UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

LECTURE NOTES IN COMPUTATIONAL ECONOMICS AND FINANCIAL ECONOMETRICS

University Course Materials

Giulio Palomba

July 2026, this version: 3rd August 2026

Premise

These pages cover several topics from the Time Series Econometrics course syllabus that were not included in Jack Lucchetti's *Notes on Time Series Analysis*. Needless to say, the following pages assume that the reader has already completed an introductory econometrics course.

Please note that the topics discussed in these pages do not follow a specific logical thread or connection, but are simply presented in no particular order. Nevertheless, these topics are fully part of the course syllabus and may therefore be included as exam questions.

Contents

1 OLS as an ML Estimator **2**
- 1.1 The ML Estimator in the OLS Model . . . 2
- 1.2 The Restricted OLS Model . . . 4

2 Classical Likelihood-Based Tests in the OLS Model **5**
- 2.1 LR Test . . . 6
- 2.2 LM Test . . . 6
- 2.3 W Test . . . 7
- 2.4 Relationship between the LR, LM, and W Tests . . . 7
- 2.5 The Relationship between the LR, LM, and W Tests and the F-Test . . . 8

3 Information Criteria **8**

4 The AR(2) Process **9**
- 4.1 Unconditional Variance . . . 9
- 4.2 Stationarity . . . 10

5 Diagnostic Tests in Time-Series Models **11**
- 5.1 Autocorrelation Tests . . . 12
  - 5.1.1 Durbin-Watson Statistic . . . 12
  - 5.1.2 Breutsch-Godfrey Test . . . 14
  - 5.1.3 Ljung-Box Test . . . 15
- 5.2 Conditional Heteroscedasticity Test: the ARCH Test . . . 16
- 5.3 Normality Tests . . . 16
  - 5.3.1 Jarque-Bera Test . . . 16
  - 5.3.2 Doornik-Hansen Test . . . 17

6 Unit Root Tests **17**
- 6.1 Distribuzione asintotica del test DF . . . 18
- 6.2 Phillips-Perron Test . . . 19
- 6.3 KPSS Test . . . 20

7 Forecasting **21**
- 7.1 Static forecast . . . 21
- 7.2 Dynamic forecast . . . 22
- 7.3 Measures of Forecast Evaluation . . . 22
  - 7.3.1 Root Mean Squared Error (RMSE) . . . 22
  - 7.3.2 Diebold-Mariano test . . . 23

8 Dummy Variables **25**

1 OLS as an ML Estimator

It is well known that one of the defining features of the Ordinary Least Squares (OLS) model is that it does not require specifying a distribution for the error term ($\boldsymbol{\varepsilon}$) within the equation:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y}$ is the $T \times 1$ dependent variable, $\boldsymbol{X}$ is the $T \times k$ matrix of regressors, and $\boldsymbol{\beta}$, a $k \times 1$ vector, is the unknown parameter vector. In this context, simply by imposing the "classical" assumptions (uncorrelatedness between regressors and the error term, linearity of the model, and full column rank of X), one obtains the estimator:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{2}$$

which, in finite samples, is unbiased and is the *Best Linear Unbiased Estimator* (BLUE). By imposing the further classical assumption of homoskedasticity, $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim i.i.d.(\boldsymbol{0}, \sigma^2\boldsymbol{I}_T)$, where $\boldsymbol{I}_T$ is a $T \times T$ identity matrix, it is straightforward to show that, as $T \to +\infty$, the estimator enjoys the fundamental properties of consistency and an asymptotic normal distribution.

1.1 The ML Estimator in the OLS Model

Estimating the values of the parameter vector $\boldsymbol{\beta}$ and the variance σ^2 using the estimator (2) is not the only applicable estimation method within the OLS framework. A viable alternative is represented by the maximum likelihood (ML) estimator which, by definition, also yields consistent and asymptotically normal estimates. However, the finite-sample properties of unbiasedness and being the BLUE no longer hold; yet, this is of minor relevance considering that statistical inference is typically conducted using asymptotic theory results. While all the classical assumptions remain valid, the use of the ML estimator necessarily requires assigning a specific distribution to the vector $\boldsymbol{\varepsilon}$. In this section, we will discuss the standard case where the error term of the OLS model follows a multivariate normal distribution

$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_T). \tag{3}$$

Imposing this condition allows the application of the ML method, thereby determining an analytical form for the likelihood function built on $\boldsymbol{\varepsilon}$. This function is given by:

$$\begin{aligned} L(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) &= \prod_{t=1}^{T} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma^2}\right\} \\ &= (2\pi)^{-\frac{T}{2}} (\sigma^2 |\boldsymbol{I}_T|)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2\right\}, \end{aligned} \tag{4}$$

where $|\boldsymbol{I}_T| = 1$ is the determinant of the identity matrix. The log-likelihood is therefore

$$\begin{aligned} \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2 \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2} \end{aligned} \tag{5}$$

or, in a more compact form,

$$\ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}. \tag{6}$$

The score function is defined as the gradient vector containing $k+1$ derivatives: k with respect to the parameters contained in $\boldsymbol{\beta}$, plus the derivative with respect to σ^2. Analytically, it is given by

$$s(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \dfrac{\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{\sigma^2} \\ -\dfrac{T}{2\sigma^2} + \dfrac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \dfrac{\boldsymbol{X}'\boldsymbol{\varepsilon}}{\sigma^2} \\ -\dfrac{T}{2\sigma^2} + \dfrac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^4} \end{bmatrix}. \tag{7}$$

Applying the first-order conditions to the score, we obtain the system:

$$\begin{cases} \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \mathbf{0} \\ \\ -T + \dfrac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^2} = 0 \end{cases} \tag{8}$$

from which the following solutions are obtained:

$$\begin{cases} \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ \\ \hat{\sigma}^2 = \dfrac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{T}, \end{cases} \tag{9}$$

where $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is the vector of residuals obtained from the OLS estimation.

The first equation clearly reveals that, for the parameters of the conditional mean $E(\boldsymbol{y} \mid \boldsymbol{X})$, the solution obtained via the ML estimator under the assumption of normal errors coincides exactly with that of the OLS estimator. From an analytical perspective, this solution stems essentially from the fact that the first equation of the score in (8) effectively coincides with the orthogonality condition imposed by the OLS method when minimizing the objective function $S(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$. In the second equation, the estimator for the unknown parameter σ^2 is given by the sample variance estimator which, as is well known, does not incorporate the degrees-of-freedom correction mechanism featured in the OLS method; this implies that the resulting estimator is biased since $E(\hat{\sigma}^2) \neq \sigma^2$, though it remains consistent, as the bias

$$\begin{aligned} \delta_{\hat{\sigma}^2} &= E(\hat{\sigma}^2 - \hat{s}^2) \\ &= \frac{T-k}{T}\sigma^2 - \sigma^2 \\ &= -\frac{k}{T}\sigma^2 \end{aligned}$$

goes to zero when $T \to +\infty$.

In this framework, the information matrix can be computed in two different ways:

(a) via the Hessian matrix

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}, \sigma^2) = -E\begin{bmatrix} \dfrac{\partial^2 \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \beta^2} & \dfrac{\partial^2 \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial (\sigma^2)'} \\ \\ \dfrac{\partial^2 \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \boldsymbol{\beta}'} & \dfrac{\partial^2 \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} & \mathbf{0} \\ \\ 0 & \dfrac{T}{2\sigma^4} \end{bmatrix}; \tag{10}$$

(b) via the Outer Product Gradient (OPG)

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}, \sigma^2) = E\left[\begin{array}{cc} \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}}' & \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}' \\ \\ \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}}' & \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \dfrac{\partial \ell(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2}' \end{array} \right] = E(\boldsymbol{S}'\boldsymbol{S}), \tag{11}$$

where $\boldsymbol{S} = \begin{bmatrix} s_1(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) & s_2(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) & \dots & s_T(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2) \end{bmatrix}'$ is a $T \times (k+1)$ matrix.

Properties:

- generally, equations (10) and (11) do not coincide. However, they do coincide if the likelihood function is correctly specified;

- equation (10) exhibits better finite-sample properties;
- equation (11) requires evaluating the individual contribution of all observations.

Consequently, the estimated parameter covariance matrix, which coincides with the Cramér-Rao (lower) bound, is given by

$$\mathrm{CR}(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} & \mathbf{0} \\ \\ 0 & \dfrac{2\sigma^4}{T} \end{bmatrix}. \tag{12}$$

From this expression, the following asymptotic distributions for the ML estimators of $\boldsymbol{\beta}$ and σ^2 are derived:

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} N\left(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\right) \qquad \text{e} \qquad \sqrt{T}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathrm{d}} N(0, 2\sigma^4),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \dfrac{\boldsymbol{X}'\boldsymbol{X}}{T} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'$.

Once all the relevant quantities have been obtained via the ML method, it is possible to exploit their properties to conduct a valid and appropriate statistical inference procedure.

1.2 The Restricted OLS Model

Equation (2) defines the estimator obtained by minimising the sum of squared residuals with respect to the values contained within the vector $\boldsymbol{\beta}$. However, it is possible to estimate the same parameter vector of the linear model in the presence of a constraint introduced by imposing a null hypothesis H_0. Consequently, the ordinary least squares estimation problem transforms into a constrained minimisation problem, namely

$$\begin{cases} \min & \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ \text{sub} & H_0 \colon \boldsymbol{g}(\boldsymbol{\beta}) = \mathbf{0}, \end{cases} \tag{13}$$

where $\boldsymbol{g}(\boldsymbol{\beta}) : \mathbb{R}^k \longrightarrow \mathbb{R}^q$ is a continuous and differentiable function of $\boldsymbol{\beta}$ that projects the k-dimensional space of the rows of $\boldsymbol{\beta}$ into a space with $q \leq k$ dimensions; in other words, q is the number of constraints imposed on the k components of $\boldsymbol{\beta}$. For simplicity, and without loss of generality, the entire analysis will be conducted assuming a linear constraint within the null hypothesis. Analytically, this constraint can be expressed by the equation

$$H_0 \colon \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \tag{14}$$

where $\boldsymbol{R}$ is a $q \times k$ matrix and $\boldsymbol{r}$ is a vector of dimension q.

