Apollo practical 2

Stephane Hess
stephane.hess@gmail.com

Apollo practical 2

Outline

1. Heterogeneity in δ
2. Screening data for heterogeneity
3. Heterogeneity in β

Heterogeneity in δ

Heterogeneity in δ

Recovery of market shares

- A question is how well the model does at recovering the market shares in the data
- In linear in parameters MNL, market shares are recovered perfectly with a full set of alternative specific constants (ASCs): $\sum_{n=1}^{N} y_{in} = \sum_{n=1}^{N} \widehat{P_n(i)}$
- This is at the sample level, how about at level of subsegments?
  - Say in subgroup g, i.e. compare $\sum_{n=1}^{N_g} y_{in}$ to $\sum_{n=1}^{N_g} \widehat{P_n(i)}$
  - This can tell us if one group is very different from the overall sample
  - Do this in `MNL_modeChoice_SP_mode_specific_time.r`

```
sharesTest_settings=list()
sharesTest_settings[["alternatives"]] = c(car=1, bus=2, air=3, rail=4)
sharesTest_settings[["choiceVar"]]    = database$choice
sharesTest_settings[["subsamples"]] = list(business=(database$business==1),
                                           leisure=(database$business==0),
                                           female=(database$female==1),
                                           male=(database$female==0)
apollo_sharesTest(model,apollo_probabilities,apollo_inputs,sharesTest_settings)
```

Heterogeneity in δ

Recovery of market shares: results

```
Prediction test for group: business (2310 observations)

                                car      bus      air      rail   All
Times chosen (data)         366.000    8.000  771.000 1165.000  2310
Times chosen (prediction)   661.554  123.023  465.595 1059.828  2310
Diff (prediction-data)      295.554  115.023 -305.405 -105.172     0
t-ratio                      16.192   11.240  -19.918   -5.327    NA
p-val                         0.000    0.000    0.000    0.000    NA
...
Prediction test for group: female (3332 observations)

                                car      bus      air      rail   All
Times chosen (data)         918.000  164.000  734.000 1516.000  3332
Times chosen (prediction)   966.045  164.092  706.325 1495.538  3332
Diff (prediction-data)       48.045    0.092  -27.675  -20.462     0
t-ratio                       2.161    0.008   -1.466   -0.852    NA
p-val                         0.031    0.994    0.143    0.394    NA
...
Prediction test for group: All data (7000 observations)

                            car  bus  air  rail   All
Times chosen (data)        1946  358 1522  3174  7000
Times chosen (prediction)  1946  358 1522  3174  7000
Diff (prediction-data)        0    0    0     0     0
```

Heterogeneity in δ

How can we improve our model?

- Use shifts in the ASCs rather than group-specific ASCs
- Means that base ASCs are still estimated with large sample
- Can simply look at t-ratios for shifts to decide on specification
- Define new parameters, and then adapt ASCs just prior to specification of utilities (inside `apollo_probabilities`)

```
...
asc_bus = asc_bus_base + asc_bus_shift_female * female + asc_bus_shift_business * business
...
V[["bus"]]  = asc_bus  + b_tt_bus  * time_bus  + b_access * access_bus  + b_cost * cost_bus
...
```

- Base this on `MNL_modeChoice_SP_mode_specific_time.r`

Heterogeneity in δ

MNL_modeChoice_SP_heterogeneity_asc.r

- Results broadly in line with expectations

```
LL(final)                                  : -5064.83

Estimates:
                                Estimate      s.e.   t.rat.(0)  Rob.s.e. Rob.t.rat.(0)
asc_car                         0.000000        NA          NA         NA           NA
asc_bus_base                   -0.034682  0.566515   -0.06122   0.554634    -0.06253
asc_bus_shift_female            0.231637  0.127793    1.81259   0.160041     1.44736
asc_bus_shift_business         -2.209101  0.365643   -6.04169   0.455506    -4.84978
asc_air_base                   -0.864407  0.362153   -2.38686   0.347832    -2.48513
asc_air_shift_female            0.282721  0.090103    3.13777   0.105003     2.69251
asc_air_shift_business          2.699196  0.103423   26.09850   0.116987    23.07265
asc_rail_base                  -2.100613  0.345312   -6.08322   0.335332    -6.26427
asc_rail_shift_female           0.185510  0.071512    2.59413   0.079163     2.34340
asc_rail_shift_business         1.423986  0.084659   16.82025   0.094026    15.14460
...
```

