Tests

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Introduction

- Impossible to determine the most appropriate model specification
- A good fit does not mean a good model
- Formal testing is necessary, but not sufficient
- No clear-cut rules can be given
- Subjective judgments of the analyst
- Good modeling = good judgment + good analysis

Introduction

Hypothesis testing. Two propositions

- H_0 null hypothesis
- H_1 alternative hypothesis

Analogy with a court trial:

- H_0: the defendant
- “Presumed innocent until proved guilty”
- H_0 is accepted, unless the data argue strongly to the contrary
- Benefit of the doubt

Introduction

- Informal tests
- Asymptotic t-test, Confidence interval
- Likelihood ratio tests
  - Test of generic attributes
  - Test of taste variations
  - Test of heteroscedasticity
- Goodness-of-fit measures
- Non nested hypotheses, Nonlinear specifications
- Prediction tests
  - Outlier analysis
  - Market segmentation tests

Informal tests

Wilkinson (1999) “The grammar of graphics”. Springer

> ... some researchers who use statistical methods pay more attention to goodness of fit than to the meaning of the model... Statisticians must think about what the models mean, regardless of fit, or they will promulgate nonsense.

- Is the sign of the coefficient consistent with expectation?
- Are the trade offs meaningful?

Informal tests

Sign of the coefficient

Example: Netherlands Mode Choice Case

Name	Value	Std err	t-test	Robust Std err	Robust t-test
ASC_CAR	-0.80	0.27	-2.95	0.28	-2.90
BETA_COST	-0.05	0.01	-4.85	0.01	-4.67
BETA_TIME	-1.33	0.34	-3.86	0.35	-3.75

Informal tests

Value of trade-offs

- How much are we ready to pay for an improvement of the level-of-service?
- Example: reduction of travel time
- The increase in cost must be exactly compensated by the reduction of travel time

$$\beta_C(C+\Delta C)+\beta_T(T-\Delta T)+\ldots=\beta_C C+\beta_T T+\ldots$$

Therefore,

$$\frac{\Delta C}{\Delta T}=\frac{\beta_T}{\beta_C}$$

Informal tests

Value of trade-offs

In general:

- Trade-off: $\frac{\partial V / \partial x}{\partial V / \partial x_C}$
- Units: $\frac{1/\text{Hour}}{1/\text{Guilder}} = \frac{\text{Guilder}}{\text{Hour}}$

Name	Value	Guilders	Euros	CHF	
ASC_CAR	-0.80	15.97	7.25	11.21	
BETA_COST	-0.05				
BETA_TIME	-1.33	26.55	12.05	18.64	(/Hour)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

t-test

Is the estimated parameter $\hat{\theta}$ significantly different from a given value θ^*?

- $H_0 : \hat{\theta} = \theta^*$
- $H_1 : \hat{\theta} \neq \theta^*$

Under H_0, if $\hat{\theta}$ is normally distributed with known variance σ^2

$$\frac{\hat{\theta} - \theta^*}{\sigma} \sim N(0, 1).$$

Therefore

$$P(-1.96 \leq \frac{\hat{\theta} - \theta^*}{\sigma} \leq 1.96) = 0.95 = 1 - 0.05$$

t-test

$$P(-1.96 \leq \frac{\hat{\theta}-\theta^*}{\sigma} \leq 1.96) = 0.95 = 1 - 0.05$$

H_0 can be rejected at the 5% level if

$$\left|\frac{\hat{\theta}-\theta^*}{\sigma}\right| \geq 1.96.$$

- If $\hat{\theta}$ asymptotically normal
- If variance unknown
- A t test should be used with n degrees of freedom.
- When $n \geq 30$, the Student t distribution is well approximated by a $N(0,1)$

Estimator of the asymptotic variance for ML

- Cramer-Rao Bound with the estimated parameters

$$\hat{V}_{CR} = -\nabla^2 \ln L(\hat{\theta})^{-1}$$

- Berndt, Hall, Hall & Haussman (BHHH) estimator

$$\hat{V}_{BHHH} = \left(\sum_{i=1}^{n} \hat{g}_i \hat{g}_i^T \right)^{-1}$$

where

$$\hat{g}_i = \frac{\partial \ln f_X(x_i;\theta)}{\partial \theta}$$

Estimator of the asymptotic variance for ML

Robust estimator:

$$\hat{V}_{CR}\hat{V}_{BHHH}^{-1}\hat{V}_{CR}$$

- The three are asymptotically equivalent
- This one is more robust when the model is misspecified
- Biogeme uses Cramer-Rao and the robust estimators

