Asymptotic Theory, Hypothesis Testing, and Linear Restrictions

1 Asymptotic inference

Question 1 relates to a linear model of the form

$$y_i = \beta_1 + x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i$$
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

with IID observations on $(\mathbf{y}, \mathbf{X})$ for $i = 1, 2, ..., n$, where y_i and u_i are scalar random variables and x_i is a $k \times 1$ random vector. You may assume that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{V})$ and that a consistent estimator $\hat{\mathbf{V}}$ of $\mathbf{V}$ is available so that we have $\hat{\boldsymbol{\beta}}_{\text{OLS}} \overset{a}{\sim} N(\boldsymbol{\beta}, \hat{\mathbf{V}}/n)$. Suppose we consider $\mathbf{X}$ as a $n \times k$ ($k = 4$) matrix partitioned as

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{\iota} & \mathbf{X}_2 & \mathbf{X}_3 & \mathbf{X}_4 \end{bmatrix},$$

where $\boldsymbol{\iota}$ is the unit vector.

Before proceeding, we need to go over the law of large numbers (LLN) and central limit theorem (CLT), as they will be vital in justifying our use of asymptotic tests rather than exact tests.

1.1 Law of large numbers and central limit theorem

Consider the following linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

where $\{u_t\}_{t=1}^n \sim IID(0, \sigma^2)$ and/or X_t is treated as predetermined with respect to the error term, i.e., $\mathbb{E}(u_t|X_t) = 0$; different from the case in which $\{u_t\}_{t=1}^n \sim NID(0, \sigma^2)$ and X_t is exogenous. Here, we cannot derive the exact distribution for the OLS estimator, nor for the statistical tests.

In order to obtain an approximation of the distribution of the test statistics, we will rely on the LLN and on the CLT. There are several variants of the LLN and the CLT. To make the exposition as simple as possible, here are two general propositions stated in White (2001).

Proposition (Law of Large Numbers): Given restrictions on the dependence, heterogeneity, and moments of a sequence of random variables $\{X_t\}_{t=1}^n$,

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{a} 0, \text{ as } n \to +\infty,$$

where

$$\bar{X}_n \equiv n^{-1} \sum_{t=1}^{n} X_t,$$
$$\bar{\mu}_n \equiv \mathbb{E}(\bar{X}_n).$$

In particular, if $\{X_t\}_{t=1}^n$ are independently distributed, such that $\mathbb{E}(X_t) = \mu$ with $\text{Var}(X_t) = \Sigma_t$ with $||\Sigma_t|| < +\infty,\ \forall t$,

$$\bar{X}_n - \mu \xrightarrow{a} 0, \text{ as } n \to +\infty.$$

In words, an independent random variable with finite variance will asymptotically tend to its expected value.

Proposition (Central Limit Theorem): Given restrictions on the dependence, heterogeneity, and moments of a sequence of random variables $\{X_t\}_{t=1}^n$,

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} (X_t - \mu_t) = n^{-\frac{1}{2}} (\bar{X}_n - \bar{\mu}_n) \xrightarrow{a} N(0, \mathbf{V}),$$

where

$$\mathbf{V} = \lim_{n \to +\infty} n^{-1} \text{Var} \left(\sum_{t=1}^{n} X_t \right).$$

In particular, if $\{X_t\}_{t=1}^n$ are independently distributed, such that $\mathbb{E}(X_t) = \mu$ with $\text{Var}(X_t) = \Sigma_t$ with $||\Sigma_t|| < +\infty,\ \forall t$,

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} (X_t - \mu) \overset{a}{\sim} N(0, \Sigma).$$

In words, an independent random variable with finite variance will asymptotically be distributed by normal distribution when we multiply its sample moments by n raised to the power of $-\frac{1}{2}$.

Now, consider (1) again, where we have $\mathbb{E}(u_t|\mathbf{X}_t) = 0$, $\mathbb{E}(u_t^2|\mathbf{X}_t) = \sigma^2$, and $\{u_t\}_{t=1}^n \sim IID(0, \sigma^2\mathbf{I})$. We know that

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

so

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{X}^\top \mathbf{u}.$$

Now, consider the variance of $n^{-\frac{1}{2}} \mathbf{X}^\top \mathbf{u}$,

$$\begin{aligned}
\text{Var}\left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \mathbf{u} \right) = \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_t^\top u_t \right) &= \frac{1}{n} \mathbb{E}\left(\sum_{t=1}^{n} u_t^2 \mathbf{X}_t^\top \mathbf{X}_t \right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left(u_t^2 \mathbf{X}_t^\top \mathbf{X}_t \right) = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\underbrace{\mathbb{E}\left(u_t^2 | \mathbf{X}_t \right)}_{\sigma^2} \mathbf{X}_t^\top \mathbf{X}_t \right] \\
&= \sigma^2 \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\mathbf{X}_t^\top \mathbf{X}_t \right].
\end{aligned}$$

