Foundations of Econometrics - Part I
Sample Exam Solutions

SampleExam1

December 7, 2025

Question 1

Jha & Sarangi (2017) article "Does Social Media Reduce Corruption?" study the effect of internet and social media penetration on corruption using cross-country analysis. Among others, the authors consider the following regression:

$$CI = \beta_1 + \beta_2 fbpen + \beta_3 netpen + \beta_4 prights + \epsilon$$

where CI = country corruption index (the higher CI, the more corrupt the country is), $fbpen$ = % of facebook users (as a proxy of social media penetration), $netpen$ = internet penetration (%), $prights$ = political rights index, index varying from 1 (best political rights) to 7 (worse).

OLS estimation output (Stata):

```
Source       SS            df      MS               Number of obs = 177
                                                     F(3, 173)     = 129.52
Model        121.351283     3      40.4504276      Prob > F      = 0.0000
Residual     54.0283574    173     .312302644      R-squared     = 0.6919
                                                     Adj R-squared = 0.6866
Total        175.37964     176     .996475229      Root MSE      = .55884

CI           Coef.         Std. Err.     t       P>|t|     [95% Conf. Interval]
fbpen       -.0101358      .004374     -2.32               -.018769   -.0015025
netpen      -.0167164      .0026357    -6.34    0.000     -.0219186  -.0115142
prights     -.1326323      .0245707    -5.40    0.000     -.1811291  -.0841354
_cons        .4844418      .136533      3.55    0.000      .2149569   .7539267
```

Part (a)

Consider that in the regression above, parameter vector $\boldsymbol{\beta}$, including all the regression coefficients, was defined as:

$$\boldsymbol{\beta} = \left[\mathbb{E}(xx')\right]^{-1} \mathbb{E}(xy),$$

with x being the column vector including all the regressors, including the constant, and y the dependent variable. Keeping this definition in mind:

(i) What would be the dimension of element $\boldsymbol{\beta}$, $\mathbb{E}(xx')$, and $\mathbb{E}(xy)$?

Answer:

(ii) Provide one reason why the authors could be interested in estimating vector $\boldsymbol{\beta}$ as just defined. Justify.

Answer:
The authors could be interested in estimating $\boldsymbol{\beta}$ because it provides the best linear approximation to the conditional expectation function (CEF). Specifically, $\boldsymbol{\beta} = \arg\min_b \mathbb{E}[(E(y|x) - x'b)^2]$ minimizes the expected squared distance between the true CEF $\mathbb{E}(y|x)$ and the linear predictor $x'b$.
This is valuable for descriptive purposes: even if the true relationship between y and x is nonlinear, $\boldsymbol{\beta}$ gives us the best linear summary of that relationship. Moreover, this $\boldsymbol{\beta}$ is also the solution to $\min \mathbb{E}[(y - x'b)^2]$, meaning it provides the best linear prediction of y given the regressors x, making it useful for forecasting and interpretation.

Part (b)

Consider running a regression of the OLS residuals from estimating the model above with respect to a constant, *fbpen*, *netpen* and *prights*. What would you expect the coefficient of determination of running this regression to be? Justify.

Answer:
I would expect the R^2 to be zero. By construction, the OLS residuals are orthogonal to all regressors. There is no linear relationship left to explain by the regressors.

Part (c)

We want to test $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$, using the exact t-test statistic.

(i) Provide the expression of the exact t-statistic and its assumed distribution under H_0.

Answer:

$$t = \frac{\hat{\beta}_k - r}{\text{se}(\hat{\beta}_k)}$$

$$t \sim t_{n-K} \quad \text{under } H_0$$

(ii) Perform test using a 1% significance level? Justify. (Useful information below.)

Answer:
Hypothesis:

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

This is a **two-sided test**. We use the given 1% significance level, $\alpha = 0.01$.
Critical Value Determination:

- The significance level must be split into two tails:

$$\frac{\alpha}{2} = \frac{0.01}{2} = 0.005 \quad \text{in each tail}$$

- We are given the critical probability from the t-distribution:

$$P(t > 2.6) = 0.005$$

- Thus, the critical value is $t_{\text{critical}} = 2.6$.

