Estimation of Price Response Functions

Outline

- **Linear Demand**
- **Constant Elasticity**
- **Logit**
  - **Given D**
  - **Joint D**
- **Multinomial Logit**

Linear Demand and Linear Regression

- Constant willingness to pay $d(p)=D-mp$
- Estimate market size D and slope m
- Use standard linear regression
  - Excel – Office Button – Excel Options – Add-Ins – Analysis ToolPak – Go
  - Wait for installation
  - Excel Toolbar Data – Click on Data Analysis – Pick Regression
- Given data (d_i,p_i)

$$\min_{D,m} \sum_i (D - mp_i - d_i)^2$$

Use standard Solver function

- Excel – Office Button – Excel Options – Add-Ins – Solver – Go
- Wait for installation
- Excel Toolbar Data – Click on Solver

<see linear.xls>

Constant Elasticity

- Constant elasticity: $d(p)=Cp^{-\varepsilon}$.
- Estimate parameters C and elasticity ε.
- Make log transformation: $\log d(p) = \log C - \varepsilon \log p$.
- Make change of variable: y=log d(p), x=log p, intercept=log C, slope = ε so that y=intercept-slope*x.
- Use standard Linear Regression to find intercept and slope.
- Given data (y_i, x_i) use Solver to find

$$\underset{\text{Intercept, Slope}}{\text{Min}} \sum_i (\text{intercept} - \text{slope} * x_i - y_i)^2$$

<see constant_elasticity.xls>

Logit Price Response Given Market Size *D*

- Logit Price Response function.

$$d(p) = D\left(1 - \frac{1}{1+e^{-(a+bp)}}\right)$$

- For given market size *D,* estimate parameters *a* and *b*.
- Make ln transformation: ln ((D-d(p))/d(p)) = a+bp.
- Make change of variable: y=ln ((D-d(p))/d(p)), x=p, intercept=a, slope = b so that y=intercept+slope*x.
- Use standard Linear Regression to find intercept and slope.
- Given data (y_i,x_i) use Solver to find

$$\underset{\text{Intercept, Slope}}{\text{Min}} \sum_i (\text{intercept} + \text{slope} * x_i - y_i)^2$$

<see logit_givenD.xls>

Logit Price Response Unknown Market Size *D*

- Logit Price Response function.

$$d(p) = D\left(1 - \frac{1}{1+e^{-(a+bp)}}\right)$$

- Estimate parameters *D*, *a* and *b*.
- No log or ln transformation possible to cast the estimation as a linear regression.
- Given data (d_i,p_i) use Solver to find

$$\underset{\text{D, a, b}}{\text{Min}} \sum_i \left(D\left(1 - \frac{1}{1+e^{-a-bp_i}}\right) - d_i \right)^2$$

- Solver does not solve this problem.

Logit Price Response Unknown Market Size D Sequential Approach

- Suppose that D is given.
- Estimate a, b.
- Update D such that it solves

$$\frac{\partial}{\partial D}\sum_i \left(D\left(1-\frac{1}{1+e^{-a-bp_i}}\right)-d_i\right)^2 = 0$$

$$\sum_i 2\left(D\left(1-\frac{1}{1+e^{-a-bp_i}}\right)-d_i\right)\left(1-\frac{1}{1+e^{-a-bp_i}}\right) = 0$$

- So

$$D(d_i, p_i; a, b) = \frac{\sum_i d_i\left(1-\frac{1}{1+e^{-a-bp_i}}\right)}{\sum_i \left(1-\frac{1}{1+e^{-a-bp_i}}\right)^2}$$

