Econometrics:
Multiple Linear Regression

Burcu Eke

UC3M

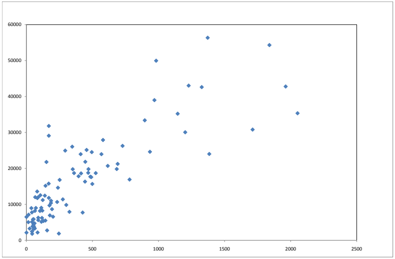
The Multiple Linear Regression Model

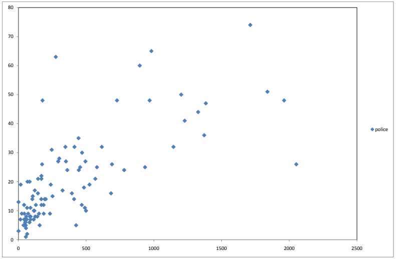
- Many economic problems involve more than one exogenous variable affects the response variable
  - Demand for a product given prices of competing brands, advertising,house hold attributes, etc.
  - Production function

- In SLR, we had $Y = \beta_0 + \beta_1 X_1 + \varepsilon$. Now, we are interested in modeling Y with more variables, such as $X_2, X_3, \ldots, X_k$
  - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$

Example 1: Crimes on campus

Consider the scatter plots: Crime vs. Enrollment and Crime vs. Police





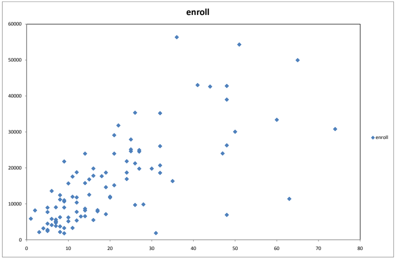
The Multiple Linear Regression Model

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$
  - The slope terms β_j, $j = 1, \ldots k$ are interpreted as the partial effects, or ceteris paribus effects, of a change in the corresponding X_j
  - When the assumptions of MLR are met, this interpretation is correct although the data do not come from an experiment $\Rightarrow$ the MLR reproduces the conditions similar to a controlled experiment (holding other factors constant) in a non-experimental setting.

- X_1 is **not** independent of $X_2, X_3, \ldots, X_k$.

Example 1, cont'd

Consider the scatter plot: Enrollment vs. Police



Example 2: Effect of education on wages

- Let Y be the wage and X_1 be the years of education
- Even though our primary interest is to assess the effects of education, we know that other factors, such as gender (X_2), work experience (X_3), and performance (X_4), can influence the wages
- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$
- Do you think education and the other factors are independent?
- β_1: the partial effect of education, holding experience, gender and performance constant. In SLR, we have no control over these factors since they are part of the error term.

Example 3: Effect of rooms on price of a house

- Suppose we have data on selling prices of apartments in different neighborhoods of Madrid
- We are interested in the effect of number of bedrooms on these prices
- But the number of bathrooms also influence the price
- Moreover, as the number of rooms increase, number of bathrooms may increase as well, i.e., it is hard to see an one-bedroom apartment with two bathrooms.

The Multiple Linear Regression Model: Assumptions

- **A1: Linear in parameters**
- **A2: $\mathbf{E}[\varepsilon|\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_k}] = \mathbf{0}$**
  - For a given combination of independent variables, $(X_1, X_2, \ldots, X_k)$, the average of the error term is equal to 0.
  - This assumption have similar implications as in SLR,
    1. $E[\varepsilon] = 0$
    2. For any function $h()$, $C(h(X_j), \varepsilon) = 0$ for all $j \in \{1, 2, \ldots, k\}$ $\Rightarrow C(X_j, \varepsilon) = 0$ for all $j \in \{1, 2, \ldots, k\}$

The Multiple Linear Regression Model: Assumptions

- **A3: Conditional Homoscedasticity** $\Rightarrow$ Given $X_1, X_2, \ldots, X_k$, variance of ε and/or Y are the same for all observations.
  - $V(\varepsilon | X_1, X_2, \ldots, X_k) = \sigma^2$
  - $V(Y | X_1, X_2, \ldots, X_k) = \sigma^2$

- **A4: No Perfect Multicolinearity**: None of the $X_1, X_2, \ldots, X_k$ can be written as linear combinations of the remaining $X_1, X_2, \ldots, X_k$.

The Multiple Linear Regression Model: Assumptions

- From **A1** and **A2**, we have a conditional expectation function (CEF) such that $E[Y|X_1, X_2, \ldots, X_k] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$
  - Then, CEF gives us the conditional mean of Y for this specific subpopulation
  - Recall that $E[Y|X_1, X_2, \ldots, X_k]$ is the best prediction (achieved via minimizing $E[\varepsilon^2]$)
  - MLR is similar to the SLR in the sense that under the linearity assumption, best prediction and best linear prediction do coincide.

The Multiple Linear Regression Model: Two Variable Case

- Let's consider the MLR model with two independent variables. Our model is of the form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- Recall the housing price example, where Y is the selling price, X_1 is the number of bedrooms, and X_2 is the number of bathrooms

The Multiple Linear Regression Model: Two Variable Case

- Then, we have $E[Y|X_1, X_2] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, hence
  - $E[Y|X_1, X_2 = 0] = \beta_0 + \beta_1 X_1$
  - $E[Y|X_1, X_2 = 1] = (\beta_0 + \beta_2) + \beta_1 X_1$
- This implies that, holding X_1 constant, the change in $E[Y|X_1, X_2 = 0] - E[Y|X_1, X_2 = 1]$ is equal to β_2, or more generally, $\beta_2 \Delta X_2$

The Multiple Linear Regression Model: Interpretation of Coefficients

- More generally, when everything else held constant, a change in X_j results in $\Delta E\left[Y|X_1,\ldots,X_k\right] = \beta_j \Delta X_j$. In other words:
$$\beta_j = \frac{\Delta E\left[Y|X_1,\ldots,X_k\right]}{\Delta X_j}$$
- Then, β_j can be interpreted as follows: When X_j increases by one unit, holding everything else constant, Y, on average, increases by varies, β_j units.
- Do you think the multiple regression of Y on $X_1,\ldots,X_k$ is equivalent to the individual simple regressions of Y on $X_1,\ldots,X_k$ seperately? WHY or WHY NOT?

