Counterfactual Study Guide

Part I: What Else Could She Ask?

Based on patterns from all 3 practice exams

Exam Structure Pattern

All three Part I exams follow this structure:

1. **Question 1 (50-60 points):** Large question with regression output - tests formulas, hypothesis testing, assumptions
2. **Question 2 (25-30 points):** Monte Carlo simulation / DGP analysis
3. **Question 3 (10-15 points):** Theoretical (asymptotic tests, derivations) - often skipped or incomplete in practice exams

CRITICAL: Formulas to Memorize

She said: Know ANOVA outputs and t-statistic BY HEART

ANOVA Table Components

SST (Total Sum of Squares):

$SST = \Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - n\bar{y}^2$

SSE (Sum of Squared Errors / Residuals):

$SSE = \Sigma e_i^2 = \Sigma(y_i - \hat{y}_i)^2$

SSR (Regression Sum of Squares / Model Sum of Squares):

$SSR = \Sigma(\hat{y}_i - \bar{y})^2 = SST - SSE$

Relationship:

$\mathbf{SST = SSR + SSE}$

Mean Squared Error (MSE = $\hat{\sigma}^2$):

$MSE = SSE / (n - K)$ where K = # of parameters

Root MSE:

Root $MSE = \sqrt{MSE} = \sqrt{(SSE/(n-K))} = \hat{\sigma}$

R^2 (Coefficient of Determination):

$\mathbf{R^2 = SSR/SST = 1 - SSE/SST}$

Adjusted R^2:

$\bar{R}^2 = 1 - (SSE/(n-K))/(SST/(n-1)) = 1 - (1-R^2)(n-1)/(n-K)$

F-statistic:

$\mathbf{F = (SSR/q) / (SSE/(n-K))}$ **where q = # of restrictions**

Alternative: $F = (R^2/q) / ((1-R^2)/(n-K))$

t-statistic and Related Formulas

t-statistic:

$t = (\hat{\beta}_j - r) / se(\hat{\beta}_j)$

Under H_0: $\beta_j = r$ (usually $r = 0$)

Distribution: $t \sim t(n-K)$ under classical assumptions

Without normality: $t \sim N(0,1)$ asymptotically

Standard Error of $\hat{\beta}_j$:

$se(\hat{\beta}_j) = \sqrt{[\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}]}$

where $\hat{\sigma}^2$ = MSE and $[(X'X)^{-1}]_{jj}$ is the jth diagonal element

Variance Decomposition with VIF:

$Var(\hat{\beta}_j|X) = \sigma^2 / SST_j \times 1/(1-R^2_j)$

VIF = $1/(1-R^2_j)$ where R^2_j is from regressing x_j on other regressors

Confidence Interval:

$\hat{\beta}_j \pm t(\alpha/2, n-K) \times se(\hat{\beta}_j)$

p-value (two-sided test):

$\text{p-value} = P(|t| > |t_{obs}| \mid H_0) = 2 \cdot P(t > |t_{obs}|)$

Topic 1: β and Estimators

What She ASKED	What She COULD ASK
Definition of $\beta = [E(xx')]^{-1}E(xy)$ and its dimensions	OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ or first-order conditions $X'(y - X\hat{\beta}) = 0$
Why estimate this β (best linear approximation to CEF)	Difference between population β and sample $\hat{\beta}$, or unbiasedness of $\hat{\beta}$
Calculate $\hat{\beta}$ from given data	Given $\hat{\beta}$, calculate fitted values or residuals

What You Must Know

- **Population parameter:** $\beta = [E(xx')]^{-1}E(xy)$ - this is what we're trying to estimate
- **OLS estimator:** $\hat{\beta} = (X'X)^{-1}X'y$ - this is how we estimate it from sample data
- **Dimensions:** If K parameters, then β is $K\times1$, $E(xx')$ is $K\times K$, $E(xy)$ is $K\times1$
- **Unbiasedness:** $E[\hat{\beta}|X] = \beta$ under Gauss-Markov assumptions
- **First-order conditions:** $X'e = 0$ or $\Sigma x_i e_i = 0$ (residuals orthogonal to regressors)

