UiO University of Oslo

ECON 4160: Econometrics-Modelling and Systems Estimation
Lecture 2: Econometric background

Ragnar Nymoen

Department of Economics University of Oslo

21 August 2018

The reference to this lecture is:

- Chapter 2 in the textbook: *Dynamic Econometrics for Empirical Macroeconomic Modelling*.
- The Lecture note: *Background to ECON 4160: Review of key concepts of statistics, with exercises*.

Concepts for modelling relationships

- For simplicity we, consider two continuous random variables, X_i and Y_i . (Combining continuous and discrete variables is not a problem, but some special considerations are required when the dependent variable is a binary variable (0,1), ie *Logit-model*).
- X_i and Y_i are characterized by a distribution function

$$F(X, Y) = P(X_i \leq x_i \text{ and } Y_i \leq y_i)$$

- In econometric analysis we usually consider the rate of increase of the distribution function, rather than the distribution function itself. This rate of change form is called the **joint probability density function** (pdf).

$$f_{X_i,Y_i}(x_i, y_i) = \frac{\partial^2}{\partial x_i \partial y_i} P(X_i \leq x_i \text{ and } Y_i \leq y_i), \tag{1}$$

Properties of pdf and correspondence to discrete distributions

Continuous:

1. $$f_{X_i,Y_i}(x_i,y_i) \geq 0, \forall x_i, y_i$$
2. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_i,Y_i}(x_i,y_i)dx_i dy_i = 1$$
3. Independence:
$$f_{X_i,Y_i}(x_i,y_i) = f_{X_i}(x_i) \cdot f_{Y_i}(y_i)$$
4. In general:
$$f_{X_i,Y_i}(x_i,y_i) = f_{Y_i}(y_i \mid x_i) \cdot f_{X_i}(x_i)$$

Discrete:

1. Non negative point probabilities:
$$p_{x_i,y_i} \geq 0$$
2. Probabilities sum to one:
$$\sum\sum p_{x_i,y_i} = 1$$
3. $$P(X_i = x_i \text{ and } Y_i = y_i) = p_{x_i} \cdot p_{y_i}$$
4. In general:
$$P(X_i = x_i \text{ and } Y_i = y_i) = p_{x_i|y_i} \cdot p_{y_i}$$

Conditional and marginal pdf

- $f_{X_i}(x_i)$ the **marginal pdf** obtained from the **joint pdf**.
- The **conditional pdf** of Y_i given $X_i = x_i$ is defined as:

$$f_{Y_i|X_i}(y_i \mid x_i) = \frac{f_{X_i,Y_i}(x_i, y_i)}{f_{X_i}(x_i)}.$$

- The conditional pdf is symmetric:

$$f_{X_i|Y_i}(x_i \mid y_i) = \frac{f_{X_i,Y_i}(x_i, y_i)}{f_{Y_i}(y_i)}.$$

PDF and Econometric Models

Econometrics models can be seen as probability density functions "put in" model equations form.

- Regression equation model: conditional pdf,
- Structural equation model: simultaneous pdf.

To keep different types of econometrics models conceptually apart, we will often refer back to the basic identity:

$$f_{X_i,Y_i}(x_i, y_i) = f_{Y_i}(y_i \mid x_i) \cdot f_{X_i}(x_i)$$

The linear relationship between Y and X

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

is:

- a *structural equation* if ϵ_i refers to a simultaneous pdf,
- a *regression equation* if ϵ_i refers to a conditional, pdf.

Model specification

The model equation is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

a. $E\left(\epsilon_i \mid X_j\right) = 0, \forall\, i$ and j, (the symbol $\forall$ is used as short-hand for "for all")
b. $Var\left(\epsilon_i \mid X_i\right) = \sigma^2, \forall\, i$
c. $Cov\left(\epsilon_i, \epsilon_j \mid X_i, X_j\right) = 0, \forall\, i \neq j,$
d. β_0, β_1 and σ^2 are constant parameters.

where E, Var and Cov are used to denote mathematical expectation, variance and covariance.

