Common mistakes in discrete choice modelling: A guide to mistakes that even the so called experts make

Danish Choice Modelling Day

JOHN ROSE| Prof.
Institute of Transport and Logistics Studies

THE UNIVERSITY OF SYDNEY

Mistake #1

Not thinking things through

Not thinking things through

And not thinking for ourselves…

- Most applied studies making use of discrete choice models report statistical tests
  - Reviewers demand these, so we supply them
- Putting aside nested versus non-nested model forms, one of the most over used, and least understood tests reported in discrete choice models is the ρ^2 statistic

$$\rho^2 = 1 - \frac{LL(M_1)}{LL(M_2)}$$

Not thinking things through

And not thinking for ourselves…

› Think about how we derive/frame the model…

$$LL_N(\beta \mid X, y) = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{jsn} \log P_{jsn}(X \mid \beta)$$

$$U_{nj} \geq U_{ni}, \quad \forall i \neq j, (i, j) \in S(n)$$

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

$$\Rightarrow P_{nj} = \Pr\left\{\varepsilon_{nj} \leq \varepsilon_{ni} + V_{nj} - V_{ni}, \; \forall i \neq j, (i, j) \in S(n)\right\}$$

$$\Rightarrow P_{nj} = \Pr\left\{U_{nj} \geq U_{ni}, \; \forall i \neq j, (i, j) \in S(n)\right\}$$

Not thinking things through

And not thinking for ourselves…

› Consider how the ρ^2 statistic can take a value of 1

$$\rho^2 = 1 - \frac{LL(M_1)}{LL(M_2)}$$

› $LL(M_1)$ must equal 0, but can it?

- For this to happen in practice, the choices will need to be deterministic and hence the error terms $\varepsilon_{nj} = 0$
- If this is the case, then for $y_{njs} = 1$, $P_{njs} = 1$ and $LL(M_1) = 0$
- But what if $\varepsilon_{nj} \neq 0$?
  - Then for all N in all S for all J, $P_{njs} \neq 1$, and $LL(M_1) \neq 0$
  - So is the upper bound of the ρ^2 statistic = 1?

Even if you could find the perfect utility specification, the estimation procedures for such a model will likely fail

Mistake #2

Not reading outside of our own literature

Not reading outside of our own literature

But you can compute the upper bounds on (variations of) ρ^2

- Hu et al. (2006) Pseudo-r^2 in logistic regression model, Statistica Sinica 16, 847-860.
- Maddala, G. S. (1983) Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, University Press, Cambridge.

THE UNIVERSITY OF SYDNEY

Not reading outside of our own literature

And not reading outside of our own narrow field…

- In linear models the interpretation of the coefficient of the interaction between two variables is straightforward...

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \delta z$$

- If x_1 and x_2 are continuous the interaction effect of the independent variables x_1 and x_2 *is the cross-derivative of the expected value of y*

$$\frac{\partial^2 E[y \mid x_1, x_2, z]}{\partial x_1 \partial x_2} = \beta_{12}$$

- If x_1 and x_2 are dichotomous, the interaction effect of change in both x_1 and x_2 from zero to one is found by taking discrete differences

$$\frac{\partial^2 E[y \mid x_1, x_2, z]}{\Delta x_1 \Delta x_2} = \beta_{12}$$

Not reading outside of our own literature

And not reading outside of our own narrow field…

- The intuition from linear models, however, does not extend to nonlinear models...
- Consider the probit model

$$E\left[y \mid x_1, x_2, z\right] = \Phi\left(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \delta z\right) = \Phi\left(.\right)$$

where Φ is the standard normal cumulative distribution

- If x_1 and x_2 are continuous

$$\frac{\partial^2 E\left[y \mid x_1, x_2, z\right]}{\partial x_1 \partial x_2} = \beta_{12} \Phi'\left(.\right) + \left(\beta_1 + \beta_{12} x_2\right)\left(\beta_2 + \beta_{12} x_1\right) \Phi''\left(.\right)$$

Not reading outside of our own literature

And not reading outside of our own narrow field…

› Note that even if $\beta_{12} = 0$

$$\frac{\partial^2 \Phi(.)}{\partial x_1 \partial x_2}\bigg|_{\beta_{12}=0} = \beta_1 \beta_2 \Phi''(.)$$

› This implies that the statistical significance of the interaction effect cannot be tested with a simple *t*-test on the coefficient of the interaction term β_{12} and that the interaction effect is conditional on the independent variables, unlike the interaction effect in linear models

Not reading outside of our own literature

And not reading outside of our own narrow field…

› Greene, W.H. (2010) Testing hypotheses about interaction terms in nonlinear models, *Economics Letters*, 107(2), 291–296.

› Ai, C. and Norton, E. (2003) Interaction terms in logit and probit models, *Economics Letters*, 80(1), 123–129.

Mistake #3

And blindly following others without thinking (or reading)…

Choice experiment

› We make use of a subset of data collected in Sydney in 2004

- 60 respondents, completed 16 choice tasks each

Sydney Road System

Practice Game

Make your choice given the route features presented in this table, thank you.

	Details of Your Recent Trip	Road A	Road B
Time in free-flow traffic (mins)	50	25	40
Time slowed down by other traffic (mins)	10	12	12
Travel time variability (mins)	+/- 10	+/- 12	+/- 9
Running costs	$ 3.00	$ 4.20	$ 1.50
Toll costs	$ 0.00	$ 4.80	$ 5.60

If you make the same trip again, which road would you choose?	Current Road	Road A	Road B (selected)
If you could only choose between the 2 new roads, which road would you choose?		Road A	Road B (selected)

$$V_{ref} = \beta_{ref} + \beta_{ff} FFT_{ref} + \beta_{sdt} SDT_{ref} + \beta_{var} VR_{ref} + \beta_{tc} PC_{ref} + \beta_{toll} TC_{ref},$$

$$V_{SC_1} = \beta_{SC_1} + \beta_{ff} FFT_{SC_1} + \beta_{sdt} SDT_{SC_1} + \beta_{var} VR_{SC_1} + \beta_{tc} PC_{SC_1} + \beta_{toll} TC_{SC_1},$$

$$V_{SC_2} = \beta_{ff} FFT_{SC_2} + \beta_{sdt} SDT_{SC_2} + \beta_{var} VR_{SC_2} + \beta_{tc} PC_{SC_2} + \beta_{toll} TC_{SC_2}.$$

Back | Next

Typical modelling process

Increase the number of draws until stability in estimates achieved

		Par.	t-ratio
Ref. ASC	Fixed par.	1.019	2.52
SP1 ASC	Fixed par.	0.429	1.52
FF time (N)	Mean	-0.115	-4.27
	Std dev.	0.147	3.95
SD time (N)	Mean	-0.185	-6.00
	Std dev.	0.132	3.57
Petrol cost (L)	Mean	-0.721	-4.12
	Std dev.	0.499	2.25
Toll cost (L)	Mean	-0.271	-1.86
	Std dev.	0.797	5.85
Trvl time var. (N)	Mean	-0.034	-2.14
	Std dev.	0.122	6.18

