FIRST MIDTERM EXAM ECON 7801 SPRING 2001

ECONOMICS DEPARTMENT, UNIVERSITY OF UTAH

Problem 1. *2 points Let $\boldsymbol{y}$ be a n-vector. (It may be a vector of observations of a random variable y, but it does not matter how the y_i were obtained.) Prove that the scalar α which minimizes the sum*

$$(y_1 - \alpha)^2 + (y_2 - \alpha)^2 + \cdots + (y_n - \alpha)^2 = \sum (y_i - \alpha)^2 \tag{1}$$

is the arithmetic mean $\alpha = \bar{y}$.

Answer. Use (**??**). □

Problem 2.

- **a.** *2 points Verify that the matrix $\boldsymbol{D} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$ is symmetric and idempotent.*

Date of exam Tuesday, February 20, 9–10:30 am.

• **b.** *1 point Compute the trace* $\operatorname{tr} \boldsymbol{D}$.

Answer. $\operatorname{tr} \boldsymbol{D} = n - 1$. One can see this either by writing down the matrix element by element, or use the linearity of the trace plus the rule that $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$. $\operatorname{tr} \boldsymbol{I} = n$ and $\operatorname{tr}(\boldsymbol{\iota}\boldsymbol{\iota}^\top) = \operatorname{tr}(\boldsymbol{\iota}^\top\boldsymbol{\iota}) = \operatorname{tr} n = n$. □

• **c.** *1 point For any vector of observations* $\boldsymbol{y}$ *compute* $\boldsymbol{D}\boldsymbol{y}$.

Answer. Element by element one can write

$$\boldsymbol{D}\boldsymbol{y} = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \tag{2}$$

There is also a more elegant matrix theoretical proof available □

• **d.** *1 point Is there a vector* $\boldsymbol{a} \neq \boldsymbol{o}$ *for which* $\boldsymbol{D}\boldsymbol{a} = \boldsymbol{o}$? *If so, give an example of such a vector.*

Answer. $\boldsymbol{\iota}$ is, up to a scalar factor, the only nonzero vector with $\boldsymbol{D}\boldsymbol{\iota} = \boldsymbol{o}$. □

• **e.** *1 point Show that the sample variance of a vector of observations* $\boldsymbol{y}$ *can be written in matrix notation as*

$$\textit{The sample variance of } \boldsymbol{y} \textit{ is } \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{1}{n}\boldsymbol{y}^\top\boldsymbol{D}\boldsymbol{y} \tag{3}$$

Answer. Let's get rid of the factor $\frac{1}{n}$ which appears on both sides: we have to show that

$$\sum (y_i - \bar{y})^2 = \boldsymbol{y}^\top \boldsymbol{D}\boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{y} \tag{4}$$

This is the squared length of the vector $\boldsymbol{D}\boldsymbol{y}$ which we computed in part c. □

Problem 3. *1 point Compute the matrix product*

$$\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 3 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 2 & 1 \\ 3 & 8 \end{bmatrix}$$

Answer.

$$\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 3 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 2 & 1 \\ 3 & 8 \end{bmatrix} = \begin{bmatrix} 1\cdot 4 + 2\cdot 2 + 4\cdot 3 & 1\cdot 0 + 2\cdot 1 + 4\cdot 8 \\ 0\cdot 4 + 3\cdot 2 + 3\cdot 3 & 0\cdot 0 + 3\cdot 1 + 3\cdot 8 \end{bmatrix} = \begin{bmatrix} 20 & 34 \\ 15 & 27 \end{bmatrix}$$

□

Problem 4. *2 points Assume that $\boldsymbol{X}$ has full column rank. Show that $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{M}\boldsymbol{y}$ where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Show that $\boldsymbol{M}$ is symmetric and idempotent.*

Answer. By definition, $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} = \Big(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}\Big)\boldsymbol{y}$. Idempotent, i.e. $\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$:

(5)

$$\boldsymbol{M}\boldsymbol{M} = \Big(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top\Big)\Big(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top\Big) \quad = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

$\square$

Problem 5. *2 points We are in the multiple regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with intercept, i.e., $\boldsymbol{X}$ is such that there is a vector $\boldsymbol{a}$ with $\boldsymbol{\iota} = \boldsymbol{X}\boldsymbol{a}$. Define the row vector $\bar{\boldsymbol{x}}^\top = \frac{1}{n}\boldsymbol{\iota}^\top\boldsymbol{X}$, i.e., it has as its jth component the sample mean of the jth independent variable. Using the normal equations $\boldsymbol{X}^\top\boldsymbol{y} = \boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}}$, show that $\bar{y} = \bar{\boldsymbol{x}}^\top\hat{\boldsymbol{\beta}}$ (i.e., the regression plane goes through the center of gravity of all data points).*

