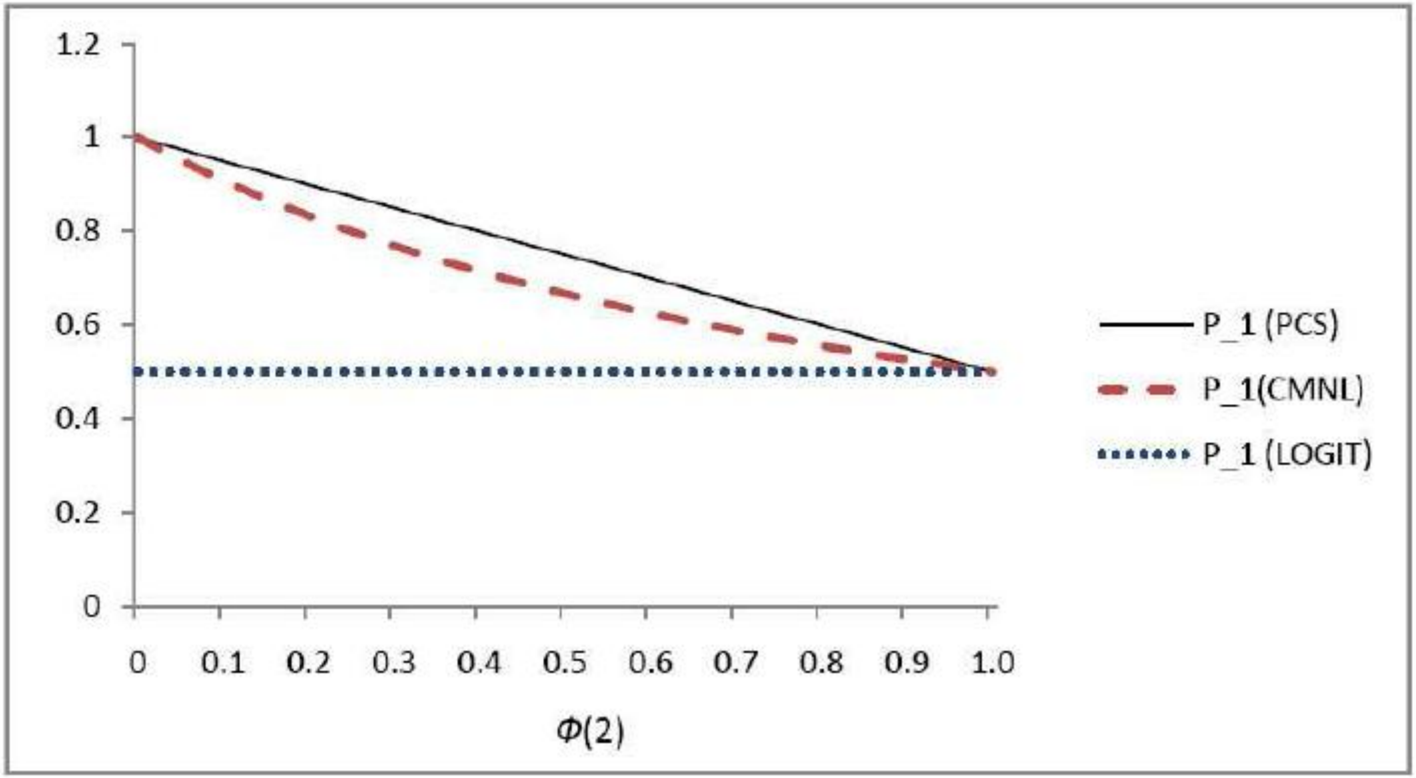
- Probability of choosing alternative 1?

- CMNL: $P(1) = \dfrac{e^{V_1}}{e^{V_1} + e^{V_1 + \ln \phi(2)}}.$

- PCS: $P(1) = \left(1 - \phi(2)\right) \cdot 1 + \phi(2) \cdot \dfrac{e^{V_1}}{e^{V_1} + e^{V_2}}$

