ECON 4160, Spring term 2013. Lecture 1

Ragnar Nymoen

University of Oslo

21 August 2013

Econometrics courses in our programmes

- ECON 2130 *Statistikk 1*
- ECON 3145/4150 *Introductory Econometrics*
- ECON 4136 *Applied Statistics and Econometrics*
- ECON 4160 **Econometrics-Modelling and System Estimation**
- ECON 4130 Statistics 2
- ECON 5101/02/03 Advanced coursed in time series, panel data, micro econometrics

References to Lecture 1

- Ch 1-3.5 in Davidson and MacKinnon, DM
- A lot of the material in these chapters is a presentation of results that are well known from any introductory course in econometrics.
- In particular OLS estimation and the properties of OLS estimators (e.g., Gauss-Markov theorem).
  - But in terms of matrix algebra rather than scalar notation,
  - and the are several sections about geometric interpretation.
- There is a good discussion of modelling concepts in C1, and the Frisch-Waugh-Lovell (FWL) theorem is lifted to the forefront late in Ch 2

The specification of regression models I

- We know that specification of regression model is complete when we have specified
  - a **regressand**
  - one or more **regressors**
  - the **functional form** of the **conditional expectation function** of the regressand given the regressors
  - The **distribution of the random distubance term** of the model
- In introductory econometrics we study how the theoretical properties of the OLS estimators of the parameters of the conditional expectation function depend on the assumed distribution of the intercept term. So called classical intercept properties lead to OLS estimators being BLUE for example.

The specification of regression models II

- In practical econometric modelling the specification of a regression model is "not always easy" (DM p 16).
- It involves several decisions that have consequences for:
  - (ir)relevance of the econometric model
  - (un)biased estimation of parameters
  - (un)stability of parameters, and the degree if autonomy of parameters wrt to policy interventions and other structural breaks
  - (un)reliability statistical inference (confidence intervals/hypothesis testing)

Pitfalls in applied regression modelling

Problem	**"Typical" cause**
Low relevance	Conditional expectation does not contain parameters of interest
Biased estimation	Wrong or incomplete conditioning
Unstable parameters	Wrong direction of regression
Unreliable statistical inference	Misspecification of functional form

- Conversely, if the conditional expectation contains the parameters of interest and it is correctly derived, and it is linear, then OLS estimation will give estimators that are at least consistent.
- If there is a regime shift (structural break) in the system, but the direction of regression is correct, then OLS *can* still give parameters that are invariant to the regime shift.
- OLS based statistical inference can be made reliable

Relevance of regression models I

- Ever since the "Probability Approach" by Haavelmo, the fundamental concept of econometric modelling is the joint probability distribution function of random variables, or the **joint probability density function** pdf:

  $$f(X_0, X_1, \ldots, X_k)$$

  for the $k+1$ random variables $(X_0, X_1, \ldots, X_k)$. (Not all need to be random, but the simplification notation)
- When we choose to use a regression model , we specify one random variable as a regressand and the other as regressors.
- Let

  $$Y = X_0$$

  define the regressand.

Relevance of regression models II

- We can always write the joint pdf as the product of a conditional pdf and a marginal pdf:

$$f(Y, X_1, \ldots, X_k) = f(Y \mid X_1, \ldots, X_k) \cdot f(X_1, \ldots, X_k) \quad (1)$$

  - Note that $f(X_1, \ldots, X_k)$ is a joint pdf in its own right, but it is **marginal** relative to the "full" pdf on the left hand side.

- The conditional pdf is the foundation of regression modelling. From $f(Y \mid X_1, \ldots, X_k)$ we can always construct the **conditional expectation function**:

$$E(Y \mid X_1, \ldots, X_k)$$

  and the **disturbance**

$$\varepsilon = Y - E(Y \mid X_1, \ldots, X_k).$$

Relevance of regression models III

- For a realization of the X-vector, the conditional expectation $E(Y \mid x_1, \dots, x_k)$ is deterministic. But we can consider the expectation for any realization of X, so $E(Y \mid X_1, \dots, X_k)$ is a deterministic function of the vector with k random X variables.
- Hence the **Law of iterated expectations** (p 14), here applied to ε;

$$E(E(\varepsilon \mid X_1, \dots, X_k)) = E(\varepsilon)$$

$$E([Y - E(Y \mid X_1, \dots, X_k)] \mid X_1, \dots, X_k) = 0 \Longrightarrow E(\varepsilon) = 0$$

Relevance of regression models IV

- From the **regression model**

$$Y = E(Y \mid X_1, \ldots, X_k) + \varepsilon$$

  We estimate the parameters of $E(Y \mid X_1, \ldots, X_k)$ by OLS, if the function is linear, or NLS (as we shall see)
- The parameters of $E(Y \mid X_1, \ldots, X_k)$ should correspond to the *parameters of interest* for our research. Examples: The average response in Y to a change in one X_j. The best prediction of Y given $x_1, \ldots, x_k$.
- If such correspondence does not exists, the relevant model is instead to model the joint pdf $f(Y, X_1, \ldots, X_k)$. We call this **system modelling.**
- Sometimes the difference can be subtle

Relevance of regression models V

- For example, we often focus on a "single equation in the joint pdf. It can "look like" a linear regression model

$$Y = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \epsilon \quad (2)$$

but because $E(\epsilon \mid X_1, X_2) \neq 0$, it is not a regression equation but a **structural equation** that belongs to the joint pdf.