Using the LLN and CLT we have:

$$\frac{1}{\sqrt{n}}(\mathbf{X}^\top \mathbf{u}) \overset{a}{\sim} N(0, \sigma^2 \mathbf{S}_{\mathbf{X}^\top\mathbf{X}}), \tag{2}$$

$$\operatorname*{plim}_{n\to+\infty} \frac{1}{n}\mathbf{X}^\top\mathbf{X} = \operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t = \mathbf{S}_{\mathbf{X}^\top\mathbf{X}}, \tag{3}$$

where

$$\mathbf{S}_{\mathbf{X}^\top\mathbf{X}} = \operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t.$$

Combining the above LLN and CLT results we derive:

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{n}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u} \\
&= \underbrace{\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}}_{\xrightarrow{p}\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}} \underbrace{\frac{1}{\sqrt{n}}\mathbf{X}^\top\mathbf{u}}_{\xrightarrow{d}N(0,\sigma^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}})} \\
\therefore \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\overset{a}{\sim} N(0, \sigma^2\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}).
\end{aligned}$$

1.1.1 The heteroskedastic case

We can extend this analysis for the heteroskedastic model. Consider

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbb{E}(u_t|\mathbf{X}_t) = 0,$$

where $\{u_t\}_{t=1}^n$ is a sequence of independent random variables (but not identically distributed!) with $\mathbb{E}(u_t^2|\mathbf{X}_t) = \sigma_t^2 < +\infty, \ \forall t$. As before, we have

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{n}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u} \\
&= \left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1} \frac{1}{\sqrt{n}}\mathbf{X}^\top\mathbf{u},
\end{aligned}$$

and using the LLN and CLT we obtain

$$\frac{1}{\sqrt{n}}(\mathbf{X}^\top \mathbf{u}) \overset{a}{\sim} N(0, \mathbf{S}_{\mathbf{X}^\top\boldsymbol{\Sigma}\mathbf{X}}), \tag{4}$$

where

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{n}}\mathbf{X}^\top\mathbf{u}\right) &= \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{X}_t^\top u_t\right) \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t u_t^2\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t\mathbb{E}\left[u_t^2|\mathbf{X}_t\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t\sigma_t^2,
\end{aligned}$$

and so we denote

$$\mathbf{S}_{\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X}} = \operatorname*{plim}_{n \to +\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{X}_t^\top \mathbf{X}_t \sigma_t^2. \tag{5}$$

Combining (3), (4), and (5), we get

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left(\frac{1}{n}\mathbf{X}^\top \mathbf{X}\right)^{-1} \frac{1}{\sqrt{n}} \mathbf{X}^\top \mathbf{u} \\ &= \underbrace{\left(\frac{1}{n}\mathbf{X}^\top \mathbf{X}\right)^{-1}}_{\xrightarrow{p} \mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1}} \underbrace{\frac{1}{\sqrt{n}} \mathbf{X}^\top \mathbf{u}}_{\xrightarrow{d} N(0, \mathbf{S}_{\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X}})} \\ \therefore \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\overset{a}{\sim} N(0, \mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X}} \mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1}). \end{aligned}$$

Note that $\mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X}} \mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1}$ is the limiting distribution of n^{-1} sandwich variance estimator.

1.2 Re-parameterisation

Consider the null hypothesis $H_0: \ \beta_3 = -\beta_2 \leftrightarrow \beta_2 + \beta_3 = 0$. Re-parameterise the model in such a way that $(\beta_2 + \beta_3)$ can be estimated directly as the coefficient on one of the explanatory in the re-parameterised model.

Let us write the linear regression model as

$$y_i = \beta_1 \iota_i + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i, \tag{6}$$

Then, since we wish to test that hypothesis that $\beta_3 = -\beta_2$, we can rewrite our model as

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

Alternatively, we could have added $\beta_3 x_{2i}$ to both LHS and RHS of (6). Rearranging our terms gives us

$$\begin{aligned} y_i &= \beta_1 \iota_i + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i + \beta_2 x_{3i} - \beta_2 x_{3i} \\ &= \beta_1 \iota_i + \beta_2 (x_{2i} - x_{3i}) + (\beta_2 + \beta_3) x_{3i} + \beta_4 x_{4i} + u_i \\ y_i &= \delta_i \iota_i + \delta_2 (x_{2i} - x_{3i}) + \delta_3 x_{3i} + \delta_4 x_{4i} + u_i, \end{aligned} \tag{7}$$

where

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

and our matrix $\mathbf{X}$ now has $\mathbf{X}_2 - \mathbf{X}_3$ as its second column vector. Thus, we can estimate $(\beta_2 + \beta_3)$ as the third row element of $\hat{\boldsymbol{\delta}}$.