Topic 2: Standard Errors and Inference

What She ASKED	What She COULD ASK
Expression for $se(\hat{\beta}_j) = \sqrt{[\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}]}$	Calculate $se(\hat{\beta}_j)$ given MSE and other ANOVA components
What information does the standard error provide?	How does $se(\hat{\beta}_j)$ change with sample size, multicollinearity, or variance?
Calculate t-statistic for $H_0: \beta_2 = 0$	Calculate t-statistic for $H_0: \beta_2 = 1$ or $H_0: \beta_2 + \beta_3 = 0$

What You Must Know

- **Standard error formula:** $se(\hat{\beta}_j) = \sqrt{[\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}]}$ where $\hat{\sigma}^2$ = MSE
- **What it measures:** The standard error quantifies the precision/uncertainty of the estimate. Smaller se → more precise estimate.
- **Factors affecting $se(\hat{\beta}_j)$:**

• Increases with σ^2 (error variance)

• Decreases with sample size n

• Increases with multicollinearity (high R^2_j)

• Decreases with variation in x_j (high SST_j)

- **Heteroskedasticity:** Standard errors are INCORRECT under heteroskedasticity. Need robust (White/Huber) standard errors.

- **$\hat{\beta}$ is unaffected:** OLS estimator $\hat{\beta}$ itself doesn't change with heteroskedasticity, only se($\hat{\beta}$) changes

Topic 3: Hypothesis Testing (t-tests)

What She ASKED	What She COULD ASK
Test H_0: $\beta_2 = 0$ using t-test at $\alpha = 0.01$	Test H_0: $\beta_2 = 1$ or H_0: $\beta_2 = \beta_3$ using t-test
Draw p-value for two-sided test	Draw p-value for one-sided test (H_1: $\beta_2 > 0$)
At what significance level is regressor significant?	Compare significance at $\alpha = 0.01$ vs $\alpha = 0.05$ vs $\alpha = 0.10$

What You Must Know

- **Two-sided test:**

H_0: $\beta_j = r$ vs H_1: $\beta_j \neq r$
Reject if $|t| > t(\alpha/2, n-K)$
p-value = $2 \cdot P(t > |t_{o\beta s}|)$

- **One-sided test (right-tailed):**

H_0: $\beta_j \leq r$ vs H_1: $\beta_j > r$
Reject if $t > t(\alpha, n-K)$
p-value = $P(t > t_{o\beta s})$

- **Decision rule:**

Reject H_0 if p-value $< \alpha$
OR reject H_0 if $|t| >$ critical value

- **Confidence interval connection:** If 95% CI doesn't contain H_0 value, reject at $\alpha = 0.05$

Topic 4: F-tests for Joint Significance

What She ASKED	What She COULD ASK
Test H_0: $\beta_3 = \beta_4 = \beta_5 = 0$ (exclusion restrictions)	Test H_0: $\beta_2 = \beta_3$ (equality restriction)
Calculate RSSE by estimating restricted model	Test H_0: $\beta_2 + \beta_3 = 1$ (linear restriction)
F-statistic = (RSSE-SSE)/q ÷ SSE/(n-K)	Calculate F from R^2 instead: $F = (R^2/q) / ((1-R^2)/(n-K))$

What You Must Know

- **General form:** H_0: $R\beta = r$ vs H_1: $R\beta \neq r$, where R is q×K matrix and r is q×1 vector
- **Exclusion restrictions (most common):**

H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$

Remove those variables and estimate restricted model to get RSSE

- **Equality restriction:**

H_0: $\beta_2 = \beta_3$ → Create new variable $w = x_2 - x_3$ and test if its coefficient is 0

OR estimate $y = \beta_1 + \beta_2(x_2 + x_3) + \ldots$ to get RSSE

- **Linear restriction:**

H_0: $\beta_2 + \beta_3 = 1$ → Rewrite as $y - x_3 = \beta_1 + \beta_2(x_2 + x_3) + \text{(other terms)} + e$