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

t-test

Example: Netherlands Mode Choice

Name	Value	Std err	t-test	Robust Std err	Robust t-test
ASC_CAR	-0.80	0.27	-2.95	0.28	-2.90
BETA_COST	-0.05	0.01	-4.85	0.01	-4.67
BETA_TIME	-1.33	0.34	-3.86	0.35	-3.75

t-test

Warning with the ASCs (ex: residential telephone)

Name	Value	Robust t-test	Value	Robust t-test
ASC_1			-1.22	-1.52
ASC_2	0.75	4.82	-0.48	-0.58
ASC_3	0.90	1.33	-0.32	-1.48
ASC_4	0.66	0.66	-0.57	-0.81
ASC_5	1.23	1.52		
B1_FCOST	-1.71	-6.25	-1.71	-6.25
B2_MCOST	-2.17	-8.90	-2.17	-8.90

t-test

Comparing two coefficients:
$H_0 : \beta_1 = \beta_2$. The t statistic is given by

$$\frac{\beta_1 - \beta_2}{\sqrt{\operatorname{var}(\beta_1 - \beta_2)}}$$

$$\operatorname{var}(\beta_1 - \beta_2) = \operatorname{var}(\beta_1) + \operatorname{var}(\beta_2) - 2\operatorname{cov}(\beta_1, \beta_2)$$

t-test

Ex: residential telephone

Coefficient1	Coefficient2	Rob. cov.	Rob. corr.	Rob. t-test
ASC_2	ASC_4	0.08	0.14	0.09
ASC_2	ASC_3	0.12	0.66	-0.22
ASC_3	ASC_4	0.03	0.21	0.36
ASC_1	ASC_4	0.08	0.14	-0.66
ASC_1	ASC_3	0.12	0.68	-1.33
B1_FCOST	B2_MCOST	0.02	0.36	1.56
ASC_1	ASC_2	0.65	0.98	-4.82

Confidence intervals

$$\Pr\left(-t_{\alpha/2} \leq \frac{\hat{\beta}_k - \beta_k}{\sqrt{\operatorname{var}(\hat{\beta}_k)}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

or, equivalently,

$$\Pr\left(\hat{\beta}_k - t_{\alpha/2}\sqrt{\operatorname{var}(\hat{\beta}_k)} \leq \beta_k \leq \hat{\beta}_k + t_{\alpha/2}\sqrt{\operatorname{var}(\hat{\beta}_k)}\right) = 1 - \alpha$$

for 95%, $\alpha = 0.05$, and $t_{0.025} = 1.96$.

Confidence intervals

When more than one parameter is considered, the quadratic form

$$(\hat{\beta} - \beta)^T \Sigma^{-1} (\hat{\beta} - \beta) \sim \chi^2_K$$

where

- $\beta \in \mathbb{R}^K$ is the vector of true parameters,
- $\hat{\beta} \in \mathbb{R}^K$ is the vector of estimates, and
- $\Sigma \in \mathbb{R}^{K \times K}$ is the covariance matrix.

$$\Pr\left((\hat{\beta} - \beta)^T \Sigma^{-1} (\hat{\beta} - \beta) \leq \chi^2_{K,\alpha}\right) = 1 - \alpha.$$

In two dimensions, the “confidence interval” is an ellipse.

Likelihood ratio test

- Used for “nested” hypotheses
- One model is a special case of the other
- H_0: the two models are equivalent

$$-2(\mathcal{L}(\hat{\beta}_R) - \mathcal{L}(\hat{\beta}_U)) \sim \chi^2_{(K_U - K_R)}$$

- $\mathcal{L}(\hat{\beta}_R)$ is the log likelihood of the restricted model
- $\mathcal{L}(\hat{\beta}_U)$ is the log likelihood of the unrestricted model

Likelihood ratio test

Example: Netherlands Mode Choice Case. 3 models:

- Null model (equal probability): $K = 0$, $\mathcal{L} = -158.04$
- Constants only (reproduces the sample shares): $K = J - 1 = 1$, $\mathcal{L} = -148.35$
- Model with cost and time: $K = 3$, $\mathcal{L} = -123.13$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood ratio test