Decision Rule Justification:
Based on the critical value, the decision rule for the two-sided test is:

- **Reject** H_0 if $|t| > 2.6$
- **Fail to reject** H_0 if $|t| \leq 2.6$

Justification: Since $\alpha = 0.01$ and $P(t > 2.6) = 0.005$, the two critical values defining the rejection region are ± 2.6. Any calculated t-statistic that falls outside this range (i.e., $|t| > 2.6$) has a p-value less than 0.01 and leads to the rejection of the null hypothesis.
Test Calculation:
From the output, we have $\hat{\beta}_2 = -0.0101358$ and $\text{se}(\hat{\beta}_2) = 0.004374$.

$$t = \frac{-0.0101358 - 0}{0.004374} = -2.32$$

Since $|t| = 2.32 < 2.6$, we **fail to reject** H_0 at the 1% significance level.

(iii) In the output above, the p-value associated with this test is missing. Given your answers to the previous two questions, what can you say about the size of the missing p-value? Be as specific as possible.

Needed information regarding $t \sim t(173)$: $P(t > 1.29) = 0.1$, $P(t > 1.65) = 0.05$, $P(t > 1.97) = 0.025$, $P(t > 2.35) = 0.01$, $P(t > 2.6) = 0.005$.

Answer:
Since we fail to reject at the 1% level (because $|t| = 2.32 < 2.6$), we know that the p-value must be greater than 0.01.
More specifically, since $2.32 < 2.35$, the p-value is greater than 0.01. Since $2.32 > 1.97$, the p-value is less than 0.05 (for a two-sided test, $P(|t| > 1.97) = 0.05$).
Therefore, $\boxed{0.01 < p\text{-value} < 0.05}$.

Part (d)

List the assumptions, regarding the dgp behind the data, that would be needed to justify the use of the exact t-test statistic. Label each assumption and next to each label, provide the corresponding statistical expression that defines the assumption, considering regressors are stochastic. No need to justify.

Answer:

A1: **Linearity** $y = X\beta + e$

A2: **Strict Exogeneity** $\mathbb{E}(e|X) = 0$

A3: **No Perfect Collinearity** $\text{rank}(X) = K$ (i.e., no linear relationship between regressors)

A4: **Conditional Homoskedasticity** $\text{Var}(e|X) = \sigma^2 I$

A5: **No Conditional Autocorrelation** $\text{Cov}(e_i, e_j|X) = 0$ for $i \neq j$

A6: **Normality** $e|X \sim N(0, \sigma^2 I)$

Part (e)

Consider testing whether $fbpen$ and $netpen$ are jointly significant at 1% significance level, using the exact F-test statistic, expressed as:

$$F = \frac{\left(R\hat{\beta} - r\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\beta} - r\right)/q}{SSE/(n-K)} \sim_{H_0} F(q, n-K).$$

(i) Detail the null and alternative hypotheses associated with this test.

Answer:
The hypotheses for the joint test are:

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_1 : \beta_2 \neq 0 \quad \text{OR} \quad \beta_3 \neq 0$$

(ii) Detail the exact values of the following 4 elements: q, r, R and SSE. No need to justify.

Answer:

- **q** (Number of restrictions): $q = 2$ (The restrictions are $\beta_2 = 0$ and $\beta_3 = 0$)
- **r** (Vector of values under H_0):

$$\mathbf{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- **R** (Restriction matrix): Since $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$, the $q \times K$ restriction matrix $\mathbf{R}$ (where $K = 4$) is:

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

- **SSE** (Sum of Squared Errors): $SSE = 54.0283574$

(iii) Draw the acceptance and rejection region associated with this test. Critical value is not available, but label it in the graph, using the notation we used in the course to properly identify it. Label both axes.

Answer:
[The graph should show an F-distribution with density on the y-axis and F-statistic values on the x-axis. The acceptance region is from 0 to $F_\alpha(q, n-K)$, and the rejection region is from $F_\alpha(q, n-K)$ to infinity, shaded on the right tail.]

(iv) Write the expression that defines the critical value associated with this test.

Answer:
$F_{0.01}(2, 173)$, which is the value c such that $P(F > c) = \alpha = 0.01$

(v) Provide the intuition of the location of the acceptance region you drew in (iii).

Answer:
A small F indicates the restricted model (with $\beta_2 = \beta_3 = 0$) fits as well as the unrestricted model, suggesting the restrictions are valid. A large F indicates the restricted model fits worse, providing evidence against $\beta_2 = \beta_3 = 0$. The acceptance region shows which part of the distribution allows us to fail to reject the null hypothesis that $\beta_2 = \beta_3 = 0$.

(vi) Knowing that the F-value $= 82.13$, detail what additional information you would need to finish the test, and how you would use it.

