Nonlinear Regression and Qualitative Response Regression Models

Lecture 11
What was a linear regression model?

- Models that are linear in the parameters or can be transformed to be linear in the parameters.

- Distinguish between linear in the parameters and linear in the variables. A linear regression model does not have to be linear in the variables.
What is a nonlinear regression model?

- A model which is not (and cannot be transformed to be) linear in the parameters
Examples

\[ Y_i = \beta_1 + \beta_2 \left( \frac{1}{X_i} \right) + u_i \]

\[ Y_i = \beta_1 + \beta_2 \ln X_i + u_i \]

\[ Y_i = e^{(\beta_1 + \beta_2 X_i + u_i)} \]
\[ \ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \]

\[ \ln Y_i = \beta_1 - \beta_2 \left( \frac{1}{X_i} \right) + u_i \]

\[ Y_i = \beta_1 + (0.75 - \beta_1) e^{-\beta_2 (X_i^2)} + u_i \]
Examples

\[ Y_i = \frac{1}{1 + e^{\beta_1 \beta_2 X_i + u_i}} \]

\[ \ln Y_i = \beta_1 + \beta_2 \left( \frac{1}{X_i} \right) + u_i \]

\[ Y_i = \beta_1 + \beta_3 X_i + u_i \]
Estimation of nonlinear models

Compare

\[ Y_i = \beta_1 + \beta_2 X_i + u_i \quad (A) \]

to

\[ Y_i = \beta_1 e^{\beta_2 X_i} + u_i \quad (B) \]
Estimation of nonlinear models

For model (A), OLS: minimize the residual sum of squares (RSS)

\[ \sum u_i^2 = \sum (Y_i - \beta_1 - \beta_2 X_i)^2 \]

\[ \begin{align*}
\frac{\partial (\sum u_i^2)}{\partial \beta_1} &= 0 \\
\frac{\partial (\sum u_i^2)}{\partial \beta_2} &= 0
\end{align*} \quad \Rightarrow \hat{\beta}_{1,OLS} \quad \hat{\beta}_{2,OLS} \]

explicit solutions
Estimation of nonlinear models

What about model (B)? OLS?

$$\sum u_i^2 = \sum \left( Y_i - \beta_1 e^{\beta_2 X_i} \right)^2$$

$$\frac{\partial (\sum u_i^2)}{\partial \beta_1} = 0$$

$$\frac{\partial (\sum u_i^2)}{\partial \beta_2} = 0$$

$$\Rightarrow$$

$$\sum Y_i e^{\beta_2 X_i} = \beta_1 e^{2\beta_2 X_i}$$

$$\sum Y_i X_i e^{\beta_2 X_i} = \beta_1 \sum X_i e^{2\beta_2 X_i}$$

no explicit solutions
Trial and error

- becomes tiresome when more than one parameter
- no guarantee to find the minimum
Search for the minimum in a more systematic way instead, using some kind of algorithm, e.g. the Newton-Raphson method (a method to find the root to an equation using taylor series expansion, starting with initial guess values successive approximations are updated until "convergence" is achieved, that is when the updated values does not differ much from the previous values)
Remember the dependent variable $Y$ can also be called *response* variable

- categorical response
  - binary/dichotomous
  - polytomous
- count data
Binary response

Examples

- Democrat/Republican
- Own a house/not own a house
- Pass/fail
- Survival/Death
- Cured/not cured
- Have insurance/not have insurance
- Repay a loan/not repay
- Employed/unemployed
- Explode/not explode
Suppose $Y = 1$ for one of the outcomes, which can be called success, and $Y = 0$ for the other outcome. What if we could find a model to describe the probability of success as a function of some explanatory variables?

Denote the probability of success given the explanatory variables $X_1, \ldots, X_k$

$$P \left( Y_i = 1 \mid X_{1i}, \ldots, X_{ki} \right) = P_i$$
Now, when $Y$ is *Bernoulli* distributed with success probability $P_i$

$$E(Y_i) = 0 \cdot (1 - P_i) + 1 \cdot P_i = P_i$$

One suggestion is to assume the "Linear Probability Model" (LPM)

$$E(Y_i) = P_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

and fit

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

with OLS as usual
Binary response

Not a good idea because

- First of all the restriction

\[ 0 \leq E(Y_i) = P_i \leq 1 \]

is not guaranteed by LPM

- Often not realistic to assume a linear relationship between the probability of success and the explanatory variable(s)
Non-normality of $u_i$

$u_i = Y_i - (\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})$ will also be Bernoulli distributed because $Y_i$ is either 1 or 0 with probabilities $P_i$ and $(1 - P_i)$, respectively.