$-2(\mathcal{L}(\beta_R) - \mathcal{L}(\beta_U))$			Unrestricted model	
			1	3
			-148.35	-123.13
Restricted	0	-158.04	19.38	69.81
model	1	-148.35		50.43

χ^2	1	3
0	3.84	7.81
1		5.99

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood ratio test

Test of generic attributes (ex: residential telephone)

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_C
BM	1	0	0	0	$\ln$(cost(BM))
SM	0	1	0	0	$\ln$(cost(SM))
LF	0	0	1	0	$\ln$(cost(LF))
EF	0	0	0	1	$\ln$(cost(EF))
MF	0	0	0	0	$\ln$(cost(MF))

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_CM	BETA_CF
BM	1	0	0	0	$\ln$(cost(BM))	0
SM	0	1	0	0	$\ln$(cost(SM))	0
LF	0	0	1	0	0	$\ln$(cost(LF))
EF	0	0	0	1	0	$\ln$(cost(EF))
MF	0	0	0	0	0	$\ln$(cost(MF))

Likelihood ratio test

- Log likelihood of the restricted model: -477.557
- Log likelihood of the unrestricted model: -476.608
- Test: 1.898
- Threshold 95% χ_1^2: 3.841
- Cannot reject that the two models are equivalent
- The simplest model is preferred

Note about the t-test: If we test BETA_CM=BETA_CF, we obtain 1.56, which is below the 1.96 threshold

Likelihood ratio test

Test of taste variations (ex: residential telephone)

- Estimate a different model for each of the 5 income groups
- Pool the results together. $K = 6 \times 5 = 30$.
- Estimate a model for the whole sample. $K = 6$
- The test is performed with 24 degrees of freedom

Likelihood ratio test

		data	loglike
Income group	1	115	-124.67
Income group	2	117	-120.86
Income group	3	104	-114.98
Income group	4	54	-59.23
Income group	5	44	-47.80
Pooled model		434	-467.55
Original model		434	-476.61
Test			18.11
Threshold	χ^2_{24}		36.42

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood ratio test

- We cannot reject the hypothesis that the two models are equivalent
- There is no sign of segmentation per income
- The simplest model is preferred

Likelihood ratio test

Test of heteroscedasticity (ex: residential telephone)
Model 1:

$$\begin{aligned}
V_{\text{BM}} &= \beta_1 + \beta_5 \ln(\text{cost}_{\text{BM}}) \\
V_{\text{SM}} &= \beta_2 + \beta_5 \ln(\text{cost}_{\text{SM}}) \\
V_{\text{LF}} &= \beta_3 + \beta_6 \ln(\text{cost}_{\text{LF}}) \\
V_{\text{EF}} &= \beta_4 + \beta_6 \ln(\text{cost}_{\text{EF}}) \\
V_{\text{MF}} &= \beta_6 \ln(\text{cost}_{\text{MF}})
\end{aligned}$$

Model 2: scale for perimeter area and non-metropolitan area

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood ratio test

	Est.	t-test against 1.
Perimeter	0.279	-3.98
Non metro.	0.306	-8.13

$$
\begin{aligned}
\mathcal{L}(\text{model1}) &= -476.608 \quad K = 6 \\
\mathcal{L}(\text{model2}) &= -464.068 \quad K = 8 \\
\text{Test} &= 25.08 \\
\text{Threshold } 95\% &= 5.99
\end{aligned}
$$

- We reject the hypothesis that the models are equivalent
- Homoscedasticity across individuals is rejected

Non-nested hypotheses

- Need to compare two different models
- If none of the models is a restricted version of the other, we talk about non-nested models
- The likelihood ratio test cannot be used
- Two possible tests:
  - Composite model
  - Davidson-MacKinnon J-test

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Composite model

- We want to test model 1 against model 2
- We generate a composite model C such that both models 1 and 2 are restricted cases of model C.
- We test 1 against C using the likelihood ratio test
- We test 2 against C using the likelihood ratio test
- Possible outcomes:
  - Only one of the two models is rejected. Keep the other.
  - Both models are rejected. Better models should be developed.
  - Both models are accepted. Use $\bar{\rho}^2$ to choose.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Goodness-of-fit

$$\rho^2 = 1 - \frac{\mathcal{L}(\hat{\beta})}{\mathcal{L}(0)}$$

- $\rho^2 = 0$: trivial model, equal probabilities
- $\rho^2 = 1$: perfect fit.