Draws	100
LL(β)	-528.56136

Median	PC	$0.49
Median	TC	$0.76

FF	PC	$14.16
FF	TC	$9.02
SD	PC	$22.85
SD	TC	$14.56
VR	PC	$4.23
VR	TC	$2.69

Issue #1

Mistake 3

Marginal utility

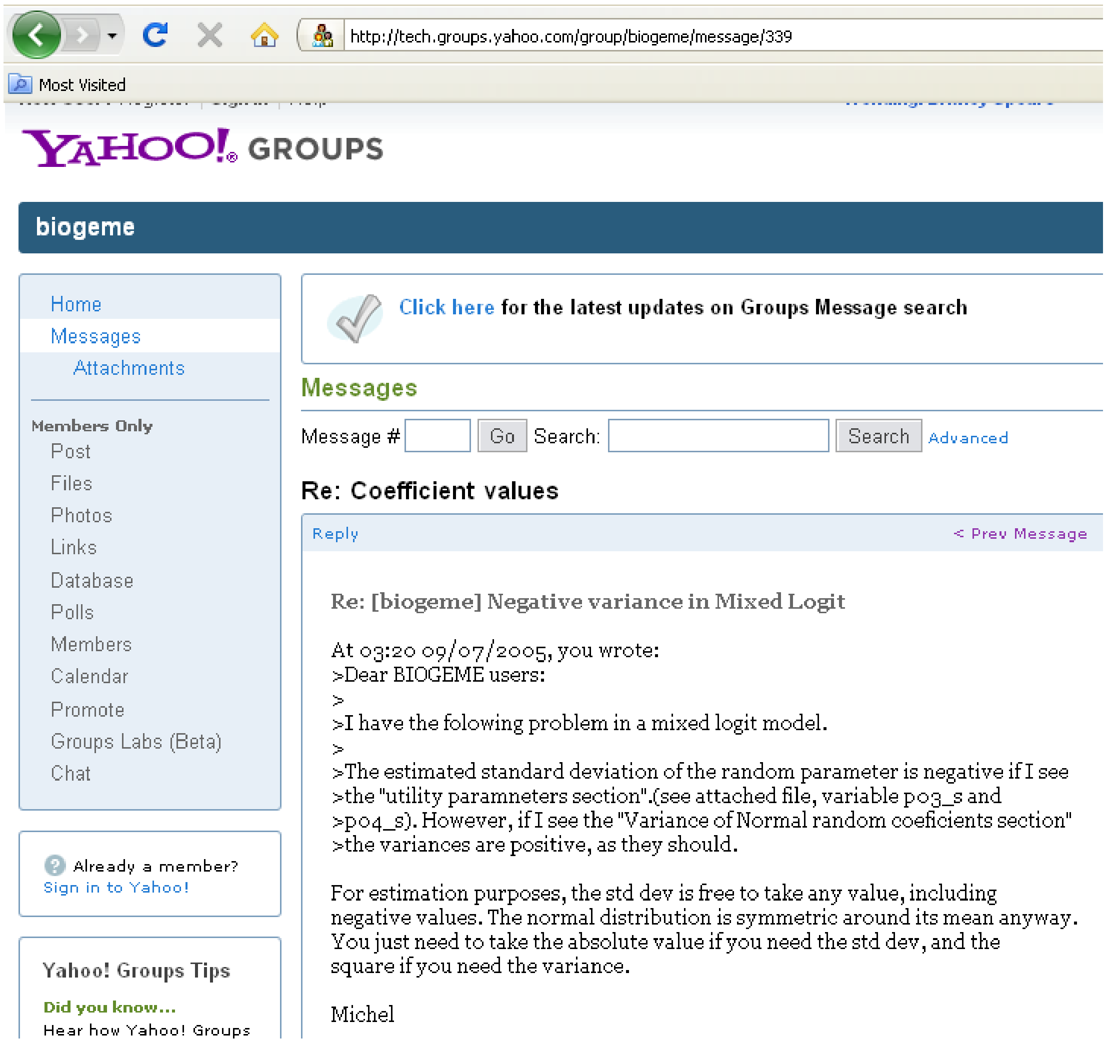
› Marginal utility may be represented via a random parameter approach

$$\beta_{nk} = \bar{\beta}_k \pm \eta_k z_{ns},$$

› where $\bar{\beta}_k$ represents the mean or some other measure of central tendency for the distribution of marginal utilities held by the sampled population and η_k represents a deviation or spread of preferences amongst sampled respondents around the mean (or other measure of central tendency) marginal utility.

› Z_{ns} represents random draws taken from a pre-specified distribution for each respondent *n* and choice task *s* (cross sectional model). Rather than assume a distribution of *s*, it is also possible to take draws over respondents only, such that Z_n.

Is it really a ±?