To determine the solution to problem (13), the method of Lagrange Multipliers[1] is applied. The Lagrangian is therefore

$$\Lambda(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}),$$

where $\boldsymbol{\lambda}$ is the q-dimensional vector containing the Lagrange multipliers. In Economics, these are defined as *shadow prices*: each element of the vector $\boldsymbol{\lambda}$ represents the increase in the objective function resulting from a "small" variation in the corresponding constraint. Applying the first-order conditions yields

$$\begin{cases} \dfrac{\partial \Lambda(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \mathbf{0} \\ \\ \dfrac{\partial \Lambda(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \lambda} = \mathbf{0} \end{cases} \quad \Rightarrow \quad \begin{cases} 2\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{R}'\boldsymbol{\lambda} = \mathbf{0} \\ \\ \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r} = \mathbf{0}, \end{cases}$$

[1] See section 2.2 for a very brief discussion.

where $\boldsymbol{R} = \dfrac{\partial(\boldsymbol{R\beta} - \boldsymbol{r})}{\partial \boldsymbol{\beta}}$ is the $q \times k$ Jacobian matrix. After some algebraic manipulation, the solutions are

$$\begin{cases} \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \dfrac{1}{2}(\boldsymbol{X'X})^{-1}\boldsymbol{R'\lambda} \\ \\ (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) - \dfrac{1}{2}(\boldsymbol{X'X})^{-1}\boldsymbol{R'\lambda} \end{cases} \Rightarrow \quad \begin{cases} \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \\ \\ \tilde{\boldsymbol{\lambda}} = 2[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \end{cases} \tag{15}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator, whilst $\tilde{\boldsymbol{\beta}}$ is the estimator relative to the restricted model. Note that the latter is an unbiased estimator only if the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$ is satisfied. Exploiting the property $\boldsymbol{X'}\hat{\boldsymbol{\varepsilon}} = \boldsymbol{0}$ and further defining the residual of the constrained model as

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}} &= \boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} \\ &= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - X\tilde{\boldsymbol{\beta}} \\ &= \hat{\boldsymbol{\varepsilon}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}), \end{aligned}$$

the following relationship regarding the sum of squared residuals is obtained

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} &= [\hat{\boldsymbol{\varepsilon}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})]'[\hat{\boldsymbol{\varepsilon}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})] \\ &= \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\boldsymbol{X'X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}). \end{aligned} \tag{16}$$

By substituting the definition of $\tilde{\boldsymbol{\beta}}$ from (15), it is straightforward to obtain the following expression

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} &= (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\boldsymbol{X'X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\ &= \{\cancel{\hat{\boldsymbol{\beta}}} - (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) - \cancel{\hat{\boldsymbol{\beta}}}\}'\boldsymbol{X'} \cdot \\ &\quad \cdot \boldsymbol{X}\{\cancel{\hat{\boldsymbol{\beta}}} - (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) - \cancel{\hat{\boldsymbol{\beta}}}\}' \\ &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \\ &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \\ &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \end{aligned} \tag{17}$$

$$\tag{18}$$

Equation (18) indicates that the difference between the sum of squared residuals in the restricted model and the sum of squared residuals in the unrestricted model can be expressed as a (positive definite) quadratic form in $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}$; this vector is null only if the estimator $\hat{\boldsymbol{\beta}}$ satisfies the constraint, and is therefore not significantly different from $\tilde{\boldsymbol{\beta}}$.

Furthermore, using the definition from (15), it also follows that

$$\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\lambda}}'\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}\tilde{\boldsymbol{\lambda}} \tag{19}$$

which corresponds to a quadratic form that is zero only if the vector $\tilde{\boldsymbol{\lambda}}$ vanishes.

2 Classical Likelihood-Based Tests in the OLS Model

This section provides the equations for the three classical likelihood-based tests: namely, the Likelihood Ratio (LR), Lagrange Multiplier (LM), and Wald (W) tests. Given their fundamental importance in econometrics, these tests are frequently referred to as the "Classical (or Holy) Trinity".

As is well known, the three tests are asymptotically equivalent and follow a χ^2_q distribution, where q denotes the number of constraints imposed by the null hypothesis H_0. However, in finite samples and under the normality assumption (3), the (alphabetical) inequality LM $\leq$ LR $\leq$ W holds. Within the context of the OLS model, it also remains true that computing the W test requires knowledge only of the unconstrained model $(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\boldsymbol{\varepsilon}})$, the LM test requires only the constrained model $(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2, \tilde{\boldsymbol{\varepsilon}})$, whilst the LR test requires both models.

2.1 LR Test

By simply applying the formal definition, it follows that

$$\begin{aligned}
\text{LR} &= 2[\ell(\boldsymbol{y}, \boldsymbol{X}; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2) - \ell(\boldsymbol{y}, \boldsymbol{X}; \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)] \\
&= 2\left[-\frac{T}{2}\ln\hat{\sigma}^2 - \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{2\hat{\sigma}^2} + \frac{T}{2}\ln\tilde{\sigma}^2 + \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{2\tilde{\sigma}^2}\right] \\
&= -T\ln\hat{\sigma}^2 - \frac{T\hat{\sigma}^2}{2\hat{\sigma}^2} + T\ln\tilde{\sigma}^2 + \frac{T\tilde{\sigma}^2}{2\tilde{\sigma}^2} \\
&= T(\ln\tilde{\sigma}^2 - \ln\hat{\sigma}^2) \\
&= T\ln\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right) = T\ln\left(\frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}\right) \xrightarrow{\text{d}} \chi^2_q. \qquad (20)
\end{aligned}$$

2.2 LM Test

Applying the definition yields

$$\begin{aligned}
\text{LM} &= s(\boldsymbol{y}, \boldsymbol{X}; \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)'\boldsymbol{\mathcal{I}}(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)^{-1}s(\boldsymbol{y}, \boldsymbol{X}; \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2) \\
&= \begin{bmatrix} \dfrac{\tilde{\boldsymbol{\varepsilon}}'\boldsymbol{X}}{\tilde{\sigma}^2} & \dfrac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{2\tilde{\sigma}^4} - \dfrac{T}{2\tilde{\sigma}^2} \end{bmatrix} \begin{bmatrix} \tilde{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1} & 0 \\ 0 & \dfrac{2\tilde{\sigma}^4}{T} \end{bmatrix} \begin{bmatrix} \dfrac{\boldsymbol{X}'\tilde{\boldsymbol{\varepsilon}}}{\tilde{\sigma}^2} \\ \dfrac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{2\tilde{\sigma}^4} - \dfrac{T}{2\tilde{\sigma}^2} \end{bmatrix} \\
&= \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} & 0 \end{bmatrix} \begin{bmatrix} \dfrac{\boldsymbol{X}'\tilde{\boldsymbol{\varepsilon}}}{\tilde{\sigma}^2} \\ \dfrac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{2\tilde{\sigma}^4} - \dfrac{T}{2\tilde{\sigma}^2} \end{bmatrix} \\
&= \frac{\tilde{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{\varepsilon}}}{\tilde{\sigma}^2} \\
&= T\frac{\tilde{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{\varepsilon}}}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}} \\
&= T\frac{\tilde{\boldsymbol{\varepsilon}}'\boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{\varepsilon}}}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}, \qquad (21)
\end{aligned}$$

where $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is the square, symmetric, and idempotent projection matrix for which $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}$ holds. In practice, the test statistic becomes

$$\text{LM} = TR^2 \xrightarrow{\text{d}} \chi^2_q, \qquad (22)$$

where R^2 is the coefficient of determination from an auxiliary regression of $\tilde{\boldsymbol{\varepsilon}}$ on $\boldsymbol{X}$. This result is crucial because several specification and diagnostic tests in econometrics are LM tests wherein the test statistic can be expressed as the product of the sample size and the R^2 index of an auxiliary regression.

An alternative equation for the LM test can be obtained as follows

$$\begin{aligned}
\text{LM} &= T\frac{\tilde{\boldsymbol{\varepsilon}}'\boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{\varepsilon}}}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}} \\
&= T\frac{(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'\boldsymbol{P}_{\boldsymbol{X}}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}} \\
&= T\frac{y'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} - 2\tilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \tilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\tilde{\boldsymbol{\beta}})}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}},
\end{aligned}$$

given that $\boldsymbol{P_X y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$, it follows that

$$\begin{aligned}
\text{LM} &= T\frac{\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - 2\tilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\tilde{\boldsymbol{\beta}})}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}, \\
&= T\frac{(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}, \\
&= T\frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}, \\
&= T\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2}.
\end{aligned} \tag{23}$$

2.3 W Test

Since the test statistic W is defined as

$$\text{W} = \boldsymbol{g}(\hat{\boldsymbol{\beta}})[\boldsymbol{J}(\hat{\boldsymbol{\beta}})Var(\hat{\boldsymbol{\beta}})\boldsymbol{J}(\hat{\boldsymbol{\beta}})']^{-1}\boldsymbol{g}(\hat{\boldsymbol{\beta}}) \xrightarrow{\text{d}} \chi^2_q, \tag{24}$$

imposing the linear constraint $\boldsymbol{g}(\boldsymbol{\beta}) = \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}$ yields

$$\begin{aligned}
\text{W} &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}Var(\hat{\boldsymbol{\beta}})\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \\
&= \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{\hat{\sigma}^2} \\
&= T\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}.
\end{aligned} \tag{25}$$

By using Equation (18), it is straightforward to show that the following also holds

$$\text{W} = T\frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} = T\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}. \tag{26}$$

2.4 Relationship between the LR, LM, and W Tests

From Equations (20), (23), and (26), it follows that

$$\begin{cases}
\text{LR} = T\ln\left(\dfrac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right) = T\ln\left(1 + \dfrac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}\right) = T\ln\left(1 + \dfrac{\text{W}}{T}\right) \\
\text{LM} = T\dfrac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2} = \dfrac{\hat{\sigma}^2}{\tilde{\sigma}^2}\text{W} \\
\text{W} = T\dfrac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2},
\end{cases}$$

therefore

- asymptotically, the three test statistics are equivalent, since it follows that
$$\lim_{T\to+\infty}(\text{LR} - \text{LM}) = \lim_{T\to+\infty}(\text{LR} - \text{W}) = \lim_{T\to+\infty}(\text{LM} - \text{W}) = 0;$$
- in finite samples, the following inequalities hold
  - $\text{W} \geq \text{LR} \quad \Rightarrow \quad \text{W} \geq T\ln\left(1 + \dfrac{\text{W}}{T}\right) \quad \Rightarrow \quad \dfrac{\text{W}}{T} \geq \ln\left(1 + \dfrac{\text{W}}{T}\right) \quad \text{per } \forall\ \text{W} > 0;$
  - $\text{LR} \geq \text{LM} \quad \Rightarrow \quad T\ln\left(1 + \dfrac{\text{W}}{T}\right) \geq \dfrac{\hat{\sigma}^2}{\tilde{\sigma}^2}\text{W} \quad \Rightarrow \quad T\ln\left(1 + \dfrac{\text{W}}{T}\right) \geq \dfrac{1}{\dfrac{\tilde{\sigma}^2}{\hat{\sigma}^2} + 1 - 1}\dfrac{\text{W}}{T}$
  $$\Rightarrow \quad T\ln\left(1 + \frac{\text{W}}{T}\right) \geq \frac{\dfrac{\text{W}}{T}}{1 + \dfrac{\text{W}}{T}} \quad \text{per } \forall\ \text{W} > 0;$$
  - $\text{W} \geq \text{LM}$ by the transitive property.