- As the title of the course suggest, we will spend quite a lot of time to develop structural modelling of the joint pdf.
- However, we begin with the regression model (in matrix notation), since the purpose of the econometric modelling is often to estimate the average response of Y to a marginal change in X_j. And this **slope parameter** can always be obtained from $E(Y \mid X_1, \ldots, X_k)$.

Biased estimation I

Problem	"Typical" cause
Low relevance	Conditional expectation does not contain parameters of interest
Biased estimation	**Wrong or incomplete conditioning**
Unstable parameters	Wrong direction of regression
Unreliable statistical inference	Misspecification of functional form

- When it is relevant to model the conditional expectation, it must still be consistent with the joint distribution, otherwise

$$f(Y, X_1, \ldots, X_k) \neq f(Y \mid X_1, \ldots, X_{k-1}) \cdot f(X_1, \ldots, X_{k-1})$$

as a consequence of "incomplete conditioning", or

$$f(Y, X_1, \ldots, X_k) \neq f(Y \mid X_1, \ldots, X_{k-1}, Z) \cdot f(X_1, \ldots, X_{k-1}, Z)$$

as a consequence of "wrong conditioning".

Biased estimation II

- The result is known as **omitted variables bias** of the OLS estimators, as when the chosen conditional model is

$$Y_i = \gamma_0 + \gamma_1 X_i + \epsilon_\iota$$

while the correct conditional model, relative to $f(Y, X_1, X_2)$ is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_\iota$$

- In practical modelling, a main cause of wrong or incomplete modelling, is that the researcher has not formulated a clear picture of the joint distribution, i.e., the economic system.

Unstable parameters (direction of regression) I

- Parameter instability may stem from:
- Irrelevance of the regression model (this is the *Lucas-critique* that we will review later)
- Wrong conditioning
- Mistaken direction of the regression:
  - If there are regime-shifts in the economic system (the joint pdf), $f(Y, X)$, then **at most** one of the two conditional regression models can have parameters that are invariant to (autonomous) the regime shift.
  - Examples:
    - Is the Phillips-curve a model for the rate of inflation or a short-run aggregate supply curve?
    - Can consumption be a stable conditional function of income when there are changes in income anticipations?

Unreliable statistical inference (functional form) I

- Even if the regression model is relevant and the conditioning is correct, the question about the linearity of the conditional expectations function
- Comment to DM example in section 1.3

$$Y_t = E(Y_t \mid X_t) + u_t$$

where

$$E(Y_t \mid X_t) = \beta_1 + \beta_2 X_t$$

It then follows that

$$E(u_t \mid X_t) = 0.$$

Unreliable statistical inference (functional form) II

Now consider that the true model is

$$Y_t = \gamma_1 + \gamma_2 X_t + \gamma_3 X_t^2 + v_t \tag{1.18}$$

with

$$E(v_t \mid X_t) = E(v_t \mid X_t^2) = 0$$

DM actually write (1.18) as

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_t^2 + v_t \tag{1.18}$$

Even then, when we consider u_t conditional on X_t and the information set in (1.18):

$$\begin{aligned} E(u_t \mid X_t, (1.18)) &= E(\beta_1 + \beta_2 X_t + \beta_3 X_t^2 + v_t - \beta_1 - \beta_2 X_t \mid X_t^2, (1.18) \\ &= \beta_3 X_t^2 + E(v_t \mid X_t^2, (1.18)) \\ &= \beta_3 X_t^2 \end{aligned}$$

Unreliable statistical inference (functional form) III

- This shows that orthogonality between disturbance and regressors only holds conditional on the **information set** of a regression model.
- The disturbance is in general correlated with variables that are part of relevant extensions of an information set.
- In this example, the extension is in terms the squared regressor, so another interpretation is that of misspecified functional form:
- Moreover, u_t is heteroskedastic which in any case makes inference about β_1 unreliable:

$$Var(u_t) = \beta_3^2 X_t^4 + \sigma_v^2$$

Simulation example I

Create a data set for the DM example with the following Data Generating Process.(DGP)

```
algebra{
eps1=rann();
eps2=rann();
X=eps2;
"X^2"=X*X;
Y = 0.5 + 1*X-0.2*X^2+eps1 ;
}
```

Then investigate $E(u_t \mid X_t) = 0$ and $E(u_t \mid X_t^2, (1.18)) = \beta_3 X_t^2$ empirically.