Example

› Consider the following choice task

	Alternative A	
S	x_{11}	x_{12}
1	4	15

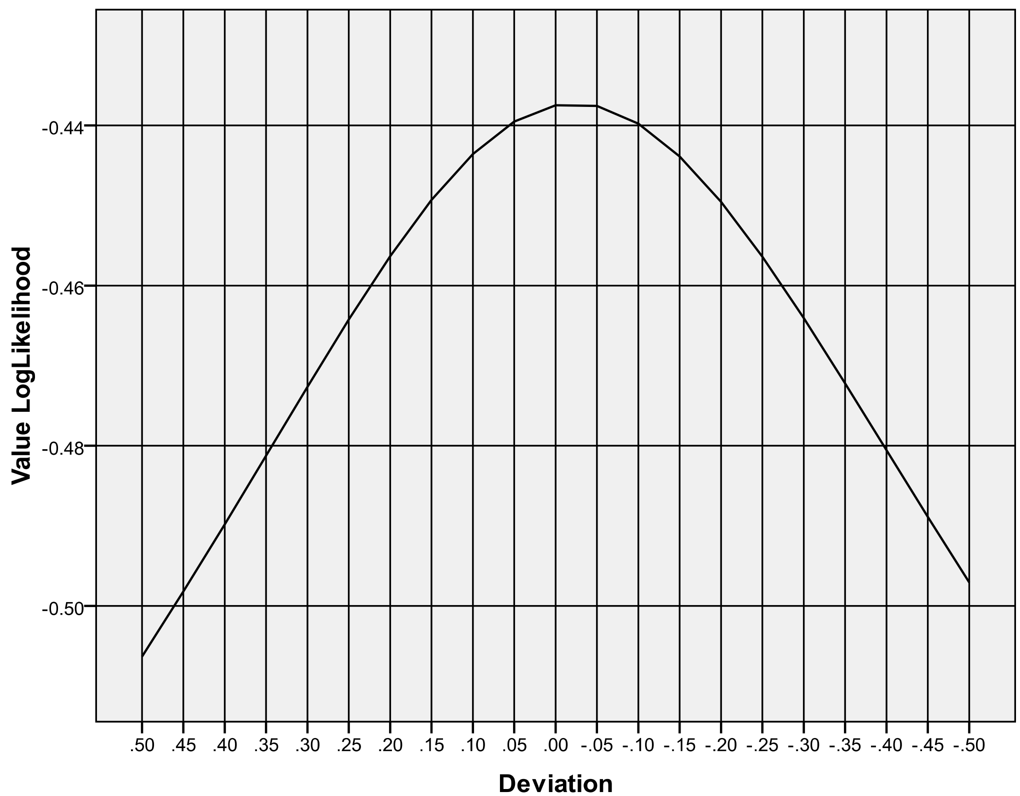
› and assuming the following

$$V_1 = \beta_1 x_{11} + \ldots$$

$$V_2 = \beta_1 x_{21} + \ldots$$

› where $\beta_1 \sim N(-0.8,\eta)$, β_2 = -0.

› Now vary the deviation para



Empirical example: Problem

				+ve sig start value		-ve sig start value	
		Par.	t-ratio	Par.	t-ratio	Par.	t-ratio
Ref. ASC	Fixed par.	0.831	2.06	0.827	2.05	0.957	2.43
SP1 ASC	Fixed par.	0.440	1.52	0.404	1.38	0.349	1.27
FF time (N)	Mean	-0.128	-4.34	-0.138	-4.49	-0.104	-3.98
	Std dev.	0.155	3.85	0.154	3.85	0.129	3.42
SD time (N)	Mean	-0.199	-5.55	-0.183	-5.61	-0.166	-5.51
	Std dev.	0.109	3.09	0.109	3.05	0.086	2.76
Petrol cost (L)	Mean	-0.581	-3.25	-0.555	-3.07	-0.530	-3.36
	Std dev.	0.448	1.30	0.479	1.40	0.474	1.36
Toll cost (L)	Mean	-0.349	-2.43	-0.213	-1.33	-0.652	-4.55
	Std dev.	0.832	6.46	0.874	6.56	0.867	6.42
Trvl time var. (N)	Mean	-0.029	-1.83	-0.030	-1.90	-0.033	-2.03
	Std dev.	0.117	5.33	0.122	5.40	0.162	4.22

Draws	100,000	5,000	5,000
LL(β)	-526.2388	-526.457	-549.227

Median	PC	$0.56	$0.57	$0.59
Median	TC	$0.71	$0.81	$0.52

FF	PC	$13.67	$14.39	$10.62
FF	TC	$10.84	$10.22	$12.00
SD	PC	$21.36	$19.14	$16.93
SD	TC	$16.94	$13.60	$19.13
VR	PC	$3.14	$3.18	$3.33
VR	TC	$2.49	$2.26	$3.77

Issue #1

Solution # 1

- 1. Asymptotically, if you take a sufficient number of draws, this problem will disappear
  - That is, if one uses a large enough number of draws, then the MSL will become symmetrical around $\pm\eta$
- 2. Use, antithetic sequences (Hammersley and Morton 1956)
  - represents a systematic modification of any other type of sequence and as such, can be applied to PMC or any QMC method
  - Therefore, it can be applied to any type of base draw (Halton, Sobol, MLHS, etc.)

Solution # 1

› 2. Antithetic sequences (Hammersley and Morton 1956)

- The process of generating antithetic draws involves taking each value drawn from an existing density and using these to construct new draws by inflecting the original values around the midpoint of the original density

$$\begin{bmatrix} d_{11,1} \\ d_{11,2} \\ d_{11,3} \\ d_{11,4} \\ d_{11,5} \\ d_{11,6} \\ d_{11,7} \\ d_{11,8} \end{bmatrix} = \begin{bmatrix} d_{11}^1 & d_{11}^2 & d_{11}^3 \\ 1-d_{11}^1 & d_{11}^2 & d_{11}^3 \\ d_{11}^1 & 1-d_{11}^2 & d_{11}^3 \\ d_{11}^1 & d_{11}^2 & 1-d_{11}^3 \\ 1-d_{11}^1 & 1-d_{11}^2 & d_{11}^3 \\ 1-d_{11}^1 & d_{11}^2 & 1-d_{11}^3 \\ d_{11}^1 & 1-d_{11}^2 & 1-d_{11}^3 \\ 1-d_{11}^1 & 1-d_{11}^2 & 1-d_{11}^3 \end{bmatrix} = \begin{bmatrix} 0.8125 & 0.7037 & 0.2800 \\ 0.1875 & 0.7037 & 0.2800 \\ 0.8125 & 0.2963 & 0.2800 \\ 0.8125 & 0.7037 & 0.7200 \\ 0.1875 & 0.2963 & 0.2800 \\ 0.1875 & 0.7037 & 0.7200 \\ 0.8125 & 0.2963 & 0.7200 \\ 0.1875 & 0.2963 & 0.7200 \end{bmatrix}$$

THE UNIVERSITY OF SYDNEY

Issue #1

Empirical example: solution

								Antithetic Halton draws			
				+ve sig start value		-ve sig start value		+ve sig start value		-ve sig start value	
		Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio
Ref. ASC	Fixed par.	0.831	2.06	0.827	2.05	0.957	2.43	0.839	2.09	0.839	2.09
SP1 ASC	Fixed par.	0.440	1.52	0.404	1.38	0.349	1.27	0.395	1.36	0.395	1.36
FF time (N)	Mean	-0.128	-4.34	-0.138	-4.49	-0.104	-3.98	-0.138	-4.50	-0.138	-4.50
	Std dev.	0.155	3.85	0.154	3.85	0.129	3.42	0.159	3.93	0.159	3.93
SD time (N)	Mean	-0.199	-5.55	-0.183	-5.61	-0.166	-5.51	-0.188	-5.55	-0.188	-5.55
	Std dev.	0.109	3.09	0.109	3.05	0.086	2.76	0.102	2.92	0.102	2.92
Petrol cost (L)	Mean	-0.581	-3.25	-0.555	-3.07	-0.530	-3.36	-0.549	-3.02	-0.549	-3.02
	Std dev.	0.448	1.30	0.479	1.40	0.474	1.36	0.444	1.36	0.444	1.36
Toll cost (L)	Mean	-0.349	-2.43	-0.213	-1.33	-0.652	-4.55	-0.325	-2.22	-0.325	-2.22
	Std dev.	0.832	6.46	0.874	6.56	0.867	6.42	0.824	6.04	0.824	6.04
Trvl time var. (N)	Mean	-0.029	-1.83	-0.030	-1.90	-0.033	-2.03	-0.025	-1.59	-0.025	-1.59
	Std dev.	0.117	5.33	0.122	5.40	0.162	4.22	0.125	5.50	0.125	5.50

