Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Examples of nonlinear least squares

Koen Jochmans François Poinas

2024 — 2025

Overview

This deck of slides goes through examples of nonlinear least squares (NLLS).

Hansen treats NLLS in Chapter 23.

For the current set of slides the relevant sections in Hansen are

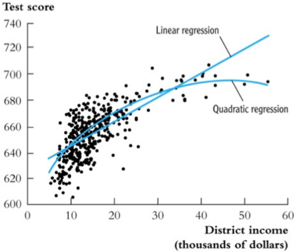
H 23.1–H23.3

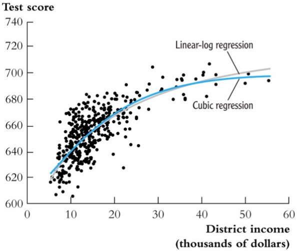
H25.1–H25.4 and H25.9

H26.11

Asymptotics for NLLS will be done later on.

Example: Explaining test scores as a function of income





The negative exponential-growth model is

$$Y = \alpha(1 - \exp(-\beta X)) + e, \qquad \mathbb{E}(e|X) = 0.$$

Parsimonious specification that yields a CEF that is

monotone increasing and

bounded.

However, this specification is nonlinear in parameters $\theta = (\alpha, \beta)'$.

Example: CES production

A CES production function for output Y with two (equally-weighted) inputs X_1, X_2 is

$$Y = A\left(X_1^\gamma + X_2^\gamma\right)^{1/\gamma}$$

with

A random factor productivity, and

γ the substitution parameter (substitution elasticity is $1/(1-\rho)$).

Let $\alpha = \mathbb{E}(A|X_1, X_2)$ so that $A = \alpha u$ for $\mathbb{E}(u|X_1, X_2) = 1$. Then

$$Y = \alpha\left(X_1^\gamma + X_2^\gamma\right)^{1/\gamma} u$$

or

$$\mathbb{E}(Y|X_1, X_2) = \alpha\left(X_1^\gamma + X_2^\gamma\right)^{1/\gamma}.$$

Here, $\theta = (\alpha, \gamma)'$.

Note that we can take logs on the left and right of the above equation to arrive at

$$\log(Y) = \log(\alpha) + \frac{1}{\gamma}\log(X_1^\gamma + X_2^\gamma) + \log(u).$$

However, $\mathbb{E}(u|X_1, X_2) = 1$ does not imply that

$$\mathbb{E}(\log(u)|X_1, X_2) = 0$$

unless u is independent of X_1, X_2.

Hence,

$$\mathbb{E}(\log(Y)|X_1, X_2) \neq \log(\alpha) + \frac{1}{\gamma}\log(X_1^\gamma + X_2^\gamma)$$

in general.

In either case, the specification is nonlinear in parameters.

Example: Exponential regression

For non-negative outcomes $Y \geq 0$ a popular specification has

$$Y = \exp(X'\theta)\, e, \qquad \mathbb{E}(e|X) = 1.$$

As before, one often sees log-linearization

$$\log(Y) = X'\theta + \log(e)$$

but the above does not imply that $\mathbb{E}(\log(Y)|X) = X'\theta$ unless u and X are independent. (The intercept in θ must in any event be redefined to include $\mathbb{E}(\log e)$ in this case.)

Note that the log operation here is not well-defined when $\mathbb{P}(Y = 0) > 0$.

Another common specification for binary outcomes $Y \in \{0, 1\}$ has

$$Y = \begin{cases} 1 & \text{if } X'\theta \geq e \\ 0 & \text{if } X'\theta < e \end{cases}$$

where e is independent of X and has CDF F.

Then

$$\mathbb{E}(Y|X) = \mathbb{P}(Y = 1|X) = \mathbb{P}(e \leq X'\theta|X) = F(X'\theta).$$

Nonlinear least squares

Let

$$\mathbb{E}(Y|X) = m(X, \theta)$$

be known up to vector θ.

Then

$$Y = m(X, \theta) + e, \qquad \mathbb{E}(e|X) = 0.$$

Therefore,

$$\mathbb{E}\left((Y - m(X, \theta))^2\right) \leq \mathbb{E}\left((Y - m(X, \tilde{\theta}))^2\right)$$

for any vector $\tilde{\theta}$.

This suggests estimating θ by minimizing

$$\sum_{i=1}^{n}(Y_i - m(X_i, \theta))^2$$

with respect to θ.

This is the nonlinear least-squares (NLLS) estimator.

Usually, a closed-form expression for it is not available.

It is defined as a root to the equation

$$\sum_{i=1}^{n} \frac{\partial m(X_i, \theta)}{\partial \theta} (Y_i - m(X_i, \theta)) = 0,$$

which can be found by numerical methods.