1.3 Hypothesis testing (single restriction)

Describe a standard normal test statistic that could be used to test this null hypothesis against the two sided alternative (i.e. $H_1: \ \beta_3 \neq -\beta_2 \leftrightarrow \beta_2 + \beta_3 \neq 0$) in a large finite sample.

We start by stating both our null and alternative hypothesis:

$$H_0 : \beta_3 = -\beta_2 \leftrightarrow \beta_2 + \beta_3 = 0,$$
$$H_1 : \beta_3 \neq -\beta_2 \leftrightarrow \beta_2 + \beta_3 \neq 0.$$

Using our re-parameterised model (7), the above hypothesis is analogous to the following:

$$H_0 : \delta_3 = 0$$
$$H_1 : \delta_3 \neq 0$$

and our test statistic would be

$$t_{\delta_3} = \frac{\hat{\delta}_3}{SE(\hat{\delta}_3)} \overset{a}{\sim} N(0,1).$$

Our test statistic t_{δ_3} is asymptotically distributed by a standard normal distribution. Recall the sufficiency conditions for exact and asymptotic distributions:

Table 1: **Exact and Asymptotic Tests**

Errors	Small Sample	Large Sample
$\mathbf{u}\vert\mathbf{X} \sim N$	$t \overset{d}{\sim} t(n-K)$ and $F \overset{d}{\sim} F(p, n-K)$	$t \overset{d}{\sim} t(n-K) \overset{d}{\sim} N(0,1)$ and $F \overset{d}{\sim} F(p, n-K) \overset{d}{\sim} \frac{\chi_p^2}{p}$
$\mathbf{u}\vert\mathbf{X} \sim$IID	N/A	Using LLN/CLT: $t \overset{a}{\sim} N(0,1)$ and $F \overset{a}{\sim} \frac{\chi_p^2}{p}$

To understand why our test statistic asymptotically normally distributed, it would be convenient to rewrite our regressor and coefficient matrices.

Let

$$\mathbf{X} = \begin{bmatrix} \tilde{\mathbf{X}}_1 & \mathbf{X}_3 \end{bmatrix},$$

where $\tilde{\mathbf{X}}_1$ is an $n \times 3$ matrix:

$$\tilde{\mathbf{X}}_1 = \begin{bmatrix} \boldsymbol{\iota} & \mathbf{X}_2 - \mathbf{X}_3 & \mathbf{X}_4 \end{bmatrix}.$$

The vector of coefficients can also be written as:

$$\boldsymbol{\delta}^\top = \begin{bmatrix} \boldsymbol{\delta}_1^\top & \delta_3 \end{bmatrix},$$

where $\boldsymbol{\delta}_1$ is a 3×1 vector of coefficients:

$$\boldsymbol{\delta}_1 = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_4 \end{bmatrix}$$

Our model can then be rewritten as:

$$\mathbf{y} = \tilde{\mathbf{X}}\boldsymbol{\delta} + \mathbf{u}. \tag{8}$$

Then we could use the Frisch-Waugh-Lovell (FWL) theorem to get an OLS estimate of δ_3 using the following regression:

$$\mathbf{M}_1\mathbf{y} = \delta_3\mathbf{M}_1\mathbf{X}_3 + \boldsymbol{\epsilon},$$

where $\mathbf{M}_1$ is the elimination matrix defined as $\mathbf{M}_1 \equiv \mathbf{I} - \tilde{\mathbf{X}}_1(\tilde{\mathbf{X}}_1^\top\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1^\top$ which projects onto the subspace orthogonal to the $\tilde{\mathbf{X}}_1$ subspace, and where $\boldsymbol{\epsilon}$ is the vector of residuals which are generally different to $\mathbf{u}$. Assuming the model is correctly specified, the parameter estimate for δ_3 is given by the standard OLS formula:

$$\begin{aligned}\hat{\delta}_3 &= (\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{-1}\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y} \\ &= \frac{\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y}}{\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3},\end{aligned}$$

and with variance:

$$\mathrm{Var}(\hat{\delta}_3) = \sigma^2(\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{-1},$$

which will then yield a test statistic:

$$z_{\delta_3} \equiv \frac{\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y}}{\sigma(\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{1/2}}.$$