Answer:
We need the critical value $F_{0.01}(2, 173)$ to determine significance. If $82.13 > F_{0.01}(2, 173)$, we reject the null hypothesis; otherwise, we fail to reject. Given that $F = 82.13$ is a very large value, it is highly likely to exceed any reasonable critical value, so we would almost certainly reject H_0 and conclude that *fbpen* and *netpen* are jointly significant.

Part (f)

If all so-called classical assumptions, except for normality, were holding:

(i) What would the distribution of the test statistic used in 1c be? Provide the expression of the test statistic and its distribution.

Answer:
The expression:

$$t = \frac{\hat{\beta}_2 - 0}{\text{se}(\hat{\beta}_2)}$$

Under the null hypothesis, asymptotically:

$$t \overset{a}{\sim} N(0, 1)$$

(ii) Would the failure of normality affect the size of the p-value you drew in 1c(iii)? If so, how? Which one would be larger? Rigorously justify.

Answer:
Yes, the failure of normality would affect the p-value in finite samples. The exact t-test relies on the assumption that errors are normally distributed. When this assumption is violated, the actual sampling distribution of the t-statistic differs from the theoretical t-distribution in finite samples (though it converges asymptotically to normal by the CLT).
Typically, with non-normal errors (especially heavy-tailed distributions), the true p-value would be larger than the one calculated assuming normality, because extreme values occur more frequently than the t-distribution predicts. This means we would be overstating the significance of our test result, potentially leading to more Type I errors than the nominal $\alpha = 0.01$ level suggests. However, this effect diminishes as sample size increases due to the Central Limit Theorem.

Part (g)

Consider the following elements: $\hat{\beta}_2$, $\text{se}(\hat{\beta}_2)$, R^2, and 95% confidence interval for β_2, whose values are reported in the output above. Consider assumption of conditional homoskedasticity did not apply.

(i) Which of the 4 elements listed would require adjustment in calculating its value and which would not? Just state Yes ('adjustment required') or No ('adjustment not required').

Answer:

- $\hat{\beta}_2$: No adjustment required
- $\text{se}(\hat{\beta}_2)$: Adjustment required
- R^2: No adjustment required
- 95% CI for β_2: Adjustment required

(ii) Select one element that would not require adjustment and justify why you selected it.

Answer:
$\hat{\beta}_2$ would not require adjustment. This is the OLS estimator of the regression coefficient. Under the Gauss-Markov assumptions (linearity, strict exogeneity, and no perfect collinearity), the OLS estimator remains unbiased and consistent even with heteroskedasticity. Heteroskedasticity affects the variance of the estimator, but it does not change the point estimate itself, which continues to be the best linear unbiased estimator (BLUE) under homoskedasticity and remains unbiased under heteroskedasticity.

(iii) Select one element that would require adjustment and justify why you selected it.

Answer:
$\text{se}(\hat{\beta}_2)$ would require adjustment. The standard error is calculated using the formula $\text{se}(\hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_2|X)}$. Under homoskedasticity, $\text{Var}(\hat{\beta}_2|X) = \sigma^2[(X'X)^{-1}]_{22}$. However, with heteroskedasticity, the variance formula changes to $\text{Var}(\hat{\beta}_2|X) = [(X'X)^{-1}X'\Omega X(X'X)^{-1}]_{22}$, where Ω is a diagonal matrix with non-constant error variances. Therefore, heteroskedasticity-robust standard errors (White/Huber standard errors) must be used to obtain valid inference.

Part (h)

(i) Under what condition could we give the OLS estimates of β_2 and β_3 a causal interpretation? Rigorously explain, using the help of a causal path diagram.

Answer:
We can give a causal interpretation to the OLS estimates of β_2 and β_3 if and only if there is no omitted variable bias and no reverse causality.
Omitted Variable Bias: For causal interpretation, we need $\mathbb{E}[\epsilon|fbpen, netpen, prights] = 0$. This requires that all confounders—variables that affect both the treatment variables ($fbpen$, $netpen$) and the outcome (CI)—are included in the regression.
Example of Confounding: Suppose economic development is omitted from the model. Economic development could simultaneously:

- Increase social media penetration ($fbpen$) and internet penetration ($netpen$) as wealthier countries have better technology infrastructure
- Decrease corruption (CI) as developed economies typically have stronger institutions

This creates spurious correlation. The causal path diagram would show:

$$\text{Economic Development} \rightarrow \begin{cases} fbpen, netpen \\ CI \end{cases}$$

Without controlling for economic development, we would attribute its effect on corruption to social media and internet penetration, biasing our estimates.
Reverse Causality: We also need to ensure that corruption does not cause changes in social media or internet penetration. For instance, if corrupt governments restrict internet access, we would observe reverse causality: $CI \rightarrow netpen$.
Conclusion: Only under the assumption of no omitted confounders and no reverse causality can we interpret β_2 and β_3 as causal effects.

