Choice task blocking and design efficiency

Mikołaj Czajkowski, Wiktor Budziński

czaj.org

Researching NGENE

- NGENE allows to optimize DCE design for different models
  - OOD / MNL / MXL
- One can use blocking
  - More choice sets in total than choice tasks presented to one respondent
- Include alternative specific constants in efficiency measures
  - ';con' – default off!
- Avoid particular attribute level combinations
    - repetition within alternatives in a choice task
    - strict attribute level dominance
    - choice task repetition
  - Matters particularly for unlabeled alternatives
  - Can me imposed by placing an asterisk next to the names of the alternatives (default off)

Monte-Carlo simulation of NGENE performance

- Goals – research the influence of:
  - Blocking
  - The inclusion of alternative specific constants in the efficiency measures
  - MXL precision
  - Model type
- 15 designs x 100 simulated datasets
  - Each design included 48 choice sets
  - 12 choice tasks per respondent (even if no blocking used in design)
  - 1200 respondents
  - Each design optimized for about 2 days

Monte-Carlo simulation of NGENE performance

	model	; rdraws	; rep	; block	; con	NGENE d-error	NGENE d-error (high precision)	Simulated d-error * 1e30			
								Mean	Median	min	max
1.1	**mxl**	**100**	**100**	1	1	0.4151	0.4439	3.9128	1.6908	0.5889	78.1367
1.2	**mxl**	**1000**	**1000**	1	1	0.4151	0.4383	4.3882	1.5867	0.6061	215.6998
1.3	**mxl**	**10000**	**1200**	1	1		0.4394	2.2341	1.8054	0.9440	9.3334
2.1	**mxl**	**100**	**100**	1	0	0.3113	0.3304	1.5419	1.3710	0.6293	3.7269
2.2	**mxl**	**1000**	**1000**	1	0	0.3421	0.3471	2.1977	1.9529	0.9850	7.4614
2.3	**mxl**	**10000**	**1200**	1	0		0.3542	6.0659	3.0184	1.3638	99.1343
3.1	**mxl**	**100**	**100**	0	1	0.3768	0.5415	4.3769	2.7256	1.1420	104.8539
3.2	**mxl**	**1000**	**1000**	0	1	0.4615	0.4993	1.7092	0.8619	0.4220	68.3785
3.3	**mxl**	**10000**	**1200**	0	1		0.5349	4.0756	3.4583	1.2738	22.4224
4.1	**mnl**			1	1			0.3070	0.2547	0.1520	1.2487
4.2	**ood**			1	1			1.1414	0.6939	0.2502	14.2980
5.1	**mnl**			1	0			0.3157	0.2413	0.1191	1.6325
5.2	**ood**			1	0			0.8741	0.4314	0.2137	31.6656
6.1	**mnl**			0	1			0.2770	0.2554	0.1190	0.7470
6.2	**ood**			0	1			1.3426	0.6011	0.1916	19.0074

Results

- Observations:
  1. MXL designs outperformed by MNL and OOD!
  2. Including '; con' or '; block' does not seem to matter
- This should not be happening
- So why?
  - NGENE extremely slow, not able to find good MXL designs in 2 days
    - C.f. 'the loss of efficiency when using MNL design for MXL models is low' reported in: Bliemer, M. C. J., and Rose, J. M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B: Methodological, 44(6):720-734.
    - Also supported by our finding that MXL designs efficiency measures very noisy – difficult to draw conclusions
  - Blocking is completely ignored by NGENE:
    - Michiel Bliemer on Mon Sep 21, 2015 11:20 pm @ http://www.choice-metrics.com/forum/viewtopic.php?f=2&t=277:
      "Ngene does NOT take blocking into account in the optimisation, blocking in Ngene is done after the design has been generated (in line with all other design types). This of course is inconsistent, but simultaneously optimising for a blocking column would add a further complication to the optimisation and is very difficult. We hope to further improve this in the future. We mentioned this issue with blocking in Rose and Bliemer (2013?) in Transportation on sample size requirements. So this means that your simulation in which you take blocking into account will yield different results. Hence my suggestion to use a single block to avoid this inconsistency."
  - Does '; con' work?

Conclusions

- Use MNL designs
- Ignore blocking
- Future
  - Develop a more efficient algorithm for MXL designs
    - … which takes blocking into account
  - Test if alternative specific constants are taken into account
    - …and how – all but one? the ones with priors provided?

One more thing …

- Does using constant blocks matter?
- Randomly select choice sets presented to each respondent from the full set of choice sets?
  - Randomize the order of choice sets for each *n* respondents and present them in that order, randomize again for the next *n* respondents, and so on… – results in more even distribution of how many times each choice set is used than random selection
    - Still not perfect because some respondents drop out
- Test setup:
  - 100 repetitions, 1200 respondents, 8 choice sets (4 per respondent), 1000 Sobol draws, designs for OOD, MNL, MXL

Results

Design		MNL		OOD		MXL	
Fixed blocks		1	0	1	0	1	0
Simulated d-error*1e24	mean	23.14	8.50	29.17	14.02	4.32	6.28
	median	11.90	3.75	16.95	5.62	2.19	3.85
	min	3.02	0.92	4.38	1.39	1.03	1.34
	max	224.56	142.96	301.01	206.17	45.15	58.89
Simulated trace	mean	0.18	0.14	0.19	0.15	0.11	0.11
	median	0.15	0.10	0.16	0.11	0.09	0.10
	min	0.09	0.07	0.10	0.08	0.07	0.08
	max	0.61	1.50	0.55	0.72	0.51	0.31

- Do not use fixed blocks, randomize choice sets for each respondent
- Side note: this time only 2 blocks and MXL designs optimized for weeks => better performance than MNL or OOD