2.5 The Relationship between the LR, LM, and W Tests and the F-Test

In addition to being interrelated, the three classical tests can be expressed as functions of the F-test, which is typically performed in finite-sample contexts. Starting from the formal definition of the latter, it follows that

$$\begin{aligned} \mathrm{F}_{q,T-k} &= \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}} - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}} \frac{T-k}{q} \\ &= \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \frac{T-k}{q} \\ &= \frac{T-k}{Tq}\mathrm{W}. \end{aligned} \tag{27}$$

By calculating the inverse function, it follows that

$$\begin{cases} \mathrm{LR} = T \ln\left(1 + \dfrac{q}{T-k}\mathrm{F}_{q,T-k}\right) \\ \mathrm{LM} = \dfrac{\hat{\sigma}^2}{\tilde{\sigma}^2}\dfrac{Tq}{T-k}\mathrm{F}_{q,T-k} = \left(\dfrac{1-\mathrm{W}}{T}\right)^{-1}\dfrac{Tq}{T-k}\mathrm{F}_{q,T-k} = \dfrac{Tq\mathrm{F}_{q,T-k}}{T-k+q\mathrm{F}_{q,T-k}} \\ \mathrm{W} = \dfrac{Tq}{T-k}\mathrm{F}_{q,T-k}. \end{cases} \tag{28}$$

From the properties of Snedecor's F-distribution, it follows that the LR, LM, and W tests follow a χ^2_q distribution as $T \to +\infty$.

3 Information Criteria

Information criteria (IC) represent highly useful tools during the specification stage of an econometric model. As is well known, choosing how many (and which) explanatory variables to include constitutes a trade-off: adding (non-collinear) variables does not worsen the explanatory power of the model, but, on the other hand, augments the risk of "overfitting" the equation with one or more variables that do not meaningfully improve the interpretability of the model. In more technical terms, adding variables implies that:

(a) the log-likelihood value does not decrease,

(b) the number of unknown parameters to be estimated increases, thereby making the model less parsimonious.

Information criteria consist of an equation designed to manage this contradiction; indeed, their general formulation is of the form

$$\mathrm{IC} = f(\underset{(-)}{\ell(\hat{\boldsymbol{\theta}})}, \underset{(+)}{k}) \tag{29}$$

where $\ell(\hat{\boldsymbol{\theta}})$ is the log-likelihood evaluated at the estimated value of the parameter vector $\hat{\boldsymbol{\theta}}$, whilst k is the number of parameters (components) contained within this vector. The signs in parentheses indicate the relationship between the values assumed by the information criterion and the variables upon which it is determined. The logic is therefore straightforward: since an increase in the number of parameters implies a non-decreasing log-likelihood, the information criterion is constructed such that it decreases as the likelihood rises, whereas it increases in value as the number of parameters k augments.
Specifically, the three most widely used information criteria in the literature are the following:

1. Akaike (1974)'s Criterion: $\mathrm{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2k$,

2. Schwarz (1978)'s Criterion or Bayesian Information Criterion: $\mathrm{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + k\log T$,

3. Hannan e Quinn (1979)'s Criterion: $\text{HQC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2k \log\log T$.

From a practical perspective, these three information criteria are usually reported by statistical and econometric software at the end of each estimated model. Among the different estimated models, the "best specification" corresponds to the one for which the information criteria achieve their minimum value. Obviously, the three proposed criteria may conflict with one another, selecting different specifications; in this case, in empirical practice, the BIC and HQC criteria tend to be preferred, as the AIC criterion is known to favour models characterised by a much larger number of parameters. The choice between the BIC and HQC criteria, which are very similar in terms of their analytical definition, is instead substantially left to the user, since there are no well-founded reasons to prefer one criterion over the other.

4 The AR(2) Process

This section strictly focus on deriving the variance of the Auto-Regressive of order 2, or AR(2) process, and provides a brief discussion on its stationarity conditions.

4.1 Unconditional Variance

Consider a generic AR(2) process defined by the equation

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,$$

where the constant μ represents the deterministic component and $\varepsilon_t \sim WN(0, \sigma^2)$; the analytical expression for the unconditional variance is obtained through a rather complex procedure, the computational difficulty of which increases with the order of the process.
To calculate the moments of an AR(p) process with $p \geq 2$, it is useful to "centre" the process around its unconditional expected value $c = E(y_t) = \dfrac{\mu}{1 - \phi_1 - \phi_2} = \dfrac{\mu}{\Phi(1)}$, namely

$$\begin{aligned} y_t &= \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ y_t - c &= \phi_1 (y_{t-1} - c) + \phi_2 (y_{t-2} - c) + \varepsilon_t \\ \tilde{y}_t &= \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \varepsilon_t, \end{aligned}$$

where $\phi_1 + \phi_2 \neq 1$ ensures the existence of $E(y_t)$. Clearly, the process $\tilde{y}_t$ is still an AR(2) process, but now $E(\tilde{y}_t) = 0$, meaning it is centred around zero.
The autocovariance of order k of the process y_t is given by

$$\begin{aligned} \gamma_k &= E[(y_t - c)(y_{t-k} - c)] \\ &= E[\tilde{y}_t \tilde{y}_{t-k}] \\ &= E[(\phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \varepsilon_t)\tilde{y}_{t-k}] \\ &= \phi_1 E(\tilde{y}_{t-1}\tilde{y}_{t-k}) + \phi_2 E(\tilde{y}_{t-2}\tilde{y}_{t-k}) + \cancel{E(\tilde{y}_{t-k}\varepsilon_t)}, \end{aligned}$$

where the condition $E(\tilde{y}_{t-k}\varepsilon_t) = 0$ relies on the property that the innovation ε_t cannot be correlated with lags of the dependent variable. The resulting expression is therefore

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \tag{30}$$

Alternatively, by dividing by $\gamma_0 = Var(y_t)$, it analogously follows that

$$\frac{\gamma_k}{\gamma_0} = \phi_1 \frac{\gamma_{k-1}}{\gamma_0} + \phi_2 \frac{\gamma_{k-2}}{\gamma_0} \qquad \Rightarrow \qquad \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}. \tag{31}$$

These two recursive autocovariances and autocorrelations allow the variance of the AR(2) process to be calculated; indeed, it follows that

$$\begin{aligned}\gamma_0 = Var(\tilde{y}_t) &= E[(\phi_1\tilde{y}_{t-1} + \phi_2\tilde{y}_{t-2} + \varepsilon_t)\tilde{y}_t] \\ &= \phi_1 E(\tilde{y}_t\tilde{y}_{t-1}) + \phi_2 E(\tilde{y}_t\tilde{y}_{t-2}) + E(\tilde{y}_t\varepsilon_t) \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2 \\ &= \phi_1\rho_1\gamma_0 + \phi_2\rho_2\gamma_0 + \sigma^2.\end{aligned}$$

Since $\rho_0 = 1$ by definition, it follows that

$$\begin{cases}\rho_1 = \phi_1 + \phi_2\rho_1 \qquad \Rightarrow \qquad \rho_1 = \dfrac{\phi_1}{1-\phi_2} \\ \\ \rho_2 = \phi_1\rho_1 + \phi_2.\end{cases}$$

By substituting, the equation for the unconditional variance is

$$\begin{aligned}\gamma_0 &= \frac{\phi_1^2}{1-\phi_2}\gamma_0 + \phi_1\phi_2\rho_1\gamma_0 + \phi_2^2\gamma_0 + \sigma^2 \\ &= \left[\frac{\phi_1^2}{1-\phi_2} + \frac{\phi_1^2\phi_2}{1-\phi_2} + \phi_2^2\right]\gamma_0 + \sigma^2 \\ &= \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}.\end{aligned} \tag{32}$$

4.2 Stationarity

The admissible parameter values for the process to be stationary are those for which the variance of the process from Equation (32) is finite, positive, and time-independent; it therefore follows that:

- $\mu \in \mathbb{R}$, since the model constant does not enter the equation for γ_0;
- the numerator of the expression is strictly positive for $\phi_2 < 1$;
- the denominator is strictly positive when

$$\begin{cases}1 + \phi_2 > 0 \\ (1-\phi_2)^2 - \phi_1^2 > 0\end{cases} \qquad \Rightarrow \qquad \begin{cases}\phi_2 > -1 \\ \phi_1^2 < (1-\phi_2)^2\end{cases} \quad \Rightarrow \quad \phi_1 + \phi_2 < 1 \ \wedge \ \phi_2 - \phi_1 < 1.$$

Combining these results yields the following scenario:

$$-1 + \phi_2 < \phi_1 < 1 - \phi_2 \qquad \wedge \qquad -1 < \phi_2 < 1 \tag{33}$$

which is illustrated graphically in Figure 1.

By observing the graph, five distinct subspaces can be identified in the plane:

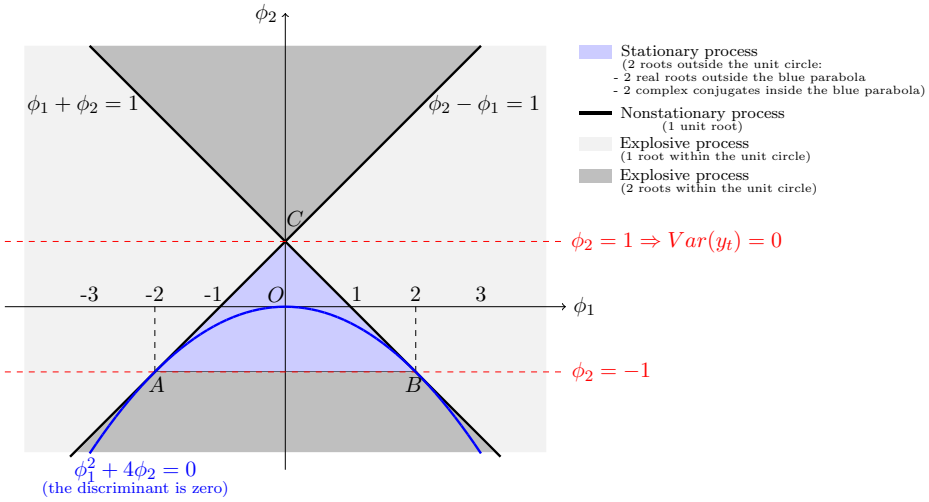
(a) the red lines defined by the equation $\phi_2 = \pm 1$, which identify the parameter combinations for which γ_0 is either zero or undefined (corresponding to a null numerator and a null denominator in Equation (32), respectively);

(b) the light-grey zones where one of the two conditions $\phi_1 + \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$ is violated;

(c) the dark-grey zones where both conditions $\phi_1 + \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$ are violated;

(d) the light-blue triangle ABC in which the stationarity conditions from (33) are simultaneously satisfied;

Figure 1: Stationarity of the AR(2) process



(e) the blue parabola plotted according to the roots of the discriminant of the second-order characteristic equation $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2$, for which the two solutions are real, coincident, and clearly outside the unit circle. It goes without saying that the half-plane "above" this parabola identifies the combinations of parameters ϕ_1 and ϕ_2 for which the discriminant is positive, whereas the half-plane "below" the parabola contains the combinations that yield a negative discriminant, thereby resulting in complex roots of the polynomial.

