Business School, Brunel University
MSc. EC5501/5509 Modelling Financial Decisions and Markets/Introduction to Quantitative Methods
Prof. Menelaos Karanasos (Room SS269, Tel. 01895265284)

Lecture Notes 5

1. The Multivariate Classical Linear Regression Model

The general form of the multiple linear regression model is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \varepsilon_t, \tag{1}$$

where $t = 1, ...T$, indicates the observation number, x_{it} and ε_t are the observations of the ith regressor and the error term at period t, respectively. y_t denotes the observation of the dependent variable at period t and the β's are (scalar) parameters. It is customary to set $x_{it} = 1$ for all t, in which case β_1 is the constant term of the regression and $\beta_2, ..., \beta_k$ are the slope coefficients of the regression:

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \varepsilon_t. \tag{1'}$$

Using matrix notation, (1') can be written as

$$y_t = \beta' x_t + \varepsilon_t, \ t = 1, ..., T, \tag{1a}$$

where β and x_t are $(k \times 1)$ vectors of parameters and exogenous variables, respectively. Alternatively,

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

where y and ε are $(T \times 1)$ vectors of the endogenous variable and the error term, respectively. X is a $(T \times k)$ matrix of exogenous variables with 1's in its first column, and β is a $(k \times 1)$ vector of parameters.

Assumptions

[**A1**] Functional form: the relationship between y and X is linear.
[**A2**] Zero mean of the disturbance: $E(\varepsilon_t) = 0$, for all t.
[**A3**] Homoscedasticity: $V(\varepsilon_t) = \sigma^2$, for all t.
[**A4**] No serial correlation: $C(\varepsilon_t, \varepsilon_s) = 0$, if $t \neq s$.
[**A5**] No perfect multicollinearity: rank$(X) = k$, $k < T$.
[**A6**] Non stochastic regressors.
[**A6'**] Uncorrelatedness of X and ε: $E(X'\varepsilon) = 0$.
[**A7**] Normality: $\varepsilon \sim N(0, \sigma^2 I)$, where I is a $(T \times T)$ identity matrix.

Assumptions [A2]-[A4] can be summarised as $\boxed{\varepsilon_t \sim iid\left(0,\sigma^2\right)}$. Furthermore, using matrix notation, assumptions [A2]-[A4] can be written as

$$
\begin{aligned}
\underset{(T\times 1)}{E\ (\varepsilon)} &= 0 \\
\underset{(T\times T)}{V\left(\varepsilon\right)} &= \underset{(T\times 1)(1\times T)}{E\ \ \left(\varepsilon\varepsilon'\right)} = \begin{bmatrix} E\left(\varepsilon_1^2\right) & E\left(\varepsilon_1\varepsilon_2\right) & . & . & . & E\left(\varepsilon_1\varepsilon_T\right) \\ E\left(\varepsilon_2\varepsilon_1\right) & E\left(\varepsilon_2^2\right) & . & . & . & E\left(\varepsilon_2\varepsilon_T\right) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ E\left(\varepsilon_T\varepsilon_1\right) & E\left(\varepsilon_T\varepsilon_2\right) & . & . & . & E\left(\varepsilon_T^2\right) \end{bmatrix} \\
&= E\left(\varepsilon\varepsilon'\right) = \begin{bmatrix} \sigma^2 & 0 & . & . & . & 0 \\ 0 & \sigma^2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sigma^2 \end{bmatrix} = \sigma^2 I
\end{aligned}
$$

So we can write: $\boxed{\varepsilon \sim D\left(0,\sigma^2 I\right)}$

1.1. Least Squares Estimation

The least squares coefficient vector $\left(\overset{\wedge}{\beta}_{OLS}\right)$ minimizes the sum of squared residuals $\left(\widehat{\varepsilon}'\widehat{\varepsilon} = \sum_{t=1}^{T}\widehat{\varepsilon}_t^2\right)$. From (1b) we have that $y = \widehat{y} + \widehat{\varepsilon}$, where y and $\widehat{y}$ are the actual and fitted values of the endogenous variable, respectively. We also have that $\widehat{y} = X\widehat{\beta}$, and so $\widehat{\varepsilon} = y - X\widehat{\beta}$. Thus the sum of squared residuals is given by

$$
\begin{aligned}
\widehat{\varepsilon}'\widehat{\varepsilon} &= \left(y - X\widehat{\beta}\right)'\left(y - X\widehat{\beta}\right) \\
&= y'y - \widehat{\beta}'X'y - y'X\widehat{\beta} + \widehat{\beta}'X'X\widehat{\beta} \\
&= y'y - 2y'X\widehat{\beta} + \widehat{\beta}'X'X\widehat{\beta}.
\end{aligned}
$$

