Model estimation and interpretation

Stephane Hess
stephane.hess@gmail.com

Model estimation and interpretation

Outline

1. Maximum likelihood estimation
2. Example of model specification and estimation
3. Interpreting outputs: model fit
4. Interpreting outputs: parameter estimates
5. Interpreting outputs: covariance matrix
6. Model comparison

Maximum likelihood estimation

Maximum likelihood estimation

Estimation of model parameters

- Observe explanatory data, say x
  - general notation, so x can also include characteristics of decision-maker (z) and choice situation (w)
- Observe choices, say Y
  - Y_{nt} is the chosen alternative for person n in choice situation t
- Specify utility functions and select model type
- Find parameter values β that best explain the choices

Likelihood and log-likelihood

- N people, T_n observations for n
- Y_{nt} chosen by n in situation t
- $y_{jnt} = \begin{cases} 1 & \text{if } Y_{nt} = j \\ 0 & \text{if } Y_{nt} \neq j \end{cases}$
- β groups together all model parameters
- Likelihood given by $L(\beta)$
- Even with modest N and J, $L(\beta) \rightarrow 0$
- Instead work with log-likelihood $LL(\beta)$

$$L(\beta) = \prod_{n=1}^{N} \prod_{t=1}^{T_n} \prod_{j=1}^{J} \left(P_{jnt}(\beta, x_{nt})\right)^{y_{jnt}}$$

$$\begin{aligned} LL(\beta) &= log\left(L(\beta)\right) \\ &= \sum_{n=1}^{N} \sum_{t=1}^{T_n} \sum_{j=1}^{J} y_{jnt} \cdot log\left(P_{jnt}(\beta, x_{nt})\right) \end{aligned}$$

$$\begin{aligned} \widehat{\beta} &= \arg\max_{\beta} L(\beta) \\ &= \arg\max_{\beta} LL(\beta) \end{aligned}$$

Maximum likelihood estimation

Implementation

Theoretical notation

- Sum over probabilities of all alternatives, but only one has a non-zero y_{jnt}

$$LL(\beta) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} \sum_{j=1}^{J} y_{jnt} \cdot log\left(P_{jnt}\left(\beta, x_{nt}\right)\right)$$

Implementation

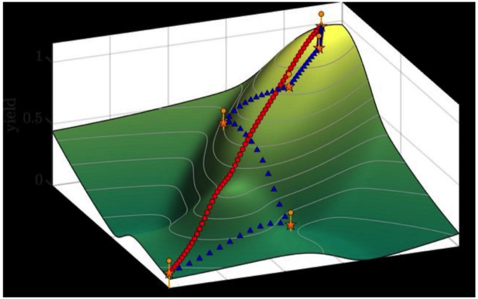
- Only need probability of chosen alternative for each observation, say j_{nt}^{*}

$$LL(\beta) = \sum_{n=1}^{N} \sum_{t=1}^{T_n} log\left(P_{j_{nt}^{*}}\left(\beta, x_{nt}\right)\right)$$

Maximum likelihood estimation

Global optimum

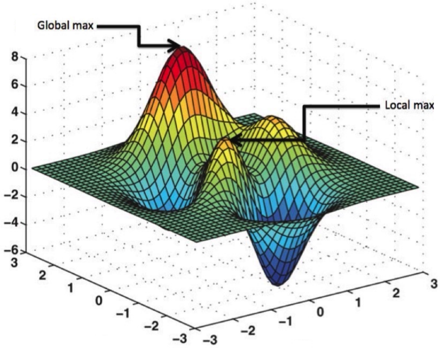
- $LL(\beta)$ for linear in parameters MNL is globally concave
- If a solution exists, it is unique



Maximum likelihood estimation

Local optima

- In more advanced models or with utility specifications that are not linear in parameters, no longer have a single global optimum
- Numerous local optima
- Starting in a *bad* location may get us trapped in one of these local optima
- Closed form choice probabilities



Maximum likelihood estimation

Aim of estimation

- Maximise LL in relation to β, find maximum likelihood estimate (MLE) of β
- At MLE, we have:

$$\frac{\partial LL\left(\beta\right)}{\partial\beta}=0$$

Maximum likelihood estimation

Additional material

Overview of process

- Starting values β_0
- $\widehat{\beta} = \arg\max_{\beta} LL(\beta)$ estimate (MLE) of β, we have: $\frac{\partial LL(\beta)}{\partial \beta} = 0$
- Question is how we get from our starting values β_0 to our MLE $\hat{\beta}$
- Need to know:
  - in which direction to move at each iteration
  - step size, i.e. how far to move
  - convergence criterion

Maximum likelihood estimation

Additional material

Gradient and Hessian

- Gradient: $g_s = \left(\frac{\partial LL(\beta)}{\partial \beta}\right)_{\beta_s}$
- Hessian: $H_s = \left(\frac{\partial^2 LL(\beta)}{\partial \beta \partial \beta'}\right)_{\beta_s}$
- At MLE: $\frac{\partial LL(\beta)}{\partial \beta} = 0$
- In practice, we use stopping criterion
- Convergence reached when: $g_s' \left(-H_s^{-1}\right) g_s < c$ where c is e.g. 10^{-6}
  - rarely changed

Example of model specification and estimation

Example of model specification and estimation

Mode choice data

- Data included with *Apollo*
- Mode choice dataset, with four alternatives (car, bus, air, rail)
- Travel time and cost for all, plus access time for non-car options
- Also a service quality attribute for air and rail (1=no frills, 2=wifi, 3=food)
- 14 stated preference choices
- 500 people, giving us $7,000$ SP observations
- Not all modes available for all individuals
- Respondent characteristics: income, gender, purpose (business *vs* leisure)

Example of model specification and estimation

MNL model specification

				
in vehicle time (TT, minutes)	275	330	80	120
cost (TC, £)	50	35	65	45
access time (AT, minutes)	-	20	55	5
service quality	-	-	wifi	no_frills

- Simple linear in attributes model, with dummy coding for service quality attribute

$$
\begin{aligned}
V_{rail} = {} & \delta_{rail} \\
& + \beta_{tt} \cdot TT_{rail} + \beta_{tc} \cdot TC_{rail} + \beta_{access} \cdot AT_{rail} \\
& + \beta_{no\_Frills} \cdot NO\_FRILLS_{rail} + \beta_{wifi} \cdot WIFI_{rail} + \beta_{food} \cdot FOOD_{rail}
\end{aligned}
$$

Example of model specification and estimation

MNL estimation.xlsx - input data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
9	ID	SP_task	av_car	av_bus	av_air	av_rail	time_car	cost_car	time_bus	cost_bus	access_bus	time_air	cost_air	access_air	service_air	time_rail	cost_rail	access_rail	service_rail	choice
10	1	1	0	0	1	1	0	0	0	0	0	50	50	55	3	170	35	5	2	4
11	1	2	0	0	1	1	0	0	0	0	0	90	65	45	1	120	75	5	3	4
12	1	3	0	0	1	1	0	0	0	0	0	70	110	40	1	155	75	25	2	4
13	1	4	0	0	1	1	0	0	0	0	0	90	80	40	1	170	35	25	2	4
14	1	5	0	0	1	1	0	0	0	0	0	90	80	35	2	130	75	25	2	3
15	1	6	0	0	1	1	0	0	0	0	0	50	95	45	3	170	35	10	2	4
16	1	7	0	0	1	1	0	0	0	0	0	60	110	40	2	120	75	15	3	4
17	1	8	0	0	1	1	0	0	0	0	0	50	95	55	3	140	55	10	1	4
18	1	9	0	0	1	1	0	0	0	0	0	70	110	40	1	140	45	25	3	4
19	1	10	0	0	1	1	0	0	0	0	0	80	65	55	3	120	35	5	1	4
20	1	11	0	0	1	1	0	0	0	0	0	60	80	55	1	130	75	15	3	4
21	1	12	0	0	1	1	0	0	0	0	0	70	110	35	2	170	35	25	3	4
22	1	13	0	0	1	1	0	0	0	0	0	90	65	45	1	140	65	15	2	4
23	1	14	0	0	1	1	0	0	0	0	0	80	50	50	1	140	55	10	3	4
24	2	1	1	1	0	1	300	40	330	25	15	0	0	0	0	170	35	25	2	4
25	2	2	1	1	0	1	390	35	390	35	5	0	0	0	0	140	45	20	1	4
26	2	3	1	1	0	1	345	35	390	20	25	0	0	0	0	130	65	5	2	4
27	2	4	1	1	0	1	275	40	360	25	25	0	0	0	0	140	55	20	1	4
28	2	5	1	1	0	1	275	45	420	20	15	0	0	0	0	130	35	15	3	1
29	2	6	1	1	0	1	250	50	420	20	20	0	0	0	0	140	55	25	1	1
30	2	7	1	1	0	1	390	30	420	20	20	0	0	0	0	130	75	10	2	1
31	2	8	1	1	0	1	250	35	420	20	15	0	0	0	0	155	55	20	2	4
32	2	9	1	1	0	1	300	35	300	30	20	0	0	0	0	140	45	20	2	4
33	2	10	1	1	0	1	390	30	390	20	10	0	0	0	0	130	75	10	1	2
34	2	11	1	1	0	1	390	35	390	15	5	0	0	0	0	140	75	20	1	1
35	2	12	1	1	0	1	275	50	360	15	5	0	0	0	0	155	35	20	1	4
36	2	13	1	1	0	1	300	40	390	20	20	0	0	0	0	170	45	5	1	1
37	2	14	1	1	0	1	275	50	420	15	5	0	0	0	0	120	75	5	3	1

Example of model specification and estimation

MNL estimation.xlsx - utilities, probabilities, likelihood

Parameter	delta_bus	delta_air	delta_rail	b_time	b_cost	b_access	b_wifi	b_food
Value	0	0	0	0	0	0	0	0