In summary:

light-blue area: the polynomial $\Phi(L)$ admits two roots outside the unit circle; hence, $y_t \sim I(0)$ and its time path is typical of a *mean-reverting* time series[2]. Furthermore:

- if $\phi_1 + 4\phi_2 \geq 0$, the AR(2) process exhibits the classical "jagged" (highly oscillatory) behaviour,
- if $\phi_1 + 4\phi_2 < 0$, the AR(2) process exhibits a smooth and cyclical behaviour;

grey areas: the polynomial $\Phi(L)$ admits at least one root inside the unit circle; hence, y_t is "explosive";

segments $\overline{AC}$ and $\overline{BC}$: the polynomial $\Phi(L)$ admits at least one unit root; hence, $y_t \sim I(1)$.

5 Diagnostic Tests in Time-Series Models

The diagnostic phase, which must immediately follow the estimation of any econometric model, plays a fundamental role in time series analysis. The objective of this investigation is essentially to perform a statistical check on the estimation, focusing on the model residuals.
Specifically, the residuals of a time series econometric model must be "clean"; that is, they must not be affected by the following problems, which are presented in descending order of severity[3]:

(a) **autocorrelation:** residuals at time t must not be correlated with residuals at time $t-k$ (where $k > 0$); that is, the past must not be used in any way to predict the present. In heuristic terms, the residuals must form a time series composed of unrelated numbers (*i.e.* memoryless). If

[2]Recall that heuristically a *mean-reverting* process is graphically characterised by being centred around its unconditional expected value, with observations generally contained within a "horizontal tube", and frequently crossing the horizontal axis $y_t = E(y_t)$.

[3]Curiously, the descending order of severity of these problems coincides with alphabetical order.

the residuals exhibit some degree of autocorrelation, the estimated model suffers from *misspecification*; that is, it has failed to capture all the dynamic relationships contained within the dependent variable. This typically occurs when relevant explanatory variables are omitted, or when the available explanatory variables themselves have limited explanatory power regarding the economic problem under analysis. In this case, any dynamic structure that is not captured by the model automatically spills over into the residuals, thereby "contaminating" them;

(b) **conditional heteroskedasticity:** in general, the variance of the residuals conditional on the information set is assumed to be constant; *i.e.*, it must hold that $Var(\varepsilon_t|\mathbb{I}_{t-1}) = E(\varepsilon_t^2|\mathbb{I}_{t-1}) = \sigma^2$. This condition represents the homoskedasticity assumption on the error terms that characterises time series models (principally ADL and ARMA models). Since the information set $\mathbb{I}_{t-1}$ could be used to predict not only the movements of the dependent variable but also those of its variance, the adjective "conditional" is appended to the terms homoskedasticity/heteroskedasticity. A violation of the conditional homoskedasticity assumption does not, in fact, invalidate the model adopted to explain the dependent variable; rather, it simply signals that a law of motion must also be specified for the conditional variance[4]. Therefore, assuming a constant conditional variance when it actually takes different values over time compromises estimation efficiency, as the estimated standard errors would fail to account for an evident loss of information;

(c) **normality:** actually, the empirical distribution of the residuals does not constitute a particularly significant problem, except in very specific cases where their normal distribution would be highly desirable.

5.1 Autocorrelation Tests

5.1.1 Durbin-Watson Statistic

The first historical attempt to conduct an autocorrelation test for a linear time series model of the type

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \qquad \text{con } \varepsilon_t \sim WN(0, \sigma^2) \tag{34}$$

is well-established in the literature and is known as the Durbin e Watson (1950) statistic. This approach is not actually a proper test, but rather represents a statistic whose values should indicate whether the residuals of the linear model exhibit any significant first-order autocorrelation. In addition to being a "classic" in the diagnostic phase of a linear time series model, the Durbin-Watson (DW) statistic undoubtedly has the merit of promptly providing an indication of the presence or absence of autocorrelation in the residuals. For this reason, almost all statistical and econometric software packages (including `Gretl`) automatically return its value within the *regression statistics* that follow the estimation of model (34). The DW statistic is characterised by the property of having a known finite sample distribution only under the following rather restrictive assumptions (Verbeek, 2004):

1. it must be possible to treat the $\boldsymbol{x}_t$ as deterministic. This assumption is fundamental as it requires all error terms ε_t to be independent of all explanatory variables, according to the well-known Gauss-Markov assumption
$$\boldsymbol{x}_t' \perp \varepsilon_t \approx E(\boldsymbol{x}_t'\varepsilon_t) = 0. \tag{35}$$
An even more relevant aspect is that this condition effectively precludes the use of lagged dependent variables within the regressors;

2. the regressors within the vector $\boldsymbol{x}_t$ must necessarily include an intercept.

[4]To simplify the discussion, one could state that, in the presence of conditional heteroskedasticity, the model for the conditional mean of the dependent variable is not misspecified, but the model for the conditional variance is. The literature on *Generalized Auto Regressive Conditional Heteroskedasticity* (GARCH) models deals with the specification of models for the conditional variance. For a specific treatment, see, for example, Bollerslev, Engle e Nelson (1994) or Palm (1996).

The structure of the hypotheses for assessing the presence of autocorrelation using the DW statistic is as follows:

$$\begin{cases} H_0 \colon \rho_1 = 0 & \rightarrow \text{ no first-order autocorrelation,} \\ H_1 \colon \rho_1 \neq 0 & \rightarrow \text{ first-order autocorrelation.} \end{cases} \tag{36}$$

Formally, the DW statistic is given by

$$\text{DW} = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}, \tag{37}$$

where $\hat{\varepsilon}_t$ is the OLS residual. In practice, this statistic compares the sample mean of the squared differences between the residual time series and its first-lagged counterpart (hence the summation starts from $t = 2$) with the sample variance or sample second moment of the residuals themselves. With some simple algebra and for "large" sample sizes T, we obtain

$$\text{DW} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t^2 - 2\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} + \sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \approx \frac{2\sum_{t=2}^{T} \hat{\varepsilon}_t^2 - 2\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \approx 2 \left(1 - \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \right),$$

therefore

$$\text{DW} = 2(1 - \hat{\rho}_1), \tag{38}$$

where $\hat{\rho}_1$ is the estimated first-order autocorrelation coefficient. Analysing the "extreme" cases, it is quite apparent that:

- under H_0, $\hat{\rho}_1 = 0$, hence DW ≈ 2;
- in the case of perfect positive correlation, $\hat{\rho}_1 = 1$, hence DW ≈ 0;
- in the case of perfect negative correlation, $\hat{\rho}_1 = -1$, hence DW ≈ 4.

In practice, a DW value close to 2 indicates the absence of first-order autocorrelation.
Unfortunately, the DW statistic also entails several issues[5]:

1. DW is unable to detect autocorrelation of orders higher than the first; this limitation is clearly indicated by the structure of the hypotheses in (36) and represents the main reason why this statistic cannot be considered a proper autocorrelation test. To overcome this drawback, it is necessary to perform more general autocorrelation test procedures, such as the Ljung e Box (1978) test or, in certain cases[6], the Breusch-Godfrey (1979-1978) test;

2. as previously mentioned, the DW statistic cannot be applied when the lagged dependent variable appears among the regressors. In the case of ARMA models, it underestimates the autocorrelation and is therefore configured as a biased estimator for ρ_1. In the presence of large samples, a correction mechanism is provided by Durbin's h-statistic

$$h = (1 - 0.5\text{DW}) \sqrt{\frac{T}{1 - T \cdot Var(\hat{\phi}_1)}}, \tag{39}$$

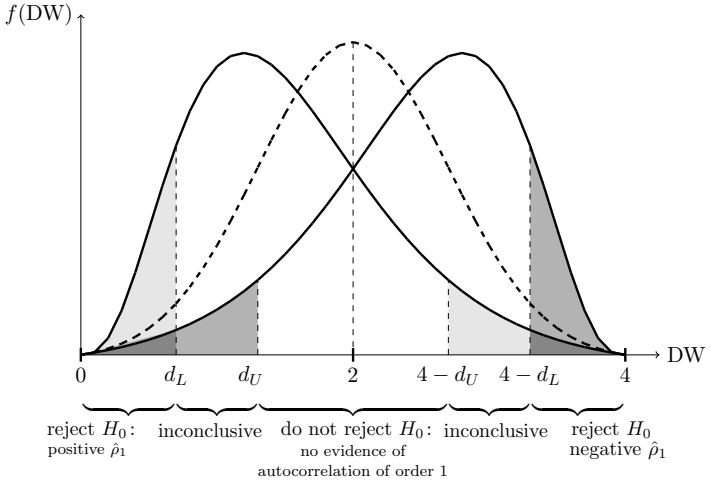
where $T \cdot Var(\hat{\phi}_1) < 1$, while $\hat{\phi}_1$ is the estimated coefficient associated with the lagged dependent variable in the linear regression model. As $T \to +\infty$, the h-statistic asymptotically converges to a standard normal distribution;

[5] See, for example, Hill, Griffiths e Lim (2013).

[6] For instance, the Breusch-Godfrey test cannot be applied in the presence of a moving average component in ARMA models.

3. the DW statistic suffers from the problem of "inconclusive zones" (or regions of indeterminacy); that is, there exist values for which it is impossible to determine with certainty whether the null hypothesis should be accepted or rejected. This occurs because the non-rejection region of H_0 and the rejection regions are separated by an interval of critical values rather than a single critical value.

Figure 2: Distribution of the DW statistic and the two distributions containing the d_L and d_U quantiles



Intuitively, this problem is illustrated in Figure 2: under the null hypothesis $H_0\colon \rho_1 = 0$, the distribution of the DW statistic, centred at 2, depends on the sample size (T), the number of regressors (k) included within $\boldsymbol{x}_t'$, and also on the actual observed values of the regressors themselves. This prevents the exact calculation of the critical values for this distribution. However, knowing T and k is sufficient to determine two distributions from which to extract the d_L and d_U quantiles, which represent the lower and upper bounds for the distribution of the DW test statistic, respectively. The d_L and d_U quantiles were tabulated by Durbin e Watson (1950) and Savin e White (1977), and are available in `Gretl` under the *Tools/Statistical tables* menu.

