Lecture 4: Logit Model I

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Random Utility Model

Today

- Logit Model
- Binary Logit Model
- Some Logit Properties
- Binary Logit Model Example in R
- Gruber and Poterba (1994)

Upcoming

- Reading for next time
  - Adamowicz et al. (1994)
- Problem set
  - Problem Set 1 is posted, due September 24

Random Utility Model Recap

A decision maker chooses the alternative that maximizes utility

- A decision maker, n, faces a choice among J alternatives
- Alternative j provides utility U_{nj} (where $j = 1, \ldots, J$)
- n chooses i if and only if $U_{ni} > U_{nj} \; \forall j \neq i$

We model utility as having observed and unobserved components

- Observed factors: V_{nj}
- Unobserved factors: ε_{nj}

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

The probability the decision maker chooses alternative i is

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj} \; \forall j \neq i) \\ &= \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \; \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \end{aligned}$$

Logit Model

Logit Model

The logit model makes a (somtimes overly) simple assumption about the joint density of unobserved utility, $f(\varepsilon_n)$

$$\varepsilon_{nj} \sim \text{i.i.d. type I extreme value (Gumbel) with } Var(\varepsilon_{nj}) = \frac{\pi^2}{6}$$

Why make this assumption about the unobserved utility of alternatives?

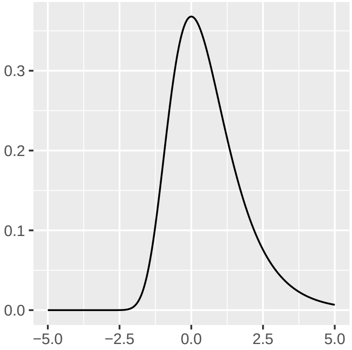
- It yields choice probabilities that have a closed-form expression

Type I Extreme Value Density and Distribution

Type I extreme value is similar to a normal distribution but with a fatter tail on one side

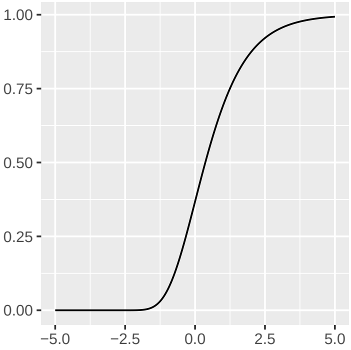
Probability density

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}$$



Cumulative distribution

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$$

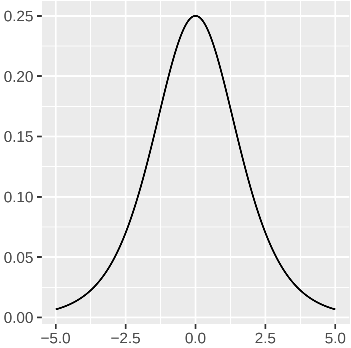


Logistic Density and Distribution

The difference of two type I extreme value draws, $\varepsilon^*_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$, follows the logistic distribution

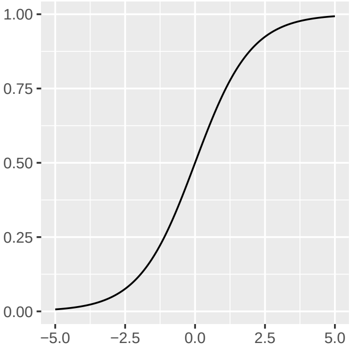
Probability density

$$f(\varepsilon^*_{nji}) = \frac{e^{\varepsilon^*_{nji}}}{\left(1 + e^{\varepsilon^*_{nji}}\right)^2}$$



Cumulative distribution

$$F(\varepsilon^*_{nji}) = \frac{e^{\varepsilon^*_{nji}}}{1 + e^{\varepsilon^*_{nji}}}$$



Logit Choice Probabilities

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj} \ \forall j \neq i) \\ &= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i) \\ &= \Pr(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \ \forall j \neq i) \end{aligned}$$

