Additional Cheat Sheet

By Marcelo Moreno - King Juan Carlos University
As part of the Econometrics Cheat Sheet Project

OLS matrix notation

The general econometric model:

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

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:

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

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

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

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

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

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

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

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

Has the following form:

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

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

The standard errors are in the diagonal of:

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

Error measurements

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

Variance-covariance matrix of u

Has the following shape:

$$\text{Var}(u) = \begin{bmatrix} \text{Var}(u_1) & \text{Cov}(u_1, u_2) & \dots & \text{Cov}(u_1, u_n) \\ \text{Cov}(u_2, u_1) & \text{Var}(u_2) & \dots & \text{Cov}(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(u_n, u_1) & \text{Cov}(u_n, u_2) & \dots & \text{Var}(u_n) \end{bmatrix}$$

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

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

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 shape:

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

where $\Omega \neq I_n$.

- Heterocedasticity: $\text{Var}(u) = \sigma_{u_i}^2 \neq \sigma_u^2$
- Auto-correlation: $\text{Cov}(u_i, u_j) = \sigma_{u_{ij}} \neq 0, \ \forall i \neq j$

Variable omission

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

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

where $\beta_2 \neq 0$, v is the error term and $\text{Cov}(v|x_1, x_2) = 0$.

The model with the available variables:

$$y = \alpha_0 + \alpha_1 x_1 + u$$

where $u = v + \beta_2 x_2$.

Relevant variable omission causes OLS estimators to be **biased** and **inconsistent**, because there is no weak exogeneity, $\text{Cov}(x_1, u) \neq 0$. Depending on the $\text{Corr}(x_1, x_2)$ and the sign of β_2, the bias on $\hat{\alpha}_1$ 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: $\hat{\alpha}_1$ will be higher than it should be (it includes the effect of x_2) $\to \hat{\alpha}_1 > \beta_1$
- (−) bias: $\hat{\alpha}_1$ will be lower than it should be (it includes the effect of x_2) $\to \hat{\alpha}_1 < \beta_1$

If $\text{Corr}(x_1, x_2) = 0$, there is no bias on $\hat{\alpha}_1$, because the effect of x_2 will be fully picked up by the error term, u.

Variable omission correction

Proxy variables

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

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

For example, the GDP per capita is a proxy variable for the life quality (non-observable).

Instrumental variables

When the variable of interest (x) is observable but **endogenous**, the proxy variables approach is no longer valid.

- An **instrumental variable** (IV) **is an observable variable** (z) that is **related** with the variable of interest that is endogenous (x), and meet the **requirements**:

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

Instrumental variables let the omitted variable in the error term, but instead of estimate the model by OLS, it utilizes a method that recognizes the presence of an omitted variable. It can also solve error measurement problems.

- **Two-Stage Least Squares** (TSLS) 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.

  The TSLS **estimation procedure** is as follows:

  1. Estimate a model regressing x by z using OLS, obtaining $\hat{x}$:
  $$\hat{x} = \hat{\pi}_0 + \hat{\pi}_1 z$$
  2. Replace x by $\hat{x}$ in the final model and estimate it by OLS:
  $$y = \beta_0 + \beta_1 \hat{x} + u$$

  There are some important things to know about TSLS:

  - TSLS estimators are less efficient than OLS when the explanatory variables are exogenous. The **Hausman test** can be used to check it:
  $$H_0\text{: OLS estimators are consistent.}$$
  If H_0 is accepted, the OLS estimators are better than TSLS and vice versa.
  - There could be some (or all) instrument that are not valid. This is known as over-identification, **Sargan test** can be used to check it:
  $$H_0\text{: all instruments are valid.}$$

Information criterion

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

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

where:

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

Is interpreted as the relative amount of information lost by the model. The p order that min. the criterion is chosen.

There are different $c_n\varphi(p)$ functions:

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

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

The non-restricted hypothesis test

Is an alternative to the F test when there are few hypothesis to test on the parameters. Let β_i, β_j be parameters, $a, b, c \in \mathbb{R}$ are constants.

- $H_0 : a\beta_i + b\beta_j = c$
- $H_1 : a\beta_i + b\beta_j \neq c$

$$\text{Under } H_0: \quad t = \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\sqrt{\mathrm{Var}(a\hat{\beta}_i + b\hat{\beta}_j)}}$$

$$= \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\sqrt{a^2\mathrm{Var}(\hat{\beta}_i) + b^2 \cdot \mathrm{Var}(\hat{\beta}_j) \pm 2ab\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j)}}$$

If $|t| > |t_{n-k-1,\alpha/2}|$, there is evidence to reject H_0.

ANOVA

Decompose the total sum of squared in sum of squared residuals and sum of squared explained: $\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$

Variation origin	Sum Sq.	df	Sum Sq. Avg.
Regression	SSE	k	SSE$/k$
Residuals	SSR	$n-k-1$	SSR$/(n-k-1)$
Total	SST	$n-1$	

The F statistic:

$$F = \frac{\text{SSA of SSE}}{\text{SSA of SSR}} = \frac{\mathrm{SSE}}{\mathrm{SSR}} \cdot \frac{n-k-1}{k} \sim F_{k,n-k-1}$$

If $F_{k,n-k-1} < F$, there is evidence to reject H_0.