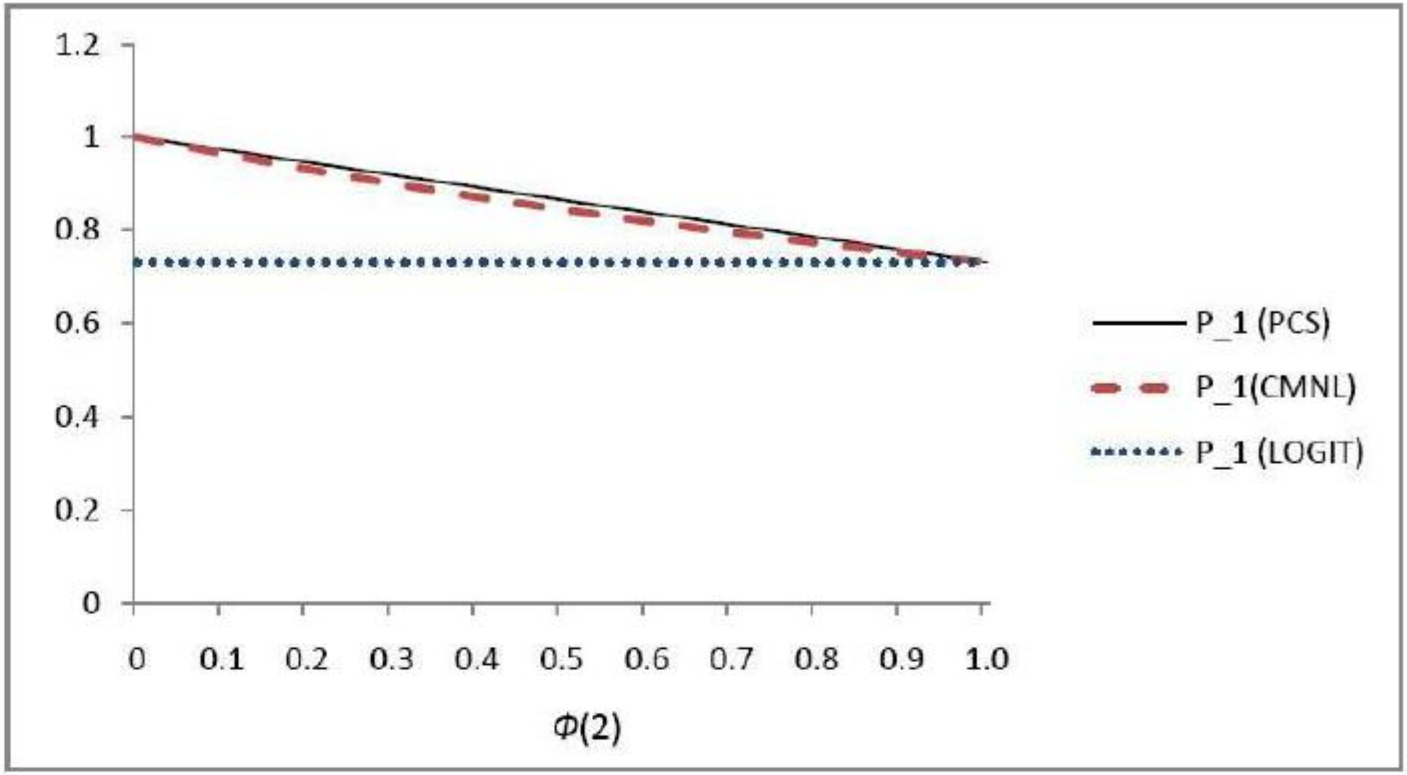
A simple example

- $P(1)$ Equal utility ($V_1 = V_2$)



