Demand Modeling: 1.202
Spring 2019
Alexandra Berke

Case Study 5 - Report

The analysis in this report builds off of the work in Case Study 4 and the best model specification reached therein. In that previous case study, the best model for the Swissmetro dataset was specified as a cross-nested logit model with utility expressions as follows.

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

The log likelihood ratio value for the above model specification without the cross-nested logit structure is -4913.23. With the cross-nested logit structure the value is -4867.59.

This report will not use a cross-nested logit model, but will build upon these previous utility expressions, and compare the results reached to those previous.

Part 1

Heteroscedasticity & Alternative Specific Variance

*Files: Part1_AlternativeSpecificVariance.**

This model assumes heteroskedasticity and randomly distributes the alternative specific constants with a normal distribution. The mean value for the train alternative is fixed to zero and the variance value for the car alternative is also fixed to zero to normalize the model. The choice of which alternative specific mean variable to fix to zero is arbitrary, but the choice for which alternative specific variance value to fix to zero was not.
The determination for which variance value to normalize to zero was made by running two versions of the model in stages, described below.

This model's resulting utility expressions are as follows.

```
V_car = ASC_CAR_RANDOM + B_car_time*CAR_TT + B_car_cost*CAR_CO +
B_senior*SENIOR

V_train = ASC_TRAIN_RANDOM + B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST +
B_he*TRAIN_HE + B_ga*GA + B_firstclass*FIRST

V_sm = ASC_SM_RANDOM + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Where and the mean value for the normally distributed train ASC random variable (`ASC_TRAIN_RANDOM`) and the variance value for the normally distributed car ASC (`ASC_CAR_RANDOM`) are fixed to zero.

Two versions of the model are run in stages. Both are run with multiple draws to simulate the random normal distribution. The first simulation runs the unidentified model and estimates all variance variables in order to determine the smallest. The smallest (closest to zero) was the variance value for the car alternative, which was also the least significant variance value among the three alternatives. This was determined via a simulation with 500 draws.

Results from unidentified model (stage 1):

Init log likelihood:	-6964.663
Final log likelihood:	-4923.924
Likelihood ratio test for the init. model:	4081.479
Rho-square for the init. model:	0.293

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR_mean	-0.259	0.149	-1.74	0.0811	0.161	-1.61	0.107
ASC_CAR_std	-0.00197	0.149	-0.0132	0.989	0.0303	-0.0649	0.948
ASC_SBB_std	-0.0323	0.175	-0.184	0.854	0.0342	-0.945	0.344
ASC_SM_mean	0.0749	0.118	0.634	0.526	0.12	0.627	0.531
ASC_SM_std	0.0258	0.176	0.147	0.883	0.0516	0.5	0.617
B_CAR_COST	-0.00753	0.00107	-7.02	2.21e-12	0.00151	-5	5.82e-07
B_CAR_TIME	-0.0131	0.000813	-16.1	0	0.00164	-7.96	1.78e-15
B_FIRST	0.262	0.0735	3.56	0.000365	0.079	3.32	0.000902
B_GA	0.516	0.188	2.74	0.00612	0.194	2.66	0.00792
B_HE	-0.00598	0.00104	-5.73	1e-08	0.00105	-5.71	1.11e-08
B_SENIOR	-1.85	0.116	-16	0	0.107	-17.3	0
B_SM_COST	-0.0116	0.000618	-18.8	0	0.000931	-12.4	0
B_SM_TIME	-0.011	0.000877	-12.5	0	0.00181	-6.05	1.42e-09
B_TRAIN_COST	-0.0294	0.00127	-23	0	0.00209	-14	0
B_TRAIN_TIME	-0.00903	0.000854	-10.6	0	0.0012	-7.52	5.31e-14

The second model fixed the car specific variance value to zero to properly normalize the model before re-estimation. It was run with many more draws than the previous (1000 draws) in order to more precisely simulate the normal distribution and estimate the parameter values. (More draws are necessary for more precise estimation, but simulations with more draws were not able to complete within reasonable time).

Results from normalized model (stage 2):

Init log likelihood:	-6964.663
Final log likelihood:	-4923.938
Likelihood ratio test for the init. model:	4081.449
Rho-square for the init. model:	0.293
Rho-square-bar for the init. model:	0.291

