1.202 Demand Modeling
Recitation 9

Kexin (Sally) Chen

April 14, 2023

Announcements

- No class next Monday
- Final exam:
  - May 22nd 9 - 12, 2-105
  - Open book and open notes
- Case study 3 due tonight
- Case study 4 released.
  - Due: April 28, 2023
  - Specification testing
  - Development of MEV models: nested logit and cross nested logit

Outline

- Test of non-nested hypotheses
- Tests of nonlinear specifications
  - Piecewise linear
  - Power series
  - Box-Cox

Case study: choice of airline itinerary

- Survey
  - Conducted by the Boeing Company (fall 2004)
  - Sample of the customers of an Internet airline booking service
- Booking (the internet service)
  - Takes a specific user request for travel in a city pair
  - Interrogates the web sites of airlines that provide service in that market
  - Returns to the user a compiled list of available itineraries

Case study: choice of airline itinerary

- Questionnaire
  - Random selection of customers for the survey
  - Three alternatives based on the origin-destination market request that the respondent entered into the itinerary search engine:
    - A non stop flight
    - A flight with 1 stop on the same airline
    - A flight with 1 stop and a change of airline
- Demographic data
  - age
  - gender
  - Income
  - Occupation
  - Education
- Context data
  - Desired departure time
  - Trip purpose
  - Who is paying for the trip
  - The number of passengers traveling together

Case study: choice of airline itinerary

FEATURES	Non-Stop (Option 1)	1 Stop (Option 2)	1 Stop (Option 3)
Departure time (local)	6:00 PM	4:30 PM	6:00 PM
Arrival time (local)	8:14 PM	8:44 PM	9:44 PM
Total time in air	4 hr 14 min	4 hr 44 min	4 hr 44 min
Total trip time	4 hr 14 min	6 hr 14 min	5 hr 44 min
Legroom	typical legroom	2-in more of legroom	4-in more of legroom
Airline [Airplane]	Depart Chicago Continental Airlines [B737] to San Diego	Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego	Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego
Fare	$565	$485	$620
1. Which is MOST attractive?	Option 1	Option 2	Option 3
2. Which is LEAST attractive?	Option 1	Option 2	Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. Yes No

Case study: choice of airline itinerary

- Sample
  - Origin-destination city pairs in the USA
  - 3609 respondents
  - 1 choice each

Outline

- **Test of non-nested hypotheses**
- Tests of nonlinear specifications
  - Piecewise linear
  - Power series
  - Box-Cox

Test of non-nested hypotheses

- Assume we want to test a model M_1 against another model M_2, which is not a restricted version of M_1
- Generate a composite model M_C such that both M_1 and M_2 are restricted cases of M_C

Test of non-nested hypotheses (cont.)

- We test M_1 against M_C and M_2 against M_C using the likelihood ratio test
- Three possible outcomes of this test:
  1. One of the two models is rejected. Then we keep the other one.
  2. Both models are rejected. Then better models should be developed. The composite model could be used as a new basis for future specifications
  3. Both models are accepted. Then we choose the model with the highest $\bar{\rho}^2$ index.

Test of non-nested hypotheses (cont.)

- Go to CaseStudies → Ch06_SpecTesting → Airline → SpecTest_airline_cox.py
- M_1 has the following systematic utilities

$$
\begin{aligned}
V_{1n} &= \underline{\beta_{\text{Fare}}\text{Fare}_{1n}} + \beta_{\text{Legroom}}\text{Legroom}_{1n} + \beta_{\text{Total\_TT\_1}}\text{Total\_TT}_{1n} \\
&\quad + \beta_{\text{SchedDE}}\text{Opt1\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt1\_SchedDelayLate}_n \\
V_{2n} &= \text{ASC}_2 + \underline{\beta_{\text{Fare}}\text{Fare}_{2n}} + \beta_{\text{Legroom}}\text{Legroom}_{2n} + \beta_{\text{Total\_TT\_2}}\text{Total\_TT}_{2n} \\
&\quad + \beta_{\text{SchedDE}}\text{Opt2\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt2\_SchedDelayLate}_n \\
V_{3n} &= \text{ASC}_3 + \underline{\beta_{\text{Fare}}\text{Fare}_{3n}} + \beta_{\text{Legroom}}\text{Legroom}_{3n} + \beta_{\text{Total\_TT\_3}}\text{Total\_TT}_{3n} \\
&\quad + \beta_{\text{SchedDE}}\text{Opt3\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt3\_SchedDelayLate}_n
\end{aligned}
$$

- Cost related coefficients are linear

Test of non-nested hypotheses (cont.)

