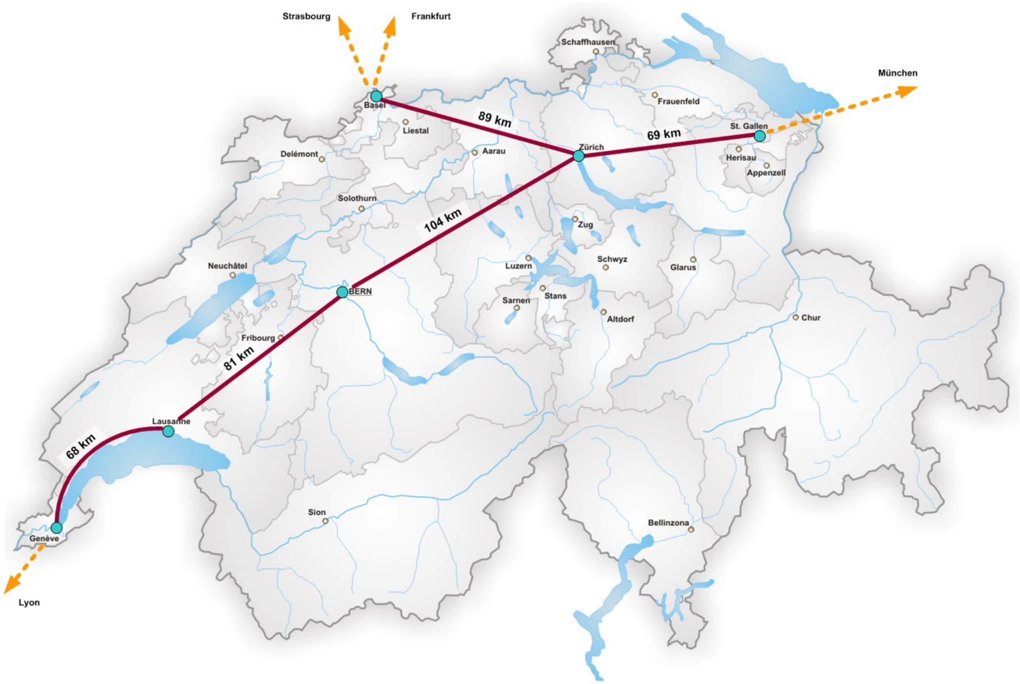
Swissmetro

- A MagLev underground system operating in partial vacuum to connect the major urban centers in Switzerland



MNL Model of Swissmetro Data

- SP data only – business and commuting trips
- Choice between **car**, **train**, and **Swissmetro (SM)**
- Considered variables:
  - Cost [sFr]
  - Travel time [min]
  - Headway [min] (Time between two vehicles at station for train and SM)
  - Annual Season Ticket (1: if individual has, 0: otherwise)
  - Senior (age > 65)

MNL Model of Swissmetro Data

- Estimation of Multinomial Logit Models (MNL)
  - Start by considering **generic** attributes

    Example code:
    Ch05_Logit/Swissmetro/logit_SM_generic.ipynb

  - Test **alternative-specific** attributes

    Example code:
    Ch05_Logit/Swissmetro/logit_SM_specific.ipynb
    Note the change in B_COST changes

  - Add and test **socio-economic characteristics**

    Example code:
    Ch05_Logit/Swissmetro/logit_SM_socioec.ipynb
    Note the new coefficients

Model Specification

Variable	Description	Parameters to be Estimated		
		Car	Train	SM
	Constant	ASC_{Car}		ASC_{SM}
TT	Travel Time	β_{Time}	β_{Time}	β_{Time}
COST	Travel Cost	$\beta_{Cost\text{-}Car}$	$\beta_{Cost\text{-}Train}$	$\beta_{Cost\text{-}SM}$
HE	Headway		β_{HE}	β_{HE}
GA	Annual Season Ticket		β_{GA}	β_{GA}
SENIOR	Age > 65	β_{Age}		β_{Age}

Model Specification

$$V_{Car} = ASC_{Car} + \beta_{Time}TT_{Car} + \beta_{Cost-Car}COST_{Car} + \beta_{Age}SENIOR$$
$$V_{Train} = 0 + \beta_{Time}TT_{Train} + \beta_{Cost-Train}COST_{Train} + \beta_{HE}HE_{Train} + \beta_{GA}GA$$
$$V_{SM} = ASC_{SM} + \beta_{Time}TT_{SM} + \beta_{Cost-SM}COST_{SM} + \beta_{HE}HE_{SM} + \beta_{Age}SENIOR + \beta_{GA}GA$$

Model Specification

$$V_{Car} = ASC_{Car} + \beta_{Time}TT_{Car} + \beta_{Cost-Car}COST_{Car} + \beta_{Age}SENIOR$$
$$V_{Train} = 0 + \beta_{Time}TT_{Train} + \beta_{Cost-Train}COST_{Train} + \beta_{HE}HE_{Train} + \beta_{GA}GA$$
$$V_{SM} = ASC_{SM} + \beta_{Time}TT_{SM} + \beta_{Cost-SM}COST_{SM} + \beta_{HE}HE_{SM} + \beta_{Age}SENIOR + \beta_{GA}GA$$