Heterogeneity in δ

Model comparison

- Clearly rejects model with generic ASCs

```
> apollo_lrTest("MNL_modeChoice_SP_mode_specific_time",model)
                                                 LL par
MNL_modeChoice_SP_mode_specific_time  -5598.90  11
MNL_modeChoice_SP_heterogeneity_asc   -5064.83  17
Difference                              534.07   6

Likelihood ratio test-value:     1068.14
Degrees of freedom:              6
Likelihood ratio test p-value:   1.63e-227
```

Heterogeneity in δ

Recovery of market shares now perfect at group-level

```
Prediction test for group: business (2310 observations)

                            car bus air rail  All
Times chosen (data)         366   8 771 1165 2310
Times chosen (prediction)   366   8 771 1165 2310
Diff (prediction-data)        0   0   0    0    0
t-ratio                       0   0   0    0   NA
p-val                         1   1   1    1   NA
...
Prediction test for group: female (3332 observations)

                            car bus air rail  All
Times chosen (data)         918 164 734 1516 3332
Times chosen (prediction)   918 164 734 1516 3332
Diff (prediction-data)        0   0   0    0    0
t-ratio                       0   0   0    0   NA
p-val                         1   1   1    1   NA
```

Screening data for heterogeneity

Screening data for heterogeneity

Insights before model estimation

- Choice modellers have become very lazy
- Estimation is too easy ...
- When computers were slow, people used to spend a lot of time studying the data
- Many insights are possible
- *Apollo* has a function for this

Screening data for heterogeneity

Example: choice_analysis.r

- Look at values for covariates when given alternatives are chosen/not chosen

```
> ### Define settings for analysis of choice data to be conducted prior to model estimation
> choiceAnalysis_settings <- list(
+   alternatives = c(car=1, bus=2, air=3, rail=4),
+   avail        = list(car=database$av_car, bus=database$av_bus, air=database$av_air, rail=
    ↪database$av_rail),
+   choiceVar    = database$choice,
+   explanators  = database[,c("female","business","income")],
+   rows         = (database$SP==1)
+ )
```

Screening data for heterogeneity

Outputs

	car	bus	air	rail
Explanator 1 (female), mean when alt is chosen:	0.4717	0.4581	0.4823	0.4776
Explanator 1 (female), mean when alt is not chosen:	0.5057	0.4731	0.4698	0.4837
Explanator 1 (female), t-test (mean if chosen - mean if not chosen)	-2.41	-0.55	0.82	-0.47
Explanator 2 (business), mean when alt is chosen:	0.1881	0.0223	0.5066	0.367
Explanator 2 (business), mean when alt is not chosen:	0.4114	0.3559	0.228	0.2891
Explanator 2 (business), t-test (mean if chosen - mean if not chosen)	-18.38	-33.41	19.16	6.52
Explanator 3 (income), mean when alt is chosen:	43573.6629	36060.5223	47232.6728	45257.018
Explanator 3 (income), mean when alt is not chosen:	45485.5074	45287.5717	42296.7974	43520.1627
Explanator 3 (income), t-test (mean if chosen - mean if not chosen)	-3.92	-10.2	9.67	3.99

Screening data for heterogeneity

Example: choice_analysis.r

- Can also look specific types of alternatives, e.g. cheapest *vs* most expensive

```
> ### create temporary copy of database
> database_temp=database
>
> ### replace costs for unavailable alternatives by NA
> database_temp[database_temp$cost_car==0,"cost_car"]=NA
> database_temp[database_temp$cost_bus==0,"cost_bus"]=NA
> database_temp[database_temp$cost_air==0,"cost_air"]=NA
> database_temp[database_temp$cost_rail==0,"cost_rail"]=NA
>
> ### find minimum available cost for each row
> database_temp$minimum_cost=apply(database_temp[,c("cost_car",
+                                                   "cost_bus",
+                                                   "cost_air","cost_rail")],
+                                  1,
+                                  function(x) min(x,na.rm=TRUE))
> database_temp$maximum_cost=apply(database_temp[,c("cost_car",
+                                                    "cost_bus",
+                                                    "cost_air","cost_rail")],
+                                  1,
+                                  function(x) max(x,na.rm=TRUE))
```