The *conditional expectation function* follows from (2) and **a**:

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i \tag{3}$$

- From **a**:

  $$Cov(\epsilon_i, X_j) = 0, \forall \; i, j.$$

  which is known as *strict exogeneity* of the regressor.
- Remark: The book makes the point (p 34-35) that **a** is a compact way of stating two assumptions:
  1. Correct functional form of the conditional expectation function .
  2. Independence between the n pairs of (X_i, Y_i)

An equivalent specification

a*. The pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are mutually independent and identically distributed, **IID** for short.

b*. The conditional distribution of Y_i given X_i has expectation $\beta_0 + \beta_1 X_i$, and variance σ^2 which is the same for all i.

c*. The model equation is
$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n.$

d*. β_0, β_1 and σ^2 are constant parameters.

a* implies that the ordering of the variables does not matter, it is up to us how we organize the data. For example, instead of $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, we can think of the variable set as:

$$\{Y_1, Y_2, \ldots, Y_n, X_1, X_2, \ldots, X_n\},$$

where there is dependency between Y_i and X_i, but all other variable pairs are independent.

IID, together with the rest of the equivalent specification gives:

- $$E(\epsilon_i \mid X_j) = 0 \ \ \forall j \ \text{ and } i \text{ (strict exogeneity}$$
- $$Var(\epsilon_i \mid X_i) = \sigma^2 \ \text{ (homoskedasticity)}$$

The Gaussian regression model

- The regression model when the data is normally (Gaussian) distributed is an important reference case
- One reason is that the results that we can derive from the Gaussian model holds for large samples, even when exact normality does not hold.
- We replace **b*** and **c*** by the more specific assumption of conditional normality:

$$(Y_i \mid X_i) \overset{D}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2), \tag{4}$$

- It then follows:

$$(\epsilon_i \mid X_i) \overset{D}{\sim} N(0, \sigma^2). \tag{5}$$

Other properties of the bivariate Gaussian model:

$$\beta_1 = \frac{\sigma_{XY}}{\sigma_X^2}, \quad (6)$$

$$\beta_0 = \mu_Y - \beta_1 \mu_X, \quad (7)$$

$$\sigma_{Y|X}^2 = \sigma_Y^2 \left(1 - \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2}\right), \quad (8)$$

- σ_X^2 and σ_Y^2 denote the unconditional variances
- σ_{XY} denotes the covariance between X and Y
- μ_X and μ_Y are the expectations of the two variables.
- $\sigma_{Y|X}^2$ is the conditional variance of Y given X
- We often use the simplified notation $\sigma_{Y|X} \equiv \sigma^2$ in the regression model specification.

Note that the expression for the conditional variance can be written as:

$$\sigma^2 = \sigma_Y^2(1 - \rho_{XY}^2), \tag{9}$$

where

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \tag{10}$$

is the theoretical correlation coefficient, which is larger than -1 and smaller than +1:

$$-1 < \rho_{XY} < 1, \tag{11}$$

so that the conditional variance is always positive. All of these parameters are population parameters. We next review their estimation.

Likelihood function

The joint pdf of $\{Y_1, Y_2, \ldots, Y_n \mid X_1, X_2, \ldots, X_n\}$ is:

$$f\{Y_1, Y_2, \ldots, Y_n \mid X_1, X_2, \ldots, X_n\} = \prod_{i=1}^{n} \frac{1}{\sigma_{Y|X}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma_{Y|X}}\right)^2\right]. \quad (12)$$

- For known β_0, β_1 and σ^2, (12) can be used to calculate the value of the joint pdf for any given values of the random variables
- But we can turn the question around and ask: For n given values of (Y_i, X_i), what are the most likely values of the parameters β_0, β_1 and σ^2? In this interpretation (12) is called the **likelihood function**.

