Lecture 8: Logit Estimation

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Numerical Optimization

Today

- Logit Estimation
- Logit Estimation Example in R

Upcoming

- Reading for next time
  - Greene textbook, Chapters 7.2.2–7.2.8 (Chapters 7.2.2–7.2.6 in the Seventh Edition)
- Problem sets
  - Problem Set 2 will be posted soon, due October 17

Course So Far

So far we have covered

- Discrete choice framework
- Random utility model
- Logit model
- Maximum likelihood estimation
- Numerical optimization

We can finally put all these pieces together and estimate a model ourselves!

Logit Estimation

Logit Likelihood Function

Under the assumptions of the logit model, the probability that decision maker n chooses alternative i

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}$$

We can express the probability of decision maker n choosing the alternative that was chosen, where $y_{ni} = 1$ if and only if n chose i

$$\prod_{i=1}^{J} (P_{ni})^{y_{ni}}$$

Then the likelihood function for the entire sample is

$$L(\beta) = \prod_{n=1}^{N} \prod_{i=1}^{J} (P_{ni})^{y_{ni}}$$

Logit Log-Likelihood Function

Then the logit log-likelihood function is

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{J} y_{ni} \ln P_{ni}$$

The MLE is the set of parameters, $\hat{\beta}$, that maximize this function

Logit Moment Condition

The first-order condition for maximization is

$$\frac{\partial LL(\beta)}{\partial \beta} = 0$$

If we assume representative utility is linear, $V_{ni} = \beta' x_{ni}$, (and after some substitution and simplification) this first-order condition is equivalent to

$$\sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) x_{ni} = 0$$

This is the sample covariance of the "residuals," $y_{ni} - P_{ni}$, and the data, x_{ni}

- This is a moment condition that we could use for GMM
  - More on that in a couple weeks...

Non-Random Sampling of Decision Makers

The previous slides all require that sampling is random (or at least exogenous to the choice)

- If your sample is endogenous to the choice, then coefficients may be inconsistent
- Example: What could go wrong if you take a survey of commute choices, but you take the survey at the bus stop?

You can recover consistent estimates from a non-random sample, but that requires a more complex estimation procedure

Logit Estimation Example in R

Multinomial Logit Estimation Example

We are again studying how consumers make choices about expensive and highly energy-consuming systems in their homes. We have data on 900 households in California and the type of heating system in their home. Each household has the following choice set, and we observe the following data

Choice set

- GC: gas central
- GR: gas room
- EC: electric central
- ER: electric room
- HP: heat pump

Alternative-specific data

- IC: installation cost
- OC: annual operating cost

Household demographic data

- income: annual income
- agehed: age of household head
- rooms: number of rooms
- region: home location

Load Dataset

```r
## Load tidyverse and mlogit
library(tidyverse)
library(mlogit)
## Load dataset from mlogit package
data('Heating', package = 'mlogit')
```

Dataset

```
## Look at dataset
as_tibble(Heating)
## # A tibble: 900 x 16
##    idcase depvar ic.gc ic.gr ic.ec ic.er ic.hp oc.gc oc.gr oc.ec oc.er
##     <dbl> <fct>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1      1 gc       866  963.  860.  996. 1136.  200.  152.  553.  506.
##  2      2 gc      728.  759.  797.  895.  969.  169.  169.  520.  486.
##  3      3 gc      599.  783.  720.  900. 1048.  166.  138.  439.  405.
##  4      4 er      835.  793.  761.  831. 1049.  181.  147.  483   425.
##  5      5 er      756.  846.  859.  986.  883.  175.  139.  404.  390.
##  6      6 gc      666.  842.  694.  863.  859.  136.  141.  398.  371.
##  7      7 gc      670.  941.  634.  952. 1087.  192.  148.  478.  446.
##  8      8 gc      778. 1022.  813. 1012.  990.  188.  159.  502.  465.
##  9      9 gc      928. 1212.  876. 1025. 1232.  169.  190.  553.  452.
## 10     10 gc      683. 1045.  776.  874.  878.  176.  136.  532.  472.
## # ... with 890 more rows, and 5 more variables: oc.hp <dbl>,
## #   income <dbl>, agehed <dbl>, rooms <dbl>, region <fct>
```

Three Models to Estimate

Base model

$$U_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj} + \varepsilon_{nj}$$

Heterogeneous coefficients model

$$U_{nj} = \alpha_j + \frac{\beta_1}{I_n} IC_{nj} + \frac{\beta_2}{I_n} OC_{nj} + \varepsilon_{nj}$$

