Econometrics Formula Reference Sheet

Key Formulas for Exam Question 1

1. t-Statistic Formula

For testing $H_0 : \beta_j = \beta_0$ versus $H_1 : \beta_j \neq \beta_0$:

$$t = \frac{\hat{\beta}_j - \beta_0}{\text{se}(\hat{\beta}_j)} \sim_{H_0} t(n-K) \tag{1}$$

where:

- $\hat{\beta}_j$ = OLS estimate of coefficient j
- β_0 = hypothesized value under H_0 (often 0)
- $\text{se}(\hat{\beta}_j)$ = standard error of $\hat{\beta}_j$
- n = sample size
- K = number of parameters (including constant)
- $n - K$ = degrees of freedom

Common case: Testing $H_0 : \beta_j = 0$

$$t = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \tag{2}$$

2. Standard Error Formula

$$\text{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}} \tag{3}$$

where:

- $\hat{\sigma}^2 = \frac{\text{SSE}}{n-K}$ = estimate of error variance
- $[(X'X)^{-1}]_{jj}$ = j-th diagonal element of $(X'X)^{-1}$
- SSE = Sum of Squared Errors (Residual SS)

Alternative expression:

$$\hat{\sigma}^2 = (\text{Root MSE})^2 \tag{4}$$

3. Confidence Interval

$(1-\alpha) \times 100\%$ confidence interval for β_j:

$$\text{CI} = \hat{\beta}_j \pm t_{\alpha/2,n-K} \times \text{se}(\hat{\beta}_j) \tag{5}$$

Common cases:

- 95% CI: $\alpha = 0.05$, so $t_{0.025,n-K}$
- 99% CI: $\alpha = 0.01$, so $t_{0.005,n-K}$
- 90% CI: $\alpha = 0.10$, so $t_{0.05,n-K}$

4. Key Relationships from Regression Output

Sum of Squares Decomposition

$$\text{Total SS} = \text{Model SS} + \text{SSE} \tag{6}$$

where:

- Total SS $= \sum_{i=1}^{n}(y_i - \bar{y})^2$
- Model SS = Explained SS $= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
- SSE = Residual SS $= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\hat{\epsilon}_i^2$

Coefficient of Determination

$$R^2 = \frac{\text{Model SS}}{\text{Total SS}} = 1 - \frac{\text{SSE}}{\text{Total SS}} \tag{7}$$

Error Variance Estimate

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n-K} \tag{8}$$

$$\text{Root MSE} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\text{SSE}}{n-K}} \tag{9}$$

IMPORTANT: Understanding the Relationship
These three quantities are related but DIFFERENT:

1. **SSE** (Sum of Squared Errors) = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
   → This is the **TOTAL SUM** of squared residuals (large number)
2. $\hat{\sigma}^2$ (Error Variance) = $\frac{\text{SSE}}{n-K}$
   → This is the **AVERAGE** squared error per df (medium number)
3. **Root MSE** (Standard Error of Regression) = $\sqrt{\hat{\sigma}^2}$
   → This is the **STANDARD DEVIATION** of errors (smaller number)

The Chain:

$$\text{SSE} \xrightarrow{\text{divide by } (n-K)} \hat{\sigma}^2 \xrightarrow{\text{take } \sqrt{}} \text{Root MSE}$$

Going Backwards:

$$\text{Root MSE} \xrightarrow{\text{square}} \hat{\sigma}^2 \xrightarrow{\text{multiply by } (n-K)} \text{SSE}$$

Example: SSE = 54.03, $n - K = 173$

- $\hat{\sigma}^2 = 54.03/173 = 0.312$
- Root MSE = $\sqrt{0.312} = 0.559$

Degrees of Freedom

- Model df = $K - 1$ (number of regressors excluding constant)
- Residual df = $n - K$
- Total df = $n - 1$

F-statistic

$$F = \frac{\text{Model SS}/(K-1)}{\text{SSE}/(n-K)} \sim_{H_0} F(K-1, n-K) \tag{10}$$

5. Hypothesis Testing Decision Rules

Using t-statistic

Two-tailed test: Reject H_0 if $|t| > t_{\alpha/2,n-K}$
One-tailed test:

- $H_1 : \beta_j > \beta_0$: Reject H_0 if $t > t_{\alpha,n-K}$
- $H_1 : \beta_j < \beta_0$: Reject H_0 if $t < -t_{\alpha,n-K}$

Using p-value

Reject H_0 if p-value $< \alpha$

Using Confidence Interval

For testing $H_0 : \beta_j = \beta_0$ at significance level α:

- Reject H_0 if β_0 is NOT in the $(1 - \alpha) \times 100\%$ CI
- Fail to reject H_0 if β_0 IS in the $(1 - \alpha) \times 100\%$ CI

6. Quick Calculations from Stata Output

Given typical output, you can calculate:

- $\hat{\sigma}^2 = (\text{Root MSE})^2$
- SSE $= \hat{\sigma}^2 \times (n - K)$
- Model SS = Total SS - SSE
- t-statistic = Coef. / Std. Err.
- Degrees of freedom $= n - K$