Generalised Least Squares and Omitted Variables

1 Conditional heteroskedasticity: GLS, FGLS, and ML estimators

Consider the model

$$
\begin{aligned}
y_i &= \begin{cases} \beta_1 + \beta_3 X_{3i} + u_i, & \text{for } i = 1, 2, ..., n^A \\ (\beta_1 + \beta_2 + \beta_3 X_{3i} + u_i, & \text{for } i = n^A + 1, n^A + 2, ..., n \\ \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, & \text{for } i = 1, 2, ..., n, \end{cases} \\
&\Leftrightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.
\end{aligned}
$$

in which X_{2i} is a binary dummy variable which takes the value zero for observations $i = 1, 2, ..., n^A$, and the value one for observations $i = n^A + 1, n^A + 2, ..., n$. You may also assume that the model satisfies the assumptions

$$
\begin{aligned}
\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta} &\leftrightarrow \mathbb{E}[\mathbf{u}|\mathbf{X}] = 0, \\
\mathrm{Var}(\mathbf{y}|\mathbf{X}) = \mathrm{Var}(\mathbf{u}|\mathbf{X}) &= \boldsymbol{\Omega},
\end{aligned}
$$

and $\mathbf{X}$ has full rank. Further assume that the variance of u_i is equal to σ_A^2 for observations $i = 1, 2, ..., n^A$ and equal to σ_B^2 for observations $i = n^A+1, n^A+2, ..., n$, that the covariance for error terms is zero for all observations, so that the $n \times n$ variance covariance matrix $\boldsymbol{\Omega}$ has the form

$$
\boldsymbol{\Omega} = \begin{bmatrix} \sigma_A^2 \mathbf{I}_{n^A} & \mathbf{O} \\ \mathbf{O} & \sigma_B^2 \mathbf{I}_{n-n^A} \end{bmatrix}.
$$

It is worth noting that we cannot divide the n observations into two samples and run separate regressions. For the second group of observations the matrix of explanatory variables, $\mathbf{X}$, is not full rank due to having two unit vectors. Thus, our results will rely on us pooling the two groups of regressors and attaining estimators. However, since X_{2i} is only there to help us distinguish between the two samples, we could simply drop it.

1.1

Explain whether or not the OLS estimator

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

is an unbiased estimator.

The OLS estimator is unbiased:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\
&= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u} \\
\mathbb{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \boldsymbol{\beta} + \mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{u}|\mathbf{X}],
\end{aligned}
$$

and by the law of iterated expectations

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}_{\beta|X}\left[\hat{\boldsymbol{\beta}}|\mathbf{X}\right]\right] &= \boldsymbol{\beta} + \mathbb{E}_X\left[\mathbb{E}_{u|X}\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}|\mathbf{X}\right]\right] \\
\mathbb{E}_X\left[\hat{\boldsymbol{\beta}}\right] &= \boldsymbol{\beta} + \mathbb{E}_X\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\underbrace{\mathbb{E}_{u|X}\left[\mathbf{u}|\mathbf{X}\right]}_{=0}\right] \\
\mathbb{E}_X\left[\hat{\boldsymbol{\beta}}\right] &= \boldsymbol{\beta}.
\end{aligned}$$

1.2

Consider the case in which σ_A^2 and σ_B^2 are known. Define the symmetry matrix $\boldsymbol{\Psi}$ to be

$$\boldsymbol{\Psi} = \begin{bmatrix} \frac{1}{\sigma_A}\mathbf{I}_A & \mathbf{O} \\ \mathbf{O} & \frac{1}{\sigma_B}\mathbf{I}_B \end{bmatrix},$$

and verify that $\boldsymbol{\Psi}^\top\boldsymbol{\Psi} = \boldsymbol{\Omega}^{-1}$. What is $\boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Psi}^\top$?