Draws		100,000	5,000	5,000	5024 (Base: 157 Halton)	5024 (Base: 157 Halton)
LL(β)		-526.2388	-526.457	-549.227	-526.003	-526.003

Median	PC	$0.56	$0.57	$0.59	$0.58	$0.58
Median	TC	$0.71	$0.81	$0.52	$0.72	$0.72

FF	PC	$13.67	$14.39	$10.62	$14.29	$14.29
FF	TC	$10.84	$10.22	$12.00	$11.42	$11.42
SD	PC	$21.36	$19.14	$16.93	$19.57	$19.57
SD	TC	$16.94	$13.60	$19.13	$15.65	$15.65
VR	PC	$3.14	$3.18	$3.33	$2.58	$2.58
VR	TC	$2.49	$2.26	$3.77	$2.07	$2.07

Issue #2

Mistake 3

Issue #2

Draws from densities

› The estimation of the MMNL model therefore involves first the generation of multidimensional infinite sequences that fill the 0-1 interval

	Halton			Sobol			MLHS		
r	***R1***	***R2***	***R3***	***R1***	***R2***	***R3***	***R1***	***R2***	***R3***
1	0.8125	0.7037	0.2800	0.7500	0.2500	0.7500	0.0481	0.3857	0.7214
2	0.1875	0.1481	0.4800	0.2500	0.7500	0.2500	0.5481	0.5857	0.0214
3	0.6875	0.4815	0.6800	0.3750	0.3750	0.6250	0.2481	0.7857	0.2214
4	0.4375	0.8148	0.8800	0.8750	0.8750	0.1250	0.4481	0.8857	0.9214
5	0.9375	0.2593	0.1200	0.6250	0.1250	0.3750	0.3481	0.9857	0.6214
6	0.0313	0.5926	0.3200	0.1250	0.6250	0.8750	0.1481	0.2857	0.1214
7	0.5313	0.9259	0.5200	0.1875	0.3125	0.3125	0.9481	0.0857	0.5214
8	0.2813	0.0741	0.7200	0.6875	0.8125	0.8125	0.8481	0.6857	0.8214
9	0.7813	0.4074	0.9200	0.9375	0.0625	0.5625	0.7481	0.4857	0.4214
10	0.1563	0.7407	0.1600	0.4375	0.5625	0.0625	0.6481	0.1857	0.3214

› Each random parameter is assigned to a different column of the generated table…

› What happens if you change the order of the assignment?

Issue #2

Example

- Consider the following choice task

	Alternative A			Alternative B			
S	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	Choice
1	4	15	20	1	20	15	1

- and assuming the following utility function

$$V_1 = \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13},$$
$$V_2 = \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23},$$

- where $\beta_1 \sim N(-0.8, 0.5)$, $\beta_2 \sim N(-0.3, 0.05)$ and $\beta_3 = 0.3$

β_2

Example

› The following choice probabilities and MSL are obtained…

Draw (*r*)	Prime 2	Prime 3	(Prime 2)	(Prime 3)	Prob(*J*=1)	Prob(*J*=2)	(Prime 3)	(Prime 2)	Prob(*J*=1)	Prob(*J*=2)
1	0.813	0.704	-0.356	-0.273	0.858	0.142	-0.532	-0.256	0.765	0.235
2	0.188	0.148	-1.244	-0.352	0.385	0.615	-1.322	-0.344	0.322	0.678
3	0.688	0.481	-0.556	-0.302	0.793	0.207	-0.823	-0.276	0.601	0.399
4	0.438	0.815	-0.879	-0.255	0.535	0.465	-0.352	-0.308	0.879	0.121
5	0.938	0.259	-0.033	-0.332	0.955	0.045	-1.123	-0.223	0.320	0.680
6	0.031	0.593	-1.731	-0.288	0.095	0.905	-0.683	-0.393	0.805	0.195
7	0.531	0.926	-0.761	-0.228	0.588	0.412	-0.077	-0.296	0.940	0.060
8	0.281	0.074	-1.090	-0.372	0.523	0.477	-1.523	-0.329	0.194	0.806
9	0.781	0.407	-0.412	-0.312	0.861	0.139	-0.917	-0.261	0.514	0.486
10	0.156	0.741	-1.305	-0.268	0.254	0.746	-0.477	-0.350	0.861	0.139
				MSL	-7.06221			MSL	-5.9134	

Empirical example: Problem

		FFT, SDT, PC, TC, VR Halton		FFT, SDT, PC, TC, VR Halton		VR,TC,PC,SDT,FFT Halton		FFT, SDT, PC, TC, VR Halton (antithetic)		VR,TC,PC,SDT,FFT Halton (antithetic)	
		Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio
Ref. ASC	Fixed par.	0.831	2.06	0.856	2.11	0.791	1.99	0.792	1.98	0.820	2.04
SP1 ASC	Fixed par.	0.440	1.52	0.414	1.43	0.430	1.50	0.403	1.39	0.393	1.35
FF time (N)	Mean	-0.128	-4.34	-0.141	-4.55	-0.133	-4.47	-0.128	-4.35	-0.139	-4.48
	Std dev.	0.155	3.85	0.157	3.88	0.147	3.79	0.153	3.84	0.149	3.81
SD time (N)	Mean	-0.199	-5.55	-0.186	-5.62	-0.185	-5.58	-0.191	-5.55	-0.186	-5.56
	Std dev.	0.109	3.09	0.109	3.07	0.091	2.74	0.106	2.99	0.107	3.08
Petrol cost (L)	Mean	-0.581	-3.25	-0.558	-3.10	-0.526	-2.94	-0.584	-3.25	-0.579	-3.18
	Std dev.	0.448	1.30	0.492	1.45	0.410	1.23	0.396	1.21	0.507	1.63
Toll cost (L)	Mean	-0.349	-2.43	-0.233	-1.49	-0.368	-2.63	-0.359	-2.52	-0.263	-1.68
	Std dev.	0.832	6.46	0.877	6.75	0.964	7.36	0.815	5.99	0.869	6.09
Trvl time var. (N)	Mean	-0.029	-1.83	-0.035	-2.17	-0.030	-1.89	-0.024	-1.53	-0.025	-1.57
	Std dev.	0.117	5.33	0.123	5.40	0.119	5.33	0.126	5.47	0.122	5.39

Draws	100,000	3840		3840 (Base: 120 Halton)	
LL(β)	-526.2388	-526.649	-531.582	-525.383	-528.809

