Bootstrapping approach for sampling of alternatives in MEV models

Ricardo Hurtubia
Gunnar Flötteröd
Michel Bierlaire

August 19-20, 2010
Lausanne

TRANSP-OR
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

1. Motivation
2. Sampling in logit models
3. Sampling methodology for MEV models
4. Bootstrapping
5. Simulation experiment
6. Results

Motivation

- Sampling is required in presence of a large choice set
- Consistent estimation is possible for MNL (McFadden, 1978)
- Sampling in non-logit models: can't be directly extended from MNL case
- Asymptotically unbiased estimator for nested logit (Guevara and Ben-Akiva, 2010)
- Bootstrapping: can be used to calculate the bias and correct for it.

Sampling in logit models

- Choice probability

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}$$

- If a sample D_n is considered:

$$P_n(i|D_n) = \frac{e^{\mu V_{in} + \ln \pi(D_n|i)}}{\sum_{j \in D_n} e^{\mu V_{jn} + \ln \pi(D_n|j)}}$$

Sampling methodology for MEV models*

- Choice probability: $$P_n(i) = \frac{e^{V_{in}+\ln G_i}}{\sum_{j\in C_n} e^{V_{jn}+\ln G_j}}$$

- with: $$G_i = \frac{\partial G(e^{V_{1n}}, e^{V_{2n}}, ..., e^{V_{J_n n}})}{\partial e^{V_{in}}}$$

$$\Rightarrow \quad P_n(i|D_n) = \frac{e^{V_{in}+\ln G_i+\ln \pi(D_n|i)}}{\sum_{j\in D_n} e^{V_{jn}+\ln G_j+\ln \pi(D_n|j)}}$$

* Guevara A. and Ben-Akiva M.E. Sampling of alternatives in Multivariate Extreme Value (MEV) Models Proceedings WCTR 2010

Sampling methodology for MEV models

- MEV formulation for nested logit:

$$G = \sum_{m=1}^{M} \left(\sum_{i \in C_{m(i)n}} e^{\mu_m V_{in}} \right)^{\frac{\mu}{\mu_m}}$$

$$\ln G_{in} = \left(\frac{\mu}{\mu_{m(i)}} - 1 \right) \left(\ln \sum_{j \in C_{m(i)n}} e^{\mu_{m(i)} V_{jn}} \right) + \ln \mu + (\mu_{m(i)} - 1) V_{in}$$

Sampling methodology for MEV models

- Approximation of the logsum:

$$\left(\ln \sum_{j \in C_{mn}} e^{\mu_m V_{jn}}\right) \approx \left(\ln \sum_{j \in D_{mn}} w_{jn} e^{\mu_m V_{jn}}\right)$$

- with

$$w_{jn} = \frac{\widetilde{n}_{jn}}{E_n(j)}$$

Number of times alternative *j* was sampled

Expected number of times alternative *j* might be sampled

Sampling methodology for MEV models

- If sampling of alternatives is without replacement:

$$\widetilde{n}_{jn} = 1$$

$$E_n(j) = P_n(j) + (1 - P_n(j)) \left(\sum_{\substack{l \in C_{m(j)} \\ l \neq j}} P_n(l) \frac{\widetilde{J}_{m(j)} - 1}{J_{m(j)} - 1} + \left(1 - \sum_{\substack{l \in C_{m(j)} \\ l \neq j}} P_n(l) \right) \frac{\widetilde{J}_{m(j)}}{J_{m(j)}} \right)$$

True choice probabilities

Sampling methodology for MEV models

- Choice probability with sampling:

$$P_n(i|D_n) = \frac{e^{V_{in} + \ln G'_i(D_n) + \ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}{\sum_{j \in D_n} e^{V_{jn} + \ln G'_j(D_n) + \ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}$$

- where: $J_{m(i)}$ is the size of the nest containing alternative *i*

  $\widetilde{J}_{m(i)}$ is the number of alternatives in the sample

  $\ln G'_i(D_n)$ considers the proposed approximation for the logsum

Bootstrapping

- Method for statistical inference

1. Re-sampling from the original sample of alternatives (with replacement)
2. Re-estimation of the logsum over the new sample
3. Estimation of the bias.

