Obtaining univariate OLS parameter estimates through elementary algebra

Consider the linear regression model

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, \dots, T.$$

Given parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ the residuals are defined as

$$\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta} x_t, \quad t = 1, \dots, T.$$

The OLS estimates of the parameters α and β are determined by minimizing the residual sum of squares

$$RSS = \sum_{t=1}^{T} \hat{u}_t^2.$$

We will show that the OLS estimates can be obtained by elementary algebra, without using calculus. First, express the residuals using the demeaned versions of y_t and x_t (that is, $y_t - \bar{y}$ and $x_t - \bar{x}$) as

$$\hat{u}_t = \left[(y_t - \bar{y}) - \hat{\beta}\left(x_t - \bar{x}\right)\right] - \left[\hat{\alpha} - \left(\bar{y} - \hat{\beta}\bar{x}\right)\right]$$

and note that since $\sum_{t=1}^{T} (y_t - \bar{y}) = \sum_{t=1}^{T} (x_t - \bar{x}) = 0$ we have

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} \left[(y_t - \bar{y}) - \hat{\beta}\left(x_t - \bar{x}\right)\right]^2 + T\left[\hat{\alpha} - \left(\bar{y} - \hat{\beta}\bar{x}\right)\right]^2,$$

as the sum of cross products vanishes. The second term is minimized, and becomes equal to zero, when

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Hence, it suffices to minimize the term $\sum_{t=1}^{T} \left[(y_t - \bar{y}) - \hat{\beta}\left(x_t - \bar{x}\right)\right]^2$ with respect to $\hat{\beta}$. Note that

$$\sum_{t=1}^{T} \left[(y_t - \bar{y}) - \hat{\beta}\left(x_t - \bar{x}\right)\right]^2 = S_{yy} + S_{xx}\hat{\beta}^2 - 2S_{yx}\hat{\beta}$$

where

$$
\begin{aligned}
S_{yy} &= \sum_{t=1}^{T} (y_t - \bar{y})^2 , \\
S_{xx} &= \sum_{t=1}^{T} (x_t - \bar{x})^2 , \\
S_{yx} &= \sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x}) .
\end{aligned}
$$

Since S_{yy} does not depend on $\hat{\beta}$, it suffices to minimize the quantity $S_{xx}\hat{\beta}^2 - 2S_{yx}\hat{\beta}$ with respect to $\hat{\beta}$. We do so, by "completing the square":

$$
S_{xx}\hat{\beta}^2 - 2S_{yx}\hat{\beta} = S_{xx}\left(\hat{\beta}^2 - 2\frac{S_{yx}}{S_{xx}}\hat{\beta}\right) = S_{xx}\left[\left(\hat{\beta} - \frac{S_{yx}}{S_{xx}}\right)^2 - \frac{S_{yx}^2}{S_{xx}^2}\right].
$$

It is clear that the expression above is minimized when the square $\left(\hat{\beta} - \frac{S_{yx}}{S_{xx}}\right)^2$ becomes zero, i.e., when $\hat{\beta} = \frac{S_{yx}}{S_{xx}}$. It follows that the OLS estimates of α and β are given by

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} = \frac{\sum_{t=1}^{T} y_t x_t - T\bar{x}\bar{y}}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2}, \\
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}.
\end{aligned}
$$

Obtaining multivariate OLS parameter estimates through elementary algebra

Consider the linear regression model

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t, \quad t = 1, \ldots, T,$$

where the parameter $\boldsymbol{\beta}$ and the regressor $\mathbf{x}_t$ are k-dimensional vectors. Given a parameter estimate $\hat{\boldsymbol{\beta}}$ the residuals are defined as

$$\hat{u}_t = y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t, \quad t = 1, \ldots, T.$$

The OLS estimates of the parameter vector $\boldsymbol{\beta}$ are determined by minimizing the residual sum of squares

$$RSS = \sum_{t=1}^{T} \hat{u}_t^2.$$

We will show that the OLS estimates can be obtained by elementary algebra, without using calculus. Note that

