

Econ 220B, Winter 2008
James Hamilton

Problem Set 1
Due Monday, January 14

1.) Let \mathbf{X} be a $(T \times k)$ matrix whose columns are linearly independent, and let $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Show that \mathbf{M} is symmetric and idempotent. Calculate the eigenvalues and rank of \mathbf{M} , and show that it is positive semidefinite.

2.) Consider a regression of y_t on \mathbf{x}_t where the first element of \mathbf{x}_t is a constant term. The R^2 or coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \mathbf{x}'_t \mathbf{b})^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

where \mathbf{b} is the OLS regression coefficient and \bar{y} is the sample mean. Show that $0 \leq R^2 \leq 1$.

3.) Let \mathbf{P} be a nonsingular symmetric $(k_1 \times k_1)$ matrix, \mathbf{Q} a nonsingular symmetric $(k_2 \times k_2)$ matrix, and \mathbf{R} an arbitrary $(k_1 \times k_2)$ matrix. Verify the following formula for the inverse of a partitioned matrix:

$$\begin{bmatrix} \mathbf{P} & \mathbf{R} \\ \mathbf{R}' & \mathbf{Q} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{W} & -\mathbf{WRQ}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{R}'\mathbf{W} & (\mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{R}'\mathbf{WRQ}^{-1}) \end{bmatrix}$$

for $\mathbf{W} = (\mathbf{P} - \mathbf{RQ}^{-1}\mathbf{R}')^{-1}$.

4.) Consider a regression of y_t on \mathbf{x}_t , where we partition the regressors into two groups: $\mathbf{x}_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t})'$ where k_1 of the variables are included in the subvector \mathbf{x}_{1t} and the remaining k_2 variables are in \mathbf{x}_{2t} :

$$y_t = \mathbf{x}'_{1t}\boldsymbol{\beta}_1 + \mathbf{x}'_{2t}\boldsymbol{\beta}_2 + \varepsilon_t.$$

The usual OLS regression coefficients are of course given by

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

for $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$ a $[T \times (k_1 + k_2)]$ matrix and \mathbf{X}_i the $(T \times k_i)$ matrix whose t th row is \mathbf{x}'_{it} . Use the results from question (3) to show that the OLS estimate \mathbf{b}_1 could equivalently be calculated as follows: (a) regress y_t on \mathbf{x}_{2t} alone and calculate the residuals e_{2t} , for e_{2t} the t th element of $\mathbf{e}_2 = \mathbf{M}_2\mathbf{y}$ with $\mathbf{M}_2 = \mathbf{I}_T - \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2$; regress each element of \mathbf{x}_{1t} on \mathbf{x}_{2t} and calculate the residuals $\tilde{\mathbf{x}}_{1t}$, where $\tilde{\mathbf{x}}'_{1t}$ is the t th row of $\mathbf{M}_2\mathbf{X}_1$; (c) regress e_{2t} on $\tilde{\mathbf{x}}_{1t}$, to obtain a $(k_1 \times 1)$ vector $\hat{\boldsymbol{\beta}}_1$ that is numerically identical to \mathbf{b}_1 given above.

Econ 220B, Winter 2008
James Hamilton

5.) In your first article submitted for publication, you report the results from a regression of the unemployment rate in city t , denoted y_t , on a constant, income per person in the city measured in dollars per capita (x_{2t}), and the fraction of the city's working-age population who are high-school graduates (x_{3t}). You obtain the following results, with regression standard errors in parentheses:

$$y_t = \underset{(0.52)}{6.37} - \underset{(0.00016)}{0.00037} x_{2t} - \underset{(8.32)}{21.56} x_{3t} \quad R^2 = 0.19.$$

You are delighted to learn that the journal has decided to accept your paper, except that the referee wants one small change: he insists that you should measure income in units of thousands of dollars per person rather than dollars per person. Unfortunately, the data you used are irretrievably lost. Fortunately, you understood the formulas for OLS sufficiently deeply that you know exactly what the new regression coefficients, standard errors, and R^2 would have turned out to be, if you were to run the new regression as instructed. What are they?