- M_2 has the following systematic utilities

$$
\begin{aligned}
V_{1n} &= \underline{\beta_{\text{LogFare}} \log(\text{Fare}_{1n})} + \beta_{\text{Legroom}} \text{Legroom}_{1n} + \beta_{\text{Total\_TT\_1}} \text{Total\_TT}_{1n} \\
&\quad + \beta_{\text{SchedDE}} \text{Opt1\_SchedDelayEarly}_n + \beta_{\text{SchedDL}} \text{Opt1\_SchedDelayLate}_n \\
V_{2n} &= \text{ASC}_2 + \underline{\beta_{\text{LogFare}} \log(\text{Fare}_{2n})} + \beta_{\text{Legroom}} \text{Legroom}_{2n} + \beta_{\text{Total\_TT\_2}} \text{Total\_TT}_{2n} \\
&\quad + \beta_{\text{SchedDE}} \text{Opt2\_SchedDelayEarly}_n + \beta_{\text{SchedDL}} \text{Opt2\_SchedDelayLate}_n \\
V_{3n} &= \text{ASC}_3 + \underline{\beta_{\text{LogFare}} \log(\text{Fare}_{3n})} + \beta_{\text{Legroom}} \text{Legroom}_{3n} + \beta_{\text{Total\_TT\_3}} \text{Total\_TT}_{3n} \\
&\quad + \beta_{\text{SchedDE}} \text{Opt3\_SchedDelayEarly}_n + \beta_{\text{SchedDL}} \text{Opt3\_SchedDelayLate}_n
\end{aligned}
$$

- Cost related coefficients are logarithmic

Test of non-nested hypotheses (cont.)

- M_C has the following systematic utilities

$$
\begin{aligned}
V_{1n} &= \underline{\beta_{\text{Fare}}\text{Fare}_{1n} + \beta_{\text{LogFare}}\log(\text{Fare}_{1n})} + \beta_{\text{Legroom}}\text{Legroom}_{1n} \\
&\quad + \beta_{\text{Total\_TT\_1}}\text{Total\_TT}_{1n} + \beta_{\text{SchedDE}}\text{Opt1\_SchedDelayEarly}_n \\
&\quad + \beta_{\text{SchedDL}}\text{Opt1\_SchedDelayLate}_n \\
V_{2n} &= \text{ASC}_2 + \underline{\beta_{\text{Fare}}\text{Fare}_{1n} + \beta_{\text{LogFare}}\log(\text{Fare}_{2n})} + \beta_{\text{Legroom}}\text{Legroom}_{2n} \\
&\quad + \beta_{\text{Total\_TT\_2}}\text{Total\_TT}_{2n} + \beta_{\text{SchedDE}}\text{Opt2\_SchedDelayEarly}_n \\
&\quad + \beta_{\text{SchedDL}}\text{Opt2\_SchedDelayLate}_n \\
V_{3n} &= \text{ASC}_3 + \underline{\beta_{\text{Fare}}\text{Fare}_{1n} + \beta_{\text{LogFare}}\log(\text{Fare}_{3n})} + \beta_{\text{Legroom}}\text{Legroom}_{3n} \\
&\quad + \beta_{\text{Total\_TT\_3}}\text{Total\_TT}_{3n} + \beta_{\text{SchedDE}}\text{Opt3\_SchedDelayEarly}_n \\
&\quad + \beta_{\text{SchedDL}}\text{Opt3\_SchedDelayLate}_n
\end{aligned}
$$

Test of non-nested hypotheses (cont.)