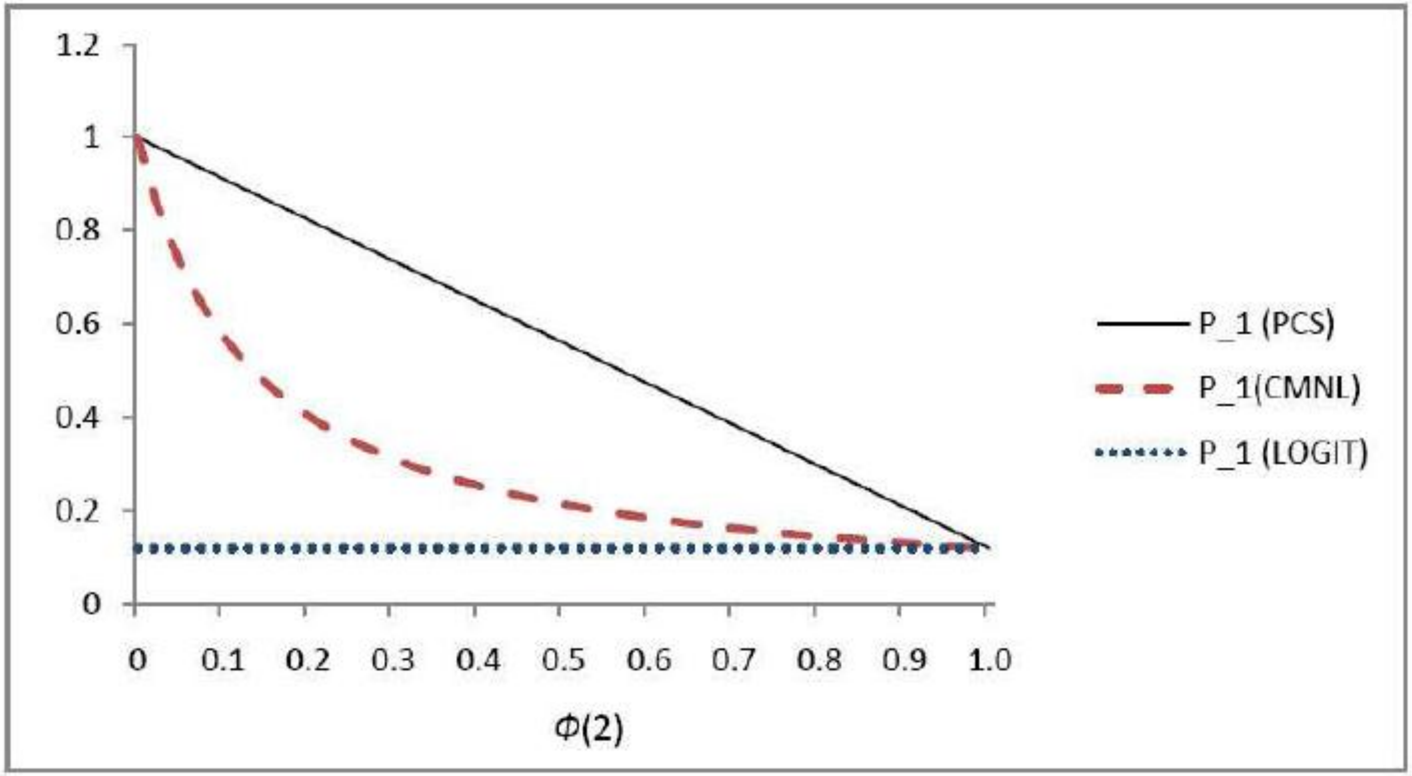
A simple example

- $P(1)$ Alternative 1 is dominant $(V_1 > V_2)$



A simple example

- $P(1)$ Alternative 2 is dominant ($V_1 < V_2$)



Synthetic data

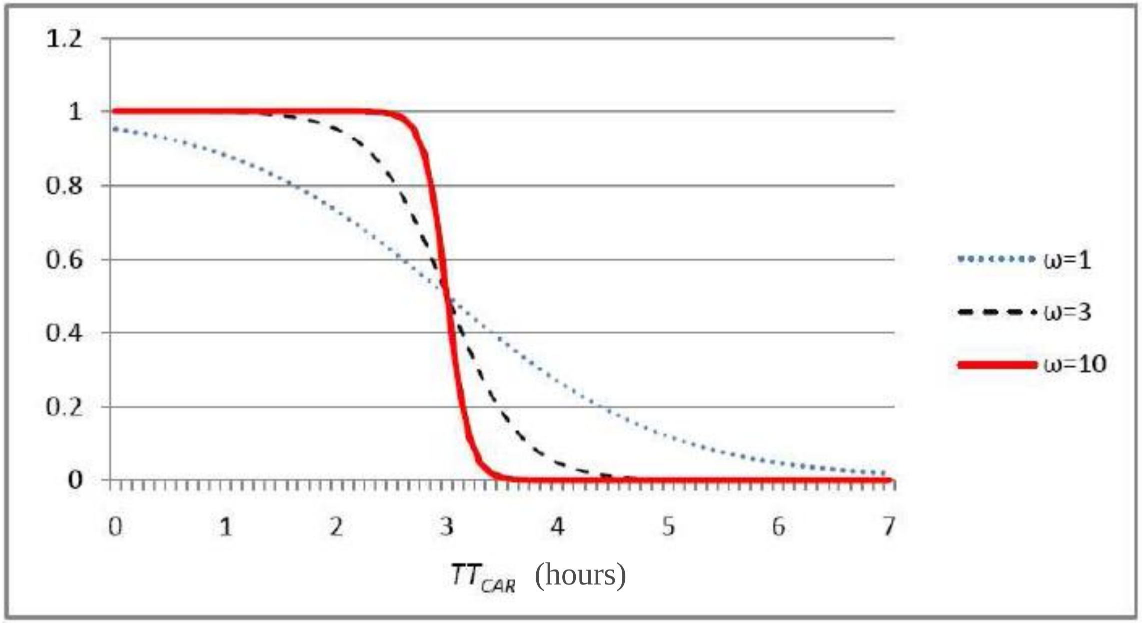
- “Swissmetro”
- Simulated choices according to Manski’s approach
- Alternatives
  - Car (not always available)
  - Train (always available)
  - Swissmetro (always available)