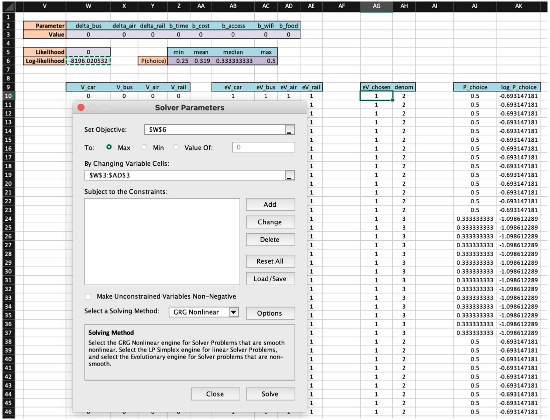
Likelihood	0
Log-likelihood	-8196.020532

	min	mean	median	max
P(choice)	0.25	0.319	0.333333333	0.5

Row	V_car	V_bus	V_air	V_rail	eV_car	eV_bus	eV_air	eV_rail	eV_chosen	denom	P_choice	log_P_choice
10	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
11	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
12	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
13	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
14	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
15	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
16	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
17	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
18	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
19	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
20	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
21	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
22	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
23	0	0	0	0	1	1	1	1	1	2	0.5	-0.693147181
24	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
25	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
26	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
27	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
28	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
29	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
30	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
31	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
32	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
33	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
34	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
35	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
36	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289
37	0	0	0	0	1	1	1	1	1	3	0.333333333	-1.098612289

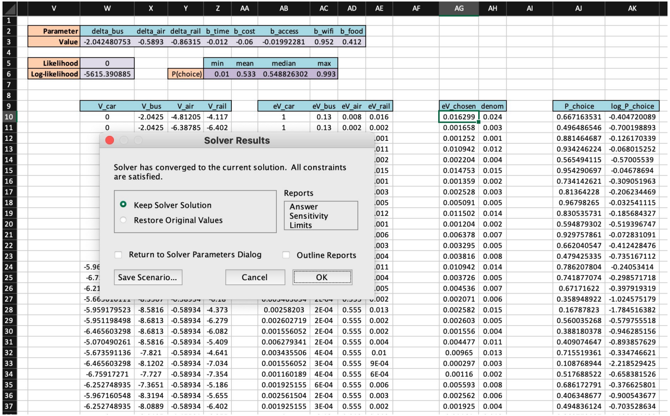
Example of model specification and estimation

MNL estimation.xlsx - using Solver



Example of model specification and estimation

MNL estimation.xlsx - solution



Parameter	delta_bus	delta_air	delta_rail	b_time	b_cost	b_access	b_wifi	b_food
Value	-2.042480753	-0.5893	-0.86315	-0.012	-0.06	-0.01992281	0.952	0.412

Likelihood	0
Log-likelihood	-5615.390885

	min	mean	median	max
P(choice)	0.01	0.533	0.548826302	0.993

Row	V_car	V_bus	V_air	V_rail	eV_car	eV_bus	eV_air	eV_rail	eV_chosen	denom	P_choice	log_P_choice
10	0	-2.0425	-4.81205	-4.117	1	0.13	0.008	0.016	0.016299	0.024	0.667163531	-0.404720089
11	0	-2.0425	-6.38785	-6.402	1	0.13	0.002	0.002	0.001658	0.003	0.496486546	-0.700198893
12								.001	0.001252	0.001	0.881464687	-0.126170339
13								.011	0.010942	0.012	0.934246224	-0.068015252
14								.002	0.002204	0.004	0.565494115	-0.57005539
15								.015	0.014753	0.015	0.954290697	-0.04678694
16								.001	0.001359	0.002	0.734142621	-0.309051963
17								.003	0.002528	0.003	0.81364228	-0.206234469
18								.005	0.005091	0.005	0.96798265	-0.032541115
19								.012	0.011502	0.014	0.830535731	-0.185684327
20								.001	0.001204	0.002	0.594879302	-0.519396747
21								.006	0.006378	0.007	0.929757861	-0.072831091
22								.003	0.003295	0.005	0.662040547	-0.412428476
23								.004	0.003816	0.008	0.479425335	-0.735167112
24	-5.96							.011	0.010942	0.014	0.786207804	-0.24053414
25	-6.7							.004	0.003726	0.005	0.741877074	-0.298571718
26	-6.21							.005	0.004536	0.007	0.67171622	-0.397919319
27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.002	0.002071	0.006	0.358948922	-1.024575179
28	-5.959179523	-8.5816	-0.58934	-4.373	0.00258203	2E-04	0.555	0.013	0.002582	0.015	0.16787823	-1.784516382
29	-5.951198498	-8.6813	-0.58934	-6.279	0.002602719	2E-04	0.555	0.002	0.002603	0.005	0.560035268	-0.579755518
30	-6.465603298	-8.6813	-0.58934	-6.082	0.001556052	2E-04	0.555	0.002	0.001556	0.004	0.388180378	-0.946285156
31	-5.070490261	-8.5816	-0.58934	-5.409	0.006279341	2E-04	0.555	0.004	0.004477	0.011	0.409074647	-0.893857629
32	-5.673591136	-7.821	-0.58934	-4.641	0.003435506	4E-04	0.555	0.01	0.00965	0.013	0.715519361	-0.334746621
33	-6.465603298	-8.1202	-0.58934	-7.034	0.001556052	3E-04	0.555	9E-04	0.000297	0.003	0.108768944	-2.218529425
34	-6.75917271	-7.727	-0.58934	-7.354	0.001160189	4E-04	0.555	6E-04	0.00116	0.002	0.517688522	-0.658381526
35	-6.252748935	-7.3651	-0.58934	-5.186	0.001925155	6E-04	0.555	0.006	0.005593	0.008	0.686172791	-0.376625801
36	-5.967160548	-8.3194	-0.58934	-5.655	0.002561504	2E-04	0.555	0.003	0.002562	0.006	0.406348677	-0.900543677
37	-6.252748935	-8.0889	-0.58934	-6.402	0.001925155	3E-04	0.555	0.002	0.001925	0.004	0.494836124	-0.703528634

Example of model specification and estimation

MNL estimation is straightforward

- Same results with *Apollo* and Excel

Parameter	delta_bus	delta_air	delta_rail	b_time	b_cost	b_access	b_wifi	b_food
Value	-2.042480753	-0.5893	-0.86315	-0.012	-0.06	-0.01992281	0.952	0.412

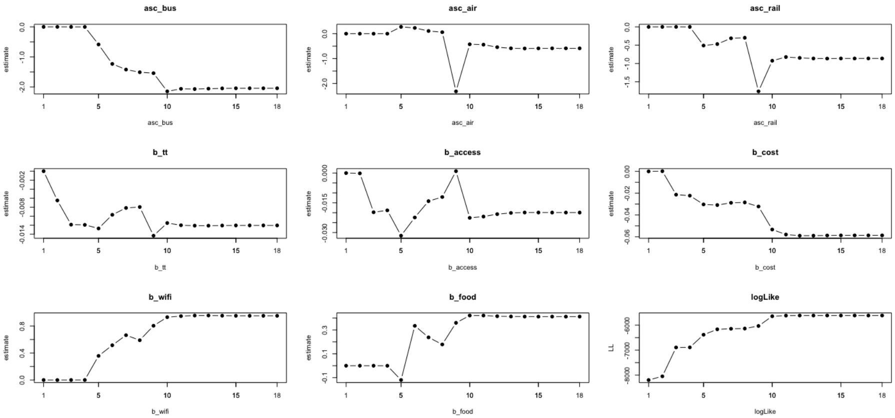
Likelihood	0
Log-likelihood	-5615.390885

	min	mean	median	max
P(choice)	0.01	0.533	0.548826302	0.993

```
LL(start)                               : -8196.02
LL at equal shares, LL(0)               : -8196.02
LL at observed shares, LL(C)            : -6706.94
LL(final)                               : -5615.39
Rho-squared vs equal shares              :  0.3149
Adj.Rho-squared vs equal shares          :  0.3139
Rho-squared vs observed shares           :  0.1627
Adj.Rho-squared vs observed shares       :  0.1616
AIC                                     :  11246.78
BIC                                     :  11301.61

Estimated parameters                    :  8
Time taken (hh:mm:ss)                   :  00:00:2.74
     pre-estimation                     :  00:00:0.61
     estimation                         :  00:00:0.74
     post-estimation                    :  00:00:1.39
Iterations                              :  19
Min abs eigenvalue of Hessian           :  23.53595

Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
              Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car        0.00000          NA          NA          NA            NA
asc_bus       -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air       -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail      -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt          -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access      -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost        -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills    0.00000          NA          NA          NA            NA
b_wifi         0.95151    0.052893      17.989    0.055165        17.248
b_food         0.41168    0.052141       7.895    0.052807         7.796
```

Example of model specification and estimation

Path towards convergence



Example of model specification and estimation

Outputs from estimation

- Three key outputs
  - model fit
  - parameter estimates
  - covariance matrix (giving standard errors and t-ratios)

Apollo outputs for MNL model

```
LL(start)                             : -8196.02
LL at equal shares, LL(0)             : -8196.02
LL at observed shares, LL(C)          : -6706.94
LL(final)                             : -5615.39
Rho-squared vs equal shares            :  0.3149
Adj.Rho-squared vs equal shares        :  0.3139
Rho-squared vs observed shares         :  0.1627
Adj.Rho-squared vs observed shares     :  0.1616
AIC                                   :  11246.78
BIC                                   :  11301.61

Estimated parameters                  :  8
Time taken (hh:mm:ss)                 :  00:00:2.74
     pre-estimation                   :  00:00:0.61
```

```
Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
               Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car         0.00000          NA          NA          NA            NA
asc_bus        -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air        -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail       -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt           -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access       -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost         -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills     0.00000          NA          NA          NA            NA
b_wifi          0.95151    0.052893      17.989    0.055165        17.248
b_food          0.41168    0.052141       7.895    0.052807         7.796
```