The solution to the above minimization problem is[1]

$$\min\left(\widehat{\varepsilon}'\widehat{\varepsilon}\right) \quad : \quad \frac{\partial\left(\widehat{\varepsilon}'\widehat{\varepsilon}\right)}{\partial\widehat{\beta}} = -2X'y + 2X'X\widehat{\beta} = 0$$

$$\Rightarrow \underset{(k\times 1)}{\widehat{\beta}} = \left(X'X\right)^{-1}X'y. \qquad (2)$$

To derive (2) we used the following matrix differentiation rules: $\frac{\partial(X\beta)}{\partial\beta} = X'$ and $\frac{\partial\left(\beta'X'X\beta\right)}{\partial\beta} = 2X'X\beta$, since $X'X$ is a symmetric matrix.

- Note that $\left(X'X\right)$ has an inverse, and consequently least-squares estimation is feasible, because of assumption [A5].

1.2. Gauss-Markov Theorem

Under assumptions [A1]-[A6] the least squares estimator $\widehat{\beta}$ is BLUE (best linear unbiased estimator)**.**

<u>Proof:</u> From eq. (2) we have that $\widehat{\beta} = Ay$, where $\underset{(k\times T)}{A} = \left(X'X\right)^{-1}X'$, i.e. $\widehat{\beta}$ is a *linear estimator.* In addition, observe that $\widehat{\beta}$ can be expressed as

$$\begin{aligned} \widehat{\beta} &= \left(X'X\right)^{-1}X'y = \left(X'X\right)^{-1}X'\left(X\beta+\varepsilon\right) \\ &= \beta + \left(X'X\right)^{-1}X'\varepsilon, \text{ i.e.} \\ \widehat{\beta} &= \beta + A\varepsilon. \qquad (2') \end{aligned}$$

Thus

$$E\left(\widehat{\beta}\right) = \beta + AE\left(\varepsilon\right) \Rightarrow E\left(\widehat{\beta}\right) = \beta, \qquad (2a)$$

[1] The second order conditions of the above minimisation problem are

$$\frac{\partial^2\left(\widehat{\varepsilon}'\widehat{\varepsilon}\right)}{\partial\widehat{\beta}\partial\widehat{\beta}'} = 2X'X,$$

where $X'X$ is a positive definite matrix. To see this consider an arbitrary $(k\times 1)$ vector α and the quadratic form $q = \alpha'X'X\alpha = v'v = \sum_i v_i^2 > 0$; since $q > 0$ we have that $X'X$ is positive definite.

and so $\widehat{\beta}$ is an *unbiased* estimator of β.
Furthermore, the variance covariance matrix of $\widehat{\beta}$ is given by

$$\begin{aligned}
\underset{(k\times k)}{V}\left(\widehat{\beta}\right) &= E\left[\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^{'}\right] \\
&= E\left[\left(A\varepsilon\right)\left(A\varepsilon\right)^{'}\right] = E\left(A\varepsilon\varepsilon^{'}A^{'}\right) \\
&= AE\left(\varepsilon\varepsilon^{'}\right)A^{'} = A\left(\sigma^{2}I\right)A^{'} = \sigma^{2}AA^{'} \Rightarrow \\
V\left(\widehat{\beta}\right) &= \sigma^{2}\left(X^{'}X\right)^{-1}, \qquad (2b)
\end{aligned}$$

since $AA' = \left(X'X\right)^{-1}X'X\left(X'X\right)^{-1} = \left(X'X\right)^{-1}$. It can also be shown that any other linear unbiased estimator b of β has greater variance than $\widehat{\beta}$. In this sense, $\widehat{\beta}$ is *best* (it has the minimum variance among all linear unbiased estimators of β)[2].