Screening data for heterogeneity

Example: choice_analysis.r

```
> database_temp$chosen_cost=with(database,
+                                (choice==1)*cost_car
+                                +(choice==2)*cost_bus
+                                +(choice==3)*cost_air
+                                +(choice==4)*cost_rail)
> ### indicators for choosing cheapest and most expensive
> database_temp$choose_cheapest=with(database_temp,(chosen_cost==minimum_cost))
> database_temp$choose_expensive=with(database_temp,(chosen_cost==maximum_cost))
>
> ### Define settings for analysis of choice data to be conducted prior to model estimation
> choiceAnalysis_settings <- list(
+   alternatives = c(cheapest=1, expensive=2, other=3),
+   choiceVar    = (1*database_temp$choose_cheapest
+                   +2*database_temp$choose_expensive
+                   +3*(database_temp$choose_cheapest==FALSE)*(database_temp$choose_expensive
    ↪==FALSE)),
+   explanators  = database[,c("female","business","income")],
+   rows         = (database$SP==1)
+ )
```

Screening data for heterogeneity

Outputs

	cheapest	expensive	other
Explanator 1 (female), mean when alt is chosen:	0.4699	0.4827	0.4736
Explanator 1 (female), mean when alt is not chosen:	0.4771	0.4728	0.4785
Explanator 1 (female), t-test (mean if chosen - mean if not chosen)	-0.44	0.78	-0.41
Explanator 2 (business), mean when alt is chosen:	0.123	0.5103	0.2793
Explanator 2 (business), mean when alt is not chosen:	0.3685	0.243	0.3844
Explanator 2 (business), t-test (mean if chosen - mean if not chosen)	-20.92	21.93	-9.37
Explanator 3 (income), mean when alt is chosen:	41054.5674	48992.2716	43198.0723
Explanator 3 (income), mean when alt is not chosen:	45435.4461	42699.5406	46411.4012
Explanator 3 (income), t-test (mean if chosen - mean if not chosen)	-7.65	14.93	-7.91

Heterogeneity in β

Heterogeneity in β

Including heterogeneity in cost sensitivity

- Cost is often the attribute for which we find the most taste heterogeneity
- Three levels of difficulty:
  - level 1 include shift in cost sensitivity for business travellers, plus LR test
  - level 2 in addition, include a continuous income elasticity on cost
    - involves multiplying cost component in utility function by $\left(\frac{inc_n}{\overline{inc}}\right)^{\lambda_{inc,TC}}$
      - inc_n = income of respondent n, with $\overline{inc}$ being a reference income, e.g. average
      - $\lambda_{inc,TC}$ gives elasticity of cost sensitivity in relation to income
    - β_{TC} gives cost sensitivity at mean income
    - need to create mean income variable outside `apollo_probabilities` to avoid a different mean across cores
  - level 3 think about what this all means for WTP computations
- Base these on `MNL_modeChoice_SP_heterogeneity_asc.r`

Heterogeneity in β

MNL_modeChoice_SP_heterogeneity_asc_cost_v1.r

- Business travellers less cost sensitive (but not zero)

```
LL(final)                                    : -5023.72

Estimates:
                            Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
...
b_cost_base                -0.075811    0.002011   -37.68868    0.002152     -35.22284
b_cost_shift_business       0.027538    0.002998     9.18620    0.003100       8.88259
> apollo_deltaMethod(model,
+                    deltaMethod_settings = list(
+                      expression=c(b_cost_business="b_cost_base+b_cost_shift_business")))
Running Delta method computation for user-defined function:

       Expression   Value Robust s.e. Rob t-ratio (0)
 b_cost_business -0.0483       0.0024          -19.75
```

Heterogeneity in β

Model comparison

- Clearly rejects model with generic cost coefficient

```
> apollo_lrTest("MNL_modeChoice_SP_heterogeneity_ASC",model)
                                                    LL par
MNL_modeChoice_SP_heterogeneity_asc          -5064.83  17
MNL_modeChoice_SP_heterogeneity_asc_cost     -5023.72  18
Difference                                      41.11   1

Likelihood ratio test-value:     82.22
Degrees of freedom:              1
Likelihood ratio test p-value:   1.218e-19
```

Heterogeneity in β

WTP differs between groups

```
> wtp=matrix(0,nrow=7,ncol=2)
> rownames(wtp)=c("VTT_car","VTT_bus","VTT_air","VTT_rail","VTT_access","WTP_wifi","WTP_food")
> colnames(wtp)=c("leisure","business")
> wtp[1,1]=60*model$estimate["b_tt_car"]/model$estimate["b_cost_base"]
...
> wtp[1,2]=60*model$estimate["b_tt_car"]/(model$estimate["b_cost_base"]+model$estimate["
    ↪b_cost_shift_business"])
...
> wtp[7,2]=-model$estimate["b_food"]/(model$estimate["b_cost_base"]+model$estimate["
    ↪b_cost_shift_business"])
>
> round(wtp,2)
            leisure  business
VTT_car       10.68     16.77
VTT_bus       15.66     24.60
VTT_air       15.29     24.01
VTT_rail       6.70     10.52
VTT_access    17.66     27.73
WTP_wifi      13.40     21.04
WTP_food       5.56      8.73
>
```