Median	PC	$0.56	$0.57	$0.59	$0.56	$0.56
Median	TC	$0.71	$0.79	$0.69	$0.70	$0.77

FF	PC	$13.67	$14.75	$13.51	$13.76	$14.84
FF	TC	$10.84	$10.65	$11.54	$10.99	$10.83
SD	PC	$21.36	$19.51	$18.77	$20.58	$19.94
SD	TC	$16.94	$14.09	$16.03	$16.43	$14.55
VR	PC	$3.14	$3.70	$3.05	$2.57	$2.64
VR	TC	$2.49	$2.68	$2.60	$2.06	$1.92

Solution # 2

- 1. Asymptotically, if you take a sufficient number of draws, this problem will disappear
- 2. Use, column permuted sequences (this paper)
  - represents a systematic modification of any other type of sequence and as such, can be applied to PMC or any QMC method
  - Therefore, it can be applied to any type of base draw (Halton, Sobol, MLHS, etc.)

$$\begin{bmatrix} d_{11,1} \\ d_{11,2} \\ d_{11,3} \\ d_{11,4} \\ d_{11,5} \\ d_{11,6} \\ d_{11,7} \\ d_{11,8} \end{bmatrix} = \begin{bmatrix} d_{11}^{1} & d_{11}^{2} & d_{11}^{3} \\ d_{11}^{1} & d_{11}^{3} & d_{11}^{2} \\ d_{11}^{2} & d_{11}^{1} & d_{11}^{3} \\ d_{11}^{2} & d_{11}^{3} & d_{11}^{1} \\ d_{11}^{3} & d_{11}^{1} & d_{11}^{2} \\ d_{11}^{3} & d_{11}^{2} & d_{11}^{1} \end{bmatrix} = \begin{bmatrix} 0.8125 & 0.7037 & 0.2800 \\ 0.8125 & 0.2800 & 0.7037 \\ 0.7037 & 0.8125 & 0.2800 \\ 0.7037 & 0.2800 & 0.8125 \\ 0.2800 & 0.8125 & 0.7037 \\ 0.2800 & 0.7037 & 0.8125 \end{bmatrix}$$

Issue #2

Empirical example: Problem

		FFT, SDT, PC, TC, VR		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT	
		Halton		Halton		Halton		Halton (antithetic)		Halton (antithetic)		Halton (colm. Perm.)		Halton (colm. Perm.)	
		Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio
Ref. ASC	Fixed par.	0.831	2.06	0.856	2.11	0.791	1.99	0.792	1.98	0.820	2.04	0.894	2.22	0.894	2.22
SP1 ASC	Fixed par.	0.440	1.52	0.414	1.43	0.430	1.50	0.403	1.39	0.393	1.35	0.378	1.35	0.378	1.35
FF time (N)	Mean	-0.128	-4.34	-0.141	-4.55	-0.133	-4.47	-0.128	-4.35	-0.139	-4.48	-0.129	-4.44	-0.129	-4.44
	Std dev.	0.155	3.85	0.157	3.88	0.147	3.79	0.153	3.84	0.149	3.81	0.146	3.97	0.146	3.97
SD time (N)	Mean	-0.199	-5.55	-0.186	-5.62	-0.185	-5.58	-0.191	-5.55	-0.186	-5.56	-0.178	-5.59	-0.178	-5.59
	Std dev.	0.109	3.09	0.109	3.07	0.091	2.74	0.106	2.99	0.107	3.08	0.112	3.20	0.112	3.20
Petrol cost (L)	Mean	-0.581	-3.25	-0.558	-3.10	-0.526	-2.94	-0.584	-3.25	-0.579	-3.18	-0.553	-3.17	-0.553	-3.17
	Std dev.	0.448	1.30	0.492	1.45	0.410	1.23	0.396	1.21	0.507	1.63	0.343	0.93	0.343	0.93
Toll cost (L)	Mean	-0.349	-2.43	-0.233	-1.49	-0.368	-2.63	-0.359	-2.52	-0.263	-1.68	-0.301	-2.02	-0.301	-2.02
	Std dev.	0.832	6.46	0.877	6.75	0.964	7.36	0.815	5.99	0.869	6.09	0.825	6.53	0.825	6.53
Trvl time var. (N)	Mean	-0.029	-1.83	-0.035	-2.17	-0.030	-1.89	-0.024	-1.53	-0.025	-1.57	-0.040	-2.39	-0.040	-2.39
	Std dev.	0.117	5.33	0.123	5.40	0.119	5.33	0.126	5.47	0.122	5.39	0.142	5.60	0.142	5.60

Draws	100,000	3840		3840 (Base: 120 Halton)		3840 (Base: 32 Halton)	
LL(β)	-526.2388	-526.649	-531.582	-525.383	-528.809	-527.626	-527.626

Median	PC	$0.56	$0.57	$0.59	$0.56	$0.56	$0.58	$0.58
Median	TC	$0.71	$0.79	$0.69	$0.70	$0.77	$0.74	$0.74

FF	PC	$13.67	$14.75	$13.51	$13.76	$14.84	$13.41	$13.41
FF	TC	$10.84	$10.65	$11.54	$10.99	$10.83	$10.42	$10.42
SD	PC	$21.36	$19.51	$18.77	$20.58	$19.94	$18.55	$18.55
SD	TC	$16.94	$14.09	$16.03	$16.43	$14.55	$14.42	$14.42
VR	PC	$3.14	$3.70	$3.05	$2.57	$2.64	$4.17	$4.17
VR	TC	$2.49	$2.68	$2.60	$2.06	$1.92	$3.24	$3.24

Issue #3 – issue 1 + issue 2

Mistake 3

Issue #3

Solution # 3

› 1. Asymptotically, if you take a sufficient number of draws, this problem will disappear

› 2. Use, antithetic column permuted sequences

- represents a systematic modification of any other type of sequence and as such, can be applied to PMC or any QMC method
- Therefore, it can be applied to any type of base draw (Halton, Sobol, MLHS, etc.)

$$\begin{bmatrix} d_{6,1} \\ d_{6,2} \\ d_{6,3} \\ d_{6,4} \\ d_{6,5} \\ d_{6,6} \\ d_{6,7} \\ d_{6,8} \end{bmatrix} = \begin{bmatrix} d_6^1 & d_6^2 \\ 1-d_6^1 & d_6^2 \\ d_6^1 & 1-d_6^2 \\ 1-d_6^1 & 1-d_6^2 \\ d_6^2 & d_6^1 \\ 1-d_6^2 & d_6^1 \\ d_6^2 & 1-d_6^1 \\ 1-d_6^2 & 1-d_6^1 \end{bmatrix} = \begin{bmatrix} 0.625 & 0.125 \\ 0.375 & 0.125 \\ 0.625 & 0.875 \\ 0.375 & 0.875 \\ 0.125 & 0.625 \\ 0.875 & 0.625 \\ 0.125 & 0.375 \\ 0.875 & 0.375 \end{bmatrix}$$