Maximum likelihood estimation (MLE)

To answer the question it is easier to work with the **log-likelihood function**:

$$\begin{aligned}\mathcal{L}(\beta_0, \beta_1, \sigma^2) &= \sum_{i=1}^{n}\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{\sigma^2}\right] \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\end{aligned} \tag{13}$$

As shown in Ch. 2.4.4 the Maximum Likelihood (ML) estimators of the model parameters of the regression function become:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \tag{14}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{15}$$

Estimation of the variance of the disturbance

ML estimator is:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2, \tag{16}$$

which is biased in finite samples, motivating the use of:

$$\tilde{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_i^2. \tag{17}$$

when we calculate the *t-value*:

$$\text{t-value} = \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}}, \tag{18}$$

are calculated in practice.

Method-of-Moments

- ML and OLS are two **estimation principles**
- Methods-of-Moments is a third (Ch. 2.4.5)
- For the Gaussian regression model, the three principles give the same estimators for β_0 and β_1
- This generalizes to multivariate regression, but not to the estimation of structural equations

Delta method

- Assume the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

and that we are interested in the parameter θ:

$$\theta = \frac{\beta_0}{\beta_1}. \tag{19}$$

- ML/OLS gives asymptotically unbiased (consistent) estimation of θ.
- Estimation of the variance can use the Delta-method in equation (2.55).
- Example(PCM natural rate):

$$\begin{aligned} INF_i &= \beta_1(U_t - \frac{-\beta_0}{\beta_1}) + \epsilon_i \\ &= \beta_1(U_t - U^{nat}) + \epsilon_i \end{aligned} \tag{20}$$

and define

$$U^{nat} \underset{def}{=} \frac{-\beta_0}{\beta_1}$$

as the natural rate of unemployment.

Assume that we have estimated the empirical PCM with OLS:

$$INF_i = \underset{(1.639)}{10.5088} - \underset{(0.4613)}{1.83147} U_i,$$

$$\hat{U}^{nat} = \frac{10.5088}{1.83147} = 5.7379.$$

and have used (2.55) formulae to obtain:

$$Var(\hat{U}^{nat}) \approx 0.44188.$$

An approximate 95 % confidence interval for U^{nat} is therefore:

$$[4.4\% \, ; 7.1\%]$$

(Reproduce this by working with Exercise 2.10, a computer exercise).

Model specification

We continue to use a linear model equation:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{21}$$

however the parameters now refer to a different pdf than was the case for regression, since the structural model is defined by:

$$Cov(X_i, \varepsilon_i) \neq 0,$$

The complete model specification is (exactly identified case):
Model equation (21) and

- a′ $E\left(\epsilon_i \mid Z_j\right) = 0, \forall\, i$ and j,
- b′ $Var\left(\epsilon_i \mid Z_i\right) = \sigma^2, \forall\, i$
- c′ $Cov\left(\epsilon_i, \epsilon_j \mid Z_i, Z_j\right) = 0, \forall\, i \neq j$,
- d′ $Cov(X_i, Z_i) \neq 0$, $\forall\, i$
- e′ β_0, β_1 and σ^2 are constant parameters.

Z_i is the **instrumental variable**

Instrument validity and relevance

Validity and relevance of Z_i is implied by the model specification:

$$Cov(\epsilon_i, Z_i) = 0, \text{ Instrument validity,} \quad (22)$$

$$Cov(X_i, Z_i) \neq 0, \text{ Instrument relevance.} \quad (23)$$

IV estimator

- Forming covariates between Y and Z, X and Z and ϵ and Z,
- and applying the Method of Moments gives the IV-estimator for β_1

$$\hat{\beta}_1^{IV} = \frac{\sum_{i=1}^{n} Y_i(Z_i - \bar{Z})}{\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z})}. \tag{24}$$

- From the properties of the model:

$$plim(\hat{\beta}_1^{IV}) = \beta_1, \tag{25}$$

- unlike the OLS estimator which is inconsistent for the parameters of the structural equation.
- Remark: Unbiasedness of IV ($E(\hat{\beta}_1^{IV}) = \beta_1$) does not hold.