Answer. Premultiply the normal equation by $\boldsymbol{a}^\top$ to get $\boldsymbol{\iota}^\top\boldsymbol{y} - \boldsymbol{\iota}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} = 0$. Premultiply by $1/n$ to get the result. $\square$

Problem 6. *2 points Let $\boldsymbol{\theta}$ be a vector of possibly random parameters, and $\hat{\boldsymbol{\theta}}$ an estimator of $\boldsymbol{\theta}$. Show that*

$$\mathcal{MSE}[\hat{\boldsymbol{\theta}};\boldsymbol{\theta}] = \mathcal{V}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] + (\mathcal{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}])(\mathcal{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}])^\top. \tag{6}$$

Don't assume the scalar result but make a proof that is good for vectors and scalars.

Answer. For any random vector $\boldsymbol{x}$ follows

$$\begin{aligned}
\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\top] &= \mathcal{E}\Big[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}] + \mathcal{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}] + \mathcal{E}[\boldsymbol{x}])^\top\Big] \\
&= \mathcal{E}\Big[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])^\top\Big] - \mathcal{E}\Big[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])\,\mathcal{E}[\boldsymbol{x}]^\top\Big] - \mathcal{E}\Big[\mathcal{E}[\boldsymbol{x}](\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])^\top\Big] + \mathcal{E}\Big[\mathcal{E}[\boldsymbol{x}]\,\mathcal{E}[\boldsymbol{x}]^\top\Big] \\
&= \mathcal{V}[\boldsymbol{x}] - \boldsymbol{O} - \boldsymbol{O} + \mathcal{E}[\boldsymbol{x}]\,\mathcal{E}[\boldsymbol{x}]^\top.
\end{aligned}$$

Setting $\boldsymbol{x} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ the statement follows. $\square$

Problem 7. *Consider two very simple-minded estimators of the unknown nonrandom parameter vector* $\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$. *Neither of these estimators depends on any observations, they are constants. The first estimator is* $\hat{\boldsymbol{\phi}} = \left[\begin{smallmatrix} 11 \\ 11 \end{smallmatrix}\right]$, *and the second is* $\tilde{\boldsymbol{\phi}} = \left[\begin{smallmatrix} 12 \\ 8 \end{smallmatrix}\right]$.

• **a.** *2 points Compute the* $\mathcal{MSE}$*-matrices of these two estimators if the true value of the parameter vector is* $\boldsymbol{\phi} = \left[\begin{smallmatrix} 10 \\ 10 \end{smallmatrix}\right]$. *For which estimator is the trace of the* $\mathcal{MSE}$ *matrix smaller?*

Answer. $\hat{\boldsymbol{\phi}}$ has smaller trace of the $\mathcal{MSE}$-matrix.

$$\hat{\boldsymbol{\phi}} - \boldsymbol{\phi} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathcal{MSE}[\hat{\boldsymbol{\phi}}; \boldsymbol{\phi}] = \mathcal{E}[(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})^\top]$$

$$= \mathcal{E}[\begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}] = \mathcal{E}[\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

$$\mathcal{MSE}[\tilde{\boldsymbol{\phi}}; \boldsymbol{\phi}] = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix}$$

Note that both $\mathcal{MSE}$-matrices are singular, i.e., both estimators allow an error-free look at certain linear combinations of the parameter vector. □

• **b.** *1 point Give two vectors* $\boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ *and* $\boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$ *satisfying* $\text{MSE}[\boldsymbol{g}^\top\hat{\boldsymbol{\phi}}; \boldsymbol{g}^\top\boldsymbol{\phi}] < \text{MSE}[\boldsymbol{g}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{g}^\top\boldsymbol{\phi}]$ *and* $\text{MSE}[\boldsymbol{h}^\top\hat{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi}] > \text{MSE}[\boldsymbol{h}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi}]$ *(*$\boldsymbol{g}$ *and* $\boldsymbol{h}$ *are not unique; there are many possibilities).*

Answer. With $\boldsymbol{g} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\boldsymbol{h} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for instance we get $\boldsymbol{g}^\top\hat{\boldsymbol{\phi}} - \boldsymbol{g}^\top\boldsymbol{\phi} = 0$, $\boldsymbol{g}^\top\tilde{\boldsymbol{\phi}} - \boldsymbol{g}^\top\boldsymbol{\phi} = 4$, $\boldsymbol{h}^\top\hat{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi} = 2$, $\boldsymbol{h}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi} = 0$, therefore $\text{MSE}[\boldsymbol{g}^\top\hat{\boldsymbol{\phi}}; \boldsymbol{g}^\top\boldsymbol{\phi}] = 0$, $\text{MSE}[\boldsymbol{g}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{g}^\top\boldsymbol{\phi}] = 16$, $\text{MSE}[\boldsymbol{h}^\top\hat{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi}$ 4, $\text{MSE}[\boldsymbol{h}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi}] = 0$. An alternative way to compute this is e.g.