For a given ε_{ni}, this is the cumulative distribuion of each ε_{nj}, and because ε_{nj} is i.i.d.

$$P_{ni} \mid \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}}$$

But ε_{ni} is random, so we have to integrate over the density of ε_{ni}

$$\begin{aligned} P_{ni} &= \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \\ &= \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \end{aligned}$$

The probability of n choosing i is a closed-form expression that depends on the representative utility (or observable components) of all alternatives

Representative Utility

We usually specify representative utility as a linear function of observable characteristics of the alternative and the agent

$$V_{nj} = \beta' x_{nj}$$

- A linear function is highly flexible and can include interactions, squared terms, etc.
- Most utility functions can be closely approximated by a function that is linear in parameters
- Non-linear utility can greatly complicate estimation

With linear representative utility, the logit choice probability is

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

Properties of Logit Choice Probabilities

P_{ni} is always within the range $(0, 1)$

- $P_{ni} \to 1$ as $V_{ni} \to \infty$
- $P_{ni} \to 0$ as $V_{ni} \to -\infty$

Choice probabilities sum to 1

$$\sum_{i=1}^{J} P_{ni} = \sum_{i=1}^{J} \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} = \frac{\sum_i e^{V_{ni}}}{\sum_j e^{V_{nj}}} = 1$$

Choice probability is a sigmoidal function of representative utility (see the logistic CDF)

- Marginal effects are small when probabilities are close to 0 or 1
- Marginal effects are largest when $P_{ni} = 0.5$

Binary Logit Model

Binary Logit Model

With only two choices, the logit choice probabilities simplify to

$$P_{n1} = \frac{e^{V_{n1}}}{e^{V_{n1}} + e^{V_{n2}}}$$
$$P_{n2} = \frac{e^{V_{n2}}}{e^{V_{n1}} + e^{V_{n2}}}$$

The log odds ratio of alternative 1 is

$$\ln\left(\frac{P_{n1}}{1 - P_{n1}}\right) = V_{n1} - V_{n2}$$

If we assume representative utility is linear, $V_{ni} = \beta' x_{ni}$

$$\ln\left(\frac{P_{n1}}{1 - P_{n1}}\right) = \beta'(x_{n1} - x_{n2})$$

The log odds ratio of each alternative is a linear function of β and x_n

- `glm()` with `family = "binomial"` can estimate this model

Binary Logit Choice Probabilities Example

A person chooses whether to take a car (c) or a bus (b) to work

- We observe the time, T, and cost, M, of each choice

We specify the representative utility of each alternative as

$$V_c = \alpha T_c + \beta M_c$$
$$V_b = \alpha T_b + \beta M_b$$

Using the logit model, the probability of driving is

$$P_c = \frac{e^{\alpha T_c + \beta M_c}}{e^{\alpha T_c + \beta M_c} + e^{\alpha T_b + \beta M_b}}$$

and the log odds ratio of driving is

$$\ln\left(\frac{P_c}{1 - P_c}\right) = (\alpha T_c + \beta M_c) - (\alpha T_b + \beta M_b)$$

Some Logit Properties

Marginal Effects

Unlike a linear probability model, the coefficients of a logit model cannot be interpreted as marginal effects on probability

- Instead, they give the marginal utility of the corresponding variable
- We know the choice probabilities, so we can derive marginal effects!

The marginal effect of z_{ni}, an observed factor of alternative i, on P_{ni}, the probability that agent n chooses alternative i

$$\begin{aligned} \frac{\partial P_{ni}}{\partial z_{ni}} &= \frac{\partial \left(e^{V_{ni}} / \sum_j e^{V_{nj}} \right)}{\partial z_{ni}} \\ &= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni}) \end{aligned}$$