Multivariate regression model

- No principal difference from simple regression. The conditioning is on two or more variables, rather than only one.
- A big difference in relevance: Allows estimation of **partial** effects of changes in regressors, while the simple model only estimates total effects.
- Remark: Absence of perfect **multi-collinearity** is sometimes stated as an assumption of the multivariate regression model. However perfect multi-collinearity in population pdf does not make logical sense.
- However, a high degree of multi-collinearity can be a problem for a particular sample.
- Ch 2.6 of the book presents the algebraic extension of simple regression, first the 2-regressor case, and then the general case with matrix notation.

Frisch-Waugh theorem

- Assume that we want to estimate the effect (on $\mathbf{y}$) of one group of variables (collected in a matrix $\mathbf{X}_2$ as in Ch. 2.6.2), given another group of variables, $\mathbf{X}_1$.
- The Frisch-Waugh theorem says that the following two methods gives numerically identical results:
  1. Estimate the full multivariate model:

$$\mathbf{y} = \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\epsilon}}.$$

  2. Estimate in two steps: Regress $\mathbf{y}$ on $\mathbf{X}_1$ and regress $\mathbf{X}_2$ on $\mathbf{X}_1$, and then regress the "Y on $\mathbf{X}_1$" residuals on the "$\mathbf{X}_2$ on $\mathbf{X}_1$" residuals, to obtain $\hat{\boldsymbol{\beta}}_2$.

Multivariate structural equation

- A principal difference from simple case: The structural equation may be **over-identified**.
- Meaning that more than one consistent IV estimator exists for the coefficient vector $\boldsymbol{\beta}$ of the structural equation.
- For over-identified structural equations, the GIVE estimator is asymptotically more efficient (lower variance) than any of the IV-estimators
- GIVE is equivalent with 2SLS (Two stage least squares)
  - In the first stage, optimal instruments are obtained as the fitted values of the endogenous X-variables, from regressions of the full set of instruments.
  - In the second stage, the fitted values are used as regressors.

Non homoskedastic models

- The reviewed estimation methods are only (asymptotically) efficient when the disturbances of the model equations are homoskedastic
- For models with heteroskedastic disturbances:
  - GLS is more efficient than OLS for regressoon models
  - GMM is more efficient than GIVE/2SLS

Model equations with time series

- Change in notation: Let T denote the sample size (replacing n)
- Time forces a unique ordering on the data:

$$\underbrace{(X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T)}_{T \text{ pairs from past (left) to present (right)}}.$$

- It is logically possible that the pairs of variables are mutually independent.
- In that case the fixed ordering forced on us by time, would be a mere formality and the results that we have reviewed above would apply also to time series data.
- However, as we saw in Lecture 1, it is typical that time series variables are autocorrelated.

- Therefore, the disturbance ϵ_t in;

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{26}$$

will in general be correlated with ϵ_{t-1} (residual autocorrelation).
- Hence, the IID regression model that we have reviewed, does not cover regression models with time series data.
- We need a different basis for time series. It will be based on the idea that conditioning on history can "break up" the dependence. For example, (X_t, Y_t) may become independent of (X_{t-2}, Y_{t-2}) once we have conditioned on (X_{t-1}, Y_{t-1}).
- In that case the joint distribution:

$$f(X_t, Y_t \mid X_{t-1}, Y_{t-1}),$$

or the conditional:

$$f(Y_t \mid X_t, X_{t-1}, Y_{t-1}),$$

represent a basis for modelling.

ARDL model equation

- Written as a model equation, $f(Y_t \mid X_t, X_{t-1}, Y_{t-1})$ becomes:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots, T, \quad (27)$$

- where:

$$E(\epsilon_{t+j} \mid Y_{t-1}) = 0 \text{ for } j = 0, 1, 2, ...,$$

- but:

$$E(\epsilon_{t-i} \mid Y_{t-1}) \neq 0 \text{ for } i = 1, 2, ...,$$

- which defines Y_{t-1} as a **pre-determined** variable.
- Again, the question is whether pre-determinedness represents “enough independence” to allow reliable estimation of the parameters in (27).

References

Nymoen, R. (2018). *Dynamic Econometrics for Empirical Macroeconomic Modelling*. Manuscript.