Warning: $\mathcal{L}(\hat{\beta})$ is a biased estimator of the expectation over all samples. Use $\mathcal{L}(\hat{\beta}) - K$ instead.

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\beta}) - K}{\mathcal{L}(0)}$$

Composite model

Example: residential telephone

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_C
BM	1	0	0	0	$\ln$(cost(BM))
SM	0	1	0	0	$\ln$(cost(SM))
LF	0	0	1	0	$\ln$(cost(LF))
EF	0	0	0	1	$\ln$(cost(EF))
MF	0	0	0	0	$\ln$(cost(MF))

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_C
BM	1	0	0	0	cost(BM)
SM	0	1	0	0	cost(SM)
LF	0	0	1	0	cost(LF)
EF	0	0	0	1	cost(EF)
MF	0	0	0	0	cost(MF)

Composite model

Composite model

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_CL	BETA_C
BM	1	0	0	0	$\ln$(cost(BM))	cost(BM)
SM	0	1	0	0	$\ln$(cost(SM))	cost(SM)
LF	0	0	1	0	$\ln$(cost(LF))	cost(LF)
EF	0	0	0	1	$\ln$(cost(EF))	cost(EF)
MF	0	0	0	0	$\ln$(cost(MF))	cost(MF)

Model	$\mathcal{L}$	K	test	conclusion
Composite	-476.80	6		
log	-477.56	5	1.51	No reject
linear	-482.72	5	11.84	Reject

Model with log is preferred

Davidson-MacKinnon J-test

$$\begin{aligned} M_0: \quad U &= f(X,\beta)+\varepsilon_0 \\ M_1: \quad U &= g(Z,\gamma)+\varepsilon_1 \end{aligned}$$

- Estimate M_1 to obtain $\hat{\gamma}$
- Consider the model obtained by convex combination

$$U = (1-\alpha)f(X,\beta)+\alpha g(Z,\hat{\gamma})+\varepsilon_0$$

- Note that α and β are estimated, not γ
- If M_0 is true, the true value of α is zero
- Perform a t-test to test α against 0.

Davidson-MacKinnon J-test

Example: residential telephone

- M_0 model with log(cost)
- M_1 model with cost

Estimate M_1

Name	Value	Robust Std err	Robust t-test
ASC_1	-0.53	0.15	-3.61
ASC_3	0.89	0.15	5.87
ASC_4	0.76	0.71	1.07
ASC_5	1.83	0.39	4.67
B1_COST	-0.15	0.02	-6.28

Davidson-MacKinnon J-test

```
[Expressions]
ASCLIN1 = -5.2704884e-01
ASCLIN3 = +8.9308708e-01
ASCLIN4 = +7.5874800e-01
ASCLIN5 = +1.8310079e+00
BETALIN = -1.4908464e-01
UTILLIN1 = ASCLIN1 + BETALIN * cost1
UTILLIN2 =           BETALIN * cost2
UTILLIN3 = ASCLIN3 + BETALIN * cost3
UTILLIN4 = ASCLIN4 + BETALIN * cost4
UTILLIN5 = ASCLIN5 + BETALIN * cost5
[Utilities]
1       BM      avail1  ALPHA * UTILLIN1
2       SM      avail2  ALPHA * UTILLIN2
3       LF      avail3  ALPHA * UTILLIN3
4       EF      avail4  ALPHA * UTILLIN4
5       MF      avail5  ALPHA * UTILLIN5
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Davidson-MacKinnon J-test

```
[GeneralizedUtilities]
1 (1 - ALPHA ) * ( ASC_1 + B1_COST * logcost1 )
2 (1 - ALPHA ) * ( ASC_2 + B1_COST * logcost2 )
3 (1 - ALPHA ) * ( ASC_3 + B1_COST * logcost3 )
4 (1 - ALPHA ) * ( ASC_4 + B1_COST * logcost4 )
5 (1 - ALPHA ) * ( ASC_5 + B1_COST * logcost5 )
```

Name	Value	Robust Std err	Robust t-test
ALPHA	0.23	0.21	1.10
ASC_1	-0.72	0.19	-3.70
ASC_3	1.22	0.22	5.67
ASC_4	1.05	0.93	1.12
ASC_5	1.77	0.38	4.68
B1_COST	-2.07	0.31	-6.73

Davidson-MacKinnon J-test

Conclusion:

- Cannot reject the hypothesis that $ALPHA = 0$.
- Cannot reject the hypothesis that the log specification is correct

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Davidson-MacKinnon J-test

- M_0 model with cost
- M_1 model with log(cost)

Estimate M_1

Name	Value	Robust Std err	Robust t-test
ASC_1	-0.72	0.15	-4.76
ASC_3	1.20	0.16	7.56
ASC_4	1.00	0.70	1.42
ASC_5	1.74	0.27	6.51
B1_COST	-2.03	0.21	-9.55

Davidson-MacKinnon J-test