Heterogeneity in β

MNL_modeChoice_SP_heterogeneity_asc_cost_v2.r

- Negative income elasticity and shift for business

```
LL(final)                                  :  -4808.87

Estimates:
                          Estimate        s.e.   t.rat.(0)    Rob.s.e. Rob.t.rat.(0)
...
b_cost_base              -0.075755    0.002081    -36.4043    0.002078     -36.4484
b_cost_shift_business     0.030066    0.002768     10.8607    0.002604      11.5452
cost_income_elast        -0.611390    0.029884    -20.4586    0.030269     -20.1988
```

Heterogeneity in β

Model comparison

- Further big improvement over model with simple cost shift for business

```
> apollo_lrTest("MNL_modeChoice_SP_heterogeneity_ASC_cost_v1",model)
                                                    LL par
MNL_modeChoice_SP_heterogeneity_asc_cost_v1 -5023.72  18
MNL_modeChoice_SP_heterogeneity_asc_cost_v2 -4808.87  19
Difference                                   214.85   1

Likelihood ratio test-value:     429.7
Degrees of freedom:              1
Likelihood ratio test p-value:   1.889e-95
```

Heterogeneity in β

WTP computations

- VTT by purpose and income group
- Can code this efficiently as a loop
- Results for car and bus

```
> round(vtt_car,2)
          Income 15490 Income 30371.5 Income 44977 Income 59155.5 Income 74891
Leisure          5.97          9.01        11.46          13.55         15.65
Business         9.90         14.94        19.00          22.46         25.95

> round(vtt_bus,2)
          Income 15490 Income 30371.5 Income 44977 Income 59155.5 Income 74891
Leisure          9.22         13.92        17.70          20.92         24.17
Business        15.29         23.08        29.34          34.69         40.07
```

Heterogeneity in β

Sample enumeration

- Can also enumerate VTT/WTP over sample to get overall values
- Why is the maximum less than on the previous slide?

```
> summary(vtt_car_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.971  10.045  12.905  13.663  15.563  25.910

> summary(vtt_bus_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  9.222  15.515  19.932  21.102  24.037  40.018
```

Heterogeneity in β

Further refinement

- Two levels of difficulty:
  - level 1 allow for a shift in time sensitivity (generic across all times, including access time) and wifi sensitivity for business travellers
  - level 2 update the sample enumeration part to compute sample level WTP
- Base these on `MNL_modeChoice_SP_heterogeneity_asc_cost_v2.r`

Heterogeneity in β

MNL_modeChoice_SP_heterogeneity_full_spec.r

- Higher time and WIFI sensitivity for business

```
LL(final)                                : -4789.62
Estimates:
                          Estimate        s.e.   t.rat.(0)     Rob.s.e. Rob.t.rat.(0)
...
b_tt_car_base          -0.013652  7.9283e-04    -17.2188  7.9439e-04     -17.1848
b_tt_bus_base          -0.021614    0.001611    -13.4200    0.001525     -14.1762
b_tt_air_base          -0.018286    0.002847     -6.4229    0.002802      -6.5268
b_tt_rail_base         -0.008544    0.001877     -4.5513    0.001845      -4.6295
b_access_base          -0.020952    0.002927     -7.1584    0.002775      -7.5488
b_time_shift_business  -0.002686    0.001433     -1.8742    0.001553      -1.7299
b_cost_base            -0.074290    0.002105    -35.2942    0.002086     -35.6060
b_cost_shift_business   0.024282    0.003144      7.7224    0.002962       8.1966
cost_income_elast      -0.610595    0.029830    -20.4690    0.030175     -20.2353
b_no_frills             0.000000          NA          NA          NA           NA
b_wifi_base             0.830402    0.066640     12.4611    0.067210      12.3554
b_wifi_shift_business   0.611864    0.100797      6.0703    0.100152       6.1094
b_food                  0.437151    0.055404      7.8903    0.056936       7.6779
```

Heterogeneity in β

Model comparison

- Further (smaller) improvement over model with generic time and wifi coefficients

```
> apollo_lrTest("MNL_modeChoice_SP_heterogeneity_ASC_cost_v2",
    ↪model)
                                                    LL par
MNL_modeChoice_SP_heterogeneity_asc_cost_v2 -4808.87  19
MNL_modeChoice_SP_heterogeneity_full_spec   -4789.62  21
Difference                                     19.25   2

Likelihood ratio test-value:    38.5
Degrees of freedom:             2
Likelihood ratio test p-value:  4.363e-09
```