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR_mean	-0.259	0.149	-1.74	0.0813	0.161	-1.61	0.107
ASC_SBB_std	0.00431	0.179	0.0241	0.981	0.0259	0.167	0.868
ASC_SM_mean	0.0749	0.118	0.635	0.526	0.119	0.627	0.531
ASC_SM_std	-0.029	0.18	-0.161	0.872	0.0384	-0.755	0.451
B_CAR_COST	-0.00753	0.00107	-7.02	2.23e-12	0.00151	-5	5.87e-07
B_CAR_TIME	-0.0131	0.000813	-16.1	0	0.00164	-7.96	1.78e-15
B_FIRST	0.262	0.0735	3.56	0.000368	0.079	3.32	0.000909
B_GA	0.517	0.188	2.75	0.00604	0.194	2.66	0.00783
B_HE	-0.00598	0.00104	-5.73	1.01e-08	0.00105	-5.71	1.12e-08
B_SENIOR	-1.85	0.116	-16	0	0.107	-17.3	0
B_SM_COST	-0.0116	0.000618	-18.8	0	0.000931	-12.4	0
B_SM_TIME	-0.011	0.000877	-12.5	0	0.00181	-6.05	1.43e-09
B_TRAIN_COST	-0.0294	0.00127	-23	0	0.00209	-14	0
B_TRAIN_TIME	-0.00903	0.000854	-10.6	0	0.0012	-7.52	5.33e-14

The two stage process used to estimate this model is superior to the simplified estimation process described in the case study handout example. That example effectively assumed the train alternative's variance would be smallest by normalizing its value to zero. This assumption could introduce error.

The signs of this model's estimated coefficients match expectations. The dummy variables representing a traveler has a GA ticket or rides first class have positive coefficients, reflecting the positive impacts of these beneficial characteristics. The coefficients for cost and travel time are negative, reflecting the negative impact of added cost and time spent.

The estimated mean and standard deviation values for the ASC's are not significantly different from zero. This indicates that little value was added by specifying the ASCs as random variables in this model. These results differ from those in the case study handout, which specified a model with randomly distributed ASCs and did find their estimated standard deviations to be significantly different from zero. However, that model's utility function specification was much simpler than the model featured in this report. It had fewer parameters and all variables were considered generic. The additional and alternative specific parameters in this report's model may have captured systematic variation that was otherwise represented in the randomly distributed ASCs in the report handout's simpler model.

Moreover, specifying the ASC's as random variables may have contributed an overall loss to this report's model, as they necessitated a simulation. Simulation compromised precision and added to the cost of computation.
The estimated log likelihood is in fact worse for this model than it was for the model that used the same utility expressions, but where the ASCs were constant rather than random variables (-4923.938 versus -4913.23).

Error Components

*Files: Part1_RailErrorComponent.**

This model uses a randomly distributed error term to capture the capture the unobserved shared attributes between the Swissmetro and train alternative. These results are compared to those for a similarly specified nested logit model developed in case study 4.

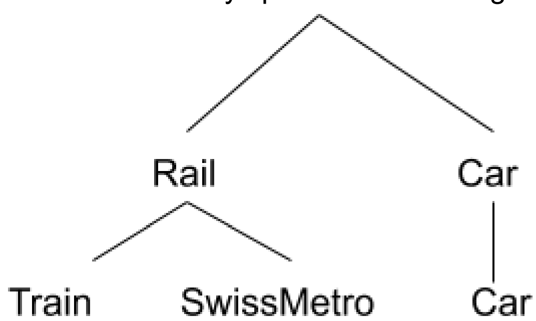

Figure: Correlation structure of NL model

A lower level logit nest included only the rail options of train and Swissmetro. The hypothesis was that the train and Swissmetro transit alternatives are quite similar versus the car alternative because they are both rail based, do not require the traveler to do the work of driving, and involve interaction with the public. This hypothesis was validated by estimating utility functions with the following expressions:

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Note that the cost and time parameters are all alternative specific.

The resulting log likelihood was -4884.805 and rho-square-bar value 0.296.

The model in this analysis uses the same utility expressions but also includes a normally distributed random error term for the rail based alternatives. This random error variable is used instead of the nested logit structure in order to capture the correlation structure.

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST + RAIL_random

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST + RAIL_random
```