Since $\boldsymbol{\Psi}$ is symmetric, $\boldsymbol{\Psi} = \boldsymbol{\Psi}^\top$. This gives us:

$$\boldsymbol{\Psi}\boldsymbol{\Psi}^\top = \begin{bmatrix} \frac{1}{\sigma_A^2}\mathbf{I}_A & \mathbf{O} \\ \mathbf{O} & \frac{1}{\sigma_B^2}\mathbf{I}_B \end{bmatrix} = \boldsymbol{\Omega}^{-1}.$$

Thus, we have

$$\boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Psi}^\top = \boldsymbol{\Psi}(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top)^{-1}\boldsymbol{\Psi}^T = \mathbf{I}_n.$$

1.3

Define $y_i^* = y_i/\sigma_A$ for observations $i = 1, 2, ..., n^A$, and $y_i^* = y_i/\sigma_B$ for observations $i = n^A + 1, n^A + 2, ..., n$. Define $\mathbf{X}^*$ and $\mathbf{u}^*$ similarly.
What is $Var(u_i^*|\mathbf{X}^*)$ for observations $i = 1, 2, ..., n^A$?
What is $Var(u_i^*|\mathbf{X}^*)$ for observations $i = n^A + 1, n^A + 2, ..., n$?
Does the error term u_i^* in the model satisfy the assumption of conditional homoskedasticity?
What is $\mathbb{E}[\mathbf{y}^*|\mathbf{X}^*]$?
What is $Var(\mathbf{y}^*|\mathbf{X}^*)$?
Assuming that rank of $\mathbf{X}^*$ is 3, write down an efficient estimator of $\boldsymbol{\beta}$ in terms of the transformed variables $\mathbf{y}^*$ and $\mathbf{X}^*$. Express this estimator in terms of the original variables $\mathbf{y}$ and $\mathbf{X}$ and the variance matrix $\boldsymbol{\Omega}$.

The covariance matrix $\boldsymbol{\Omega}$ is non-singular, and therefore so is the matrix $\boldsymbol{\Psi}$ as well, and so the transformed regression model

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^* \iff \boldsymbol{\Psi}\mathbf{y} = \boldsymbol{\Psi}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Psi}\mathbf{u}, \tag{1}$$

is the same as the original model. The estimator is given by minimising the GLS criterion function

$$\arg\max_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \arg\max_{\boldsymbol{\beta}} \mathbf{u}^\top \boldsymbol{\Omega} \mathbf{u}$$
$$\frac{\partial \mathbf{u}^\top \boldsymbol{\Omega} \mathbf{u}}{\partial \boldsymbol{\beta}} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}} \frac{\partial \mathbf{u}^\top \boldsymbol{\Omega} \mathbf{u}}{\partial \mathbf{u}},$$

and by our first order condition we attain:

$$\begin{aligned} 0 &= 2\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{u} \\ &= 2\mathbf{X}^\top \boldsymbol{\Omega} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \implies \hat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{y} \iff (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} \mathbf{y}^*. \end{aligned} \tag{2}$$

We can show the covariance matrix of $\boldsymbol{\Psi}\mathbf{u}$:

$$\begin{aligned} \mathbb{E}[\boldsymbol{\Psi}\mathbf{u}\mathbf{u}^\top \boldsymbol{\Psi}^\top | \mathbf{X}] &= \boldsymbol{\Psi} \mathbb{E}[\mathbf{u}\mathbf{u}^\top | \mathbf{X}] \boldsymbol{\Psi}^\top \\ &= \boldsymbol{\Psi} \boldsymbol{\Omega} \boldsymbol{\Psi}^\top \\ &= \mathbf{I}_n, \end{aligned}$$

as proven above. It then follows that this transformed variance matrix satisfies conditional homoskedasticity, and for both samples A and B variance is unity. It then follows that

$$\begin{aligned} \mathbb{E}[\mathbf{y}^* | \mathbf{X}^*] &= \mathbf{X}^* \boldsymbol{\beta}, \\ \mathrm{Var}[\mathbf{y}^* | \mathbf{X}^*] &= \mathrm{Var}[\mathbf{u}^* | \mathbf{X}^*] = \mathbf{I}_n. \end{aligned}$$