- Apply the likelihood ratio test for M_1 against M_C
- Null hypothesis: $\beta_{\text{LogFare}} = 0.$
- Reject H_0 if $-2\left(L(\hat{\beta}_{M_1}) - L(\hat{\beta}_{M_C})\right) \geq \chi^2_{K,(1-\alpha)}$

M_1 results

Summary statistics
Number of observations=3609
$\mathcal{L}(0) = -3964.892$
$\mathcal{L}(\hat{\beta}) = -2320.447$
$\bar{\rho}^2 = 0.412$

M_C results

Summary statistics
Number of observations=3609
$\mathcal{L}(0) = -3964.892$
$\mathcal{L}(\hat{\beta}) = -2271.656$
$\bar{\rho}^2 = 0.425$

$$-2(-2320.447 + 2271.656) = 97.583. \qquad \chi^2_{1,95\%} = 3.84$$

- We reject H_0

Test of non-nested hypotheses (cont.)

- Apply the likelihood ratio test for M_2 against M_C
- Null hypothesis: $\beta_{\text{Fare}} = 0.$
- Reject H_0 if $-2\left(L(\hat{\beta}_{M_2}) - L(\hat{\beta}_{M_C})\right) \geq \chi^2_{K,(1-\alpha)}$

M_2 results

Summary statistics
Number of observations=3609
$\mathcal{L}(0) = -3964.892$
$\mathcal{L}(\hat{\beta}) = -2283.103$
$\bar{\rho}^2 = 0.422$

M_C results

Summary statistics
Number of observations=3609
$\mathcal{L}(0) = -3964.892$
$\mathcal{L}(\hat{\beta}) = -2271.656$
$\bar{\rho}^2 = 0.425$

$-2(-2283.103 + 2271.656) = 22.895.$ $\chi^2_{95\%,1} = 3.84$

- We reject H_0

Outline

- Test of non-nested hypotheses
- **Tests of nonlinear specifications**
  - Piecewise linear
  - Power series
  - Box-Cox

Test of Nonlinear Specifications

- Consider a variable x of the model (travel time, say)
- Restricted model: V is a linear function of x
- Unrestricted model: V is a nonlinear function of x
  - Piecewise linear
  - Power series
  - Box-Cox transforms
- For each of them, the linear specification is obtained using simple restrictions on the nonlinear specification

Outline

- Test of non-nested hypotheses
- **Tests of nonlinear specifications**
  - **Piecewise linear**
  - Power series
  - Box-Cox

Piecewise linear specification

- Model
  - Partition the range of values *x* into *M* intervals

    $[a_m, a_{m+1}]$, m=1,…,M
  - For example, the partition [0-2], (2-3], (3-) corresponds to

    M=3, $a_1=0$, $a_2=2$, $a_3=3$, $a_4=+\infty$
  - The slope of the utility function may vary across intervals
  - Therefore, there will be *M* parameters instead of 1
  - The function must be continuous

Piecewise linear specification

- Specifications
  - Linear Specification

$$V_i = \beta x_i + \cdots$$

  - Piecewise linear specification

$$V_i = \sum_{m=1}^{M} \beta_m x_{im} + \cdots$$

where $$x_{im} = \max(0, \min(x - a_m, a_{m+1} - a_m))$$

that is
$$x_{im} = \begin{cases} 0 & \text{if } x < a_m \\ x - a_m & \text{if } a_m \leq x < a_{m+1} \\ a_{m+1} - a_m & \text{if } a_{m+1} \leq x \end{cases}$$

Piecewise linear specification

```
from biogeme.models import piecewiseVariables
thresholds = [Null,2,3,Null]
pw_tt = piecewiseVariables(TripTimeHours_1 ,thresholds)
TripTimeHours_1_1 = DefineVariable('TripTimeHours_1_1', \
pw_tt[0],database)
TripTimeHours_1_2 = DefineVariable('TripTimeHours_1_2', \
pw_tt[1],database)
TripTimeHours_1_3 = DefineVariable('TripTimeHours_1_3', \
pw_tt[2],database)
```