- Sequential Algorithm

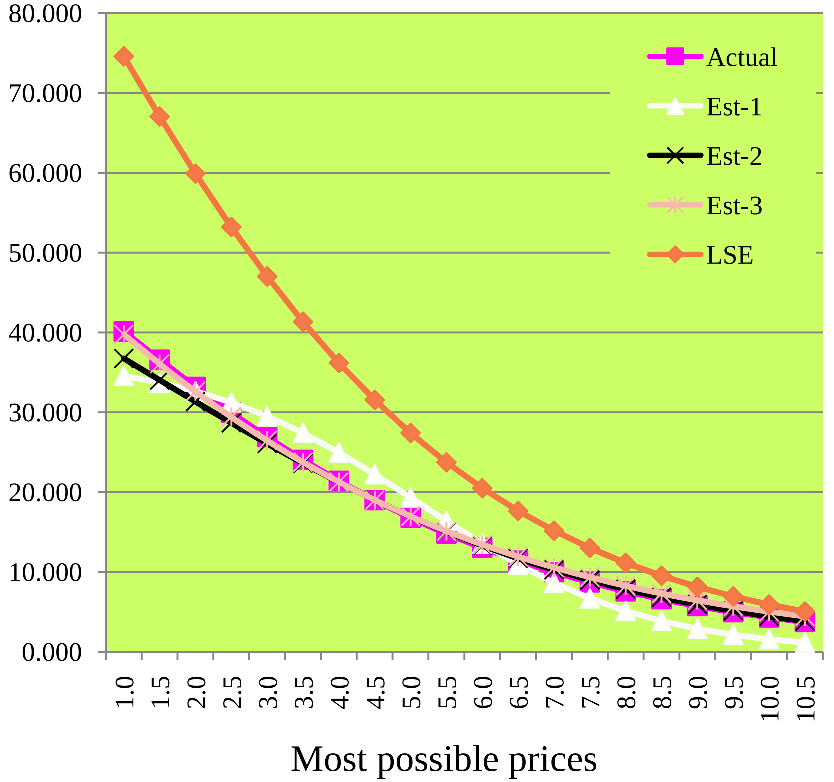
0. Start with an initial guess for D
1. Use current D to estimate a and b by using the procedure of Logit with given D.
2. Update D so that it sets the derivative of the sum of squares of errors to zero
3. Stop or go to 1.

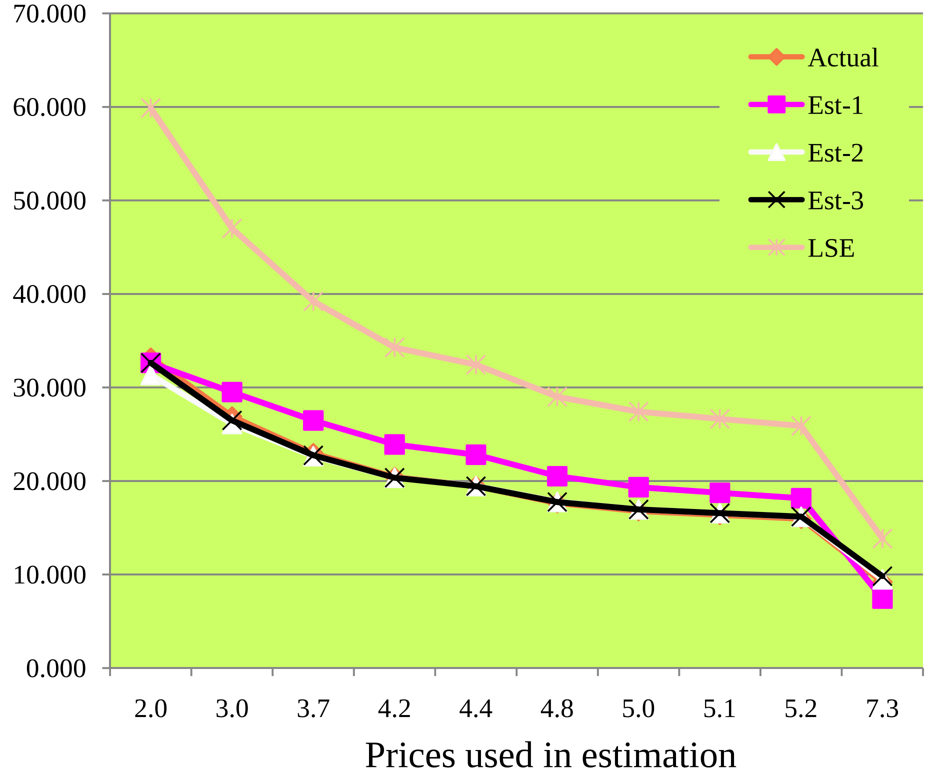
Logit Price Response Sequential Algorithm