Bootstrapping

- Bias of the approximation:

$$\left(\ln \sum_{j \in D_{mn}} w_{jn} e^{\mu_m V_{jn}}\right) - \left(\ln \sum_{j \in C_{mn}} e^{\mu_m V_{jn}}\right)$$

- Bootstrap estimator of the bias:

$$\rho_{mn} = \frac{1}{K}\left(\sum_{k=1}^{K} \ln \sum_{j \in D_{mn}^{k}} w_{jn} e^{\mu_m V_{jn}}\right) - \left(\ln \sum_{j \in D_{mn}} w_{jn} e^{\mu_m V_{jn}}\right)$$

- where K is the number of re-sampling instances

Bootstrapping

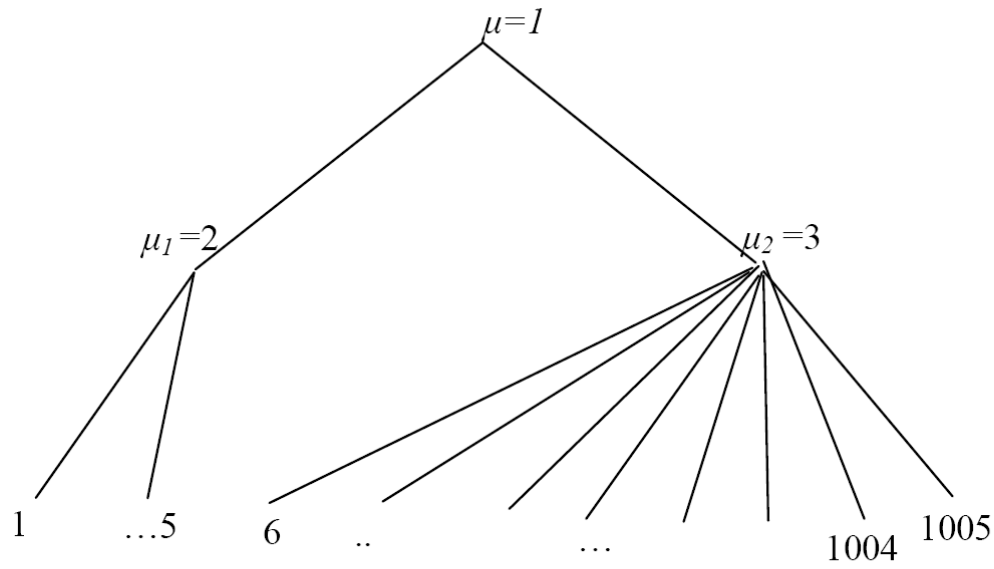
- Probability for estimation after bootstrapping:

$$P_n(i|D_n) = \frac{e^{V_{in}+\ln G_i''(D_n)+\ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}{\sum_{j\in D_n} e^{V_{jn}+\ln G_j''(D_n)+\ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}$$

$$\ln G_i''(D_n) = \left(\frac{\mu}{\mu_{m(i)}} - 1\right)\left[\left(\ln \sum_{j\in D_{m(i)n}} w_{jn} e^{\mu_{m(i)} V_{jn}}\right) - \rho_{mn}\right] + \ln \mu + (\mu_{m(i)} - 1) V_{in}$$

Experiment

- Nested logit:



- Utility functions: $V_{in} = \beta_a a_{in} + \beta_b b_{in}$
- Attributes: $a_{in}, b_{in} \sim U(-1,1)$
- True parameters: $\beta_a = \beta_b = 1$

Experiment

- Simulation of choices considering all alternatives in nest 2, following

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

- Sample of alternatives without replacement for nest 2 and estimation following:

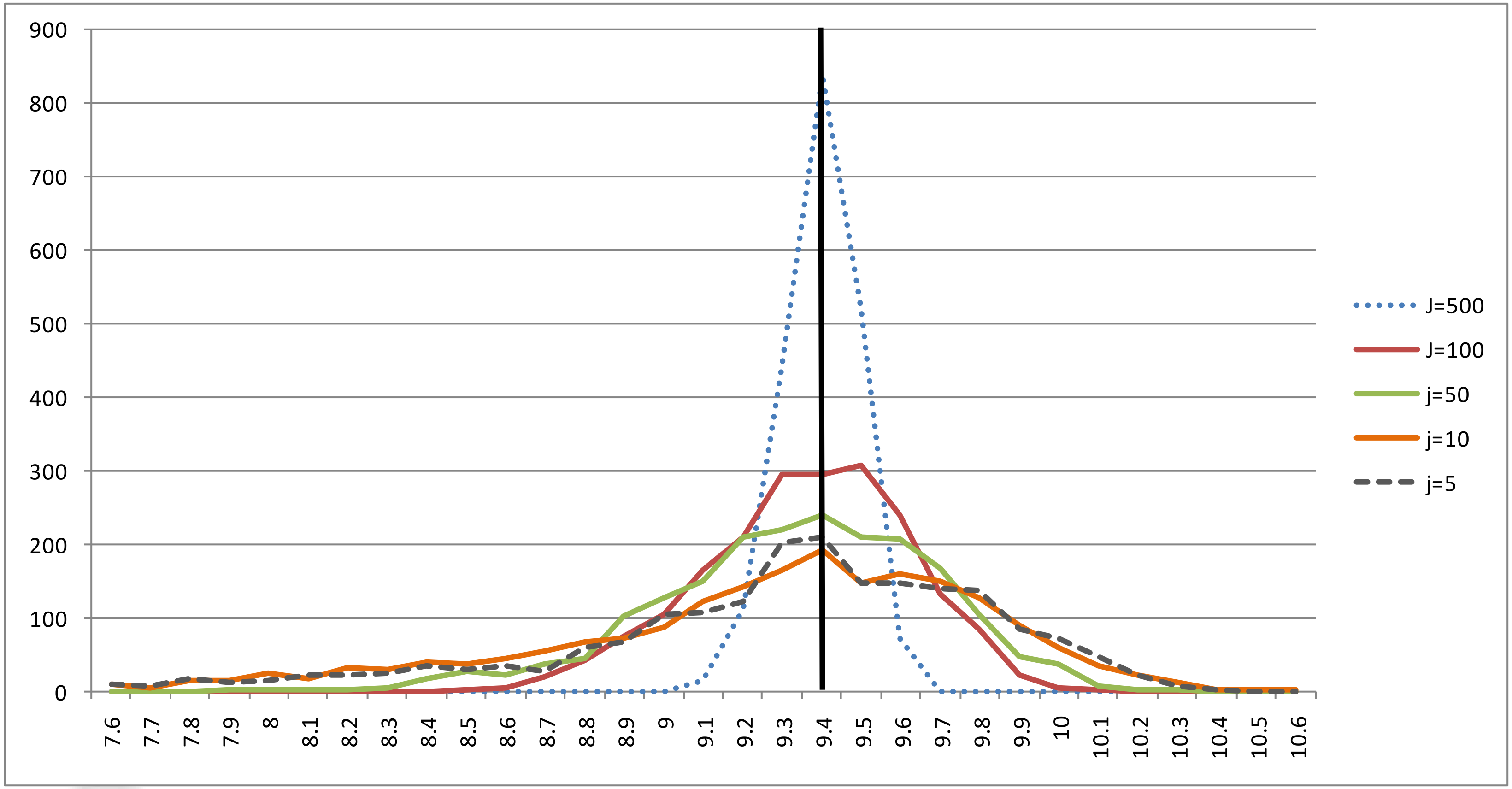
$$P_n(i|D_n) = \frac{e^{V_{in} + \ln G'_i(D_n) + \ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}{\sum_{j \in D_n} e^{V_{jn} + \ln G'_j(D_n) + \ln \frac{J_{m(i)}}{\widetilde{J}_{m(i)}}}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Experiment

The bias of the approximated logsum decreases with the sample size



	logsum	st_dev
full	9.347	-
J_2=500	9.346	0.092
J_2=100	9.312	0.256
J_2=50	9.286	0.346
J_2=10	9.196	0.580
J_2=5	9.168	0.685