Piecewise linear specification

- The systematic utility expressions used in this model

$$
\begin{aligned}
V_{1n} &= \beta_{\text{Fare}}\text{Fare}_{1n} + \beta_{\text{Legroom}}\text{Legroom}_{1n} \\
&\quad +\beta_{\text{SchedDE}}\text{Opt1\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt1\_SchedDelayLate}_n \\
&\quad \underline{+\beta_{\text{Total\_TT}_1\text{\_1}}\text{Total\_TT\_1}_{1n} + \beta_{\text{Total\_TT}_1\text{\_2}}\text{Total\_TT\_2}_{1n}} \\
&\quad \underline{+\beta_{\text{Total\_TT}_1\text{\_3}}\text{Total\_TT\_3}_{1n}} \\
V_{2n} &= \text{ASC}_2 + \beta_{\text{Fare}}\text{Fare}_{2n} + \beta_{\text{Legroom}}\text{Legroom}_{2n} \\
&\quad +\beta_{\text{SchedDE}}\text{Opt2\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt2\_SchedDelayLate}_n \\
&\quad +\beta_{\text{Total\_TT}_2}\text{Total\_TT}_{2n} \\
V_{3n} &= \text{ASC}_3 + \beta_{\text{Fare}}\text{Fare}_{3n} + \beta_{\text{Legroom}}\text{Legroom}_{3n} \\
&\quad +\beta_{\text{SchedDE}}\text{Opt3\_SchedDelayEarly}_n + \beta_{\text{SchedDL}}\text{Opt3\_SchedDelayLate}_n \\
&\quad +\beta_{\text{Total\_TT}_3}\text{Total\_TT}_{3n}
\end{aligned}
$$

Estimation Results: piecewise linear specification

Estimation results

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	ASC2	-2.33	0.412	-5.65	1.6e-08
2	ASC3	-2.55	0.438	-5.83	5.57e-09
3	Fare	-0.0193	0.000799	-24.1	0.0
4	Legroom	0.227	0.0267	8.51	0.0
5	SchedDE	-0.14	0.0165	-8.47	0.0
6	SchedDL	-0.105	0.0137	-7.64	2.22e-14
7	Total_TT1_1	-0.825	0.238	-3.46	0.000531
8	Total_TT1_2	-0.443	0.188	-2.36	0.0184
9	Total_TT1_3	-0.228	0.0889	-2.57	0.0102
10	Total_TT2	-0.3	0.0701	-4.28	1.84e-05
11	Total_TT3	-0.301	0.0701	-4.29	1.82e-05