Since the distribution of the DW statistic is symmetric, by observing Figure 2 (where d_L and d_U are the quantiles to the left of which the probability under the distribution curves is 5%), the following situations emerge:

- if $0 \leq \text{DW} \leq d_L \Rightarrow \rho_1 > 0$, hence H_0 is rejected;
- if $d_L < \text{DW} < d_U \Rightarrow$ the test does not provide any clear indication (the test is inconclusive);
- if $d_U \leq \text{DW} \leq 4 - d_U \Rightarrow \rho_1 = 0$, hence H_0 is not rejected;
- if $4 - d_U < \text{DW} < 4 - d_L \Rightarrow$ the test does not provide any clear indication (the test is inconclusive);
- if $4 - d_L \leq \text{DW} \leq 4 \Rightarrow \rho_1 < 0$, hence H_0 is rejected.

The inconclusive zones are therefore given by the intervals (d_L, d_U) and $(4 - d_U, 4 - d_L)$, which decrease in width as the value of T increases or the value of k decreases. In general, due to the symmetry that characterises the distribution of the DW test statistic, the $4 - d_U$ and $4 - d_L$ quantiles are those to the right of which the area under the distribution curves corresponds to a probability of 5%;

4. the power of the DW test is typically low when the alternative hypothesis specification is other than an AR(1) process $\varepsilon_t = \phi_1 \varepsilon_{t-1} + u_t$ (consider, for instance, $H_1\colon \varepsilon_t \sim \text{MA}(1)$ or $H_1\colon \varepsilon_t \sim \text{RW}$).

5.1.2 Breutsch-Godfrey Test

This test, developed independently by Breusch (1979) and Godfrey (1978), henceforth the BG test, is useful for determining whether some serial dependence exists within the changes of the dependent

variable in a dynamic linear model. In contrast to the test by Durbin e Watson (1950), this test is general, as it is capable of testing different orders of serial autocorrelation and can also be used when lags of the dependent variable are included as regressors.
From a technical point of view, the starting point is the dynamic linear model given in equation (34), from which the residuals are obtained

$$\hat{\varepsilon}_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}. \tag{40}$$

The underlying logic of the test is as follows: if there is an autocorrelation that is not "captured" by the model, then the residuals should follow an AR process of some order $q > 0$; therefore, consider the auxiliary regression

$$\hat{\varepsilon}_t = \boldsymbol{x}_t'\boldsymbol{\delta} + \rho_1\hat{\varepsilon}_{t-1} + \rho_2\hat{\varepsilon}_{t-2} + \ldots + \rho_m\hat{\varepsilon}_{t-m} + \eta_t \tag{41}$$

in which the dependent variable is the time series of the residuals $\hat{\varepsilon}_t$, while the list of regressors is identical to that of the baseline model, with the addition of all residual lags up to the maximum order m.
The structure of the hypotheses for assessing the presence of autocorrelation using the BG test is therefore as follows:

$$\begin{cases} H_0\colon \rho_1 = \rho_2 = \ldots = \rho_q = 0 & \rightarrow \text{ no autocorrelation up to order } m, \\ H_1\colon \exists\, \rho_i \neq 0, \text{ with } i = 1, 2, \ldots, m & \rightarrow \text{ at least one autocorrelation coefficient is non-zero,} \end{cases} \tag{42}$$

where ρ_i represents the autocorrelation of order i between the observations. The BG test is obtained as a LM test computed via the auxiliary regression (41): in particular, the test statistic is obtained through a convenient asymptotic approximation given by

$$\text{BG} = TR^2 \xrightarrow{\text{d}} \chi^2_m, \tag{43}$$

where the R^2 index refers to the auxiliary regression and T is the sample size[7] relative to the estimation of equation (34). Since it is conceived within the context of the OLS model, the BG test has the limitation of being applicable only to dynamic linear models. For instance, this test is never applicable in the presence of moving average terms: in this case, perfect collinearity would occur in the auxiliary regression since the lags of ε_t are included within the vector $\boldsymbol{x}_t'$.

5.1.3 Ljung-Box Test

The test by Ljung e Box (1978), hereafter LB test, represents a test for determining whether the observations of a given time series exhibit an autocorrelation of an order lower than, or at most equal to, a predetermined order m. Analogously to other autocorrelation tests, the null hypothesis specifies the absence of autocorrelation; this implies that:

$$\begin{cases} H_0\colon \rho_1 = \rho_2 = \ldots = \rho_m = 0 & \rightarrow \text{ no autocorrelation up to order } m, \\ H_1\colon \exists\, \rho_i \neq 0, \text{ con } i = 1, 2, \ldots, m & \rightarrow \text{ at least one autocorrelation coefficient is non-zero,} \end{cases} \tag{44}$$

where ρ_i is the autocorrelation of order i. The test statistic is

$$\text{LB} = T(T+2)\sum_{i=1}^{m} \frac{\hat{\rho}_i^2}{T-i} \xrightarrow{\text{d}} \chi^2_m, \tag{45}$$

where $\hat{\rho}_i$ represents the i-th estimated autocorrelation.
Compared to the BG test, the LB test takes on a more general connotation, as it is also applicable to any ARMA-type model. In particular, when the LB test is performed as a diagnostic test on the residuals of an estimated ARMA(p,q) model, an adjustment for the d.o.f. must be made; in this context, there is a physiological loss of d.o.f. because the residuals were obtained using $p + q$ lags, so the limiting distribution to be used in the test procedure is that of the $\chi^2_{m-(p+q)}$ random variable. It follows that the minimum order for the LB test must be equal to $p + q + 1$.

[7] See equation (22). In many textbooks, equation (43) is presented in the version $\text{LM}_{\text{BG}} = (T-m)R^2 \xrightarrow{\text{d}} \chi^2_q$, since the auxiliary regression is run on a sample of $(T-m)$ observations. In fact, for the test statistic to be asymptotically distributed as a chi-squared random variable, the first q missing observations must be replaced with zeros.

5.2 Conditional Heteroscedasticity Test: the ARCH Test

The ARCH (*AutoRegressive Conditional Heteroscedasticity*) test, developed by Engle (1982), is the most widely used conditional heteroscedasticity test within the diagnostic phase of time series models. From a strictly formal viewpoint, it resembles the Breusch-Godfrey autocorrelation test; indeed,

1. there is a linear regression model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}+\varepsilon_t$ in which $\boldsymbol{x}_t$ contains exogenous or lagged variables;
2. the time series consist of T observations;
3. an auxiliary regression, analogous to (41), is set up on the squared residuals of the type
$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1\hat{\varepsilon}_{t-1}^2 + \alpha_2\hat{\varepsilon}_{t-2}^2 + \ldots + \alpha_m\hat{\varepsilon}_{t-m}^2 + u_t; \tag{46}$$
4. the structure of the hypotheses is as follows:
$$\begin{cases} H_0\colon \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0 \\ H_1\colon \exists\, \alpha_i \neq 0, \text{ with } i = 1, 2, \ldots, m, \end{cases} \tag{47}$$
where α_i represents the autocorrelation of order i between the squared residuals at time t and the squared residuals at time $t-i$.

   Observing these hypotheses, it can be noticed that under H_0, there is an absence of autocorrelation up to order m between the squared residuals of the estimated model, which consequently turn out to be constant over time. In other words, the hypotheses of the ARCH test in (47) can be rewritten as follows
$$\begin{cases} H_0\colon \varepsilon_t|\mathbb{I}_{t-1} \sim i.i.d.(0,\sigma^2) & \rightarrow \text{ the variance conditional to the information set is constant over time (conditional homoscedasticity),} \\ H_1\colon \varepsilon_t|\mathbb{I}_{t-1} \sim i.i.d.(0,\sigma_t^2) & \rightarrow \text{ the variance conditional to the information set varies over time (conditional heteroscedasticity),} \end{cases} \tag{48}$$
   where $\sigma_t^2 = Var(\varepsilon_t|\mathbb{I}_{t-1}) = E(\varepsilon_t^2|\mathbb{I}_{t-1})$;
5. the ARCH test is also configured as a LM test computed via the auxiliary regression (46); hence, the test statistic is obtained through the usual asymptotic approximation given in (22), which is expressed as
$$\text{ARCH} = TR^2 \xrightarrow{\text{d}} \chi_q^2, \tag{49}$$
   where the R^2 index refers to the auxiliary regression and T is the sample size.

5.3 Normality Tests

5.3.1 Jarque-Bera Test

The test by Jarque e Bera (1980), hereafter the JB test, is a test for determining whether the empirical distribution of a time series can be approximated by a normal distribution or not. In short, the structure of the hypotheses is
$$\begin{cases} H_0\colon \text{ the series } y_t \text{ has a normal marginal distribution,} \\ H_1\colon \text{ the series } y_t \text{ has a non-normal marginal distribution.} \end{cases} \tag{50}$$
The JB statistic is given by the following expression
$$\text{JB} = T\left[\frac{\gamma_3^2}{3!} + \frac{(\gamma_4-3)^2}{4!}\right] \xrightarrow{\text{d}} \chi_2^2, \tag{51}$$
where γ_3 and γ_4 represent the sample skewness and kurtosis indices, respectively. The asymptotic distribution is that of a χ_2^2 random variable, since the null hypothesis of normality requires the joint restriction to zero of both the skewness index and the excess kurtosis $(\gamma_4 - 3)$.

5.3.2 Doornik-Hansen Test

All the diagnostic tests proposed in this section can be conducted in a univariate framework, that is, on one single time series; the test by Doornik e Hansen (2008), however, is capable of testing the null hypothesis of multinormality, and can therefore be applied in a multivariate context, *i.e.* for multiple time series simultaneously. Obviously, this test finds application in the diagnostic phase following the estimation of a VAR(q) model.
Given a multivariate stochastic process $\boldsymbol{y}_t$ containing T observations for n time series, the structure of the hypotheses for the Doornik-Hansen test mirrors the Jarque-Bera test; indeed, we have

$$\begin{cases} H_0\colon \text{ the series } \boldsymbol{y}_t \text{ have a multivariate normal marginal distribution,} \\ H_1\colon \text{ the series } \boldsymbol{y}_t \text{ have a non-multivariate normal marginal distribution.} \end{cases} \tag{52}$$

Doornik e Hansen (2008) show that, once the time series in $\boldsymbol{y}_t$ are transformed into the standardised and independent variables $\boldsymbol{z}_t$[8], generalising the Jarque-Bera test statistic to a multivariate framework is fairly straightforward; indeed, it yields the equation

$$\text{JB} = T\left[\frac{\boldsymbol{G}_3'\boldsymbol{G}_3}{3!} + \frac{(\boldsymbol{G}_4 - 3\boldsymbol{\iota})'(\boldsymbol{G}_4 - 3\boldsymbol{\iota})}{4!}\right] \xrightarrow{\text{d}} \chi^2_{2n}, \tag{53}$$

where $\boldsymbol{G}_3 = \begin{bmatrix}\gamma_{3,1} & \gamma_{3,2} & \dots \gamma_{3,n}\end{bmatrix}'$ and $\boldsymbol{G}_4 = \begin{bmatrix}\gamma_{4,1} & \gamma_{4,2} & \dots \gamma_{4,n}\end{bmatrix}'$ contain the n skewness and kurtosis indices of the individual standardised time series $\boldsymbol{z}_t$, respectively, whereas $\boldsymbol{\iota}$ is the "sum vector", that is, the vector of dimension n in which every element has a unit value. This test statistic therefore imposes $2n$ restrictions (n restrictions on the skewness indices and n restrictions on the excess kurtosis); hence, for large samples, the χ^2_{2n} asymptotic distribution holds. For the distribution to remain the same even in the presence of finite samples, certain corrections are imposed on equation (53), yielding the following sum of quadratic forms

$$\text{DH} = \boldsymbol{B}_1'\boldsymbol{B}_1 + \boldsymbol{B}_2'\boldsymbol{B}_2 \xrightarrow{\text{d}} \chi^2_{2n}, \tag{54}$$

where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are vectors of dimension $n \times 1$. Their precise definition is contained within the Appendix of the paper by Doornik e Hansen (2008).