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} \left(y_t - \hat{\boldsymbol{\beta}}'\mathbf{x}_t\right)^2 = S_{yy} + \hat{\boldsymbol{\beta}}'\mathbf{S}_{xx}\hat{\boldsymbol{\beta}} - 2\mathbf{S}_{yx}'\hat{\boldsymbol{\beta}}$$

where

$$\begin{aligned} S_{yy} &= \sum_{t=1}^{T} y_t^2, \\ \mathbf{S}_{xx} &= \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t', \\ \mathbf{S}_{yx} &= \sum_{t=1}^{T} \mathbf{x}_t y_t. \end{aligned}$$

Since S_{yy} does not depend on $\hat{\boldsymbol{\beta}}$, it suffices to minimize the quantity $\hat{\boldsymbol{\beta}}'\mathbf{S}_{xx}\hat{\boldsymbol{\beta}} - 2\mathbf{S}_{yx}'\hat{\boldsymbol{\beta}}$ with respect to $\hat{\boldsymbol{\beta}}$. As in the case of the univariate regression, we do so by "completing the square".

Consider the function $Q(\mathbf{z}) = \mathbf{z}'\mathbf{A}\mathbf{z} - 2\mathbf{b}'\mathbf{z}$, where $\mathbf{z}$ is an k-dimensional vector, $\mathbf{A}$ is a symmetric and invertible $k \times k$ matrix, and $\mathbf{b}$ is an k-dimensional vector. Note that

$$(\mathbf{z} - \mathbf{c})' \mathbf{A} (\mathbf{z} - \mathbf{c}) = \mathbf{z}'\mathbf{A}\mathbf{z} - 2\mathbf{c}'\mathbf{A}\mathbf{z} + \mathbf{c}'\mathbf{A}\mathbf{c}$$

and so if we choose $\mathbf{c}$ so that $\mathbf{c}'\mathbf{A} = \mathbf{b}' \Leftrightarrow \mathbf{c} = \mathbf{A}^{-1}\mathbf{b}$ we obtain

$$\left(\mathbf{z} - \mathbf{A}^{-1}\mathbf{b}\right)' \mathbf{A} \left(\mathbf{z} - \mathbf{A}^{-1}\mathbf{b}\right) = \mathbf{z}'\mathbf{A}\mathbf{z} - 2\mathbf{b}'\mathbf{z} + \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}.$$

It follows that

$$\mathbf{z}'\mathbf{A}\mathbf{z} - 2\mathbf{b}'\mathbf{z} = \left(\mathbf{z} - \mathbf{A}^{-1}\mathbf{b}\right)' \mathbf{A} \left(\mathbf{z} - \mathbf{A}^{-1}\mathbf{b}\right) - \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}$$

and so the function $Q(\mathbf{z}) = \mathbf{z}'\mathbf{A}\mathbf{z} - 2\mathbf{b}'\mathbf{z}$ is minimized when $\mathbf{z} = \mathbf{A}^{-1}\mathbf{b}$.

Therefore, the quantity $\hat{\boldsymbol{\beta}}'\mathbf{S}_{xx}\hat{\boldsymbol{\beta}} - 2\mathbf{S}'_{yx}\hat{\boldsymbol{\beta}}$ is minimized for $\hat{\boldsymbol{\beta}} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{yx}$ and so the OLS estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{yx} = \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right).$$

The more frequently used form of the OLS estimate, based on vector-matrix notation, is obtained as follows. Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_T' \end{bmatrix}.$$

It then follows that $\sum_{t=1}^{T} \mathbf{x}_t y_t = \mathbf{X}'\mathbf{y}$ and $\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' = \mathbf{X}'\mathbf{X}$ and, therefore, we have

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y}.$$

Asymptotic Normality of OLS Estimators

Consider the linear regression model

$$y_t = \boldsymbol{\gamma}' \mathbf{z}_t + u_t, \quad t = 1, \ldots, T,$$

where the parameter $\boldsymbol{\gamma}$ and the regressor $\mathbf{z}_t$ are k-dimensional vectors. We assume that $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[\mathbf{z}_t u_t] = \mathbf{0}_k$, that the moment matrices $\mathbb{E}[\mathbf{z}_t \mathbf{z}_t']$ and $\mathbb{E}[(\mathbf{z}_t \mathbf{z}_t') u_t^2]$ are finite, and denote $\sigma^2 = \mathbb{E}[u_t^2] = \mathbb{V}[u_t]$. The OLS estimator of $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_t y_t \right) = \boldsymbol{\gamma} + \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t u_t \right).$$