Incorrect functional form

To check if the model **functional form** is correct, we can use **Ramsey's RESET** (Regression Specification Error Test). It test the original model vs. a model with variables in powers.

$$H_0\text{: the model is correctly specified.}$$

Test procedure:

1. Estimate the original model and obtain $\hat{y}$ and R^2:
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$
2. Estimate a new model adding powers of $\hat{y}$ and obtain the new R^2_{new}:
$$\tilde{y} = \hat{y} + \tilde{\gamma}_2\hat{y}^2 + \cdots + \tilde{\gamma}_l\hat{y}^l$$
3. Define the test statistic, under $\gamma_2 = \cdots = \gamma_l = 0$ as null hypothesis:
$$F = \frac{R^2_{\text{new}} - R^2}{1 - R^2_{\text{new}}} \cdot \frac{n-(k+1)-l}{l} \sim F_{l,n-(k+1)-l}$$

If $F_{l,n-(k+1)-l} < F$, there is evidence to reject H_0.

Logistic regression

When there is a binary (0, 1) dependent variable, the linear regression model is no longer valid, we can use logistic regression instead. For example, a **logit model**:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i + u_i)}} = \frac{e^{\beta_0 + \beta_1 x_i + u_i}}{1 + e^{\beta_0 + \beta_1 x_i + u_i}}$$

where $P_i = \mathrm{E}(y_i = 1 \mid x_i)$ and $(1 - P_i) = \mathrm{E}(y_i = 0 \mid x_i)$

The **odds ratio** (in favor of $y_i = 1$):

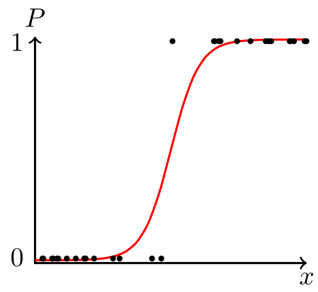
$$\frac{P_i}{1 - P_i} = \frac{1 + e^{\beta_0 + \beta_1 x_i + u_i}}{1 + e^{-(\beta_0 + \beta_1 x_i + u_i)}} = e^{\beta_0 + \beta_1 x_i + u_i}$$

Taking the natural logarithm of the odds ratio, we obtain the **logit**:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 x_i + u_i$$

P_i is between 0 and 1, but L_i goes from $-\infty$ to $+\infty$.

If L_i is positive, it means that when x_i increments, the probability of $y_i = 1$ increases, and vice versa.



Statistical definitions

Let ξ, η be random variables, $a, b \in \mathbb{R}$ constants, and P denotes probability.

Mean

Definition: $\quad E(\xi) = \sum_{i=1}^n \xi_i \cdot P[\xi = \xi_i]$

Population mean: $\quad \mathrm{E}(\xi) = \dfrac{1}{N}\sum_{i=1}^N \xi_i$

Sample mean: $\quad \mathrm{E}(\xi) = \dfrac{1}{n}\sum_{i=1}^n \xi_i$

Some properties:

- $\mathrm{E}(a) = a$
- $\mathrm{E}(\xi + a) = \mathrm{E}(\xi) + a$
- $\mathrm{E}(a \cdot \xi) = a \cdot \mathrm{E}(\xi)$
- $\mathrm{E}(\xi \pm \eta) = \mathrm{E}(\xi) + \mathrm{E}(\eta)$
- $\mathrm{E}(\xi \cdot \eta) = \mathrm{E}(\xi) \cdot \mathrm{E}(\eta)$ only if ξ and η are independent.
- $\mathrm{E}(\xi - \mathrm{E}(\xi)) = 0$
- $\mathrm{E}(a \cdot \xi + b \cdot \eta) = a \cdot \mathrm{E}(\xi) + b \cdot \mathrm{E}(\eta)$

Variance

Definition: $\quad \mathrm{Var}(\xi) = \mathrm{E}(\xi - \mathrm{E}(\xi))^2$

Population variance: $\quad \mathrm{Var}(\xi) = \dfrac{\sum_{i=1}^N (\xi_i - \mathrm{E}(\xi))^2}{N}$

Sample variance: $\quad \mathrm{Var}(\xi) = \dfrac{\sum_{i=1}^n (\xi_i - \mathrm{E}(\xi))^2}{n-1}$

Some properties:

- $\mathrm{Var}(a) = 0$
- $\mathrm{Var}(\xi + a) = \mathrm{Var}(\xi)$
- $\mathrm{Var}(a \cdot \xi) = a^2 \cdot \mathrm{Var}(\xi)$
- $\mathrm{Var}(\xi \pm \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta) \pm 2 \cdot \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Var}(a \cdot \xi \pm b \cdot \eta) = a^2 \cdot \mathrm{Var}(\xi) + b^2 \cdot \mathrm{Var}(\eta) \pm 2ab \cdot \mathrm{Cov}(\xi, \eta)$