- **F-statistic formulas:**

F = (RSSE - SSE)/q ÷ SSE/(n-K)

Alternative: $F = (R^2_{ur} - R^2_r)/q \div (1 - R^2_{ur})/(n-K)$

Overall significance: $F = R^2/q \div (1-R^2)/(n-K)$ where $q = K-1$

- **Distribution:** $F \sim F(q, n-K)$ under H_0 with classical assumptions
- **Decision rule:** Reject H_0 if $F > F(\alpha, q, n-K)$
- **Relationship to t-test:** For single restriction (q=1), $F = t^2$ and $F(1,n-K) = [t(n-K)]^2$

Topic 5: Classical Assumptions

What She ASKED	What She COULD ASK
List all 6 classical assumptions (linearity, strict exogeneity, etc.)	Which specific assumption is violated in this scenario?
What happens if we drop homoskedasticity?	What happens if we drop normality or strict exogeneity?
Which statistics need adjustment under heteroskedasticity?	What if both heteroskedasticity AND autocorrelation are present?

The 6 Classical Assumptions

4. **A1 - Linearity:** $y = X\beta + e$
5. **A2 - Strict Exogeneity:** $E[e|X] = 0$
6. **A3 - No Perfect Collinearity:** $\text{rank}(X) = K$
7. **A4 - Conditional Homoskedasticity:** $Var(e|X) = \sigma^2 I$
8. **A5 - No Conditional Autocorrelation:** $Cov(e_i, e_j|X) = 0$ for $i \neq j$
9. **A6 - Normality:** $e|X \sim N(0, \sigma^2 I)$

Effects of Dropping Assumptions

Drop HOMOSKEDASTICITY (A4):

- **$\hat{\beta}$ remains unbiased and consistent**
- **se($\hat{\beta}$) is WRONG** → need heteroskedasticity-robust (White/Huber) standard errors
- **t-tests and F-tests are INVALID** using regular standard errors
- **Confidence intervals are WRONG**
- **R^2 is UNAFFECTED**
- OLS is no longer BLUE (not efficient)

Drop NORMALITY (A6):

- **$\hat{\beta}$ remains unbiased and consistent**
- **Exact t and F distributions NO LONGER HOLD**
- **Use ASYMPTOTIC distributions instead:**

$t \sim N(0,1)$ asymptotically (not t-distribution)

$F \sim \chi^2(q)/q$ asymptotically

- Standard errors are still CORRECT (if homoskedasticity holds)
- R^2 is UNAFFECTED

Drop STRICT EXOGENEITY (A2):

- **$\hat{\beta}$ is BIASED** - omitted variable bias, measurement error, simultaneity
- **May or may not be consistent** depending on the violation
- ALL inference is compromised
- **Need IV or other methods**

Perfect COLLINEARITY (violate A3):

- **(X'X) is SINGULAR** - cannot invert
- **$\hat{\beta}$ is NOT DEFINED** (infinitely many solutions)
- Cannot estimate individual coefficients

Topic 6: R^2 and Model Fit

What She ASKED	What She COULD ASK
Why is R^2 called in-sample predictive power?	Calculate R^2 from ANOVA components
Regress OLS residuals on all regressors - what R^2?	What's the relationship between R^2, adjusted R^2, and sample size?
How does adding a regressor affect R^2?	When can $R^2 = 1$? When can it be negative?

What You Must Know

- **Definition:** $R^2 = SSR/SST = 1 - SSE/SST$
- **Interpretation:** Proportion of variance in y explained by the model. Ranges from 0 to 1.
- **In-sample predictive power:** Measures how well the model fits the data used to estimate it, NOT out-of-sample prediction
- **Adding regressors:** R^2 NEVER decreases when adding variables (SSE can only decrease or stay same)
- **Adjusted R^2:** $\bar{R}^2 = 1 - (SSE/(n-K))/(SST/(n-1))$ - penalizes for additional regressors, can decrease
- **Properties of OLS residuals:**

$\Sigma e_i = 0$ (residuals sum to zero with intercept)

$X'e = 0$ (orthogonal to regressors)

→ Regressing e on X gives $R^2 = 0$

- **What R^2 does NOT tell you:**

Causality (high $R^2 \neq$ causal relationship)

Whether model is correctly specified

Out-of-sample prediction quality

- **Perfect fit:** $R^2 = 1$ when $SSE = 0$ (all points on regression line)
- **Can R^2 be negative?:** NO if model has intercept. YES if no intercept (SSE could $> SST$).