Information set. Encompassing I

- DM uses the concept **information set**, symbolized by Ω_t, to denote a *longer list* of variables than $(X_0, X_1, \ldots, X_k)$ that are used in the (final) regression model
- With the aid of joint pdf decomposition, we can express this as

$$
\begin{aligned}
& f(Y, X_1, \ldots, X_k, X_{k+1}, \ldots, X_l) \\
& = f(Y, X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_l) \cdot f(X_{k+1}, \ldots, X_l)
\end{aligned}
$$

instead of (1) which we started with above.

Information set. Encompassing II

- If $(Y, X_1, \ldots, X_k)$ are independent from $(X_{k+1}, \ldots, X_l)$ we have

$$f(Y, X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_l) = f(Y, X_1, \ldots, X_k)$$

and we have

$$\begin{aligned} f(Y, X_1, \ldots, X_k, X_{k+1}, \ldots, X_l) &= f(Y, X_1, \ldots, X_k) \cdot f(X_{k+1}, \ldots, X_l) \\ &= \underbrace{f(Y \mid X_1, \ldots, X_k)}_{\substack{\downarrow \\ Y = E(Y \mid X_1, \ldots, X_k) + \varepsilon}} \cdot f(X_1, \ldots, X_k) \cdot f(X_{k+1}, \ldots, X_l) \end{aligned}$$

- It it easy to predict that if information sets are imprecisely formulated, or even implicit, different researchers will end up with different regression models for the same variable.

Information set. Encompassing III

- Therefore an important part of applied research is to **encompass** earlier models and findings: A new model should explain the results of existing models!
- Encompassing has both informal (in your master thesis you will review existing studies and discuss you results in the light of those studies), and formal aspects: later in the course we will introduce **encompassing tests**.

Endogenous and exogenous variables I

- The distinction between endogenous and exogenous variables is important in both economics and in econometrics
- In regression models, all regressors are exogenous, or at least **predetermined,** in the sense that they are uncorrelated with the disturbances, at least asymptotically
- Endogeneity problems, inconsistent parameter estimation, do however arise if we wrongly apply OLS to a structural equation of the joint distribution, i.e., the economic systemic.
- We will have plenty of opportunity to train ourselves in being precise about this.

The regression model in matrix notation I

Let $\mathbf{X}$ be a $n \times k$ matrix with the regressors of the model

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

where $\mathbf{y}$ is $n \times 1$and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector with disturbances and the parameter vector $\boldsymbol{\beta}$ is $k \times 1$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

A typical row in this equation is

$$Y_i = \mathbf{X}_i\, \boldsymbol{\beta} + \varepsilon_i = \sum_{j=1}^{k} X_{ij}\beta_j + \varepsilon_i,\; i = 1, 2, \dots, n \tag{4}$$

The regression model in matrix notation II

if we let $\mathbf{X}_i$ denote the i^{th} row in $\mathbf{X}$. ($1 \times k$ matrix)
As usual, unless both the regressand and all the regressors are measured as deviations from mean, there us an intercept in the model. When we need to make this explicit, we partition $\mathbf{X}$ as

$$\mathbf{X} = \left[\begin{array}{ccc} \boldsymbol{\iota} & \vdots & \mathbf{X}_2 \end{array} \right]$$

The regression model in matrix notation III

where

$$\boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}$$

$$\mathbf{X}_2 = \begin{bmatrix} X_{12} & \dots & X_{1k} \\ X_{22} & \dots & X_{2k} \\ \vdots & & \vdots \\ X_{n2} & \dots & X_{nk} \end{bmatrix}$$

OLS estimator I

- By solving the question B to the first seminar, you will show that both the Method-of-Moments (MM) and the Ordinary Least Squares (OLS) principle gives the estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5}$$

  for $\boldsymbol{\beta}$.
- Here, $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of the $\mathbf{X}'\mathbf{X}$ matrix with (uncentered) moments between the regressors.
- For the inverse to exist, $rank(\mathbf{X}'\mathbf{X}) = k$, (full rank). This is the generalization of the "absence of perfect multicollinearity" condition.