(ii) Does the value of R^2 play any role in helping us determining whether we can use the estimates of β_2 and β_3 for causal inference? And the size of the sample? Rigorously discuss.

Answer:
R^2 **and Causality:** The value of R^2 does **not** play any role in determining causality. R^2 measures the goodness of fit—the proportion of variance in the dependent variable explained by the regressors. However, a high R^2 does not imply causation; it only indicates strong correlation or predictive power. Even with $R^2 = 1$, the relationship could be entirely spurious due to omitted variable bias or reverse causality. Conversely, a low R^2 does not preclude a causal relationship; it simply means other factors also influence the outcome. Causality depends on the validity of the strict exogeneity assumption $\mathbb{E}[\epsilon|X] = 0$, not on model fit.
Sample Size and Causality: The size of the sample also does **not** determine whether we can make causal inferences. Increasing the sample size improves the precision of our estimates (reduces standard errors) and ensures consistency, but it does not eliminate bias. If omitted variable bias or reverse causality exists, the estimates will remain biased regardless of sample size. As $n \to \infty$, the OLS estimator converges to the probability limit, which equals the true parameter plus the bias term. Therefore, a large sample gives us a precise estimate of a potentially wrong (biased) quantity.
Conclusion: Neither R^2 nor sample size addresses the fundamental requirements for causal inference: exogeneity and absence of confounding. These are assumptions about the data generating process, not properties of the estimation procedure or sample.

Question 2

Consider the following dgp:

$$y_i = 1 + 1 \cdot x_{i2} + 1 \cdot x_{i3} + \epsilon_i \quad \epsilon_i|X \sim \text{i.i.}N(0, 64)$$

$$x_{i2} \sim \text{i.i.}U[0, 20] \quad x_{i3} = x_{i2} + v_i \quad v_i \sim \text{i.i.}N(0, 4)$$

Using this dgp, we generated 10,000 samples of 50 observations each ($n = 50$), keeping the 50 observations of regressors x_2 and x_3 the same across all samples. With each of the generated samples, we estimated by OLS the following regression:

$$y = \beta_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \epsilon.$$

With the OLS estimates of parameter β_2 the following density histogram was produced.
[Histogram would be displayed here showing distribution of OLS estimates]

Part (a)

"The histogram above illustrates the concept of the conditional sampling distribution of the OLS estimator under classical assumptions". Would you agree with this statement? Rigorously discuss.

Answer:
I would **partially agree** with this statement, but with important caveats.
What the histogram does illustrate: The histogram does illustrate the conditional sampling distribution of $\hat{\beta}_2$ given X, because the regressors x_2 and x_3 are held fixed across all 10,000 samples. This is the definition of conditional sampling distribution: the distribution of the estimator for a fixed realization of X. The histogram appears to demonstrate normality of the estimator, which is consistent with the classical assumption A6 (normality of errors).
What the histogram does NOT illustrate: However, the histogram alone cannot verify all classical assumptions. Specifically:

- The histogram cannot tell us about **homoskedasticity** (constant variance of errors conditional on X). We would need to examine whether the spread of the distribution remains constant.
- The histogram cannot verify **no autocorrelation** of error terms.
- The histogram shows the distribution is centered (suggesting unbiasedness), but doesn't directly verify **strict exogeneity** $\mathbb{E}[\epsilon|X] = 0$.

Conclusion: The histogram illustrates normality and the conditional nature of the sampling distribution, which are part of the classical assumptions, but it does not illustrate or verify the complete set of classical assumptions required for exact inference.

Part (b)

Using the described simulation exercise, (i) identify a property of OLS estimator that can be illustrated and (ii) a property that cannot be illustrated. Rigorously justify both, by providing the expression that defines the property and how you can, or cannot, illustrate it with this experiment.