Topic 7: Variance Decomposition and Multicollinearity

What She ASKED	What She COULD ASK
Expression: $Var(\hat{\beta}_j\|X) = \sigma^2/SST_j \times 1/(1-R^2_j)$	Calculate VIF given R^2_j
What is VIF and what does it measure?	How does multicollinearity affect $se(\hat{\beta})$, t-stats, and confidence intervals?
In Monte Carlo, how does changing σ_v^2 (collinearity) affect histogram?	What happens to variance as $R^2_j \to 1$? Why can't we estimate coefficients?

What You Must Know

- **Variance decomposition:**

$$Var(\hat{\beta}_j|X) = \sigma^2 / SST_j \times 1/(1-R^2_j)$$

- **Components:**

σ^2 = error variance (from MSE)

SST_j = total variation in x_j

R^2_j = R^2 from regressing x_j on other regressors

- **VIF (Variance Inflation Factor):**

$$VIF = 1/(1-R^2_j)$$

- **Interpretation:**

VIF = 1 → no multicollinearity ($R^2_j = 0$)

VIF = 5 → variance is 5 times larger than with no collinearity

VIF → ∞ as $R^2_j \to 1$ (perfect collinearity)

- **Effects of multicollinearity:**

$\hat{\beta}$ remains UNBIASED

Var($\hat{\beta}$) INCREASES → larger standard errors

t-statistics DECREASE → harder to reject H_0

Confidence intervals WIDER

Estimates become unstable/sensitive to small data changes

R^2 typically UNAFFECTED (joint fit can be good even if individual coefficients imprecise)

- **Perfect collinearity:**

$R^2_j = 1$ → VIF = ∞ → Var($\hat{\beta}_j$) = ∞

Cannot estimate individual coefficients (infinitely many solutions)

Topic 8: Monte Carlo Simulations

What She ASKED	What She COULD ASK
Histogram centered at true β value - illustrates unbiasedness	How would histogram change if we increase sample size n?
Can/cannot illustrate consistency with fixed n=50	What if we change error variance σ^2 (from 16 to 64)?
What happens to histogram width when σ^2_V changes (collinearity)?	Perfect collinearity ($\sigma^2_V=0$) vs no collinearity ($\sigma^2_V \to \infty$)?

What You Must Know

- **Setup:** Generate M samples from DGP, estimate $\hat{\beta}$ for each, create histogram of M estimates
- **What CAN be illustrated:**

Unbiasedness: Histogram centered at true $\beta \to E[\hat{\beta}] = \beta$

Sampling distribution shape: Normally distributed if errors normal

Variance/precision: Width of histogram shows $Var(\hat{\beta})$

Effects of multicollinearity: Compare histogram widths with different R^2_j

- **What CANNOT be illustrated:**

Consistency: Requires $n \to \infty$, but simulation uses fixed n

Would need to vary n and show $Var(\hat{\beta}) \to 0$ as n increases

- **Effects of parameter changes:**

↑ **σ^2 (error variance)** → histogram WIDER (less precise estimates)

↓ **$\sigma^2 \to 0$** → histogram becomes spike at true value (perfect estimation)

↑ **n (sample size)** → histogram NARROWER (more precise)

↑ **R^2_j (collinearity)** → histogram WIDER (VIF increases)

$R^2_j \to 1$ (perfect collinearity) → no defined histogram (cannot estimate)

- **Conditional vs unconditional:**

Fixed X across samples: Conditional sampling distribution (what's typically shown)