Where `RAIL_random` is a normally distributed random variable.
`RAIL_random = RAIL_mean + RAIL_std * bioDraws('RAIL_random','NORMAL')`

The `RAIL_mean` is fixed to zero and the standard deviation component (`RAIL_std`) is estimated.

The model was run with 1000 draws to simulate the normally distributed error component.

Results:

Init log likelihood:	-6964.663
Final log likelihood:	-4923.933
Likelihood ratio test for the init. model:	4081.459
Rho-square for the init. model:	0.293
Rho-square-bar for the init. model:	0.291

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.259	0.149	-1.74	0.0814	0.161	-1.61	0.107
ASC_SM	0.0749	0.118	0.635	0.525	0.119	0.627	0.531
B_CAR_COST	-0.00753	0.00107	-7.02	2.22e-12	0.00151	-5	5.85e-07
B_CAR_TIME	-0.0131	0.000813	-16.1	0	0.00164	-7.96	1.78e-15
B_FIRST	0.262	0.0735	3.56	0.000369	0.079	3.32	0.000912
B_GA	0.518	0.188	2.75	0.00595	0.194	2.66	0.00774
B_HE	-0.00598	0.00104	-5.73	1.01e-08	0.00105	-5.71	1.12e-08
B_SENIOR	-1.85	0.116	-16	0	0.107	-17.3	0
B_SM_COST	-0.0116	0.000618	-18.8	0	0.000931	-12.4	0
B_SM_TIME	-0.011	0.000877	-12.5	0	0.00181	-6.05	1.42e-09
B_TRAIN_COST	-0.0294	0.00127	-23	0	0.00209	-14	0
B_TRAIN_TIME	-0.00903	0.000854	-10.6	0	0.0012	-7.52	5.33e-14
RAIL_std	-0.0287	0.153	-0.188	0.851	0.0231	-1.24	0.213

The estimated cost, headway, and time coefficients are negative as expected, reflecting the negative impacts of cost and time spent on travel alternatives. The estimated coefficients for the GA and first class dummy variables are positive, which was also expected.

The estimated value for the rail error component's standard deviation is of most interest in this model. Its estimated value is not significantly different from zero.

This result differs from the case study handout, which also estimated a model with a normally distributed random error component for the rail alternatives, and also fixed its mean to zero. The handout's model resulted in an estimated standard deviation that was significantly different from zero. However, that model was simpler than the model in this report. Its utility functions did not include parameters for the GA and first class variables. The model in this report did include dummy variables for GA and first class. Since these variables are only used in the utility functions for the rail alternatives (train and Swissmetro), it is possible that these variables, together with headway (headway is also only included in the utility functions for rail alternatives), captured a relationship between the rail alternatives that was otherwise captured by the rail error random variable in the simpler model. The hypothesis for the simpler model was that the rail error random variable would capture correlation in the rail alternatives due them sharing unobserved attributes. Those unobserved attributes could include attributes captured by the first class and GA dummy variables present in this report's model.

This model should be compared to the nested logit model that was estimated in case study 4, which used the same utility function specifications but without the random error term. The scale parameter for the rail nest was used to capture the potential correlation between the rail alternatives that the random rail error variable in this model meant to capture.
That nested logit model did estimate a scale parameter for the rail nest that was significantly different from 1. An important question is why that scale parameter was statistically significant, while the standard deviation in this report's model was not.

However, while these two model specifications may seem interchangeable, the assumptions they make to capture the correlation between rail alternatives are different, as shown below.

Nested logit assumption:

Variance/Covariance	Swissmetro	Train	Car
Swissmetro	σ^2	$\sigma^2 * \rho_{rail}$	0
Train	$\sigma^2 * \rho_{rail}$	σ^2	0
Car	0	0	σ^2

Where ρ_{rail} captures the covariance between the rail alternatives (Swissmetro and train) due to their nesting in the model.

Logit mixture assumption (this report's model):

Variance/Covariance	Swissmetro	Train	Car
Swissmetro	$\sigma^2 + \sigma^2_{rail}$	σ^2_{rail}	0
Train	σ^2_{rail}	$\sigma^2 + \sigma^2_{rail}$	0
Car	0	0	σ^2

Where σ^2_{rail} is the variance shared across rail alternatives due to the normally distributed rail random variable.

A key difference is that the nested logit model assumes that the variance for each alternative is the same, while the mixture model assumes that variance for the rail alternatives is the same, but different than the variance for the car alternative.

The two models are not directly comparable, but since the nested logit model had a better log likelihood (-4884.805 versus -4923.933) and rho-square-bar (0.296 versus 0.291), in this case the nested logit model seems to be the superior model for capturing a simple correlation between the rail alternatives.

The next model in this report will use this finding in order to build from this nested logit model and create a mixture of MEV models with nested-logit and alternative specific variance.

Mixture of MEV and Alternative Specific Variance models

*Files: Part1_MEV_Mixture.**

Results from the previous model suggested that a simple nested logit model with a nest for the rail alternatives was a good way to capture the correlation between the rail alternatives.

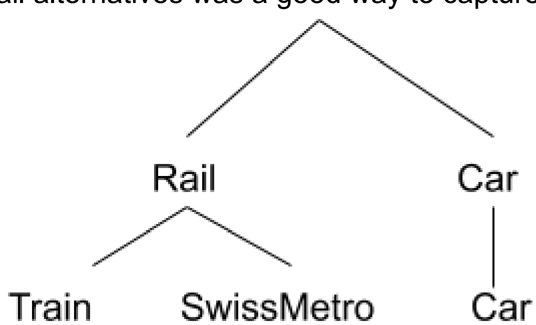

Figure: Correlation structure of NL model

This model is mixture of MEV models that builds on that nested logit model. It uses the same specification of utility functions and nest structure, but also includes alternative specific variance. The hypothesis is that while the rail alternatives are correlated, each alternative also has a different variance.   The difference between this hypothesis and the assumption of a simple nested logit model can be shown in variance/covariance matrices, as below.