- Sequential algorithm is implemented in logit_jointD.xls.
- Unfortunately, $D\text{-}D(d_i,p_i;a,b)$ is small.
- In other words, updates do not change the market size much.
- Hypothesis: a and b are sufficient to represent market size.
- Graph actual demand vs. estimated demand for

	Sequential			
Actual	**Est-1**	**Est-2**	**Est-3**	**LSE**
D=100	37	70	150	200
a=0.1	-3.29	-0.41	0.76	0.19
b=0.3	0.64	0.31	0.26	0.33

Logit Price Response
Sequential Algorithm Graphs





Conclusion: LSE estimates are significantly worse than the Sequential ones.
Reason: Excel Solver fails to minimuize sum of squares, a highly nonlinear objective.
Suggestion: Select 2-3 values of D and use Sequantial algorithm.

Multinomial Logit

Most common consumer choice model is multinomial logit (MNL)

$$f_i(p=[p_1, p_2, ..., p_n]) = \frac{e^{-b_i p_i}}{\sum_{j=1}^{n} e^{-b_j p_j}}$$

Parameters are $b_1, ..., b_n$. Parameter b_i is large when product i is price sensitive.
No standard methodology to estimate parameters.

- We develop our own maximum likelihood estimators
- Given a sample where each individual makes a choice in response to prices $[p_1,p_2,\ldots,p_n]$, what are the most likely values of parameters $[b_1,b_2,\ldots,b_n]$?
- That is, what parameter values maximize the probability of observing the choices?
- Data: For individual k: $[y_{1k},y_{2k},\ldots,y_{nk};\ p_{1k},p_{2k},\ldots,p_{nk}]$.
  - Indicator variable y_{ik}=1 if individual k chooses product i; otherwise zero.

<see Generate and Pure Data sheets of concert.xls>

Maximizing Likelihood Function

- From the multinomial probability frequency function, the likelihood function is

$$L([b_1, b_2, ..., b_n]) = [\text{Constant}] \prod_{i,k} f_i^{y_{ik}} = [\text{Constant}] \prod_{i,k} \left(\frac{e^{-b_i p_i}}{\sum_{j=1}^{n} e^{-b_j p_j}} \right)^{y_{ik}}$$

- We can instead maximize the logarithm of the likelihood

$$\log L([b_1, b_2, ..., b_n]) = \sum_{i,k} y_{ik} \log \left(\frac{e^{-b_i p_i}}{\sum_{j=1}^{n} e^{-b_j p_j}} \right) = \text{sum y} * \text{log(exp/sum) in MLE of concert.xls.}$$

- Use solver to find parameters $[b_1,b_2,\ldots,b_n]$ that maximize $\log L([b_1,b_2,\ldots,b_n])$.

<see MLE sheet of concert.xls>

R Estimation Example: Another Multinomial Example: Fishing Mode

- Consider the Fishing mode and Income example on
  - Page 494 of Microeconometrics: Methods and Applications (2005) by Cameron, Adrian Colin.; Trivedi, P. K. Book available as an ebook from UTD Library.

- Fishing modes are: Beach, Pier, Private Boat, Charter (boat) .
- Individuals choose one of these.
- Each individual reports own income and choice.
- Data: For individual k: [$y_{\text{Beach},k}$, $y_{\text{Pier},k}$, $y_{\text{Private},k}$, $y_{\text{Charter},k}$; income_k]
- Statistical model with parameters *a* and *b*:

$$P(y_{Beach}=1)=\frac{e^{a_{Beach}+b_{Beach}*\text{Income}}}{e^{a_{Beach}+b_{Beach}*\text{Income}}+e^{a_{Pier}+b_{Pier}*\text{Income}}+e^{a_{\text{Private}}+b_{\text{Private}}*\text{Income}}+e^{a_{Charter}+b_{Charter}*\text{Income}}}$$

$$P(y_{Pier}=1)=\frac{e^{a_{Pier}+b_{Pier}*\text{Income}}}{e^{a_{Beach}+b_{Beach}*\text{Income}}+e^{a_{Pier}+b_{Pier}*\text{Income}}+e^{a_{\text{Private}}+b_{\text{Private}}*\text{Income}}+e^{a_{Charter}+b_{Charter}*\text{Income}}}$$

Downloading R

1. Download standard >= R-2.9.2 version from www.r-project.org
2. This should create R-2.9.2 directory and, underneath it, directories

   bin; doc; etc; include; library; modules; share; src; Tcl
3. Among those directories, library is important to us as we shall add estimation specific and other packages to this library. Library directory should have 27 subdirectories.