6 Unit Root Tests

This section is dedicated to unit root tests. Specifically, it contains an analytical proof regarding the asymptotic distribution of the OLS estimator of an AR(1) process with a unit root, and briefly illustrates the main characteristics of two alternative tests to the Dickey e Fuller (1979) or ADF test.

[8] The transformation of n variables $\boldsymbol{y}_t$ into the corresponding independent standardised variables $\boldsymbol{z_t}$ requires only some simple algebra; indeed, it is sufficient to set up the equation

$$\boldsymbol{Z}_t = \boldsymbol{Z}_t^* \boldsymbol{H},$$

where $\boldsymbol{y}_t'$ and $\boldsymbol{z}_t'$ are the t-th row of the matrices $\boldsymbol{y}_t$ and $\boldsymbol{Z}_t$, both of dimension $T \times n$, and

- $\boldsymbol{Z}_t^* = \boldsymbol{M}_{\boldsymbol{\iota}}\boldsymbol{Y}_t\boldsymbol{D}^{-1/2}$ is the $T \times n$ matrix containing the standardised variables,
- $\boldsymbol{D} = < \boldsymbol{V} >$ is the diagonal matrix containing the variances of the n time series $\boldsymbol{y}_t$,
- $\boldsymbol{M}_{\boldsymbol{\iota}} = \boldsymbol{I}_T - \boldsymbol{\iota}(\boldsymbol{\iota}'\boldsymbol{\iota})^{-1}\boldsymbol{\iota}'$ is the projection matrix that returns the deviations from the mean value,
- $\boldsymbol{H} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'$ is the product of the eigenvector matrices and the square root of the eigenvalues of the inverse correlation matrix $\boldsymbol{R}^{-1} = (\boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{D}^{-1/2})^{-1}$.

6.1 Distribuzione asintotica del test DF

The OLS estimator of an AR(1) model $y_t = \phi y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim WN(0, \sigma^2)$, is given by

$$\hat{\phi} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_t^2} = \phi + \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}.$$

Under stationarity ($|\phi| < 1$), the asymptotic distribution of the estimator is

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{\text{d}} N(0, 1 - \phi^2),$$

hence, in the case of a unit root, **the variance collapses to zero** and the t-test cannot be performed.

Under $H_0 \colon \phi = 1$, the quantity $\hat{\phi} - 1$ must be multiplied by T rather than $\sqrt{T}$ in order to converge in distribution (rather than in probability!). To prove this, it is necessary to examine the behavior of the distributions of the estimator's numerator and denominator separately.

1. **Numerator**
   Given that $y_t^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2$, it follows that

$$\begin{aligned}
\sum_{t=1}^{T} y_{t-1}\varepsilon_t &= \sum_{t=1}^{T} \frac{1}{2}(y_t^2 - y_{t-1}^2 - \varepsilon_t^2) \\
&= \frac{1}{2}(\cancel{y_1^2} + \cancel{y_2^2} + \ldots + \cancel{y_{T-1}^2} + y_T^2 - y_0^2 - \cancel{y_1^2} - \cancel{y_2^2} - \ldots - \cancel{y_{T-1}^2}) - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t^2 \\
&= \frac{1}{2}(y_t^2 - y_0^2) - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t^2.
\end{aligned}$$

   By setting $y_0 = 0$, we simply assign an initial value from which the time series starts, thus with no loss of generality; under this assumption, we obtain

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = \frac{1}{2}\left(\frac{y_t}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2 T}\sum_{t=1}^{T} \varepsilon_t^2;$$

   this allows us to define the first moment of the random walk as $E(y_t) = 0$. Since $Var(y_t) = T\sigma^2$, standardising yields

$$\begin{cases}
\dfrac{y_t}{\sigma\sqrt{T}} \xrightarrow{\text{d}} N(0,1) \qquad \Rightarrow \qquad \dfrac{y_t^2}{T\sigma^2} \xrightarrow{\text{d}} \chi_1^2 \\[2ex]
\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{\text{pr}} \sigma^2 \qquad \Rightarrow \qquad \dfrac{1}{T\sigma^2}\displaystyle\sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{\text{pr}} \dfrac{\sigma^2}{\sigma^2} = 1,
\end{cases}$$

   then, by Slutsky's Theorem, it follows that the numerator has a **shifted and scaled** chi-squared distribution with 1 *d.o.f.*, that is

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{\text{d}} \frac{1}{2}(\chi_1^2 - 1).$$

2. **Denominatore**
   Putting the pieces together, under H_0 it therefore follows that Since we have

$$E(y_{t-1}^2) = (t-1)\sigma^2 \qquad \Rightarrow \qquad E\left(\sum_{t=1}^{T} y_t^2\right) = \sigma^2 \sum_{t=1}^{T}(t-1) = \sigma^2 \frac{T(T-1)}{2},$$

   the denominator does <u>not</u> diverge as $T \to +\infty$ only if it is divided by T^2, that is,

$$E\left(\frac{1}{T^2}\sum_{t=1}^{T} y_t^2\right) = \sigma^2 \frac{\not{T}(T-1)}{2\not{T}^2} = \frac{\sigma^2}{2}\left(\frac{T-1}{T}\right).$$

   Hence, the denominator has a **non-standard distribution**.

Therefore, putting the pieces together, under H_0 it follows that

$$\sqrt{T}(\hat{\phi}-1) = \sqrt{T}\frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2} \qquad \Rightarrow \qquad T(\hat{\phi}-1) = \frac{T\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2} = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{\text{d}} \text{DF}.$$

6.2 Phillips-Perron Test

The test proposed by Phillips e Perron (1988), henceforth the PP test, enables testing for the presence of a unit root even when dealing with dynamics that are more general than the AR(p) process used to derive the ADF test. From an analytical perspective, the hypothesis framework is a one-tailed test of the form

$$\begin{cases} H_0 \colon \Delta y_t = \varepsilon_t \\ H_1 \colon \Delta y_t = (\phi - 1) y_{t-1} + \varepsilon_t \end{cases} \tag{55}$$

where $\phi < 1$ represents the coefficient of the AR(1) model $y_t = n_t + \phi y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \text{WN}(0, \sigma^2)$. In this model, $n_t = \delta_0 + \delta_1 t$ is the deterministic component, which reduces to a simple constant if $\delta_1 = 0$. The PP test statistic essentially consists of a modification of the estimator test statistic by Dickey e Fuller (1979), obtained via a rather complex expression[9] given by

$$Z_\tau = \frac{\hat{\phi}-1}{\text{se}(\hat{\phi})}\sqrt{\frac{\hat{\gamma}_0\hat{\phi}}{\hat{\omega}_m(\hat{\varepsilon}_t)}} - \frac{\hat{\omega}_m(\hat{\varepsilon}_t) - \hat{\gamma}_0}{2\,\hat{\omega}_m(\hat{\varepsilon}_t)}\frac{T\,\text{se}(\hat{\phi})}{\hat{s}^2},$$

where

- $\hat{\phi}$ is the OLS estimate of the parameter ϕ in the AR(1) process,
- $\text{se}(\hat{\phi})$ is the standard error of the parameter ϕ estimated via the OLS method,
- $\hat{s}^2 = \dfrac{1}{T-k}\displaystyle\sum_{t=1}^{T}\hat{\varepsilon}_t^2$ is the OLS estimator for the variance of the innovations, where k denotes the number of parameters estimated in the AR(1) model: if the deterministic component contains only a constant, the estimated parameters are $k = 2$; otherwise, if a linear trend is also included, the estimated parameters are $k = 3$,
- $\hat{\gamma}_0 = \dfrac{\sigma^2}{1-\hat{\phi}}$, where $\hat{\sigma}^2 = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\hat{\varepsilon}_t^2$ is the second estimator as defined in equation (9),

[9] Phillips e Perron (1988) also provide the alternative test statistic $Z_\rho = T(\hat{\phi}-1) - \dfrac{TVar(\hat{\phi})}{\hat{s}^2}[\hat{\omega}_m(\hat{\phi}) - \hat{\gamma}_0]$.

- $\hat{\omega}_m(\hat{\varepsilon}_t)$ is the *long-run variance* of the residuals.

Long-run variance:

Given a stochastic process x_t, its long-run variance is given by the expression

$$\hat{\omega}_m(x_t) = \frac{1}{T}\sum_{t=m}^{T-m}\left[\sum_{i=-m}^{m}\kappa_i(x_t-\bar{x})(x_{t-i}-\bar{x})\right] = \sum_{i=-m}^{m}\kappa_i\left[\frac{1}{T}\sum_{t=m}^{T-m}(x_t-\bar{x})(x_{t-i}-\bar{x})\right] = \sum_{i=-m}^{m}\kappa_i\hat{\gamma}_i,$$

where $\hat{\gamma}_i$ is the i-th estimated autocovariance, κ_i is the weight assigned to each autocovariance, and m is a truncation parameter that can be specified by the user, although the first integer below $4(T/100)^{2/9}$ is frequently adopted as the default value.
A weight-determination mechanism (*kernel*) known as the "Bartlett triangle" (1946) is often adopted, given by equation

$$\kappa_i = \begin{cases} 1 - \dfrac{|i|}{m+1} & \text{if } |i| \leq m \\ \\ 0 & \text{otherwise.} \end{cases}$$

This mechanism assigns a unit weight to the variance and weights that decrease linearly as $|i|$ increases, thereby forming an isosceles triangle shape that resembles a camping tent. In this case, the long-run variance equation can be written as

$$\hat{\omega}_m(x_t) = \hat{\gamma}_0 + 2\sum_{i=1}^{m}\kappa_i\hat{\gamma}_i,$$

since the symmetry property $\hat{\gamma}_i = \hat{\gamma}_{-i}$ holds.

