Econometrics I - Basic Multiple Regression

Ryan T. Godwin

University of Manitoba

Basic Multiple Regression

The population model is:

$$\boldsymbol{y} = f\left(x_1, x_2, \ldots, x_k; \boldsymbol{\theta}\right) + \boldsymbol{\epsilon} \quad (1)$$

- $\boldsymbol{y}$ is the dependent variable or "regressand"
- $x_1, x_2, \ldots, x_k$ are the explanatory variables or "regressors"
- $\boldsymbol{\theta}$ is a parameter vector
- $\boldsymbol{\epsilon}$ is the disturbance term or the random "error"

In general f may be:

- linear or non-linear in the variables
- linear or non-linear in the parameters
- parametric or non-parametric

Questions

The population model is:

$$\boldsymbol{y} = f\left(x_1, x_2, \ldots, x_k; \boldsymbol{\theta}\right) + \boldsymbol{\epsilon}$$

1. What is the role of the error term?
2. What is random, and what is deterministic?
3. What is observable, and what is unobservable?

Classic population models

Try to determine the components of the model, and whether or not it is linear/non-linear in the regressors/parameters.

Keynes' consumption function

$$C = \beta_1 + \beta_2 Y + \epsilon$$

Cobb-Douglas production function

$$Y = AK^{\beta_2} L^{\beta_3} e^{\epsilon}$$

By taking logs, the Cobb-Douglas production function can be rewritten as:

$$\log Y = \beta_1 + \beta_2 \log K + \beta_3 \log L + \epsilon$$

where $\beta_1 = \log A$.

Classic population models

Gravity model of trade

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \epsilon_{ij}$$

Mincer earnings equation

$$\ln w = \ln w_0 + \rho s + \beta_1 x + \beta_2 x^2$$

CES production function

$$Y = \varphi \left(aK^r + (1-a)L^r \right)^{1/r} e^{\epsilon}$$

Taking logs, the CES production function is written as:

$$\log Y = \log \varphi + \frac{1}{r} \log \left(aK^r + (1-a)L^r \right) + \epsilon$$

Sample information

Suppose that we have a *sample* of n observations:

$$\{y_i; x_{i1}, x_{i2}, \ldots, x_{ik}\}; \quad i = 1, 2, \ldots, n$$

Assume:

- observed values generated by pop. model
- model is *linear in the parameters*

then:

$$\begin{aligned} y_i &= \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \\ i &= 1, \ldots, n \end{aligned} \tag{2}$$

Recall that the βs and ϵ are unobservable. So, y_i is generated by two components:

1. Deterministic component: $\Sigma_{j=1}^{k} \beta_j x_{ij}$
2. Stochastic component: ϵ_i

So, the y_i must be "realized values" of a random variable!

Interpreting the parameters in a model

How to interpret the estimated parameters? The βs in equation 2 have an important economics interpretation. For example:

$$\frac{\partial y_i}{\partial x_{1i}} = \beta_1$$

The parameters are the marginal effects of the x on y, with other factors held constant.

From Keynes' consumption function:

$$\frac{\partial C}{\partial Y} = \beta_2 = \text{MPC}$$

Depending on how the population model is specified, however, the β might not be interpreted as marginal effects.

$$\log Y = \beta_1 + \beta_2 \log K + \beta_3 \log L + \epsilon,$$

and

$$\begin{aligned} \beta_2 &= \frac{\partial \log Y}{\partial \log K} \\ &= \frac{\partial \log Y}{\partial Y} \times \frac{\partial Y}{\partial K} \times \frac{\partial K}{\partial \log K} \\ &= \frac{1}{Y} \times \frac{\partial Y}{\partial K} \times K = \frac{\partial Y / Y}{\partial K / K}, \end{aligned}$$

β_2 is the elasticity of output with respect to capital. Need to be careful about interpreting parameters.

Questions:

1. How are the parameters interpreted in a *log-linear* model? For example:

$$\log(y_i) = \beta_0 + \beta_1 x_{i1} + ... + \epsilon_i$$

2. How are the parameters interpreted in a *linear-log* model? For example:

$$y_i = \beta_0 + \beta_1 \log(x_{i1}) + ... + \epsilon_i$$

Assumptions of the Classical Linear Regression Model

- Make six "classical" assumptions, and refer back to them throughout the course.
- These simplifying assumptions are a starting point - likely not satisfied in real data.
- Main objective of this course: re-consider these assumptions:
  - are they realistic
  - can they be tested
  - what if they are wrong
  - can they be "relaxed"?

When these assumptions are violated, and we consider how to fix the resulting consequence, we will be led to different *estimation strategies*, such as *instrumental variables* estimation or *generalized least squares*.

A.1: Linearity

The model is linear in the parameters:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

Linearity in the param. allows model to be written in matrix form:

$$\boldsymbol{y} = \underset{(n \times 1)}{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} \; ; \quad \boldsymbol{\beta} = \underset{(n \times 1)}{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}} \; ; \quad X = \underset{(n \times k)}{\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}}$$

$$\boldsymbol{\epsilon} = \underset{(n \times 1)}{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

Then, we can write the model, for the full sample, as:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

If we take the i^{th} row (observation) of this model we have:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i \qquad \text{(scalar)}$$

- Vectors are in bold.
- The dimensions of vectors/matrices are written (rows $\times$ columns).
- The first subscript denotes the row, the second subscript the column.
- The above equation sometimes expressed as $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$.

Question: Which of the "classic" population models are linear in the parameters, and which can be *linearized*?

