Economics 300–01: Quantitative Methods in Economics
Wesleyan University, Spring 2002

Summary of Formulas for the Final Exam

1. **Ordinary Least Squares (OLS) Estimation:**

   - Model (Data Generating Process):

   $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + e_i$$

   - Model Assumptions:

     (a) $\sum_{i=1}^{n} \left(X_{ji} - \bar{X}_j\right)^2 > 0 \quad \forall\, j = 1, \ldots, k\,; \quad X_j$'s are not perfectly collinear.

     (b) $\mathrm{E}[e_i] = 0$.

     (c) $\mathrm{var}(e_i) = \sigma^2$.

     (d) $\mathrm{cov}(e_i, e_j) = 0 \quad \forall\, i \neq j$.

     (e) $e_i \sim N(0, \sigma^2)$.

   - Estimated Regression Line:

   $$Y_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki} + \widehat{e}_i$$

   - Fitted Values of Regressand (Dependent Variable):

   $$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}$$

   - Estimated Residual:

   $$\widehat{e}_i = Y_i - \widehat{Y}_i$$

2. **Formulas for Least Squares Estimates:**

   - $k = 1$: (Simple Regression)

   $$\begin{aligned} \widehat{\beta}_1 &= \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum \left(X_i - \bar{X}\right)^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i X_i} \\ \widehat{\beta}_0 &= \bar{Y} - \widehat{\beta}_1 \bar{X} \end{aligned}$$

   - $k = 2$:

   $$\begin{aligned} \widehat{\beta}_1 &= \frac{\sum x_1 y \sum x_2^2 - \sum x_2 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2} \\ \widehat{\beta}_2 &= \frac{\sum x_2 y \sum x_1^2 - \sum x_1 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2} \\ \widehat{\beta}_0 &= \bar{Y} - \widehat{\beta}_1 \bar{X}_1 - \widehat{\beta}_2 \bar{X}_2 \end{aligned}$$

   - Standard Error of the Estimate (SE of the regression):

   $$s = \sqrt{s^2}; \quad s^2 = \widehat{\sigma^2} = \frac{\sum \widehat{e}_i^2}{n - k - 1}$$

3. **Total vs. Direct Effect:** ($k = 2$ case)

$$\underbrace{\frac{\sum x_1 y}{\sum x_1^2}}_{\text{Total Effect}} \quad = \quad \underbrace{\widehat{\beta}_1}_{\text{Direct Effect}} \quad + \quad \underbrace{\widehat{\beta}_2 \frac{\sum x_1 x_2}{\sum x_1^2}}_{\text{Indirect Effect}}$$

4. **Coefficient of Determination:**

- Analysis of Variance:

$$\underbrace{\sum \left(Y_i - \bar{Y}\right)^2}_{SST} \;=\; \underbrace{\sum \left(Y_i - \widehat{Y}\right)^2}_{SSR} \;+\; \underbrace{\sum \left(\widehat{Y} - \bar{Y}\right)^2}_{SSE}$$

  - Special case: $k = 1$ (Simple Regression):

$$\underbrace{\sum y_i^2}_{SST} \;=\; \underbrace{\sum \widehat{e}_i^2}_{SSR} \;+\; \underbrace{\widehat{\beta}_1^2 \sum x_i^2}_{SSE}$$

- R–squared:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum \widehat{e}_i^2}{\sum y_i^2}$$

- Adjusted R–squared:

$$\bar{R}^2 = 1 - \left(\frac{SSR}{SST}\right)\left(\frac{n-1}{n-k-1}\right) = 1 - \frac{\sum \widehat{e}_i^2/(n-k-1)}{\sum y_i^2/(n-1)}$$

5. **Statistical Inference: Overview**

(a) Gauss-Markov Theorem:

i. If assumptions (a) – (d) hold:
- $\widehat{\beta}_j$ is the BLUE of β_j.
- s^2 is an unbiased estimator of σ^2.

ii. If assumptions (a) – (e) hold:
- $\widehat{\beta}_j$ is the BUE of β_j.
- s^2 is the BUE of σ^2.