If V_{ni} is linear in z_{ni} with coefficient β_z, then the marginal effect is

$$\frac{\partial P_{ni}}{\partial z_{ni}} = \beta_z P_{ni}(1 - P_{ni})$$

Cross Marginal Effects

The marginal effect of z_{nj}, an observed factor of alternative j, on P_{ni}, the probability that agent n chooses alternative i

$$\begin{aligned}\frac{\partial P_{ni}}{\partial z_{nj}} &= \frac{\partial \left(e^{V_{ni}} / \sum_k e^{V_{nk}}\right)}{\partial z_{nj}} \\ &= -\frac{\partial V_{nj}}{\partial z_{nj}} P_{ni} P_{nj}\end{aligned}$$

If V_{nj} is linear in z_{nj} with coefficient β_z, then the marginal effect is

$$\frac{\partial P_{ni}}{\partial z_{nj}} = -\beta_z P_{ni} P_{nj}$$

In a binary logit model, these marginal effect expressions are negatives of each another

Elasticities

Sometimes elasticities are more informative than marginal effects, especially when considering price changes

The elasticity of P_{ni}, the probability that agent n chooses alternative i, with respect to z_{ni}, an observed factor of alternative i

$$\begin{aligned} E_{iz_{ni}} &= \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} \\ &= \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni}(1 - P_{ni}) \end{aligned}$$

If V_{nj} is linear in z_{nj} with coefficient β_z, then the elasticity is

$$E_{iz_{ni}} = \beta_z z_{ni}(1 - P_{ni})$$

Cross Elasticities

The elasticity of P_{ni}, the probability that agent n chooses alternative i, with respect to z_{nj}, an observed factor of alternative j

$$
\begin{aligned}
E_{iz_{nj}} &= \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} \\
&= -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj}
\end{aligned}
$$

If V_{nj} is linear in z_{nj} with coefficient β_z, then the elasticity is

$$E_{iz_{nj}} = -\beta_z z_{nj} P_{nj}$$

This elasticity depends only on features of alternative j and not on features of alternative i

Taste Variation

Decision makers' tastes can vary for many reasons, some of which are observable and others are not

- The logit model can only explicitly capture taste variation due to observable characteristics

Consider some sources of taste variation in the car vs. bus example

- Some people hate driving and some people love it, but we do not directly observe this preference
  - We cannot explicitly include this taste variation in the model ... yet!
- People with higher incomes care less about the cost of each alternative

$$\beta_n = \frac{\beta}{I_n} \quad \Rightarrow \quad U_{nc} = \alpha T_{nc} + \beta \frac{M_{nc}}{I_n} + \varepsilon_{nc}$$

Scale Parameter

We assume the utility of unobserved factors has variance $\pi^2/6$

- We can use a scale parameter, σ, to allow for a different variance

Suppose the unobserved portion of utility actually has variance $\sigma^2 \times (\pi^2/6)$

$$U^*_{nj} = V_{nj} + \varepsilon^*_{nj}$$

Dividing by σ gives a scaled model

$$U_{nj} = \frac{V_{nj}}{\sigma} + \varepsilon_{nj} \text{ where } \varepsilon_{nj} = \frac{\varepsilon^*_{nj}}{\sigma}$$

The variance of this scaled unobserved component is

$$Var(\varepsilon_{nj}) = \frac{1}{\sigma^2} Var(\varepsilon^*_{nj}) = \frac{\pi^2}{6}$$

Logit Choice Probabilities with a Scale Parameter

In the scaled model, choice probabilities are

$$P_{ni} = \frac{e^{V_{ni}/\sigma}}{\sum_j e^{V_{nj}/\sigma}}$$

If V_{nj} is linear in parameters with coefficients β^*

$$P_{ni} = \frac{e^{(\beta^*/\sigma)' x_{ni}}}{\sum_j e^{(\beta^*/\sigma)' x_{nj}}}$$

But β^* and σ are not separately identified, so we can only estimate their ratio, $\beta = \beta^*/\sigma$, which gives the standard logit expression

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

Parameters are estimated relative to the variance of unobserved utility

Heteroskedasticity and the Scale Parameter

Different subsets of decision makers may each have a different variance of unobserved utility