Alternative-specific coefficients model

$$U_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj} + \beta_{3j} R_n + \varepsilon_{nj}$$

Optimization in R

```r
### Optimization in R
## Help file for the optimization function, optim
?optim
## Arguments for optim function
optim(par, fn, gr, ..., method, lower, upper, control, hessian)
```

Function to Calculate Log-Likelihood

```
### Calculate MLE for multinomial logit heating choice using cost data and
### alternative effects
## Function to calculate log-likelihood using cost data and alternative effects
log_likelihood_wide <- function(parameters, data){
  ## Extract individual parameters
  alpha_1 <- parameters[1]
  alpha_2 <- parameters[2]
  alpha_3 <- parameters[3]
  alpha_4 <- parameters[4]
  beta_1 <- parameters[5]
  beta_2 <- parameters[6]
  ## Calculate utility for each alternative given the parameters
  data <- data %>%
    mutate(utility_ec = alpha_1 + beta_1 * ic.ec + beta_2 * oc.ec,
           utility_er = alpha_2 + beta_1 * ic.er + beta_2 * oc.er,
           utility_gc = alpha_3 + beta_1 * ic.gc + beta_2 * oc.gc,
           utility_gr = alpha_4 + beta_1 * ic.gr + beta_2 * oc.gr,
           utility_hp = beta_1 * ic.hp + beta_2 * oc.hp)
  ## Calculate logit probability denominator given the parameters
  data <- data %>%
    mutate(probability_denominator = exp(utility_ec) + exp(utility_er) +
             exp(utility_gc) + exp(utility_gr) + exp(utility_hp))
  ## Calculate logit probability numerator given the parameters
  data <- data %>%
    mutate(probability_numerator = exp(utility_ec) * (depvar == 'ec') +
             exp(utility_er) * (depvar == 'er') +
             exp(utility_gc) * (depvar == 'gc') +
             exp(utility_gr) * (depvar == 'gr') +
             exp(utility_hp) * (depvar == 'hp'))
  ## Calculate log of logit choice probability given the parameters
  data <- data %>%
    mutate(probability_choice = probability_numerator / probability_denominator,
           log_probability_choice = log(probability_choice))
  ## Calculate the log-likelihood for these parameters
```

Estimate Logit Model

```
## Maximize the log-likelihood function
model_wide <- optim(rep(0, 6), log_likelihood_wide, data = Heating,
                    method = 'BFGS', hessian = TRUE)
```

Estimation Results

```
## Show optimization results
model_wide
## $par
## [1]  1.811460030  1.986682245  1.661661689  0.255606904 -0.001603602
## [6] -0.007689238
##
## $value
## [1] 1008.337
##
## $counts
## function gradient
##       76       12
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##              [,1]         [,2]         [,3]         [,4]         [,5]
## [1,]    58.682876    -6.268791    -39.79257    -9.042317    -974.0483
## [2,]    -6.268791    74.971017    -52.12634   -11.874898    11305.5416
## [3,]   -39.792565   -52.126345    205.37492   -81.829543   -31025.7987
## [4,]    -9.042317   -11.874898    -81.82954   109.932020    10530.5731
```

Maximum Likelihood Estimator and Hessian

```
## Show MLE parameters
model_wide$par
## [1]  1.811460030  1.986682245  1.661661689  0.255606904 -0.001603602
## [6] -0.007689238

## Show Hessian at the MLE
model_wide$hessian
##              [,1]         [,2]         [,3]         [,4]         [,5]
## [1,]    58.682876    -6.268791    -39.79257    -9.042317    -974.0483
## [2,]    -6.268791    74.971017    -52.12634   -11.874898   11305.5416
## [3,]   -39.792565   -52.126345    205.37492   -81.829543  -31025.7987
## [4,]    -9.042317   -11.874898    -81.82954   109.932020   10530.5731
## [5,]  -974.048313 11305.541577 -31025.79874 10530.573092 8813294.1157
## [6,] 15359.360795 16493.309579 -24006.72752 -7914.243457 2781602.8033
##             [,6]
## [1,]   15359.361
## [2,]   16493.310
## [3,]  -24006.728
## [4,]   -7914.243
## [5,] 2781602.803
## [6,] 9017721.751
```