The Multiple Linear Regression Model: Interpretation of Coefficients

- Recall Example 3. In the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 + \varepsilon$, where X_1 is the number of bedrooms, and X_2 is the number of bathrooms
  - β_1 is the increase in housing prices, on average, for an additional bedroom while holding the number of bathrooms constant, in other worlds, for the apartments with the same number of bathrooms

- If we were to perform a SLR, $Y = \alpha_0 + \alpha_1 X_1 + \varepsilon$, where X_1 is the number of bedrooms
  - α_1 is the increase in housing prices, on average, for an additional bedroom.
  - Notice that in this regression we have no control over number of bathrooms. In other words, we ignore the differences due to the number of bathrooms

- β_j partial regression coefficient

The Multiple Linear Regression Model: Full Model vs Reduced Model

- Recall our model with two independent variables:
  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where
  $E[Y|X_1, X_2] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- In this model, the parameters satisfy:

$$E[\varepsilon] = 0 \quad \Rightarrow \quad \beta_0 = E[Y] - \beta_1 E[X_1] - \beta_2 E[X_2] \tag{1}$$

$$C(X_1, \varepsilon) = 0 \quad \Rightarrow \quad \beta_1 V(X_1) + \beta_2 C(X_1, X_2) = C(X_1, Y) \tag{2}$$

$$C(X_2, \varepsilon) = 0 \quad \Rightarrow \quad \beta_1 C(X_1, X_2) + \beta_2 V(X_2) = C(X_2, Y) \tag{3}$$

The Multiple Linear Regression Model: Full Model vs Reduced Model

- From these equations, we then get

$$\beta_1 = \frac{V(X_2)C(X_1,Y) - C(X_1,X_2)C(X_2,Y)}{V(X_1)V(X_2) - [C(X_1,X_2)]^2}$$
$$\beta_2 = \frac{V(X_1)C(X_2,Y) - C(X_1,X_2)C(X_1,Y)}{V(X_1)V(X_2) - [C(X_1,X_2)]^2}$$

- Notice that if $C(X_1, X_2) = 0$, these parameters are equivalent to SLR of Y on X_1 and X_2, respectively, i.e.,
  - $\beta_1 = \frac{C(X_1,Y)}{V(X_1)}$ and $\beta_2 = \frac{C(X_2,Y)}{V(X_2)}$

The Multiple Linear Regression Model: Full Model vs Reduced Model

- Assume we are interested in the effect of X_1 on Y, and concerned that X_1 and X_2 may be correlated. Then a SLR does **not** give us the effect we want.

- Define the model $L(Y|X_1) = \alpha_0 + \alpha_1 X_1$ as the **reduced model** (We also could have considered reduced model with X_2 as our independent variable, results would have been the same). Then α_0 and α_1 satisfy the following equations:

$$E[\varepsilon] = 0 \quad \Rightarrow \quad \alpha_0 = E[Y] - \alpha_1 E[X_1] \tag{4}$$

$$C(X_1, \varepsilon) = 0 \quad \Rightarrow \quad \alpha_1 = \frac{C(X_1, Y)}{V(X_1)} \tag{5}$$

The Multiple Linear Regression Model: Full Model vs Reduced Model

- Using equations (2) and (5) we get:

$$\alpha_1 = \frac{C(X_1, Y)}{V(X_1)} = \frac{\beta_1 V(X_1) + \beta_2 C(X_1, X_2)}{V(X_1)} = \beta_1 + \beta_2 \frac{C(X_1, X_2)}{V(X_1)}$$

- Notice the following:
  1. $\alpha_1 = \beta_1$ **only if** when $C(X_1, X_2) = 0$ OR when $\beta_2 = 0$
  2. $\dfrac{C(X_1, X_2)}{V(X_1)}$ is the slope for the prediction $L(X_2|X_1) = \gamma_0 + \gamma_1 X_1$

Review

- Our model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$
  - We want X_1 and $X_2, \ldots, X_k$ NOT to be independent
- Assumptions:
  - **A1** to **A3** $\Rightarrow$ similar to SLR
  - **A4**: No perfect multicollinearity
    - For example, we don't have a case like $X_{1i} = 2X_{2i} + 0.5X_{3i}$ for all $i = 1, \ldots, n$

Review

- Interpretation

Model	Coefficient of X_1
$Y_i = \alpha_0 + \alpha_1 X_{1i} + \varepsilon_i$	For one unit increase in X_1, Y, on average, increases by α_1 units
$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$	For one unit increase in X_1, holding everything else constant, Y, on average, increases by β_1 units

Review

- Full Model vs Reduced Model - focusing on two variable case
  - Full Model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$
  - Reduced Model: $Y_i = \alpha_0 + \alpha_1 X_{1i} + \varepsilon_i$
  - Reduced model cannot control for X_2, cannot hold it constant, and see the effects of X_1, holding X_2 constant

- $\alpha_1 = \dfrac{C(X_1, Y)}{V(X_1)} = \dfrac{\beta_1 V(X_1) + \beta_2 C(X_1, X_2)}{V(X_1)} =$
  $\beta_1 + \beta_2 \dfrac{C(X_1, X_2)}{V(X_1)}$
  - α_1 reflects the effect of X_1, holding X_2 constant, plus $\beta_2 \dfrac{C(X_1, X_2)}{V(X_1)}$

Review: An Example

- Model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, where Y is the University GPA, X_1 is the high school GPA, and X_2 is the total SAT score.

	Coefficient	Std. Error	t-ratio	p-value
const	−0.0881312	0.286664	−0.3074	0.7592
HighSGPA	0.407113	0.0905946	4.4938	0.0000
SATtotal	0.00121666	0.000301086	4.0409	0.0001

Review: An Example

	Coefficient	Std. Error	t-ratio	p-value
const	0.822036	0.190689	4.3109	0.0000
HighSGPA	0.565491	0.0878337	6.4382	0.0000

	Coefficient	Std. Error	t-ratio	p-value
const	0.151892	0.307993	0.4932	0.6230
SATtotal	0.00180201	0.000296847	6.0705	0.0000

Review: An Example

Table : Dependent variable: HighSGPA

	Coefficient	Std. Error	t-ratio	p-value
const	0.589574	0.314040	1.8774	0.0634
SATtotal	0.00143780	0.000302675	4.7503	0.0000