Solution # 3 – empirical example

		FFT, SDT, PC, TC, VR		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT		FFT, SDT, PC, TC, VR		VR,TC,PC,SDT,FFT			
		Halton		Halton		Halton		Halton (antithetic)		Halton (antithetic)		Halton (colm. Perm.)		Halton (colm. Perm.)		Halton (colm. Perm.)	
		Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio	Par.	t-ratio
Ref. ASC	Fixed par.	0.831	2.06	0.856	2.11	0.791	1.99	0.792	1.98	0.820	2.04	0.894	2.22	0.894	2.22	0.852	2.16
SP1 ASC	Fixed par.	0.440	1.52	0.414	1.43	0.430	1.50	0.403	1.39	0.393	1.35	0.378	1.35	0.378	1.35	0.386	1.37
FF time (N)	Mean	-0.128	-4.34	-0.141	-4.55	-0.133	-4.47	-0.128	-4.35	-0.139	-4.48	-0.129	-4.44	-0.129	-4.44	-0.136	-4.52
	Std dev.	0.155	3.85	0.157	3.88	0.147	3.79	0.153	3.84	0.149	3.81	0.146	3.97	0.146	3.97	0.127	3.72
SD time (N)	Mean	-0.199	-5.55	-0.186	-5.62	-0.185	-5.58	-0.191	-5.55	-0.186	-5.56	-0.178	-5.59	-0.178	-5.59	-0.186	-5.71
	Std dev.	0.109	3.09	0.109	3.07	0.091	2.74	0.106	2.99	0.107	3.08	0.112	3.20	0.112	3.20	0.094	3.39
Petrol cost (L)	Mean	-0.581	-3.25	-0.558	-3.10	-0.526	-2.94	-0.584	-3.25	-0.579	-3.18	-0.553	-3.17	-0.553	-3.17	-0.634	-3.62
	Std dev.	0.448	1.30	0.492	1.45	0.410	1.23	0.396	1.21	0.507	1.63	0.343	0.93	0.343	0.93	0.408	1.32
Toll cost (L)	Mean	-0.349	-2.43	-0.233	-1.49	-0.368	-2.63	-0.359	-2.52	-0.263	-1.68	-0.301	-2.02	-0.301	-2.02	-0.498	-4.10
	Std dev.	0.832	6.46	0.877	6.75	0.964	7.36	0.815	5.99	0.869	6.09	0.825	6.53	0.825	6.53	0.568	5.45
Trvl time var. (N)	Mean	-0.029	-1.83	-0.035	-2.17	-0.030	-1.89	-0.024	-1.53	-0.025	-1.57	-0.040	-2.39	-0.040	-2.39	-0.030	-1.91
	Std dev.	0.117	5.33	0.123	5.40	0.119	5.33	0.126	5.47	0.122	5.39	0.142	5.60	0.142	5.60	0.114	5.64

Draws	100,000	3840		3840 (Base: 120 Halton)		3840 (Base: 32 Halton)		3840 (Base 1 Halton)
LL(β)	-526.2388	-526.649	-531.582	-525.383	-528.809	-527.626	-527.626	-527.787

Median	PC	$0.56	$0.57	$0.59	$0.56	$0.56	$0.58	$0.58	$0.53
Median	TC	$0.71	$0.79	$0.69	$0.70	$0.77	$0.74	$0.74	$0.61

FF	PC	$13.67	$14.75	$13.51	$13.76	$14.84	$13.41	$13.41	$15.41
FF	TC	$10.84	$10.65	$11.54	$10.99	$10.83	$10.42	$10.42	$13.45
SD	PC	$21.36	$19.51	$18.77	$20.58	$19.94	$18.55	$18.55	$21.03
SD	TC	$16.94	$14.09	$16.03	$16.43	$14.55	$14.42	$14.42	$18.36
VR	PC	$3.14	$3.70	$3.05	$2.57	$2.64	$4.17	$4.17	$3.40
VR	TC	$2.49	$2.68	$2.60	$2.06	$1.92	$3.24	$3.24	$2.97