- We can use scale parameters to account for this group-wise heteroskedasticity
- We can estimate the relative scale parameters of each group compared to one reference group

Suppose we have commute data for both Amherst (A) and Boston (B)

- The scale parameters for each city are σ^A and σ^B with $k = (\sigma^B/\sigma^A)^2$

$$\text{Amherst:} \quad P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

$$\text{Boston:} \quad P_{ni} = \frac{e^{(\beta/\sqrt{k})' x_{ni}}}{\sum_j e^{(\beta/\sqrt{k})' x_{nj}}}$$

Binary Logit Model Example in R

Binary Logit Model Example

We are studying how consumers make choices about expensive and highly energy-consuming appliances in their homes. We have data on whether they choose to purchase a window air conditioning unit. For each household, we observe the purchase price of the air conditioner and its annual operating cost. (To simplify things, we assume there is only one "representative" air conditioner for each household and how much the household operates the air conditioner is fixed.) We can use a binary logit model to see how the purchase price and the operating cost affect the decision to purchase.

$$U_n = \beta_0 + \beta_1 P_n + \beta_2 C_n + \varepsilon_n$$

$$Y_n = \begin{cases} 1 & \text{if } U_n > 0 \\ 0 & \text{otherwise} \end{cases}$$

Load Dataset

```r
### Load and look at dataset
## Load tidyverse
library(tidyverse)
```

```r
## Load dataset
data <- read_csv('ac_renters.csv')

## Parsed with column specification:
## cols(
##  air_conditioning = col_logical(),
##  cost_system = col_double(),
##  cost_operating = col_double(),
##  income = col_double(),
##  residents = col_double(),
##  city = col_double()
## )
```

Dataset

```
## Look at dataset
data
## # A tibble: 600 x 6
##    air_conditioning cost_system cost_operating income residents  city
##    <lgl>                  <dbl>          <dbl>  <dbl>     <dbl> <dbl>
##  1 FALSE                    620            258     74         1     1
##  2 FALSE                    685            141     74         1     1
##  3 FALSE                    570            152     57         1     1
##  4 TRUE                     497            193     81         1     1
##  5 TRUE                     541            162     59         2     1
##  6 FALSE                    663            160     50         2     1
##  7 FALSE                    579            185     60         1     1
##  8 FALSE                    502            158     61         1     1
##  9 TRUE                     562            132     48         3     1
## 10 FALSE                    495            111     44         1     1
## # ... with 590 more rows
```

Binary Logit Model Regression

```r
### Model air conditioning as a binary logit
## Regress air conditioning on cost variables
reg_logit <- data %>%
  glm(formula = air_conditioning ~ cost_system + cost_operating,
      family = 'binomial')
```

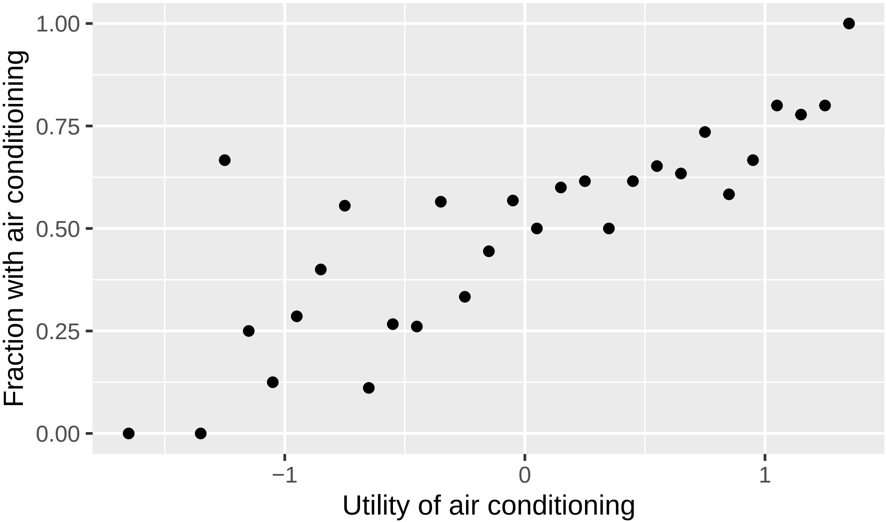
Regression Summary

```r
## Summarize regression results
reg_logit %>%
  summary()
##
## Call:
## glm(formula = air_conditioning ~ cost_system + cost_operating,
##     family = "binomial", data = .)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7294  -1.1778   0.8018   1.0572   1.7400
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     4.991258   0.940686   5.306 1.12e-07 ***
## cost_system    -0.005485   0.001474  -3.721 0.000198 ***
## cost_operating -0.010045   0.002159  -4.653 3.27e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 826.91  on 599  degrees of freedom
## Residual deviance: 787.99  on 597  degrees of freedom
## AIC: 793.99
##
## Number of Fisher Scoring iterations: 4
```