But the test statistic z_{δ_3} assumes that we know σ^2, which is unrealistic. Therefore we need an estimator for σ, s, the standard error from the regression (8), which is given by:

$$s^2 = \frac{\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}}{n-k},$$

giving the test statistic:

$$t_{\delta_3} = \frac{\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y}}{s(\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{1/2}} = \left(\frac{\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}}{n-k}\right)^{-1/2}\frac{\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y}}{(\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{1/2}}$$

We are almost ready to justify using an asymptotic normal distribution. Because our error terms are assumed to be IID, and not NID, we need to use a LLN and CLT to justify our asymptotic distribution when we have a large finite sample. So we rewrite the above test statistic as a function of quantities to which we can apply either the law of large numbers (LLN) or central limit theorem (CLT) to check its asymptotic distribution. So rewrite the above test statistic as:

$$t_{\delta_3} = \left(\frac{\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}}{n-k}\right)^{-1/2}\frac{n^{-1/2}\mathbf{X}_3^\top\mathbf{M}_1\mathbf{y}}{(n^{-1}\mathbf{X}_3^\top\mathbf{M}_1\mathbf{X}_3)^{1/2}}, \tag{9}$$

where the numerator and denominator of the second term on the RHS has been multiplied by $n^{-1/2}$. So long as the model has been correctly specified, we know that $s^2 \to \sigma^2$ as $n \to \infty$, which is equivalent to saying that the OLS error variance estimator s^2 is consistent. Combining this with the following: i) s^2 is $n/(n-k)$ times the average of the $\hat{u}_i^2$; ii) $\text{plim}\frac{n}{n-k} = 1$; and iii) the average of the $\hat{u}_i^2 \to \sigma^2$ by a LLN, it follows that asymptotically

$$\left(\frac{\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}}{n-k}\right)^{-1/2} \to \frac{1}{\sigma}.$$

When the data is generated by (8) with $\delta_3 = 0$ under the null hypothesis, we have $\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{u}$, and so

$$t_{\delta_3} = \left(\frac{\mathbf{y}^\top \mathbf{M}_\mathbf{X}\mathbf{y}}{n-k}\right)^{-1/2} \frac{n^{-1/2}\mathbf{X}_3^\top \mathbf{M}_1\mathbf{y}}{(n^{-1}\mathbf{X}_3^\top \mathbf{M}_1 \mathbf{X}_3)^{1/2}} = \frac{n^{-1/2}\mathbf{X}_3^\top \mathbf{M}_1\mathbf{u}}{\sigma(n^{-1}\mathbf{X}_3^\top \mathbf{M}_1 \mathbf{X}_3)^{1/2}}.$$

So long as regressors are exogenous, we can work conditionally on $\mathbf{X}$, which implies that the only stochastic part of the above expression is $\mathbf{u}$ in the numerator. The expectation of the numerator is zero, and its conditional variance is:

$$\mathbb{E}\left[\mathbf{X}_3\mathbf{M}_1\mathbf{u}\mathbf{u}^\top \mathbf{M}_1\mathbf{X}_3|\mathbf{X}\right] = \sigma^2\mathbf{X}_3\mathbf{M}_1\mathbf{X}_3,$$

and so it's evident that t_{δ_3} has mean 0 and variance 1, conditional on $\mathbf{X}$. But since these values do not rely on $\mathbf{X}$, these are the unconditional mean and variance of the asymptotic distribution of t_{δ_3}. Applying a CLT to the numerator of t_{δ_3}, the numerator must be asymptotically normally distributed, and so under the null hypothesis with exogenous regressors

$$t_{\delta_3} \overset{a}{\sim} N(0,1).$$

This concludes the justification for using the standard normal distribution for large samples in place of the finite sample t distribution. What critical value do we choose for the test at the 5% critical value? For the two sided test we pick $z = 1.96$. If our test statistic in absolute value is less than 1.96, we do not reject the null hypothesis.