Compared to the ADF test, the PP test offers the following advantages:

(a) it is robust to heteroskedasticity in the error term ε_t,

(b) it does not require any specification of the lag length within the model for y_t.

On the other hand, Davidson e MacKinnon (2004) show that the ADF test is preferable to the PP test in finite samples.
Analogously to the ADF test, the asymptotic distribution of the PP test statistic is non-standard; therefore, its critical values are drawn from a specific distribution tailored to this test.

6.3 KPSS Test

The KPSS test (an acronym derived from the authors' initials, Kwiatkowski, Phillips, Schmidt e Shin, 1992) is a *non-parametric* unit root test based on the following system of equations:

$$\begin{cases} y_t = bt + \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + u_t, \end{cases} \tag{56}$$

where t is the linear trend, $\mu_t \sim \text{RW}$, and $u_t \sim \text{WN}(0, \sigma_u^2)$, whereas ε_t is a zero-mean process with a variance that is not necessarily constant over time (heteroskedasticity). The KPSS test is essentially configured as a test of whether the variance σ_u^2 is zero: under the null hypothesis, therefore, the process μ_t is constant over time, and y_t is consequently stationary. The hypothesis framework is thus the reverse of that of the ADF and PP tests. In practice, it features the following one-tailed test structure:

$$\begin{cases} H_0\colon \sigma_u^2 = 0 & \Rightarrow \; y_t \sim I(0) \\ H_0\colon \sigma_u^2 > 0 & \Rightarrow \; y_t \sim I(1). \end{cases} \tag{57}$$

The test statistic is given by

$$\text{KPSS} = \frac{1}{T^2\gamma_0^*}\sum_{t=1}^{T}S_t^2 \tag{58}$$

where

- $S_t = \sum_{i=1}^{t} \hat{\varepsilon}_i$ is a *Brownian bridge* formed by the cumulative sums of the residuals (note, indeed, that $S_0 = S_T = 0$),
- $\hat{\varepsilon}_t = y_t - \hat{\mu} - \hat{b}t$ (μ_t is constant under H_0),
- γ_0^* is the long-run variance of y_t calculated via the same procedure used for the PP test.

The asymptotic distribution of this statistic is non-standard; therefore, the critical values for the KPSS test are computed from a specific distribution tailored to this test.

7 Forecasting

As is well known, the objective of time series models is very often to provide forecasts regarding the path over time of one or more dependent variables contained in the vector y_t. Given a sample of T observations, y_t denotes the value of a variable of interest at time t, and f_t defines the time series of some forecast obtained following the estimation of a given model. The forecast carried out in this manner can be:

in-sample, meaning that the time series f_t consists of T forecasts given by the fitted values of the model applied to all observations within the sample ($t = 1, 2, \ldots, T$);

out-of-sample, meaning that the series f_t consists of a number of forecasts h, which represents the forecast horizon, *i.e.*, the number of steps ahead that the analyst has chosen to consider ($t = T+1, T+2, \ldots, T+h$). In practice, a model is estimated on the available sample, and the information obtained is then exploited to forecast the path of y_t from the $(T+1)$-th observation onwards. In the specific case where the final sample observation y_T refers to the present day, this type of analysis constitutes a genuine "forecast of the future".

Another important distinction is between a *static forecast* and a *dynamic forecast*. The difference between these two methods essentially lies in the alternative ways of updating the information set $\mathbb{I}_{t-1}$ on the basis of which the forecasts are obtained. In particular, since this information set is given by the history of the series y_t, in a static forecast it is updated whenever a new data point is observed. In a dynamic forecast, by contrast, the forecasts are based on knowledge of the phenomenon (information) under investigation up to a specific period $t = T_0$. In this case, sample data subsequent to period T_0 do not flow into $\mathbb{I}_{t-1}$.

7.1 Static forecast

Consider the following partitioned vector

$$y_t = [y_1 \quad y_2 \quad \ldots \quad y_{T_0} \mid \underbrace{y_{T_0+1} \quad y_{T_0+2} \quad \ldots \quad y_T}_{h \text{ observations}}]',$$

where $h = T - T_0$. The static forecast is obtained by dividing the available sample into two consecutive subsamples. The estimation of the time series model takes place within the first subsample ($t = 1, 2, \ldots, T_0$), while the forecasts are generated for the h observations of the second subsample. In this context, it is important to emphasise that the sample observations of the second subsample are available; hence, each forecast f_t ($t = T_0 + 1, T_0 + 2, \ldots, T$) will be produced by exploiting this information.
Defining the forecast error as $\hat{\varepsilon}_t|\mathbb{I}_{t-1} = y_t - f_t|\mathbb{I}_{t-1}$, where $\mathbb{I}_{t-1}$ is the information set at time $t-1$, the following forecasts can be obtained via the static forecasting method:

t	forecast error	information set
one step ahead (T_0+1)	$\hat{\varepsilon}_{T_0+1}\|\mathbb{I}_{T_0} = y_{T_0+1} - f_{T_0+1}\|\mathbb{I}_{T_0}$	$\mathbb{I}_{T_0} = \{y_{T_0}, y_{T_0-1}, y_{T_0-2}, \ldots\}$
two step ahead (T_0+2)	$\hat{\varepsilon}_{T_0+2}\|\mathbb{I}_{T_0+1} = y_{T_0+2} - f_{T_0+2}\|\mathbb{I}_{T_0+1}$	$\mathbb{I}_{T_0+1} = \{y_{T_0+1}, y_{T_0}, y_{T_0-1}, \ldots\}$
$\vdots$	$\vdots$	$\vdots$
h steps ahead (T_0+h)	$\hat{\varepsilon}_T\|\mathbb{I}_{T-1} = y_T - f_T\|\mathbb{I}_{T-1}$	$\mathbb{I}_{T-1} = \{y_{T-1}, y_{T-2}, y_{T-3}, \ldots\}$

Observing this sequence, it can be noticed that at each step ahead:

- the information set upon which the conditioning is based is updated by adding the sample information $y_{T_0+1}, y_{T_0+2}, \ldots, y_T$;
- obviously, the forecasts $f_{T_0+1}, f_{T_0+2}, \ldots, f_T$ exploit this information. From a strictly technical perspective, this is equivalent to applying the econometric model estimated up to observation T_0 to the data collected from period T_0 to period T;
- the entire procedure consists of obtaining a sequence of one-step-ahead forecasts; indeed, each forecast f_t exploits all information available up to y_{t-1}.

7.2 Dynamic forecast

In this context, the forecasting mechanism is as follows:

t	forecast errore	information set
one step ahead (T_0+1)	$\hat{\varepsilon}_{T_0+1}\|\mathbb{I}_{T_0} = y_{T_0+1} - f_{T_0+1}\|\mathbb{I}_{T_0}$	$\mathbb{I}_{T_0} = \{y_{T_0}, y_{T_0-1}, y_{T_0-2}, \ldots\}$
two steps ahead (T_0+2)	$\hat{\varepsilon}_{T_0+2}\|\mathbb{I}_{T_0+1} = y_{T_0+2} - f_{T_0+2}\|\mathbb{I}_{T_0+1}$	$\mathbb{I}_{T_0+1} = \{f_{T_0+1}, y_{T_0}, y_{T_0-1}, \ldots\}$
$\vdots$	$\vdots$	$\vdots$
h steps ahead (T_0+h)	$\hat{\varepsilon}_T\|\mathbb{I}_{T-1} = y_T - f_T\|\mathbb{I}_{T-1}$	$\mathbb{I}_{T-1} = \{f_{T-1}, f_{T-2}, f_{T-3}, \ldots\}$

The crucial difference compared to the static forecasting case is that the information set is not updated by adding the sample observations from $t = T_0 + 1$ onwards; instead, the forecasts themselves are used as they are generated. For instance, when forecasting y_{T_0+5}, the sample observations y_{T_0+1}, y_{T_0+2}, y_{T_0+3}, and y_{T_0+4} cannot be used, and the previously computed forecasts f_{T_0+1}, f_{T_0+2}, f_{T_0+3}, and f_{T_0+4} flow into $\mathbb{I}_{t-1}$ in their place. Under this mechanism:

- only the one-step-ahead forecast is identical to the one obtained in the static framework, because the conditioning set $\mathbb{I}_{T_0}$ is the same in both cases;
- the dynamic forecast updates the information set using the forecasts f_t; consequently, the resulting forecasted values are based on forecasts instead of the sample observations y_t. This augments the overall uncertainty, thereby inherently increasing the variance of the forecast error, which is the primary reason why, in general, less accurate forecasts are obtained via dynamic forecasting.

7.3 Measures of Forecast Evaluation

In this section, only one forecast performance indicator and one test particularly used within the time series context will be presented.

7.3.1 Root Mean Squared Error (RMSE)

Given the time series y_t with size T observations and the sequence of forecasts f_t having the same dimension, the quadratic mean of the forecast error defines the Root Mean Squared Error (RMSE). In formulas[10]

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2}, \tag{59}$$

[10]For completeness, it should be noted that the list of potential forecast performance indicators is rather extensive. For a comprehensive overview of the topic, reading section 29.4 within the software *User Guide* for `Gretl` (Cottrell e Lucchetti, 2015) is strongly recommended.

where $\hat{\varepsilon}_t = y_t - f_t$. Clearly, this indicator can take only positive values as it is constructed as a mean of squares. Its theoretical minimum value is zero, which would occur if the forecasts perfectly matched the observations of the dependent variable, *i.e.* $f_t = y_t$ for $\forall\, t$. Based on these properties, it is evident that, among alternative forecasts, the preferred one is that associated with the lowest RMSE.

On the other hand, like all indicators, the RMSE suffers from two limitations:

(a) there is no threshold value beyond which the forecast f_t can be considered "far" from the time series y_t,

(b) assuming that such a threshold exists in a specific analysis, it is not necessarily applicable to other contexts. This depends on the unit of measurement (order of magnitude) in which the series $\hat{\varepsilon}_t$ is expressed. For instance, RMSE = 100 would be a very high value if the forecast errors all fell within the interval $[-1, 1]$, whereas it would be considerably "small" if those errors were all 10 times larger.