Interpreting outputs: model fit

Interpreting outputs: model fit

Model fit measures

```
LL(start)                                  : -8196.02
LL at equal shares, LL(0)                  : -8196.02
LL at observed shares, LL(C)               : -6706.94
LL(final)                                  : -5615.39
Rho-squared vs equal shares                 :  0.3149
Adj.Rho-squared vs equal shares             :  0.3139
Rho-squared vs observed shares              :  0.1627
Adj.Rho-squared vs observed shares          :  0.1616
AIC                                        :  11246.78
BIC                                        :  11301.61

Estimated parameters                       :  8
Time taken (hh:mm:ss)                      :  00:00:2.74
     pre-estimation                        :  00:00:0.61
```

Interpreting outputs: model fit

Log-likelihoood at convergence

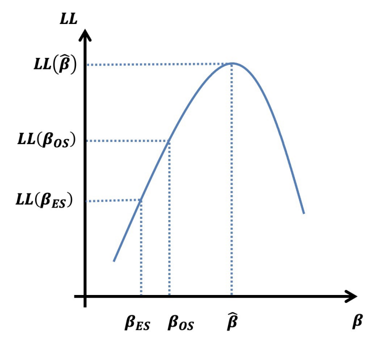
- Likelihood $L(\beta)$ shows how likely choices in our data are
  - conditional on chosen model and at parameter values β
- Classical estimation maximises log-likelihood
  - Log-likelihood $LL(\beta) = log\left[L(\beta)\right]$
- At convergence, obtain parameters values $\hat{\beta}$
- $LL\left(\hat{\beta}\right)$ used extensively during specification search

Interpreting outputs: model fit

Obtain metrics other than just $LL(\hat{\beta})$

- $LL\left(\hat{\beta}\right)$: LL at convergence (MLE)
- $LL\left(\beta_0\right)$: LL at starting values
- $LL\left(\beta_{ES}\right)$: LL at equal shares
  - random model, $P_j = \frac{1}{J}\ \forall j$
  - often written as $LL\left(0\right)$
  - often same as $LL\left(\beta_0\right)$
- $LL\left(\beta_{OS}\right)$: LL at observed shares
  - replicates aggregate shares in the data
  - helps understand how much other parameters contribute to understanding choices especially in labelled settings

Different LL measures



Interpreting outputs: model fit

Additional material

Information criteria

Akaike Information Criterion (AIC)

$$\boxed{AIC = -2 \cdot LL\left(\hat{\beta}\right) + 2K}$$

K is number of estimated parameters, penalises LL for model complexity

Bayesian Information Criterion (BIC)

$$\boxed{BIC = -2 \cdot LL\left(\hat{\beta}\right) + K \cdot log\left(O\right)}$$

O is number of observations in the sample, i.e. $O = \sum_{n=1}^{N} T_n$

- Penalty term of BIC is more stringent than for AIC, as, with $N \geqslant 8$, $K \cdot log\left(O\right) > 2K$
- BIC tends to favour models with smaller K compared to AIC

Interpreting outputs: model fit

Model performance

- In regression, evaluate performance using e.g. R^2
- In choice modelling, focus is more on relative performance
  - how well does one model on a dataset do compared to another
- Or look at out-of-sample validation
  - compare fit on validation data to estimation data (often $20\% - 80\%$ split)
  - checks for overfitting
  - limited insight if validation data from same sample

Interpreting outputs: model fit

Goodness of fit

- LL, AIC and BIC depend on sample size and choice set size
- ρ^2 is one option to compare performance across models/data
- Can calculate against equal shares or observed shares
- Should use adjusted ρ^2 ($\bar{\rho}^2$) to accounts for number of parameters

$$\bar{\rho}^2(ES) = 1 - \frac{LL(\hat{\beta}) - K}{LL(\beta_{ES})}$$

$$\bar{\rho}^2(OS) = 1 - \frac{LL(\hat{\beta}) - K + J - 1}{LL(\beta_{OS})}$$

(K is number of parameters, J is number of alternatives)

Interpreting outputs: model fit

ρ^2 and R^2

- ρ^2 also referred to as pseudo-R^2
- Many people see 0.3 as being a good value, but depends on data
- Very high ρ^2 can also indicate data problems

  Key reference: *Mokhtarian, P.L. (2016), Discrete choice models' ρ^2: A reintroduction to an old friend, Journal of Choice Modelling, 21, pp. 60-65*

Relationship between $\rho 2$ and R^2

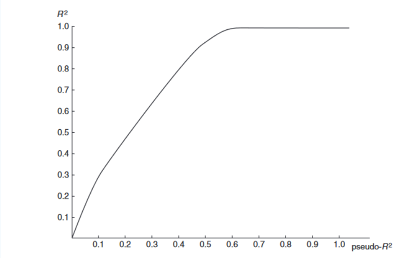


Figure reference: *Domencich, T.A. & McFadden, D. (1975), Urban Travel Demand: A Behavioral Analysis, North-Holland Publishing Co.*

Interpreting outputs: model fit

Hit rate (and why note use it)

Assigns choice to highest P_j

		Model 1		Model 2	
Person	**Choice**	P_A	P_B	P_A	P_B
1	A	0.8	0.2	0.55	0.45
2	A	0.75	0.25	0.53	0.47
3	B	0.3	0.7	0.48	0.52
4	A	0.85	0.15	0.54	0.46
5	B	0.25	0.75	0.49	0.51
6	B	0.2	0.8	0.49	0.51
7	B	0.6	0.4	0.44	0.56
8	A	0.45	0.55	0.54	0.46
9	A	0.85	0.15	0.51	0.49
10	A	0.35	0.65	0.55	0.45
	LL	**-4.47**		**-6.32**	
	Av prob choice	**0.67**		**0.53**	
	Hit rate	**0.7**		**1**	

Why is this a bad idea?

"This statistic incorporates a notion that is opposed to the meaning of probabilities and the purpose of specifying choice probabilities. The statistic is based on the idea that the decision maker is predicted by the researcher to choose the alternative for which the model gives the highest probability."

(Kenneth Train, Discrete Choice Methods with Simulation, Cambridge University Press)

Interpreting outputs: parameter estimates

Interpreting outputs: parameter estimates

Model outputs

- Model estimation returns maximum likelihood estimates, or MLE, given by $\hat{\beta}$
- These are the estimates that give us the log-likelihood at convergence
- Parameter types
  1. parameters capturing impact of changes in an attribute on utility
  2. parameters relating to model structure (e.g. nesting parameters)
  3. parameters capturing socio-demographic interactions
- Our focus for now is on the first of these

Interpreting outputs: parameter estimates

Illustrative example

- Can get initial insights from signs
  - Each £ in cost loses us 0.005 units in utility
  - Each GB in memory gains us 0.004 units in utility
  - Moving from 3G to 4G and 5G increases utility
  - Highest utility for Apple ahead of Samsung & Huawei
- But what about the size of the estimates?

Mobile phone choice

Parameters	Estimate ($\hat{\beta}_k$)
$\beta_{cost,£}$	-0.005
$\beta_{memory,GB}$	0.004
β_{3G}	0
β_{4G}	0.5
β_{5G}	0.75
β_{Huawei}	0
$\beta_{Samsung}$	1
β_{Apple}	1.75

Interpreting outputs: parameter estimates

Scale matters

- Different studies have different levels of noise
- Greater noise means smaller β, and vice versa
- We thus cannot say that cost matters more in our study than in a study where $\beta_{cost,£} = -0.002$

Mobile phone choice

Parameters	Estimate ($\hat{\beta}_k$)
$\beta_{cost,£}$	-0.005
$\beta_{memory,GB}$	0.004
β_{3G}	0
β_{4G}	0.5
β_{5G}	0.75
β_{Huawei}	0
$\beta_{Samsung}$	1
β_{Apple}	1.75

Interpreting outputs: parameter estimates

Units, levels and ranges

- Cannot say cost is more important than memory
  - with continuous attributes, units matter
- Cannot say that Apple is better than 5G
  - with categorical attributes, levels and ranges matter
  - remember that only differences in utility matter, in this case differences against the base
- Also incorrect to say that brand is *more important* than download speed
  - can at best say that with the specific levels used here, brand can influence choice more than download speed

Mobile phone choice

Parameters	Estimate ($\hat{\beta}_k$)
$\beta_{cost,£}$	-0.005
$\beta_{memory,GB}$	0.004
β_{3G}	0
β_{4G}	0.5
β_{5G}	0.75
β_{Huawei}	0
$\beta_{Samsung}$	1
β_{Apple}	1.75

Interpreting outputs: parameter estimates

Estimation results

- Initial interpretation:
  - Each min of in vehicle time loses us 0.01205 units in utility
  - Each min of access time loses us 0.01992 units in utility
  - Each £ in cost loses us 0.0587 units in utility
  - Moving from no frills to wifi gains us 0.9515 units in utility
  - Moving from no frills to food gains us 0.41168 units in utility
- Care required:
  - Units matter
  - Scale might differ across studies
- So how can we interpret these results?

```
Estimates:
              Estimate
asc_car        0.00000
asc_bus       -2.04288
asc_air       -0.58780
asc_rail      -0.86198
b_tt          -0.01205
b_access      -0.01992
b_cost        -0.05870
b_no_frills    0.00000
b_wifi         0.95150
b_food         0.41168
```

Interpreting outputs: parameter estimates

Marginal rate of substitution (MRS)

- Absolute values of estimates have no meaning
- Can only look at relative impacts on utility of changes in attributes
  - given by partial derivatives
- Let us start by looking at continuous attributes with a linear specification
  - $\frac{\partial V_j}{\partial x_{j,k}} = \beta_k$ is impact of a one unit change in attribute $x_{j,k}$
- MRS represents relative impact on utility of unit changes in two attributes
  - e.g. attributes $x_{j,k}$ and $x_{j,m}$

$$\boxed{MRS_{x_{j,k},x_{j,m}} = \frac{\partial V_j}{\partial x_{j,k}} \Big/ \frac{\partial V_j}{\partial x_{j,m}} = \frac{\beta_k}{\beta_m}}$$