Nested logit model assumption (model from case study 4):

Variance/Covariance	Swissmetro	Train	Car
Swissmetro	σ^2	$\sigma^2 * \rho_{rail}$	0
Train	$\sigma^2 * \rho_{rail}$	σ^2	0
Car	0	0	σ^2

Mixture model assumption (this report's model):

Variance/Covariance	Swissmetro	Train	Car
Swissmetro	$\sigma^2_{sm} + \sigma^2$	$\sigma^2 * \rho_{rail}$	0
Train	$\sigma^2 * \rho_{rail}$	$\sigma^2_{train} + \sigma^2$	0
Car	0	0	$\sigma^2_{car} + \sigma^2$

Where ρ_{rail} captures the covariance between the rail alternatives (Swissmetro and train) due to their nesting in the model.

The resulting model is a combination of the nested logit model and the first model specified in this report (*Part1_AlternativeSpecificVariance.**). The utility expressions are as follows.

```
V_car = ASC_CAR_RANDOM + B_car_time*CAR_TT + B_car_cost*CAR_CO +
B_senior*SENIOR

V_train = ASC_TRAIN_RANDOM + B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST +
B_he*TRAIN_HE + B_ga*GA + B_firstclass*FIRST

V_sm = ASC_SM_RANDOM + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Where and the mean value for the normally distributed train ASC random variable (`ASC_TRAIN_RANDOM`) is fixed to zero.

There is also a scale parameter estimated for the logit rail nest.

The model is estimated in two stages.   The first stage estimates the unidentified model, estimating all alternative specific variance values (none are fixed to zero), in order to determine the smallest value.   The smallest estimated variance value is then fixed to zero in the second stage of estimation.

Stage 1 of estimation

In the first stage of estimation, the starting (default) values were set to the values that resulted from estimation of the nested logit model in case study 4 (see appendix for estimated values). This was done to help the model more quickly converge and make it more likely for it to reach a global optima rather than a local optima.

Stage 1 estimation results

Stage 1 estimation was run with 20 draws and then 40 draws. In both cases, the smallest (closest to zero) estimated value for the alternative specific variance term was that for train (`ASC_SBB_std`). This value was then fixed to zero to normalize the model before re-estimation in stage 2. Results for estimation for the unidentified model with 40 draws are shown below.

Init log likelihood:	-6936.568
Final log likelihood:	-4884.238
Likelihood ratio test for the init. model:	4104.661
Rho-square for the init. model:	0.296
Rho-square-bar for the init. model:	0.294

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR_mean	-0.255	0.0936	-2.73	0.00641	0.103	-2.49	0.0129
ASC_CAR_std	-0.191	0.116	-1.65	0.0991	0.0927	-2.06	0.0395
ASC_SBB_std	0.0451	0.0399	1.13	0.258	0.0345	1.31	0.191
ASC_SM_mean	0.0247	0.0285	0.868	0.386	0.0283	0.875	0.381
ASC_SM_std	-0.0574	0.0413	-1.39	0.165	0.0392	-1.46	0.143
B_CAR_COST	-0.00609	0.00107	-5.71	1.15e-08	0.0015	-4.07	4.77e-05
B_CAR_TIME	-0.0111	0.000801	-13.9	0	0.00161	-6.91	4.89e-12
B_FIRST	0.247	0.0728	3.4	0.000681	0.0791	3.13	0.00177
B_GA	0.643	0.188	3.42	0.000629	0.201	3.2	0.00135
B_HE	-0.00159	0.000327	-4.86	1.18e-06	0.000413	-3.85	0.000117
B_SENIOR	-0.5	0.0631	-7.91	2.44e-15	0.101	-4.95	7.45e-07
B_SM_COST	-0.0125	0.000593	-21.2	0	0.000935	-13.4	0
B_SM_TIME	-0.00408	0.000577	-7.08	1.49e-12	0.00108	-3.76	0.000167
B_TRAIN_COST	-0.0196	0.000998	-19.6	0	0.00172	-11.4	0
B_TRAIN_TIME	-0.00293	0.000455	-6.43	1.27e-10	0.000783	-3.74	0.000185
MU_rail	4.03	0.445	9.06	0	0.764	5.27	1.33e-07

Stage 2 estimation

The smallest variance value estimated in stage 1 was for the train alternative. This was normalized to zero for the second stage of estimation. The updated model was run with 100[1] draws.

Stage 2 estimation results

Init log likelihood:	-6936.271
Final log likelihood:	-4884.714
Likelihood ratio test for the init. model:	4103.114
Rho-square for the init. model:	0.296
Rho-square-bar for the init. model:	0.294

[1] More than 100 draws could not complete within 8 hours on my MacBook Air with this complex model.

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR_mean	-0.252	0.0927	-2.72	0.00659	0.102	-2.48	0.0133
ASC_CAR_std	-0.0457	0.141	-0.325	0.745	0.062	-0.737	0.461
ASC_SM_mean	0.024	0.028	0.858	0.391	0.0277	0.866	0.386
ASC_SM_std	-0.0196	0.0694	-0.282	0.778	0.0501	-0.39	0.696
B_CAR_COST	-0.00601	0.00106	-5.68	1.37e-08	0.00149	-4.04	5.33e-05
B_CAR_TIME	-0.0111	0.000794	-13.9	0	0.00159	-6.97	3.19e-12
B_FIRST	0.245	0.0723	3.39	0.00071	0.0785	3.11	0.00184
B_GA	0.64	0.187	3.42	0.000634	0.2	3.2	0.00137
B_HE	-0.00156	0.000321	-4.86	1.16e-06	0.000404	-3.87	0.000109
B_SENIOR	-0.49	0.0615	-7.97	1.55e-15	0.0979	-5.01	5.52e-07
B_SM_COST	-0.0125	0.000583	-21.4	0	0.000925	-13.5	0