Hence,

$$\sqrt{T} \left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \right) = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{z}_t u_t \right).$$

Since $\frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'$ converges to $\mathbb{E}[\mathbf{z}_t \mathbf{z}_t']$ in probability and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{z}_t u_t$ converges in distribution to a multivariate normal with mean $\mathbf{0}_k$ and covariance matrix $\mathbb{E}[(\mathbf{z}_t \mathbf{z}_t') u_t^2]$, we have that $\sqrt{T} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ converges in distribution to a multivariate normal with mean $\mathbf{0}_k$ and covariance matrix $(\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1} \cdot \mathbb{E}[(\mathbf{z}_t \mathbf{z}_t') u_t^2] \cdot (\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1}$.

Note that if $\mathbf{z}_t$ and u_t are assumed to be independent, then the covariance matrix simplifies to $\sigma^2 \cdot (\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1}$.

Let us now specialize to the case in which there is an intercept and only one pure regressor x_t. Then $k = 2$ and $\mathbf{z}_t = \begin{bmatrix} 1 & x_t \end{bmatrix}'$. It follows that

$$(\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1} = \begin{bmatrix} 1 & \mathbb{E}[x_t] \\ \mathbb{E}[x_t] & \mathbb{E}[x_t^2] \end{bmatrix}^{-1} = \frac{1}{\mathbb{V}[x_t]} \begin{bmatrix} \mathbb{E}[x_t^2] & -\mathbb{E}[x_t] \\ -\mathbb{E}[x_t] & 1 \end{bmatrix}$$

and

$$\begin{aligned} & (\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1} (\mathbf{z}_t \mathbf{z}_t') (\mathbb{E}[\mathbf{z}_t \mathbf{z}_t'])^{-1} \\ = \ & \frac{1}{\mathbb{V}[x_t]^2} \begin{bmatrix} \mathbb{E}[x_t^2] & -\mathbb{E}[x_t] \\ -\mathbb{E}[x_t] & 1 \end{bmatrix} \begin{bmatrix} 1 & x_t \\ x_t & x_t^2 \end{bmatrix} \begin{bmatrix} \mathbb{E}[x_t^2] & -\mathbb{E}[x_t] \\ -\mathbb{E}[x_t] & 1 \end{bmatrix} \\ = \ & \frac{1}{\mathbb{V}[x_t]^2} \begin{bmatrix} \mathbb{E}[x_t^2] - \mathbb{E}[x_t] x_t & \mathbb{E}[x_t^2] x_t - \mathbb{E}[x_t] x_t^2 \\ -\mathbb{E}[x_t] + x_t & -\mathbb{E}[x_t] x_t + x_t^2 \end{bmatrix} \begin{bmatrix} \mathbb{E}[x_t^2] & -\mathbb{E}[x_t] \\ -\mathbb{E}[x_t] & 1 \end{bmatrix} \end{aligned}$$

$$= \frac{1}{\mathbb{V}[x_t]^2} \begin{bmatrix} \left(\mathbb{E}\left[x_t^2\right] - x_t\mathbb{E}\left[x_t\right]\right)^2 & -\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right) \\ -\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right) & \left(x_t - \mathbb{E}\left[x_t\right]\right)^2 \end{bmatrix},$$

and so

$$\begin{aligned} &\left(\mathbb{E}\left[\mathbf{z}_t\mathbf{z}_t'\right]\right)^{-1} \cdot \mathbb{E}\left[\left(\mathbf{z}_t\mathbf{z}_t'\right)u_t^2\right] \cdot \left(\mathbb{E}\left[\mathbf{z}_t\mathbf{z}_t'\right]\right)^{-1} \\ = &\frac{1}{\mathbb{V}\left[x_t\right]^2} \begin{bmatrix} \mathbb{E}\left[\left(\mathbb{E}\left[x_t^2\right] - x_t\mathbb{E}\left[x_t\right]\right)^2 u_t^2\right] & -\mathbb{E}\left[\left(\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right)\right)u_t^2\right] \\ -\mathbb{E}\left[\left(\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right)\right)u_t^2\right] & \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2 u_t^2\right] \end{bmatrix}. \end{aligned}$$