Answer:
(i) Property that CAN be illustrated: Unbiasedness
The property of unbiasedness states that $\mathbb{E}[\hat{\beta}_2|X] = \beta_2 = 1$.
How we can illustrate it: With 10,000 samples, we can compute the average of all $\hat{\beta}_2$ estimates:

$$\bar{\hat{\beta}}_2 = \frac{1}{10000} \sum_{s=1}^{10000} \hat{\beta}_2^{(s)}$$

By the Law of Large Numbers, $\bar{\hat{\beta}}_2 \to \mathbb{E}[\hat{\beta}_2|X]$ as the number of simulations increases. If the histogram is centered around the true value $\beta_2 = 1$, this demonstrates that $\mathbb{E}[\hat{\beta}_2|X] = 1$, confirming unbiasedness.
(ii) Property that CANNOT be illustrated: Consistency
The property of consistency states that $\hat{\beta}_2 \xrightarrow{p} \beta_2$ as $n \to \infty$.
Why we cannot illustrate it: Consistency is a large-sample (asymptotic) property that requires examining the behavior of the estimator as the sample size n increases to infinity. In this simulation, we fix $n = 50$ across all samples. While we generate many samples (10,000), each sample has only 50 observations. The histogram shows the sampling distribution for a fixed sample size, not the limiting behavior as $n \to \infty$.
How we would need to illustrate it: To illustrate consistency, we would need to generate samples with varying sample sizes (e.g., $n = 50, 100, 500, 1000, 5000$) and show that the variance of $\hat{\beta}_2$ decreases and its distribution collapses around the true value as n increases.

Part (c)

If with each of the generated samples we calculated the 99% confidence interval for parameter β_2, how many of these intervals would you expect not to include a 1? Justify.

Answer:
We would expect approximately **100 intervals** (or 1% of 10,000) not to include the true value $\beta_2 = 1$.
Justification: A 99% confidence interval has a coverage probability of 0.99, meaning that if we construct the interval using the correct procedure (which requires the classical assumptions to hold), 99% of the intervals constructed from repeated samples will contain the true parameter value. Equivalently, 1% of the intervals will fail to contain the true value.
Since we are generating 10,000 samples:

$$\text{Expected number of intervals not containing } \beta_2 = 10000 \times 0.01 = 100$$

This is the expected number based on the definition of a confidence interval. The actual number in any finite simulation will vary slightly due to randomness, but should be close to 100.

Part (d)

Recall the expression of the variance decomposition of OLS estimator under Gauss-Markov assumptions:

$$\text{var}(\hat{\beta}_2|X) = \sigma^2 \cdot \frac{1}{SST_2} \cdot \frac{1}{1 - R_2^2}.$$

(i) Change one of the elements of the dgp above to create perfect collinearity. Clearly indicate the element that you changed and how. Using the expression of the variance decomposition provided, rigorously argue why we would not be able to uniquely estimate β_2. Additionally, describe how the histogram provided above would look like.

Answer:
Element changed:
Change $x_{i3} = x_{i2} + v_i$ to $x_{i3} = x_{i2}$ (i.e., set $v_i = 0$ for all i, or equivalently, set the variance of v_i to zero).
Explanation using variance decomposition:
With this change, x_3 is perfectly collinear with x_2 because $x_3 = x_2$ exactly. In the regression $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$, we can substitute to get:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_2 + \epsilon = \beta_1 + (\beta_2 + \beta_3) x_2 + \epsilon$$

When we regress x_2 on x_3 (and a constant), we get $R_2^2 = 1$ because x_2 and x_3 are perfectly correlated. Substituting into the variance decomposition:

$$\text{var}(\hat{\beta}_2 | X) = \sigma^2 \cdot \frac{1}{SST_2} \cdot \frac{1}{1 - R_2^2} = \sigma^2 \cdot \frac{1}{SST_2} \cdot \frac{1}{1 - 1} = \frac{\sigma^2}{SST_2 \cdot 0} = \infty$$

The variance becomes infinite, which means the OLS estimator is undefined. Geometrically, the $X'X$ matrix becomes singular (non-invertible) under perfect collinearity, so $(X'X)^{-1}$ does not exist, and we cannot compute $\hat{\beta} = (X'X)^{-1} X' y$.
We can only estimate the combined effect $\beta_2 + \beta_3$, not the individual coefficients β_2 and β_3 separately. There exist infinitely many combinations of (β_2, β_3) that yield the same fitted values.
Description of histogram:
The histogram would **not exist** or would show an **error/undefined behavior**. In practice, statistical software would either fail to compute the estimates (returning an error about singularity) or return arbitrary/unstable values that depend on numerical precision. If the software did return values, the histogram would show extremely large variance (or no consistent pattern) because the estimates would be numerically unstable.

(ii) Now, go back to the original dgp provided and change another element so that with any sample generated, we could perfectly estimate β_2. Again, clearly indicate the element that you changed and how. Using the expression provided, explain why in this case we would be able to perfectly estimate β_2. Additionally, describe how the histogram provided above would look like.