MULTIPLE LINEAR REGRESSION: THE ESTIMATION

The Multiple Linear Regression Model: Estimation

- Our Objective: Estimate the population parameters, i.e., $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$ and σ^2
- Our Model: $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$, where $E[\varepsilon|X] = 0$ and $V(\varepsilon|X) = \sigma^2$
- Our Data: Pairs of $(y_i, x_{1i}, x_{2i}, \ldots, x_{ki})$ $i = 1, 2, ..., n$ are a **random sample** of population data $(Y_i, X_{1i}, \ldots, X_{ki})$:

y	x_1	$\ldots$	x_k
y_1	x_{11}	$\ldots$	x_{k1}
y_2	x_{12}	$\ldots$	x_{k2}
.	.	.	.
.	.	.	.
.	.	.	.
y_n	x_{1n}	$\ldots$	x_{kn}

The Multiple Linear Regression Model: Estimation

- Given our model and the sample data, we can write our model as: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$, for $i = 1, \ldots, n$
- Our model satisfies our assumptions **A1-A4**

OLS for The Multiple Linear Regression Model

- We obtain the estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ by solving the equation below

$$\min_{\beta_0,\ldots,\beta_k} \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \text{ where } \hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}$$

OLS for The Multiple Linear Regression Model

- The first-order conditions are: $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$, $\sum_{i=1}^{n} \hat{\varepsilon}_i X_{1i} = 0$, $\sum_{i=1}^{n} \hat{\varepsilon}_i X_{2i} = 0' \ldots$, $\sum_{i=1}^{n} \hat{\varepsilon}_i X_{ki} = 0$
- Or equivalently, by using $x_{ji} = (X_{ji} - \bar{X}_j)$ (where $j = 1, \ldots, k$ and $i = 1, \ldots, n$) the FOC can be expressed as: $\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i = 0$, and

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i x_{1i} &= 0 \\
\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i x_{2i} &= 0 \\
\ldots \quad \ldots & \quad \ldots \\
\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i x_{ki} &= 0
\end{aligned}
$$

- Notice that (1) implies that sample mean of the residuals is zero, and the rest imply that the sample covariance between the residuals and X_j's are equal to 0.

OLS for The Multiple Linear Regression Model

- Notice that these first order conditions are sample analogue of the first-order conditions for classical regression model with population β's: $E[\varepsilon] = 0$

$$\begin{aligned} C\left(X_1, \varepsilon\right) &= 0 \\ C\left(X_2, \varepsilon\right) &= 0 \\ \dots \quad \dots & \quad \dots \\ C\left(X_k, \varepsilon\right) &= 0 \end{aligned}$$

OLS for The Multiple Linear Regression Model

- The system of equations, also known as normal equations, are

$$
\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1 \sum_i X_{1i} + \hat{\beta}_2 \sum_i X_{2i} + \ldots + \hat{\beta}_k \sum_i X_{ki} &= \sum_i Y_i \\
\hat{\beta}_1 \sum_i x_{1i}^2 + \hat{\beta}_2 \sum_i x_{1i}x_{2i} + \ldots + \hat{\beta}_k \sum_i x_{1i}x_{ki} &= \sum_i x_{1i}Y_i \\
\hat{\beta}_1 \sum_i x_{1i}x_{2i} + \hat{\beta}_2 \sum_i x_{2i}^2 + \ldots + \hat{\beta}_k \sum_i x_{2i}x_{ki} &= \sum_i x_{2i}Y_i \\
\ldots \quad & \ldots \quad \ldots \\
\hat{\beta}_1 \sum_i x_{1i}x_{ki} + \hat{\beta}_2 \sum_i x_{ki}x_{2i} + \ldots + \hat{\beta}_k \sum_i x_{ki}^2 &= \sum_i x_{ki}Y_i
\end{aligned}
$$

OLS for The Multiple Linear Regression Model

- In general, for k variables, we will have a linear system with $k+1$ equations with unknowns $k+1$: $(\beta_0, \beta_1, \ldots, \beta_k)$
- This system of linear equations will have a **unique** solution only if **A4** holds $\Rightarrow$ no perfect multicollinearity
- If **A4** does NOT hold, then we will have infinitely many solutions

OLS for The Multiple Linear Regression Model: Two Variable Case

- Let's consider model with two variables:
  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

- Then the population parameters will have the following form:

$$
\begin{aligned}
\beta_0 &= E[Y] - \beta_1 E[X_1] - \beta_2 E[X_2] \\
\beta_1 &= \frac{V(X_2)C(X_1, Y) - C(X_1, X_2)C(X_2, Y)}{V(X_1)V(X_2) - [C(X_1, X_2)]^2} \\
\beta_2 &= \frac{V(X_1)C(X_2, Y) - C(X_1, X_2)C(X_1, Y)}{V(X_1)V(X_2) - [C(X_1, X_2)]^2}
\end{aligned}
$$

OLS for The Multiple Linear Regression Model: Two Variable Case

- When we apply the analogy principle, we get:

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{S^2_{X_2} S_{X_1,Y} - S_{X_1,X_2} S_{X_2,Y}}{S^2_{X_1} S^2_{X_2} - (S_{X_1,X_2})^2}$$

$$\hat{\beta}_2 = \frac{S^2_{X_1} S_{X_2,Y} - S_{X_1,X_2} S_{X_1,Y}}{S^2_{X_1} S^2_{X_2} - (S_{X_1,X_2})^2}$$

OLS for The Multiple Linear Regression Model: Two Variable Case

- where

$$S_{X_1}^2 = \frac{1}{n}\sum_i (X_{1i} - \bar{X}_1)^2 \qquad S_{X_2}^2 = \frac{1}{n}\sum_i (X_{2i} - \bar{X}_2)^2$$

$$S_{X_1,X_2} = \frac{1}{n}\sum_i (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)$$

$$S_{X_1,Y} = \frac{1}{n}\sum_i (X_{1i} - \bar{X}_1)(Y_i - \bar{Y})$$

$$S_{X_2,Y} = \frac{1}{n}\sum_i (X_{2i} - \bar{X}_2)(Y_i - \bar{Y})$$

- The slope estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ measure the partial effect of X_1 and X_2 on the mean value of Y, respectively.

Properties of the OLS Estimators

The properties of OLS estimators under multiple linear regression model is similar to those under simple linear regression model. The verification of these properties is similar to SLR case.