Summary statistics

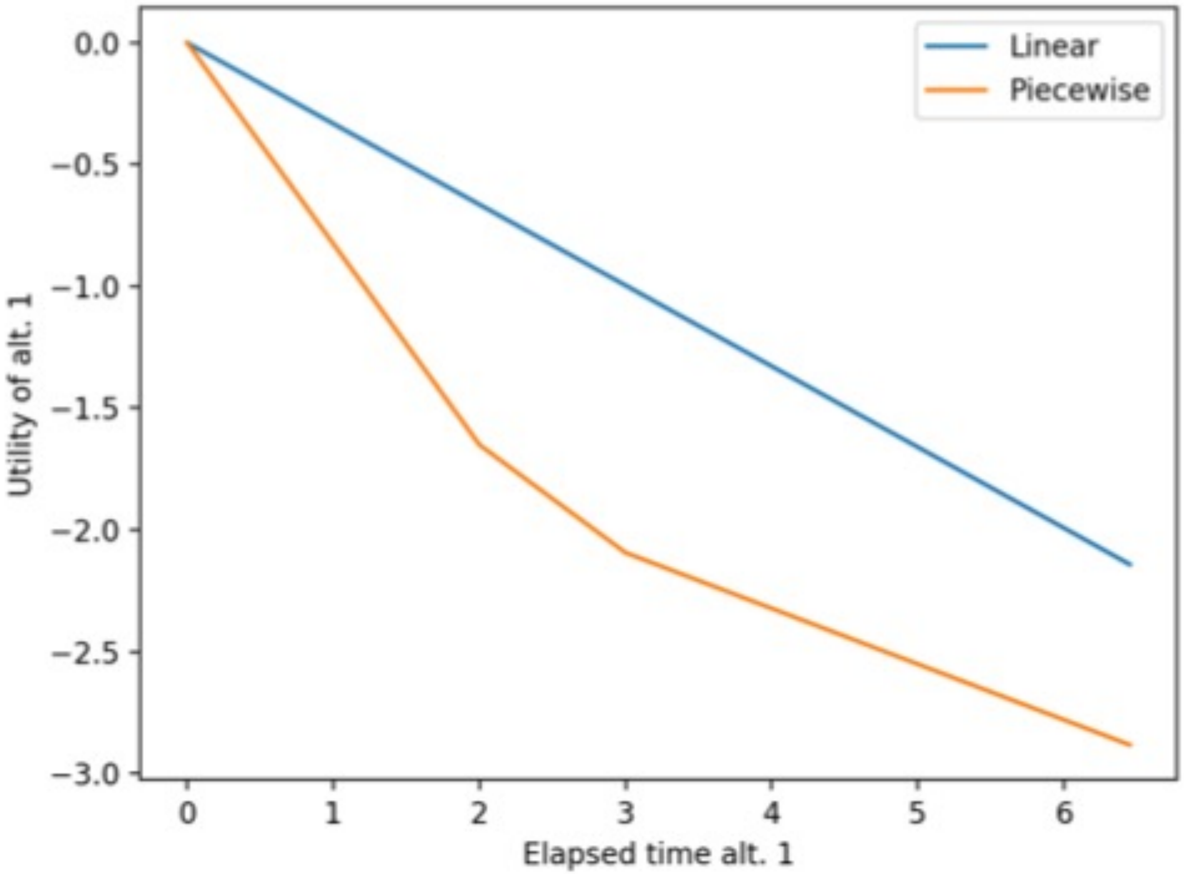
Number of observations=3609

$\mathcal{L}(0) = -3964.892$

$\mathcal{L}(\hat{\beta}) = -2315.041$

$\bar{\rho}^2 = 0.413$

Piecewise linear specification



Estimation Results: linear specification

	Estimation results				
	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	ASC2	-1.43	0.183	-7.81	5.55e-15
2	ASC3	-1.64	0.192	-8.53	0.0
3	Fare	-0.0193	0.000802	-24.0	0.0
4	Legroom	0.226	0.0267	8.45	0.0
5	SchedDE	-0.139	0.0163	-8.53	0.0
6	SchedDL	-0.104	0.0137	-7.59	3.29e-14
7	Total_TT1	-0.332	0.0735	-4.52	6.27e-06
8	Total_TT2	-0.299	0.0696	-4.29	1.77e-05
9	Total_TT3	-0.302	0.0699	-4.31	1.6e-05

Summary statistics

Number of observations=3609

$\mathcal{L}(0) = -3964.892$

$\mathcal{L}(\hat{\beta}) = -2320.447$

$\bar{\rho}^2 = 0.412$

H_0: the linear specification is the correct model

- Test restrictions $\beta_{\text{Total\_TT}_1\text{\_1}} = \beta_{\text{Total\_TT}_1\text{\_2}} = \beta_{\text{Total\_TT}_1\text{\_3}}.$

- Statistic $-2(-2320.447 + 2315.041) = 10.812.$

- Threshold. $\chi^2_{2,0.05} = 5.99$

- The linear specification is rejected

Outline

- Test of non-nested hypotheses
- **Tests of nonlinear specifications**
  - Piecewise linear
  - **Power series**
  - Box-Cox

Power series

- Idea: if the utility function is nonlinear in x, it can be approximated by a polynomial of degree M
- Linear specification

$$V_i = \beta x_i + \cdots$$

- Power series

$$V_i = \sum_{m=1}^{M} \beta_m x_i^m + \cdots$$

Power series specification

Option 1 utility with square and cubic terms

$$
\begin{aligned}
V_{1n} \quad = \quad & \beta_{\text{Fare}}\text{Fare}_{1n} + \beta_{\text{Legroom}}\text{Legroom}_{1n} + \\
& \beta_{\text{SchedDE}}\text{Opt1\_SchedDelayEarly}_{n} + \beta_{\text{SchedDL}}\text{Opt1\_SchedDelayLate}_{n} \\
& +\beta_{\text{Total\_TT}}\text{Total\_TT}_{1n} + \underline{\beta_{\text{Total\_TT\_1\_sq}}\text{Total\_TT}^{2}_{1n}} \\
& \underline{+\beta_{\text{Total\_TT}_1\text{\_cu}}\text{Total\_TT}^{3}_{1n}}
\end{aligned}
$$