- MRS: what increase (decr.) in $x_{j,k}$ cancels out impact of decrease (incr.) in $x_{j,m}$

Interpreting outputs: parameter estimates

MRS for in vehicle travel time (TT) and access time (AT)

- We have a linear in attributes specification:
  - $\frac{\partial V_j}{\partial TT_j} = \beta_{tt} = -0.01205$
  - $\frac{\partial V_j}{\partial AT_j} = \beta_{access} = -0.01992$
  - $MRS_{TT_j,AT_j} = \frac{\partial V_j}{\partial TT_j} \Big/ \frac{\partial V_j}{\partial AT_j} = \frac{\beta_{tt}}{\beta_{access}} = \frac{-0.01205}{-0.01992} = 0.6049$
- Each minute of in vehicle time is equivalent of 0.6049 minutes in access time
- What does this mean?

```
Estimates:
              Estimate
asc_car        0.00000
asc_bus       -2.04288
asc_air       -0.58780
asc_rail      -0.86198
b_tt          -0.01205
b_access      -0.01992
b_cost        -0.05870
b_no_frills    0.00000
b_wifi         0.95150
b_food         0.41168
```

Interpreting outputs: parameter estimates

MRS for in vehicle travel time (TT) and access time (AT)

- MRS shows changes in **opposite** direction that cancel each other out in terms of utility impact
  - Inherent feature of compensatory models
- In our case:
  - if we reduce in vehicle time by one minute
    - can increase access time by around 36 seconds
  - if we increase access time by 1 minute
    - need to reduce in vehicle time by 1.65 minutes ($\frac{1}{0.6049}$)

Estimates:

	Estimate
asc_car	0.00000
asc_bus	-2.04288
asc_air	-0.58780
asc_rail	-0.86198
b_tt	-0.01205
b_access	-0.01992
b_cost	-0.05870
b_no_frills	0.00000
b_wifi	0.95150
b_food	0.41168

Interpreting outputs: parameter estimates

Willingness-to-pay (WTP)

- MRS where the denominator is a monetary attribute
- Let us say $x_{j,c}$ is the cost attribute for alternative j

$$\boxed{WTP_{x_{j,k},x_{j,c}} = \frac{\partial V_j}{\partial x_{j,k}} \Big/ \frac{\partial V_j}{\partial x_{j,c}} = \frac{\beta_k}{\beta_c}}$$

- Represents monetary value of changes in attribute k
- Can relate to both:
  - willingness-to-pay (WTP) for an improvement in an attribute
  - willingness-to-accept (WTA) a worse value in return for lower cost
- Also need to consider whether the monetary attribute is a cost attribute or a compensation/savings attribute

Interpreting outputs: parameter estimates

WTP for change in travel time (TT)

- Most common transport MRS: value of travel time (VTT)
- MRS for in vehicle time and travel cost
  - $\frac{\beta_{tt}}{\beta_{cost}} = \frac{-0.01205}{-0.0587} = 0.2053$
- One minute in travel time has a monetary value of £0.2053
- Often expressed per hour:
  - $60 \cdot £0.2053 = £12.32$

Estimates:

	Estimate
asc_car	0.00000
asc_bus	-2.04288
asc_air	-0.58780
asc_rail	-0.86198
b_tt	-0.01205
b_access	-0.01992
b_cost	-0.05870
b_no_frills	0.00000
b_wifi	0.95150
b_food	0.41168

Interpreting outputs: parameter estimates

WTP interpretation

- Obtained monetary value of £12.32/hr of travel time
- Remember that MRS relates to changes in opposite direction for the two attributes
- A reduction in travel time by one hour can compensate for an increase in cost by £12.32
  - Willingness to pay (WTP)
- A reduction in cost by £12.32 can compensate for an increase in travel time by one hour
  - Willingness to accept (WTA)

```
Estimates:
               Estimate
asc_car         0.00000
asc_bus        -2.04288
asc_air        -0.58780
asc_rail       -0.86198
b_tt           -0.01205
b_access       -0.01992
b_cost         -0.05870
b_no_frills     0.00000
b_wifi          0.95150
b_food          0.41168
```

Interpreting outputs: parameter estimates

Categorical attributes

- Can only look at changes in utility between levels, and compare to other attributes
- Let us say $x_{j,q}$ and $x_{j,r}$ are categorical variables
  - Relative impact (second to third level for q vs first to second for r)

$$\boxed{\frac{\beta_{q,\mathbf{3}}-\beta_{q,\mathbf{2}}}{\beta_{r,\mathbf{2}}-\beta_{r,\mathbf{1}}}}$$

- Can of course combine with continuous attributes too
  - MRS (relative to 1 unit in k): $\frac{\beta_{q,\mathbf{3}}-\beta_{q,\mathbf{2}}}{\beta_k}$
  - WTP: $\frac{\beta_{q,\mathbf{3}}-\beta_{q,\mathbf{2}}}{\beta_c}$ if $x_{j,c}$ is the cost attribute

Interpreting outputs: parameter estimates

Results for our example

- WTP for changes in the quality of service attribute
- From no frills to wifi
  - $\frac{\beta_{wifi}-\beta_{no\_frills}}{\beta_{cost}} = \frac{0.95150}{-0.0587} = -£16.21$
- From no frills to food
  - $\frac{\beta_{food}-\beta_{no\_frills}}{\beta_{cost}} = \frac{0.41168}{-0.0587} = -£7.01$
- From food to wifi
  - $\frac{\beta_{wifi}-\beta_{food}}{\beta_{cost}} = \frac{0.95150-0.41158}{-0.0587} = \frac{0.53992}{-0.0587} = -£9.20$

Question

Why are these negative?

```
Estimates:
             Estimate
asc_car       0.00000
asc_bus      -2.04288
asc_air      -0.58780
asc_rail     -0.86198
b_tt         -0.01205
b_access     -0.01992
b_cost       -0.05870
b_no_frills   0.00000
b_wifi        0.95150
b_food        0.41168
```

Interpreting outputs: parameter estimates

Interpretation

- Meaning of an MRS helps to understand the sign
- We saw
  - $\frac{\beta_{wifi}-\beta_{no\_frills}}{\beta_{cost}} = \frac{0.95150}{-0.0587} = -£16.21$
- Going from no frills to wifi has same impact as a **reduction** in cost by £16.21
- Means a WTP for wifi (compared to no frills) of £16.21
- Or a WTA of a journey with no frills compared to wifi in return for a reduction in cost by £16.21

```
Estimates:
              Estimate
asc_car        0.00000
asc_bus       -2.04288
asc_air       -0.58780
asc_rail      -0.86198
b_tt          -0.01205
b_access      -0.01992
b_cost        -0.05870
b_no_frills    0.00000
b_wifi         0.95150
b_food         0.41168
```

Interpreting outputs: parameter estimates

MRS with non-linear utility functions

Linear in attributes

- $V_{in} = \cdots + \beta_{tt}\ TT_{in} + \beta_{cost}\ C_{in} + \dots$
- $VTT = \frac{\beta_{tt}}{\beta_{cost}}$

Logarithmic

- $V_{in} = \cdots + \beta_{log\_tt}\ log(TT_{in}) + \beta_{log\_cost}\ log(C_{in}) + \dots$
- $\frac{\partial V_i n}{\partial TT_{in}} = \frac{\beta_{log\_tt}}{TT_{in}}\ \&\ \frac{\partial V_i n}{\partial C_{in}} = \frac{\beta_{log\_cost}}{C_{in}} \Rightarrow VTT = \frac{\beta_{log\_tt}}{\beta_{log\_cost}} \cdot \frac{C_{in}}{TT_{in}}$

Polynomial

- $V_{in} = \cdots + \beta_{tt}\ TT_{in} + \beta_{tt2}\ TT_{in}^2 + \beta_{cost}\ C_{in} + \beta_{cost2}\ C_{in}^2 + \dots$
- $VTT = \frac{\beta_{tt} + \beta_{tt2}\ T_{in}}{\beta_{cost} + \beta_{cost2}\ C_{in}}$

Interpreting outputs: parameter estimates

MRS with non-linear utility functions

- With non-linear specifications, MRS becomes context-dependent
- Non-linearities in preferences are commonly observed
- But implications for policy work are difficult
  - what value should VTT be calculated at?

Interpreting outputs: parameter estimates

Log transform for our example

- Log transform applied to travel time and cost
- Implies that each additional minute in time or £ in cost matters less than the one before
  - Decreasing marginal utilities
- Impacts trade-offs: $VTT = \frac{\beta_T}{\beta_C} \cdot \frac{C_{in}}{T_{in}}$
- Example at 30 mins and £20
  - $VTT = \frac{\beta_{log\_tt}}{\beta_{log\_cost}} \frac{C_{in}}{TT_{in}} = \frac{-1.93451}{-.83805} \cdot \frac{20}{30} = £0.4544/min$
  - VTT per hour: $60 \cdot £0.4544 = £27.27/hr$

```
Estimates:
              Estimate
asc_car        0.00000
asc_bus       -3.15708
asc_air       -0.96113
asc_rail      -0.30838
b_log_tt      -1.93451
b_access      -0.01639
b_log_cost    -2.83805
b_no_frills    0.00000
b_wifi         0.72844
b_food         0.29205
```

```
VTT at different time/cost levels (in £/hr)
          £20   £30   £40   £50   £60
30 mins 27.27 40.90 54.53 68.16 81.80
40 mins 20.45 30.67 40.90 51.12 61.35
50 mins 16.36 24.54 32.72 40.90 49.08
60 mins 13.63 20.45 27.27 34.08 40.90
```

Interpreting outputs: parameter estimates

MRS/WTP calculations: a common confusion

- Positive MRS if impact of both attributes has same sign
  - i.e. if both attributes are desirable, or if both attributes are undesirable
- With oppositely signed impacts, MRS is negative
- Makes sense: trade-off relates to changes of an attribute in one direction cancelling out changes of an attribute in the other direction

- Trade-off between time and cost is positive
  - Reduced time to accept increased cost (and v.v.)
- Trade-off between reliability and cost is negative
  - Increased reliability to accept increased cost (and v.v.)