Heterogeneity in β

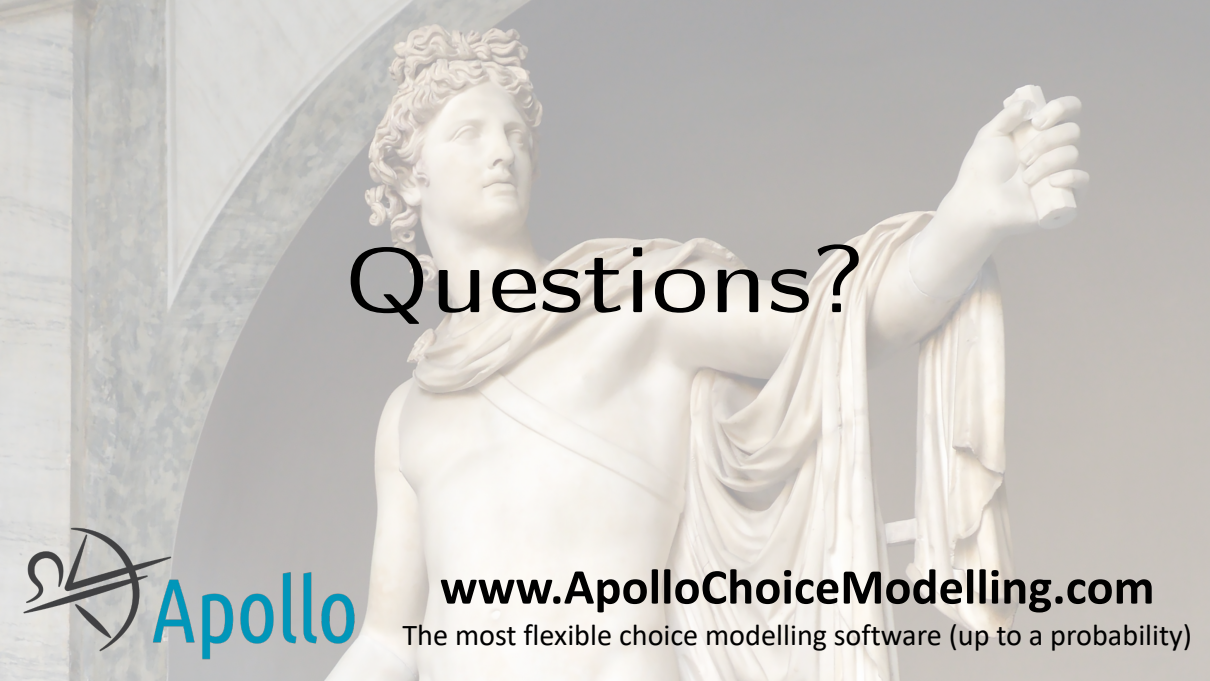
Level 2: VTT/WTP by group

```
> round(vtt_car,2)
          Income 15490 Income 30371.5 Income 44977 Income 59155.5 Income 74891
Leisure           5.77           8.70          11.06          13.07         15.10
Business         10.26          15.47          19.66          23.24         26.84
...

> round(vtt_bus,2)
          Income 15490 Income 30371.5 Income 44977 Income 59155.5 Income 74891
Leisure           9.13          13.78          17.51          20.70         23.91
Business         15.25          23.01          29.25          34.57         39.93
...

> round(wtp_wifi,2)
          Income 15490 Income 30371.5 Income 44977 Income 59155.5 Income 74891
Leisure           5.85           8.82          11.21          13.26         15.31
Business         15.09          22.76          28.93          34.20         39.50
```

Heterogeneity in β

Level 2: Sample enumeration outputs

```
> summary(vtt_car_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.769   9.699  12.634  13.620  15.602  26.807
> summary(vtt_bus_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  9.133  15.356  19.788  20.950  23.776  39.871
> summary(vtt_air_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.728  12.992  16.784  17.887  20.240  34.412
> summary(vtt_rail_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.610   6.070   8.140   8.917  10.724  18.425
> summary(vtt_access_sample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  8.854  14.886  19.185  20.341  23.072  38.786
> summary(wtp_wifi)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.848   9.833  13.374  16.735  22.956  39.441
> summary(wtp_food)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.079   5.094   6.576   6.707   7.774  11.954
```


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