Under the assumption that x_t and u_t are independent, we further have

$$\mathbb{E}\left[\left(x_t\mathbb{E}\left[x_t\right] - \mathbb{E}\left[x_t^2\right]\right)^2 u_t^2\right] = \mathbb{E}\left[\left(x_t\mathbb{E}\left[x_t\right] - \mathbb{E}\left[x_t^2\right]\right)^2\right]\mathbb{E}\left[u_t^2\right] = \mathbb{E}\left[x_t^2\right] \cdot \mathbb{V}\left[x_t\right] \cdot \sigma^2,$$

$$\begin{aligned} \mathbb{E}\left[\left(\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right)\right)u_t^2\right] &= \mathbb{E}\left[\left(\left(\mathbb{E}\left[x_t\right] - x_t\right)\left(\mathbb{E}\left[x_t^2\right] - \mathbb{E}\left[x_t\right]x_t\right)\right)\right]\mathbb{E}\left[u_t^2\right] \\ &= \mathbb{E}\left[x_t\right] \cdot \mathbb{V}\left[x_t\right] \cdot \sigma^2, \end{aligned}$$

$$\mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2 u_t^2\right] = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2\right]\mathbb{E}\left[u_t^2\right] = \mathbb{V}\left[x_t\right] \cdot \sigma^2,$$

and the covariance matrix $\sigma^2 \cdot \left(\mathbb{E}\left[\mathbf{z}_t\mathbf{z}_t'\right]\right)^{-1}$ simplifies to

$$\frac{\sigma^2}{\mathbb{V}\left[x_t\right]} \cdot \begin{bmatrix} \mathbb{E}\left[x_t^2\right] & -\mathbb{E}\left[x_t\right] \\ -\mathbb{E}\left[x_t\right] & 1 \end{bmatrix}.$$

Univariate OLS: Unbiased Estimator of Error Variance

Consider the univariate linear regression model

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, \dots, T.$$

where the regressors are non-stochastic (fixed) and the disturbances have zero mean and are uncorrelated and homoscedastic with variance equal to σ^2, i.e., $\mathbb{E}[u_t] = 0$, $\mathbb{C}[u_t, u_s] = 0$, $t \neq s$, and $\mathbb{V}[u_t] = \mathbb{E}[u_t^2] = \sigma^2$. We will show that

$$s^2 = \frac{1}{T-2} \sum_{t=1}^{T} \hat{u}_t^2$$

is an unbiased estimator of the disturbance variance σ^2, where $\hat{u}_t = y_t - \hat{y}_t$, $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$, and $\hat{\alpha}$ and $\hat{\beta}$ are the OLS estimators of α and β, respectively. The OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ are given by

$$\hat{\beta} = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x},$$

which can also be expressed as

$$\hat{\beta} = \sum_{t=1}^{T} w_t y_t, \quad \hat{\alpha} = \sum_{t=1}^{T} q_t y_t,$$

where

$$w_t = \frac{x_t - \bar{x}}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, \quad q_t = \frac{1}{T} - \bar{x} \cdot w_t.$$

The following properties are easily verified:

$$\sum_{t=1}^{T} w_t = 0, \quad \sum_{t=1}^{T} w_t x_t = 1, \quad \sum_{t=1}^{T} q_t = 1, \quad \sum_{t=1}^{T} q_t x_t = 0,$$

and

$$\sum_{t=1}^{T} w_t^2 = \frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, \quad \sum_{t=1}^{T} q_t^2 = \frac{\frac{1}{T} \sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, \quad \sum_{t=1}^{T} q_t w_t = \frac{-\bar{x}}{\sum_{t=1}^{T} (x_t - \bar{x})^2}.$$