Alternatively, we could prove this by the following

$$\mathrm{Var}\left(u_i^* | \mathbf{X}^*\right) = \begin{cases} \mathrm{Var}\left(\frac{u_i}{\sigma_A} | \mathbf{X}^*\right) = \frac{1}{\sigma_A^2} \mathrm{Var}\left(u_i | \mathbf{X}^*\right) = \frac{\sigma_A^2}{\sigma_A^2} = 1 & \forall i \in A, \\ \mathrm{Var}\left(\frac{u_i}{\sigma_B} | \mathbf{X}^*\right) = \frac{1}{\sigma_B^2} \mathrm{Var}\left(u_i | \mathbf{X}^*\right) = \frac{\sigma_B^2}{\sigma_B^2} = 1 & \forall i \in B, \end{cases}$$

and

$$\begin{aligned} \mathrm{Var}\left(\mathbf{y}^* | \mathbf{X}^*\right) &= \boldsymbol{\Psi} \mathrm{Var}\left(\mathbf{y} | \mathbf{X}\right) \boldsymbol{\Psi}^\top \\ &= \boldsymbol{\Psi} \boldsymbol{\Omega} \boldsymbol{\Psi}^\top \\ &= \mathbf{I}_n, \end{aligned}$$

since $\boldsymbol{\Psi}$ is not random.

1.4

Suppose further that $\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega})$. What is the distribution of $\mathbf{y}^*|\mathbf{X}^*$? What is the distribution of $\mathbf{u}^*|\mathbf{X}^*$?

Since we know the conditional expectation and variance of $\mathbf{y}^*$, we know that

$$\mathbf{y}^* | \mathbf{X}^* \overset{d}{\sim} N(\mathbf{X}^* \boldsymbol{\beta}, \mathbf{I}).$$

We also derived the information needed to determine the distribution of $\mathbf{u}^*|\mathbf{X}^*$:

$$\mathbf{u}^*|\mathbf{X}^* \overset{d}{\sim} N(\mathbf{0}, \mathbf{I}).$$

This implies that we can apply standard Gauss-Markov Theorem assumptions – all we needed was that the conditional expectation of our error variable was zero, that there was no serial correlation between our errors, and that errors were homoskedastic. Note, if we had serial correlation then $\mathbf{\Psi}$ would be a triangle matrix.

1.4.1

Still assuming that σ_A^2 and σ_B^2 are known, write down the log likelihood function $L(\boldsymbol{\beta})$. How does the (conditional) ML estimator of $\boldsymbol{\beta}$ in this model compare to the GLS estimator of $\boldsymbol{\beta}$?

To attain the conditional maximum likelihood estimator, begin by constructing the likelihood function

$$L_t(y_t^*, \boldsymbol{\beta}, \mathbf{I}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_t^* - \mathbf{X}_t^* \boldsymbol{\beta})^2}{2}\right), \tag{3}$$

and then taking logs:

$$l_t(y_t^*, \boldsymbol{\beta}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}(y_t^* - \mathbf{X}_t^* \boldsymbol{\beta})^2. \tag{4}$$

The t observations are assumed to be independent, so the log likelihood function is just the sum of these equations:

$$l(\mathbf{y}^*, \boldsymbol{\beta}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta})^\top(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}). \tag{5}$$

Recall the difference between a PDF and a likelihood function: a PDF maps the probability of observing a random variable for a given set of parameters, whereas a likelihood function gives the ‘probability’ of your parameters given the random variables that were observed. Differentiating the log likelihood function and using our first order condition gives

$$\frac{\partial l(\mathbf{y}^*, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{1}{2}\frac{\partial \mathbf{u}^\top \mathbf{u}}{\partial \boldsymbol{\beta}} = 0,$$

and after applying the chain rule of matrix calculus we attain:

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}^{*\top}\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{u}^*.$$

This conditional MLE coincides with the GLS estimator attained beforehand.

1.5

Now suppose σ_A^2 and σ_B^2 are not known. Suggest an estimator of σ_A^2 and σ_B^2 that can be calculated using the OLS residuals. Suggest a condition on the behaviour of the ration n_A/n as $n \to \infty$ that would be required to show that these estimators are consistent as $n \to \infty$.