1. OLS estimators are linear in Y
2. OLS estimators are unbiased (Expected value of the estimator is equal to the parameter itself) under **A1**, **A2**, and **A4**.
3. Under assumptions **A1** to **A4**, the Gauss-Markov Theorem states that the $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ have the minimum variance among all linear unbiased estimators
4. OLS estimators are consistent

Properties of the OLS Estimators: The Variance

- Using **A1, A2** and **A3**, we have:
- $V(\hat{\beta}_j) = \dfrac{\sigma^2}{nS_j^2\left(1 - R_j^2\right)} = \dfrac{\sigma^2}{\sum_i x_{ji}^2\left(1 - R_j^2\right)}$, where $j = 1, \ldots, k$
  - $S_j^2 = \dfrac{1}{n}\sum_i x_{ji}^2 = \dfrac{1}{n}\sum_i \left(X_{ji} - \bar{X}_j\right)^2$
  - R_j^2 is the coefficient of determination for the linear projection of X_j on all other independent variables: $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$ such that:

$$X_j = \theta_0 + \theta_1 X_1 + \ldots + \theta_{j-1} X_{j-1} + \theta_{j+1} X_{j+1} + \ldots + \theta_k X_k + u$$

Properties of the OLS Estimators: The Variance

- R_j^2 measures the proportion of information of variable X_j that is already contained in other variables.

- Them $\left(1 - R_j^2\right)$ measures the proportion of information of variable X_j that is NOT explained by any other variable.
  - $R_j^2 = 1$ is NOT possible, because it would imply that X_j is an exact linear combination of the other independent variables. This is ruled out by **A4**
  - If R_j^2 is close to 1, however, $V(\hat{\beta}_j)$ would be very high.
  - On the contrary, if $R_j^2 = 0$, then this implies that the correlation of X_j with the other independent variables is 0. In this case, then $V(\hat{\beta}_j)$ would be minimum.

Properties of the OLS Estimators: The Variance

- Intuitively,
  - When S_j^2 is higher $\Rightarrow$ Sample variance of X_j is higher AND $V(\hat{\beta}_j)$ would be lower $\Rightarrow$ more precise estimator
  - When the sample size, n, increases, $V(\hat{\beta}_j)$ will be smaller $\Rightarrow$ more precise estimator
  - When R_j^2 is higher, $V(\hat{\beta}_j)$ will be bigger $\Rightarrow$ less precise estimator

- For proof, see Wooldridge (Since the order of the explanatory variables is arbitrary, the proof for $\hat{\beta}_1$ can be extended to $\hat{\beta}_2, \ldots, \hat{\beta}_k$, without loss of generality)

Estimation of the Variance

- Estimation of σ^2 is similar to the case in simple linear regression.
- One attempt to estimate σ^2 would be to use $(1\backslash n)\sum_i \varepsilon_i^2$ instead of $E[\varepsilon^2]$. However, we do not know the population values for ε
- We could use sample values of the error, $\hat{\varepsilon}$ where

$$\hat{\varepsilon}_i = Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}\right)$$

- We can use $\hat{\varepsilon}_i$'s to estimate the sample variance:

$$\tilde{\sigma}^2 = (1\backslash n)\sum_i \hat{\varepsilon}_i^2$$

The Variance of The OLS Estimators

- However, $\tilde{\sigma}^2$ is a biased estimator, because the residuals satisfy $(k+1)$ linear restrictions:

$$\sum_i \hat{\varepsilon}^2 = 0, \sum_i \hat{\varepsilon}^2 X_{1i} = 0, \ldots, \sum_i \hat{\varepsilon}^2 X_{ki} = 0$$

  Thus the degrees of freedom left for residuals is $(n-(k+1))$

- When we account for this bias, we get an **unbiased** estimate:

$$\hat{\sigma}^2 = \sum_i \frac{\varepsilon_i^2}{n-k-1}$$

- Notice that both of $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ are consistent and for larger samples, their value is very close to each other.

Estimation of the Variance

- Recall that $V(\hat{\beta}_j) = \dfrac{\sigma^2}{nS_j^2\left(1 - R_j^2\right)}$

- We estimate $V(\hat{\beta}_j)$ by:

$$\hat{V}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{nS_j^2\left(1 - R_j^2\right)} \quad \text{where}$$
$$S_j^2 = \frac{1}{n}\sum_i x_{ji}^2 = \frac{1}{n}\sum_i \left(X_{ji} - \bar{X}_j\right)^2$$

Goodness of Fit Measures

- One GOF measure is **the standard error of the regression**:

$$\sqrt{\hat{\sigma}^2} = \hat{\sigma} = \sqrt{\sum_i \frac{\varepsilon_i^2}{n-k-1}}$$

Goodness of Fit Measures

- R^2, the **coefficient of determination**, is another measure of goodness of fit, similar to SLR case.

$$R^2 = \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2} = 1 - \frac{\sum_i \varepsilon_i^2}{\sum_i y_i^2},\ 0 \leq R^2 \leq 1$$

  - where $y_i = Y_i - \bar{Y}$, $\hat{y}_i = \hat{Y}_i - \bar{Y}$, and $\varepsilon_i = Y_i - \hat{Y}_i$

- The interpretation of R^2 is similar to the SLR model. It measures the proportion of the variability of Y explained by our model.

Goodness of Fit Measures

- The R^2 can be used to compare different models with the same dependent (endogenous) variable Y
- The R^2 weakly increases as the number of independent variables increases, i.e., by adding one more variable, we get $R^2_{new} \geq R^2_{old}$
- Hence, when comparing models with the same dependent variable Y, but with different number of independent variables, R^2 is not a good measure

Goodness of Fit Measures

- Adjusted-R^2, $\bar{R}^2$, is a goodness of fit measure that adjusts for the different degrees of freedom, i.e., for adding more variables in the model
- It is calculated as:

$$\bar{R}^2 = 1 - \left[\left(1 - R^2\right) \frac{n-1}{n-k-1} \right] = 1 - \frac{\frac{\sum_i \hat{\varepsilon}_i^2}{n-k-1}}{\frac{\sum_i y_i^2}{n-1}}$$

- What happens to the relationship between $\bar{R}^2$ and R^2 as sample size increases? What about if we have relatively small sample size with big number of independent variables?