Newton's algorithm

Newton-Raphson is a popular root-finding algorithm.

Want to solve $\varphi(x) = 0$. Let x_0 be an initial guess. For a new guess x_1 we have

$$\frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \approx \left.\frac{\partial \varphi(x)}{\partial x}\right|_{x=x_0} = \varphi'(x_0).$$

So,

$$\varphi(x_0) + (x_1 - x_0)\,\varphi'(x_0) \approx \varphi(x_1).$$

We want that $\varphi(x_1) = 0$. Solving for x_1 yields

$$x_1 = x_0 - \varphi(x_0)/\varphi'(x_0)$$

as our new guess.

In practice, if the function does not improve at x_1 we re-evaluate in $x_1' = x_0 - h(x_0 - x_1)$ for $h \in (0, 1)$ a step size and re-evaluate. We then iterate this procedure until no further improvement (up to some specified tolerance level) is found.

Example: Linear model

In the linear CEF specification

$$m(X, \theta) = X'\theta, \qquad \frac{\partial m(X_i, \theta)}{\partial \theta} = X$$

and so we solve

$$\sum_{i=1}^{n} X_i(Y_i - X_i'\theta) = 0,$$

which are the normal equations for OLS.

Example: Exponential regression model

Here,

$$m(X,\theta) = \exp(X'\theta), \qquad \frac{\partial m(X,\theta)}{\partial \theta} = X \exp(X'\theta)$$

and so we solve

$$\sum_{i=1}^{n} \exp(X_i'\theta) X_i (Y_i - \exp(X_i'\theta)) = 0.$$

Here,

$$m(X,\theta) = F(X'\theta), \qquad \frac{\partial m(X,\theta)}{\partial \theta} = Xf(X'\theta)$$

for $f(a) = \partial F(a)/\partial a$ and so we solve

$$\sum_{i=1}^{n} f(X_i'\theta)X_i(Y_i - F(X_i'\theta)) = 0.$$

For example for probit, $F(a) = \Phi(a)$ and we solve

$$\sum_{i=1}^{n} \phi(X_i'\theta)X_i(Y_i - \Phi(X_i'\theta)) = 0.$$

We will do asymptotics for NLLS later.

Typically, NLLS is not asymptotically efficient.

An example is OLS, which is efficient only under homoskedasticity.

A GLS argument suggests that the efficient estimating equation is

$$\sum_{i=1}^{n} \frac{\partial m(X_i, \theta)}{\partial \theta} \frac{1}{\mathrm{var}(Y_i | X_i)} (Y_i - m(X_i, \theta)) = 0.$$

Example: Exponential regression

If

$$Y = \exp(X'\theta)\, e, \qquad \mathbb{E}(e|X) = 1$$

then

$$\operatorname{var}(Y|X) = \operatorname{var}(e|X)\ \exp(X'\theta)^2$$

so that, if $\operatorname{var}(Y|X) = \sigma^2$ we have that NLLS is optimal; in this case

$$\operatorname{var}(e|X) = \frac{\sigma^2}{\exp(X'\theta)^2}$$

so that e is heteroskedastic. If, on the other hand, $\operatorname{var}(e|X) = \sigma^2$ then the solution to

$$\sum_{i=1}^{n} \frac{X_i}{\exp(X_i'\theta)}(Y_i - \exp(X_i'\theta)) = 0$$

is asymptotically efficient.

If $Y|X$ is Poisson distributed then

$$\mathbb{E}(Y|X) = \text{var}(Y|X)$$

so that

$$\sum_{i=1}^{n} X_i(Y_i - \exp(X_i'\theta)) = 0$$

is optimal.

We will derive this estimator as the MLE for this model.

Example: Probit

If $Y \in \{0,1\}$ with

$$\mathbb{E}(Y|X) = \Phi(X'\theta)$$

we have

$$\text{var}(Y|X) = \Phi(X'\theta)\,(1 - \Phi(X'\theta))$$

and so the optimal estimator for θ here solves

$$\sum_{i=1}^{n} \frac{\phi(X_i'\theta)}{\Phi(X_i'\theta)\,(1 - \Phi(X_i'\theta))} X_i\,(Y_i - \Phi(X_i'\theta)) = 0.$$

This is the probit estimator.

We will derive this estimator as the MLE for this model.

Nevertheless,

$$Y = m(X, \theta) + e, \qquad \mathbb{E}(e|X) = 0$$

implies that

$$\mathbb{E}(g(X)\, e) = 0$$

and so

$$\sum_{i=1}^{n} g(X_i)\,(Y_i - m(X_i, \theta)) = 0$$

is a valid estimating equation for any choice of function g.