Sign of MRS

		Monetary attribute: Cost	Monetary attribute: Savings
Other attribute	Desirable	MRS<0	MRS>0
	Undesirable	MRS>0	MRS<0

Interpreting outputs: parameter estimates

Discussion

- MRS calculations are key output for interpretation of random utility models
- Used universally across disciplines (unlike some other metrics)
- Computation easy in simple models, but becomes more complex with
  - non-linearity
  - heterogeneity
- Point-of-indifference in utility does not mean probabilities stay the same in all models
- Meaning of MRS different for non-RUM models (context dependent)

WTP space: introduction

- Can reparameterise model in WTP space as opposed to utility/preference space
- By rescaling marginal utilities of (some) non-cost attributes by cost coefficient
- Means that WTP is directly estimated, rather than needing to be calculated as a ratio of partial derivatives
- Not a different model, simply a reparameterisation of the utility function
- For fixed parameter models, the two specifications are mathematically equivalent

Interpreting outputs: parameter estimates

Understanding reparameterisation of utilities

- Simple example where T is time in minutes, and C is cost in £

Specification 1

$$V_{ni} = \dots + \beta_T\, T_{ni} + \beta_C\, C_{ni} + \dots$$

Marginal utilities wrt one minute in T and £1 in C given by β_T and β_C

Specification 2

$$V_{ni} = \dots + \frac{1}{60} \cdot \beta'_T \cdot T_{ni} + 100 \cdot \beta'_C \cdot C_{ni} + \dots$$

Marginal utilities wrt one min in T and £1 in C given by $\frac{1}{60} \cdot \beta'_T$ and $100 \cdot \beta'_C$

- Models mathematically equivalent, so we have that $\beta'_T = 60 \cdot \beta_T$ (i.e., expressed in hours) and $\beta'_C = \frac{1}{100} \cdot \beta_C$ (i.e., expressed in pence)
- Can use the same rationale to reparameterise the actual numeraire

Interpreting outputs: parameter estimates

WTP space: implementation

- Rescale marginal utilities of (some) non-cost attributes by cost coefficient

Utility space

$$V_{ni} = \dots + \beta_T \, T_{ni} + \beta_C \, C_{ni} + \dots$$

$$\frac{\partial V_{ni}}{\partial T_{ni}} = \beta_T \;\&\; \frac{\partial V_{ni}}{\partial C_{ni}} = \beta_C$$

$$\frac{\partial V_{ni}}{\partial T_{ni}} \Big/ \frac{\partial V_{ni}}{\partial C_{ni}} = \frac{\beta_T}{\beta_C}$$

WTP space

$$V_{ni} = \dots + \beta'_C \, \beta'_{VT} \, T_{ni} + \beta'_C \, C_{ni} + \dots$$

$$\frac{\partial V_{ni}}{\partial T_{ni}} = \beta'_C \, \beta'_{VT} \;\&\; \frac{\partial V_{ni}}{\partial C_{ni}} = \beta'_C$$

$$\frac{\partial V_{ni}}{\partial T_{ni}} \Big/ \frac{\partial V_{ni}}{\partial C_{ni}} = \beta'_{VT}$$

- With fixed coefficients, this is a simple rescaling, and the two models are thus mathematically equivalent, with $\beta_T = \beta'_C \, \beta'_{VT}$, $\beta_C = \beta'_C$, and thus $\frac{\beta_T}{\beta_C} = \beta'_{VT}$

Key reference: *Train, K. & Weeks, M. (2006). Discrete Choice Models in Preference Space and Willingness-to Pay Space, in Environmental Resource Economics, A. Alberini and R. Scarpa, eds., Springer*

Interpreting outputs: parameter estimates

Models in preference space and WTP space

- Mathematically equivalent

```
LL(start)                                : -8196.02
LL at equal shares, LL(0)                : -8196.02
LL at observed shares, LL(C)             : -6706.94
LL(final)                                : -5615.39
Rho-squared vs equal shares               :  0.3149
Adj.Rho-squared vs equal shares           :  0.3139
Rho-squared vs observed shares            :  0.1627
Adj.Rho-squared vs observed shares        :  0.1616
AIC                                      :  11246.78
BIC                                      :  11301.61

Estimated parameters                     :  8
Time taken (hh:mm:ss)                    :  00:00:2.74
     pre-estimation                      :  00:00:0.61
```

```
Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
               Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car         0.00000          NA          NA          NA            NA
asc_bus        -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air        -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail       -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt           -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access       -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost         -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills     0.00000          NA          NA          NA            NA
b_wifi          0.95151    0.052893      17.989    0.055165        17.248
b_food          0.41168    0.052141       7.895    0.052807         7.796
```

```
LL(start)                                : -8196.02
LL at equal shares, LL(0)                : -8196.02
LL at observed shares, LL(C)             : -6706.94
LL(final)                                : -5615.39
Rho-squared vs equal shares               :  0.3149
Adj.Rho-squared vs equal shares           :  0.3139
Rho-squared vs observed shares            :  0.1627
Adj.Rho-squared vs observed shares        :  0.1616
AIC                                      :  11246.78
BIC                                      :  11301.61
```

```
Estimates:
               Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car         0.00000          NA          NA          NA            NA
asc_bus        -2.04289    0.075132     -27.191    0.092220       -22.152
asc_air        -0.58784    0.180223      -3.262    0.197274        -2.980
asc_rail       -0.86200    0.107217      -8.040    0.117824        -7.316
v_tt            0.20533    0.008783      23.379    0.009523        21.563
v_access        0.33932    0.042442       7.995    0.042270         8.027
b_cost         -0.05870    0.001463     -40.118    0.001680       -34.951
v_no_frills     0.00000          NA          NA          NA            NA
v_wifi        -16.20831    0.896311     -18.083    1.003304       -16.155
v_food          7.01268    0.881982       7.951    0.894944         7.836
```

Interpreting outputs: parameter estimates

Same findings for MRS, and same standard errors

- Understand the reason for the signs of MRS?

```
Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
              Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car        0.00000          NA          NA          NA            NA
asc_bus       -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air       -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail      -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt          -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access      -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost        -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills    0.00000          NA          NA          NA            NA
b_wifi         0.95151    0.052893      17.989    0.055165        17.248
b_food         0.41168    0.052141       7.895    0.052807         7.796
```

```
> apollo_deltaMethod(model,
+                    deltaMethod_settings = list(
+                      expression=c(VTT="b_tt/b_cost",
+                                   VAT="b_access/b_cost",
+                                   VWIFI="(b_wifi-b_no_frills)/b_cost",
+                                   VFOOD="(b_food-b_no_frills)/b_cost")))
The expression VWIFI includes parameters that were fixed in estimation: b_no_frills
These have been replaced by their fixed values, giving: (b_wifi-0)/b_cost

The expression VFOOD includes parameters that were fixed in estimation: b_no_frills
These have been replaced by their fixed values, giving: (b_food-0)/b_cost

Running Delta method computation for user-defined function:

 Expression    Value Robust s.e. Rob t-ratio (0)
        VTT   0.2053      0.0095           21.56
        VAT   0.3393      0.0423            8.03
      VWIFI -16.2084      1.0033          -16.15
      VFOOD  -7.0127      0.8949           -7.84
```

```
Estimates:
              Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car        0.00000          NA          NA          NA            NA
asc_bus       -2.04289    0.075132     -27.191    0.092220       -22.152
asc_air       -0.58784    0.180223      -3.262    0.197274        -2.980
asc_rail      -0.86200    0.107217      -8.040    0.117824        -7.316
v_tt           0.20533    0.008783      23.379    0.009523        21.563
v_access       0.33932    0.042442       7.995    0.042270         8.027
b_cost        -0.05870    0.001463     -40.118    0.001680       -34.951
v_no_frills    0.00000          NA          NA          NA            NA
v_wifi       -16.20831    0.896311     -18.083    1.003304       -16.155
v_food         7.01268    0.881982       7.951    0.894944         7.836
```

Interpreting outputs: parameter estimates

WTP space: what is the benefit?

- With simple models, it is of course straightforward to calculate WTP as ratios of partial derivatives
- Calculation of standard errors using Delta method is not too hard either
- And this avoids the need for a non-linear utility specification, which can lead to convergence problems
- Benefits of WTP space arise specifically in the context of models with random heterogeneity, where the ratio of marginal utilities may not be well defined

WTP space: common confusions

Confusion 1 WTP space is a different model

Truth No, WTP space is not a model, it is a different parameterisation of the utility function

Confusion 2 Papers stating that the WTP space model fits better than utility space, or vice versa

Truth This only happens as a result of different distributional assumptions in mixed logit, not because one specification is superior to another

Interpreting outputs: parameter estimates

Additional material

WTP space: common confusions

Confusion 3 Some people say that cost coefficient in WTP space is constrained to 1

Truth This confusion arises from $V_i = \mu \left(\delta_{wtp,i} + \sum_k^K \beta_{wtp,k} x_{k,i} + cost_i \right)$, where k are all non-cost attributes. But $\mu = \beta_{cost}$, everything in the utility is multiplied by the cost coefficient

Confusion 4 All parameters and ASC need to be included in scaling by β_{cost}

Truth No, up to the user which parameters to express in monetary terms, and these 4 specifications are all equivalent, and final 3 are all WTP space

$$V_i = \delta_i + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_{cost} cost_i$$

$$V_i = \beta_{cost} \left(\delta_{wtp,i} + \beta_{wtp,1} x_{1,i} + \beta_{wtp,2} x_{2,i} + cost_i \right)$$

$$V_i = \delta_i + \beta_{cost} \left(\beta_{wtp,1} x_{1,i} + \beta_{wtp,2} x_{2,i} + cost_i \right)$$

$$V_i = \delta_i + \beta_2 x_{2,i} + \beta_{cost} \left(\beta_{wtp,1} x_{1,i} + cost_i \right)$$