Downloading Useful Packages

Download the following packages from http://cran.r-project.org/web/packages/

xlsReadWrite.zip, maxLik.zip, mlogit.zip, Ecdat.zip; nls2.zip

into new directories that you create under R-2.9.2\library with names xlsReadWrite, maxLik, mlogit, Ecdat.

The roles of these packages are:

xlsReadWrite; ## Required for im(ex)porting excel available
maxLik; ## Required for Maximum Likelihood Estimation
mlogit; ## Required for Multinomial Logit model
Ecdat; ## Interesting Econometric data files
nls2; ## Nonlinear least squares
Check the directories in R-2.9.2\library

You should have directories 27 standard R directories, plus 5 that you have manually added above. At this point the entire R-2.9.2 directory takes 86,392,832 bytes on my hard disk.

Starting R

1. Start R (click on the icon on your desktop, on the quick start button on your Start menu, or click on R-2.9.2\bin\Rgui.exe). This will start R with a home directory of R-2.9.2.
2. I suggest that you keep your data in a different directory, say C:\Demreman\R\.
3. You have to tell R which directory you want to work in. Go to File menu and then click on to Change dir(ectory)

Make R read your packages

In R, issue commands

```
> library(xlsReadWrite);
 > library(maxLik);
> library(mlogit);
> library(Ecdat, nls2);
> library(nls2);
```

Multinomial Example: Fishing Mode

- Issue command

  > data("Fishing",package="mlogit");

  to read the Fishing data.
- Briefly Fishing data are about 1182 individuals' fishing mode choices. Data come from a survey conducted by Thomson and Crooke (1991).
- Issue

  > fix(Fishing);

  to see what is inside the Fishing dataframe. It has 1 row for each individual and 1182 rows in total.
- It has 12 columns = 3 columns + 4 columns + 4 columns + 1 column.
  - The first 3 columns have the chosen mode of fishing. Its price and the probability of catching a fish.
  - The next four columns have the price for each mode of fishing. These prices change from one individual to another as they can depend on the location and access of the individual.
  - The next four columns have the catch probability for each mode of fishing.
  - The last column contains the monthly income of the individual.

Multinomial Example: Fishing Mode Data Manipulation

- Issue command

  > Fish <- mlogit.data(Fishing,varying=c(4:11),shape="wide",choice="mode")

  to prepare Fishing data for multinomial logit regression.
  - Immediately the number of rows become 4728, so one row for each individual and each alternative. See the Fish dataframe by issuing

    >fix(Fish);
  - The first four rows are now for the first individual whose identity is in the column named chid. The next column is the alternative. By putting individual id and alternative together we obtain the first column, named row.names.
- The important columns for our purpose are named mode and income.
- Manipulation on the model

$$\frac{\mathrm{P}(y_{Pier}=1)}{\mathrm{P}(y_{Beach}=1)}=\frac{e^{a_{Pier}+b_{Pier}*\mathrm{Income}}}{e^{a_{Beach}+b_{Beach}*\mathrm{Income}}}=\exp(a_{Pier}-a_{Beach}+(b_{Pier}-b_{Beach})*Income),$$

$$\log\left(\frac{\mathrm{P}(y_{Pier}=1)}{\mathrm{P}(y_{Beach}=1)}\right)=a_{Pier}-a_{Beach}+(b_{Pier}-b_{Beach})*Income$$

- So a_{Beach} and b_{Beach} can be assumed to be zero to estimate the other parameters with respect to these two.