```
[Expressions]
ASCLOG1 = -7.2124491e-01
ASCLOG3 = +1.2012643e+00
ASCLOG4 = +9.9917468e-01
ASCLOG5 = +1.7364214e+00
COSTLOG = -2.0261980e+00
UTILLOG1 = ASCLOG1 + COSTLOG * logcost1
UTILLOG2 =           COSTLOG * logcost2
UTILLOG3 = ASCLOG3 + COSTLOG * logcost3
UTILLOG4 = ASCLOG4 + COSTLOG * logcost4
UTILLOG5 = ASCLOG5 + COSTLOG * logcost5
[Utilities]
1       BM      avail1  ALPHA * UTILLOG1
2       SM      avail2  ALPHA * UTILLOG2
3       LF      avail3  ALPHA * UTILLOG3
4       EF      avail4  ALPHA * UTILLOG4
5       MF      avail5  ALPHA * UTILLOG5
```

Davidson-MacKinnon J-test

```
[GeneralizedUtilities]
1 (1 - ALPHA ) * ( ASC_1 + B1_COST * cost1 )
2 (1 - ALPHA ) * ( ASC_2 + B1_COST * cost2 )
3 (1 - ALPHA ) * ( ASC_3 + B1_COST * cost3 )
4 (1 - ALPHA ) * ( ASC_4 + B1_COST * cost4 )
5 (1 - ALPHA ) * ( ASC_5 + B1_COST * cost5 )
```

Name	Value	Robust Std err	Robust t-test
ALPHA	0.79	0.21	3.70
ASC_1	-0.51	0.69	-0.73
ASC_3	0.95	0.69	1.38
ASC_4	0.91	3.37	0.27
ASC_5	1.96	1.44	1.36
B1_COST	-0.16	0.09	-1.88

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Davidson-MacKinnon J-test

Conclusions:

- Reject the hypothesis that ALPHA=0
- Reject the hypothesis that the linear specification is correct

Non linear specification

Three approaches

- Piecewise linear specifications
- Power series expansion
- Box-Cox transforms

Piecewise linear specification

- A coefficient may have different values
- For example

$$V_i = \beta_{T1} x_{T1} + \beta_{T2} x_{T2} + \beta_{T3} x_{T3} + \beta_{T4} x_{T4} + \ldots$$

where

