Additional Cheat Sheet

By Marcelo Moreno - King Juan Carlos University

As part of the Econometrics Cheat Sheet Project

OLS in matrix notation

The general econometric model:

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + u_i$$

Can be written in matrix notation as:

$$y = X\beta + u$$

Let's call $\hat{u}$ the vector of estimated residuals ($\hat{u} \neq u$):

$$\hat{u} = y - X\hat{\beta}$$

The objective of OLS is to minimize the SSR:

$$\text{Min SSR} = \text{Min} \sum_{i=1}^n \hat{u}_i^2 = \text{Min } \hat{u}^T \hat{u}$$

- Defining $\hat{u}^T \hat{u}$:

$$\hat{u}^T \hat{u} = (y - X\hat{\beta})^T (y - X\hat{\beta}) =$$
$$= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta}$$

- Minimizing $\hat{u}^T \hat{u}$:

$$\frac{\partial \hat{u}^T \hat{u}}{\partial \hat{\beta}} = -2X^T y + 2X^T X \hat{\beta} = 0$$
$$\hat{\beta} = (X^T X)^{-1} (X^T y)$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \dots & \sum x_k \\ \sum x_1 & \sum x_1^2 & \dots & \sum x_1 x_k \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_k & \sum x_k x_1 & \dots & \sum x_k^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y \\ \sum y x_1 \\ \vdots \\ \sum y x_k \end{bmatrix}$$

The second derivative is $\frac{\partial^2 \hat{u}^T \hat{u}}{\partial \hat{\beta}^2} = X^T X > 0$ (is a min.)

Variance-covariance matrix of $\hat{\beta}$

$$\text{Var}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1} = \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \dots & \text{Cov}(\hat{\beta}_0, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \dots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \dots & \text{Var}(\hat{\beta}_k) \end{bmatrix}$$

where: $\hat{\sigma}^2 = \frac{\hat{u}^T \hat{u}}{n-k-1}$

The standard errors are in the diagonal of:

$$\text{se}(\hat{\beta}) = \sqrt{\text{Var}(\hat{\beta})}$$

Error measures

- $\text{SSR} = \hat{u}^T \hat{u} = y^T y - \hat{\beta}^T X^T y = \sum (y_i - \hat{y}_i)^2$
- $\text{SSE} = \hat{\beta}^T X^T y - n\overline{y}^2 = \sum (\hat{y}_i - \overline{y})^2$
- $\text{SST} = \text{SSR} + \text{SSE} = y^T y - n\overline{y}^2 = \sum (y_i - \overline{y})^2$

R-squared

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$$

The R-squared corrected by degrees of freedom:

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2)$$

Variance-covariance matrix of $\hat{u}$

$$\text{Var}(\hat{u}) = \begin{bmatrix} \text{Var}(\hat{u}_1) & \text{Cov}(\hat{u}_1, \hat{u}_2) & \dots & \text{Cov}(\hat{u}_1, \hat{u}_n) \\ \text{Cov}(\hat{u}_2, \hat{u}_1) & \text{Var}(\hat{u}_2) & \dots & \text{Cov}(\hat{u}_2, \hat{u}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{u}_n, \hat{u}_1) & \text{Cov}(\hat{u}_n, \hat{u}_2) & \dots & \text{Var}(\hat{u}_n) \end{bmatrix}$$

When there is no heterocedasticity and no auto-correlation, the variance-covariance matrix of u has the form:

$$\text{Var}(\hat{u}) = \hat{\sigma}_u^2 \cdot I_n = \begin{bmatrix} \hat{\sigma}_u^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\sigma}_u^2 \end{bmatrix}$$

where I_n is an identity matrix of $n \times n$ elements.