Goodness of Fit Measures: Example

Recall the University GPA example

Model	R^2	$\bar{R}^2$
$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X_1$	0.297241	0.290070
$\hat{Y} = \hat{\delta}_0 + \hat{\delta}_2 X_2$	0.273272	0.265857
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$	0.398497	0.386095

Interpretation of Coefficients: A Review

Model	β_j
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$	For one unit increase in X_j, holding everything else constant, Y, on average, increases by $\hat{\beta}_j$ units
$\widehat{\ln Y_i} = \hat{\beta}_0 + \hat{\beta}_1 \ln X_{1i} + \ldots + \hat{\beta}_k \ln X_{ki}$	For one percent (%) increase in X_j, holding everything els constant, Y, on average, increases by $\hat{\beta}_j$ percents (%)
$\widehat{\ln Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$	For one unit increase in X_j, holding everything else constant, Y, on average, increases by $(\hat{\beta}_j * 100)\%$

Interpretation of Coefficients: A Review

Model	β_j
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \ln X_{1i} + \ldots + \hat{\beta}_k \ln X_{ki}$	For one percent (%) increase in X_j, holding everything else constant, Y, on average, increases by $\frac{\hat{\beta}_j}{100}$ units
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{X_{1i}}$	For one unit increase in X_j, Y, on average, increases by $-\hat{\beta}_1 \frac{1}{X_1^2}$ units

Interpretation of Coefficients: A Review

Model	β_j
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{1i}^2$	For one unit increase in X_j, Y, on average, increases by $\hat{\beta}_1 + 2\hat{\beta}_2 X_1$ units
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ $+\hat{\beta}_3 X_{1i} X_{2i}$	For one unit increase in X_j, holding everything else constant, Y, on average, increases by $\hat{\beta}_1 + \hat{\beta}_3 X_2$ units

Inferences in the Linear Regression Models: Hypothesis Tests

- Wooldridge: Chapters 4 and 5 (5.2) OR Goldberger: Chapters 7, 10 (10.3), 11, and 12 (12.5 & 12.6).

- A **hypothesis test** is a statistical inference technique that assesses whether the information provided by the data (sample) supports or does not support a particular conjecture or assumptions about the population.

- In general, there are two types of statistical hypotheses:
  1. **Nonparametric** hypotheses are about the properties of the population distribution (independent observations, normality, symmetry, etc.).
  2. **Parametric** hypotheses are about conditions or restrictions of population parameter values.

Inferences in the Linear Regression Models: Hypothesis Tests

- **Null hypothesis** is a hypothesis to be tested, denoted by H_0
- **Alternate hypothesis** is a hypothesis to be considered as an alternate to the null hypothesis, denoted by H_1, or H_a
- **Hypothesis test** is to decide whether the null hypothesis should be rejected in favor of the alternative hypothesis.
- The classical approach to hypothesis testing, based on Neyman-Pearson, divides the sample space given H_0 in two regions:
  - Rejection (Critical) region and Do not reject (DNR) region
  - If the observed data fall into the rejection region, the null hypothesis is rejected

Inferences in the Linear Regression Models: Hypothesis Tests

Steps to perform Hypothesis test is:

1. Define H_0 and H_1
2. Define a **test statistic** that measures the discrepancy between the sample data and the null hypothesis H_0. This statistic:
   - is a fuction of H_0 and the sample data, that is, the test statistics is a random variable
   - must have a known distribution (exact or approximate) under H_0 (in case H_0 is true)
   - If the discrepancy between the sample data and H_0 is "big", the value of the test statistic will be within a range of values unlikely under H_0 $\Rightarrow$ evidence against H_0
   - If the discrepancy between the sample data and H_0 is "small", the value of the test statistic will be within a range of values unlikely under H_0 $\Rightarrow$ evidence for H_0

Inferences in the Linear Regression Models: Hypothesis Tests

3. Determine the "major discrepancies", ie, the critical (rejection) region. This region is defined by a critical value, given the distribution of the test statistic.
4. Given the sample data, calculate the value of the test statistic and check if you are in the rejection region

Inferences in the Linear Regression Models: Hypothesis Tests

- Since the sample data used in the test is random, the test statistic, which is a function of the sample data, is also a random variable.
  - Therefore, the test statistic may lead to different conclusions for different samples.
  - When the hypothesis test is carried out, we conclude either in favor of H_0 or H_1, and we will be in one of four situations:

	H_0 is True	H_0 is False
Do not reject (DNR) H_0	Correct Decision	Type II error
Reject (R) H_0	Type I error	Correct Decision

Inferences in the Linear Regression Models: Hypothesis Tests

- We define the significance level as
  $\alpha = P(\text{Type I error}) = P\left(\text{Reject } H_0 | H_0 \text{ is true}\right)$

- We would like to minimize both types of errors, but it is not possible to minimize the probability of both types of error for a given sample size.
  - The only way to reduce the probability of both errors is by increasing the sample size

Inferences in the Linear Regression Models: Hypothesis Tests

- The classical procedure is as follows:
- Set α (significance level of the test), ie, establish the maximum probability of rejecting H_0 when it is true
  - we set a value as small as desired: usual values are 10%, 5%, or 1%
- Minimize
  $P(\text{Type II error}) = P\,(\text{DNR Reject } H_0 | H_0 \text{ is false})$ or, equivalently, maximize $1 - P(\text{Type II error})$ =Power of the test
- Since power is defined under the alternative hypothesis, its value depends on the true value of the parameter (which is unknown).

Inferences in the Linear Regression Models: Hypothesis Tests

- A hypothesis test is **consistent** if
  $\lim_{n\to\infty}(1 - P(\text{type II error})) = 1$

- Probabilities of Type I and Type II Errors:
  - Significance level α: The probability of making a Type I error (rejecting a true null hypothesis)
  - Power of a hypothesis test: Power = 1-P(Type II error) = 1-β
    - The Probability of rejecting a false null hypothesis
    - Power near 0: the hypothesis test is not good at detecting a false null hypothesis.
    - Power near 1: the hypothesis is extremely good at detecting a false null hypothesis

Inferences in the Linear Regression Models: Hypothesis Tests

- Relation between Type I and Type II error probabilities: For a fixed sample size, the smaller we specify the significance level, α, the larger will be the probability, β, of not rejecting a false null hypothesis

- Since we don't know β, we don't know the probability of not rejecting a false null hypothesis. Thus when we DNR H_0, we **never say** "the data support H_0". We always say "the data do not provide sufficient evidence support H_a"

- If we reject H_0 at significance level α, we say our results are "statistically significant at level α". Likewise, if we do not reject H_0 at significance level α, we say our results are "not statistically significant at level α".

