Tutorial 1

October 14, 2020

1 Linear regression model

We have a linear regression in matrix form

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \epsilon \sim N(0, \sigma^2 \mathbf{I}_n), \tag{1}$$

and from the lecture we know the OLS estimator for both β and σ^2 is

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

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{n-k}. \tag{3}$$

How do we get the projection matrix? First, we can see that the fitted values are

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}, \tag{4}$$

and we can substitute (2) into (4) then it becomes

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

where the projection matrix is $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
The annihilation matrix or the residual maker can be derived through the residuals

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}, \tag{5}$$

and we can substitute (2) into (5)

$$\mathbf{e} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

$$\mathbf{e} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y},$$

$$\mathbf{e} = \mathbf{M}\mathbf{y},$$

where $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ can be seen as a matrix that makes the residuals out of $\mathbf{y}$. We want to show for both the projection and annihilation matrix they are symmetric and idempotent. Let define a matrix $\mathbf{A}$ then the matrix $\mathbf{A}$ is symmetric if

$$\mathbf{A} = \mathbf{A}^T,$$

and $\mathbf{A}$ is idempotent if

$$\mathbf{A} = \mathbf{A}^2$$

Therefore, let's prove $\mathbf{M}$ and $\mathbf{P}$ are symmetric

$$\mathbf{P}' = (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'$$

$$\mathbf{P}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Note $\mathbf{X}'\mathbf{X}$ is always symmetric. A similar logic can be applied to the $\mathbf{M}$ matrix. To show that $\mathbf{M}$ is idempotent then

$$\mathbf{M}^2 = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

$$\mathbf{M}^2 = \mathbf{I}^2 - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M}^2 = \mathbf{I}^2 - 2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M}^2 = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M}^2 = \mathbf{M}.$$

A similar logic can be applied to $\mathbf{P}$.
Next, we want to show that the $MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2$. Let's assume that we know $E(\hat{\theta}) = \bar{\theta}$. Thus, the MSE can be written in the form

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2],$$

Next we can add and subtract the $\bar{\theta}$ from the term

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \bar{\theta} + \bar{\theta} + \theta)^2],$$

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \bar{\theta})^2 - 2(\bar{\theta} - \theta)(\hat{\theta} - \bar{\theta}) + (\bar{\theta} - \theta)^2)],$$

We can use the linear expectation operator rule

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \bar{\theta})^2] - 2E[(\bar{\theta} - \theta)(\hat{\theta} - \bar{\theta})] + E[(\bar{\theta} - \theta)^2)],$$

Note if X is a random variable then the variance is $var(X) = E(X - E(X))^2$. Therefore $E[(\hat{\theta} - \bar{\theta})^2] = Var(\hat{\theta})$. Note here $E(\hat{\theta}) - \bar{\theta} = 0$ (from the above assumption), thus the term $2E[(\bar{\theta} - \theta)(\hat{\theta} - \bar{\theta})] = 0$. As a result,

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + E[(\bar{\theta} - \theta)^2)],$$

By definition $E[(\bar{\theta} - \theta)]$ is defined as the bias for $\hat{\theta}$. Thus,

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2.$$

2 OLS estimator

From (2) we know the OLS estimator for β is

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

We can substitute (1) into (2)

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon),$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon,$$

Since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$, then

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon,$$

Thus the sampling error is

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon.$$

For the OLS estimator β to be unbiased, we want to show $E(\hat{\beta} - \beta|\mathbf{X}) = 0$. Thus,

$$E(\hat{\beta} - \beta|\mathbf{X}) = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon|\mathbf{X}),$$

Since the matrix $\mathbf{X}$ is known, the only random variable in the above term is ϵ and from (1) we know that it is normally distributed with a mean equal to zero, $E(\epsilon|\mathbf{X}) = 0$

$$E(\hat{\beta} - \beta|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon|\mathbf{X}),$$

Thus,

$$E(\hat{\beta} - \beta|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0},$$

$$E(\hat{\beta} - \beta|\mathbf{X}) = 0.$$

We now want to find the variance of the OLS estimator β, $var(\hat{\beta})$, from (2)

$$var(\hat{\beta}) = var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}),$$

$$var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'var(\mathbf{y}|\mathbf{X}),$$

Note assume $z \sim N(0, \Omega)$ and $\mathbf{A}$ is a known matrix. Then the variance of $V(\mathbf{A}z) = \mathbf{A}V(z)\mathbf{A}' = \mathbf{A}\Omega\mathbf{A}'$. You can see in the above term that $\mathbf{A}$ is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Therefore, we know $((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Remember $\mathbf{X}'\mathbf{X}$ is a symmetric matrix. Thus,

$$var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'var(\mathbf{y}|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

From (1) we know $var(\mathbf{y}|\mathbf{X}) = \sigma^2\mathbf{I}$. Thus,

$$var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

Since σ^2 is scalar, we can move it to the front

$$var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

Since, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$, then

$$var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

3 GLS estimator

If we violate the assumption 1.4, that is the conditional homoskedasticity and independence, the linear regression model becomes