Additional material

WTP space: common confusions

Confusion 5 Non-cost attributes should have a negative sign in front of them

Truth A user may find this convenient, but it is not a requirement - only implies that WTP is for increases in attribute (remember the earlier point)

$$V_i = \beta_{cost}\left(\delta_{wtp,i} - \beta_{wtp,1}x_{1,i} - \beta_{wtp,2}x_{2,i} + cost_i\right)$$

Confusion 6 Only cost can be used for scaling

Truth A model can be parameterised in any other valuation space, using whatever attribute the user would use as the denominator in MRS

Interpreting outputs: covariance matrix

Interpreting outputs: covariance matrix

Covariance matrix and standard errors

- Estimation gives us estimates ($\hat{\beta}$) and asymptotic variance-covariance matrix Ω
- Our key interest is in standard errors (σ)
  - given by square root of diagonal elements of covariance matrix
  - expression of precision of parameter estimates
  - a smaller standard error means we can be surer about the estimated value
- We use standard errors for asymptotic confidence intervals and for statistical tests

Covariance matrix and standard errors

$$\Omega = I\left(\beta\right)^{-1}$$

$$I\left(\beta\right) = -E\left(H\left(\beta\right)\right)$$

$$H\left(\beta\right) = \left(\frac{\partial^2 LL\left(\beta\right)}{\partial\beta\partial\beta'}\right)$$

$$\sigma_{\beta_k} = \sqrt{\Omega_{k,k}}$$

Interpreting outputs: covariance matrix

Standard errors and sample size

- β_k is one parameter in model
- True value given by β_k^*
- Maximum likelihood estimate (MLE): $\widehat{\beta}_k$
- Incomplete data leads to sampling error
- Asymptotic normality:

$$\boxed{\sqrt{N}\left(\widehat{\beta}-\beta^*\right) \to \mathcal{N}\left(0,\Omega\right)}$$

Interpreting outputs: covariance matrix

Estimates, covariance matrix and standard errors

Estimates:

	Estimate	s.e.	t.rat.(0)
asc_car	0.00000	NA	NA
asc_bus	-2.04288	0.075131	-27.191
asc_air	-0.58781	0.180223	-3.262
asc_rail	-0.86199	0.107216	-8.040
b_tt	-0.01205	5.5356e-04	-21.775
b_access	-0.01992	0.002507	-7.946
b_cost	-0.05870	0.001463	-40.118
b_no_frills	0.00000	NA	NA
b_wifi	0.95151	0.052893	17.989
b_food	0.41168	0.052141	7.895

Classical covariance matrix:

	asc_bus	asc_air	asc_rail	b_tt	b_access	b_cost	b_wifi	b_food
asc_bus	0.005645	8.0367e-04	-3.9811e-04	-1.106e-05	-7.224e-05	1.964e-05	-5.033e-05	8.338e-06
asc_air	8.0367e-04	0.032480	0.017187	6.376e-05	-2.9608e-04	-8.730e-06	-0.002116	-0.001694
asc_rail	-3.9811e-04	0.017187	0.011495	4.812e-05	-9.259e-05	2.006e-05	-0.002069	-0.001724
b_tt	-1.106e-05	6.376e-05	4.812e-05	3.064e-07	4.220e-08	3.187e-07	-6.011e-06	-3.221e-06
b_access	-7.224e-05	-2.9608e-04	-9.259e-05	4.220e-08	6.284e-06	4.757e-07	-5.420e-06	-3.216e-06
b_cost	1.964e-05	-8.730e-06	2.006e-05	3.187e-07	4.757e-07	2.141e-06	-1.825e-05	-1.021e-05
b_wifi	-5.033e-05	-0.002116	-0.002069	-6.011e-06	-5.420e-06	-1.825e-05	0.002798	0.001522
b_food	8.338e-06	-0.001694	-0.001724	-3.221e-06	-3.216e-06	-1.021e-05	0.001522	0.002719

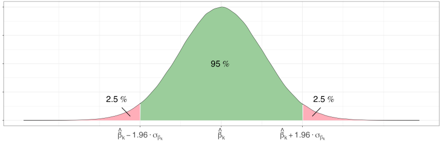
With covariance matrix Ω, have $\sqrt{\Omega_{k,k}} = \sigma_{\beta_k}$

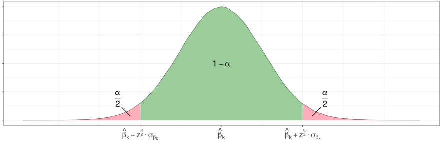
$$e.g.: \sqrt{0.005645} = 0.0751$$

Interpreting outputs: covariance matrix

Confidence intervals

- With new data, estimates would change
- Often report 95% confidence intervals
  - 95% chance that the true value for a parameter lies in that range
  - with smaller standard errors, CIs will be narrower
- Calculation of CIs uses estimated value, standard error, and critical value from a $N(0,1)$ distribution
  - e.g. for 95%, we use $\widehat{\beta}_k \pm 1.96\sigma_{\beta_k}$





CI	α	$\frac{\alpha}{2}$	$z^{\frac{\alpha}{2}}$
99%	0.01	0.005	2.57
95%	0.05	0.025	1.96
90%	0.1	0.05	1.64

Interpreting outputs: covariance matrix

Confidence intervals: example

- 95% confidence interval for cost coefficient β_{cost}:
  - lower 95% CI limit:
    $-0.0587 - 1.96 * 0.001463 = -0.06156748$
  - upper 95% CI limit:
    $-0.0587 + 1.96 * 0.001463 = -0.05583252$

Estimates:

	Estimate	s.e.
asc_car	0.00000	NA
asc_bus	-2.04288	0.075131
asc_air	-0.58781	0.180223
asc_rail	-0.86199	0.107216
b_tt	-0.01205	5.5356e-04
b_access	-0.01992	0.002507
b_cost	-0.05870	0.001463
b_no_frills	0.00000	NA
b_wifi	0.95151	0.052893
b_food	0.41168	0.052141

Interpreting outputs: covariance matrix

Review of statistical hypothesis testing

- Type I error: false positive
  - p-value of 0.05 means we accept a 5% chance of wrongly rejecting H_0
- Type II error: false negatives
  - probability of wrongly failing to reject H_0 is β, relates to power of statistical test
  - especially with small sample sizes, there is a significant risk of wrongly failing to reject H_0

Type I and type II errors

- H_0: null hypothesis
- H_1: alternate hypothesis

		Decision	
		Accept H_0	Reject H_0
Truth	H_0 True		Type I error probability α
	H_0 False	Type II error probability β	

Interpreting outputs: covariance matrix

Statistical hypothesis testing: process

Formulate hypotheses

- H_0: null hypothesis
- H_1: alternate hypothesis

Carry out statistical test

1. Set acceptable value for α
2. Compute value for test statistic
3. Test conclusion (two options)
   - 3.1 Compare test statistic to critical value
   - 3.2 Compute *p*-value for test and compare to α

Remember: *p*-value is probability of result occurring by chance even if H_0 is true

Interpreting outputs: covariance matrix

Statistical tests: t-ratios

- Used (typically) to test whether parameter is different from zero ($H_0 : \beta_k = 0$)
- Test statistic: $t_{\widehat{\beta}_k} = \frac{\widehat{\beta}_k}{\sigma_{\beta_k}}$
- We compare the test statistic to a critical value and see if it exceeds it
  - e.g. 1.96 for a 95% two-sided test, or 1.64 for a one-sided test

confidence level	1 sided critical value	2 sided critical value
99%	2.33	2.58
95%	1.64	1.96
90%	1.28	1.64

```
Estimates:
              Estimate        s.e.   t.rat.(0)
asc_car        0.00000          NA          NA
asc_bus       -2.04288    0.075131     -27.191
asc_air       -0.58781    0.180223      -3.262
asc_rail      -0.86199    0.107216      -8.040
b_tt          -0.01205  5.5356e-04     -21.775
b_access      -0.01992    0.002507      -7.946
b_cost        -0.05870    0.001463     -40.118
b_no_frills    0.00000          NA          NA
b_wifi         0.95151    0.052893      17.989
b_food         0.41168    0.052141       7.895
```

Interpreting outputs: covariance matrix

Statistical tests: t-ratios and p-values

- Can compute p-value (probability that our result was obtained by chance if H_0 is true)
- Should always report either standard error or t-ratio alongside p-value
- And state whether one-sided or two-sided

t-ratio	p-value (1 sided)	p-value (2 sided)
2.58	0.005	0.01
2.33	0.01	0.02
1.96	0.025	0.05
1.64	0.05	0.1
1.28	0.1	0.2

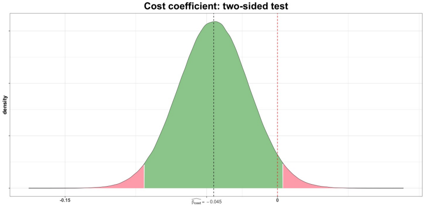
```
Estimates:
                Estimate        s.e.   t.rat.(0)  p(2-sided)
asc_car          0.00000          NA          NA          NA
asc_bus         -2.04288    0.075131     -27.191    0.000000
asc_air         -0.58781    0.180223      -3.262    0.001108
asc_rail        -0.86199    0.107216      -8.040   8.882e-16
b_tt            -0.01205  5.5356e-04     -21.775    0.000000
b_access        -0.01992    0.002507      -7.946   1.998e-15
b_cost          -0.05870    0.001463     -40.118    0.000000
b_no_frills      0.00000          NA          NA          NA
b_wifi           0.95151    0.052893      17.989    0.000000
b_food           0.41168    0.052141       7.895   2.887e-15
```

```
Estimates:
                Estimate        s.e.   t.rat.(0)  p(1-sided)
asc_car          0.00000          NA          NA          NA
asc_bus         -2.04288    0.075131     -27.191       0.000
asc_air         -0.58781    0.180223      -3.262  5.5398e-04
asc_rail        -0.86199    0.107216      -8.040   4.441e-16
b_tt            -0.01205  5.5356e-04     -21.775       0.000
b_access        -0.01992    0.002507      -7.946   9.992e-16
b_cost          -0.05870    0.001463     -40.118       0.000
b_no_frills      0.00000          NA          NA          NA
b_wifi           0.95151    0.052893      17.989       0.000
b_food           0.41168    0.052141       7.895   1.443e-15
```

Interpreting outputs: covariance matrix

One-sided or two-sided tests

- $\widehat{\beta}_{cost} = -0.045$
- $\sigma_{\beta_{cost}} = 0.025$
- $t_{\widehat{\beta}_{cost}} = -1.8$
- Typical critical value: 1.96
- Using 1.96 makes sense in absence of sign assumptions
- But often we know the sign
- Consider using one-sided tests
  - otherwise we would say that very negative values are also unacceptable



Interpreting outputs: covariance matrix

What do statistical tests tell us?