$$= \begin{cases} t & \text{if } t < 90 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T2} = \begin{cases} 0 & \text{if } t < 90 \\ t - 90 & \text{if } 90 \leq t < 180 \\ 90 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 0 & \text{if } t < 180 \\ t - 180 & \text{if } 180 \leq t < 270 \\ 90 & \text{otherwise} \end{cases} \qquad x_{T4} = \begin{cases} 0 & \text{if } t < 270 \\ t - 270 & \text{otherwise} \end{cases}$$

Piecewise linear specification

Note: coding in Biogeme

$$x_{Ti} = \begin{cases} 0 & \text{if } t < a \\ t-a & \text{if } a \leq t < a+b \\ b & \text{otherwise} \end{cases} \qquad x_{Ti} = \max(0, \min(t-a, b))$$

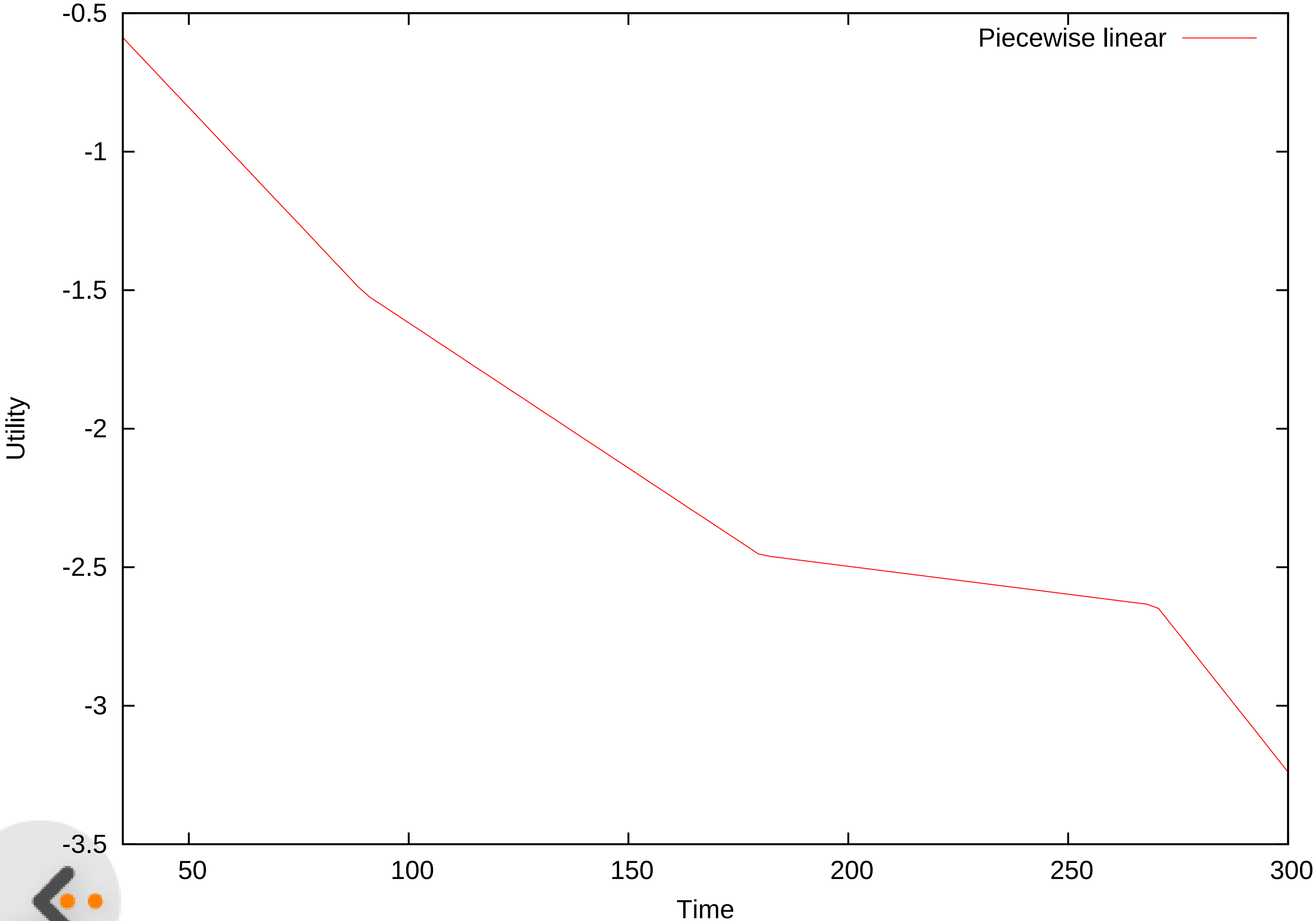
$$\begin{aligned} x_{T1} &= \min(t, 90) \\ x_{T2} &= \max(0, \min(t-90, 90)) \\ x_{T3} &= \max(0, \min(t-180, 90)) \\ x_{T4} &= \max(0, t-270) \end{aligned}$$

Piecewise linear specification

Examples:

t	TT1	TT2	TT3	TT4
40	40	0	0	0
100	90	10	0	0
200	90	90	20	0
300	90	90	90	30

Piecewise linear specification



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Piecewise linear specification

- Perform a likelihood ratio test
- Example: Swissmetro
- Linear model: $\mathcal{L} =$-5031.87 ($K = 12$)
- Piecewise linear model: $\mathcal{L} =$-5025 ($K = 15$)
- Test = -2(-5031.87 + 5025) = 13.74
- Threshold 95% χ^2_3= 7.81
- Reject the linear model

Power series

$$V_i = \beta_1 T + \beta_2 T^2 + \beta_3 T^3 + \dots$$

- In practice, these terms can be very correlated
- Example: Swissmetro with 2 terms
  - Linear model: $\mathcal{L} =$-5031.87 ($K = 12$)
  - Power series model: $\mathcal{L} =$ -5031.36 ($K = 13$)
  - Test = -2(-5031.87 +5031.36) = 1.02
  - Threshold 95% χ^2_1= 3.84
  - Cannot reject the linear model

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Power series

- Example: Swissmetro with 3 terms
  - Linear model: $\mathcal{L} =$-5031.87 ($K = 12$)