Estimation Results: power series specification

	Estimation results				
	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	ASC2	-2.32	0.528	-4.39	1.11e-05
2	ASC3	-2.55	0.55	-4.63	3.66e-06
3	Fare	-0.0193	0.0008	-24.1	0.0
4	Legroom	0.227	0.0267	8.52	0.0
5	SchedDE	-0.14	0.0165	-8.46	0.0
6	SchedDL	-0.105	0.0137	-7.63	2.31e-14
7	Total_TT1	-0.994	0.516	-1.93	0.0542
8	Total_TT1_cu	-0.0036	0.0145	-0.249	0.804
9	Total_TT1_sq	0.113	0.155	0.728	0.466
10	Total_TT2	-0.301	0.0701	-4.29	1.75e-05
11	Total_TT3	-0.302	0.0702	-4.3	1.68e-05

Summary statistics

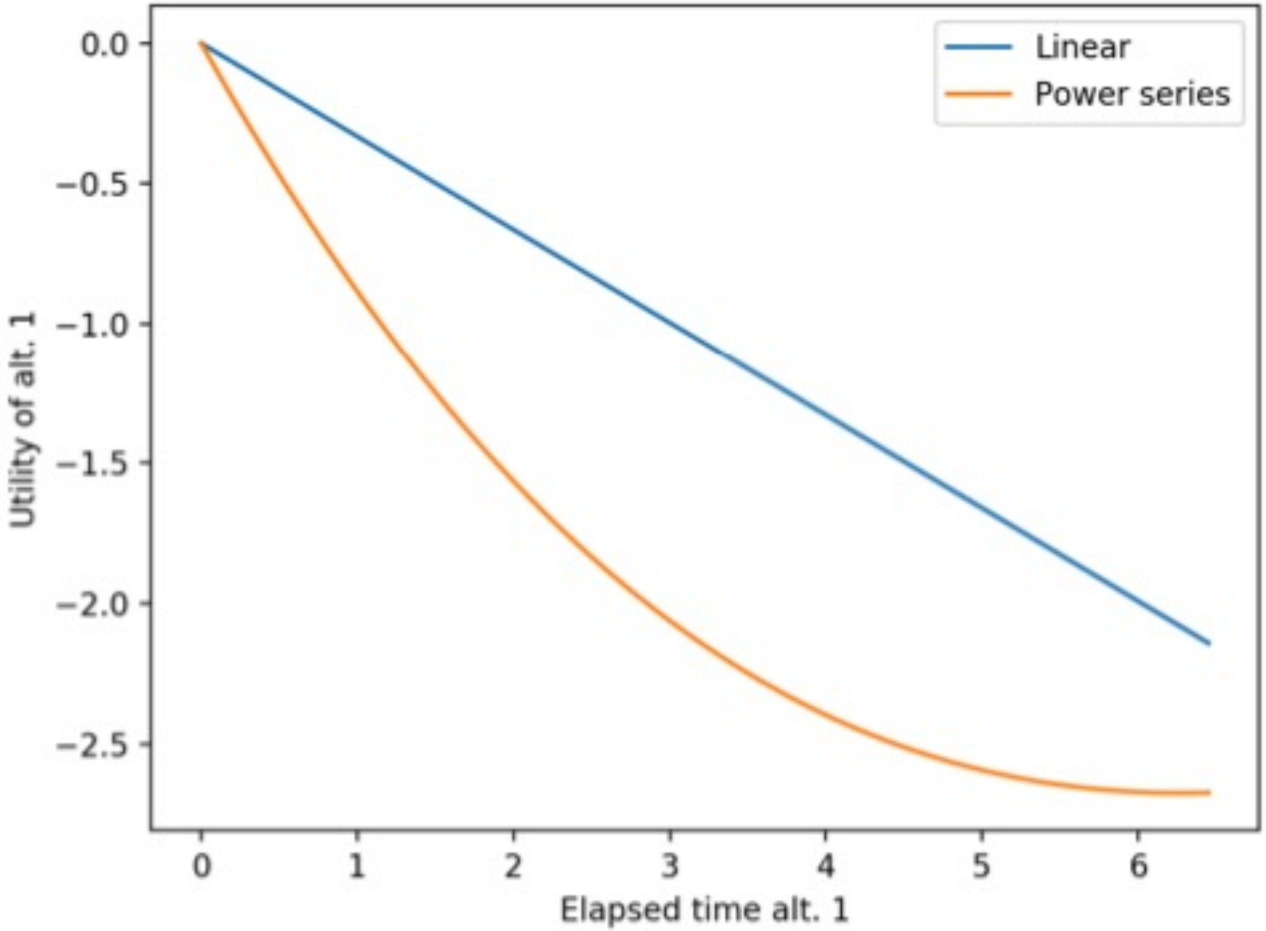
Number of observations=3609

$\mathcal{L}(0) = -3964.892$

$\mathcal{L}(\hat{\beta}) = -2314.402$

$\bar{\rho}^2 = 0.414$

Power series: *M=3*



H_0: the linear specification is the correct model

- Test restrictions $\beta_{\text{Total\_TT}_1\text{\_sq}} = \beta_{\text{Total\_TT}_1\text{\_cu}} = 0.$
- Statistic $-2(-2320.447 + 2314.402) = 12.090.$
- Threshold. $\chi^2_{2,0.05} = 5.9$
- The linear specification is rejected

Outline

- Test of non-nested hypotheses
- **Tests of nonlinear specifications**
  - Piecewise linear
  - Power series
  - **Box-Cox**

Box-Cox transform

- Definition
  - Let $x > 0$ be a positive variable
  - Its Box-Cox transform is defined as

$$B(x,\lambda) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln x & \text{if } \lambda = 0. \end{cases}$$

where $\lambda \in \mathbb{R}$ is a parameter

- Continuity $$\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} = \ln x.$$

Box-Cox transform