Variance-Covariance Matrix

```
## Calculate MLE variance-covariance matrix
model_wide$hessian %>%
  solve()
##                [,1]          [,2]         [,3]          [,4]
## [1,]  1.980437e-01  1.424109e-01 1.860571e-02 -5.521102e-03
## [2,]  1.424109e-01  1.284812e-01 7.940390e-03 -5.846949e-03
## [3,]  1.860571e-02  7.940390e-03 5.124370e-02  3.792083e-02
## [4,] -5.521102e-03 -5.846949e-03 3.792083e-02  4.244292e-02
## [5,]  9.512776e-05  3.493597e-05 9.747095e-05  4.399295e-05
## [6,] -5.824413e-04 -4.723201e-04 9.342135e-05  1.447288e-04
##                [,5]          [,6]
## [1,]  9.512776e-05 -5.824413e-04
## [2,]  3.493597e-05 -4.723201e-04
## [3,]  9.747095e-05  9.342135e-05
## [4,]  4.399295e-05  1.447288e-04
## [5,]  3.843712e-07 -4.639197e-08
## [6,] -4.639197e-08  2.356832e-06
```

Standard Errors

```
## Calculate MLE standard errors
model_wide$hessian %>%
  solve() %>%
  diag() %>%
  sqrt()
## [1] 0.4450210320 0.3584427563 0.2263707190 0.2060168046 0.0006199767
## [6] 0.0015351975
```

Heterogeneous Coefficients

```r
### Calculate MLE for multinomial logit heating choice using cost data scaled
### by income and alternative effects
## Function to calculate log-likelihood using two variables and four
## alternative effects
log_likelihood_long <- function(parameters, data){
  ## Extract individual parameters
  alphas <- parameters[1:4]
  beta_1 <- parameters[5]
  beta_2 <- parameters[6]
  ## Assign constant parameters to alternatives
  data <- data %>%
    group_by(idcase) %>%
    arrange(idcase, alt) %>%
    mutate(constant = c(alphas, 0)) %>%
    ungroup()
  ## Calculate utility for each alternative given the parameters
  data <- data %>%
    mutate(utility = constant + beta_1 * var_1 + beta_2 * var_2)
  ## Calculate logit probability denominator given the parameters
  data <- data %>%
    group_by(idcase) %>%
    mutate(probability_denominator = sum(exp(utility))) %>%
    ungroup()
  ## Filter on only the alternative that was chosen
  data <- data %>%
    filter(choice == TRUE)
  ## Calculate log of logit choice probability given the parameters
  data <- data %>%
    mutate(probability_choice = exp(utility) / probability_denominator,
           log_probability_choice = log(probability_choice))
  ## Calculate the log-likelihood for these parameters
  log_likelihood <- sum(data$log_probability_choice)
  return(-log_likelihood)
}
```

Heterogeneous Coefficients Optimization

```r
## Gather heating dataset into a long dataset
heating_long <- Heating %>%
  gather(key, value, starts_with('ic.'), starts_with('oc.')) %>%
  separate(key, c('cost', 'alt')) %>%
  spread(cost, value) %>%
  mutate(choice = (depvar == alt)) %>%
  select(-depvar) %>%
  mutate(var_1 = ic / income, var_2 = oc / income)
## Maximize the log-likelihood function
model_long <- optim(rep(0, 6), log_likelihood_long, data = heating_long,
                    method = 'BFGS', hessian = TRUE)
```

Heterogeneous Coefficients MLE Results

```r
## Show MLE parameters
model_long$par
## [1]  0.459588535  0.758278893  2.220618539  0.788015418 -0.002274941
## [6] -0.005378071

## Calculate MLE standard errors
model_long$hessian %>%
  solve() %>%
  diag() %>%
  sqrt()
## [1] 0.286591829 0.230307487 0.194141433 0.179223237 0.001958449
## [6] 0.002754184
```

Alternative-Specific Coefficients

```r
### Calculate MLE for multinomial logit heating choice using cost data,
### alternative effects, and alternative-specific coefficients on rooms
## Function to calculate log-likelihood using flexible matrices
log_likelihood_matrix <- function(parameters, data){
  ## Extract explanatory variables
  data_x <- data %>%
    map(1)
  ## Extract choice data
  data_y <- data %>%
    map(2)
  ## Calculate utility for each alternative given the parameters
  utility <- data_x %>%
    map(~ .x %*% parameters)
  ## Calculate logit probability denominator given the parameters
  probability_denominator <- utility %>%
    map(~ sum(exp(.x))) %>%
    unlist()
  ## Calculate logit probability numerator given the parameters
  probability_numerator <- utility %>%
    map2(data_y, ~ exp(sum(.x * .y))) %>%
    unlist()
  ## Calculate logit choice probability given the parameters
  probability_choice <-  probability_numerator / probability_denominator
  ## Calculate log of logit choice probability given the parameters
  log_probability_choice <- log(probability_choice)
  ## Calculate the log-likelihood for these parameters
  log_likelihood <- sum(log_probability_choice)
  return(-log_likelihood)
}
```