  - Power series model: $\mathcal{L} =$ -5023.79 ($K = 14$)
  - Test = -2(-5031.87 +5023.79) = 16.16
  - Threshold 95% χ^2_2= 5.99
  - Reject the linear model

Box-Cox transforms

- Box-Cox transforms

$$\beta \frac{x^\lambda - 1}{\lambda}, \; x > 0$$

- Box-Tukey transforms

$$\beta \frac{(x+\alpha)^\lambda - 1}{\lambda}, \; x + \alpha > 0$$

where β, α and λ must be estimated

Box-Cox transforms

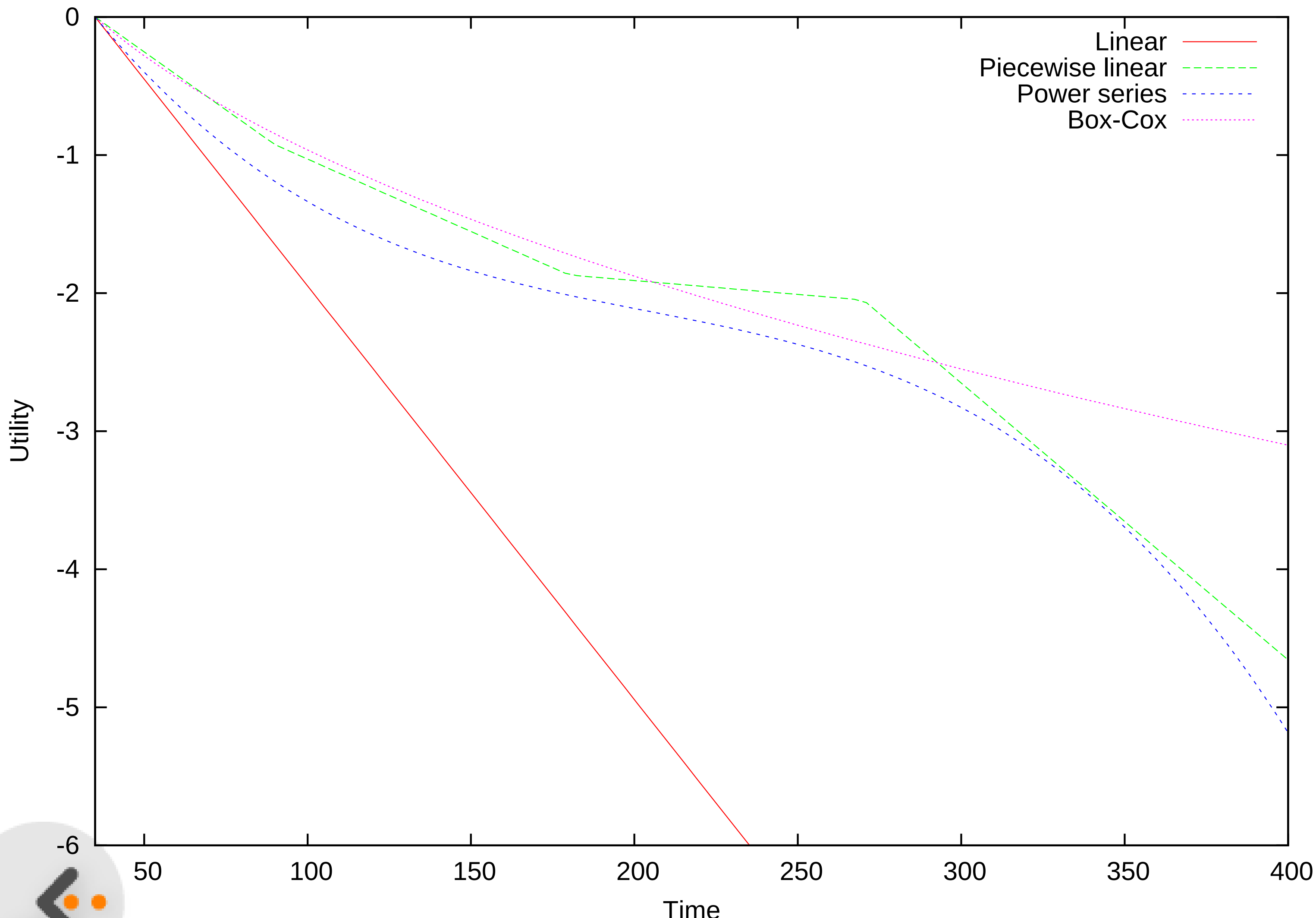
Example: Swissmetro

- Linear model: $\mathcal{L} =$-5031.87 ($K = 12$)
- Box-Cox model: $\mathcal{L} =$ -5029.83 ($K = 13$)
- Test = -2(-5031.87+5029.83) = 4.08
- Threshold 95% χ^2_3= 3.84
- Reject the linear model

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Comparison



Linear
Piecewise linear
Power series
Box-Cox
Utility
Time
0
-1
-2
-3
-4
-5
-6
50
100
150
200
250
300
350
400

Outlier analysis

- Apply the model on the sample
- Examine observations where the predicted probability is the smallest for the observed choice
- Test model sensitivity to outliers, as a small probability has a significant impact on the log likelihood
- Potential causes of low probability:
  - Coding or measurement error in the data
  - Model misspecification
  - Unexplainable variation in choice behavior

Outlier analysis

- Coding or measurement error in the data
  - Look for signs of data errors
  - Correct or remove the observation
- Model misspecification
  - Seek clues of missing variables from the observation
  - Keep the observation and improve the model
- Unexplainable variation in choice behavior
  - Keep the observation
  - Avoid over fitting of the model to the data

Outlier analysis

Example: residential telephone