What is the test?

- Common description: test of significance
- Actual meaning: test to see whether the null hypothesis of a parameter being equal to zero can be rejected

What is significance?

- Statistical significance of test relates to probability of H_0 being rejected when it is in fact true (type I error)
- Incorrect to talk about a parameter being 95% significant
- Significance level in this case is 5%, and we can reject H_0 with a 95% level of confidence, not significance

Interpreting outputs: covariance matrix

Recap: classical standard errors

- Classical covariance matrix Ω is a conservative estimate of sampling error
- Relies on number of assumptions
  - model is correct
  - no unmodelled correlation across choices

Classical covariance matrix

$$\Omega = I(\beta)^{-1}$$

$$I(\beta) = -E(H(\beta))$$

$$H(\beta) = \left(\frac{\partial^2 LL(\beta)}{\partial\beta\partial\beta'}\right)$$

Interpreting outputs: covariance matrix

Robust covariance matrix

- Given by sandwich estimator involving the BHHH matrix (B)
- Robust se tend to be larger
- Using LL at person level in BHHH gives different robust standard errors from working at observation level
  - T observations each from N people is not the same as 1 observation each from NT people

Robust covariance matrix

$$\Omega_{robust} = \Omega \, B \, \Omega$$

$$B_{jk} = \sum_{n}^{N} \frac{\partial LL_n(\beta)}{\partial \beta_j} \frac{\partial LL_n(\beta)}{\partial \beta_k}$$

Estimates:

	Estimate	s.e.	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0)
asc_car	0.00000	NA	NA	NA	NA
asc_bus	-2.04288	0.075131	-27.191	0.092220	-22.152
asc_air	-0.58781	0.180223	-3.262	0.197274	-2.980
asc_rail	-0.86199	0.107216	-8.040	0.117824	-7.316
b_tt	-0.01205	5.5356e-04	-21.775	5.9548e-04	-20.242
b_access	-0.01992	0.002507	-7.946	0.002489	-8.003
b_cost	-0.05870	0.001463	-40.118	0.001680	-34.951
b_no_frills	0.00000	NA	NA	NA	NA
b_wifi	0.95151	0.052893	17.989	0.055165	17.248
b_food	0.41168	0.052141	7.895	0.052807	7.796

Interpreting outputs: covariance matrix

Bootstrapping uses fewest assumptions

- Works by repeated sampling
- Approach with least assumptions, and lowest risk of numerical issues
- But computationally most demanding

Bootstrapped covariance matrix

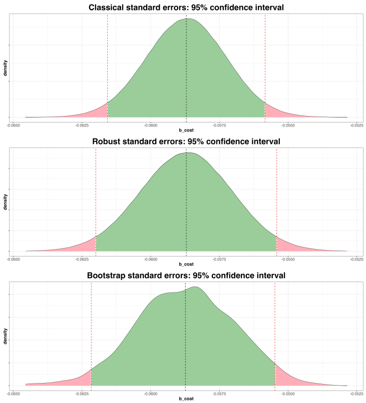
1. Draw S versions of data with replacement
2. S sets of β: $\beta_s = \langle \beta_{1,s}, \ldots, \beta_{K,s} \rangle$, $s = 1, \ldots, S$
3. $\frac{\sum_{s=1}^{S} \beta_{k,s}}{S} \rightarrow \widehat{\beta}_k =$ with large S
4. $\Omega_{bootstrap} = var\left(\beta\right)$

```
Estimates:
              Estimate    Rob.s.e. Rob.t.rat.(0) Bootstrap.s.e. Bootstrap.t.rat.(0)
asc_car        0.00000          NA            NA             NA                  NA
asc_bus       -2.04288    0.092220       -22.152       0.092188             -22.160
asc_air       -0.58781    0.197274        -2.980       0.195013              -3.014
asc_rail      -0.86199    0.117824        -7.316       0.116588              -7.393
b_tt          -0.01205  5.9548e-04       -20.242     6.0600e-04             -19.891
b_access      -0.01992    0.002489        -8.003       0.002432              -8.192
b_cost        -0.05870    0.001680       -34.951       0.001763             -33.294
b_no_frills    0.00000          NA            NA             NA                  NA
b_wifi         0.95151    0.055165        17.248       0.057051              16.678
b_food         0.41168    0.052807         7.796       0.053916               7.636
```

Interpreting outputs: covariance matrix

Confidence interval comparison

- Asymptotic CI for classical and robust
  - symmetrical by definition
  - wider with robust se
- With bootstrapping, can look at empirical CI
  - No longer necessarily symmetrical
  - lower limit in this case 5.1% further from mean than upper limit



Interpreting outputs: covariance matrix

Standard errors for derived measures

- Obtain estimates and standard errors for β
- Key interest is in functions of individual elements of β
  - MRS and WTP
  - difference between two parameters
  - demand forecasts and elasticities
  - welfare measures
  - moments of distributions
  - correlation between randomly distributed coefficients
- Need standard errors for derived quantities

```
Estimates:
        Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
b_tt    -0.05977    0.004257     -14.040    0.006742        -8.865
b_tc    -0.13182    0.013506      -9.760    0.023638        -5.576
b_hw    -0.03745    0.001848     -20.269    0.002317       -16.161
b_ch    -1.15207    0.043419     -26.534    0.061373       -18.772

> diff_tt_hw=-0.05977-(-0.03745)
> diff_tt_hw
[1] -0.02232
> VTT_per_hour=60*(-0.05977/-0.13182)
> VTT_per_hour
[1] 27.20528
```

The Delta method

- Delta method is a first-derivative calculation
- Often described as an approximation
- Shown to be exact rather than an approximation by Daly et al. (2012)

Delta method calculations

- Let Φ be a function of β
- Estimates $\hat{\beta}$ and AVC matrix Ω
- $cov\left(\Phi\right) = \Phi'^{\,T}\Omega\Phi'$
- Φ' gives first derivatives of Φ against β

Key reference: *Daly, A.J., Hess, S. & de Jong, G. (2012), Calculating errors for measures derived from choice modelling estimates, Transportation Research Part B 46(2), pp. 333-341.*

Interpreting outputs: covariance matrix

Examples: difference and ratio

- Difference: $\Phi = \beta_1 - \beta_2$
  - i.e. $\phi'_1 = 1$ and $\phi'_2 = -1$
  - and $var\left(\beta_1 - \beta_2\right) = \omega_{11} + \omega_{22} - 2\,\omega_{12}$
- Ratio: $\Phi = \frac{\beta_1}{\beta_2}$
  - i.e. $\phi'_1 = \frac{1}{\beta_2}$ and $\phi'_2 = -\frac{\beta_1}{\beta_2^2}$
  - and

$$var\left(\frac{\beta_1}{\beta_2}\right) = \left(\frac{\beta_1}{\beta_2}\right)^2\left(\frac{\omega_{11}}{\beta_1^2} + \frac{\omega_{22}}{\beta_2^2} - 2\,\frac{\omega_{12}}{\beta_1\beta_2}\right)$$

```
Estimates:
     Estimate       s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
b_tt  -0.05977   0.004257    -14.040    0.006742        -8.865
b_tc  -0.13182   0.013506     -9.760    0.023638        -5.576
b_hw  -0.03745   0.001848    -20.269    0.002317       -16.161
b_ch  -1.15207   0.043419    -26.534    0.061373       -18.772
```

```
> model$robvarcov
             b_tt         b_tc         b_hw         b_ch
b_tt 4.545565e-05 1.176310e-04 2.920627e-06 1.379524e-04
b_tc 1.176310e-04 5.587382e-04 6.516358e-06 3.154629e-04
b_hw 2.920627e-06 6.516358e-06 5.370123e-06 4.961005e-05
b_ch 1.379524e-04 3.154629e-04 4.961005e-05 3.766619e-03
```

```
> diff_tt_hw=-0.05977-(-0.03745)
> se_diff_tt_hw=sqrt(4.545565e-05+5.370123e-06-2*2.920627e-06)
> diff_tt_hw
[1] -0.02232
> se_diff_tt_hw
[1] 0.00670705
> VTT_per_hour=60*(-0.05977/-0.13182)
> se_VTT_per_hour=sqrt((VTT_per_hour)^2*(4.545565e-05/((-0.05977)^2)+5.58
7382e-04/((-0.13182)^2)-2*1.176310e-04/((-0.05977)*(-0.13182))))
> VTT_per_hour
[1] 27.20528
> se_VTT_per_hour
[1] 3.334052
```

```
> deltaMethod_settings=list(expression=c(diff_tt_hw="b_tt-b_hw",
+                                        VTT_per_hour="60*b_tt/b_tc"))
> apollo_deltaMethod(model, deltaMethod_settings)
Running Delta method computation for user-defined function:

   Expression   Value Robust s.e. Rob t-ratio (0)
   diff_tt_hw -0.0223      0.0067           -3.33
 VTT_per_hour 27.2065      3.3343            8.16
```

Interpreting outputs: covariance matrix

Why am I getting `Inf` or `NaN` for standard errors?