A.2: Full Rank

We assume that there are no exact linear dependencies among the columns of X (if there were, then one or more regressor is redundant). Note that X is $(n \times k)$ and $\text{rank}(X) = k$. So we are also implicitly assuming that $n > k$, since $\text{rank}(A) \leq \min\{\#rows, \#cols\}$.
What does this assumption really mean? Suppose we had:

$$y_i = \beta_1 x_{i1} + \beta_2 (2x_{i1}) + \epsilon_i$$

We can only identify, and estimate, the one function, $(\beta_1 + 2\beta_2)$. In this model, $\text{rank}(X) = k - 1 = 1$. An example which is commonly found in undergraduate textbooks, of where A.2 is violated, is the dummy variable trap.

A.3: Errors have a zero mean

Assume that, *in the population*, $\mathrm{E}(\epsilon_i) = 0 \quad ; \quad i = 1, 2, \ldots, n$. So,

$$E(\boldsymbol{\epsilon}) = E\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \mathbf{0}$$

A.4: Spherical errors

Assume that, in the population, the disturbances are generated by a process whose variance is constant (σ^2), and that these disturbances are uncorrelated with each other:

$$\text{var}(\epsilon_i) = \sigma^2; i = 1, 2, \ldots, n \quad \text{(Homoskedasticity)}$$

$$\text{cov}(\epsilon_i, \epsilon_j) = 0; \forall i \neq j \quad \text{(no Autocorrelation)}$$

Putting these assumptions together we can determine the form of the "covariance matrix" for the random vector, $\boldsymbol{\epsilon}$.

$$V(\epsilon) = E\left[(\boldsymbol{\epsilon} - E(\boldsymbol{\epsilon}))(\boldsymbol{\epsilon} - E(\boldsymbol{\epsilon}))'\right] = E\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right] = \begin{bmatrix} E(\epsilon_1\epsilon_1) & \cdots & E(\epsilon_1\epsilon_n) \\ \vdots & \ddots & \vdots \\ E(\epsilon_n\epsilon_1) & \cdots & E(\epsilon_n\epsilon_n) \end{bmatrix}$$

but...

$$E\left(\epsilon_i \epsilon_i\right) = E\left(\epsilon_i^2\right) = E\left[\left(\epsilon_i - 0\right)^2\right] = \text{var}\left(\epsilon_i\right) = \sigma^2$$

and

$$E\left(\epsilon_i \epsilon_j\right) = E\left[\left(\epsilon_i - 0\right)\left(\epsilon_j - 0\right)\right] = \text{cov}\left(\epsilon_i, \epsilon_j\right) = 0.$$

So:

$$V(\boldsymbol{\epsilon}) = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

a scalar matrix.

A.5: Generating process for X

The classical regression model assumes that the regressors are "fixed in repeated samples" (laboratory situation). Very strong assumption.

Alternatively, allow X to be random, but restrict the form of their randomness – the process that generates X is unrelated to the process that generates $\boldsymbol{\epsilon}$ in the population.

This is likely the most important assumption. We will soon see that it is imperative that X and $\boldsymbol{\epsilon}$ are statistically independent.

So, if X is random, we need to assume *strict exogeneity*:

$$\mathrm{E}(\boldsymbol{\epsilon}|X) = \mathbf{0} \tag{4}$$

which implies that X and $\boldsymbol{\epsilon}$ are uncorrelated:

$$\operatorname{cov}(\boldsymbol{x_j}, \boldsymbol{\epsilon}) = \mathbf{0} \quad ; \quad \text{for } j = 1, \ldots, k$$

Note that independence implies zero correlation, but not necessarily the other way around due to correlation measuring only *linear* dependencies between variables.

Prove that the zero correlation assumption is equivalent to each column of X being *orthogonal* to $\boldsymbol{\epsilon}$, that is, prove that A.5 at least implies that:

$$\mathrm{E}(X'\boldsymbol{\epsilon}) = \mathbf{0} \tag{5}$$

A.6: Normality of errors

$$(\boldsymbol{\epsilon}|X) \sim N\left[\mathbf{0}, \sigma^2 I_n\right]$$

This assumption is not as strong as it seems:

- often reasonable due to the Central Limit Theorem (C.L.T.)
- often not needed
- when some distributional assumption is needed, often a more general one is ok

Summary

The classical linear regression model is:

- $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $(\boldsymbol{\epsilon}|X) \sim N\left[\mathbf{0}, \sigma^2 I_n\right]$
- $\text{rank}(X) = k$
- The data generating process (DGP) of X and $\boldsymbol{\epsilon}$ are unrelated

Implications for y (if X is non-random; or conditional on X):

$$E(\boldsymbol{y}) = X\boldsymbol{\beta} + E(\boldsymbol{\epsilon}) = X\boldsymbol{\beta}$$

$$V(\boldsymbol{y}) = V(\boldsymbol{\epsilon}) = \sigma^2 I_n$$

Because linear transformations of a Normal random variable are themselves Normal, we also have:

$$\boldsymbol{y} \sim N\left[X\boldsymbol{\beta}, \sigma^2 I_n\right]$$

Some questions:

- How reasonable are the assumptions associated with the classical linear regression model?
- How do these assumptions affect the estimation of the model's parameters?
- How do these assumptions affect the way we test hypotheses about the model's parameters?
- Which of these assumptions are used to establish the various results we'll be concerned with?
- Which assumptions can be "relaxed" without affecting these results?