Inferences in the Linear Regression Models: Hypothesis Tests

- For a given H_0
  1. A test statistic, C, is defined
  2. A critical region is defined using the distribution of C under H_0, given a significance level α
  3. The test statistic is calculated using sample data, $\hat{C}$
  4. If the value of test statistic $\hat{C}$ is within the critical region, we reject H_0 with a significance level α; otherwise, we do not reject (DNR) H_0

Inferences in the Linear Regression Models: Hypothesis Tests

- **Test statistic** is the statistic used as a basis for deciding whether the null hypothesis should be rejected
- **Rejection region** is the set of values for the test statistic that leads to rejection of the null hypothesis
- **Non-rejection region** is the set of values for the test statistic that leads to nonrejection of the null hypothesis
- **Critical values** are the values of the test statistic that separate the rejection and nonrejection regions

Inferences in the Linear Regression Models: Hypothesis Tests

- Instead of considering critical value to decide whether to reject the null hypothesis, we can consider $p-$value.
- Assuming H_0 is true, the $p-$value indicates how likely it is to observe the test statistic, i.e.,
$p-$value$= P\left(\hat{C} \in \{ \text{ critical region } \} \mid H_0\right)$
- Can be interpreted as the probability of observing a value at least as extreme as the test statistic.

Inferences in the Linear Regression Models: Hypothesis Tests

- Corresponding to an observed value of a test statistic, the **p-value** is the lowest level of significance at which the null hypothesis can be rejected.
- Decision criterion for p-value approach to hypothesis tests:
  - if $p-$value$\leq \alpha$, reject H_0
  - if $p-$value$> \alpha$, do NOT reject H_0

Hypothesis Tests: An Example, Test for Population Mean

- Given a random variable, Y, we can test whether its mean, $E[Y] = \mu$ is equal to some constant quantity μ_0: $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$

- To test this hypothesis, we have data on Y from a sample of size n, $\{Y_i\}_{i=1}^{n}$, where Y_i's are independent. Then we can estimate the sample mean as:

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Hypothesis Tests: An Example, Test for Population Mean

- The expected value and variance of $\bar{Y}$ are:

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \mu$$

$$V(\bar{Y}) = V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i) = \frac{\sigma^2}{n}$$

Hypothesis Tests: An Example, Test for Population Mean

- Suppose $Y \sim N(\mu, \sigma^2)$, and σ^2 is known. Then:
  - Since Y is normally distributed, the so are the sample observations of Y, Y_i.
  - Since $\bar{Y}$ is a linear combination of Y_i's, which are normally distributed, $\bar{Y}$, is also normally distributed:
    $\bar{Y} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$
  - Under H_0, $\bar{Y} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$
  - In this case, our test statistic is defined as: $\hat{C} = \dfrac{\bar{Y} - \mu_0}{\sigma^2/n})$
  - Under H_0, $\hat{C} \sim N(0, 1)$

Hypothesis Tests: An Example, Test for Population Mean

- The test aims to assess whether the discrepancy, measured by the test statistic, is statistically "large" in absolute value (we are interested in the magnitude of the discrepancy, not the direction)
- We apply the classic procedure by determining a level of significance: α
- In this case, we have a two-tailed test, with respective significant levels of $\frac{\alpha}{2}$ in each tail. The rejection region corresponds to extreme differences, whether positive or negative.

Hypothesis Tests: An Example, Test for Population Mean

- How to we determine rejection region?
- How do we find the $p-$value?
- What happens when σ^2 is **unknown**?

Hypothesis Tests in Regression Setting: Sampling Distributions of OLS Estimators

- In the MLR (also in SLR), we are interested in hypothesis tests on parameters $\beta_0, \beta_1, \ldots, \beta_k$
- To do this, we need:
  - A test statistic and its distribution
- We have the OLS estimators of the parameters and their properties. However, to make inferences we must characterize the sampling distribution of these estimators.

Hypothesis Tests in Regression Setting: Sampling Distributions of OLS Estimators

- To derive exact distribution of $\hat{\beta}_j$, we would need to assume that Y, given all X_j's is normally distributed, i.e.,

$$Y_i | X1i, \ldots, X_{ki} \sim N\left(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}, \sigma^2\right)$$

$$\Longleftrightarrow \varepsilon_i | X_{1i}, \ldots, X_{ki} \sim N\left(0, \sigma^2\right)$$

- In this case it is possible to show that $\hat{\beta}_j$'s follow normal distribution $\Rightarrow \hat{\beta}_j \sim N\left(\beta_j, V(\hat{\beta}_j)\right)$

- In general, this assumption is not verifiable, and it is difficult to be met. In that case, the distribution of the estimators $\hat{\beta}_j$ will be unknown.

Hypothesis Tests in Regression Setting: Sampling Distributions of OLS Estimators

- Even if normality assumption is not met, we can base our inferences on the asymptotic distribution
- So, we will use the asymptotic distribution of $\hat{}_j$'s
- It can be proven that $\hat{\beta}_j \overset{\text{asy}}{\sim} N\left(\beta_j, V(\hat{\beta}_j)\right)$. This implies that

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \overset{\text{asy}}{\sim} N(0, 1)$$

Hypothesis Tests in Regression Setting: Sampling Distributions of OLS Estimators

- Substituting $V(\hat{\beta}_j)$ by a consistent estimator $\hat{V}(\hat{\beta}_j) = \dfrac{\hat{\sigma}^2}{nS_j^2(1 - R_j^2)}$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \overset{\text{asy}}{\sim} N(0, 1)$$

Hypothesis Test of the Value of a Parameter

- We are interested in tests on β_j's: $H_0 : \beta_j = \beta_j^0$ against the alternative $H_1 : \beta_j \neq \beta_j^0$, where β_j^0 is a constant number

- Since σ^2 is unknown, we employ a $t-$test, with the following test statistic:

$$t = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\hat{V}(\hat{\beta}_j)}} = \frac{\hat{\beta}_j - \beta_j^0}{s_{\hat{\beta}_j}}$$

- Under H_0, this test statistic asymptotically follows $N(0,1)$

Hypothesis Test of the Value of a Parameter

- Almost all econometric programs and/or packages present the estimated coefficients with the corresponding standard errors, $t-$statistic and the $p-$value associated with the $t-$statistic.