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Results

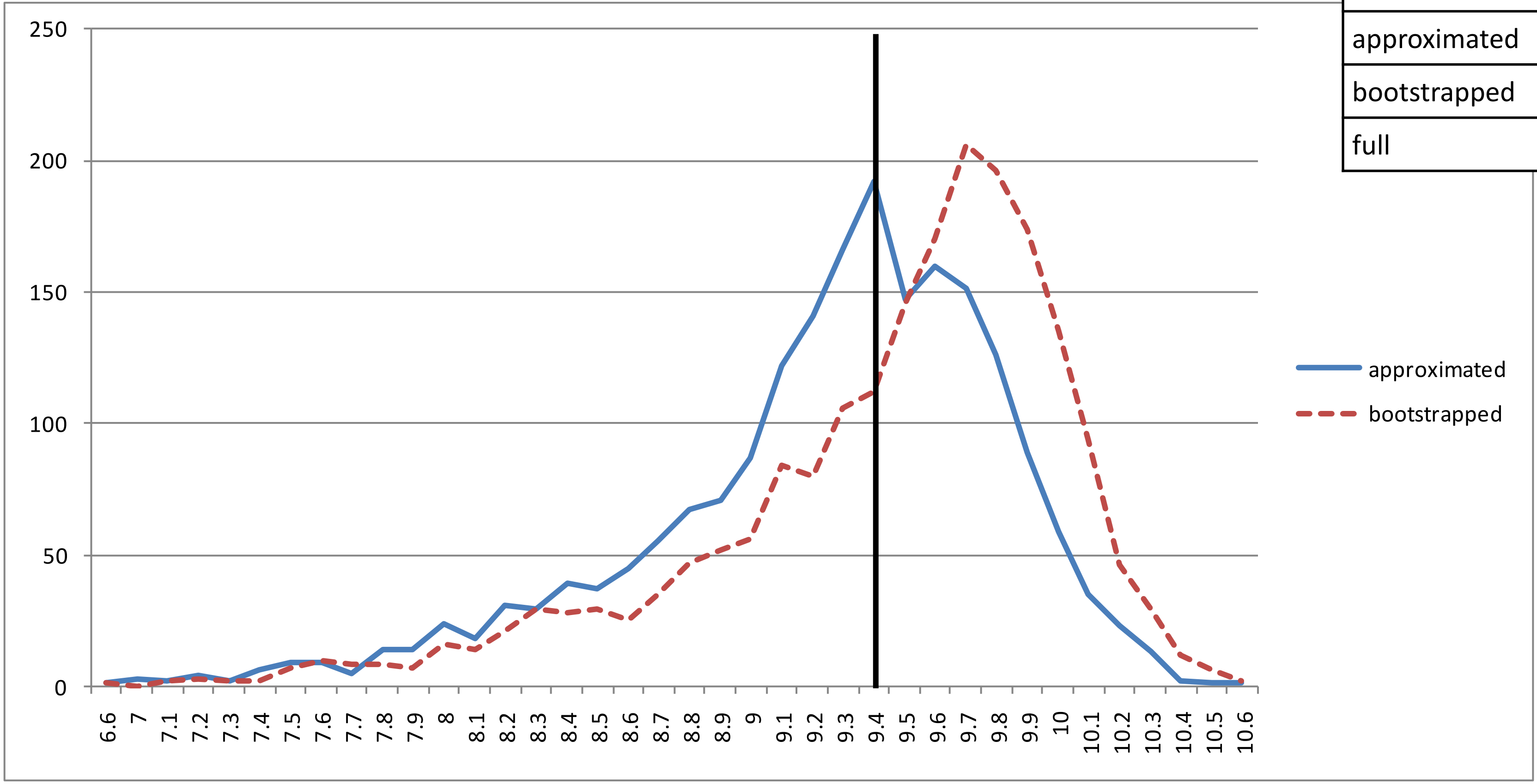
- Estimation results using true probabilities for the logsum approximation (sampling 10 alternatives for nest 2)

Parameter	Value	stdError	t-test (against 0)	True Parameters	t-test (against TP)
beta_a	0.8862	0.0603	14.6904	1	1.8874
beta_b	0.8328	0.0550	15.1306	1	3.0371*
mu1	2.4220	0.5168	4.6870	2	0.8167
mu2	3.4989	0.2186	16.0048	3	2.2820*
loglikelihood	-2658.21				