(b) Sampling Distributions for Simple Regression:

i. If assumptions (a) – (e) hold:
- $\widehat{\beta}_1 \sim N\left(\beta_1,\, \frac{\sigma^2}{\sum x^2}\right)$
- $\widehat{\beta}_0 \sim N\left(\beta_0,\, \sigma^2 \cdot \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}\right)\right)$

6. **Statistical Inference: Single Coefficient Tests**

(a) Hypotheses (two-sided):

- *Null Hypothesis*: $\mathrm{H}_0 : \beta_j = \beta_j^0$.
- *Alternative Hypothesis*: $\mathrm{H}_1 : \beta_j \neq \beta_j^0$.

(b) Approaches:

i. Classical:

- *Test Statistic*:

$$\widehat{t} = \frac{\widehat{\beta}_j - \beta_j^0}{SE_{\widehat{\beta}_j}} \sim t_{n-k-1}$$

- *Decision Rule*: Reject H_0 at the $\frac{\lambda}{2}$ level of significance if $\left|\widehat{t}\right| > t_{n-k-1,\,\lambda/2}$

ii. Confidence Interval:

- *Test Statistic*:

$$\beta_j = \widehat{\beta}_j \pm t_{n-k-1,\,\lambda/2} \cdot SE_{\widehat{\beta}_j}$$

- *Decision Rule*: Reject H_0 at the $\frac{\lambda}{2}$ level of significance if β_j^0 is not covered by the above confidence interval

iii. p-value:

- *Test Statistic*:

$$p = 2 \cdot \Pr\left(t > \widehat{t}\right)$$

- *Decision Rule*: Reject H_0 at the p level of significance.

(c) Computing the Standard Error of $\widehat{\beta}_j$:

- $k = 1$ (Simple Regression):

$$SE_{\widehat{\beta}_1} = \frac{s}{\sqrt{\sum x^2}}$$

- $k = 2$:

$$SE_{\widehat{\beta}_1} = \frac{s}{\sqrt{\sum x_1^2 - \frac{\left(\sum x_1 x_2\right)^2}{\sum x_2^2}}}$$

$$SE_{\widehat{\beta}_2} = \frac{s}{\sqrt{\sum x_2^2 - \frac{\left(\sum x_1 x_2\right)^2}{\sum x_1^2}}}$$

- $k \geq 2$:

$$SE_{\widehat{\beta}_j} = \frac{s}{\sqrt{\sum x_j^2}} \cdot \frac{1}{\sqrt{1 - R_j^2}},$$

where R_j^2 is the coefficient of determination of the regression of X_j on all other X's.

Note: The above formula (for all of part 6) are only valid for $j = 1, \ldots, k$; these formula do not apply to the estimated constant term, $\widehat{\beta}_0$.

7. **Statistical Inference: Joint Tests and Tests of Linear Restrictions**

(a) General case:

i. Hypotheses:

- *Null Hypothesis*: Restricted regression is true; yields SSR_0.
- *Alternative Hypothesis*: Unrestricted regression is true; yields SSR_1.

ii. Test Statistic:

$$\widehat{F} = \frac{(SSR_0 - SSR_1)\,/r}{SSR_1/(n-k-1)}$$

iii. Decision Rule:

Reject H_0 at the λ level of significance if $\widehat{F} > F_{r,\,n-k-1,\,\lambda}$.

(b) Special case: Test of no linear relationship

i. Hypotheses:

- *Null Hypothesis*: All $\beta_j = 0$, for $j = 1, \ldots, k$.
- *Alternative Hypothesis*: At least one $\beta_j \neq 0$.

ii. Test Statistic:

$$\widehat{F} = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k}$$

iii. Decision Rule:

Reject H_0 at the λ level of significance if $\widehat{F} > F_{k,\,n-k-1,\,\lambda}$.

8. **Prediction**

(a) Simple Regression ($k = 1$):

- Predicted (forecasted) value for Y_0, given X_0:

$$\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 X_0$$

- Standard Error of the Prediction:

$$SE_{\widehat{Y}_0} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x^2}}$$

- Prediction (forecast) interval for Y_0:

$$Y_0 = \widehat{Y}_0 \pm t_{n-k-1,\,\lambda/2} \cdot SE_{\widehat{Y}_0}$$

(b) Multiple Regression ($k \geq 2$):

- Predicted (forecasted) value for Y_0, given $X_{10}, \ldots, X_{k0}$:

$$\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 X_{10} + \cdots + \widehat{\beta}_k X_{k0}$$

9. **Other Topics:**

   (a) Multicollinearity

   (b) Dummy Variables

   (c) Non-linear Regression

   (d) Heteroskedasticity

   (e) Autocorrelation