Alternative-Specific Coefficients Estimation Data

```r
## Split heating dataset into list of household data frames
heating_split <- heating_long %>%
  group_by(idcase) %>%
  arrange(idcase, alt) %>%
  group_split()
## Crate matrix of dummy variables (intercepts) to bind to data
constant_matrix <- diag(4) %>%
  rbind(c(0, 0, 0, 0))
## Function to create list of datasets for estimation
create_data_matrix <- function(data){
  data_x <- data %>%
    select(ic, oc) %>%
    as.matrix()
  data_x <- constant_matrix %>%
    cbind(data_x)
  rooms <- data$rooms[1]
  data_x <-
    data_x %>%
    cbind(rooms * constant_matrix)
  data_y <- data %>%
    select(choice) %>%
    mutate(choice = 1 * choice) %>%
    pull(choice)
  return(list(x = data_x, y = data_y))
}
## Create list of datasets for estimation
heating_matrix <- heating_split %>%
  map(.x = ., .f = ~ create_data_matrix(.x))
```

Alternative-Specific Coefficients Optimization

```r
## Maximize the log-likelihood function
model_matrix <- optim(rep(0, 10), log_likelihood_matrix,
                      data = heating_matrix,
                      method = 'BFGS', hessian = TRUE)
```

Alternative-Specific Coefficients MLE Results

```r
## Show MLE parameters
model_matrix$par
##  [1]  1.628301941  1.913566330  1.741159930  0.418940436 -0.001591614
##  [6] -0.007539289  0.032805692  0.013956764 -0.014443707 -0.032992137

## Calculate MLE standard errors
model_matrix$hessian %>%
  solve() %>%
  diag() %>%
  sqrt()
##  [1] 0.6708095328 0.5862586030 0.4458930224 0.4732882974 0.0006203237
##  [6] 0.0015350175 0.1086191569 0.1027954371 0.0850252568 0.0959267775
```

Likelihood Ratio Test for Joint Significance

```
### Test that alternative-specific coefficients are jointly significant
## Calculate likelihood ratio test statistic
lr_test_statistic <- -2 * (model_matrix$value - model_wide$value)
lr_test_statistic
## [1] 0.8856428

## Find chi-squared critical value for df = 4
qchisq(0.95, 4)
## [1] 9.487729

## Test if likelihood ratio test statitsic is greater than critical value
lr_test_statistic > qchisq(0.95, 4)
## [1] FALSE

## Calculate p-value of test
1 - pchisq(lr_test_statistic, 4)
## [1] 0.9266115
```

Bayer et al. (2009)

Bayer et al. (2009)

Research question

- How much do people value air quality in the place where they choose to live?

Empirical methods

- Model the location choice of households as a multinomial logit model to recover composite city-level valuations
- Regress the composite city-level value on city attributes, including PM10 concentration, to estimate the value of air quality

Results

- Willingness to pay for air quality is \$149–\$185 per $\mu g/m^3$ of PM10, or an elasticity of 0.34–0.42

Residential Location Choice

The hedonic pricing model is the standard model used to estimate the value of air quality

- The hedonic model is based on the idea that a home is a bundle of attributes, each of which contributes to the price of the home, so we can estimate attribute values from home prices
- An underlying assumption of this model is that moving is costless

An alternative approach developed by Bayer et al. uses multinomial logit

- The choice of where to live is clearly a discrete choice
- But air pollution is endogenous, and it is challenging to instrument within a multinomial model
- Bayer et al. develop this approach to first estimate a composite city-specific valuation, and then regress this value on attributes, which does allow for IV

General Comments and Questions

Clever combination of structural estimation (first stage) and a linear IV model (second stage)

- This is becoming more common in the absence of exogenous variation: recover structural parameters and then deal with endogeneity
- Or you could use GMM...more on this later

Some assumptions and implications of this model

- Income counterfactuals can be estimated from others with similar characteristics
- All households spend the same fraction of income on housing (20%)
- Standard logit assumption of i.i.d. errors

Do we believe

- These assumptions?
- The IV strategy?

Announcements

Reading for next time

- Greene textbook, Chapters 7.2.2–7.2.8 (Chapters 7.2.2–7.2.6 in the Seventh Edition)

Upcoming

- Problem Set 2 will be posted soon, due October 17