$$\mathcal{MSE}[\boldsymbol{h}^\top\tilde{\boldsymbol{\phi}}; \boldsymbol{h}^\top\boldsymbol{\phi}] = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 16$$

□

• **c.** *1 point Show that neither* $\mathcal{MSE}[\hat{\boldsymbol{\phi}}; \boldsymbol{\phi}] - \mathcal{MSE}[\tilde{\boldsymbol{\phi}}; \boldsymbol{\phi}]$ *nor* $\mathcal{MSE}[\tilde{\boldsymbol{\phi}}; \boldsymbol{\phi}] - \mathcal{MSE}[\hat{\boldsymbol{\phi}}; \boldsymbol{\phi}]$ *is a nonnegative definite matrix. Hint: you are allowed to use the mathematical fact that if a matrix is nonnegative definite, then its determinant is nonnegative.*

Answer.

$$\mathcal{MSE}[\tilde{\boldsymbol{\phi}}; \boldsymbol{\phi}] - \mathcal{MSE}[\hat{\boldsymbol{\phi}}; \boldsymbol{\phi}] = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \tag{7}$$

Its determinant is negative, and the determinant of its negative is also negative. □

Problem 8.

• **a.** *2 points Show that the sampling error of the OLS estimator is*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon} \tag{8}$$

• **b.** *2 points Derive from this that $\hat{\boldsymbol{\beta}}$ is unbiased and that its $\mathcal{MSE}$-matrix is*

$$\mathcal{MSE}[\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}] = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \tag{9}$$

Problem 9.

• **a.** *2 points Show that*

$$SSE = \boldsymbol{\varepsilon}^\top \boldsymbol{M} \boldsymbol{\varepsilon} \qquad \text{where} \qquad \boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \tag{10}$$

Answer. $SSE = \hat{\varepsilon}^\top \hat{\varepsilon}$, where $\hat{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{M}\boldsymbol{y}$ where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$. From $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{O}$ follows $\hat{\varepsilon} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{M}\boldsymbol{\varepsilon}$. Since $\boldsymbol{M}$ is idempotent and symmetric, it follows $\hat{\varepsilon}^\top \hat{\varepsilon} = \boldsymbol{\varepsilon}^\top \boldsymbol{M} \boldsymbol{\varepsilon}$. □

• **b.** *1 point Is SSE observed? Is $\boldsymbol{\varepsilon}$ observed? Is $\boldsymbol{M}$ observed?*

• **c.** *3 points Under the usual assumption that $\boldsymbol{X}$ has full column rank, show that*

$$\mathrm{E}[SSE] = \sigma^2 (n - k) \tag{11}$$

Answer. $\mathrm{E}[\hat{\varepsilon}^\top \hat{\varepsilon}] = \mathrm{E}[\operatorname{tr} \boldsymbol{\varepsilon}^\top \boldsymbol{M} \boldsymbol{\varepsilon}] = \mathrm{E}[\operatorname{tr} \boldsymbol{M} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top] = \sigma^2 \operatorname{tr} \boldsymbol{M} = \sigma^2 \operatorname{tr}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top) = \sigma^2 (n - \operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{X}) = \sigma^2 (n - k)$. □

Problem 10. *The prediction problem in the Ordinary Least Squares model can be formulated as follows:*

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X}_0 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_0 \end{bmatrix} \qquad \mathcal{E}[\begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_0 \end{bmatrix}] = \begin{bmatrix} \boldsymbol{o} \\ \boldsymbol{o} \end{bmatrix} \qquad \mathcal{V}[\begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_0 \end{bmatrix}] = \sigma^2 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix}. \tag{12}$$

$\boldsymbol{X}$ and $\boldsymbol{X}_0$ are known, $\boldsymbol{y}$ is observed, $\boldsymbol{y}_0$ is not observed.

• **a.** *4 points Show that $\boldsymbol{y}_0^* = \boldsymbol{X}_0\hat{\boldsymbol{\beta}}$ is the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{y}_0$ on the basis of $\boldsymbol{y}$, where $\hat{\boldsymbol{\beta}}$ is the OLS estimate in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.*