- Theoretical identification issues
  - e.g. missing normalisation for ASCs
- Empirical identification issues
  - e.g. parameters going towards $-\inf$ or $+\inf$, if one group of people never or always chooses a given option
- Calculation of numerical derivatives could lead to some zero probabilities
  - use analytical derivatives, and if not possible, use bootstrapping

```
Estimates:
              Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car        0.73984          NA          NA          NA            NA
asc_bus       -1.30304          NA          NA          NA            NA
asc_air        0.41869          NA          NA          NA            NA
asc_rail       0.14451          NA          NA          NA            NA
b_tt          -0.01205          NA          NA          NA            NA
b_access      -0.01992          NA          NA          NA            NA
b_cost        -0.05870          NA          NA          NA            NA
b_no_frills   -0.26666          NA          NA          NA            NA
b_wifi         0.68484          NA          NA          NA            NA
b_food         0.14502          NA          NA          NA            NA
```

```
Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
              Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car        0.00000          NA          NA          NA            NA
asc_bus       -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air       -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail      -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt          -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access      -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost        -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills    0.00000          NA          NA          NA            NA
b_wifi         0.95151    0.052893      17.989    0.055165        17.248
b_food         0.41168    0.052141       7.895    0.052807         7.796
```

Interpreting outputs: covariance matrix

Additional material

How should significance be reported?

- Minimum output should be estimates and standard errors, or estimates and t-ratios, as s.e. can be calculated from t-ratios
- Common practice in some fields to report estimates and *p*-values only
  - This is bad practice, for two reasons
  - *p*-values imply an analyst decision on whether a one-sided or two-sided test is used, and this is often not reported
  - *p*-values are often reported with a numerical precision that prevents an analyst from recovering standard errors (e.g. $p < 0.001$)
- Even worse is the reliance on $*$ measures in some fields, e.g. using * for 90% confidence, ** for 95% confidence and *** for 99% confidence
  - The same issues apply as for *p*-values, but they are further compounded by the fact that e.g. $***$ could mean a *t*-ratio of 4 or 40
- *p*-values and $*$ measures should never replace s.e. or *t*-ratios

Interpreting outputs: covariance matrix

Additional material

Recommendations

- Wasserstein et al. (2019) conclude *"that it is time to stop using the term 'statistically significant' entirely. Nor should variants such as 'significantly different', 'p<0.05', and 'nonsignificant' survive, whether expressed in words, by asterisks in a table, or in some other way."*
- And *"[analysts should not] believe that an association or effect exists just because it was statistically significant [or] that an association or effect is absent just because it was not statistically significant."*

Wasserstein, R.L., Schirm, A.L., Lazar, N.A. (2019), Moving to a world beyond "p<0.05". The American Statistician 73, 1–19.

Interpreting outputs: covariance matrix

Additional material

Recommendations (continued)

- In health, "*clinical significance*" measures whether a treatment has noticeable effect on health outcomes. Choice modellers may wish to consider "*behavioural significance*", i.e. does a parameter change predictions and "*policy significance*", and does it have a significant impact on outcome of any process using the results
- Finally, note that removing a parameter that is "*not significant*" may have undesirable impact on other parameters
  - useful approximation to say that removal of parameter 1 will change parameter 2 by $-t_1 * \frac{r_{12}}{t_2}$, where t are the respective t-ratios and r_{12} is the correlation

Model comparison

Model comparison

Specification testing strategy

Model building

- Generate a set of different models
- Gradually build up complexity
- Driven by a-priori considerations and intermediate results

Model comparison

- Increased complexity leads to new insights and better fit
- But also estimate additional parameters
- Need to test whether gains justify additional complexity

Should not rely exclusively on mathematical performance to select among competing models

Model comparison

Improving our model

- We return to the mode choice MNL model
- But we now add some journey purpose interactions to our model

Model comparison

Example with purpose interaction

```
LL(start)                                   : -8196.02
LL at equal shares, LL(0)                   : -8196.02
LL at observed shares, LL(C)                : -6706.94
LL(final)                                   : -5615.39
Rho-squared vs equal shares                  :  0.3149
Adj.Rho-squared vs equal shares              :  0.3139
Rho-squared vs observed shares               :  0.1627
Adj.Rho-squared vs observed shares           :  0.1616
AIC                                         :  11246.78
BIC                                         :  11301.61

Estimated parameters                        :  8
Time taken (hh:mm:ss)                       :  00:00:2.74
     pre-estimation                         :  00:00:0.61
```

```
Unconstrained optimisation.

These outputs have had the scaling used in estimation applied to them.
Estimates:
               Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car         0.00000          NA          NA          NA            NA
asc_bus        -2.04288    0.075131     -27.191    0.092220       -22.152
asc_air        -0.58781    0.180223      -3.262    0.197274        -2.980
asc_rail       -0.86199    0.107216      -8.040    0.117824        -7.316
b_tt           -0.01205  5.5356e-04     -21.775  5.9548e-04       -20.242
b_access       -0.01992    0.002507      -7.946    0.002489        -8.003
b_cost         -0.05870    0.001463     -40.118    0.001680       -34.951
b_no_frills     0.00000          NA          NA          NA            NA
b_wifi          0.95151    0.052893      17.989    0.055165        17.248
b_food          0.41168    0.052141       7.895    0.052807         7.796
```

```
LL(start)                                   : -8196.02
LL at equal shares, LL(0)                   : -8196.02
LL at observed shares, LL(C)                : -6706.94
LL(final)                                   : -5085.13
Rho-squared vs equal shares                  :  0.3796
Adj.Rho-squared vs equal shares              :  0.3782
Rho-squared vs observed shares               :  0.2418
Adj.Rho-squared vs observed shares           :  0.2402
AIC                                         :  10192.26
BIC                                         :  10267.65
```

```
Estimates:
                       Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
asc_car                 0.00000          NA          NA          NA            NA
asc_bus                -1.95572    0.079729     -24.530    0.091610       -21.348
shift_bus_business     -2.18638    0.365073      -5.989    0.455484        -4.800
asc_air                -1.51256    0.194741      -7.767    0.203458        -7.434
shift_air_business      2.71303    0.103702      26.162    0.119770        22.652
asc_rail               -1.38040    0.116286     -11.871    0.122307       -11.286
shift_rail_business     1.41871    0.084527      16.784    0.095600        14.840
b_tt                   -0.01347  5.9169e-04     -22.761  6.1258e-04       -21.985
b_access               -0.02140    0.002626      -8.149    0.002578        -8.299
b_cost                 -0.06633    0.001624     -40.840    0.001761       -37.662
b_no_frills             0.00000          NA          NA          NA            NA
b_wifi                  1.06257    0.055861      19.022    0.058752        18.086
b_food                  0.44020    0.054719       8.045    0.056982         7.725
```

Model comparison

Likelihood ratio test

- Situation where one model is nested within a more general version
- H_0 can be expressed as constraints on parameters of the general model
  - e.g. general model uses gender specific coefficients, while constrained model uses generic coefficients
- Compare test statistic against critical value of χ^2 distribution
  - e.g. 3.84 for one additional parameter

LR test

- Two likelihoods: $L\left(\hat{\beta}^{H_1}\right)$ & $L\left(\hat{\beta}^{H_0}\right)$
- Likelihood ratio: $R = \frac{L\left(\hat{\beta}^{H_0}\right)}{L\left(\hat{\beta}^{H_1}\right)}$
- Test statistic: $-2 \cdot log\left(R\right)$
  $\Rightarrow -2\left(LL\left(\hat{\beta}^{H_0}\right) - LL\left(\hat{\beta}^{H_1}\right)\right)$
- $-2 \cdot log\left(R\right) \sim \chi^2_d$, where d is difference in number of parameters

Model comparison

Model with purpose interaction: LR test

```
> apollo_lrTest("MNL_base_SP",model)
                         LL par
MNL_base_SP        -5615.39   8
MNL_purpose_split -5085.13   11
Difference          530.26    3

Likelihood ratio test-value:    1060.52
Degrees of freedom:             3
Likelihood ratio test p-value:  1.337e-229
```

Model comparison

Nested *vs.* non-nested tests

- Nested hypotheses:
  - Restricted model is special case of general model
  - When H_0 is true, general model simplifies to restricted model
    - we can then use a likelihood ratio test
- Non-nested hypotheses:
  - Neither model is a special case of the other
    - model 1: $V_i = \ldots + \beta_k x_{k,i}$
    - model 2: $V_i = \ldots + \beta_k x_{k,i}^2$
- Can compare using goodness of fit criteria that penalise for complexity
  - e.g. adj. ρ^2, AIC, BIC
- Also some formal tests

Model comparison

Test based on adjusted ρ^2

- Well known Ben-Akiva & Swait test is based on adjusted ρ^2

$$\boxed{P(\bar{\rho}_1^2 - \bar{\rho}_2^2 \geqslant z) \leqslant \Phi\left[-\left(2 \cdot z \cdot LL\left(\beta_{ES}\right) + df_1 - df_2\right)^{\frac{1}{2}}\right]}$$

where:

- z is observed difference in adjusted ρ^2
- Φ is cumulative standard normal distribution
- $LL\left(\beta_{ES}\right)$ is log-likelihood at equal shares
- df_1 and df_2 are the number of parameters for models 1 and 2

Key reference: *Ben-Akiva, M. & Swait, J. (1986). The Akaike Likelihood Ratio Index. Transportation Science, 20(2), 133-136.*

Summary

Key points from this class

- Estimation of MNL is straightforward, but becomes harder with more complex models
- Parameters estimates themselves can only be interpreted relative to each other, and only differences matter for categorical variables
- Care is required in interpretation of statistical tests, including reporting

Suggested reading

- Train, K.E. (2009), Discrete Choice Methods with Simulation, Cambridge University Press, free online access https://eml.berkeley.edu/books/choice2.html
  - Chapters 3 and 8


Questions?
Apollo
www.ApolloChoiceModelling.com
The most flexible choice modelling software (up to a probability)