[2] To complete the proof of the Gauss-Markov theorem we need to show that any other linear unbiased estimator, b, has greater variance than $\widehat{\beta}$. Without loss of generality we can write:

$$\begin{aligned}
b &= \left(A+C\right)y = \widehat{\beta}+Cy = \widehat{\beta}+CX\beta+C\varepsilon, \\
E\left(b\right) &= E\left(\widehat{\beta}\right)+CX\beta+CE\left(\varepsilon\right) \Rightarrow \\
E\left(b\right) &= \beta \text{ if } CX=0.
\end{aligned}$$

Observe that

$$\begin{aligned}
b-\beta &= \left(A+C\right)y-\beta = \left(A+C\right)\left(X\beta+\varepsilon\right)-\beta \\
&= AX\beta+CX\beta+\left(A+C\right)\varepsilon-\beta = \left(A+C\right)\varepsilon,
\end{aligned}$$

since $CX=0$ and $AX = \left(X^{'}X\right)^{-1}X^{'}X = I$. Therefore,

$$\begin{aligned}
V\left(b\right) &= E\left[\left(b-\beta\right)\left(b-\beta\right)^{'}\right] = E\left[\left(A+C\right)\varepsilon\right]\left[\left(A+C\right)\varepsilon\right]^{'} \\
&= E\left[\left(A+C\right)\varepsilon\varepsilon^{'}\left(A+C\right)^{'}\right] = \left(A+C\right)E\left(\varepsilon\varepsilon^{'}\right)\left(A+C\right)^{'} \\
&= \sigma^{2}\left(A+C\right)\left(A+C\right)^{'} = \sigma^{2}\left[\left(X^{'}X\right)^{-1}+CC^{'}\right] \\
&= V\left(\widehat{\beta}\right)+\sigma^{2}CC^{'} \Rightarrow V\left(b\right) \geq V\left(\widehat{\beta}\right),
\end{aligned}$$

since CC' is positive semidefinite.

1.3. Maximum Likelihood Estimation

The joint density of the random variables y_t and x_t can be written as the product of the conditional and marginal densities

$$f(y_t, x_t) = f(y_t/x_t)\, f(x_t).$$

We assume that X has *rank k* and is weakly exogenous with respect to the parameters of interest θ. *Weak exogeneity* requires that the parameters of interest are only functions of the parameters of the conditional distribution and there are no restrictions linking them with the parameters of the marginal distribution. In other words, weak exogeneity allows us to ignore the marginal distribution of x_t; weak exogeneity is required for inference (estimation and testing). Essentially, a variable x_t in a model is defined to be weakly exogenous for estimating a set of parameters θ if inference on θ conditional on x_t involves no loss of information.

For the first two moments of the conditional distribution we assume that (i) the conditional expectation is a linear function of x_t,

$$E(y_t/x_t) = \beta' x_t, \text{ (assumption [A1])} \tag{3a}$$

and (ii) the conditional variance is a constant (independent of x_t),

$$V(y_t/x_t) = \sigma^2. \tag{3b}$$

In addition, we assume that the parameters of interest, $\theta = (\beta, \sigma^2)$, are time invariant.

- The conditional expectation constitutes the *systematic part* of y_t, whereas the *unsystematic part* of y_t is defined as

  $$\varepsilon_t = y_t - E(y_t/x_t) \Rightarrow E(\varepsilon_t) = 0 \text{ (assumption [A2])}, \tag{3c}$$

  since $E(\varepsilon_t) = E(y_t) - E[E(y_t/x_t)] \Rightarrow$(using the law of iterated expectations) $E(\varepsilon_t) = E(y_t) - E(y_t) = 0$.