When there is heterocedasticity and auto-correlation, the variance-covariance matrix of u has the form:

$$\text{Var}(\hat{u}) = \hat{\sigma}_u^2 \cdot \Omega = \begin{bmatrix} \hat{\sigma}_{u_1}^2 & \hat{\sigma}_{u_{12}} & \dots & \hat{\sigma}_{u_{1n}} \\ \hat{\sigma}_{u_{21}} & \hat{\sigma}_{u_2}^2 & \dots & \hat{\sigma}_{u_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{u_{n1}} & \hat{\sigma}_{u_{n2}} & \dots & \hat{\sigma}_{u_n}^2 \end{bmatrix}$$

where $\Omega \neq I_n$

Incorrect functional form

- Ramsey's RESET test for functional from. H_0: the model is correctly specified when compared to a model. The procedure:
  1. Estimate the original model and obtain $\hat{y}$ and R^2.
  $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k$$
  2. Estimate a new model adding powers of $\hat{y}$ and obtain the new R^2_{new} For example, adding m powers of $\hat{y}$:
  $$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + ... + \tilde{\beta}_k x_k + \tilde{\beta}_{k+1} \hat{y}^2 + ... + \tilde{\beta}_{k+(m-1)} \hat{y}^m$$
  3. Define the statistic of contrast:
  $$F = \frac{R^2_{new} - R^2}{1 - R^2_{new}} \frac{n - (k+m) - 1}{m} \sim F_{m-1, n-k-m}$$
  where m = number of powers of $\hat{y}$, k = number of parameters associated with x

Variable omission

Most of the time, is hard to get all relevant variables for an analysis. For example, a true model with all variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

The estimated model (with the available variables):

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + u$$

Omission of variables provoke OLS bias and inconsistency. Depending of the correlation between x_1 and x_2 and the sign of β_2, the bias on $\tilde{\beta}_1$ could be:

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	$(+)$ bias	$(-)$ bias
$\beta_2 < 0$	$(-)$ bias	$(+)$ bias

$(+)$ bias $\to$ $\tilde{\beta}_1$ will be higher than it should be (it includes the effect of x_2). $\tilde{\beta}_1 > \beta_1$

$(-)$ bias $\to$ $\tilde{\beta}_1$ will be lower than it should be (it includes the effect of x_2). $\tilde{\beta}_1 < \beta_1$

If $\text{Corr}(x_1, x_2) = 0$, there is no bias on β_1, because the effect of x_2 will be fully picked up by the error term, u.

There are two approaches to solve the problem:

- Make use of proxy variables.
- Make use of instrumental variables (IV).

Proxy variables

Is the approach when a relevant variable is not available for the model because is non-observable, and there is no data.

- A proxy variable is something related with the non-observable variable that has data available.

For example, the intellectual coefficient is a proxy variable for a subject's capacity (non-observable).

Instrumental variables (IV)

When proxy variables are not available, the alternative approach is to look for a variable, let's call it z, that has a relation with x. The z variable must meet the following requirements to be called an Instrumental Variable (IV):

$$\text{Cov}(z, u) = 0 \text{ (exogeneity)}$$
$$\text{Cov}(z, x) \neq 0 \text{ (relevance)}$$

This method let the omitted variable in the error term, but instead of estimate the model by OLS, it utilizes a method that recognize the presence of an omitted variable. This method can also solve error measurements.

TSLS

This is a method to estimate a model with multiple instrumental variables.

The $\text{Cov}(z, u) = 0$ requirement can be relaxed, but there has to be a minimum of variables that satisfies it.

Can have multicollinearity problems.

Some tests:

- Endogeneity tests: is TSLS better than OLS when there are no endogenous variables? Do we really need TSLS? $\to$ Hausman test $\to$ H_0: OLS is consistent (it is better to use OLS).
- Is there too many IV? $\to$ Sagan test $\to$ H_0: all IV seem ok

Information criterion

It is used to compare models with different number of parameters (k). The general formula:

$$\mathrm{Cr}(k) = \log(\tfrac{\mathrm{SSR}}{n}) + c_n \varphi(k)$$

where:

- SSR is the Sum of Squared Residuals from a model of order k.
- c_n is a sequence indexed by the sample size.
- $\varphi(k)$ is a function that penalizes large k orders.

The order of k that minimizes the criterion is chosen.

The $c_n \varphi(k)$ function:

- Akaike: $\mathrm{AIC}(k) = \log(\frac{\mathrm{SSR}}{n}) + \frac{2}{n} k$
- Hannan-Quinn: $\mathrm{HQ}(k) = \log(\frac{\mathrm{SSR}}{n}) + \frac{2\log(\log(n))}{n} k$
- Schwarz: $\mathrm{Sc}(k) = \log(\frac{\mathrm{SSR}}{n}) + \frac{\log(n)}{n} k$

$\mathrm{Sc}(k) \leq \mathrm{HQ}(k) \leq \mathrm{AIC}(k)$

Incorrect functional forms

Ramsey RESET test: it test the specification errors of a regression. H_0: the model is correctly specified.