$$
\begin{aligned}
V_n(\mathrm{CAR}) &= \mathrm{ASC}_{\mathrm{CAR}}+\beta_{\mathrm{cost}} \cdot \mathrm{COST}_{\mathrm{CAR}} + \beta_{\mathrm{tt}} \cdot \mathrm{TT}_{\mathrm{CAR}} \\
V_n(\mathrm{TRAIN}) &= \beta_{\mathrm{cost}} \cdot \mathrm{COST}_{\mathrm{TRAIN}} + \beta_{\mathrm{tt}} \cdot \mathrm{TT}_{\mathrm{TRAIN}}+\beta_{\mathrm{he}} \cdot \mathrm{HE}_{\mathrm{TRAIN}} \\
V_n(\mathrm{SM}) &= \mathrm{ASC}_{\mathrm{SM}}+\beta_{\mathrm{cost}} \cdot \mathrm{COST}_{\mathrm{SM}} + \beta_{\mathrm{tt}} \cdot \mathrm{TT}_{\mathrm{SM}} + \beta_{\mathrm{he}} \cdot \mathrm{HE}_{\mathrm{SM}}
\end{aligned}
$$

Synthetic data

- Car availability:

$$\phi(\mathrm{CAR}) = \frac{1}{1+\exp(\omega(TT_{\mathrm{CAR}}/60 - a))}$$



Synthetic data

- 2 possible choice sets:
  - Car, Train, Swissmetro

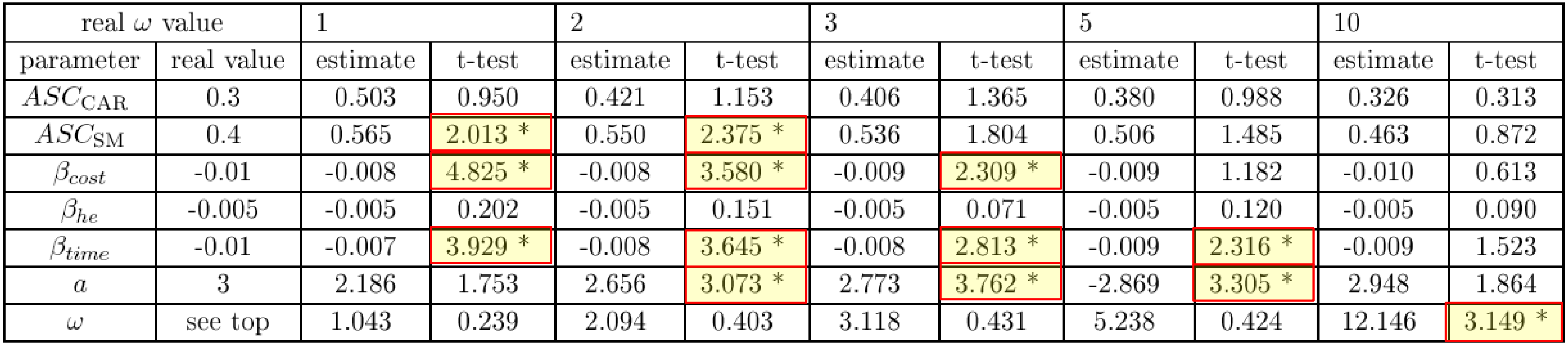
$$P(CAR, TRAIN, SM) = \phi(CAR)$$

  - Train, Swissmetro

$$P(TRAIN, SM) = 1 - \phi(CAR)$$

Synthetic data

- Results for CMNL (mean of each parameter over 100 estimations)