Answer:
Element changed:
Change $\epsilon_i|X \sim$ i.i.$N(0, 64)$ to $\epsilon_i = 0$ for all i (i.e., set $\sigma^2 = 0$).
Explanation using variance decomposition:
With $\sigma^2 = 0$, there is no random error in the model, so:

$$y_i = 1 + 1 \cdot x_{i2} + 1 \cdot x_{i3}$$

This is a deterministic relationship. Substituting into the variance decomposition:

$$\text{var}(\hat{\beta}_2|X) = \sigma^2 \cdot \frac{1}{SST_2} \cdot \frac{1}{1 - R_2^2} = 0 \cdot \frac{1}{SST_2} \cdot \frac{1}{1 - R_2^2} = 0$$

The variance of $\hat{\beta}_2$ is exactly zero, meaning there is no sampling variability. Every sample (with the same X) will yield exactly the same estimate $\hat{\beta}_2 = 1$. We can perfectly estimate β_2 because there is no noise in the data—the relationship is exact.
Description of histogram:
The histogram would show a **single spike at** $\beta_2 = 1$ with zero spread. All 10,000 estimates would be identical ($\hat{\beta}_2 = 1$), so there would be no distribution—just a single vertical line or bar at the true value. This reflects perfect estimation with no uncertainty.

Question 3

A linear regression model with 4 regressors (const, x_2, x_3, x_4) is set. We want to test: $H_0 : \beta_2 = \beta_3$ versus $H_1 : \beta_2 \neq \beta_3$. To perform this test we want to use estimator tilde, $\tilde{\beta}$, (i.e., not the OLS estimator!), which has the following asymptotic distribution:

$$\sqrt{n}(\tilde{\beta} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, \Omega).$$

Part (a)

Specify the dimension of the following elements of the asymptotic distribution above: $\tilde{\beta}$, $\boldsymbol{\beta}$, 0, Ω. No need to justify.

Answer:

Part (b)

Derive, step by step, a test statistic to perform the test above using this estimator tilde, for the case where matrix Ω is known.

Answer:
Step 1: Express the null hypothesis in matrix form
The null hypothesis $H_0 : \beta_2 = \beta_3$ can be written as:

$$H_0 : \beta_2 - \beta_3 = 0$$

This is equivalent to:

$$R\boldsymbol{\beta} = r$$

where $R = \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix}$ (a 1×4 matrix) and $r = 0$ (a scalar).
Step 2: Apply the restriction to the estimator
Under the null hypothesis, we have:

$$R\tilde{\beta} = r$$

The test is based on how far $R\tilde{\beta}$ deviates from r. We form:

$$R\tilde{\beta} - r = R\tilde{\beta} - 0 = R\tilde{\beta}$$

Step 3: Derive the asymptotic distribution of $R\tilde{\beta}$
Given that $\sqrt{n}(\tilde{\beta} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, \Omega)$, we can multiply both sides by the restriction matrix R:

$$\sqrt{n}R(\tilde{\beta} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, R\Omega R')$$

Under H_0 (where $R\boldsymbol{\beta} = r = 0$):

$$\sqrt{n}R\tilde{\beta} \overset{a}{\sim} N(0, R\Omega R')$$

Step 4: Standardize to obtain a chi-squared statistic
Since $\sqrt{n}R\tilde{\beta}$ is asymptotically normal with mean 0 and variance $R\Omega R'$, we can standardize it. Note that $R\Omega R'$ is a 1×1 matrix (a scalar in this case), so its inverse is simply the reciprocal. The Wald statistic is:

$$W = n(R\tilde{\beta} - r)'[R\Omega R']^{-1}(R\tilde{\beta} - r)$$

Substituting $r = 0$:

$$W = n(R\tilde{\beta})'[R\Omega R']^{-1}(R\tilde{\beta})$$

Since $R\tilde{\beta}$ is 1×1 (a scalar), we can write:

$$W = n\frac{(R\tilde{\beta})^2}{R\Omega R'}$$

Step 5: State the asymptotic distribution
Under H_0, the Wald statistic follows:

$$W \overset{a}{\sim} \chi^2(1)$$

where the degrees of freedom equals the number of restrictions ($q = 1$).
Final test statistic:

$$\boxed{W = n(R\tilde{\beta})'[R\Omega R']^{-1}(R\tilde{\beta}) \overset{a}{\sim} \chi^2(1) \text{ under } H_0}$$

Or more explicitly:

$$\boxed{W = n\frac{(\tilde{\beta}_2 - \tilde{\beta}_3)^2}{\omega_{22} - 2\omega_{23} + \omega_{33}} \overset{a}{\sim} \chi^2(1) \text{ under } H_0}$$

where ω_{ij} denotes the (i, j) element of Ω.