Visualize Predicted Utility

```
### Visualize utility of air conditioning adoption using bins
## Calculate predicted utility of air conditioning
data <- data %>%
  mutate(utility_ac_logit = predict(reg_logit))
## Plot fraction vs. utility of air conditioning using bins
data %>%
  mutate(bin = cut(utility_ac_logit,
                   breaks = seq(-1.7, 1.4, 0.1),
                   labels = 1:31)) %>%
  group_by(bin) %>%
  summarize(fraction_ac = mean(air_conditioning)) %>%
  mutate(bin = as.numeric(bin),
         bin_mid = 0.1 * (bin - 1) - 1.65) %>%
  ggplot(aes(x = bin_mid, y = fraction_ac)) +
  geom_point() +
  xlab('Utility of air conditioning') +
  ylab('Fraction with air conditioining')
```

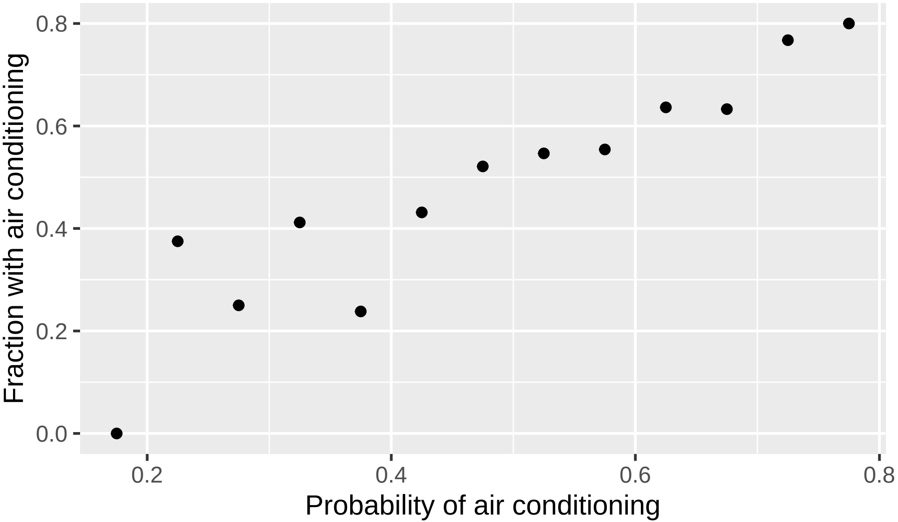
Predicted Utility Plot



Visualize Choice Probability

```r
### Visualize probability of air conditioning using bins
## Calculate predicted probability of air conditioning
data <- data %>%
  mutate(prob_ac_logit = exp(utility_ac_logit) /
           (1 + exp(utility_ac_logit)))
## Plot fraction vs. probability of air conditioning using bins
data %>%
  mutate(bin = cut(prob_ac_logit,
                   breaks = seq(0, 1, 0.05),
                   labels = 1:20)) %>%
  group_by(bin) %>%
  summarize(fraction_ac = mean(air_conditioning)) %>%
  mutate(bin = as.numeric(bin),
         bin_mid = 0.05 * (bin - 1) + 0.025) %>%
  ggplot(aes(x = bin_mid, y = fraction_ac)) +
  geom_point() +
  xlab('Probability of air conditioning') +
  ylab('Fraction with air conditioning')
```