However, OLS estimators still unbiased.
Heteroscedastic variance of $u_i$

The variance of the disturbance term is

$$\text{var} (u_i) = P_i (1 - P_i)$$

and since $P_i$ is a function of the explanatory variables by the model the variance is different for different levels

Transformation? $P_i$ unknown and needs to be estimated in such case
Why not use a model were

1. the probability of success $P_i$ never ends up outside the interval $(0, 1)$

2. the relationship between $P_i$ and $X_i$ is nonlinear

directly
The cumulative distribution function (CDF) of a random variable
Binary response

Which CDF?

- Logistic CDF ⇒ Logit
- Normal CDF ⇒ Probit
The Logit model

Use the logistic distribution function

\[ P_i = E(Y_i) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \]

and let \( z = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} \)

For example with one explanatory variable

\[ P_i = \frac{e^{\beta_1 + \beta_2 X_i}}{1 + e^{\beta_1 + \beta_2 X_i}} \]
The Logit model

can also be expressed as

$$\log \left( \frac{P_i}{1 - P_i} \right) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

where

$$\frac{P_i}{1 - P_i}$$

is the odds of success
The Logit model

For example

\[
\log \left( \frac{P_i}{1 - P_i} \right) = \beta_1 + \beta_2 X_i
\]

\[
\log \left( \frac{P_i}{1 - P_i} \right) = \beta_1 + \beta_2 X_i + \beta_3 X_i^2
\]
Estimation for the Logit model

The method of Maximum Likelihood (ML)

Maximize the likelihood

\[
L = \prod_{i=1}^{n} P_i^{Y_i} (1 - P_i)^{1-Y_i}
\]

or equivalently

\[
\ln L = \sum_{i=1}^{n} \left[ Y_i \ln \frac{P_i}{1 - P_i} + \ln (1 - P_i) \right]
\]
Estimation for the Logit model

Set

\[ \frac{\partial \ln L}{\partial \beta} = 0 \]

for each parameter, no explicit solutions \( \Rightarrow \) use some nonlinear estimation technique to obtain the numerical solutions for \( \hat{\beta}_1, \hat{\beta}_2, \ldots \)

The ML method is a large-sample method, i.e. the ML estimators are asymptotically normal distributed
Consider

\[ P_i = \frac{e^{\beta_1 + \beta_2 X_i}}{1 + e^{\beta_1 + \beta_2 X_i}} \]

\( \beta_1 \) is the log-odds of success at \( X = 0 \)
\( \beta_2 \) is the change in the log-odds for a one-unit increase in \( X \), perhaps more meaningful to interpret \( e^{\beta_2} \) in terms of the odds, \( 100 \cdot (e^{\beta_2} - 1) \) then gives the percent change in the odds for a one-unit increase in \( X \)

When there is more than one explanatory variable the interpretation of a regression coefficient \( \beta \) is *holding the other variables constant* (as usual for a partial regression coefficients)
Conventional goodness of fit measures like $R^2$ are not appropriate, instead there are several kinds of pseudo $R^2$, e.g.

$$\text{Count } R^2 = \frac{\text{number of correct predictions}}{\text{total number of predictions}}$$

Observations for which the predicted probability of success $\hat{P}_i$ is greater than (less than) 0.5 are classified as 1 (0)
The equivalence to the overall F test for linear regression is the **Likelihood ratio (LR) test**

The LR test statistic involves the ratio of the likelihood function for the model of interest and the minimal model (the model with only an intercept term). Under the null hypothesis of the minimal model the LR statistic follows the $\chi^2$ distribution with df equal to the number of explanatory variables, i.e. a large observed value of the LR statistic leads to the rejection of the minimal model.
If the normal CDF $\Phi$ is used instead

$$P_i = \Phi (\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})$$

or alternatively

$$\Phi^{-1} (P_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

The shapes of the response functions for logit and probit models are quite similar for many applications, the differences are largest in the tails.
Count Data

The number of events in a time interval for example:
- Visits to a physician per year
- Number of phone calls received in 5 minutes
- Number of patents received by a firm in a year

Can be modeled with the **Poisson regression model**
Poisson regression

The response variable $Y_i$ follows the poisson distribution with mean $E(Y_i) = \mu_i$, the poisson regression model can be written as

\[
E(Y_i) = \mu_i = e^{\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}}
\]

\[
\ln(\mu_i) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}
\]