Lecture 9: Nonlinear Least Squares I

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Logit Estimation Using MLE

Today

- Nonlinear Least Squares
- Properties of the Nonlinear Least Squares Estimator
- Computing the Nonlinear Least Squares Estimator

Upcoming

- Reading for next time
  - Optional: Schaefer (1998)
- Problem sets
  - Problem Set 2 is posted, due October 17

Maximum Likelihood Recap

If we are willing to make a strong assumption about the density of our data

$$f(y \mid \beta)$$

then the MLE gives us the "best" consistent estimator of β

- "Best" means the estimator with the lowest variance

But what if we are not willing to make such a strong assumption about our data?

- Note: For the logit model, we are already making some strong assumptions, so the MLE assumption is not unreasonable. But in other settings, this assumption may be overly restrictive.

Nonlinear Least Squares

Nonlinear Regression Models

The general formula for a nonlinear regression model

$$y_i = h(x_i, \beta) + \varepsilon_i$$

- OLS (linear model) is a special case
- Some models that appear to be nonlinear can be simplified to a linear model
  - See the Cobb-Douglas production function example from Lecture 1

Examples of Nonlinear Regression Models

Binary discrete choice

$$u_i = \beta' x_i + \varepsilon_i$$

$$y_i = \begin{cases} 1 & \text{if } u_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

CES production function

$$\ln y_i = \ln \gamma - \frac{\upsilon}{\rho} \ln \left[\delta K_i^{-\rho} + (1 - \delta) L_i^{-\rho} \right] + \varepsilon_i$$

Exponential regression

$$y_i = \beta_0 + \beta_1 e^{\beta_2 x_{i1} + \beta_3 x_{i2}} + \varepsilon_i$$

Nonlinear Least Squares Assumption

NLS Assumption

For the nonlinear regression model $y_i = h(x_i, \beta) + \varepsilon_i$, we assume that

1. $\frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0}\right)\left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0'}\right) \xrightarrow{p} Q^0$
2. $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial h(x_i,\beta_0)}{\partial \beta_0}\varepsilon_i \xrightarrow{p} 0$
3. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial h(x_i,\beta_0)}{\partial \beta_0}\varepsilon_i \xrightarrow{d} N(0, \sigma^2 Q^0)$

where Q^0 is a positive definite matrix and σ^2 is a finite constant

What is a further assumption that almost guarantees this to be true?

- Strict independence of ε_i and x_i: $f(\varepsilon_i) = f(\varepsilon_i \mid x_i)$
- ε_i is normally distributed: $\varepsilon_i \sim N(0, \sigma^2)$

Nonlinear Least Squares Estimator

Just as with ordinary least squares (OLS), we seek to minimize the sum of squared errors

$$S(\beta) = \sum_{i=1}^{n} [y_i - h(x_i, \beta)]^2$$

The nonlinear least squares estimator is defined as

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, S(\beta)$$

The first-order condition for this minimization is

$$\sum_{i=1}^{n} [y_i - h(x_i, \beta)] \frac{\partial h(x_i, \beta)}{\partial \beta} = 0$$

- Unlike OLS, this estimator may not have a closed-form expression
- This is a moment condition that we could use for GMM
  - More on that next week...

Nonlinear Least Squares Example

We want to estimate the parameters of

$$y_i = \beta_1 + \beta_2 e^{\beta_3 x_i} + \varepsilon_i$$

Nonlinear least squares will find the set of parameters that minimizes

$$S(\beta) = \sum_{i=1}^{n} [y_i - (\beta_1 + \beta_2 e^{\beta_3 x_i})]^2$$

The first-order conditions are

$$\frac{\partial S(\beta)}{\partial \beta_1} = -\sum_{i=1}^{n} [y_i - (\beta_1 + \beta_2 e^{\beta_3 x_i})] = 0$$

$$\frac{\partial S(\beta)}{\partial \beta_2} = -\sum_{i=1}^{n} [y_i - (\beta_1 + \beta_2 e^{\beta_3 x_i})] e^{\beta_3 x_i} = 0$$

$$\frac{\partial S(\beta)}{\partial \beta_3} = -\sum_{i=1}^{n} [y_i - (\beta_1 + \beta_2 e^{\beta_3 x_i})] \beta_2 x_i e^{\beta_3 x_i} = 0$$

Nonlinear Least Squares for the Logit Model

We can think of regressing a choice on logit choice probabilities

$$y_{ni} = P_{ni} + \omega_{ni}$$

If we substitute in P_{ni} and assume linear representative utility, we get

$$y_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} + \omega_{ni}$$

We can use nonlinear least squares to find the set of parameters, $\hat{\beta}$, that minimizes

$$S(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{J} \left[y_{ni} - \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} \right]^2$$

How is this different from our MLE logit estimation?