A few final things to note: The unbiased estimator of the variance of $\hat{\delta}_3$, assuming that the residuals are homoskedastic is

$$\hat{\text{Var}}\left(\hat{\delta}_3\right) = \frac{s^2}{\mathbf{X}_3\mathbf{M}_1\mathbf{X}_3},$$

therefore we can rewrite (9) compactly as

$$t_{\delta_3} = \frac{\hat{\delta}_3 - \delta_3^0}{\sqrt{\hat{\text{Var}}\left(\hat{\delta}_3\right)}} \overset{a}{\sim} N(0,1). \tag{10}$$

Secondly, if we assume that the residuals are heteroskedastic, the asymptotic distribution of t_{δ_3} would be similar to (10), however, the estimator of the variance would be

$$\hat{\text{Var}}_h\left(\hat{\delta}_3\right) = (\mathbf{X}_3^\top \mathbf{M}_1\mathbf{X}_3)^{-1}\hat{\text{Var}}_h\left(\mathbf{X}_3\mathbf{M}_1\mathbf{u}\right)(\mathbf{X}_3^\top \mathbf{M}_1\mathbf{X}_3)^{-1},$$

where

$$\hat{\text{Var}}_h\left(\mathbf{X}_3\mathbf{M}_1\mathbf{u}\right) = \begin{bmatrix}-\mathbf{X}_3^\top \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} & 1\end{bmatrix} \underbrace{\begin{bmatrix}\mathbf{X}_1^\top \hat{\mathbf{\Sigma}}\mathbf{X}_1 & \mathbf{X}_1^\top \hat{\mathbf{\Sigma}}\mathbf{X}_3 \\ \mathbf{X}_1^\top \hat{\mathbf{\Sigma}}\mathbf{X}_3 & \mathbf{X}_3^\top \hat{\mathbf{\Sigma}}\mathbf{X}_3\end{bmatrix}}_{\mathbf{X}^\top \hat{\mathbf{\Sigma}}\mathbf{X}} \begin{bmatrix}-(\mathbf{X}_1^\top \mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X}_3^\top \\ 1\end{bmatrix},$$

and

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix}\hat{u}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & \ddots & \cdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{u}_n^2\end{bmatrix}. \tag{11}$$

1.4 Multiple restrictions

Consider the join null hypothesis $H_0: \ \beta_3 = -\beta_2$ and $\beta_3 = \beta_4 \leftrightarrow \beta_3 = -\beta_2 = \beta_4$. Write down the restricted model implied by this null hypothesis.

The restricted model can be written as

$$\begin{aligned}\mathbf{y} &= \beta_1 \boldsymbol{\iota} + \beta_2 \mathbf{X}_2 - \beta_2 \mathbf{X}_3 - \beta_2 \mathbf{X}_4 + \mathbf{u} \\ &= \beta_1 \boldsymbol{\iota} + \beta_2 (\mathbf{X}_2 - \mathbf{X}_3 - \mathbf{X}_4) + \mathbf{u}.\end{aligned}$$

Notice that the number of parameters to estimate has gone from 4 to 2.

1.4.1 Re-parameterisation with multiple restrictions

Re-parameterise the model in such a way that $(\beta_2 + \beta_3)$ can be estimated directly as the coefficient on one of the explanatory variables in the re-parameterised model.

Begin by adding $\beta_2 \mathbf{X}_3$ to the LHS and RHS of (6):

$$\mathbf{y} + \beta_2 \mathbf{X}_3 = \beta_1 \boldsymbol{\iota} + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3 + \beta_4 \mathbf{X}_4 + \mathbf{u} + \beta_2 \mathbf{X}_3,$$

and then add $\beta_2 \mathbf{X}_4$ to both the LHS and RHS of the above equation:

$$\mathbf{y} + \beta_2 \mathbf{X}_3 + \beta_2 \mathbf{X}_4 = \beta_1 \boldsymbol{\iota} + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3 + \beta_4 \mathbf{X}_4 + \mathbf{u} + \beta_2 \mathbf{X}_3 + \beta_2 \mathbf{X}_4$$

Rearranging the above equation will give us our re-parameterised model:

$$\mathbf{y} = \beta_1 \boldsymbol{\iota} + \beta_2 (\mathbf{X}_2 - \mathbf{X}_3 - \mathbf{X}_4) + (\beta_2 + \beta_3) \mathbf{X}_3 + (\beta_2 + \beta_4) \mathbf{X}_4 + \mathbf{u}. \tag{12}$$

Testing the exclusion of $\mathbf{X}_3$ is akin to testing if $\beta_2 + \beta_3 = 0$ or that $\beta_3 = -\beta_2$. Likewise, testing the exclusion of $\mathbf{X}_4$ is akin to testing $\beta_2 + \beta_4 = 0$ or that $\beta_4 = -\beta_2$. Testing the exclusion of both $\mathbf{X}_3$ and $\mathbf{X}_4$ tests two restrictions; that $\beta_3 = -\beta_2$ and $\beta_4 = -\beta_2$, or equivalently that $\beta_3 = -\beta_2 = \beta_4$.

As before, there were multiple ways to go about re-parameterising our initial linear model.