OLS estimator II

- DM uses $\mathbf{X}^{\mathsf{T}}$ as symbol for the **transpose**. We use the more common $'$ notation, which also avoids confusion with T for the sample size of a sample with time series data.
- By solving the first Exercise B to the first seminar, you can show that $\hat{\boldsymbol{\beta}}' = \begin{pmatrix} \hat{\beta}_1 & \vdots & \hat{\boldsymbol{\beta}}_2' \end{pmatrix}$ and

$$\hat{\boldsymbol{\beta}}_2 = \left[(\mathbf{X}_2 - \bar{\mathbf{X}}_2)'(\mathbf{X}_2 - \bar{\mathbf{X}}_2)\right]^{-1} (\mathbf{X}_2 - \bar{\mathbf{X}}_2)'\mathbf{y} \tag{6}$$

  In $\bar{\mathbf{X}}_2$, the typical row is $\boldsymbol{\iota}\bar{X}_k$, $k = 1, 2, \ldots, k-1$.
- (6) is the generalization of our old friend the "x-deviation from mean" form of OLS estimators that we have studied in detail for the case of $k = 1$ and $k = 2$!

Properties of OLS estimators I

- (5) and (6) are "only" matrix formulations of the OLS estimators for multiple regression that we know from before, it is clear that all the properties that we know from an introductory course still hold.
- Specifically

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

  reminds us that the conditional expectation $E(\hat{\boldsymbol{\beta}} \mid \mathbf{X})$ and variance $Var(\hat{\boldsymbol{\beta}} \mid \mathbf{X})$ depend on the assumption about the disturbances in $\boldsymbol{\varepsilon}$.
- Ch 3.1-3.5 in DM is good exposition for review.

Two important matrices I

- From Ch 2.3 we highlight two important matrices in regression theory
- The "residual maker"

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{7}$$

  plays a central role in many derivations.
- The name stems for the fact that

$$\begin{aligned}\mathbf{My} &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y} - \mathbf{X}\hat{\beta} \equiv \hat{\varepsilon}\end{aligned}$$

Two important matrices II

The following properties are worth noting:

$$\begin{aligned}
\mathbf{M} &= \mathbf{M}', \text{ symmetric matrix} \\
\mathbf{M}^2 &= \mathbf{M}, \text{ idempotent matrix} \\
\mathbf{M}\mathbf{X} &= \mathbf{0}, \text{ regression of } X \text{ on } X \text{ gives perfect fit}
\end{aligned}$$

- The prediction matrix

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{8}$$

gives

$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

It is also symmetric and idempotent.

Two important matrices III

- **M** and **P** are orthogonal:

$$\mathbf{MP} = \mathbf{PM} = \mathbf{0}$$

DM say that the **annihiliate** each other.

M and **P** are **complementary projections**

$$\mathbf{M} + \mathbf{P} = \mathbf{I} \tag{9}$$

which gives

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\varepsilon} = \mathbf{Py} + \mathbf{My}$$

TSS, SSE, and all that I

From

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$$

we get

$$\begin{aligned}\mathbf{y}'\mathbf{y} &= (\mathbf{y}'\mathbf{P} + \mathbf{y}'\mathbf{M})(\mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}) \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\end{aligned}$$

Written out, this is:

$$\underbrace{\sum_{i=1}^{n} Y_i{}^2}_{\text{"}TSS\text{"}} = \underbrace{\sum_{i=1}^{n} \hat{Y}_i{}^2}_{\text{"}ESS\text{"}} + \underbrace{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}_{RSS} \tag{10}$$

TSS, SSE, and all that II

You may be more used to write this famous decompostion as:

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})^2}_{ESS} + \underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}_{RSS} \qquad (11)$$

There is no conflict here, since

$$\mathbf{X}'\hat{\varepsilon} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) \underset{\substack{MM \\ OLS}}{=} \mathbf{0},$$

and, when there is an intercept in the regression:

$$\mathbf{X} = \left[\; \boldsymbol{\iota} \;\vdots\; \mathbf{X}_2 \;\right]$$

TSS, SSE, and all that III

this gives

$$\boldsymbol{\iota}'\hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \tag{12}$$

in the first row of $\mathbf{X}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, and therefore:

$$\bar{Y} = \overline{\hat{Y}}. \tag{13}$$

Using this, you can show that (11) can be written as (10).

TSS, SSE, and all that IV

- Memo: Alternative ways of writing *RSS*:

$$\hat{\varepsilon}'\hat{\varepsilon} = (\mathbf{y}'\mathbf{M}')\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\hat{\varepsilon} = \hat{\varepsilon}'\mathbf{y} \tag{14}$$

$$\hat{\varepsilon}'\hat{\varepsilon} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} \tag{15}$$

$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}.$$

TSS, SSE, and all that V

- The coefficient of determination (R^2) is defined with reference to (11):

$$R^2 = 1 - \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$
$$= 1 - \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{(\mathbf{M}_\iota \mathbf{y})'(\mathbf{M}_\iota \mathbf{y})}$$

$$\mathbf{M}_\iota = \mathbf{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'$$

where we have "sneaked in" the idempotent centering matrix $\mathbf{M}_\iota$, which we will come back to when we talk about the Frisch-Waugh theorem.