- In general, these programs calculate the $p-$value based on the $t-$distribution with $(n-k-1)$ degrees of freedom.

- For relatively large values of n, the critical values for the $t-$ distribution, and the asymptotic distribution are very similar.
  - For example, for a two-tailed test with significance level α, with $(n-k-1)$ degrees of freedom, $|t| > t_{1-\alpha/2}$, is:
    - For $\alpha = 0.05$, the critical value is 1.98 for the $t-$distribution and 1.96 for the $N(0,1)$
    - For $\alpha = 0.10$, the critical values is 1.658 for the $t-$distribution and 1.645 for the $N(0,1)$

Hypothesis Test of the Value of a Parameter

- Therefore, we can base our rejection decision on the $p-$values shown in the output tables of econometric programs when n is relatively large.

- Also, using the output on the parameters, we can construct approximate confidence intervals.

Hypothesis Tests in Regression Setting: Example

- Demand for money for USA and economic activity (Goldberger, p. 107). (data file: TIM1.GDT)
- Model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where
  - $Y = \ln(100 * V4/V3)$ =log of real quantity demanded of money (M1)
  - $X_1 = \ln(V2)$ =log of real GDP
  - $X + 2 = \ln(V6)$ =log of the interest rate for treasury bill

Hypothesis Tests in Regression Setting: Example

	Coefficient	Std. Error	t-ratio	p-value
const	2.32959	0.205437	11.3397	0.0000
X1	0.557290	0.0263835	21.1227	0.0000
X2	−0.203154	0.0210047	−9.6719	0.0000

Mean dependent var	6.628638	S.D. dependent var	0.172887
Sum squared resid	0.080411	S.E. of regression	0.047932
R^2	0.927291	Adjusted R^2	0.923137
$F(2, 35)$	223.1866	P-value(F)	1.20e–20
Log-likelihood	63.08600	Akaike criterion	−120.1720
Schwarz criterion	−115.2592	Hannan–Quinn	−118.4241
$\hat{\rho}$	0.627702	Durbin–Watson	0.727502

Hypothesis Tests in Regression Setting: Example

- Interpretation:
  - $\hat{\beta}_1$ is the estimator of the elasticity of money demand with respect to the GDP, holding the interest rate constant. If the GDP increases by 1%, holding the interest rate constant, on average, the demand for money grows by 0.6%
  - $\hat{\beta}_2$ is the estimator of the elasticity of money demand with respect to the interest rate, holding GDP constant. If the interest rate increases by 1% holding the GDP constant, on average, the demand for money grows by 0.2%

Hypothesis Tests in Regression Setting: Example

- We would like to test the following two hypotheses:
  - Money demand is inelastic to the interest rate
  - The elasticity of money demand with respect to the GDP is 1

- $H_0 : \beta_2 = 0$ (money demand w.r.t. interest rate is 0) versus $H_1 : \beta_2 \neq 0$. Then, under H_0, $t = \dfrac{\hat{\beta}_2 - 0}{s_{\hat{\beta}_2}} \overset{asy}{\sim} N(0,1)$ and

  $$|t| = \left| \frac{-0.2032}{0.021} \right| = 9.676 > Z^* = 1.96$$

  - Note that $|t| \approx 9.7 \Rightarrow$ $p-$value is almost 0 (for a Z = 3.09, the p-value for a normal distribution is 0.001)
  - $\Rightarrow$ we reject H_0 at a significance level of 1%?
  - **A** 95% confidence interval for β_2 is:

    $$\hat{\beta}_2 \pm s_{\hat{\beta}_2} * 1.96 \Rightarrow -0.2032 \pm 0.021 * 1.96 \Rightarrow [-0.244, -$$

Hypothesis Tests in Regression Setting: Example

- $H_0 : \beta_1 = 1$ (elasticity of money demand w.r.t. GDP is 1) versus $H_1 : \beta_1 \neq 1$. Then, under H_0, $t = \dfrac{\hat{\beta}_1 - 1}{s_{\hat{\beta}_1}} \overset{\text{asy}}{\sim} N(0,1)$ and $|t| = \left| \dfrac{0.5573 - 1}{0.0264} \right| = 16.769 > Z^* = 1.96$
  - Note that $|t| \approx 16.8 \Rightarrow$ $p-$value is almost 0
  - $\Rightarrow$ we reject H_0 at a significance level of 1%?
  - **A** 95% confidence interval for β_1 is:

$$\hat{\beta}_1 \pm s_{\hat{\beta}_1} * 1.96 \Rightarrow 0.5573 \pm 0.0264 * 1.96 \Rightarrow [0.505, 0.609]$$

Hypothesis Tests in Regression Setting: Linear Hypothesis Tests

- Let's Consider the null hypothesis of the following form:
  - $H_0 : \lambda_0\beta_0 + \lambda_1\beta_1 + \ldots + \lambda_k\beta_k = \mu^0$

- To test this hypothesis, we construct our test statistic as

$$t = \frac{\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k - \mu^0}{\sqrt{\hat{V}\left(\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k\right)}} = \frac{\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k - \mu^0}{s_{\left(\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k\right)}}$$

- Using the asymptotic approximation we have that under H0: $t = \frac{\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k - \mu^0}{s_{\left(\lambda_0\hat{\beta}_0 + \lambda_1\hat{\beta}_1 + \ldots + \lambda_k\hat{\beta}_k\right)}} \overset{asy}{\sim} N(0,1)$

Linear Hypothesis Test: An Example

- Example: Production technology is estimated as follows:

$$\begin{array}{lllll} \hat{Y} = 2.37+ & 0.632X_1 & + & 0.452X_2 & n = 31 \\ & (0.257) & & (0.219) & \\ s_{\hat{\beta}_1,\hat{\beta}_2} & \hat{C}(\hat{\beta}_1, \hat{\beta}_2) & = 0.055 & & \end{array}$$

  - Y =logarithm of output
  - X_1 =logarithm of labor
  - X_2 =logarithm of capital

Linear Hypothesis Test: An Example

- Interpretation:
  - $\hat{\beta}_1$ is the estimator of the elasticity of output with respect to labor, holding capital constant. If the labor increases by 1%, holding the capital constant, on average, the output grows by 0.63%
  - $\hat{\beta}_2$ is the estimator of the elasticity of output with respect to capital, holding labor constant. If the capital increases by 1% holding the labor constant, on average, the output grows by 0.45%