Multinomial Example: Fishing Mode Logit Regression

- Issue command

  > summary(mlogit(mode~1|income,data=Fish));

  to estimate *a,b* parameters. R outputs

- Call: mlogit(formula = mode ~ 1 | income, data = Fish)
- Frequencies of alternatives:

beach	boat	charter	pier
0.11337	0.35364	0.38240	0.15059

- Newton-Raphson maximisation gradient close to zero. May be a solution 5 iterations, 0h:0m:0s g'(-H)^-1g = 9.47E-30
- Coefficients :

	Estimate	Std. Error	t-value	Pr(>|t|)
altboat	7.3892e-01	1.9673e-01	3.7560	0.0001727 ***
altcharter	1.3413e+00	1.9452e-01	6.8955	5.367e-12 ***
altpier	8.1415e-01	2.2863e-01	3.5610	0.0003695 ***
altboat:income	9.1906e-05	4.0664e-05	2.2602	0.0238116 *
altcharter:income	-3.1640e-05	4.1846e-05	-0.7561	0.4495908
altpier:income	-1.4340e-04	5.3288e-05	-2.6911	0.0071223 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Note that numbers in red are the coefficients in column MNL of Table 15.2 of Cameron and Trivedi (2005).

- Log-Likelihood: -1477.2; McFadden R^2: 0.77463 ; Likelihood ratio test : chisq = 41.145 (p.value=6.0931e-09)

Multinomial Example: Fishing Mode Interpretation of Logit Regression

- Inserting the coefficients

$$\log\left(\frac{P(y_{Boat}=1)}{P(y_{Beach}=1)}\right)=0.739+0.092*Income/1000$$

$$\log\left(\frac{P(y_{Charter}=1)}{P(y_{Beach}=1)}\right)=1.341-0.032*Income/1000$$

$$\log\left(\frac{P(y_{Pier}=1)}{P(y_{Beach}=1)}\right)=0.814-0.143*Income/1000$$

- As the income increases, one is much less likely to fish on the pier, less likely to fish on a charter boat and more likely to fish on a private boat.

Another Example: Transportation Mode

- Issue command

  > data("Mode",package="mlogit");

  to read the transportation Mode data.
- Transportation modes are Car, Carpool, Bus, Rail.
- Data are from 453 individuals.
- Issue

  > fix(Mode)

  to see what is inside the Mode dataframe. It has 1 row for each individual and 453 rows in total.
- It has 12 columns = 1 column + 4 columns + 4 columns.
  – The first column is the chosen mode of transportation.
  – The next four columns are the price of each mode.
  – The next four columns are the duration of each mode.
- Issue command

  > market <- mlogit.data(Mode, alt.levels=c("car", "carpool", "bus", "rail"), shape="wide", choice="choice");

  to prepare Fishing data for multinomial logit regression.

Transportation Mode Results

- Issue command

  > summary(mlogit(choice~1|cost.car+cost.carpool+cost.bus+cost.rail, data=market));

  to estimate parameters. R outputs

- Coefficients :

	Estimate
altcarpool	-5.26903
altbus	-2.07793
altrail	-3.65221
altcarpool:cost.car	0.82269
altbus:cost.car	0.95693
altrail:cost.car	0.98898
altcarpool:cost.carpool	-0.41234
altbus:cost.carpool	0.17596
altrail:cost.carpool	0.13842
altcarpool:cost.bus	0.54418
altbus:cost.bus	-1.74465
altrail:cost.bus	0.21560
altcarpool:cost.rail	-0.47507
altbus:cost.rail	-0.19281
altrail:cost.rail	-1.11535

- Estimations are with respect to Car probability.
- If cost.car increases, the probability of choosing another mode is higher.
- If cost.carpool increases, probability of carpool driops, those of bus and rail increase.
- If cost.bus increases, probability of bus drops, those of carpool and rail increase.
- If cost.rail increases, probability of carpool, bus and rail all drop with respect to car. That is the probability of car increases.

Summary

Outline

- **Linear Demand**
- **Constant Elasticity**
- **Logit**
  - **Given D**
  - **Joint D**
- **Multinomial Logit**
  - **Concert, Fishing and Transportation data**