Properties of the Nonlinear Least Squares Estimator

Consistency of the NLS Estimator

If the following assumptions hold

1. The parameter space containing β_0 is compact (no gaps or non-concave regions)
2. For any vector β in that parameter space $(1/n)S(\beta) \xrightarrow{p} q(\beta)$, a continuous and differentiable function
3. $q(\beta)$ has a unique minimum at the true parameter vector, β_0

then the nonlinear least squares estimator is consistent

$$\hat{\beta} \xrightarrow{p} \beta_0$$

where $\hat{\beta}$ is the NLS estimator and β_0 is the true parameter value(s)

Asymptotic Normality of the NLS Estimator

The asymptotic distribution of the NLS estimator, $\hat{\beta}$, is normal with mean at the true parameter value(s), β_0, and known variance

$$\hat{\beta} \overset{a}{\sim} N\left[\beta_0, \frac{\sigma^2}{n}(Q^0)^{-1}\right]$$

where

$$Q^0 = \text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0}\right)\left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0'}\right)$$

NLS Variance Estimator

From the asymptotic normality of the NLS estimator, we have

$$Var(\hat{\beta}) = \frac{\sigma^2}{n}\left[\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0}\right)\left(\frac{\partial h(x_i,\beta_0)}{\partial \beta_0'}\right)\right]^{-1}$$

- Variance of NLS estimator depends on β_0 and σ^2, which we do not know, so we need an estimator for the variance of the NLS estimator

We can estimate this variance using estimators, $\hat{\beta}$ and $\hat{\sigma}^2$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2\left[\sum_{i=1}^{n}\left(\frac{\partial h(x_i,\beta)}{\partial \beta}\right)_{\hat{\beta}}\left(\frac{\partial h(x_i,\beta)}{\partial \beta'}\right)_{\hat{\beta}}\right]^{-1}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[y_i - h(x_i,\hat{\beta})]^2$$

- More robust estimators exist, but we will stick with this most basic variance estimator for now

Estimating the NLS Variance-Covariance Matrix

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2 \left[\sum_{i=1}^{n} \left(\frac{\partial h(x_i, \beta)}{\partial \beta} \right)_{\hat{\beta}} \left(\frac{\partial h(x_i, \beta)}{\partial \beta'} \right)_{\hat{\beta}} \right]^{-1}$$

Steps to estimate this variance-covariance matrix

1. Write down the derivative of the nonlinear regression model with respect to each of the K parameters.
2. Calculate this $K \times 1$ vector of derivatives, at the estimated parameters, for each decision maker.
3. Calculate the $K \times K$ matrix that is the product of the above vector and its transpose for each decision maker.
4. Sum these matrices for all decision makers.
5. Estimate the variance of the econometric error as the mean sum of squares at the estimated parameters.
6. Calculate the variance-covariance matrix, which is a function of the above $K \times K$ matrix and the estimated error variance.

NLS Goodness of Fit

There is no universally agreed upon measure to summarize the goodness of fit of a nonlinear least squares model

Some people calculate the traditional R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y_i - h(x_i, \hat{\beta})]^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- But R^2 for a nonlinear model may not have a range of $[0, 1]$

Many alternatives exist, and some may be useful depending on what you want your R^2 to summarize

Wald Test

A Wald test is used to test hypotheses about parameters fit by NLS

$$H_0\colon\ r(\beta_0) = q$$

where r is a vector of J functions of parameters

The Wald test statistic is

$$W = [r(\hat{\beta}) - q]'[R(\hat{\beta})\widehat{Var}(\hat{\beta})R'(\hat{\beta})]^{-1}[r(\hat{\beta}) - q]$$

where R is the $J \times K$ Jacobian

$$R(\hat{\beta}) = \left(\frac{\partial r(\beta)}{\partial \beta'}\right)_{\hat{\beta}}$$