Suppose σ_A^2 and σ_B^2 are not known. For the data that is generated by a DGP belonging to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbb{E}[\mathbf{u}|\mathbf{X}] = 0,\ \mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}] = \boldsymbol{\Omega},$$

the exogeneity assumption implies that $\hat{\boldsymbol{\beta}}$ is unbiased, which in no way depends on assumptions about the covariance matrix of the error terms. Whatever the form of the error covariance matrix $\boldsymbol{\Omega}$, the covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ is equal to:

$$\begin{aligned}\mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top|\mathbf{X}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}. \end{aligned} \tag{6}$$

We cannot evaluate the above expression, but we can estimate it. For the purposes of asymptotic theory, we want to consider the covariance matrix not of $\hat{\boldsymbol{\beta}}$, but of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Asymptotic covariance of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is

$$\underbrace{\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}}_{=\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}} \quad \underbrace{\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\boldsymbol{\Omega}\mathbf{X}}_{\text{Can be estimated by } \frac{1}{n}\mathbf{X}^\top\hat{\boldsymbol{\Omega}}\mathbf{X}} \quad \underbrace{\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}}_{=\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}},$$

where $\hat{\boldsymbol{\Omega}}$ is an inconsistent estimator with typical t-th element as $\hat{u}_t^2$. $\frac{1}{n}\mathbf{X}^\top\hat{\boldsymbol{\Omega}}\mathbf{X}$ is a $k \times k$ matrix and it has $\frac{1}{2}(k^2 + k)$ distinct elements independent of sample size, and where k is the number of columns of $\mathbf{X}$.[1]

For the A subsample, an estimator for the variance is given by

$$s_A^2 = \frac{1}{n_A - k}\sum_{t=1}^{n_A}\hat{u}^2_{\text{OLS},t},\ k = 3, \tag{7}$$

and for the B subsample,

$$s_B^2 = \frac{1}{n - n_A - k}\sum_{t=n_A+1}^{n}\hat{u}^2_{\text{OLS},t},\ k = 3. \tag{8}$$

A condition for the ratio n_A/n to ensure that both estimators remain consistent would be that the ratio remain constant and in between (0,1). Why? This is to ensure that if we are able to apply a LLN to the RHS of (7) and (8) that we able to evaluate a probability limit, showing that both estimators are consistent as their sample sizes tend to infinity.

1.6

How could you use these estimators of σ_A^2 and σ_B^2 to obtain a transformed version of the original model in which the feasible GLS estimator

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^\top\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

[1] Using this estimator and evaluating the sandwich equation (6) yields the heteroskedasticity-consistent covariance matrix estimator:

$$\hat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Omega}}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}.$$

can be computed as an OLS estimator using the transformed variables? Does this feasible GLS estimator coincide with the (conditional) ML estimator of $\boldsymbol{\beta}$ *obtained under the assumption that* $\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega})$*, in the case in which* σ_A^2 *and* σ_B^2 *(and hence* $\boldsymbol{\Omega}$*) are not known?*

We know that the GLS estimator is given by (2). Using our estimators for σ_A^2 and σ_B^2, rewrite (2) as

$$\hat{\boldsymbol{\beta}}_{\text{FGLS}} = (\mathbf{X}^\top \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Omega}}^{-1} \mathbf{u}.$$

We can do this as $\hat{\boldsymbol{\Omega}}$ is a diagonal matrix with t-th element $\frac{1}{\hat{\omega}_t^2}$, and $\hat{\boldsymbol{\Psi}}$ can be chosen as the diagonal matrix with t-th element $\frac{1}{s_t}$, where s_t^2 is corresponds to our error variance estimators. Then, for typical observation t:

$$s_t^{-1} y_t = s_t^{-1} \mathbf{X}_t \boldsymbol{\beta} + s_t^{-1} u_t,$$

which then implies

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{OLS}} &= (\mathbf{X}^\top \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Psi}}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Psi}}^\top \mathbf{y} \\
&= (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} \mathbf{y}^* \\
\hat{\boldsymbol{\beta}}_{\text{FGLS}} &= (\mathbf{X}^\top \hat{\boldsymbol{\Omega}} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Omega}} \mathbf{y}.
\end{aligned}$$

