Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Examples of nonlinear models

Koen Jochmans François Poinas

2025 — 2026

Test score data

We had earlier observed a nonlinear relationship between test scores and income in the California test score data.

At that point we had dealt with this by using either a simple polynomial regression or a linear-log transformation.

An issue embedded in the test score problem is that test scores are restricted to a bounded interval.

Outcome variables that exhibit restricted support naturally lead to nonlinearity.

The negative exponential-growth model with parameters (α, β) is

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

Predicted test score for income level x is

$$\hat{y}(x) = \alpha(1 - e^{-x\beta}).$$

Here,

$$\lim_{x \downarrow 0} \hat{y}(x) = 0 \qquad \lim_{x \uparrow +\infty} \hat{y}(x) = \alpha,$$

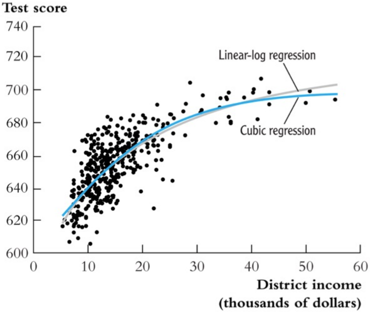
Marginal increase in income changes the predicted test score by

$$\frac{\partial \hat{y}(x)}{\partial x} = \alpha\beta e^{-x\beta}.$$

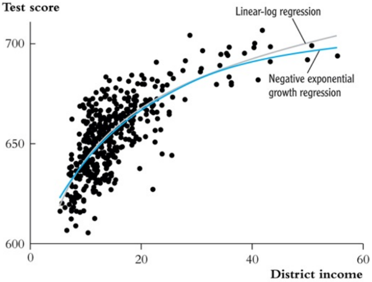
This reflects decreasing returns;

$$\frac{\partial \hat{y}(x)/\partial x}{\hat{y}(x)} = \beta \, \frac{e^{-x\beta}}{1 - e^{-x\beta}}$$

is monotonically decreasing in x.



Test score
740
720
700
680
660
640
620
600
0
10
20
30
40
50
60
Linear-log regression
Cubic regression
District income
(thousands of dollars)



Test score
700
650
600
0
20
40
60
Linear-log regression
Negative exponential
growth regression
District income

Production functions

A two factor CES production function is

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

where A is (random) factor productivity, and γ is the substitution parameter.

Can normalize latent A with $\mathbb{E}(A|X_1, X_2) = \alpha$ to get

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

This is a nonlinear function.

Can linearise and write

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

but

$$\mathbb{E}(e|X_1, X_2) = 1 \not\Leftrightarrow \mathbb{E}(\log(e)|X_1, X_2) = 0,$$

so the nonlinear conditional-mean model and log-log conditional-mean model are not compatible.

Hence,

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

A related issue is what to do with zero output in the log-linearized specification.

Censoring

Data on wages, taxes, hours is often top or bottom coded.

An example of left-censoring is when we observe

$$Y = \max(Y^*, 0),$$

but not Y^* itself.

Censoring introduces nonlinearity.

To see this consider a classical linear model for $Y^*|X$,

$$Y^* = X'\beta + e, \qquad e|X \sim N(0, \sigma^2).$$

Left-censoring introduces a mass point at zero.

We have

$$\mathbb{E}(Y^*|X, Y^* > 0) = X'\beta + \mathbb{E}(e|X, e > -X'\beta).$$

and this last term is not zero, in general.

Under normality can calculate

$$\mathbb{E}(e|X, e > -X'\beta) = -\sigma\frac{\phi(X'\beta/\sigma)}{\Phi(X'\beta/\sigma)},$$

so conditional-mean in non-censored population is a nonlinear function of X.

Parameters of interest

With censoring we may also be interested in different parameters.

We still have that

$$\frac{\partial \mathbb{E}(Y^*|X)}{\partial X} = \beta,$$

but this may not be the ultimate parameter of interest here.

The marginal effect of X on Y is nonlinear. A calculation reveals that

$$\frac{\partial \mathbb{E}(Y|X)}{\partial X} = \frac{\partial \mathbb{E}(Y|X, Y > 0)\, \mathbb{P}(Y > 0|X)}{\partial X} = \beta\, \Phi\left(\frac{X'\beta}{\sigma}\right).$$

This takes into account both the intensive and the extensive margin.

Binary outcomes

A severe form of coding would be

$$Y = \operatorname{sign} Y^* = \begin{cases} 1 & \text{if } Y^* > 0 \\ 0 & \text{if } Y^* \leq 0 \end{cases},$$

along with

$$Y^* = X'\beta + e, \qquad e|X \sim N(0, \sigma^2).$$

Then

$$\mathbb{E}(Y|X) = \mathbb{P}(Y = 1|X) = \mathbb{P}(e > -X'\beta) = \Phi(X'\beta/\sigma)$$

This is the probit model.

Here, again,

$$\frac{\partial \mathbb{E}(Y|X)}{\partial X} = \frac{\beta}{\sigma}\phi(X'\beta/\sigma).$$

Need to impose a normalization; e.g., $\sigma = 1$.

Consumer chooses among goods $\{1, 2, \ldots, m\}$.

Can interpret one choice as the outside option (do not buy anything).

Utility of choosing option c is

$$u(X_c, e_c).$$

Then choice c is made when

$$u(X_c, e_c) \geq u(X_{c'}, e_{c'}), \qquad \text{for all } c' \neq c.$$

The observed choice probabilities are

$$\mathbb{P}(Y = c | X_1, \ldots, X_m)$$

which according to revealed preference are

$$\mathbb{P}(u(X_c, e_c) \geq u(X_1, e_1), \ldots, u(X_c, e_c) \geq u(X_m, e_m) | X_1, \ldots, X_m).$$

These probabilities are generally difficult to compute in closed form.

A popular specification has

$$u(X_c, e_c) = X_c' \theta_c + e_c$$

with $(e_1, \ldots, e_m)$ following a generalized extreme value distribution.

Then

$$\mathbb{P}(Y = c | X_1, \ldots, X_m) = \frac{e^{X_c' \theta_c}}{\sum_{c'=1}^{m} e^{X_{c'}' \theta_{c'}}}.$$

Here, everything can only be interpreted as being relative to the outside option.