1.4.2 Hypothesis test

Describe a $\chi^2(2)$ test statistic that could be used to test this null hypothesis against the alternative hypothesis $H_1: \ \beta_3 \neq -\beta_2$ or $\beta_3 \neq \beta_4$ in a large finite sample.

Just as with our test of a single restriction, it's worth first formally writing our null and alternative hypotheses. This time, however, since we are testing multiple restrictions, we will write our hypotheses in terms of restriction matrices. To make this easier, let's first rewrite our model in (12) as

$$\begin{aligned}&\mathbf{y} = \theta_1 \boldsymbol{\iota} + \theta_2 (\mathbf{X}_2 - \mathbf{X}_3 - \mathbf{X}_4) + \theta_3 \mathbf{X}_3 + \theta_4 \mathbf{X}_4 + \mathbf{u} \\ &\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{u},\end{aligned}$$

where the second column vector $\mathbf{X}$ is $\mathbf{X}_2 - \mathbf{X}_3 - \mathbf{X}_4$. We can then setup our hypothesis test as

$$H_0: \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix},$$

where our restriction matrices are written in the form

$$\mathbf{r} = \mathbf{R}\boldsymbol{\theta}.$$

It's also worth noting that if we didn't bother with re-parameterisation, we could write our hypothesis test based on (6) as

$$H_0: \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}.$$

We now have to justify why we can use an asymptotic test distributed with a χ^2 distribution. Conditional on our model being correctly specified, and our assumptions in regards to the error terms are valid, then

$$\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, \sigma^2 \mathbf{R}\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}\mathbf{R}^\top),$$

since

$$\begin{aligned}
\text{Var}\left(\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) &= \text{Var}\left(\sqrt{n}\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}\right) \\
&= \mathbb{E}\left[\sqrt{n}\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}\mathbf{u}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\sqrt{n}\right] \\
&= n\mathbb{E}\left[\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\right] \\
&= n\sigma^2\mathbb{E}\left[\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\right] \\
&= \sigma^2\mathbf{R}\underbrace{\mathbb{E}\left[\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right]}_{\overset{p}{\to}\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}}\mathbf{R}^\top.
\end{aligned}$$

Recall our following theorem that if a random variable $\mathbf{x}$ is normally distributed, then we have

$$\mathbf{x} \overset{d}{\sim} N(\underset{n\times 1}{\mathbf{0}}, \boldsymbol{\Omega}) \implies \mathbf{x}^\top\boldsymbol{\Omega}^{-1}\mathbf{x} \overset{d}{\sim} \chi^2(n),$$

as the product of two normally distributed random variables is itself chi-square distributed. Therefore, since we have q restrictions here due to the rank of $\mathbf{R}$:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\mathbf{R}^\top\underbrace{\left[\sigma^2\mathbf{R}\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}\mathbf{R}^\top\right]^{-1}}_{\left[\text{Var}_\infty(\mathbf{R}\hat{\boldsymbol{\beta}})\right]^{-1}}\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \chi^2(q).$$

A consistent estimator of the asymptotic variance $\text{Var}_\infty(\mathbf{R}\hat{\boldsymbol{\beta}})$ is

$$\hat{\text{Var}}_\infty(\mathbf{R}\hat{\boldsymbol{\beta}}) = s^2\mathbf{R}\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{R}^\top, \tag{13}$$

therefore, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\mathbf{R}^\top\underbrace{\left[s^2\mathbf{R}\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{R}^\top\right]^{-1}}_{\left[\hat{\text{Var}}_\infty(\mathbf{R}\hat{\boldsymbol{\beta}})\right]^{-1}}\sqrt{n}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \chi^2(q), \tag{14}$$

or

$$F_{\mathbf{R}\boldsymbol{\beta}} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{R}^\top \left[\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\right]^{-1} \mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{q} \Big/ \frac{\hat{\mathbf{u}}\hat{\mathbf{u}}}{n-k} \overset{a}{\sim} \frac{1}{q}\chi^2(q). \tag{15}$$

Note, like the t-test, if we assume that the residuals are heteroskedastic, the asymptotic distribution of the F-stat would be similar to the one in (14), however, the estimator of the variance (13) would be replaced by

$$\hat{\text{Var}}_\infty(\mathbf{R}\hat{\boldsymbol{\beta}}) = n\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R},$$

where $\hat{\boldsymbol{\Sigma}}$ was defined in (11).

Returning back to our question, our test statistic is distributed as $\chi^2(2)$, and since our test statistic is non-negative, we only consider values from the upper tail of the distribution to be evidence against the validity of the null hypothesis. Values of the test statistic above 0.5×5.99 would lead us to reject the null hypothesis.