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \epsilon \sim N(0, \sigma^2\mathbf{V}), \tag{6}$$

where $\mathbf{V}$ is a known, symmetric and positive definite matrix. The inverse of $\mathbf{V}^{-1}$ also exist and can decomposed as $\mathbf{V}^{-1} = \mathbf{C}'\mathbf{C}$. We can premultiply (6) with $\mathbf{C}$

$$\mathbf{Cy} = \mathbf{CX}\beta + \mathbf{C}\epsilon, \epsilon \sim N(0, \sigma^2),$$

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\beta + \tilde{\epsilon}, \tilde{\epsilon} \sim N(0, \sigma^2 \mathbf{CC}'),$$

where $\tilde{\mathbf{y}} = \mathbf{Cy}$, $\tilde{\mathbf{X}} = \mathbf{CX}$ and $\mathbf{V} = \mathbf{CC}'$. Note $(\mathbf{V}^{-1})^{-1} = (\mathbf{C}'\mathbf{C})^{-1} = (\mathbf{V}^{-\frac{1}{2}'}\mathbf{V}^{-\frac{1}{2}})^{-1} = \mathbf{V}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}'} = \mathbf{CC}' = \mathbf{V}$. Therefore the GLS estimator for β is

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{7}$$

What is sampling error $\hat{\beta}_{GLS} - \beta$? We can substitute (6) into (7), therefore

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{X}\beta + \epsilon),$$

$$\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta + (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\epsilon,$$

since $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \mathbf{I}$,then

$$\hat{\beta}_{GLS} = \beta + (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\epsilon,$$

$$\hat{\beta}_{GLS} - \beta = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\epsilon.$$

Next, we want to show that $\hat{\beta}_{GLS}$ is an unbiased estimator which implies $E(\hat{\beta}_{GLS}|\mathbf{X}) = \beta$. Therefore we want to show $E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = 0$, and $E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta$ can be equivalent to $E(\hat{\beta}_{GLS} - \beta|\mathbf{X})$. Given that we already know $\hat{\beta}_{GLS} - \beta$, then

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = E(\hat{\beta}_{GLS} - \beta|\mathbf{X}),$$

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = E((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\epsilon|\mathbf{X}),$$

Note that $\mathbf{X}$ and $\mathbf{V}$ are known matrices, they are not random variables. The only random variable defined in this term is ϵ and we know from the normality assumption that $E(\epsilon|\mathbf{X}) = 0$. Therefore,

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\epsilon|\mathbf{X}),$$

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{0},$$

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = \mathbf{0} \iff E(\hat{\beta}_{GLS}|\mathbf{X}) = \beta.$$

Next, we want to show the variance of the $V(\hat{\beta}_{GLS}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. Using (7) we want to find

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = V((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}),$$

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = ((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}V(\mathbf{y}|\mathbf{X})),$$

Note: assume $z \sim N(0, \Omega)$ and $\mathbf{A}$ is a known matrix. Then the variance of $V(\mathbf{A}z) = \mathbf{A}V(z)\mathbf{A}' = \mathbf{A}\Omega\mathbf{A}'$. You can see in the above term that $\mathbf{A}$ is $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. Therefore, we know $((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})' = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. Remember both $\mathbf{V}$and $\mathbf{X}'\mathbf{X}$ are known and symmetric matrices. Thus,

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = ((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}V(\mathbf{y}|\mathbf{X})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}),$$

From (6) we know $V(\mathbf{y}|\mathbf{X}) = \sigma^2\mathbf{V}$,

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = ((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\sigma^2\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}),$$

since $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}$,

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = \sigma^2((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}),$$

and $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \mathbf{I}$, then we get

$$V(\hat{\beta}_{GLS}|\mathbf{X}) = \sigma^2((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

4 FGLS

In GLS, the matrix $\mathbf{V}$ is assumed be known but in practice usually $\mathbf{V}$is unknown and infeasible. As a result, an estimator of $\mathbf{V}$, that is $\hat{\mathbf{V}}$, has be undertaken. When estimation of $\hat{\mathbf{V}}$ is involved then it becomes FGLS. Essentially the FGLS is a two step estimation, that is first estimate $\hat{\mathbf{V}}$ and then plug this $\hat{\mathbf{V}}$ back into (6) and compute (7). However, the issue now is that $\hat{\mathbf{V}}$ becomes a random variable as it is typically a function of the data. Therefore, the standard proof of unbiasedness and the variance described above for the GLS estimator cannot be implemented for the FGLS estimator. For example, in the above section we showed that

$$E(\hat{\beta}_{GLS}|\mathbf{X}) - \beta = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\epsilon|\mathbf{X}),$$

However, for the FGLS estimator this is not true since $\hat{\mathbf{V}}$ is a random variable.

$$E(\hat{\beta}_{FGLS}|\mathbf{X}) - \beta \neq (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}E(\epsilon|\mathbf{X}),$$

Note before in the GLS, the matrix $\mathbf{V}$ was assumed to be known and not a random variable. Similarly, the variance of the FGLS estimator does not hold too

$$V(\hat{\beta}_{FGLS}|\mathbf{X}) \neq (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}V(\mathbf{y}|\mathbf{X})).$$