Conclusions

Mistake 3

THE UNIVERSITY OF SYDNEY

THE UNIVERSITY OF SYDNEY

Future research

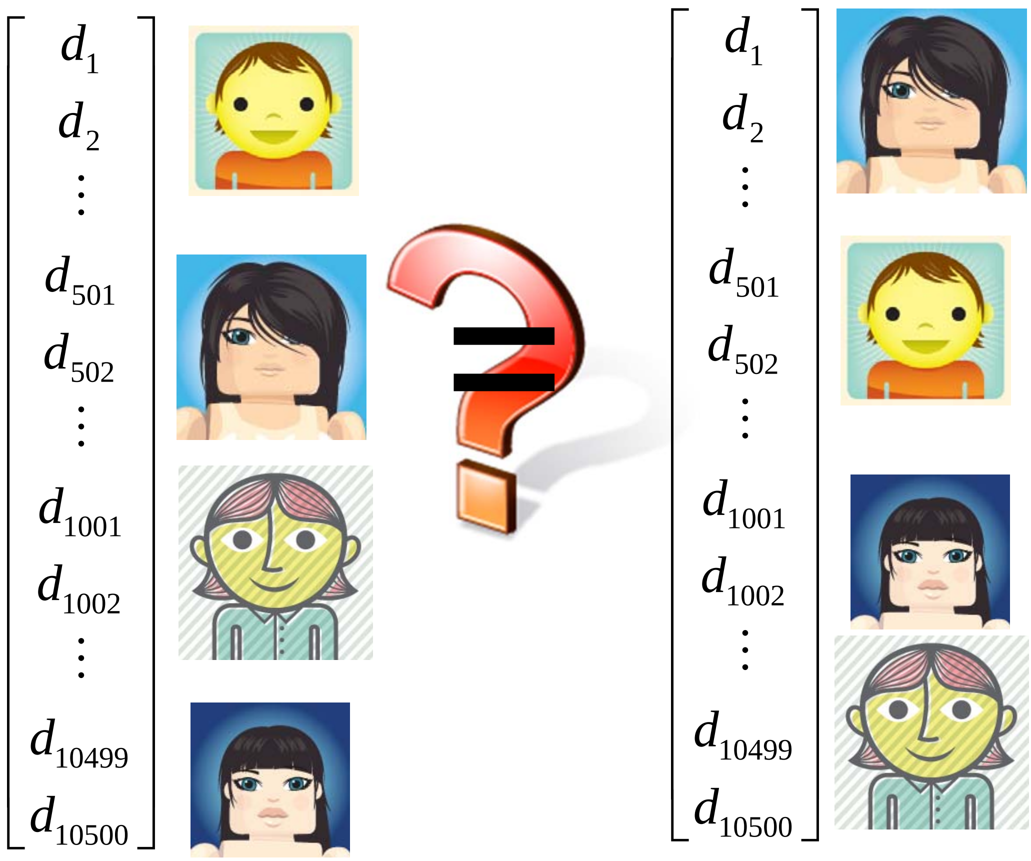
Limitations and what is next

- Likely that the majority of studies reported within the literature represent SML maxima that suffer from the issues identified within this paper.
  - most papers use between 100 and 1,000 draws
- Whilst the results here do not show large differences in SML values, significant differences in other data sets have been found
  - One person in this audience has differences of 38 units in SML just by changing the order the random parameters were estimated in
  - This creates issues with tests based on SML values such as -2 LL ratio tests, AIC, BIC, etc.
- The draws discussed herein explode exponentially as the number of random parameter estimates increase, and hence may not be practical in practice.
- They also don’t address issue number 4

Future research

Limitations and what is next

› Issue # 4



Mistake #4

Thinking linearly in a non-linear world

THE UNIVERSITY OF SYDNEY

What is the relationship between orthogonality and the model outputs

- Orthogonality is about the correlation structure of the data (design)
- What does it mean for the model outputs?
- In linear regression models, the variance-covariance matrix is given as

$$\frac{\sigma^2}{(X'X)}$$ where X is the data or design

Orthogonality and the AVC matrix I

Design					
S	A	B	C	D	E
1	1	0	1	1	-1
2	1	1	-1	-1	-1
3	0	-1	1	-1	0
4	-1	-1	0	0	-1
5	0	1	1	0	1
6	1	-1	-1	1	1
7	-1	1	0	1	0
8	-1	0	-1	0	0
9	0	0	0	-1	1

	Correlation Matrix				
	A	B	C	D	E
A	1	0	0	0	0
B	0	1	0	0	0
C	0	0	1	0	0
D	0	0	0	1	0
E	0	0	0	0	1

	X'X				
	A	B	C	D	E
A	6	0	0	0	0
B	0	6	0	0	0
C	0	0	6	0	0
D	0	0	0	6	0
E	0	0	0	0	6

	(X'X)-1				
	A	B	C	D	E
A	0.166667	0	0	0	0
B	0	0.166667	0	0	0
C	0	0	0.166667	0	0
D	0	0	0	0.166667	0
E	0	0	0	0	0.166667

Orthogonality and the AVC matrix II

Design					
S	**A**	**B**	**C**	**D**	**E**
1	1	0	1	1	-1
2	0	-1	-1	0	-1
3	1	1	0	1	0
4	-1	-1	1	0	-1
5	0	1	1	0	1
6	1	-1	-1	1	1
7	-1	1	0	1	0
8	-1	0	-1	0	0
9	0	0	0	-1	1

Correlation Matrix					
	A	**B**	**C**	**D**	**E**
A	**1**	0	0	0.408248	0.166667
B	0	**1**	0.333333	0.204124	0.333333
C	0	0.333333	**1**	0	-0.16667
D	0.408248	0.204124	0	**1**	-0.20412
E	0.166667	0.333333	-0.16667	-0.20412	**1**

X'X					
	A	**B**	**C**	**D**	**E**
A	**6**	0	0	2	1
B	0	**6**	2	1	2
C	0	2	**6**	0	-1
D	2	1	0	**5**	-1
E	1	2	-1	-1	**6**

(X'X)-1					
	A	**B**	**C**	**D**	**E**
A	**0.218023**	0.056202	-0.03198	-0.11434	-0.07946
B	0.056202	**0.264858**	-0.11047	-0.10207	-0.13307
C	-0.03198	-0.11047	**0.218023**	0.052326	0.087209
D	-0.11434	-0.10207	0.052326	**0.288114**	0.109819
E	-0.07946	-0.13307	0.087209	0.109819	**0.257106**

Orthogonality and the AVC matrix: interaction effects II

Design

S	A	B	C	AB
1	1	0	1	0
2	1	1	-1	1
3	0	-1	1	0
4	-1	-1	0	1
5	0	1	1	0
6	1	-1	-1	-1
7	-1	1	0	-1
8	-1	0	-1	0
9	0	0	0	0

Correlation Matrix

	A	B	C	AB
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
AB	0	0	0	1

X'X

	A	B	C	AB
A	6	0	0	0
B	0	6	0	0
C	0	0	6	0
AB	0	0	0	4

(X'X)-1

	A	B	C	AB
A	0.166667	0	0	0
B	0	0.166667	0	0
C	0	0	0.166667	0
AB	0	0	0	0.25

Orthogonality and the AVC matrix: interaction effects II

Design

S	A	B	C	AB
1	1	1	1	1
2	1	0	1	0
3	0	-1	-1	0
4	0	-1	0	0
5	-1	1	1	-1
6	1	-1	-1	-1
7	-1	1	0	-1
8	-1	0	-1	0
9	0	0	0	0

Correlation Matrix

	A	B	C	AB
A	1	-0.33333	0.166667	0.433013
B	-0.33333	1	0.666667	0
C	0.166667	0.666667	1	0.216506
AB	0.433013	0	0.216506	1

X'X

	A	B	C	AB
A	6	-2	1	2
B	-2	6	4	0
C	1	4	6	1
AB	2	0	1	4

(X'X)-1

	A	B	C	AB
A	**0.308333**	0.225	-0.18333	-0.10833
B	0.225	**0.475**	-0.35	-0.025
C	-0.18333	-0.35	**0.433333**	-0.01667
AB	-0.10833	-0.025	-0.01667	**0.308333**

Loss of orthogonality

Mistake 4

Loss of orthogonality

Plans are made for failure...