- The $(T \times T)$ conditional variance-covariance matrix of y is equal to the variance-covariance matrix of ε:

  $$V(y/X) = V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I \text{ (assumptions [A3]-[A4])}. \tag{3d}$$

- The conditional distribution is assumed *normal* (assumption [A7]):

$$f\left(y_t/x_t\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(y_t - \beta^{'}x_t\right)^2}{2\sigma^2}\right\}. \tag{3e}$$

- Observe that in the above discussion we expressed assumptions [A1]-[A7] in terms of the observable y rather than the unobservable ε :

$$(y/X) \sim N\left(X\beta, \sigma^2 I\right).$$

Following Kennedy(1985), "the maximum likelihood principle of estimation is based on the idea that the sample of data at hand is more likely to have come from a real world characterized by one particular set of parameter values than from a real world characterized by any other set of parameter values...The maximum likelihood estimate (MLE) is the particular set of values, $\widehat{\beta}$ and $\widehat{\sigma}^2$, that creates the greatest probability of having obtained the sample in question".

In order to estimate the parameters of interest, β and σ^2, we need to write down the likelihood function of y_t. Since the sample is independent, the likelihood function is proportional to the product of the distributions of each y_t, i.e.

$$L = \prod_{t=1}^{T} f\left(y_t/x_t\right)$$

The log-likelihood is given by

$$\begin{aligned} LL &= \sum_{t=1}^{T} \log f\left(y_t/x_t\right), \text{ where} \\ \log f\left(y_t/x_t\right) &= -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\log 2\pi - \frac{\left(y_t - \beta^{'}x_t\right)^2}{2\sigma^2}. \end{aligned}$$

So the log-likelihood can be written as

$$\begin{aligned} LL &= -\frac{T}{2}\log\sigma^2 - \frac{T}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \beta^{'}x_t\right)^2, \text{ or} \\ LL &= -\frac{T}{2}\log\sigma^2 - \frac{T}{2}\log 2\pi - \frac{1}{2\sigma^2}\left(y^{'}y - 2y^{'}X\beta + \beta^{'}X^{'}X\beta\right). \end{aligned}$$

Maximization of the log-likelihood gives:

$$\begin{aligned}\frac{\partial LL}{\partial \beta} &= -\frac{1}{2\sigma^2}\left(-2X^{'}y + 2X^{'}X\beta\right) = 0 \Rightarrow \boxed{\widehat{\beta} = \left(X^{'}X\right)^{-1}X^{'}y}, \qquad (4)\\ \frac{\partial LL}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{\varepsilon^{'}\varepsilon}{2\sigma^4} = 0 \Rightarrow \boxed{\widehat{\sigma}^2 = \left(\widehat{\varepsilon}^{\prime}\widehat{\varepsilon}\right)/T}.\end{aligned}$$

1.4. Sampling distribution of $\widehat{\beta}$

$\widehat{\beta}$ is a linear function of the normally distributed random vector y, $\widehat{\beta} = Ay$, where $A = \left(X^{'}X\right)^{-1}X^{\prime}$. Hence we have that

$$\widehat{\beta} \sim N\left(AX\beta,\ \sigma^2 AA^{'}\right) \Rightarrow\ \widehat{\beta} \sim N\left(\beta,\ \sigma^2\left(X^{'}X\right)^{-1}\right),$$

1.5. Sampling distribution of $\widehat{\sigma}^2$

(The following results are presented without proof)

$$\textit{(Optional)}\ \frac{T\widehat{\sigma}^2}{\sigma^2} \sim \chi^2\left(T-k\right),\ E\left(\widehat{\sigma}^2\right) = \frac{T-k}{T}\sigma^2,\ V\left(\widehat{\sigma}^2\right) = \frac{2\left(T-k\right)}{T^2}\sigma^4,$$

so $\widehat{\sigma}^2$ is a biased estimator. An unbiased estimator of σ^2 is given by

$$\begin{aligned}\textit{(Important)}\ s^2 &= \frac{\widehat{\varepsilon}^{\prime}\widehat{\varepsilon}}{T-k} \text{ or } s^2 = \frac{T}{T-k}\widehat{\sigma}^2, \qquad (5)\\ \textit{(Optional)}\ E\left(s^2\right) &= \sigma^2,\ V\left(s^2\right) = \frac{2\sigma^4}{T-k},\ V\left(s^2\right) > V\left(\widehat{\sigma}^2\right).\end{aligned}$$

Note that there is a trade-off between unbiasdness and lower variance.