- Linear specification

$$V_i = \beta x_i + \cdots$$

- Box-Cox specification

$$V_i = \beta B(x, \lambda) + \cdots$$

- Properties
  - Convex if $\lambda > 1$
  - Linear if $\lambda = 1$
  - Concave if $\lambda < 1$

Box-Cox transform

- Estimation
  - λ is estimated from data
  - Utility function not linear-in-parameters
- Testing the linear specification
  - Restriction: $\lambda = 1$
  - Likelihood ratio test
  - *t*-test can also be used

Box-Cox transform specification

Box-Cox transformation for TT in Option 1

$$
\begin{aligned}
V_1 = \; & \beta_{\text{Fare}}\text{Fare}_1 + \beta_{\text{Legroom}}\text{Legroom}_1 \\
& + \beta_{\text{SchedDE}}\text{Opt1\_SchedDelayEarly} \\
& + \beta_{\text{SchedDL}}\text{Opt1\_SchedDelayLate} + \beta_{\text{Total\_TT}_1}\frac{\text{Total\_TT}_1^{\lambda} - 1}{\lambda}
\end{aligned}
$$

```
from biogeme.models import boxcox
M2_Opt1 = Constant1 + Fare * Fare_1 + Legroom * Legroom_1 \
+ SchedDE * Opt1_SchedDelayEarly \
+ SchedDL * Opt1_SchedDelayLate \
+ Total_TT1 * boxcox(TripTimeHours_1, LAMBDA) \
```

Estimation Results: Box-Cox specification

Estimation results

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	ASC2	-1.51	0.263	-5.76	8.2e-09
2	ASC3	-1.74	0.28	-6.21	5.16e-10
3	Fare	-0.0193	0.000799	-24.1	0.0
4	LAMBDA	-0.139	0.338	-0.412	0.68
5	Legroom	0.227	0.0267	8.52	0.0
6	SchedDE	-0.14	0.0165	-8.47	0.0
7	SchedDL	-0.105	0.0137	-7.63	2.26e-14
8	Total_TT1	-1.24	0.373	-3.34	0.000839
9	Total_TT2	-0.306	0.068	-4.49	7.03e-06
10	Total_TT3	-0.306	0.0682	-4.48	7.34e-06