- Generic Time Coefficient

Model Specification

$$V_{Car} = ASC_{Car} + \beta_{Time}TT_{Car} + \beta_{Cost-Car}COST_{Car} + \beta_{Age}SENIOR$$
$$V_{Train} = 0 + \beta_{Time}TT_{Train} + \beta_{Cost-Train}COST_{Train} + \beta_{HE}HE_{Train} + \beta_{GA}GA$$
$$V_{SM} = ASC_{SM} + \beta_{Time}TT_{SM} + \beta_{Cost-SM}COST_{SM} + \beta_{HE}HE_{SM} + \beta_{Age}SENIOR + \beta_{GA}GA$$

- Generic Time Coefficient
- Alternative Specific Cost Coefficient

Estimation Results

- Initial log-likelihood: -6958.42
- Final log-likelihood: -4927.17
- $\bar{\rho}^2$: 0.291
- Number of observations: 6759

Parameter	Estimate	Rob. t-stat
ASC_{Car}	-0.608	-4.24
ASC_{SM}	-0.135	-1.26
β_{Time}	-1.11	-9.19
$\beta_{Car\text{-}Cost}$	-0.936	-8.02
$\beta_{Train\text{-}Cost}$	-2.68	-15.2
$\beta_{SM\text{-}Cost}$	-1.04	-14.0
β_{HE}	-0.586	-5.55
β_{GA}	0.557	2.91
β_{Age}	-1.88	-17.3

Estimation Results

- Initial log-likelihood: -6958.42
  - Initial LL vs. $L(0)$
- Final log-likelihood: -4927.17
- $\bar{\rho}^2$: 0.291
- Number of observations: 6759

Parameter	Estimate	Rob. t-stat
ASC_{Car}	-0.608	-4.24
ASC_{SM}	-0.135	-1.26
β_{Time}	-1.11	-9.19
$\beta_{Car\text{-}Cost}$	-0.936	-8.02
$\beta_{Train\text{-}Cost}$	-2.68	-15.2
$\beta_{SM\text{-}Cost}$	-1.04	-14.0
β_{HE}	-0.586	-5.55
β_{GA}	0.557	2.91
β_{Age}	-1.88	-17.3

Estimation Results

- What are the preferences over the three alternatives, all else being the same?
- What do the time and cost parameters show?
- How does frequency affect the train utility?
- Socio-economic characteristics?

Parameter	Estimate	Rob. t-stat
ASC_{Car}	-0.608	-4.24
ASC_{SM}	-0.135	-1.26
β_{Time}	-1.11	-9.19
$\beta_{Car\text{-}Cost}$	-0.936	-8.02
$\beta_{Train\text{-}Cost}$	-2.68	-15.2
$\beta_{SM\text{-}Cost}$	-1.04	-14.0
β_{HE}	-0.586	-5.55
β_{GA}	0.557	2.91
β_{Age}	-1.88	-17.3

Estimation Results

- Train is preferred over Car and SM
- Time and cost coefficients are negative and significant
- Sensitivity to cost is the highest for train
- If rail-based modes are more frequent they have higher utility
- Having a season ticket increases the utility towards rail-based modes
- Seniors prefer train

Parameter	Estimate	Rob. t-stat
ASC_{Car}	-0.608	-4.24
ASC_{SM}	-0.135	-1.26
β_{Time}	-1.11	-9.19
$\beta_{Car\text{-}Cost}$	-0.936	-8.02
$\beta_{Train\text{-}Cost}$	-2.68	-15.2
$\beta_{SM\text{-}Cost}$	-1.04	-14.0
β_{HE}	-0.586	-5.55
β_{GA}	0.557	2.91
β_{Age}	-1.88	-17.3

Specification Testing: Generic vs. Alternative-Specific Test

- How can we justify the model specification with alternative specific cost coefficients as opposed to a generic specification?
- Likelihood ratio test
- **Restricted model:** Generic

$$V_{Car} = ASC_{Car} + \beta_{Time}TT_{Car} + \beta_{Cost}COST_{Car} + \beta_{Age}SENIOR$$
$$V_{Train} = 0 + \beta_{Time}TT_{Train} + \beta_{Cost}COST_{Train} + \beta_{HE}HE_{Train} + \beta_{GA}GA$$
$$V_{SM} = ASC_{SM} + \beta_{Time}TT_{SM} + \beta_{Cost}COST_{SM} + \beta_{HE}HE_{SM} + \beta_{Age}SENIOR + \beta_{GA}GA$$

- **Unrestricted model:** Alternative-Specific

$$V_{Car} = ASC_{Car} + \beta_{Time}TT_{Car} + \beta_{Cost-Car}COST_{Car} + \beta_{Age}SENIOR$$
$$V_{Train} = 0 + \beta_{Time}TT_{Train} + \beta_{Cost-Train}COST_{Train} + \beta_{HE}HE_{Train} + \beta_{GA}GA$$
$$V_{SM} = ASC_{SM} + \beta_{Time}TT_{SM} + \beta_{Cost-SM}COST_{SM} + \beta_{HE}HE_{SM} + \beta_{Age}SENIOR + \beta_{GA}GA$$

