

Testing Ordinary Least Squares Results in the Small and Large - Normality in Small Samples: Lecture XIII

Charles B. Moss ¹

¹University of Florida

February 1, 2012

1 Readings

2 Gauss Markov

- Distribution of the Estimates
- Normal Distribution of Errors
- Estimating the Variance
- Chi-Squared Distributions

3 Readings - Next Time

Readings

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 4, Section 4.3, pp. 63-87)
- * Morgan, M.S. 1992. *The History of Econometric Ideas*. Cambridge University Press. (Chapter 3: Random Shocks Enter the Business Cycle pp. 73 - 100).
- * Theil, H. 1971. *Principles of Econometrics* John Wiley & Sons. (Chapter 3 Section 3.7)

Gauss Markov

Theorem 4.2: Gauss-Markov Theorem *In the linear regression model with regressor matrix $\mathbf{X}$, the least squares estimator $\mathbf{b}$ is the minimum variance linear unbiased estimator of β . For any vector of constants $\mathbf{w}$, the minimum variance linear unbiased estimator of $\mathbf{w}'\beta$ in the regression model is $\mathbf{w}'\mathbf{b}$, where $\mathbf{b}$ is the least squares estimator.*

Theorem 4.3: Gauss-Markov Theorem (Concluded) *In the linear regression model, the least squares estimator $\mathbf{b}$ is the minimum variance linear unbiased estimator of β whether $\mathbf{X}$ is stochastic or nonstochastic, so long as the other assumptions of the model continued to hold.*

Distribution of the Estimates

- In our development of Gauss-Markov, we derived the variance of the $\hat{\beta}$ vector (the estimated coefficient vector)

$$V(\hat{\beta}) = (X'X)^{-1} X'E[\epsilon\epsilon'] X (X'X)^{-1}. \quad (1)$$

- Under assumption of homoscedasticity $E[\epsilon\epsilon'] = \sigma^2 I$ so the variance becomes

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

- In this construction, we take the term "variance" to imply that there are distribution involved - $\hat{\beta}$ is a random variable of some time.
- The real question for our analysis is: What type of random variable is $\hat{\beta}$? How do we prove the specification of the distribution.

Normal Distribution of Errors

- As a starting point, consider the regression equation

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \quad \ni: \quad \epsilon_t \sim N(0, \sigma^2) \quad (3)$$

- The normality assumption was not required for Gauss-Markov, but we may want it for testing.
- Remembering our Mathematical Statistics - if $x_t \sim N(\mu, \sigma^2)$

$$Z = \sum_{t=1}^T x_t \sim N(T\mu, T^2\sigma^2) \quad (4)$$

given that the x_t s are independently and identically distributed (heavy on the independent).

Normal Distribution of Errors - Continued

- In addition, similarly, we have

$$\tilde{Z} = \sum_{t=1}^T a_t x_t \sim N \left(\mu \sum_{t=1}^T a_t, \sigma^2 \sum_{t=1}^T a_t^2 \right). \quad (5)$$

- Going back to Lecture IX

$$\hat{\beta} = \sum_{t=1}^T d_t E[y_t] \quad (6)$$

Normal Distribution of Errors - Continued

- Backing off the estimator side a little

$$\begin{aligned}\sum_{t=1}^T d_t y_t &= \sum_{t=1}^T d_t (x_t \beta + \epsilon_t) \\ &= \sum_{t=1}^T d_t x_t \beta + \sum_{t=1}^T d_t \epsilon_t \sim N(\beta, \sigma_b^2)\end{aligned}\tag{7}$$

Normal Distribution of Errors - Continued

- Substituting the value for d_t yields

$$\sum_{t=1}^T \frac{(x_t - \bar{x}) \epsilon_t}{\sum_{s=1}^T (x_s - \bar{x}) x_s} \quad (8)$$

so that the variance matrix is inversely proportional to the variance of the independent variable.