VAR

The VAR model general form:

$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + B_0 x_t + ... + B_q x_{t-q} + C D_t + u_t$

where:

- $y_t = (y_{1t}, ..., y_{Kt})'$ is a vector of K observable endogenous variables.
- $x_t = (x_{1t}, ..., x_{Mt})'$ is a vector of M observable exogenous or unmodelled variables.
- D_t contains all deterministic variables which may consist of a constant, a linear trend, seasonal dummy variables...
- u_t is a K-dimensional unobservable zero mean white noise process with positive definite covariance matrix $E(u_t u_t' = \Sigma_u)$.
- The A_i, B_j and C are parameter matrices of suitable dimension.

For example, a model with two endogenous variables (with two lags), an exogenous contemporaneous variable, a constant ($const$) and a trend ($Trend_t$):

$$y_{1t} = a_{11,1} y_{1,t-1} + a_{12,1} y_{2,t-1} + a_{11,2} y_{1,t-2} + a_{12,2} y_{2,t-2} + b_{11} x_t + c_{11} + c_{12} Trend_t + u_{1t}$$

$$y_{2t} = a_{21,1} y_{2,t-1} + a_{22,1} y_{1,t-1} + a_{21,2} y_{2,t-2} + a_{22,2} y_{1,t-2} + b_{21} x_t + c_{21} + c_{22} Trend_t + u_{2t}$$

For example, the equations:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = A_1 \cdot \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + A_2 \cdot \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + B_0 \cdot \begin{bmatrix} x_t \end{bmatrix} + C \cdot \begin{bmatrix} const \\ Trend_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \cdot \begin{bmatrix} x_t \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \cdot \begin{bmatrix} const \\ Trend_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

VECM

$$\Delta y_t = \Pi^* \begin{bmatrix} y_{t-1} \\ D^{co}_{t-1} \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_p \Delta y_{t-p} + B_0 x_t + ... + B_q x_{t-q} + C D_t + u_t$$

where:

- $\Pi^* = \alpha \beta^{T*}$.
- D^{co}_t contains all deterministic terms included in the cointegration relations.
- D_t contains all remaining deterministic variables.

For example. The matrix form:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \Pi^* \begin{bmatrix} y_{t-1} \\ const \end{bmatrix} + \Gamma_1 \cdot \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \end{bmatrix} + B_0 \cdot \begin{bmatrix} x_t \end{bmatrix} + C \cdot \begin{bmatrix} Trend_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \alpha \left[\beta^T \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + C^* \begin{bmatrix} const \end{bmatrix} \right] + \Gamma_1 \cdot \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \end{bmatrix} + B_0 \cdot \begin{bmatrix} x_t \end{bmatrix} + C \cdot \begin{bmatrix} Trend_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \end{bmatrix} \left[\begin{bmatrix} \beta_{11} & \beta_{21} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} c^* \end{bmatrix} \begin{bmatrix} const \end{bmatrix} \right] + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \cdot \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \cdot \begin{bmatrix} x_t \end{bmatrix} + \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} \cdot \begin{bmatrix} Trend_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

The VECM model general form:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + ... + \Gamma_p \Delta y_{t-p} + B_0 x_t + ... + B_q x_{t-q} + C D_t + u_t$$

where:

- $y_t = (y_{1t}, ..., y_{Kt})'$ is a vector of K observable endogenous variables.
- $x_t = (x_{1t}, ..., x_{Mt})'$ is a vector of M observable exogenous or unmodelled variables.
- $\Pi = \alpha \beta'$. Suppose $\mathrm{rk}(\Pi) = r = \mathrm{rk}(\alpha) = \mathrm{rk}(\beta)$
- D_t contains all deterministic variables which may consist of a constant, a linear trend, seasonal dummy variables...
- u_t is a K-dimensional unobservable zero mean white noise process with positive definite covariance matrix $E(u_t u_t' = \Sigma_u)$.
- The A_i, B_j and C are parameter matrices of suitable dimension.