This FGLS estimator does not coincide with the conditional maximum likelihood estimator given by equations (3) and (5). GLS uses σ_A^2 and σ_B^2 not s_A^2 and s_B^2; also, even with unknown σ_A^2 and σ_B^2, MLE in this case will be different from FGLS due to the usage of different residuals.

2 Omitted variables

The data generation process for a scalar y_i is

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i,$$

with $\mathbb{E}[u_i] = 0$, in which the coefficients $\beta_k > 0$ for $k = 1, 2, 3, 4$ and $\mathbb{E}[x_{ki} u_i] = 0$ for $k = 2, 3, 4$. It is known that $Corr(x_{2i}, x_{4i}) > 0$ and that $Corr(x_{3i}, x_{4i}) > 0$, but also that in the linear projection of x_{4i} on x_{2i} and x_{3i}

$$x_{4i} = \delta_1 + \delta_2 x_{2i} + \delta_3 x_{3i} + v_i$$

the coefficients $\delta_2 > 0$ and $\delta_3 < 0$. Lacking data on the variable x_{4i}, a researcher estimates the model

$$y_i = \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \epsilon_i.$$

Assuming that a large sample with IID observations is used, explain whether the OLS estimator of the coefficient γ_2 in this model is likely to over-estimate or under-estimate the coefficient β_2 in the data generating process. Does the same apply to the OLS estimator of γ_3, interpreted as an estimator of β_3?

$\hat{\gamma}_2$ overestimates β_2. To see this we begin by multiplying the linear projection of x_{4i} on x_{2i} and x_{3i} with β_4:

$$\beta_4 x_{4i} = \beta_4 \delta_1 + \beta_4 \delta_2 x_{2i} + \beta_4 \delta_3 x_{3i} + \beta_4 v_i,$$

and then substitute into the equation depicting the true DGP:

$$y_i = \underbrace{(\beta_1 x_1 + \beta_4 \delta_1)}_{\gamma_1} + \underbrace{(\beta_2 + \delta_2 \beta_4)}_{\gamma_2} x_{2i} + \underbrace{(\beta_3 + \beta_4 \delta_4)}_{\gamma_3} x_{3i} + \underbrace{(\beta_4 v_i + u_i)}_{\epsilon_i}.$$

We know $\beta_2 > 0$, $\beta_4 > 0$, and $\delta_2 > 0$, and this implies that $\gamma_2 > \beta_2$.

By symmetry we can do the same exercise for γ_3 and β_3. We find that $\gamma_3 < \beta_3$, and hence we conclude that γ_3 underestimates β_3.

The proof uses a simply application of the LLN. Start with

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \implies \hat{\boldsymbol{\gamma}}_{\text{OLS}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= \boldsymbol{\gamma} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon} \\ \therefore \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon}, \end{aligned}$$

then, use LLN and take the probability limit of $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$:

$$\begin{aligned} \operatorname*{plim}_{n \to +\infty} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) &= \underbrace{\left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1}}_{\xrightarrow{p} \mathbf{S}^{-1}_{\mathbf{X}^\top \mathbf{X}}} \underbrace{\frac{1}{n} \mathbf{X}^\top \boldsymbol{\epsilon}}_{\xrightarrow{p} 0} \\ \implies \hat{\boldsymbol{\gamma}} &\xrightarrow{p} \boldsymbol{\gamma}. \end{aligned}$$

This of course implies then that

$$\begin{aligned} \hat{\gamma}_2 &\xrightarrow{p} \beta_2 + \delta_2 \beta_4 > \beta_2, \\ \hat{\gamma}_3 &\xrightarrow{p} \beta_3 + \delta_4 \beta_4 < \beta_4. \end{aligned}$$