B_SM_TIME	-0.00402	0.000567	-7.09	1.39e-12	0.00106	-3.78	0.000156
B_TRAIN_COST	-0.0195	0.000983	-19.8	0	0.0017	-11.4	0
B_TRAIN_TIME	-0.00287	0.000446	-6.44	1.23e-10	0.000764	-3.76	0.000172
MU_rail	4.04	0.442	9.14	0	0.759	5.32	1.02e-07

The model hypothesis can be evaluated by separating the evaluations for the rail nest's scale parameter and alternative specific variances.

The scale parameter for the rail nest (`MU_rail` or ρ_{rail}) is greater than 1. This is expected because the scale parameter for the higher nest (`MU`) was implicitly set to 1, and we expect `MU/MU_rail` < 1. This inequality holds in our estimated results.
Moreover, the scale parameter for the rail nest is significantly different from 1.

t-statistic = (MU_rail - 1)/std. error = (4.04 - 1)/0.759 = 4.01

This result is consistent with the model previously developed in case study 4, which this model builds off of.

The difference between this model and the case study 4 model is the addition of the alternative specific variance assumption. The results show that the addition of alternative specific variance added computational complexity that was not justified.
The final estimated log likelihood value for this model (-4884.714) is about the same as the simpler (case study 4) model that did not assume alternative specific variance (-4884.805).
In addition, the estimated alternative specific variance values are not significantly different from zero.

Other than the alternative specific variance values, the estimated parameters match expectations, and are consistent with the previous model results. Cost, time, and headway coefficients are negative, while coefficients for the GA and first class dummy variables are positive.

Part 2

File: Part2_Bayesian.ipynb
Results spreadsheet:
https://docs.google.com/spreadsheets/d/1W_ZEaaYesM1y84D_vmoaqCdP1yphbQIVU7_lH0GcHIU

This work compares the results of 4 different model specifications.

0. (MNL) Multinomial logit.

1. (MXL 1) Logit mixture model where cost and time are normally distributed random variables, with ASCs.

2. (MXL 2) Logit mixture model where cost, time, and alternative specific parameters are *normally* distributed random variables.

3. (MXL 3) Logit mixture model where cost and time are *log-normally* distributed random variables, and alternative specific parameters are constants (ASC's).

Overall the model with the best results is MXL 2. This model is also the most complex, but the improved results justify its added complexity.

In the analysis, each model is estimated using the first 8 choices of each respondent in the Swissmetro data. These estimates are then used to predict the 9th choice for each respondent. The data from the first 8 choices is considered the *training data*, while the 9th choice is considered the *test data.*

For the logit mixture models (MXL), two sets of results are computed: (1) *unconditional* and (2) *conditional*. The first results are for the *unconditional* probability, where the posteriors of the individual specific coefficients *are not used*. The second does use the posteriors of individual specific coefficients; results are *conditional* on these values as they are used to make predictions for the same individuals these coefficients were estimated from.

The compared results are summarized in the table below.

		MXL 1		MXL 2		MXL 3	
	MNL	Unconditional	Conditional	Unconditional	Conditional	Unconditional	Conditional
Likelihood	-649.12	-652.37	-480.96	-647.8	-455.775	-659.11	-572.99
Predicted Probability	0.497	0.512	0.694	0.519	0.743	0.504	0.659

Table: Comparison of model results

The probability of the cost and time coefficients taking positive values when estimated as normally distributed random variables was computed (for both MXL 1 and MXL 2) based on the estimated mean and standard deviation values. The number of individuals with individual specific time and cost coefficients was also counted directly.

MXL 1 - Probability of coefficients greater than 0
Cost: ~ 0.19
Time: ~ 0.14

MXL 1 - Fraction of individual specific coefficients estimated with values greater than 0 (direct counts)
Cost: 90 / 752 ~ 0.12
Time: 89 / 752 ~ 0.12

MXL 2 - Probability of coefficients greater than 0
Cost: ~ 0.07
Time: ~ 0.06

MXL 2 - Fraction of individual specific coefficients estimated with values greater than 0 (direct counts)
Cost: 8 / 752 ~ 0.01
Time: 26 / 752 ~ 0.03

Below are more detailed descriptions of each model's specification, estimation process, and results.

Models and Estimation

All models are defined and estimated in *Part2_Bayesian.ipynb*

0. (MNL) Multinomial logit

This model is the same as specified in Part 1 of the handout.