- The major point is that the normality of the estimated parameter flows from the normality of the residuals**
- Thus, the ordinary least squares estimates are distributed

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \quad (9)$$

- The only problem is that we don't know what σ^2 is?

Estimating the Variance

- The "natural" estimation of σ^2 involves the estimated regression errors

$$\hat{\sigma}^2 = \epsilon(\beta)' \epsilon(\beta) = (X\beta + \epsilon)' (X\beta + \epsilon) \quad (10)$$

which we have seen before.

- This time, we want a little different approach (Greene p. 62) letting $e = My = M[X\beta + \epsilon] = M\epsilon$ where $MX = 0$. Again, here think about $M = I - X(X'X)^{-1}X'$

$$\begin{aligned} MX &= (I - X(X'X)^{-1}X')X \\ &= X - X(X'X)^{-1}X'X \\ &= X - X = 0 \end{aligned} \quad (11)$$

Estimating the Variance - Continued

- To understand what is going on, consider the following "simple" 4×4 case

$$\begin{bmatrix} e_1 & e_2 & e_3 & e_4 \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \\
 \begin{bmatrix} e_1 & e_2 & e_3 & e_4 \end{bmatrix} \begin{bmatrix} e_1 M_{11} + e_2 M_{12} + e_3 M_{13} + e_4 M_{14} \\ e_1 M_{21} + e_2 M_{22} + e_3 M_{23} + e_4 M_{24} \\ e_1 M_{31} + e_2 M_{32} + e_3 M_{33} + e_4 M_{34} \\ e_1 M_{41} + e_2 M_{42} + e_3 M_{43} + e_4 M_{44} \end{bmatrix} \quad (12) \\
 \sum_{i=1}^4 \sum_{j=1}^4 M_{ij} e_i e_j$$

Estimating the Variance - Continued

- Moving to the next series of results relies on the trace results (Greene p. 999) $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CBA)$ (given that each multiplication is defined).
- Applying this to the variance estimate

$$e'e = \epsilon' M \epsilon \Rightarrow \text{tr} [M \epsilon \epsilon'] \quad (13)$$

- Working on the matrices

$$\text{tr} \left(\begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \begin{bmatrix} e_1 e_1 & e_1 e_2 & e_1 e_3 & e_1 e_4 \\ e_2 e_1 & e_2 e_2 & e_2 e_3 & e_2 e_4 \\ e_3 e_1 & e_3 e_2 & e_3 e_3 & e_3 e_4 \\ e_4 e_1 & e_4 e_2 & e_4 e_3 & e_4 e_4 \end{bmatrix} \right) \quad (14)$$

Estimating the Variance - Continued

- Focusing on the diagonal of this multiplication

$$\begin{bmatrix} M_{11} \\ M_{12} \\ M_{13} \\ M_{14} \end{bmatrix}' \begin{bmatrix} e_1 e_1 \\ e_1 e_2 \\ e_1 e_3 \\ e_1 e_4 \end{bmatrix} \quad \begin{bmatrix} M_{21} \\ M_{22} \\ M_{23} \\ M_{24} \end{bmatrix}' \begin{bmatrix} e_1 e_2 \\ e_2 e_2 \\ e_2 e_3 \\ e_2 e_4 \end{bmatrix} \quad \dots \quad \begin{bmatrix} M_{41} \\ M_{42} \\ M_{43} \\ M_{44} \end{bmatrix}' \begin{bmatrix} e_1 e_4 \\ e_2 e_4 \\ e_3 e_4 \\ e_4 e_4 \end{bmatrix} \quad (15)$$

$$\Rightarrow \text{tr}(Mee') = \sum_{i=1}^4 \sum_{j=1}^4 M_{ij} e_i e_j = e' M e$$

Estimating the Variance - Continued

- Given that the errors are independently and identically distributed

$$\begin{aligned} E [\text{tr} (Mee')] &= \text{tr} (ME [ee']) \\ &= \text{tr} (M [\sigma^2 I]) \\ &= \sigma^2 \text{tr} (M) \end{aligned} \tag{16}$$

- Last game with the traces

$$\begin{aligned} \text{tr} [M] &= \text{tr} [I - X (X'X)^{-1} X'] = \text{tr} [I] - \text{tr} [(X'X)^{-1} X'X] \\ &= n - k \end{aligned} \tag{17}$$

- Thus $E [e'e] = (n - k) \sigma^2$.