This Wald test statistic converges in distribution to a chi-squared

$$W \xrightarrow{d} \chi^2(J)$$

Conducting a NLS Wald Test

Hypothesis: H_0: $r(\beta_0) = q$

Test statistic: $W = [r(\hat{\beta}) - q]'[R(\hat{\beta})\widehat{Var}(\hat{\beta})R'(\hat{\beta})]^{-1}[r(\hat{\beta}) - q]$

Jacobian matrix: $R(\hat{\beta}) = \left(\frac{\partial r(\beta)}{\partial \beta'}\right)_{\hat{\beta}}$

Steps to conduct a Wald Test using NLS estimators

1. Create a vector of J parameter restrictions, $r(\beta_0) = q$.
2. Calculate the $J \times K$ Jacobian matrix by differentiating each restriction with respect to each of the K parameters.
3. Calculate the Wald test statistic, which is a function of the vector of restrictions, the Jacobian matrix, and the variance-covariance matrix.
4. Conduct the Wald test using this test statistic, which is distributed χ^2.

Computing the Nonlinear Least Squares Estimator

Computing the Nonlinear Least Squares Estimator

One option for minimizing the sum of squared errors in a nonlinear regression is to use the numerical optimization techniques we talked about previously

An alternative option is to take advantage of the fact that we already know how to minimize sum of squared errors for a linear model using OLS, so if we can sufficiently approximate our nonlinear regression as a linear regression, we can use OLS to help us find the NLS estimator

1. Begin with some initial parameter values, β_1
2. Approximate the regression model as a linear model given a set of parameters, β_t
3. Use OLS to find the next iteration of parameters, β_{t+1}
4. Repeat (2) and (3) until the parameters converge

Linear Approximation of a Nonlinear Regression Model

A common method for minimizing the sum of squares in a NLS regression is the Gauss-Newton method, which approximates the nonlinear regression model, $h(x_i, \beta_{t+1})$, as a linear Taylor approximation given a vector of parameters, β_t

$$h(x_i, \beta_{t+1}) \approx h(x_i, \beta_t) + (\beta_{t+1} - \beta_t)' \left(\frac{\partial h(x_i, \beta_t)}{\partial \beta_t} \right)$$

which implies that

$$y_i - h(x_i, \beta_t) + \beta_t' \left(\frac{\partial h(x_i, \beta_t)}{\partial \beta_t} \right) = \beta_{t+1}' \left(\frac{\partial h(x_i, \beta_t)}{\partial \beta_t} \right) + \varepsilon_{it}$$

where ε_{it} is an error term that includes the original econometric error, ε_i, and error due to the linear Taylor approximation

Iterate Parameters Using OLS

In that final linear equation

$$y_i - h(x_i, \beta_t) + \beta_t' \left(\frac{\partial h(x_i, \beta_t)}{\partial \beta_t} \right) = \beta_{t+1}' \left(\frac{\partial h(x_i, \beta_t)}{\partial \beta_t} \right) + \varepsilon_{it}$$

the only unknowns are the next vector of parameters, β_{t+1}, and the composite error term, ε_{it}

- We start from a given vector of parameters, β_t, so $h(x_i, \beta_t)$ and $\partial h(x_i, \beta_t)/\partial \beta_t$ can be calculated or numerically approximated
- We can think of this as a linear regression equation

To get the next set of parameters, β_{t+1}

1. Calculate the left-hand side of this regression equation
2. Regress those values on $\partial h(x_i, \beta_t)/\partial \beta_t$

β_{t+1} is the result of this OLS regression

Iteration and Convergence

We have a closed-form solution for the parameters that result from an OLS regression, so we can more succinctly write

$$\beta_{t+1} = \beta_t + \left\{ \sum_{i=1}^{n} \left(\frac{\partial h(x_i, \beta)}{\partial \beta} \right)_{\beta_t} \left(\frac{\partial h(x_i, \beta)}{\partial \beta'} \right)_{\beta_t} \right\}^{-1} \times \left\{ \sum_{i=1}^{n} \left(\frac{\partial h(x_i, \beta)}{\partial \beta} \right)_{\beta_t} [y_i - h(x_i, \beta_t)] \right\}$$

We use this formula to iterate through parameters until they converge

- See notes on numerical optimization for convergence criteria

Alternatively, we can use the other numerical optimization methods we discussed previously

Announcements

Reading for next time

- Optional: Schaefer (1998)

Upcoming

- Problem Set 2 is posted, due October 17