Part (c)

Justify why the statistic you derived is a proper test statistic.

Answer:
The statistic W is a proper test statistic because it satisfies the following requirements:
1. Known distribution under H_0: Under the null hypothesis, $W \overset{a}{\sim} \chi^2(1)$. This asymptotic distribution is well-defined and does not depend on unknown parameters (since Ω is assumed known). This allows us to determine critical values and compute p-values.
2. Measures deviation from H_0: The statistic W measures the squared deviation of $R\tilde{\beta}$ from the hypothesized value $r = 0$, weighted by the inverse of its variance $R\Omega R'$. Large values of W indicate that the data are inconsistent with the null hypothesis.
3. Asymptotic validity: The statistic is based on the asymptotic normality of $\sqrt{n}(\tilde{\beta} - \boldsymbol{\beta})$. By the continuous mapping theorem and Slutsky's theorem, the transformation to the Wald statistic preserves asymptotic validity. As $n \to \infty$, the distribution of W under H_0 converges to $\chi^2(1)$, ensuring correct Type I error control in large samples.
4. Proper size: For a test at significance level α, we reject H_0 if $W > \chi^2_{1,\alpha}$ (the α critical value of the $\chi^2(1)$ distribution). Asymptotically, $P(W > \chi^2_{1,\alpha} | H_0) = \alpha$, so the test has the correct size.
5. Power against alternatives: Under the alternative hypothesis $H_1 : \beta_2 \neq \beta_3$, the statistic W diverges to infinity as $n \to \infty$ (assuming $\beta_2 - \beta_3 \neq 0$), ensuring the test has asymptotic power approaching 1.
Therefore, W is a valid Wald test statistic for testing $H_0 : \beta_2 = \beta_3$.

Question 4

We set the following data generating process (dgp):

$$y = 2 + 1 \cdot x_2 + 1 \cdot x_3 + \epsilon \quad \epsilon | X \sim N(0, 16)$$

$$\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad \rho \in [-1, 1]$$

Part (a)

If we set $\rho = -1$, and generate a sample of 50 observations, show that there would exist an infinite number of OLS estimates for parameters β_2 and β_3 in regression:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon.$$

Answer:
When the correlation coefficient between x_2 and x_3 is set to $\rho = -1$, a state of **perfect negative collinearity** exists.
Step 1: Show the correlation matrix is singular
The correlation matrix for (x_2, x_3) is:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

The determinant is:

$$\det(\mathbf{C}) = (1)(1) - (-1)(-1) = 1 - 1 = 0$$

Since the determinant is zero, the covariance matrix is singular (non-invertible).
Step 2: Explain the perfect linear relationship
With $\rho = -1$, the variables x_2 and x_3 are perfectly negatively correlated. This means there exists an exact linear relationship between them:

$$x_3 = a + b \cdot x_2$$

where $b < 0$ (negative relationship). Specifically, for standardized variables with equal variances, $x_3 = c - x_2$ for some constant c. In our case with means of 10 and variances of 1, we have $x_3 = 20 - x_2$.
Step 3: Show the regression model is underidentified
Substituting the linear relationship into the regression:

$$y = \beta_1 + \beta_2 x_2 + \beta_3(20 - x_2) + \epsilon$$

$$y = (\beta_1 + 20\beta_3) + (\beta_2 - \beta_3)x_2 + \epsilon$$

We can only identify the combined parameters:

- Intercept: $\beta_1^* = \beta_1 + 20\beta_3$
- Slope: $\beta_2^* = \beta_2 - \beta_3$

But we have 2 equations with 3 unknowns (β_1, β_2, β_3), so the system is underidentified.
Step 4: Show $(X'X)$ is singular
The matrix of sample moments $X'X$ becomes singular under perfect collinearity:

$$\det(X'X) = 0$$

Consequently, $(X'X)^{-1}$ does not exist, making the OLS estimator undefined:

$$\hat{\beta} = (X'X)^{-1}X'y \quad \text{cannot be computed}$$

Step 5: Infinitely many solutions
For any arbitrary value of $\beta_3 = k$, we can set:

- $\beta_2 = \beta_2^* + k$
- $\beta_1 = \beta_1^* - 20k$

All such combinations yield identical fitted values $\hat{y}_i$ and the same sum of squared residuals. Therefore, there exist **infinitely many combinations** of $(\beta_1, \beta_2, \beta_3)$ that minimize the OLS objective function.