7.3.2 Diebold-Mariano test

Very often in empirical applications, a time series is analysed using several alternative forecasting models. In this context, the need inherently arises to determine which of these models exhibits the superior forecasting performance regarding the path over time of a time series y_t. In other words, the objective is to identify the model that provides the most accurate forecast[11]. If we consider two time series f_{1t} and f_{2t} corresponding to two different forecasts obtained from two distinct econometric models, the test proposed by Diebold e Mariano (1995), henceforth the DM test, was conceived as a test for predictive accuracy; hence, it attempts to answer the following question: are f_{1t} and f_{2t} equally capable of forecasting y_t, or is one preferable to the other?
From a technical perspective, it is necessary to define the forecast error given by the expression

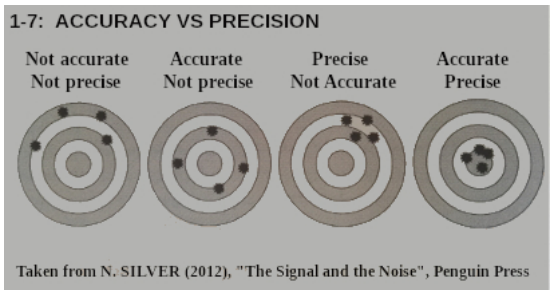
$$e_{it} = f_{it} - y_t,$$

with $i = 1, 2$. At this stage, $L(e_{it})$ defines a *loss function* on this forecast error, characterized by the following properties:

1. $L(0) = 0$, meaning that the loss function is zero if the forecast error is zero, and thus if $f_{it} = y_t$ for all $t = 1, 2, \ldots, T$;
2. $L(e_{it}) > 0$ when $f_{it} \neq y_t$ for at least one value of $t = 1, 2, \ldots, T$;
3. $L(e_{it})$ is monotonically non-decreasing with respect to the forecast error magnitude. In other words, the loss function does not decrease if $|e_{it}|$ increases.

Typical examples of loss functions satisfying these properties are[12]

[11] The DM test seeks to identify the most accurate forecasting model, rather than the most precise one. Heuristically, one could think of an accurate model as one that, at each step t, yields "large" forecast errors but with alternating signs; a precise model, by contrast, commits "smaller" forecast errors, but with signs that tend to persist. The figure presented here, taken from the bestseller by Silver (2012), illustrates this concept very clearly.



[12] In fact, loss functions can be defined in many other ways. Several examples are implemented in the `Gretl` software within the additional package *DiebMar*, which can be downloaded from the website
`https://gretl.sourceforge.net/cgi-bin/gretldata.cgi?opt=SHOW_FUNCS`.

- *U-shaped* loss function: $L(e_{it}) = e_{it}^2$,
- *V-shaped* loss function: $L(e_{it}) = |e_{it}|$.

The DM test is based on the difference between the loss functions computed for the two forecasts; hence, it is given by

$$d_t = L(e_{1t}) - L(e_{2t}) = L(f_{1t} - y_t) - L(f_{2t} - y_t). \tag{60}$$

The two forecasts f_{1t} and f_{2t} are equally accurate if the expected value of this difference is zero for every value of t. Formally, the hypothesis structure is as follows:

$$\begin{cases} H_0 \colon E(d_t) = 0 & f_{1t} \text{ and } f_{2t} \text{ have the same predictive accuracy} \\ H_1 \colon E(d_t) \neq 0 & \text{one forecast between } f_{1t} \text{ and } f_{2t} \text{ is more accurate} \end{cases} \tag{61}$$

Technically, the test is constructed by exploiting the Lindeberg-Lévy Central Limit Theorem (under the assumption that the series d_t is stationary and has short memory). Therefore, consider the following limiting distribution:

$$\sqrt{h}(\bar{d} - \mu) \xrightarrow{\text{d}} N(0, \hat{\omega}_m(d_t)),$$

where

- $\bar{d} = \dfrac{1}{h}\sum_{t=1}^{h} d_t$ is the sample mean of the difference d_t,
- $h = T - T_0$ is the number of steps ahead considered when making the forecasts,
- $\text{LRV}(d_t)$ is the long-run variance of d_t,
- $\mu = E(d_t)$. Clearly, under H_0, it follows that $\mu = 0$.

Test statistic:

$$\text{DM} = \frac{\bar{d}}{\sqrt{\dfrac{\hat{\omega}_m(d_t)}{h}}} = \frac{\bar{d}}{\sqrt{\dfrac{1}{h}\displaystyle\sum_{i=-m}^{m} \kappa_i \hat{\gamma}_i}} \xrightarrow{\text{d}} N(0,1),$$

where $\hat{\gamma}_i = \dfrac{1}{T}\displaystyle\sum_{t=m}^{T-m} (d_t - \bar{d})(d_{t-|i|} - \bar{d})$ are the sample autocovariances of the series d_t, while the weights κ_i are determined via the "Bartlett triangle", a mechanism already illustrated in the box on page 20. Given that under the null hypothesis the DM test statistic is asymptotically distributed as a standard normal random variable, the quantiles of this distribution ($z_{\alpha/2}$) are used as critical values for the test. Since the DM test is a two-tailed test, H_0 is rejected when

$$|\text{DM}| > z_{\alpha/2} \qquad \text{or} \qquad p\text{-value} = 2Pr(Z > \text{DM}) < \alpha,$$

where $Z \sim N(0,1)$ and α is the significance level of the test (generally, $\alpha = 0.05$).

Finally, the following aspects should be considered:

(a) it is possible to use the DM test for *in-sample* forecasts simply by setting $T_0 = 1$. In this manner, the forecasts are generated by exploiting the entire sample size ($h = T$);

(b) it is possible to use the DM test to evaluate the forecasting performance of a single series f_t. In this context, it is sufficient to set $f_{1t} = f_t$ and $f_{2t} = y_t$ within equation (60).

8 Dummy Variables

In the context of econometric time series models, dummy variables are frequently employed as explanatory variables to capture the impact of exceptional events, such as wars, crises, or currency devaluations, whose effects are fully contained within a few sample observations (generally one).
Technically, dummy variables are configured as dichotomous variables that take a value of one when the extraordinary event (or *outlier*) is observed, and a value of zero when such event does not occur. With reference to a generic dynamic model linear in the parameters of the form $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \delta d_t + \varepsilon_t$, it follows that

$$\hat{y}_t = E(y_t|\mathbb{I}_{t-1}) = \begin{cases} \boldsymbol{x}_t'\hat{\boldsymbol{\beta}} & \text{se } d_t = 0 \text{ (the exceptional event does not occur at time } t) \\ \boldsymbol{x}_t'\hat{\boldsymbol{\beta}} + \hat{\delta} d_t & \text{se } d_t = 1 \text{ (the exceptional event occurs at time } t), \end{cases}$$

where d_t is a scalar dummy variable, and $\mathcal{I}_{t-1}$ represents the information set at time $t-1$. Obviously, should more than one dummy variable be used, the second term within the model would be given by the scalar product $\boldsymbol{d}_t'\boldsymbol{\delta}$. The sign of the estimated parameter $\hat{\delta}$ indicates whether, relative to the conditional expectation of y_t, the *outlier* takes a significantly higher or lower value.

Although dummy variables are a useful tool for interpreting the peaks and/or troughs observed over time within time series plots, it is nevertheless bad practice to overuse them. This stems from the fact that these variables are not genuine explanatory variables, because they are created *ad hoc* by the user to interpret those movements that the "true" explanatory variables fail to capture.
Another recommendation is that a separate dummy variable should be used for each unusual event. This choice is essentially driven by two reasons: from a numerical perspective, the coefficient of a dummy variable employed jointly for both a peak and a trough could suffer from a cancellation effect, such that the coefficient itself might turn out to be statistically insignificant; from the perspective of economic interpretation, using the same dummy variable for more than one outlier would invalidate the analysis, as it would be impossible to isolate the effect caused by each individual unusual event.

References

Akaike, H. (1974). *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, 19(6), pp. 716–723.

Bartlett, M. S. (1946). *On the theoretical specification and sampling properties of autocorrelated time series.* Supplement to the Journal of the Royal Statistical Society, 8, pp. 27–41.

Bollerslev, T., Engle, R. F. e Nelson, D. B. (1994). *ARCH models.* in R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, Elsevier, North Holland.

Breusch, T. S. (1979). *Testing for autocorrelation in dynamic linear models.* Australian Economic Papers, 17, pp. 334–355.

Cottrell, A. e Lucchetti, R. (2015). *Gretl User's Guide.* software Gretl.

Davidson, R. e MacKinnon, J. G. (2004). *Econometric Theory and Methods.* Oxford University Press, New York.

Dickey, D. A. e Fuller, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root.* Journal of the American Statistical Association, 74(366), pp. 427–431.

Diebold, F. X. e Mariano, R. M. (1995). *Comparing predictive accuracy.* Journal of Business and Economic Statistics, 13, pp. 253–263.

Doornik, J. A. e Hansen, H. (2008). *An omnibus test for univariate and multivariate normality.* Oxford Bulletin of Economics and Statistics, 70, pp. 927–939.

Durbin, J. e Watson, G. (1950). *Testing for serial correlation in least squares regression.* Biometrika, 37, pp. 409–428.

Engle, R. F. (1982). *Autoregressive conditional heteroskedasticity with estimates of the U.K. inflation.* Econometrica, 50, pp. 987–1008.

Godfrey, L. G. (1978). *Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables.* Econometrica, 46, pp. 1293–1302.

Hannan, E. J. e Quinn, B. G. (1979). *The determination of the order of an autoregression.* Journal of the Royal Statistical Society B, 41, pp. 190–195.

Hill, C. R., Griffiths, W. E. e Lim, G. C. (2013). *Prinicìpi di Econometria.* Zanichelli, Bologna.

Jarque, C. M. e Bera, A. K. (1980). *Efficient test for normality, homoscedasticity and serial independence of regression residuals.* Economics Letters, 6(3), pp. 255–259.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. e Shin, Y. (1992). *Testing the null hypothesis of stationarity against the alternative of a unit root.* Journal of Econometrics, 54(1-3), pp. 159–178.

Ljung, G. M. e Box, G. E. P. (1978). *On a measure of a lack of fit in time series models.* Biometrika, 65(2), pp. 297–303.

Lucchetti, R. (2026). *Notes on Time Series Analysis.* disponibile su `www2.econ.univpm.it/servizi/hpp/lucchetti/didattica/matvario/stocproc.pdf`.

Palm, F. C. (1996). *GARCH models of volatility.* in G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, vol. 14, Statistical Methods in Finance, Elsevier, North Holland.

Phillips, P. C. B. e Perron, P. (1988). *Testing for a unit root in time series regression.* Biometrika, 75(2), pp. 335–346.

Savin, N. E. e White, K. J. (1977). *The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors*. Econometrica, 45, pp. 1989–1996.

Schwarz, G. E. (1978). *Estimating the dimension of a model*. Annals of Statistics, 6(2), pp. 461–464.

Silver, N. (2012). *Il segnale e il rumore*. Fandango edizioni.

Verbeek, M. (2004). *A guide to modern econometrics*. John Wiley & Sons, Rotterdam, 2nd edition.