Functional Form

Econometrics. ADEi.

1. Introduction

We have employed the linear function in our model specification. Why?

- It is simple and has good mathematical properties.
- It could be reasonable approximation, although is incorrect.
- Most of the times, the functional form is not explicit in the theoretical model, but rather we know that:

$$y_t = f\left(x_t^{'}; \beta\right) + u_t$$

1. Introduction

In practice, most of the economic relationships are non linear.

Examples:

– Entry of a country into a supra-economy (EU)

– Asymmetrical behavior of the agents (recession vs. expansive cycle)

– The DGP are truly more complicated than our simple theoretical models.

1. Introduction

What can we do if we have a non-linear specification?

– If we can linearize it, we can do it

– Otherwise, we should employ non-linear estimators

– Let us see some examples

2. Linearizable models

The most used example is that of a Cobb-Douglas function. This is a non-linear model

$$Y = AK^{\alpha}L^{\beta}, \quad \alpha > 0, \beta > 0$$

Where α and β are the input elasticities.

2. Linearizable models

But, we can linearize it by simply taking logs:

$$\ln(Y) = \beta_1 + \beta_2 \ln(K) + \beta_3 \ln(L)$$

Where β_2 and β_3 the corresponding input elasticities and $\beta_1 = \ln(A)$.

Remark: The coefficients of the explanatory variables of a double-log model are ELASTICITIES

2. Linearizable models

Sometimes the response of the dependent variable has bounds.

Examples:

- Learning curve
- Wage curve
- Effect of advertising on sales

How can we specify this relationship?

2. Linearizable models

LEARNING CURVE

Let us consider that a firm wishes to know how much time it takes to produce a good (cars, for instance). Is this a linear relationship?

Yt = hours per produced car

Xt = number of produced cars

$$y_t = \alpha + \beta x_t$$

2. Linearizable models

INVERSE MODEL

If we admit that the larger the number of produced cars, the shorter the period of time to produce them, then we can use the following relationship:

$$y_t = \alpha + \beta \frac{1}{x_t}$$

if $\beta > 0$, then α is a lower asymptote that corresponds with the minimum time to produce a car (on average)

2. Linearizable models

RECIPROCAL MODEL

$$y_t = \alpha + \beta \frac{1}{x_t}$$

But, if $\beta<0$, then α is an upper asymptote and, consequently, the relationship exhibit an upper bound (ceiling wage).

2. Linearizable models

SEMILOG (LOG-LIN) MODEL

$$\ln(y_t) = \alpha + \beta x_t$$

The slope coefficient measures the relative change in Y for a given absolute change in the value of the explanatory variable.

If $x_t = t$, then β means the growth rate per year.

2. Linearizable models

LIN-LOG MODEL

$$y_t = \alpha + \beta \ln(x_t)$$

This is not very helpful. We can use it in those situations where the growth rate of x explains the variable y.

β represents the absolute change in y by the relative change in x. A unit increase in the log of x increases y by β units

2. Linearizable models

POLYNOMIAL MODELS

$$y_t = \alpha + \beta x_t + \gamma x_t^2$$

This is not very helpful.

- U-shape function (Long Run Average Cost function)
- Some demand functions where the income-elasticity may vary (i.e. some luxury goods).

3. Non-linear models

CES production function

- NLS, we minimize the non-linear sum of the squared residuals
- Non-linear ML, we maximize the non-linear likelihood function.
- We should employ some mathematical algorithms to this end.
- Most popular are: Gauss-Newton, Newton-Rapshon, BHHH and EM (expectation-maximization).

3. Non-linear models

Non linear estimations

$$Q = F\left[a K^{r} + (1-a) L^{r} \right]^{\frac{1}{r}}$$

Where Q is the output, L and K are labor and capital inputs, respectively, F is the productivity factor, s=1/(1-r) is the elasticity of substitution between the two inputs.

If r=1, this is a linear function. (r=0 is the Cobb-Douglas) Otherwise, we cannot linearize it and we should use non-linear techniques

3. Non-linear models

https://me.ucsb.edu/~moehlis/21.html

3. Non-linear models

Non linear estimations

Assuming the algorithm works correctly (convergence criterion is held), then the non-linear estimator has the following properties:

- It is biased
- It can be consistent
- It can be asymptotically efficient (not easy to prove)

4. Box-Cox transformations

- We are imposing a particular functional form. Instead of this, it seems sensible to let data talk and use that functional form that provide us a better fit.
- To that end, we can follow the Box-Cox approach.
- These authors suggest to use the following transformation:

$$z^{(\lambda)} = \frac{z^{\lambda}-1}{\lambda}$$

- If λ=0, the $z^{(\lambda)}$= ln(z)

4. Box-Cox transformation

Given that the values of the parameters λ are unknown, we can estimate them by using non-linear methods.

In the very simple bivariate case, this implies to estimate the following model

$$y^{(\lambda_1)} = \beta_1 + \beta_2 x^{(\lambda_2)} + u_t$$

We can do it by simply optimizing the likelihood function

4. Box-Cox transformation

We can test for particular values of the Box-Cox parameters λ, by using a LR statistic

$$LR = -2ln\ \frac{L_R}{L_{UR}} \xrightarrow{As} {\chi_r}^2$$

Where L_R and L_{UR} are the restricted (under the null hypothesis) and the unrestricted likelihood function, respectively.

4. Box-Cox transformation

However, we should note that the (concentrated) log-likelihood function can be stated as follows:

$$\ell^* = -\frac{T}{2}\ln\tilde{\sigma}^2 + (\lambda_1 - 1)\sum_{\mathrm{i}=1}^{T}\ln y_i$$

Where the last term corresponds to the Jacobian of the transformation. We can remove it by dividing the values of the dependent variable by its geometric mean.

5. Ramsey's RESET test

- Ramsey proposes a statistic to test the possible existence of misspecification problems
- RESET stands for Regression Specification Error Test
- It does not directly test for functional form problems
- But it can be used to detect if there are any neglected nonlinearities in our specification

5. Ramsey's RESET test

- The statistic is obtained as follows.Let us first estimate our model

$$y_t = x'_t\, \beta + u_t$$

and get the fitted values $\hat{y}_t = x'_t \hat{\beta}$

- We extend the original model by including squares and cubes of the fitted values ${\hat{y}_t}^2, {\hat{y}_t}^3$, … into the model and test for joint significance of added terms with a general F test

5. Ramsey's RESET test

The auxiliary regression can be written as follows:

$$y_t = x'_t\, \beta + \gamma_1 \hat{y}_t^{\,2} + \gamma_2 \hat{y}_t^{\,3} + \cdots + \gamma_m \hat{y}_t^{\,m+1} + u_t$$

The null hypothesis of the RESET test is that the model is correctly specficied

$$H_o: \gamma_1 = \gamma_2 = \ldots = \gamma_m = 0$$

Under this null hypothesis, the F-statistic follows and F(m,T-k-m) distribution

5. Ramsey's RESET test

- A drawback with RESET test is that it provides no real direction on how to proceed if the null hypothesis is rejected.

- Some have argued that RESET is a very general test for model misspecification, including unobserved omitted variables and heteroscedasticity.

- So, it should be used with some caution