

Derivation of the Least Squares Model: Lecture VIII

Charles B. Moss ¹

¹University of Florida

January 26, 2012

- 1 Readings
- 2 Derivation of Ordinary Least Squares
- 3 Normal Equations
 - Implications of Normal Equations
- 4 Partial Regression Analysis
 - Partial Regression Adjustment
 - Frisch-Waugh Theorem
 - Regression With and Without a Constant
- 5 Goodness of Fit Analysis
- 6 Readings - Next Time

Readings

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 3)
- * Hansen, B.E. 2011. *Econometrics* Draft Textbook Website: <http://www.ssc.wisc.edu/~bhansen>. (Chapter 5).
- Hayashi, F. 2000. *Econometrics* Oxford University Press (Section 1.2 - 1.4 pp. 15-46)
- * Moss, C.B. 2012. *Mathematical Statistics for Applied Econometrics* Draft Textbook Website: <http://ricardo.ifas.ufl.edu/aeb6571/econometrics/MathStat2012.pdf>. (Chapter 10, pp. 149 - 159).
- * Popper, K. 2010. *The Logic of Scientific Discovery* Routledge Classics [Chapter 8: Probability, 133-208].

Readings - Continued

- Stewart, G.W. 1995. Gauss, Statistics, and Gaussian Elimination. *Journal of Computational and Graphical Statistics* 4(1), 1-11. (<http://www.jstor.org/stable/1390624>).
- Stigler S.M. 1986. *The History of Statistics*. Harvard University Press.
- Theil, H. 1971. *Principles of Econometrics* John Wiley & Sons. (Chapter 3)

Derivation of Ordinary Least Squares

- We trudged through the derivation of the least squares estimator in matrix form in Lecture VI.
- In connecting back to Greene (Chapter 3) it worth a brief look at normal equations

$$e'e = y'y - b_0'X'y - y'Xb_0 + b_0'X'Xb_0$$

$$S(b_0) = y'y - 2y'Xb_0 + b_0'X'Xb_0 \quad (1)$$

$$\Rightarrow \frac{\partial S(b_0)}{\partial b_0} = -2X'y + 2X'Xb_0 = 0.$$

Normal Equations

- This last equation implies the **least squares normal equations**

$$X'Xb = X'y \quad (2)$$

- These normal equations are actually traced back to Carl Friedrich Gauss and his work on the Gaussian or normal distribution (Stigler 1986, Stewart 1995).
- The normal equations are derived from the first-order conditions of the normal distribution

Normal Equations - Continued

- Using the data from KLEM-Food.xls (the KLEM data for Food and Kindred products) the $X'X$ matrix (with the vector of ones for the intercept) becomes

$$X'X = \begin{bmatrix} 45.0000 & 1.3889 & 2.4383 \\ 1.3889 & 0.2929 & 0.1707 \\ 2.4383 & 0.1707 & 0.6559 \end{bmatrix} \quad (3)$$

$$X'y = \begin{bmatrix} 0.5578 \\ 0.1446 \\ 0.0420 \end{bmatrix} \quad (4)$$

Normal Equations - Continued

- The normal equations then become

$$\begin{bmatrix} 45.0000 & 1.3889 & 2.4383 \\ 1.3889 & 0.2929 & 0.1707 \\ 2.4383 & 0.1707 & 0.6559 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.5578 \\ 0.1446 \\ 0.0420 \end{bmatrix} \quad (5)$$

$$45.0000\alpha_0 + 1.3889\alpha_1 + 2.4383\alpha_2 = 0.5578$$

$$1.3889\alpha_0 + 0.2929\alpha_1 + 0.1707\alpha_2 = 0.1446$$

$$2.4383\alpha_0 + 0.1707\alpha_1 + 0.6559\alpha_2 = 0.0420$$

- Greene (pp. 28-29) presents an heuristic solution for these normal equations.