New X each sample: Unconditional sampling distribution

Topic 9: Asymptotic Theory and Tests

What She ASKED	What She COULD ASK
Derive Wald test statistic for H_0: $R\beta = r$	Derive LM (Lagrange Multiplier) test
Distribution: $\sqrt{n}(\hat{\beta}-\beta) \sim N(0,\Omega)$	Derive LR (Likelihood Ratio) test
What changes without normality assumption?	Show equivalence of Wald, LM, LR asymptotically

What You Must Know

- **Asymptotic normality:**

$$\sqrt{n}(\hat{\beta}-\beta) \to^d N(0,V)$$

- **Wald test:**

For H_0: $R\beta = r$

$$W = n(R\hat{\beta}-r)'[R\Omega R']^{-1}(R\hat{\beta}-r) \sim \chi^2(q)$$

where Ω is asymptotic variance matrix

- **Without normality:**

t-statistic ~ N(0,1) asymptotically (not t-distribution)

F-statistic ~ $\chi^2(q)/q$ asymptotically

Critical values different from exact distributions

- **Key differences:**

Exact (finite sample): Requires normality, uses t and F distributions

Asymptotic: No normality needed, uses normal and χ^2 distributions

Quick Reference Cheat Sheet

Core Formulas - MEMORIZE THESE

OLS Estimator:

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

Standard Error:

$$se(\hat{\beta}_j) = \sqrt{[MSE \cdot [(X'X)^{-1}]_{jj}]}$$

t-statistic:

$$t = (\hat{\beta}_j - r) / se(\hat{\beta}_j) \sim t(n-K)$$

F-statistic:

$$F = [(RSSE-SSE)/q] / [SSE/(n-K)] \sim F(q,n-K)$$

$$F = (R^2/q) / [(1-R^2)/(n-K)]$$

R^2:

$$R^2 = SSR/SST = 1 - SSE/SST$$

Variance Decomposition:

$$Var(\hat{\beta}_j|X) = \sigma^2/SST_j \times VIF$$

$$VIF = 1/(1-R^2_j)$$

ANOVA Decomposition:

$$SST = SSR + SSE$$

Common Exam Tricks - WATCH OUT

- **β indices:** Carefully check which β you're testing. Is it β_2, β_3, β_4? Don't mix them up!
- **Two-sided vs one-sided:** Two-sided uses α/2 in each tail. One-sided uses α in one tail.
- **Degrees of freedom:** t has n-K, F has (q, n-K) where K = total # parameters including intercept
- **What's affected by heteroskedasticity:** se($\hat{\beta}$), t-stats, CIs, F-stats YES. $\hat{\beta}$ and R^2 NO.
- **What's affected by dropping normality:** Distribution changes (t→N, F→χ^2/q). $\hat{\beta}$ and se($\hat{\beta}$) still OK if homoskedasticity holds.
- **Restricted model:** Remove the variables being tested to get RSSE (RSSE ≥ SSE always)
- **R matrix dimensions:** q × K where q = # restrictions, K = # parameters
- **Perfect collinearity:** Cannot estimate at all (VIF = ∞). Imperfect collinearity: can estimate but imprecisely (high VIF).

Final Exam Day Strategy

10. **READ CAREFULLY:** Check β indices, check if test is one/two-sided, check which variables in F-test
11. **WRITE FORMULAS FIRST:** Before calculating, write the formula. Shows you know the method even if arithmetic wrong.

12. **SHOW YOUR WORK:** Partial credit is possible. Write steps clearly.
13. **DON'T SKIP HARD QUESTIONS:** Write something. You scored 0/15 on asymptotic tests - at least try!
14. **TIME MANAGEMENT:** Question 1 is worth most points. Don't spend too long on Question 3.
15. **DOUBLE-CHECK ANOVA CALCULATIONS:** Verify SST = SSR + SSE. If this doesn't hold, you made an arithmetic error.

You've got this! Focus on the gaps and you'll improve significantly.

Good luck! 🍀