Choice Probability Plot



Marginal Effects and Elasticities

```r
### Calculate marginal effects and elasticities
## Calculate the average marginal effect of each cost variable
coef(reg_logit)[2:3] *
  mean(data$prob_ac_logit * (1 - data$prob_ac_logit))
##     cost_system cost_operating
##    -0.001273955   -0.002333158

## Calculate the elasticity of each cost variable
coef(reg_logit)[2:3] *
  c(mean(data$cost_system * (1 - data$prob_ac_logit)),
    mean(data$cost_operating * (1 - data$prob_ac_logit)))
##     cost_system cost_operating
##      -1.5074074     -0.7448268
```

Cost Tradeoffs

How do consumers trade off the purchase price and the annual operating cost?

- What reduction in purchase price offsets a $1 increase in the annual operating cost?

$$U = \beta_0 + \beta_1 P + \beta_2 C + \varepsilon$$

$$dU = \beta_1 dP + \beta_2 dC$$

$$dU = 0 \Rightarrow \frac{dP}{dC} = -\frac{\beta_2}{\beta_1}$$

```
### Calculate the tradeoff between system cost and operating cost
## Calculate system cost equivalence of an increse in operating cost
-coef(reg_logit)[3] / coef(reg_logit)[2]
## cost_operating
##      -1.831429
```

Implied Discount Factor

What is the implied discount factor of consumers for a two-year time horizon? A three-year time horizon?

$$U = \beta_0 + \beta_1 P + \beta_2 C + \varepsilon$$

Divide by β_1 to express in terms of present day dollars

$$\frac{U}{\beta_1} = \frac{\beta_0}{\beta_1} + P + \frac{\beta_2}{\beta_1} C + \frac{\varepsilon}{\beta_1}$$

Re-express using a sum of discounted operating costs

$$\frac{U}{\beta_1} = \frac{\beta_0}{\beta_1} + P + \sum_{t=1}^{T} \gamma^{t-1} C + \frac{\varepsilon}{\beta_1}$$

The last two equations give that

$$\frac{\beta_2}{\beta_1} = \sum_{t=1}^{T} \gamma^{t-1}$$

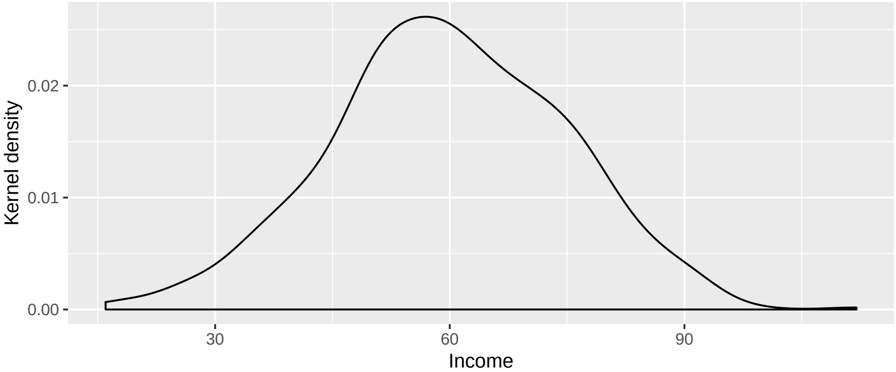
Implied Discount Factor Calculation

```r
### Calculte the implied discount factor of consumers
## Calculate the implied discount factor for two-year horizon
coef(reg_logit)[3] / coef(reg_logit)[2] - 1
## cost_operating
##      0.8314289

## Calculate the implied discount factor for three-year horizon
(-1 + sqrt(1 - 4 * (1 - coef(reg_logit)[3] / coef(reg_logit)[2]))) / 2
## cost_operating
##      0.5399177
```