```
V_car = ASC_car + BETA_time * TIME_car + BETA_cost * COST_car
V_SM = ASC_SM + BETA_time * TIME_SM + BETA_cost * COST_SM
V_train = BETA_time*TIME_train + BETA_cost * COST_train
```

The model's parameters are estimated using MCMC (on the training data):

	mean	std. dev.	2.5%	97.5%	Rhat
0	0.766398	0.058094	0.651846	0.881184	1.006574
1	0.602931	0.048624	0.507730	0.699931	1.009486
2	-1.067394	0.054099	-1.175383	-0.962256	1.007610
3	-1.261752	0.059792	-1.382350	-1.148355	1.006176

All values of R-hat are close to 1, indicating the separate Markov Chains converged to stationary values.

Results from predictions on the test data:
Log-likelihood (simulated at posterior means): -649.12
Predicted Probability: 0.497

1. (MXL 1)

This is a logit mixture model where cost and time are normally distributed random variables, with ASCs.
This model is the same as that specified in Part 2 of the handout.

```
V_car = ASC_car + BETA_time * TIME_car + BETA_cost * COST_car
V_SM = ASC_SM + BETA_time * TIME_SM + BETA_cost * COST_SM
V_train = BETA_time*TIME_train + BETA_cost * COST_train
```

Where `BETA_time` and `BETA_cost` are normally distributed random variables, and `ASC_car` and `ASC_SM` are alternative specific constants (the alternative specific constant for the train alternative is normalized to zero).

Estimated results
Estimates are again obtained by running MCMC on the training data. The resulting R-hat values are close to 1, indicating the estimated means and variance values for the random variables converged to stationary values.

```
Fixed parameters:
       mean  std. dev.      2.5%     97.5%      Rhat
0  0.403494   0.104516  0.199164  0.611025  1.009599
1  0.832593   0.089524  0.659136  1.006793  1.013095

Random parameters (means):
       mean  std. dev.      2.5%     97.5%      Rhat
0 -0.406494   0.028875 -0.464803 -0.351088  1.008459
1 -0.485257   0.026475 -0.537479 -0.434834  1.010140

Random parameters (standard deviations):
       mean  std. dev.      2.5%     97.5%      Rhat
0  0.468416   0.031896  0.410759  0.534444  1.009343
1  0.451258   0.023075  0.407650  0.497703  1.017810
```

Results from predictions on the test data:
Unconditional log likelihood: -652.37
Conditional log likelihood: -480.96

Unconditional predicted probability: 0.512
Conditional predicted probability: 0.694

A problem with using normally distributed random variables for the cost and time coefficients is that while we understand that these coefficients should always be negative, they are not guaranteed to be negative.
The probability that they have positive values can be computed from their estimated mean and standard deviation values. The number of individual specific coefficients with values greater than zero can also be counted directly from the posterior estimates obtained from the Gibbs sampler.

Probability of values greater than zero based on normal distribution parameters

Cost:
Cost mean (estimated): -0.406494
Cost std. deviation (estimated): 0.468416

Calculation of Z-Score: $(x - \mu) / \sigma = (0 + 0.406) / 0.468 = 0.8675$
Probability of value > 0 is about 1 - 0.81 = 0.19

Time:
Time mean (estimated): -0.485257

Time std. deviation (estimated): 0.451258

Calculation of Z-Score: $(x - \mu) / \sigma = (0 + 0.485) / 0.451 = 1.075$
Probability of value > 0 is about 1 - 0.86 = 0.14

Directly counting positive cost and time individual specific coefficients
This is done in the notebook code (above line 25 in *Part2_Bayesian.ipynb*).

The total number of estimated individual specific coefficients was 752.

Cost
The number of estimated individual specific **cost** coefficients greater than zero: 90
Fraction of total: 90 / 752 ~ 0.12

Time
The number of estimated individual specific **time** coefficients greater than zero: 89
Fraction of total: 89 / 752 ~ 0.12

2. (MXL 2)

This is a logit mixture model where *all* variables: cost, time, and alternative specific parameters are *normally* distributed random variables.
I.e. the utility expressions are:

```
V_car = ASC_car + BETA_time * TIME_car + BETA_cost * COST_car
V_SM = ASC_SM + BETA_time * TIME_SM + BETA_cost * COST_SM
V_train = BETA_time*TIME_train + BETA_cost * COST_train
```

Where `BETA_time`, `BETA_cost`, `ASC_car`, and `ASC_SM` are normally distributed random variables.
Note that in this model specification, both the mean and variance for the train alternative are fixed to zero in order to normalize the model. Proper model estimation would be done in two stages. The first stage would estimate the unidentified model with all variance values as random variables in order to find the smallest variance value. The smallest value would then be normalized to zero for the second stage of estimation.

Initially Estimated results
The initially estimated parameter values, estimated by MCMC on the training data are below.

```
Random parameters (means):
       mean  std. dev.      2.5%     97.5%      Rhat
0 -0.014269   0.190994 -0.387485  0.360423  1.018961
1  0.693842   0.234670  0.230365  1.151820  1.012315
2 -0.566741   0.038004 -0.641798 -0.494913  1.406679
3 -0.720039   0.039134 -0.802991 -0.646028  1.394247

Random parameters (standard deviations):
       mean  std. dev.      2.5%     97.5%      Rhat
0  2.236346   0.224754  1.805045  2.689522  1.054404
1  3.605029   0.277003  3.091226  4.159134  1.036569
2  0.390697   0.040099  0.316297  0.471971  2.116572
3  0.473488   0.033744  0.409751  0.542184  2.444560
```

The mean values for the cost and time coefficients have the expected signs, but their R-hat values are not close to one, indicating that they should not be trusted to have reached stationary values between the two Markov Chains they were estimated with.

To deal with this, I increased the number of Markov Chain iterations by an order of magnitude and reran the MCMC estimation.

```
mcmc_nChain = 2
# Note the extra iterations
mcmc_iterBurn = 200000  # vs: 20000
mcmc_iterSample = 200000  # vs: 20000
mcmc_thin = 5
mcmc_iterMem = 20000
mcmc_disp = 1000
seed = 4711
simDraws = 10000
```

Final Estimated Results

```
Random parameters (means):
       mean  std. dev.       2.5%      97.5%       Rhat
0 -0.006811   0.192645 -0.380114  0.375000  1.001087
1  0.700808   0.228102  0.254882  1.151687  1.000768
2 -0.564432   0.037921 -0.641728 -0.493622  1.028296
3 -0.715819   0.037580 -0.791738 -0.644111  1.028836

Random parameters (standard deviations):
       mean  std. dev.       2.5%      97.5%       Rhat
0  2.227488   0.215013  1.818990  2.661734  1.003090
1  3.603926   0.269848  3.098442  4.152504  1.001957
2  0.387283   0.041217  0.309406  0.471667  1.081652
3  0.469561   0.032644  0.407820  0.536623  1.127215
```

The estimated mean values for cost and time were negative, as expected. All R-hat values were satisfactorily close to 1.

The estimated model was then run on the test data.

Results from predictions on the test data:
Unconditional log likelihood: -647.80
Conditional log likelihood: -455.775

Unconditional predicted probability: 0.519
Conditional predicted probability: 0.743

These results are a large improvement over the results for MNL and the mixed logit model that did not specify alternative specific variance.

		MXL 1		MXL 2	
	MNL	Unconditional	Conditional	Unconditional	Conditional
Likelihood	-649.12	-652.37	-480.96	-647.8	-455.775
Predicted Probability	0.497	0.512	0.694	0.519	0.743

However, these improvements are at the cost of computational complexity. This additional complexity is not only observed through the need to estimate additional random variables, but also in the number of iterations needed by the Markov Chain in order to reach stationary values. The added iterations add computational time.

This model suffers from the same issue of MXL 1 in that the cost and time coefficients are not guaranteed to be negative due to their specification as normally distributed random variables.

The probability of these coefficients taking on positive values is calculated below.

Cost:
Cost mean (estimated): -0.564432
Cost std. deviation (estimated): 0.387283

Calculation of Z-Score: $(x - \mu) / \sigma = (0 + 0.564) / 0.387 = 1.457$
Probability of value > 0 is about 1 - 0.93 = 0.07

Time:
Time mean (estimated): -0.715819
Time std. deviation (estimated): 0.469561

Calculation of Z-Score: $(x - \mu) / \sigma = (0 + 0.716) / 0.47 = 1.523$
Probability of value > 0 is about 1 - 0.94 = 0.06

Directly counting positive time and cost coefficients
The number of individual specific coefficients for time and cost that took on positive values was also counted directly within the notebook code (see line 40 in *Part2_Bayesian.ipynb*).

The total number of estimated individual specific coefficients was 752.

Cost
The number of estimated individual specific **cost** coefficients greater than zero: 8
Fraction of total: 8 / 752 ~ 0.01

Time
The number of estimated individual specific **time** coefficients greater than zero: 26
Fraction of total: 26 / 752 ~ 0.03

These values present an additional improvement of MXL 2 over MXL 1: The probabilities and direct counts of the time and cost coefficients taking on positive values are much lower for MXL 2 than MXL 1.

The improvements to MXL 2 over MXL 1 seem to justify the added complexity of MXL 2. However, it is possible that there are are other, less complex models that can reach similar improvements. For example, MXL 3 is an attempt to find such a model. It specifies the time and cost coefficients as *log-normally* distributed random variables instead of normally distributed. It also simplifies MXL 2 by removing the normally distributed alternative specific random variables.

3. (MXL 3)

This is a logit mixture model where cost and time are *log-normally* distributed random variables, and the alternative specific parameters are constants (ASC's).
I.e. the utility expressions are:

```
V_car = ASC_car + BETA_time * TIME_car + BETA_cost * COST_car
V_SM = ASC_SM + BETA_time * TIME_SM + BETA_cost * COST_SM
V_train = BETA_time*TIME_train + BETA_cost * COST_train
```

Where `BETA_time` and `BETA_cost` are *log-normally* distributed random variables, and `ASC_car` and `ASC_SM` alternative specific constants (the alternative specific constant for the train alternative is normalized to zero).

The benefit of this model specification over MXL 2 is that it more closely matches our hypothesis that the cost and time coefficients are negative, because their *log-normal* distribution constrains them to negatives values.

The model's parameters are estimated using MCMC (on the training data):

```
Fixed parameters:
       mean   std. dev.      2.5%     97.5%      Rhat
0 -0.230958   0.079589 -0.387074 -0.076887  1.008193
1  0.562897   0.071738  0.422578  0.706022  1.010078

Random parameters (means):
       mean   std. dev.      2.5%     97.5%      Rhat
0 -1.643367   0.125988 -1.898574 -1.398229  1.002418
1 -0.841716   0.068357 -0.973698 -0.705896  1.008828

Random parameters (standard deviations):
       mean   std. dev.      2.5%     97.5%      Rhat
0  1.747499   0.123857  1.516690  1.998678  1.000737
1  1.355614   0.072767  1.222574  1.503236  1.002607
```

All R-hat values are close enough to zero to be confident in the estimated values.

The estimated model was then run on the test data.

Results from predictions on the test data:

Unconditional log likelihood: -659.11
Conditional log likelihood: -572.99

Unconditional predicted probability: 0.504
Conditional predicted probability: 0.659

The resulting log likelihood and prediction values are worse than those for all of the previously specified models, including MNL.

		MXL 1		MXL 2		MXL 3	
	MNL	Unconditional	Conditional	Unconditional	Conditional	Unconditional	Conditional
Likelihood	-649.12	-652.37	-480.96	-647.8	-455.775	-659.11	-572.99
Predicted Probability	0.497	0.512	0.694	0.519	0.743	0.504	0.659

This model specification should therefore be rejected.

Code & Results

See attached zip file for all model and output files.

Code is also on Github:
https://github.com/aberke/MIT-Demand-Modeling/tree/master/case-study-5

Results spreadsheet:
https://docs.google.com/spreadsheets/d/1W_ZEaaYesM1y84D_vmoaqCdP1yphbQIVU7_IH0GcHIU/edit?usp=sharing

Appendix

Table: Results from simple nested logit model estimated in case study 4.

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.252	0.0926	-2.72	0.00651	0.102	-2.48	0.0131
ASC_SM	0.0239	0.0279	0.855	0.393	0.0277	0.863	0.388
B_CAR_COST	-0.00601	0.00106	-5.68	1.37e-08	0.00149	-4.04	5.36e-05
B_CAR_TIME	-0.0111	0.000793	-13.9	0	0.00159	-6.97	3.14e-12
B_FIRST	0.244	0.0722	3.38	0.000713	0.0785	3.11	0.00185
B_GA	0.64	0.187	3.42	0.000635	0.2	3.2	0.00137
B_HE	-0.00156	0.000321	-4.87	1.14e-06	0.000404	-3.87	0.000109
B_SENIOR	-0.49	0.0613	-7.98	1.33e-15	0.0977	-5.01	5.34e-07
B_SM_COST	-0.0125	0.000582	-21.4	0	0.000925	-13.5	0
B_SM_TIME	-0.00401	0.000566	-7.09	1.37e-12	0.00106	-3.78	0.000155
B_TRAIN_COST	-0.0194	0.000982	-19.8	0	0.0017	-11.4	0
B_TRAIN_TIME	-0.00287	0.000445	-6.44	1.23e-10	0.000763	-3.76	0.000171
MU_rail	4.04	0.442	9.15	0	0.76	5.32	1.04e-07