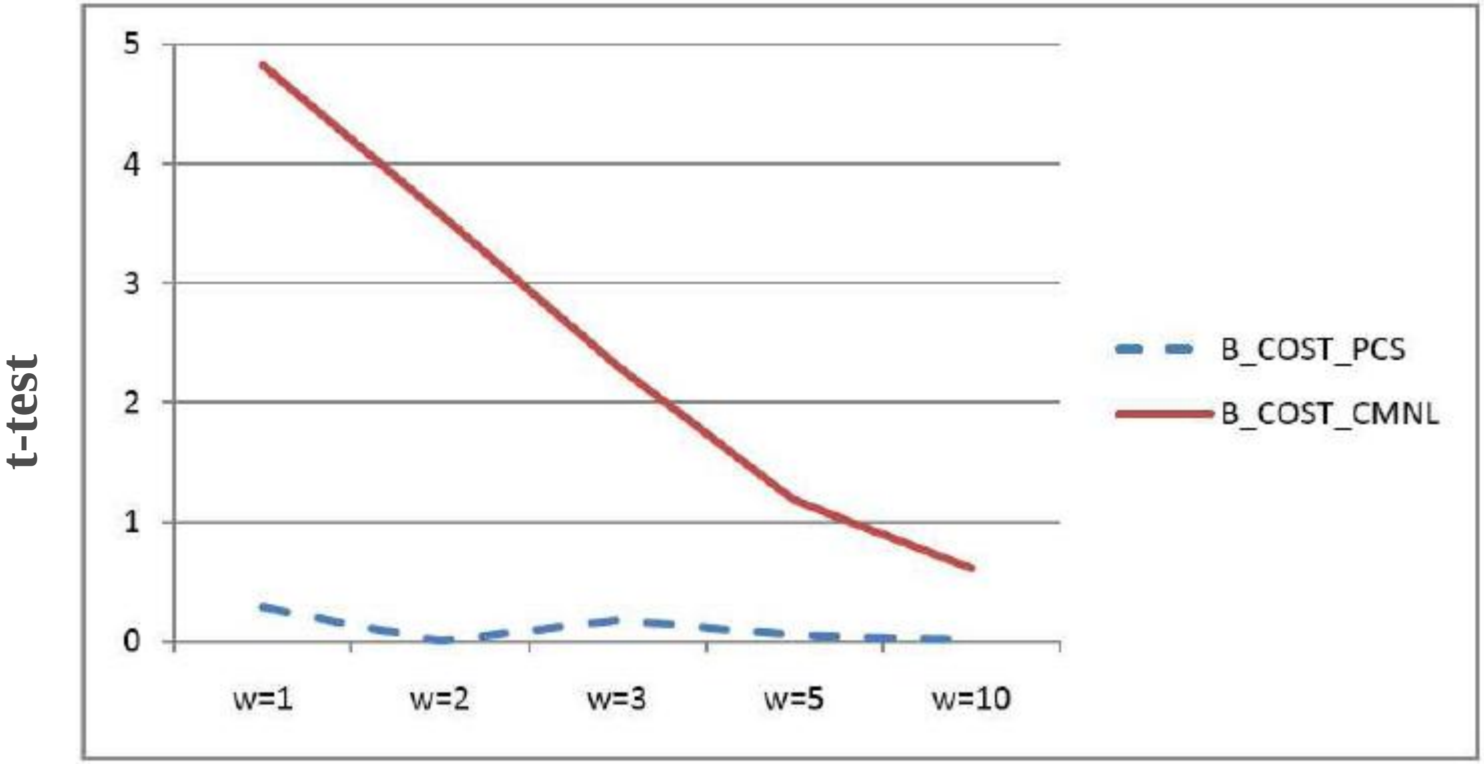


real ω value		1		2		3		5		10	
parameter	real value	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test
ASC_{CAR}	0.3	0.503	0.950	0.421	1.153	0.406	1.365	0.380	0.988	0.326	0.313
ASC_{SM}	0.4	0.565	2.013 *	0.550	2.375 *	0.536	1.804	0.506	1.485	0.463	0.872
β_{cost}	-0.01	-0.008	4.825 *	-0.008	3.580 *	-0.009	2.309 *	-0.009	1.182	-0.010	0.613
β_{he}	-0.005	-0.005	0.202	-0.005	0.151	-0.005	0.071	-0.005	0.120	-0.005	0.090
β_{time}	-0.01	-0.007	3.929 *	-0.008	3.645 *	-0.008	2.813 *	-0.009	2.316 *	-0.009	1.523
a	3	2.186	1.753	2.656	3.073 *	2.773	3.762 *	-2.869	3.305 *	2.948	1.864
ω	see top	1.043	0.239	2.094	0.403	3.118	0.431	5.238	0.424	12.146	3.149 *

- The quality of the estimates improves when the dispersion decreases

Synthetic data

- t-test over dispersion



Conclusions

- CMNL approach generates biased results when compared with Manski's approach
- The CMNL is a valid approximation when the constraints tend to be deterministic
- Still, is convenient for big choice-set problems (considered as a model on its own)

- Further work
  - Identify more specifically when is recommendable to use the CMNL
  - Justification from the behavioral approach?
  - Possible correction to the model?

Thank you