Visualize Income Data

```
### Visualize income variable
## Plot kernel density of income
data %>%
  ggplot(aes(x = income)) +
  geom_density() +
  xlab('Income') +
  ylab('Kernel density')
```



Binary Logit with Heterogeneous Coefficients

```
### Model air conditioning with heterogeneous cost coefficients
## Regress air conditioning on costs divided by income
reg_logit_income <- data %>%
  glm(formula = air_conditioning ~ I(cost_system / income) +
        I(cost_operating / income),
      family = 'binomial')
```

Heterogeneous Coefficients Regression Summary

```
## Summarize regression results
reg_logit_income %>%
  summary()
##
## Call:
## glm(formula = air_conditioning ~ I(cost_system/income) + I(cost_operating/income),
##     family = "binomial", data = .)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8159  -0.8297   0.3987   0.7277   2.2674
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               6.62710    0.58417  11.345  < 2e-16 ***
## I(cost_system/income)    -0.35861    0.05389  -6.655 2.83e-11 ***
## I(cost_operating/income) -1.02016    0.15801  -6.456 1.07e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 826.91  on 599  degrees of freedom
## Residual deviance: 580.45  on 597  degrees of freedom
## AIC: 586.45
##
## Number of Fisher Scoring iterations: 5
```

Marginal Effects and Elasticities with Heterogeneity

```r
### Calculate marginal effects and elasticities
## Calculate predicted utility and probability
data <- data %>%
  mutate(utility_ac_logit_income = predict(reg_logit_income),
         prob_ac_logit_income = exp(utility_ac_logit_income) /
                                   (1 + exp(utility_ac_logit_income)))
## Calculate the average marginal effect of each cost variable
coef(reg_logit_income)[2:3] *
  mean(data$prob_ac_logit_income * (1 - data$prob_ac_logit_income) /
         data$income)
##    I(cost_system/income) I(cost_operating/income)
##            -0.0009860421            -0.0028050473

## Calculate the average elasticity of each cost variable
coef(reg_logit_income)[2:3] *
  c(mean(data$cost_system * (1 - data$prob_ac_logit_income) /
          data$income),
    mean(data$cost_operating * (1 - data$prob_ac_logit_income) /
          data$income))
##    I(cost_system/income) I(cost_operating/income)
##                -2.066112                -1.556797
```

Gruber and Poterba (1994)

Gruber and Poterba (1994)

Research question

- How does the price of health insurance affect the choice to purchase health insurance?

Empirical methods

- Combine a binary probit model with a difference-in-differences design to estimate the own-price elasticity of health insurance
- Exploit a change in federal tax policy that effectively reduced the price of health insurance for self-employed individuals

Results

- Own-price elasticity of health insurance is -1.8 for self-employed single individuals

Binary Probit Model

Similar to a binary logit model but with a different assumption about the joint density of unobserved utility (or regression error term)

$$\varepsilon_{nj} \sim \text{i.i.d. } N(0,1)$$

This assumption, plus the assumption of linear representative utility, yields

$$\Phi^{-1}(P_n) = \beta' x_n$$

- `glm()` with `family = binomial(link = "probit")` can estimate this model

Probit is more common than logit for binary models, but logit is more common than probit for multinomial models

- Results tend to be very similar
- Probit choice probabilities do not have closed-form expressions

General Comments and Questions

25 years later, we are still researching how federal policies affect health insurance coverage

- Gruber and Sommers (2019)
- Hot topic for research in economics and more broadly

Combination of “structural estimation” (probit) and research design (DiD)

- What are the identifying assumptions?
- Do we believe them?

Why use a probit instead of a linear probability model?

- What do they gain from having structural parameters?

Announcements

Reading for next time

- Adamowicz et al. (1994)

Upcoming

- Problem Set 1 is posted, due September 24