Conclusion: Under perfect collinearity ($\rho = -1$), the OLS estimates are not uniquely defined, and there exist infinitely many OLS estimates for β_2 and β_3.

Part (b)

We should simulate through different values of ρ $(-1, -0.5, 0, 0.5, 1)$.

Additional elements to set:

- Number of Monte Carlo simulations: $M = 10,000$ (or another large number)
- Sample size: $n = 50$ (same as before)
- Values of ρ to test: $\rho \in \{-0.99, -0.5, 0, 0.5, 0.99\}$ (avoid exact ± 1 to prevent singularity errors)
- True parameter values: $\beta_1 = 2$, $\beta_2 = 1$, $\beta_3 = 1$
- Error variance: $\sigma^2 = 16$

Main steps of Monte Carlo experiment:

1. **For each value of ρ:**
   a. Initialize storage for estimates: Create empty arrays to store $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\text{se}(\hat{\beta}_2)$ for each simulation
2. **For each simulation $s = 1, \ldots, M$:**
   a. Generate regressors: Draw (x_2, x_3) from the bivariate normal distribution with correlation ρ
   b. Generate errors: Draw $\epsilon \sim N(0, 16)$
   c. Generate dependent variable: Compute $y = 2 + 1 \cdot x_2 + 1 \cdot x_3 + \epsilon$
   d. Run OLS regression: Estimate $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$
   e. Store results: Save $\hat{\beta}_1^{(s)}$, $\hat{\beta}_2^{(s)}$, $\hat{\beta}_3^{(s)}$, and $\text{se}(\hat{\beta}_2^{(s)})$
3. **Calculate summary statistics for each ρ:**
   a. Mean of estimates: $\bar{\hat{\beta}}_2 = \frac{1}{M} \sum_{s=1}^{M} \hat{\beta}_2^{(s)}$ (check for unbiasedness)
   b. Variance of estimates: $\text{Var}(\hat{\beta}_2) = \frac{1}{M-1} \sum_{s=1}^{M} (\hat{\beta}_2^{(s)} - \bar{\hat{\beta}}_2)^2$
   c. Average standard error: $\overline{\text{se}(\hat{\beta}_2)} = \frac{1}{M} \sum_{s=1}^{M} \text{se}(\hat{\beta}_2^{(s)})$
   d. Variance Inflation Factor: Compute $VIF = \frac{1}{1-R_2^2}$ where R_2^2 is from regressing x_2 on x_3

What to look at/analyze:

- **Unbiasedness across ρ values:** Verify that $\bar{\hat{\beta}}_2 \approx 1$ for all values of ρ. OLS should remain unbiased regardless of correlation between regressors (as long as there's no perfect collinearity).
- **Variance as a function of ρ:** Plot $\mathrm{Var}(\hat{\beta}_2)$ against ρ. We should observe:
  - Minimum variance at $\rho = 0$ (no correlation)
  - Increasing variance as $|\rho| \to 1$ (approaching perfect collinearity)
  - Variance should be symmetric around $\rho = 0$ (i.e., $\mathrm{Var}(\hat{\beta}_2|\rho) = \mathrm{Var}(\hat{\beta}_2|-\rho)$)
- **Variance Inflation Factor (VIF):** The theoretical relationship is:

$$\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2}{SST_2(1-R_2^2)} = \frac{\sigma^2}{SST_2} \cdot VIF$$

  As $\rho^2 \to 1$, we have $R_2^2 \to 1$, so $VIF \to \infty$. This shows that high correlation inflates the variance of estimates, making them unstable and imprecise.
- **Distribution stability:** Create histograms of $\hat{\beta}_2$ for each ρ value. As $|\rho|$ increases, the histograms should become wider (higher variance) while remaining centered at the true value $\beta_2 = 1$.
- **Practical implications:** Even though estimates remain unbiased, high multicollinearity (large $|\rho|$) makes estimates unreliable for inference because:
  - Standard errors become very large
  - Confidence intervals become very wide
  - Small changes in the data can lead to large changes in estimates
  - Individual coefficients may not be statistically significant even if they are jointly significant

Expected outcome: The simulation should confirm that OLS estimates stay unbiased across all values of ρ, but the variance increases dramatically as $|\rho| \to 1$, demonstrating that multicollinearity does not cause bias but severely reduces precision and makes the estimates unstable.