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_C
BM	1	0	0	0	$\ln$(cost(BM))
SM	0	1	0	0	$\ln$(cost(SM))
LF	0	0	1	0	$\ln$(cost(LF))
EF	0	0	0	1	$\ln$(cost(EF))
MF	0	0	0	0	$\ln$(cost(MF))

Outlier analysis

- Observation with lowest probability of choice = 3.83%
- Choice: Metro Area Flat
- Costs: BM (5.39), SM (5.78), LF (8.48), EF (n.a.), MF (38.28)
- Area of residence: perimeter (without extended)
- Number of users in the household: 2 (20-29 years)
- Income: 30K–40K
- Conclusion: the model can be improved

Market segments

- Compared predicted vs. observed shares per segment
- Let N_j be the set of samples individuals in segment j
- Observed share for alt. i and segment j

$$S(i,j) = \sum_{n \in N_j} y_{in}/N$$

- Predicted share for alt. i and segment j

$$\hat{S}(i,j) = \sum_{n \in N_j} P_n(i)/N$$

Market segments

Example: residential telephone

	ASC_BM	ASC_SM	ASC_LF	ASC_EF	BETA_C
BM	1	0	0	0	$\ln$(cost(BM))
SM	0	1	0	0	$\ln$(cost(SM))
LF	0	0	1	0	$\ln$(cost(LF))
EF	0	0	0	1	$\ln$(cost(EF))
MF	0	0	0	0	$\ln$(cost(MF))

- Two segments: up to 2 users, more than 2 users

Market segments

	Predicted				Observed		
	<=2	> 2	Total		<=2	> 2	Total
1	57	16	73	1	61	12	73
2	92	31	123	2	102	21	123
3	120	58	178	3	108	70	178
4	2	1	3	4	3	0	3
5	33	24	57	5	29	28	57
	303	131	434		303	131	434

Market segments

Error	$<=2$	> 2
1	-7.0%	35.8%
2	-10.2%	49.5%
3	11.2%	-17.3%
4	-37.6%	∞
5	12.9%	-13.4%

Market segments

Note:

- With a full set of constants: $\sum_{n \in N_j} y_{in} = \sum_{n \in N_j} P_n(i)$
- Do not saturate the model with constants

Conclusions

- Tests are designed to check meaningful hypotheses
- Do not test hypotheses that do not make sense
- Do not apply the tests blindly
- Always use your judgment.