1.6. Degrees of Freedom

The residuals are the observable estimates of the unobservable disturbances. Even when the disturbances are assumed uncorrelated (assumption [A4]), the residuals are linearly dependent. This is just another way of looking at the fact that they must sum to zero (if there is a constant term) and be uncorrelated with the regressors. With a sample of T observations we estimate T residuals. However, k parameters are employed to estimate these residuals, and so only $(T-k)$ of the residuals are linearly independent. That is, given any $(T-k)$ residuals, and the original data, it is possible to deduce the other k residuals.

We say that the sum of squared residuals has $(T-k)$ *degrees of freedom*. The number of degrees of freedom of a sum of squares is just the number of independent observations (in this case residuals) from which the sum of squares is constructed.

1.7. R^2, and Corrected R^2

The total variation of Y can be decomposed as follows:

$$\underbrace{\sum_{t=1}^{T}(y_t-\overline{y})^2}_{\text{Total Variation of } y} = \underbrace{\sum_{t=1}^{T}(\widehat{y}_t-\overline{y})^2}_{\text{Explained Variation of } y} + \underbrace{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}_{\text{Residual Variation of } y}$$

R^2 measures the proportion of the variation in y which is explained by the multiple regression equation:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}{\sum_{t=1}^{T}(y_t-\overline{y})^2}.$$

It is called the (squared) multiple correlation coefficient because it measures the (square of) the correlation between y and a particular linear combination of the x's. If there is a constant term then $0 \leq R^2 \leq 1$.

- R^2 can be used to compare regressions with the same dependent variable.
- The addition of more independent variables to the regression equation can never lower R^2 and is likely to increase it, since adding more variables cannot worsen the explanation of y.

The difficulty with R^2 as a measure of goodness of fit is that R^2 pertains to explained and unexplained *variation* in y and therefore does not account for the number of degrees of freedom. A natural solution is to use *variances,* not variations, thus eliminating the dependence of goodness of fit on the number of exogenous variables in the model (variance equals variation divided by the degrees of freedom). We define the corrected R^2 as

$$\overline{R}^2 = 1 - \frac{\text{variance of the regression}}{\text{sample variance of } y} = 1 - \frac{s^2}{s_y^2} = 1 - \frac{\frac{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}{T-k}}{\frac{\sum_{t=1}^{T}(y_t-\overline{y})^2}{T-1}}.$$

The relationship between R^2 and $\overline{R}^2$ is given by

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\left(\frac{T-1}{T-k}\right).$$

Note that

- If $k = 1$, then $R^2 = \overline{R}^2$.
- If $k > 1$, then $R^2 \geq \overline{R}^2$.
- $\overline{R}^2$ can be negative.
- When new variables are added to the regression model, R^2 always increases, while $\overline{R}^2$ may rise or fall ($\overline{R}^2$ always increases when the $t - ratio$ of an additional explanatory variable is greater than 1).

1.8. Testing hypotheses about linear restrictions

1.8.1. Testing a single linear restriction which involves one parameter

In the context of eq. (1') we wish to test

$$\begin{aligned} H_0 &: \beta_i = \beta_i^* \text{ , where } \beta_i^* \text{ is some constant} \\ H_1 &: \beta_i \neq \beta_i^*. \end{aligned}$$

The sample evidence on β_i is given by $\widehat{\beta}_i$. The obvious way to measure the distance between the hypothesized value (β_i^*) and the sample evidence $\left(\widehat{\beta}_i\right)$ is the difference $\left(\widehat{\beta}_i - \beta_i^*\right)$. We need to divide the latter by the standard error (s_i) of $\widehat{\beta}_i$, so that our distance measure does not depend on the units of measurement of the data. The resulting test-statistic is called a $t - test$, since it can be shown that, under H_0, it follows a t-distribution with $T - k$ degrees of freedom:

$$t - test = \frac{\widehat{\beta}_i - \beta_i^*}{s_i} \sim t\left(T - k\right).$$

Decision rule*:* if (negative critical value)$<t - test<$(positive critical value), then we accept H_0. (Note that as the degrees of freedom increase the distribution of the $t - test$ approximates the standard normal.). Alternatively, we can use an $F - test$ (see below).