› Many things can and do go wrong when it comes to attempts to maintain orthogonality in stated choice data

Non-evenly spaced attribute levels

› A common error is to have unevenly spaced attribute levels in the design

Levels:
A – {5,10,12}
B – {1,2,4}
C – {0,2,3}

	A	B	C
1	0	0	0
2	0	1	1
3	0	2	2
4	1	0	1
5	1	1	2
6	1	2	0
7	2	0	2
8	2	1	0
9	2	2	1

correlation matrix

	A	B	C
A	1		
B	0	1	
C	0	0	1

	A	B	C
1	5	1	0
2	5	2	2
3	5	4	0
4	10	1	3
5	10	2	0
6	10	4	3
7	12	1	2
8	12	2	3
9	12	4	2

correlation matrix

	A	B	C
A	1		
B	0	1	
C	0.57	0	1

Loss of orthogonality

Missing data I

› The most common problem however occurs when choice tasks are missing from the data set, e.g., due to non-response

	A	B	C
1	0	0	0
2	0	1	1
3	0	2	2
4	1	0	1
5	1	1	2
6	1	2	0
7	2	0	2
8	2	1	0
9	2	2	1

correlation matrix

	A	B	C
A	1		
B	0	1	
C	0	0	1

	A	B	C
1	0	0	0
2	0	1	1
3	0	2	2
4	1	0	1
5	1	1	2
6	1	2	0
7	2	0	2
8	2	1	0
9	2	2	1

missing (rows 8 and 9)

correlation matrix

	A	B	C
A	1		
B	-0.32	1	
C	0.32	0.03	1

Loss of orthogonality

Missing data II

› The most common problem however occurs when choice tasks are missing from the data set, e.g., due to missing blocks

blocked design

	A	B	C	D
1	0	0	0	0
2	0	1	0	1
3	1	0	1	0
4	1	1	1	1
5	0	0	1	1
6	0	1	1	0
7	1	0	0	1
8	1	1	0	0

correlation matrix

	A	B	C	D
A	1			
B	0	1		
C	0	0	1	
D	0	0	0	1

respondent 1, respondent 2, respondent 3

		A	B	C	D
respondent 1	1	0	0	0	0
	4	1	1	1	1
	6	0	1	1	0
	7	1	0	0	1
respondent 2	2	0	1	0	1
	3	1	0	1	0
	5	0	0	1	1
	8	1	1	0	0
respondent 3	1	0	0	0	0
	4	1	1	1	1
	6	0	1	1	0
	7	1	0	0	1

correlation matrix

	A	B	C	D
A	1			
B	0	1		
C	0	0.33	1	
D	0.33	0	0	1

Dummy coding

linear:

$$U_1 = \beta_1 + \beta_2 A$$
$$U_2 = \beta_3 B$$

	A	B
1	2	2
2	3	1
3	1	2
4	3	3
5	1	0
6	0	3
7	2	0
8	0	1

correlation matrix:

	A	B
A	1	
B	0	1

nonlinear:

$$U_1 = \beta_1 + \beta_2^1 A_1 + \beta_2^2 A_2 + \beta_2^3 A_3$$
$$U_2 = \beta_3^1 B_1 + \beta_3^2 B_2 + \beta_3^3 B_3$$

	A1	A2	A3	B1	B2	B3
1	0	0	1	0	0	1
2	0	0	0	0	1	0
3	0	1	0	0	0	1
4	0	0	0	0	0	0
5	0	1	0	1	0	0
6	1	0	0	0	0	0
7	0	0	1	1	0	0
8	1	0	0	0	1	0

dummy coding:

0	1	0	0
1	0	1	0
2	0	0	1
3	0	0	0

correlation matrix:

	A1	A2	A3	B1	B2	B3
A1	1					
A2	-0.3	1				
A3	-0.3	-0.3	1			
B1	-0.3	0.3	0.3	1		
B2	0.3	-0.3	-0.3	-0.3	1	
B3	-0.3	0.3	0.3	-0.3	-0.3	1

Loss of orthogonality

Effects coding

$$U_1 = \beta_1 + \beta_2 A$$
$$U_2 = \beta_3 B$$

$$U_1 = \beta_1 + \beta_2^1 A_1 + \beta_2^2 A_2 + \beta_2^3 A_3$$
$$U_2 = \beta_3^1 B_1 + \beta_3^2 B_2 + \beta_3^3 B_3$$

	A	B
1	2	2
2	3	1
3	1	2
4	3	3
5	1	0
6	0	3
7	2	0
8	0	1

	A1	A2	A3	B1	B2	B3
1	0	0	1	0	0	1
2	-1	-1	-1	0	1	0
3	0	1	0	0	0	1
4	-1	-1	-1	-1	-1	-1
5	0	1	0	1	0	0
6	1	0	0	-1	-1	-1
7	0	0	1	1	0	0
8	1	0	0	0	1	0

effects coding:

0	1	0	0
1	0	1	0
2	0	0	1
3	-1	-1	-1

correlation matrix:

	A	B
A	1	
B	0	1

correlation matrix:

	A1	A2	A3	B1	B2	B3
A1	1					
A2	0.5	1				
A3	0.5	0.5	1			
B1	0	0.5	0.5	1		
B2	0	0	0	0.5	1	
B3	0	0.5	0.5	0.5	0.5	1

Orthogonal coding

$$U_1 = \beta_1 + \beta_2 A$$
$$U_2 = \beta_3 B$$

	A	B
1	2	2
2	3	1
3	1	2
4	3	3
5	1	0
6	0	3
7	2	0
8	0	1

correlation matrix:

	A	B
A	1	
B	0	1

$$U_1 = \beta_1 + \beta_2^1 A_1 + \beta_2^2 A_2 + \beta_2^3 A_3$$
$$U_2 = \beta_3^1 B_1 + \beta_3^2 B_2 + \beta_3^3 B_3$$

	A1	A2	A3	B1	B2	B3
1	1	-1	-3	1	-1	-3
2	3	1	1	-1	-1	3
3	-1	-1	3	1	-1	-3
4	3	1	1	3	1	1
5	-1	-1	3	-3	1	-1
6	-3	1	-1	3	1	1
7	1	-1	-3	-3	1	-1
8	-3	1	-1	-1	-1	3

orthogonal coding:

0	-3	1	-1
1	-1	-1	3
2	1	-1	-3
3	3	1	1

correlation matrix:

	A1	A2	A3	B1	B2	B3
A1	1					
A2	0	1				
A3	0	0	1			
B1	0	0	0	1		
B2	0	0	0	0	1	
B3	0	0	0	0	0	1

Asymptotic variance covariance matrix

MNL model AVC

› Example: MNL model with generic parameters (McFadden, 1974)

$$I_N(\beta \mid X) = -\sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} X_{jk_1sn} P_{jsn}(X \mid \beta)\left(X_{jk_2sn} - \sum_{i=1}^{J} X_{ik_2sn} P_{isn}(X \mid \beta) \right)$$

Assuming that all responds observe the same choice situations,

$$\begin{aligned} I_N(\beta \mid X) &= -N \cdot \sum_{s=1}^{S}\sum_{j=1}^{J} X_{jk_1s} P_{js}(X \mid \beta)\left(X_{jk_2s} - \sum_{i=1}^{J} X_{ik_2s} P_{is}(X \mid \beta) \right) \\ &= N \cdot I_1(\beta \mid X) \end{aligned}$$

Therefore, the AVC matrix becomes:

$$\begin{aligned} \Omega_N(\beta \mid X) &= I_N^{-1}(\beta \mid X) \\ &= \tfrac{1}{N} \cdot I_1^{-1}(\beta \mid X) \end{aligned}$$