Using these properties, we observe that

$$\hat{\beta} = \sum_{t=1}^{T} w_t y_t = \sum_{t=1}^{T} w_t \left(\alpha + \beta x_t + u_t\right) = \beta + \sum_{t=1}^{T} w_t u_t,$$

and

$$\hat{\alpha} = \sum_{t=1}^{T} q_t y_t = \sum_{t=1}^{T} q_t \left(\alpha + \beta x_t + u_t\right) = \alpha + \sum_{t=1}^{T} q_t u_t.$$

Note that the residuals can be expressed as

$$\hat{u}_t = y_t - \hat{y}_t = y_t - \left(\hat{\alpha} + \hat{\beta} x_t\right) = \left(\alpha + \beta x_t + u_t\right) - \left(\hat{\alpha} + \hat{\beta} x_t\right),$$

or equivalently

$$\hat{u}_t = -\left[\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t\right] + u_t.$$

Hence,

$$\hat{u}_t^2 = \left[\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t\right]^2 + u_t^2 - 2\left[\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t\right] u_t$$

where

$$\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t = \left(\sum_{s=1}^{T} q_s u_s\right) + \left(\sum_{s=1}^{T} w_s u_s\right) x_t.$$

It follows that

$$\begin{aligned}
& \mathbb{E}\left[\left[\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t\right] u_t\right] \\
= \;& \mathbb{E}\left[\left[\left(\sum_{s=1}^{T} q_s u_s\right) + \left(\sum_{s=1}^{T} w_s u_s\right) x_t\right] u_t\right] \\
= \;& \left(q_t + w_t x_t\right) \sigma^2,
\end{aligned}$$

and so

$$\sum_{t=1}^{T} \mathbb{E}\left[\left[\left(\hat{\alpha} - \alpha\right) + \left(\hat{\beta} - \beta\right) x_t\right] u_t\right] = \sum_{t=1}^{T} \left(q_t + w_t x_t\right) \sigma^2 = (1+1)\sigma^2 = 2\sigma^2.$$

Moreover,

$$\begin{aligned}
& \mathbb{E}\left[\left[(\hat{\alpha}-\alpha)+\left(\hat{\beta}-\beta\right)x_t\right]^2\right] \\
= & \mathbb{E}\left[\left[\left(\sum_{s=1}^{T} q_s u_s\right)+\left(\sum_{s=1}^{T} w_s u_s\right)x_t\right]^2\right] \\
= & \mathbb{E}\left[\left(\sum_{s=1}^{T} q_s u_s\right)^2\right]+\mathbb{E}\left[\left(\sum_{s=1}^{T} w_s u_s\right)^2\right]\cdot x_t^2+2\cdot\mathbb{E}\left[\left(\sum_{s=1}^{T} q_s u_s\right)\left(\sum_{s=1}^{T} w_s u_s\right)\right]\cdot x_t \\
= & \sigma^2\left[\left(\sum_{s=1}^{T} q_s^2\right)+\left(\sum_{s=1}^{T} w_s^2\right)\cdot x_t^2+2\cdot\left(\sum_{s=1}^{T} q_s w_s\right)\cdot x_t\right] \\
= & \sigma^2\left[\frac{\frac{1}{T}\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}+\frac{x_t^2}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}-2\frac{\bar{x}\cdot x_t}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}\right],
\end{aligned}$$

and so

$$\begin{aligned}
& \sum_{t=1}^{T}\mathbb{E}\left[\left[(\hat{\alpha}-\alpha)+\left(\hat{\beta}-\beta\right)x_t\right]^2\right] \\
= & \sigma^2\left[\frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}+\frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}-2\frac{T\cdot\bar{x}^2}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}\right], \\
= & 2\sigma^2\left[\frac{\frac{1}{T}\sum_{t=1}^{T} x_t^2-\bar{x}^2}{\frac{1}{T}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}\right] \\
= & 2\sigma^2.
\end{aligned}$$