Specification Testing: Generic vs. Alternative-Specific Test

- H_0: $\beta_{Cost-Car} = \beta_{Cost-Train} = \beta_{Cost-SM}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

Restricted Model

Model:	Multinomial Logit
Number of estimated parameters:	5
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5315.386
Likelihood ratio test:	3298.553
Rho-square:	0.237
Adjusted rho-square:	0.236
Final gradient norm:	+2.161e-002
Diagnostic:	Convergence reached...
Iterations:	69
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Unrestricted Model

Model:	Multinomial Logit
Number of estimated parameters:	7
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5068.559
Likelihood ratio test:	3792.209
Rho-square:	0.272
Adjusted rho-square:	0.271
Final gradient norm:	+1.262e-002
Diagnostic:	Convergence reached...
Iterations:	81
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Specification Testing: Generic vs. Alternative-Specific Test

- H_0: $\beta_{Cost-Car} = \beta_{Cost-Train} = \beta_{Cost-SM}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

Restricted Model

Model:	Multinomial Logit
Number of estimated parameters:	5
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5315.386
Likelihood ratio test:	3298.553
Rho-square:	0.237
Adjusted rho-square:	0.236
Final gradient norm:	+2.161e-002
Diagnostic:	Convergence reached...
Iterations:	69
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Unrestricted Model

Model:	Multinomial Logit
Number of estimated parameters:	7
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5068.559
Likelihood ratio test:	3792.209
Rho-square:	0.272
Adjusted rho-square:	0.271
Final gradient norm:	+1.262e-002
Diagnostic:	Convergence reached...
Iterations:	81
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Specification Testing: Generic vs. Alternative-Specific Test

- H_0: $\beta_{Cost-Car} = \beta_{Cost-Train} = \beta_{Cost-SM}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

Restricted Model

Model:	Multinomial Logit
Number of estimated parameters:	$\boxed{5}$
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5315.386
Likelihood ratio test:	3298.553
Rho-square:	0.237
Adjusted rho-square:	0.236
Final gradient norm:	+2.161e-002
Diagnostic:	Convergence reached...
Iterations:	69
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Unrestricted Model

Model:	Multinomial Logit
Number of estimated parameters:	$\boxed{7}$
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5068.559
Likelihood ratio test:	3792.209
Rho-square:	0.272
Adjusted rho-square:	0.271
Final gradient norm:	+1.262e-002
Diagnostic:	Convergence reached...
Iterations:	81
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

- The degrees of freedom is $df = K_U - K_R$

Specification Testing: Generic vs. Alternative-Specific Test

- H_0: $\beta_{Cost-Car} = \beta_{Cost-Train} = \beta_{Cost-SM}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

Restricted Model

Model:	Multinomial Logit
Number of estimated parameters:	5
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5315.386
Likelihood ratio test:	3298.553
Rho-square:	0.237
Adjusted rho-square:	0.236
Final gradient norm:	+2.161e-002
Diagnostic:	Convergence reached...
Iterations:	69
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

Unrestricted Model

Model:	Multinomial Logit
Number of estimated parameters:	7
Number of observations:	6768
Number of individuals:	6768
Null log-likelihood:	-6964.663
Init log-likelihood:	-6964.663
Final log-likelihood:	-5068.559
Likelihood ratio test:	3792.209
Rho-square:	0.272
Adjusted rho-square:	0.271
Final gradient norm:	+1.262e-002
Diagnostic:	Convergence reached...
Iterations:	81
Run time:	00:05
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

- The degrees of freedom is $df = K_U - K_R$
- A typical threshold is $\alpha = 0.05$

Specification Testing: Generic vs. Alternative-Specific Test

- H_0: $\beta_{Cost-Car} = \beta_{Cost-Train} = \beta_{Cost-SM}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

  $-2(L_R - L_U) = -2(-5315.386 + 5068.559) = 493.654$
  $\chi^2_{((1-0.05),2)} = 5.991$

- We reject H_0. There exists sufficient statistical evidence that our model should include alternative specific cost coefficients

Specification Testing: Market Segmentation Test

- Is there variation across different market segments?
  - E.g. education levels, income categories, etc.,

- Estimate model on the full data set. Then estimate the same model on each group separately

Example code:
Ch06_SpecTesting/Swissmetro/
SpecTest_SM_segmentation.ipynb
Note how the database is duplicated and restricted into Male and Female observations

Specification Testing: Market Segmentation Test

- H_0: $\boldsymbol{\beta_{Male}} = \boldsymbol{\beta_{Female}}$
- Reject H_0 if $-2(L_R - L_U) > \chi^2_{((1-\alpha),df)}$

Model	Log likelihood	Number of coefficients
Male	-3680.002	9
Female	-1110.618	9
Restricted model	-4927.167	9

$$-2(L_R - L_U) = -2(-4927.167 + 3680.002 + 1110.618)$$
$$= 273.094$$
$$\chi^2_{((1-0.05),9)} = 16.919$$

- The log likelihood of the unrestricted model is the sum of the log likelihoods of the models estimated on each of the 2 segments