Summary statistics

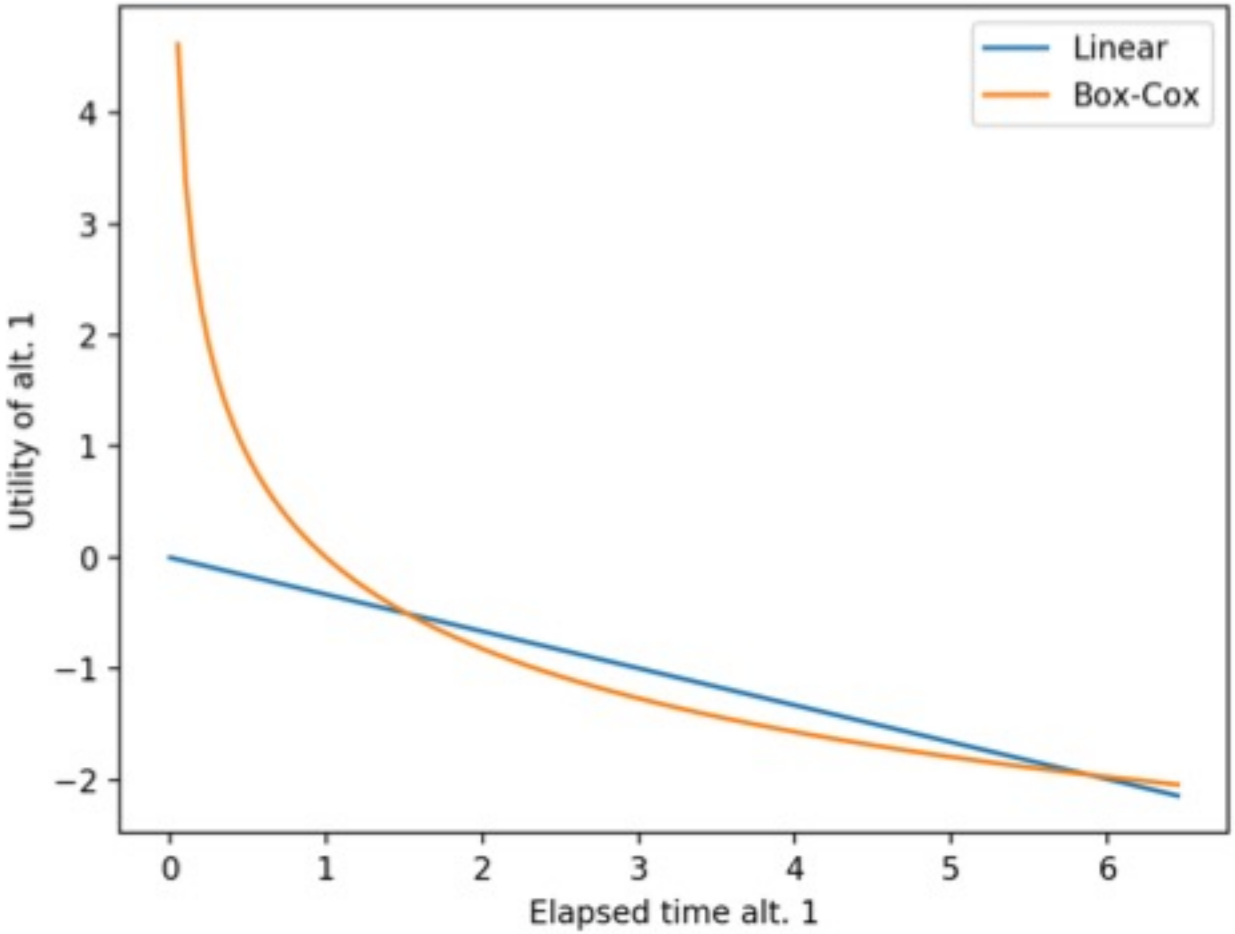
Number of observations=3609

$\mathcal{L}(0) = -3964.892$

$\mathcal{L}(\hat{\beta}) = -2314.574$

$\bar{\rho}^2 = 0.414$

Box-Cox transform



H_0: the linear specification is the correct model

- *t*-test
  - λ = -0.139
  - Robust asymptotic standard error = 0.338
  - $H_0 : \lambda = 1$
  - Test:

$$\frac{-0.139 - 1}{0.338} = -3.37$$

  - The hypothesis can be rejected at the 5% level

H_0: the linear specification is the correct model

- Likelihood ratio test
  - Unrestricted model: -2314.574
  - Restricted model: -2320.447
  - Test: $-2(-2320.447 + 2314.574) = 11.747$
  - Threshold: $\chi^2_{1,0.05} = 3.84$
  - The hypothesis can be rejected at the 5% level