

CONSIDER 2 VARIABLE REGRESSION

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

THE OLS ESTIMATORS OF THE COEFFICIENTS ARE

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

THE VARIANCE OF $\hat{\beta}_2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$

VARIANCE OF $\hat{\beta}_1 = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right] \sigma^2$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\bar{x}}{\sum (x_i - \bar{x})^2}$$

[HINT: RECALL $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ &
 $X'X$ IN THIS 2-VAR REGRESSION CASE IS
 $X'X = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$]

NOW SUPPOSE THAT WE WANT TO USE THE
 VALUE OF y FOR $x = x_0$. THE OLS PREDICTION IS

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$$

THE ACTUAL VALUE y_0 WILL RELATE TO x_0 AS

$$y_0 = \beta_1 + \beta_2 x_0 + u_0$$

THUS THE PREDICTION ERROR WILL BE

$$e_0 = y_0 - \hat{y}_0 = -(\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2)x_0 + u_0$$

$$E(e_0) = -E[\hat{\beta}_1 - \beta_1] - x_0 E[\hat{\beta}_2 - \beta_2] + E(u_0) = 0$$

THUS \hat{Y}_0 IS AN UNBIASED PREDICTOR OF Y_0 .

11-2
EC 411

CONSIDER THE VARIANCE OF THE PREDICTION ERROR

$$\text{Var}(e_0) = E[(\hat{\beta}_1 - \beta_1)^2] + x_0^2 E[(\hat{\beta}_2 - \beta_2)^2] + 2x_0 E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)] + E[u_0^2]$$

(NOTE: u_0 IS INDEPENDENT PAST u_i 'S)

$$= \text{Var}(\hat{\beta}_1) + x_0^2 \text{Var}(\hat{\beta}_2) + 2x_0 \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) + \sigma_u^2$$

$$= \sigma_u^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2} \right] + \sigma_u^2 \frac{x_0^2}{\sum(x_i - \bar{x})^2} - \frac{2x_0\bar{x}}{\sum(x_i - \bar{x})^2} \sigma_u^2 + \sigma_u^2$$

$$= \sigma_u^2 \left[1 + \frac{1}{n} + \frac{x^2 + x_0^2 - 2x_0\bar{x}}{\sum(x_i - \bar{x})^2} \right]$$

$$= \sigma_u^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right]$$

OF COURSE, WE DO NOT KNOW σ_u^2 & MUST USE σ_u^2 FOR CONFIDENCE INTERVALS & SIGNIFICANCE TESTS.

$$\frac{e_0 - E(e_0)}{se(e_0)} \sim t_{n-2}$$

THUS
$$\frac{Y_0 - \hat{Y}_0}{se(e_0)} = t \sim t_{n-2}$$

NOTE THAT $\Pr[-t_{\alpha/2}^* \leq t \leq t_{\alpha/2}^*] = 1 - \alpha$

HENCE
$$\Pr\left[-t_{\alpha/2}^* \leq \frac{Y_0 - \hat{Y}_0}{se(e_0)} \leq t_{\alpha/2}^*\right] = 1 - \alpha$$

HENCE
$$Pr \left[-se(e_0) t_{\alpha/2}^* \leq Y_0 - \hat{Y}_0 \leq se(e_0) t_{\alpha/2}^* \right] = 1 - \alpha$$

OR
$$Pr \left[\hat{Y}_0 - se(e_0) t_{\alpha/2}^* \leq Y_0 \leq \hat{Y}_0 + se(e_0) t_{\alpha/2}^* \right] = 1 - \alpha$$

THIS DEFINES A CONFIDENCE INTERVAL FOR THE PREDICTION.

WE NOW CONSIDER THE K-VARIABLE REGRESSION

$$y = X\beta + u$$

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

$$var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

NOW CONSIDER THE PREDICTION OF y FOR $x = x_0$:

$$\hat{y}_0 = x_0 \hat{\beta}$$

AGAIN,
$$y_0 = x_0 \beta + u_0$$

$$\hat{y}_0 - y_0 = x_0 (\hat{\beta} - \beta) + u_0$$

$$E(\hat{y}_0 - y_0) = x_0 E(\hat{\beta} - \beta) + E(u_0) = 0$$

$$var(\hat{y}_0 - y_0) = x_0 E \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right] x_0' + E(u_0^2)$$

$$= \sigma^2 \left[1 + x_0 (X'X)^{-1} x_0' \right]$$

AS BEFORE, WE USE $\hat{\sigma}^2$ & DEFINE

THE STATISTIC
$$\frac{\hat{y}_0 - y_0}{se[(\hat{y}_0) - y_0]} \sim t_{n-k}$$
 & GET

THE RELEVANT CONFIDENCE INTERVAL