2 Use of dummy variables

(*See the problem set for a full description of the question*)

2.1

Begin by stating our hypothesis test:

$$H_0 : \beta = 0,$$
$$H_a : \beta \neq 0.$$

Our test statistic, t_β, is given by the following:

$$t_\beta = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \overset{a}{\sim} N(0,1),$$

since we do not assume normality of the error term, we must use the asymptotic theory to conduct our test. Although we are unsure about the dimensions of $\boldsymbol{\gamma}_S$ and $\boldsymbol{\gamma}_F$, since we have a relatively large sample size ($n = 5000$), we know from the previous question that our test statistic will be asymptotically distributed as a standard normal distribution, $N(0,1)$. So we need no concern ourselves with the exact t distribution.

Our test statistic is $t_\beta = -1.54$, which does not fall within our rejection region ($|t_\beta| \geq 1.96$ at the 5% significance level), and so we are unable to reject the null hypothesis. This suggests some kind of [negative] correlation between a student's academic performance and class size, although this is not statistically significant.

2.2

Once we introduce gender dummy variables into our model we have coefficients α^g, β^g, $g \in \{B, G\}$. What do these mean? α^B and α^G are intercept coefficients for the male and female dummy variables, respectively. Positive values imply a positive correlation between gender and academic performance. Similarly, β^g, $g \in \{B, G\}$ are coefficients for the interaction term between class size and gender. Positive values imply that there is a positive correlation between class size and academic performance for the gender g. Negative values suggest a negative correlation.

2.3

The results obtained by the researcher:

$$\hat{\beta}^B = -0.009, \; SE(\hat{\beta}^B) = 0.010,$$
$$\hat{\beta}^G = -0.042, \; SE(\hat{\beta}^G) = 0.012,$$

suggest that there exists a negative correlation between academic performance and class size for both genders, with the effect being stronger for girls than boys. The fact that these coefficients are both negative is inline with our previous estimate of β.

Note: It is worth noting that a 5% significance level, $\hat{\beta}^B$ is statistically insignificant (cannot reject a null hypothesis that $\beta^B = 0$) and $\hat{\beta}^G$ is statistically significant (reject a null hypothesis that $\beta^G = 0$).

2.4

We now wish to test if the β^g coefficients are equivalent. Our test is

$$H_0 : \beta^B = \beta^G,$$
$$H_a : \beta^B \neq \beta^G,$$

and much like the previous question, we can define a parameter, $\delta = \beta^B - \beta^G$, and rewrite the model as (after adding $\beta^G(D_i CS_i)$ to the LHS and RHS of the equation):

$$A_i = \alpha^B D_i + \alpha^G(1 - D_i) + (\beta^B - \beta^G)(D_i CS_i) + \beta^G CS_i + (\mathbf{x}_i^S)^\top \boldsymbol{\gamma}_S + (\mathbf{x}_i^F)^\top \boldsymbol{\gamma}_F + u_i,$$

and run the following test:

$$H_0 : \delta = 0,$$
$$H_a : \delta \neq 0.$$

Our test statistic is:

$$t_\delta = \frac{\hat{\delta} - \delta}{SE(\hat{\delta})} \overset{a}{\sim} N(0, 1),$$

and we compute the test statistic to be $t_\delta = 2.0625$, and therefore reject the null hypothesis at the 5% significance level. This suggests that there may be correlation between gender, class size, and academic performance. Since we defined $\delta = \beta^B - \beta^G$, this then implies that the detrimental effect of class size on academic performance for females is greater than for males.

v)

Additional dummy variable regressors we could include would be the effect of social economic and/or family factors on academic performance for males and females. For example, females from lower socio economic backgrounds may have lower academic performance than their male counterparts because they may be expected to help the household more. Another example could be that males from single parent households may perform worse than their single parent female counterparts.

3 Consistent estimation of the error variance

Consider the classic linear regression model

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

where the data on $\mathbf{y}$ and $\mathbf{X}$ are IID, with $\mathbb{E}[u_i|\mathbf{X}_i] = 0$ and $\mathbb{E}[u_i^2|\mathbf{X}_i] = \sigma^2$, and the $k \times k$ matrix $\mathbb{E}[\mathbf{X}^\top \mathbf{X}]$ exists and is non-singular. You may use the results that $\mathbb{E}[\mathbf{u}] = 0$ and $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$ in this setting.