Implications of Normal Equations

- Greene (pp. 30-31) also presents three implications of the least squares solution:

$$X'Xb - X'y = -X'(y - Xb) = -X'e = 0. \quad (6)$$

- *The least squares residuals sum to zero.*
 $x_1'e = i'e = \sum_i e_i = 0.$
- *The regression hyperplane passes through the point of means of the data. $\hat{y} = \bar{x}'b.$*
- *The mean of the fitted values from the regression equals the mean of the actual values. $\hat{y} = Xb.$*

Partial Regression

- Greene 3.3
- Regressions based on projected variance matrices

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \Rightarrow \hat{\beta} = \Sigma_{22}^{-1} \Sigma_{21} \quad (7)$$

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \Rightarrow \hat{\beta} = \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{21} \\ \Sigma_{31} \end{bmatrix} \quad (8)$$

Partial Regression - Continued

- Returning to the KLEM - Food data, consider two sets of regressor variables:
 - ind1 - the constant, log change in the price of output, and log change in the price of energy.
 - ind2 - the log change in the price capital, the log change in the price of labor, and the log change in the price of materials.

$$\begin{bmatrix} \sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \quad (9)$$

Partial Regression - Continued

$$\begin{bmatrix} \sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} =$$

$$\begin{bmatrix} 0.006351 & 0.012396 & 0.003214 & 0.000934 & 0.004917 & 0.001854 & 0.001875 \\ 0.012396 & 1.000000 & 0.030864 & 0.054183 & 0.053743 & 0.061489 & 0.032448 \\ 0.003214 & 0.030864 & 0.006509 & 0.003793 & 0.009262 & 0.002469 & 0.003637 \\ 0.000934 & 0.054183 & 0.003793 & 0.014576 & 0.005233 & 0.004138 & 0.003843 \\ 0.004917 & 0.053743 & 0.009262 & 0.005233 & 0.027972 & 0.002668 & 0.004370 \\ 0.001854 & 0.061489 & 0.002469 & 0.004138 & 0.002668 & 0.006271 & 0.002317 \\ 0.001875 & 0.032448 & 0.003637 & 0.003843 & 0.004370 & 0.002317 & 0.002746 \end{bmatrix} \quad (10)$$

Partial Regression Results

Variables	Full Regression	Regression 1	Regression 2
Constant	-0.03916	-0.00013	
Output Price	-0.09970	0.53834	
Energy Price	-0.18999	-0.07553	
Capital Price	0.12513		0.09610
Labor Price	0.42759		0.08564
Material Price	0.98355		0.45761

Partial Regression Adjustment

- From our discussion of partitioned inverses:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

$$\begin{bmatrix} \Sigma_{22}^{-1} + \Sigma_{22}^{-1} \Sigma_{23} (\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \Sigma_{32} \Sigma_{22}^{-1} & -\Sigma_{22}^{-1} \Sigma_{23} (\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \\ -(\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \Sigma_{32} \Sigma_{22}^{-1} & (\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \end{bmatrix} \\ \times \begin{bmatrix} \Sigma_{21} \\ \Sigma_{31} \end{bmatrix}$$

$$\hat{\beta}_3 = -(\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{21} + (\Sigma_{33} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{23})^{-1} \Sigma_{31}$$

(11)

Partial Regression Adjustment - Continued

- Implementing this expression yields

$$\hat{\beta}_3 = \begin{bmatrix} 0.12513 \\ 0.42759 \\ 0.98356 \end{bmatrix} \quad (12)$$

- Which is consistent with the results for the full regression.

In the linear least squares regression of vector y on two sets of variables X_1 and X_2 , the subvector b_2 is the set of coefficients obtained when the residuals from regression y on X_1 alone are regressed on the set of residuals obtained when each column of X_2 is regressed on X_1 .