Finally, $\sum_{t=1}^{T}\mathbb{E}\left[u_t^2\right]=T\sigma^2$ and so

$$\sum_{t=1}^{T}\mathbb{E}\left[\hat{u}_t^2\right]=2\cdot\sigma^2+T\cdot\sigma^2-2\cdot 2\cdot\sigma^2=(T-2)\cdot\sigma^2$$

which yields

$$\mathbb{E}\left[\frac{1}{T-2}\sum_{t=1}^{T}\hat{u}_t^2\right]=\sigma^2.$$

R^2 : a goodness of fit measure

Consider the linear regression model

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t, \quad t = 1, \dots, T,$$

where the parameter $\boldsymbol{\beta}$ and the regressor $\mathbf{x}_t$ are k-dimensional vectors. The regressor $\mathbf{x}_t$ includes an intercept. In vector-matrix form, the system is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

The OLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

and the vector of residuals is

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

The R^2 is defined as the square of the correlation between the actual data $\mathbf{y}$ and the fitted data $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, that is

$$R^2 = \frac{\left[\sum_{t=1}^{T}\left(y_t - \bar{y}\right)\left(\hat{y}_t - \bar{\hat{y}}\right)\right]^2}{\left[\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2\right]\left[\sum_{t=1}^{T}\left(\hat{y}_t - \bar{\hat{y}}\right)^2\right]},$$

where the averages $\bar{y}$ and $\bar{\hat{y}}$ are defined as

$$\begin{aligned} \bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t, \\ \bar{\hat{y}} &= \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t. \end{aligned}$$

First, we show that $\bar{\hat{y}} = \bar{y}$. Note that the matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \left[\begin{array}{cc} \mathbf{1}_T & \mathbf{Z} \end{array}\right]$$

where $\mathbf{1}_T$ is the $T \times 1$ vector of ones and $\mathbf{Z}$ is the $T \times (k-1)$ matrix of “pure” regressors.

The vector of residuals is expressed

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{y} = \left[\mathbf{I}_T - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y} = \left(\mathbf{I}_T - \mathbf{P}_X\right)\mathbf{y}$$

where $\mathbf{I}_T$ is the $T \times T$ identity matrix and

$$\mathbf{P}_X = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'.$$

Note that $\mathbf{X}'\left(\mathbf{I}_T - \mathbf{P}_X\right) = \mathbf{X}' - \mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{X}' - \mathbf{X}' = \mathbf{0}_{k\times T}$ and so $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}_k$. This implies that $\mathbf{1}_T'\hat{\mathbf{u}} = 0$, or $\frac{1}{T}\sum_{t=1}^T \hat{u}_t = \bar{\hat{u}} = 0$. Since $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$, we have $\bar{y} = \bar{\hat{y}}$. It follows that

$$R^2 = \frac{\left[\sum_{t=1}^T \left(y_t - \bar{y}\right)\left(\hat{y}_t - \bar{y}\right)\right]^2}{\left[\sum_{t=1}^T \left(y_t - \bar{y}\right)^2\right]\left[\sum_{t=1}^T \left(\hat{y}_t - \bar{y}\right)^2\right]}.$$

Moreover,

$$\sum_{t=1}^T \left(y_t - \bar{y}\right)^2 = \left(\mathbf{y} - \bar{y}\mathbf{1}_T\right)'\left(\mathbf{y} - \bar{y}\mathbf{1}_T\right),$$

$$\sum_{t=1}^T \left(\hat{y}_t - \bar{y}\right)^2 = \left(\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T\right)'\left(\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T\right),$$

and

$$\sum_{t=1}^T \left(y_t - \bar{y}\right)\left(\hat{y}_t - \bar{y}\right) = \left(\mathbf{y} - \bar{y}\mathbf{1}_T\right)'\left(\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T\right).$$

Note that

$$\mathbf{y} - \bar{y}\mathbf{1}_T = \left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y}$$

where $\mathbf{J}_T$ is the $T \times T$ matrix of ones, and

$$\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T = \left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y}.$$