Covariance

Definition: $\quad \mathrm{Cov}(\xi, \eta) = \mathrm{E}[(\xi - E(\xi)) \cdot (\eta - E(\eta))]$

Population covariance:

$$\frac{\sum_{i=1}^N (\xi_i - \mathrm{E}(\xi)) \cdot (\eta_i - \mathrm{E}(\eta))}{N}$$

Sample covariance:

$$\frac{\sum_{i=1}^n (\xi_i - \mathrm{E}(\xi)) \cdot (\eta_i - \mathrm{E}(\eta))}{n-1}$$

Some properties:

- $\mathrm{Cov}(\xi, a) = 0$
- $\mathrm{Cov}(\xi + a, \eta + b) = \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Cov}(a \cdot \xi, b \cdot \eta) = ab \cdot \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Cov}(\xi, \xi) = \mathrm{Var}(\xi)$
- $\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(\eta, \xi)$

VAR (Vector Autoregressive)

A VAR model captures **dynamic interactions** between time series variables. The VAR(p):

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

where:

- $y_t = (y_{1t}, \ldots, y_{Kt})^\intercal$ is a vector of K observable endogenous time series variables.
- A_i's are $K \times K$ coefficient matrices.
- $x_t = (x_{1t}, \ldots, x_{Mt})^\intercal$ is a vector of M observable exogenous time series variables.
- B_j's are $K \times M$ coefficient matrices.
- D_t is a vector that contains all deterministic terms, that may be a: constant, linear trend, seasonal dummy, and/or any other user specified dummy variables.
- C is a coefficient matrix of suitable dimension.
- $u_t = (u_{1t}, \ldots, u_{Kt})^\intercal$ is a vector of K white noise series.

The process is **stable** if:

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1$$

this is, there are **no roots** in and on the complex unit circle.

For example, a VAR model with two endogenous variables ($K = 2$), two lags ($p = 2$), an exogenous contemporaneous variable ($M = 1$), a constant (const) and a trend (Trend_t):

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

Visualizing the separate equations:

$$y_{1t} = a_{11,1} y_{1,t-1} + a_{12,1} y_{2,t-1} + a_{11,2} y_{1,t-2} + a_{12,2} y_{2,t-2} + b_{11} x_t + c_{11} + c_{12} \text{Trend}_t + u_{1t}$$
$$y_{2t} = a_{21,1} y_{2,t-1} + a_{22,1} y_{1,t-1} + a_{21,2} y_{2,t-2} + a_{22,2} y_{1,t-2} + b_{21} x_t + c_{21} + c_{22} \text{Trend}_t + u_{2t}$$

If there is an unit root, the determinant is zero for $z = 1$, then some or all variables are integrated and a VAR model is no longer appropiate (is unstable).

VECM (Vector Error Correction Model)

If **cointegrating relations** are present in a system of variables, the VAR form is not the most convenient. It is better to use a VECM, that is, the levels VAR substracting y_{t-1} from both sides. The VECM($p-1$):

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

where:

- y_t, x_t, D_t and u_t are as specified in VAR.
- $\Pi = -(I_K - A_1 - \cdots - A_p)$ for $i = 1, \ldots, p-1$; Πy_{t-1} is referred as the **long-term** part.
- $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ for $i = 1, \ldots, p-1$ is referred as the **short-term** parameters.
- A_i, B_j and C are coefficient matrices of suitable dimensions.

If the VAR(p) process is unstable (there are roots), Π can be written as a product of ($K \times r$) matrices α (**loading matrix**) and β (**cointegration matrix**) with $\text{rk}(\Pi) = \text{rk}(\alpha) = \text{rk}(\beta) = r$ (**cointegrating rank**) as follows $\Pi = \alpha \beta^\intercal$.

- $\beta^\intercal y_{t-1}$ contains the cointegrating relations.

For example, if there are three endogenous variables ($K = 3$) with two cointegratig relations ($r = 2$), the long term part of the VECM:

$$\Pi y_{t-1} = \alpha \beta^\intercal y_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} = \begin{bmatrix} \alpha_{11} ec_{1,t-1} + \alpha_{12} ec_{2,t-1} \\ \alpha_{21} ec_{1,t-1} + \alpha_{22} ec_{2,t-1} \\ \alpha_{31} ec_{1,t-1} + \alpha_{32} ec_{2,t-1} \end{bmatrix}$$

where:

$$ec_{1,t-1} = \beta_{11} y_{1,t-1} + \beta_{21} y_{2,t-1} + \beta_{31} y_{3,t-1}$$
$$ec_{2,t-1} = \beta_{12} y_{1,t-1} + \beta_{22} y_{2,t-1} + \beta_{32} y_{3,t-1}$$

Note: this is a very basic introduction, there is a lot more literature about the correct specific use of this models and more advanced ones. For example, the VECM with deterministic terms inside the cointegrating relations, the Structural VAR model, etc.