* parameters statistically different from the true parameters

Results

Bootstrap: logsum values distribution

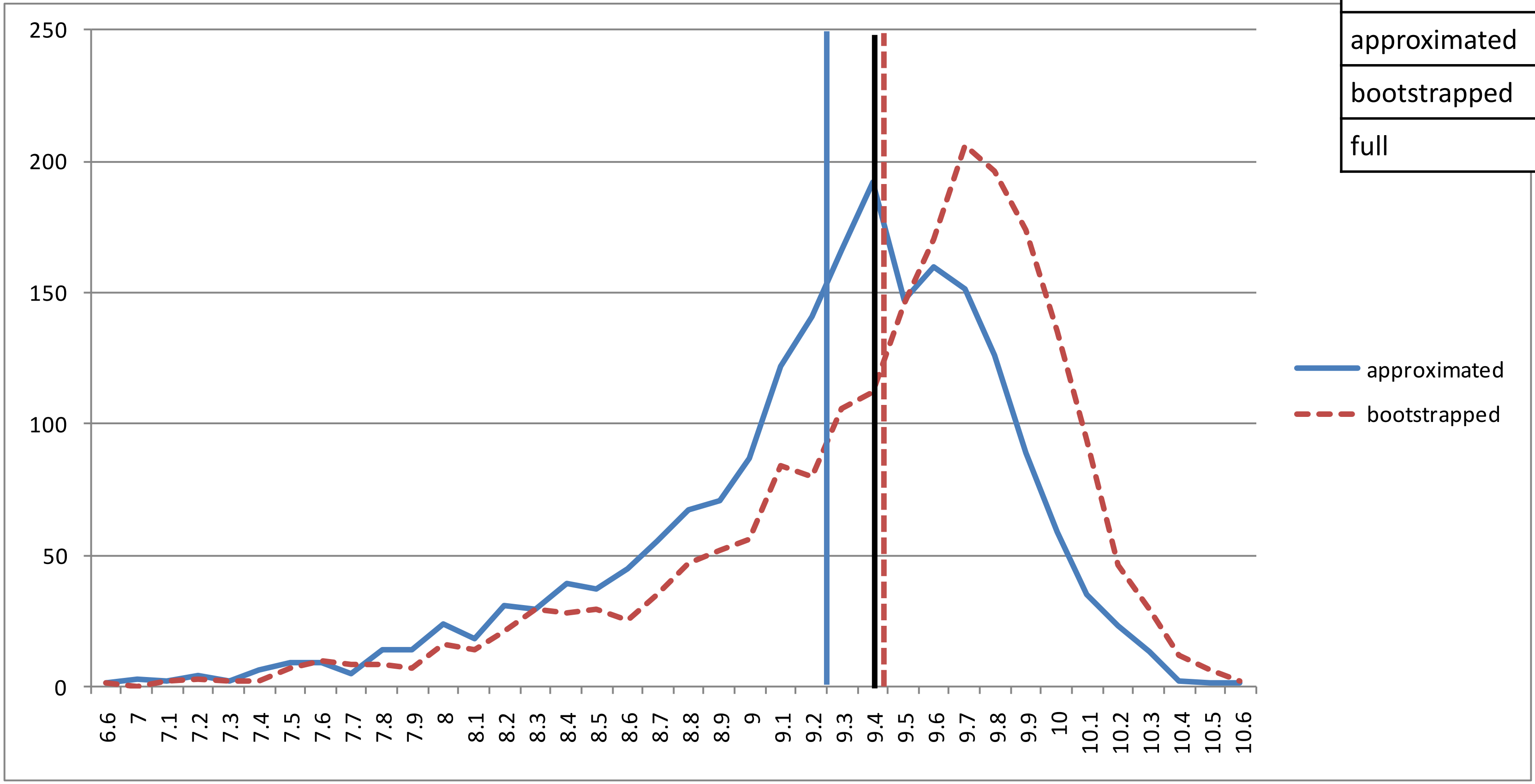


	logsum	stdev
approximated	9.196	0.580
bootstrapped	9.397	0.585
full	9.347	-

Results

Bootstrap: logsum values distribution

	logsum	stdev
approximated	9.196	0.580
bootstrapped	9.397	0.585
full	9.347	-



Results

- Estimation results after bootstrapping

Parameter	Value	stdError	t-test (against 0)	True Parameter	t-test (against TP)	estimates before bootstrapping
beta_a	0.9476	0.0630	15.0505	1	0.8325	0.8862
beta_b	0.8906	0.0574	15.5135	1	1.9060	0.8328
mu1	2.2446	0.4865	4.6136	2	0.5027	2.4220
mu2	3.2761	0.1985	16.5009	3	1.3907	3.4989
loglikelihood	-2659.15					-2658.21

Conclusions

- Bootstrapping allows to correct the estimated value of the sampled logsum.
- Unbiased parameter estimates are obtained
- Further work:
  - Sensitivity analysis
  - Does improvement happens with more realistic approximations of the logsum?
  - Extension to other MEV models