The matrices $\mathbf{J}_T$ and $\mathbf{P}_X$ are symmetric and have the following properties $\mathbf{J}_T\mathbf{J}_T = T\mathbf{J}_T$, $\mathbf{P}_X\mathbf{P}_X = \mathbf{P}_X$ and $\mathbf{P}_X\mathbf{X} = \mathbf{X}$. From the last property it follows that $\mathbf{P}_X\mathbf{1}_T = \mathbf{1}_T$ and so $\mathbf{P}_X\mathbf{J}_T = \mathbf{J}_T$. Hence, the matrices $\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T$, $\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T$, and $\mathbf{I}_T - \mathbf{P}_X$ are idempotent, i.e.,

$$\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right) = \mathbf{I}_T - \frac{1}{T}\mathbf{J}_T$$

$$\begin{aligned}\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right) &= \mathbf{P}_X - \frac{1}{T}\mathbf{J}_T \\ \left(\mathbf{I}_T - \mathbf{P}_X\right)\left(\mathbf{I}_T - \mathbf{P}_X\right) &= \mathbf{I}_T - \mathbf{P}_X.\end{aligned}$$

Using the above properties, we obtain

$$\begin{aligned}\sum_{t=1}^{T}(y_t - \bar{y})^2 &= (\mathbf{y} - \bar{y}\mathbf{1}_T)'(\mathbf{y} - \bar{y}\mathbf{1}_T) \\ &= \mathbf{y}'\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y} = \mathbf{y}'\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y},\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 &= (\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T)'(\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T), \\ &= \mathbf{y}'\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y} = \mathbf{y}'\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y},\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{T}(y_t - \bar{y})(\hat{y}_t - \bar{y}) &= (\mathbf{y} - \bar{y}\mathbf{1}_T)'(\hat{\mathbf{y}} - \bar{y}\mathbf{1}_T) \\ &= \mathbf{y}'\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y} = \mathbf{y}'\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y}.\end{aligned}$$

The total sum of squares is defined as

$$TSS \equiv \sum_{t=1}^{T}(y_t - \bar{y})^2 = \mathbf{y}'\left(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y},$$

the explained sum of squares is defined as

$$ESS = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 = \mathbf{y}'\left(\mathbf{P}_X - \frac{1}{T}\mathbf{J}_T\right)\mathbf{y},$$

and the residual sum of squares as

$$RSS = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}\hat{u}_t^2.$$

Since

$$\mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_T - \mathbf{P}_X)\mathbf{y}$$

we have

$$RSS=\sum_{t=1}^{T}\left(y_t-\hat{y}_t\right)^2=\mathbf{y}'\left(\mathbf{I}_T-\mathbf{P}_X\right)\left(\mathbf{I}_T-\mathbf{P}_X\right)\mathbf{y}=\mathbf{y}'\left(\mathbf{I}_T-\mathbf{P}_X\right)\mathbf{y}.$$

It follows that

$$TSS=ESS+RSS$$

and

$$\begin{aligned}
R^2 &= \frac{\left[\mathbf{y}'\left(\mathbf{P}_X-\frac{1}{T}\mathbf{J}_T\right)\mathbf{y}\right]^2}{\left[\mathbf{y}'\left(\mathbf{I}_T-\frac{1}{T}\mathbf{J}_T\right)\mathbf{y}\right]\left[\mathbf{y}'\left(\mathbf{P}_X-\frac{1}{T}\mathbf{J}_T\right)\mathbf{y}\right]}\\
&= \frac{\mathbf{y}'\left(\mathbf{P}_X-\frac{1}{T}\mathbf{J}_T\right)\mathbf{y}}{\mathbf{y}'\left(\mathbf{I}_T-\frac{1}{T}\mathbf{J}_T\right)\mathbf{y}}=\frac{ESS}{TSS}=1-\frac{RSS}{TSS}.
\end{aligned}$$

Note that ESS and RSS are both nonnegative quantities and so $0\leq ESS\leq TSS$ and $0\leq RSS\leq TSS$. Equivalently, R^2 takes values between 0 and 1. If the regression provides a perfect fit, then $RSS=0\Leftrightarrow ESS=TSS\Leftrightarrow R^2=1$. If the regression does not have any explanatory power, then $ESS=0\Leftrightarrow R^2=0$.