3.1

What is the unconditional variance of $\mathbf{u}$*?*

The law of iterated expectations (LIE) states:

$$\mathbb{E}_B[\mathbb{E}_{A|B}[A|B]] = \mathbb{E}[A].$$

So if

$$\mathbb{E}[\mathbf{u}|\mathbf{X}] = 0 \implies \mathbb{E}[\mathbf{u}] = 0.$$

As we have

$$\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}] = \sigma^2\mathbf{I},$$

by LIE:

$$\begin{aligned}\mathbb{E}[\mathbf{u}\mathbf{u}^\top] &= \mathbb{E}_X[\mathbb{E}_{u|X}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]] = \mathbb{E}_X[\sigma^2] \\ &= \sigma^2\mathbf{I}.\end{aligned}$$

Alternatively, by the law of total variance:

$$\begin{aligned}\mathrm{Var}[\mathbf{u}] &= \mathbb{E}[\mathrm{Var}[\mathbf{u}|\mathbf{X}]] + \mathrm{Var}[\mathbb{E}[\mathbf{u}|\mathbf{X}]] \\ &= \mathbb{E}[\sigma^2\mathbf{I}] + \mathrm{Var}[0] \\ &= \sigma^2\mathbf{I}.\end{aligned}$$

3.2

What is the plim of $\frac{1}{n}\sum u_i^2$*?*

By the LLN under IID and a finite mean:

$$\text{plim } \frac{1}{n}\sum_{i=1}^{n} u_i^2 = \mathbb{E}[u_i^2] = \sigma^2.$$

3.3

Prove that $\mathbf{u} = \hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$.

Start with

$$\begin{aligned}\mathbf{u} &= \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta} + \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).\end{aligned}$$

3.4

For any $k \times 1$ vector b, the scalar $\mathbf{X}b = (\mathbf{X}b)^\top = b^\top \mathbf{X}^\top$. Use this to express the squared error term $\mathbf{u}\mathbf{u}^\top$ in the form

$$\mathbf{u}\mathbf{u}^\top = \hat{\mathbf{u}}\hat{\mathbf{u}}^\top + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \hat{\mathbf{u}}$$

Start with $\mathbf{u}\mathbf{u}^\top = (\hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}))(\hat{\mathbf{u}} - \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}))^\top$, and expand to get

$$\begin{aligned}
&\hat{\mathbf{u}}\hat{\mathbf{u}} - (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \hat{\mathbf{u}} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \hat{\mathbf{u}}\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
=&\hat{\mathbf{u}}\hat{\mathbf{u}} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \hat{\mathbf{u}}.
\end{aligned}$$

3.5

Summing both sides of this expression and dividing by the sample size n gives the relation

$$\frac{1}{n}\sum_{i=1}^n u_i^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i \right) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top \hat{u}_i$$

Use properties of the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS}$ an the OLS residuals $\hat{u}_i$ to find the probability limits of the product terms on the right-hand side of this expression.

We know by the LLN (and despite not being explicitly stated, the Slutsky Theorem/Central Mapping Theorem):

$$\begin{aligned}
\operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n u_i^2 = &\underbrace{\operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2}_{=\sigma^2} + \underbrace{\operatorname*{plim}_{n\to+\infty} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top}_{=0} \underbrace{\left(\operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i \right)}_{=\mathbf{S}_{\mathbf{X}^\top \mathbf{X}}} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
&- 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \underbrace{\operatorname*{plim}_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top \hat{u}_i}_{=0}.
\end{aligned}$$

3.6

What does this tell you about the probability limits of the estimators

$$\begin{aligned}
\hat{\sigma}^2_{\text{ML}} &= \frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^\top}{n}, \\
\hat{\sigma}^2_{\text{OLS}} &= \frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^\top}{n-k}.
\end{aligned}$$

This tells us that

$$\operatorname*{plim}_{n\to+\infty} \hat{\sigma}^2_{\text{ML}} = \operatorname*{plim}_{n\to+\infty} \frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^\top}{n} = \operatorname*{plim}_{n\to+\infty} \frac{\mathbf{u}\mathbf{u}^\top}{n} = \sigma^2,$$

and we know that

$$\hat{\sigma}^2_{\text{OLS}} = \frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^\top}{n-k} = \frac{n}{n-k}\frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^\top}{n} = \frac{n}{n-k}\hat{\sigma}^2_{\text{ML}},$$

and since the probability limit of $n/(n-k) = \lim_{n\to+\infty} \frac{n}{n-k} = 1$, then using Slutsky's Theorem/CMT, we also have

$$\operatorname*{plim}_{n\to+\infty} \hat{\sigma}^2_{\text{OLS}} = 1 \times \operatorname*{plim}_{n\to+\infty} \hat{\sigma}^2_{\text{ML}} = \sigma^2.$$

Both estimators are consistent estimators of σ^2 with conditional homoskedasticity.