Linear Hypothesis Test: An Example

- Consider $H_0 : \beta_1 + \beta_2 = 1$ (Constant returns to scale) versus $H_1 : \beta_1 + \beta_2 \neq 1$
- Under H_0,

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{s_{\hat{\beta}_1 + \hat{\beta}_2}} \overset{asy}{\sim} N(0, 1)$$

  with

  $s_{\hat{\beta}_1 + \hat{\beta}_2} = \sqrt{\hat{V}(\hat{\beta}_1 + \hat{\beta}_2)} = \sqrt{\hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\beta}_2) + 2\hat{C}(\hat{\beta}_1 + \hat{\beta}_2)}$
- $|t| = \left| \frac{0.632 + 0.452 - 1}{\sqrt{(0.257)^2 + (0.219)^2 + 2 * 0.055}} \right| = 0.177 < Z^* = 1.96$
- Therefore, do not reject the null hypothesis of constant returns to scale.

Tests with Multiple Linear Constraints

- How can you compare more than one hypothesis about the parameters of the model?
- For example:
  - $H_0 : \beta_1 + \beta_2 + \ldots + \beta_k = 0$
  - Or,
    $$H_0 : \begin{aligned} \beta_1 &= 0 \\ \beta_2 &= 1 \end{aligned}$$
  - Or,
    $$H_0 : \begin{aligned} \beta_1 + \beta_3 &= 0 \\ \beta_2 &= -1 \\ \beta_4 - 2\beta_5 &= 1 \end{aligned}$$

Tests with Multiple Linear Constraints

- Previous concepts:
  - **Unrestricted** model is the model on which you want to make a hypothesis testing (under H_1).
  - **Restricted** model is the model that has imposed the linear constraint(s) under H_0
  - Example 1: $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1 : \beta_1 \neq 0$ and/or $\beta_2 = 0$

Unrestricted model	Restricted model
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	$Y = \beta_0 + \varepsilon$

Tests with Multiple Linear Constraints

- Example 2: $H_0 : \beta_2 + 2\beta_1 = 1$ vs. $H_1 : \beta_2 + 2\beta_1 \neq 1$

Unrestricted model	Restricted model
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	$Y^* = \beta_0 + +\beta_1 X^* + \varepsilon$
	$Y^* = Y - X_2$
	$X^* = X_1 - 2X_2$

Tests with Multiple Linear Constraints: Definitions

	Unrestricted model	Restricted model
Coefficient of Determination	R^2_{un}	R^2_r
Sums of Squares of the Residuals	SSR_{un}	SSR_r

Tests with Multiple Linear Constraints

- Using the asymptotic distribution we have under H_0 with q linear constraints,

$$W^0 = n\frac{SSR_r - SSR_{un}}{SSR_{un}} \overset{\text{asy}}{\sim} \chi_q^2$$

- Or equivalently, provided that the restricted and unrestricted models have the **same** the dependent

$$W^0 = n\frac{R_{un}^2 - R_r^2}{1 - R_{un}^2} \overset{\text{asy}}{\sim} \chi_q^2$$

- Note that we use n instead of $(n - k - 1)$, for large samples, these values are very close.

Tests with Multiple Linear Constraints

- Most econometric programs perform automatically if the null hypothesis is written
  - Typically, these programs provide the F statistic, which assumes conditional normality of the observations
  - This statistic has the form:

$$F = \frac{SSR_r - SSR_{un}}{SSR_{un}} \frac{n-k-1}{q} \sim F_{q,n-k-1}$$

    and therefore relates to the asymptotic test statistic W_0 as $W_0 \approx qF$
  - For n big enough both test will provide the same conclusions. But we perform the asymptotic test
- Note that it is easy to show that $(SSR_r - SSR_{un}) \geq 0$ and $(R^2_{un} - R^2_r) \geq 0$
- All tests seen previously are special cases of this test.

Tests for Global Significance

- $H_0 : \beta_1 + \beta_2 + \ldots + \beta_k = 0$ versus $H_1 : \beta_j \neq 0$ at least for some $j \in \{1, 2, \ldots, k\}$
- Unrestricted Model: $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$, with R^2_{un}
- Restricted Model: $Y = \beta_0 + \varepsilon$with $R^2_r = 0$
- Using the asymptotic distribution under H_0,

$$W^0 = n \frac{R^2_{un}}{1 - R^2_{un}} \overset{\text{asy}}{\sim} \chi^2_k$$

Tests for Global Significance

- In general, econometric packages provide the F statistic for global significance:

$$F = \frac{R^2_{un}}{1 - R^2_{un}} \frac{n-k-1}{k} \sim F_{k,n-k-1}$$

and therefore relates to the asymptotic test statistic W_0 as $W_0 \approx kF$

Tests for Global Significance: Example

- Recall Model Demand example that uses data for U.S. TIM1.GDT
- Model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where
  - $Y = \ln(100 * V4/V3)$ = log of real quantity demanded of money (M1)
  - $X_1 = \ln(V2)$ = log of real GDP
  - $X + 2 = \ln(V6)$ = log of the interest rate for the treasury bill

Tests for Global Significance: Example

	Coefficient	Std. Error	t-ratio	p-value
const	2.32959	0.205437	11.3397	0.0000
X1	0.557290	0.0263835	21.1227	0.0000
X2	−0.203154	0.0210047	−9.6719	0.0000

Mean dependent var	6.628638	S.D. dependent var	0.172887
Sum squared resid	0.080411	S.E. of regression	0.047932
R^2	0.927291	Adjusted R^2	0.923137
$F(2, 35)$	223.1866	P-value(F)	1.20e–20
Log-likelihood	63.08600	Akaike criterion	−120.1720
Schwarz criterion	−115.2592	Hannan–Quinn	−118.4241
$\hat{\rho}$	0.627702	Durbin–Watson	0.727502

Tests for Global Significance: Example

- We consider $H_0 : \beta_1 + \beta_2 = 0$ versus $H_1 : \beta_j \neq 0$ at least for some $j \in \{1, 2\}$
- The asymptotic test then is:

$$W^0 = 38 * \frac{0.9273}{1 - 0.9273} = 482.55 > \chi^2_2 = 5.99$$

- Gretl output gives the F statistic as: $F(2, 35) = 223.1866$, which is quite close to $W^0/2$
- Clearly, we reject H_0