Chi-Squared Distributions

- The σ^2 statistic is based on the square of normal distributions (defined as the χ_n^2 distribution where n is the number of independent chi^2 distributions).
- As a starting place, assume that we only have two points - the standard chi^2 formulation would be

$$(y_1 - \mu_y)^2 + (y_2 - \mu_y)^2 = y_1^2 - 2y_1\mu_y + \mu_y^2 + y_2^2 - 2y_2\mu_y + \mu_y^2 \quad (18)$$

taking the expectation

$$E[y_1^2 - 2y_1\mu_y + \mu_y^2 + y_2^2 - 2y_2\mu_y + \mu_y^2] = y_1^2 + y_2^2 - 2\mu_y^2 \quad (19)$$

Chi-Squared Distributions - Continued

- Taking a slightly different notion

$$(y_1 - y_2)^2 \Rightarrow E[y_1^2 - 2y_1y_2 + y_2^2] \quad (20)$$

- If I can make the assumption that the two variables are independent $E[y_1y_2] = \mu_y\mu_y$ implying

$$E[(y_1 - \mu_y)^2 + (y_2 - \mu_y)^2] = y_1^2 + y_2^2 - 2\mu_y^2 = E[(y_1 - y_2)^2] \quad (21)$$

- Thus $E[(y_1 - \mu_y)^2 + (y_2 - \mu_y)^2] \sim \chi_1^2$.

Chi-Squared Distributions - Continued

- Extending this to three observations

$$E \left[(y_1 - \mu_y)^2 + (y_2 - \mu_y)^2 + (y_3 - \mu_y)^2 \right] \sim \chi_2^2 \quad (22)$$

- Next, suppose that we are actually dealing with a single variable regression so that $\mu_y(x) \Rightarrow \alpha_0 + \alpha_1 x$. In this case looking at two observations

$$E \left[(y_1 - (\alpha_0 + \alpha_1 x_1))^2 + (y_2 - (\alpha_0 + \alpha_1 x_2))^2 \right] = 0 \quad (23)$$

or we can solve for a combination of α_0 and α_1 which exactly explains x_1 and x_2 .

Chi-Squared Distributions - Continued

- If we had three observations

$$E \left[(y_1 - (\alpha_0 + \alpha_1 x_1))^2 + (y_2 - (\alpha_0 + \alpha_1 x_2))^2 + (y_3 - (\alpha_0 + \alpha_1 x_3))^2 \right] \sim \chi^2_3 \quad (24)$$

- Note also by extension of the reasoning that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^3 (y_i - (\alpha_0 + \alpha_1 x_i))^2}{3 - 2} \quad (25)$$

Chi-Squared Distributions - Continued

- In its general form

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T (y_i - (\alpha_0 + \alpha_1 X'_{i.}))^2}{T - (k + 1)} \quad (26)$$

Readings - Next Time

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 4 Sections 4.4-4.3 pp. 63 - 87)
- * Hansen, B.E. 2011. *Econometrics* Draft Textbook Website: <http://www.ssc.wisc.edu/bhansen>. (Chapter 6).
- * Morgan, M.S. 1992. *The History of Econometric Ideas*. Cambridge University Press. (Chapter 4: Tinbergen and Macrodynamic Models pp. 101-130).
- Theil, H. 1971. *Principles of Econometrics* John Wiley & Sons. (Chapter 8)