Regression With and Without a Constant

- To this point I have used the specification including a vector of ones for the constant.
- Another approach is to "normalize" the regression by using the variance matrix and then imputing the constant.

$$\tilde{X} = \begin{bmatrix} 1 & X \end{bmatrix} \Rightarrow \hat{\beta} = (\tilde{X}'\tilde{X})^{-1} (\tilde{X}'y)$$

$$\Sigma_{XX} = \frac{(X - \mu_X)'(X - \mu_X)}{N}, \quad \Sigma_{Xy} = \frac{(X - \mu_X)'(y - \mu_y)}{N}$$

$$\Rightarrow \hat{\beta}_{2:k} = \Sigma_{XX}^{-1} \Sigma_{Xy}, \quad \hat{\beta}_1 = \mu_y - \mu_X \hat{\beta}_{2:k}$$

(13)

- Returning to the simple regression example (log change in energy use on log change in output price and energy price).

$$\hat{\Sigma} = \begin{bmatrix} \sigma_y y & \Sigma_{yX} \\ \Sigma_{Xy} & \Sigma_{XX} \end{bmatrix}$$

$$= \begin{bmatrix} 0.00634 & 0.00290 & 0.00027 \\ 0.00290 & 0.00568 & 0.00217 \\ 0.00027 & 0.00217 & 0.01190 \end{bmatrix} \quad (14)$$

$$\hat{\beta} = \begin{bmatrix} 0.00568 & 0.00217 \\ 0.00217 & 0.01190 \end{bmatrix}^{-1} \begin{bmatrix} 0.00290 \\ 0.00027 \end{bmatrix} = \begin{bmatrix} 0.53834 \\ -0.075523 \end{bmatrix}$$

$$\alpha_0 = \bar{y} - \mu'_X \hat{\beta} = -0.00013$$

Goodness of Fit Analysis

- Most statistics on the Goodness of Fit compare the explanatory power of the regression model against a simple mean.
- The first thing we want to define is the **total variation**

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (15)$$

- This is difficult to mesh-up with the least squares formulation we have been using

$$\begin{aligned} SSE &= \left(y - X(X'X)^{-1}X'y \right)' \left(y - X(X'X)^{-1}X'y \right) \\ &= y'y - y'X(X'X)^{-1}X'y \Rightarrow SSE = y'My \ni M = \left(I - X(X'X)^{-1}X' \right) \end{aligned} \quad (16)$$

- The M^0 in Greene is somewhat different. M^0 subtracts out the means.
- In this case M^0 is a sweep matrix

$$Q_T = I_{T \times T} - \frac{1}{T} ee' \quad (17)$$

- Take a 3×3

$$Q_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (18)$$

$$x_1 Q_3 = \begin{bmatrix} x_{11} - \frac{1}{3}(x_{11} + x_{12} + x_{13}) \\ x_{12} - \frac{1}{3}(x_{11} + x_{12} + x_{13}) \\ x_{13} - \frac{1}{3}(x_{11} + x_{12} + x_{13}) \end{bmatrix}'$$

- SST can then be defined as

$$SST = Y' M^0 y \ni M^0 = \frac{Q_T Q_T}{T} \Rightarrow M^0 = \frac{Q_T}{T} \quad (19)$$

with the last equality due to the fact that Q_T is idempotent.

- The R^2 is then defined as

$$R^2 = \frac{SSR}{SST} = \frac{b' X' M^0 X b}{y' M^0 y} = 1 - \frac{e' e}{y' M^0 y} \quad (20)$$

Readings - Next Time

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 4.1-4.3 pp. 51 - 63)
- * Hansen, B.E. 2011. Econometrics Draft Textbook Website: <http://www.ssc.wisc.edu/bhansen>. (Chapter 5).
- Hayashi, F. 2000. *Econometrics* Oxford University Press (Section 1.5 pp. 47-54)
- * Popper, K. 2010. *The Logic of Scientific Discovery* Routledge Classics [Chapter 10: Corroboration, or How a Theory Stands Up to Tests, 248-282].