1.8.2. Testing a single linear restriction which involves two parameters

As an example consider that in the context of eq. (1') we wish to test

$$\begin{aligned} H_0 &: \beta_2 + \beta_3 = 0, \\ H_1 &: \beta_2 + \beta_3 \neq 0. \end{aligned}$$

In this case the $t - test$ is

$$t - test = \frac{\widehat{\beta}_2 + \widehat{\beta}_3}{s_{\left(\widehat{\beta}_2 + \widehat{\beta}_3\right)}} = \frac{\widehat{\beta}_2 + \widehat{\beta}_3}{\sqrt{s_2^2 + s_3^2 + s_{2,3}}} \sim t\left(T - k\right),$$

where s_2^2 is the sample variance of $\widehat{\beta}_2$, s_3^2 is the sample variance of $\widehat{\beta}_3$, and $s_{2,3}$ is the sample covariance between $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Alternatively, we can use an $F - test$ (see below).

1.8.3. Testing more than one linear restrictions

In this case we can employ an $F - test$. To carry out this test we must formulate the restricted model, by imposing the linear restrictions on the unrestricted one, and estimate it. Suppose that the maintained or unrestricted model is

$$(UR) : y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t, \; t = 1, ..., T$$

We wish to test the joint significance of the exogenous variables x_{2t}, x_{3t}, i.e.

$$\begin{aligned} H_0 &: \beta_2 = 0, \text{ and } \beta_3 = 0 \\ H_1 &: \beta_2 \neq 0, \beta_3 \neq 0. \end{aligned}$$

The restricted model is

$$(R) : y_t = \beta_0 + \beta_1 x_{1t} + u_t, \; t = 1, ..., T.$$

The $F - test$ is computed as follows:

$$F - test = \frac{\left(\sum_{t=1}^{T} \widehat{u}_t^2 - \sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/2}{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/\left(T - 4\right)} \sim F\left(2, T - 4\right).$$

Generally, the $F - test$ is given by

$$F - test = \frac{(RRSS - URSS)/r}{URSS/(T-k)} \sim F(r, T-k),$$

where $RRSS$ and $URSS$ are the restricted and unrestricted residual sums of squares, respectively; r is the number of restrictions, $(T-k)$ are the degrees of freedom of the UR model.
Decision rule*:* if $F - test$ <critical value, then accept H_0.

1.8.4. The $F - statistic$

The $F-statistic$, reported by most regression programs, tests the joint hypothesis that none of the explanatory variables helps to explain the variation of y about its mean. In other words, it tests the hypothesis that all the slope coefficients of regression (1') are equal to zero:

$$\begin{aligned} H_0 &: \beta_2 = \beta_3 = ... = \beta_k = 0, \\ H_1 &: \beta_2 \neq 0, \text{ and } \beta_3 \neq 0, ..., \text{ and } \beta_k \neq 0. \end{aligned}$$

Therefore, the unrestricted and restricted models (recall that the restricted model is obtained by imposing the restrictions on the unrestricted one) are:

$$\begin{aligned} \text{Unrestricted model } (UR) &: y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \varepsilon_t, \\ \text{Restricted model } (R) &: y_t = \beta_1 + u_t. \end{aligned}$$

We can test the above null hypothesis by using an $F - test$, which we call the $F - statistic$:

$$\begin{aligned} F - statistic &= \frac{\left(\sum_{t=1}^{T} \widehat{u}_t^2 - \sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/(k-1)}{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2\right)/(T-k)} \sim F(k-1, T-k), \\ F - statistic &= \frac{(RRSS - URSS)/(k-1)}{URSS/(T-k)}, \end{aligned}$$

where $URSS$ denotes the unrestricted residual sum of squares, $\sum_{t=1}^{T} \widehat{\varepsilon}_t^2$, and $RRSS$ denotes the restricted residual sum of squares, $\sum_{t=1}^{T} \widehat{u}_t^2$.

Decision rule: if $F - statistic$ <critical value, then accept H_0.
It can be shown that

$$F - statistic = \frac{R_{UR}^2 (T-k)}{(1-R_{UR}^2)(k-1)}.$$

So the $F - statistic$ can be used in the multiple regression model to test the statistical significance of the R^2 statistic.