Answer. Take any other predictor $\tilde{\boldsymbol{y}}_0 = \tilde{\boldsymbol{B}}\boldsymbol{y}$ and write $\tilde{\boldsymbol{B}} = \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D}$. Unbiasedness means $\mathcal{E}[\tilde{\boldsymbol{y}}_0 - \boldsymbol{y}_0] = \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_0\boldsymbol{\beta} = \boldsymbol{o}$, from which follows $\boldsymbol{D}\boldsymbol{X} = \boldsymbol{O}$. Because of unbiasedness we know $\mathcal{MSE}[\tilde{\boldsymbol{y}}_0; \boldsymbol{y}_0] = \mathcal{V}[\tilde{\boldsymbol{y}}_0 - \boldsymbol{y}_0]$. Since the prediction error can be written $\tilde{\boldsymbol{y}}_0 - \boldsymbol{y} = \begin{bmatrix} \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_0 \end{bmatrix}$, one obtains

$$\begin{aligned}
\mathcal{V}[\tilde{\boldsymbol{y}}_0 - \boldsymbol{y}_0] &= \begin{bmatrix} \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D} & -\boldsymbol{I} \end{bmatrix} \mathcal{V}[\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_0 \end{bmatrix}] \begin{bmatrix} \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_0^\top + \boldsymbol{D}^\top \\ -\boldsymbol{I} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_0^\top + \boldsymbol{D}^\top \\ -\boldsymbol{I} \end{bmatrix} \\
&= \sigma^2 \Big(\boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D}\Big)\Big(\boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{D}\Big)^\top + \sigma^2\boldsymbol{I} \\
&= \sigma^2 \Big(\boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_0^\top + \boldsymbol{D}\boldsymbol{D}^\top + \boldsymbol{I}\Big).
\end{aligned}$$

This is smallest for $\boldsymbol{D} = \boldsymbol{O}$. □

• **b.** *2 points The same* $\boldsymbol{y}_0^* = \boldsymbol{X}_0\hat{\boldsymbol{\beta}}$ *is also the Best Linear Unbiased Estimator of* $\boldsymbol{X}_0\boldsymbol{\beta}$, *which is the expected value of* $\boldsymbol{y}_0$. *You are not required to re-prove this here, but you are asked to compute* $\mathcal{MSE}[\boldsymbol{X}_0\hat{\boldsymbol{\beta}}; \boldsymbol{X}_0\boldsymbol{\beta}]$ *and compare it with* $\mathcal{MSE}[\boldsymbol{y}_0^*; \boldsymbol{y}_0]$. *Can you explain the difference?*

Answer. Estimation error and $\mathcal{MSE}$ are

$$\boldsymbol{X}_0\hat{\boldsymbol{\beta}} - \boldsymbol{X}_0\boldsymbol{\beta} = \boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon} \quad \text{due to (8)}$$

$$\mathcal{MSE}[\boldsymbol{X}_0\hat{\boldsymbol{\beta}}; \boldsymbol{X}_0\boldsymbol{\beta}] = \mathcal{V}[\boldsymbol{X}_0\hat{\boldsymbol{\beta}} - \boldsymbol{X}_0\boldsymbol{\beta}] = \mathcal{V}[\boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}] = \sigma^2\boldsymbol{X}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_0^\top.$$

It differs from the prediction $\mathcal{MSE}$ matrix by $\sigma^2\boldsymbol{I}$, which is the uncertainty about the value of the new disturbance $\boldsymbol{\varepsilon}_0$ about which the data have no information. □

Problem 11. *Given a regression with a constant term and two explanatory variables which we will call x and z, i.e.,*

$$y_t = \alpha + \beta x_t + \gamma z_t + \varepsilon_t \tag{13}$$

• **a.** *1 point How will you estimate β and γ if it is known that $\beta = \gamma$?*

Answer. Write

$$y_t = \alpha + \beta(x_t + z_t) + \varepsilon_t \tag{14}$$

□

• **b.** *1 point How will you estimate β and γ if it is known that $\beta+\gamma=1$?*

Answer. Setting $\gamma = 1-\beta$ gives the regression

$$y_t - z_t = \alpha + \beta(x_t - z_t) + \varepsilon_t \tag{15}$$

□

Problem 12. *2 points Use the simple matrix differentiation rules $\partial(\boldsymbol{w}^\top\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top = \boldsymbol{w}^\top$ and $\partial(\boldsymbol{\beta}^\top\boldsymbol{M}\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top = 2\boldsymbol{\beta}^\top\boldsymbol{M}$ to compute $\partial L/\partial\boldsymbol{\beta}^\top$ where L is the Lagrange function for the constrained OLS problem*

$$L(\boldsymbol{\beta}) = (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{R}\boldsymbol{\beta}-\boldsymbol{u})^\top\boldsymbol{\lambda} \tag{16}$$

Answer. Write the objective function as $\boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\lambda}^\top\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{\lambda}^\top\boldsymbol{u}$ to get $\partial L/\partial\boldsymbol{\beta}^\top = -2\boldsymbol{y}^\top\boldsymbol{X} + 2\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{\lambda}^\top\boldsymbol{R}$. □

Problem 13. *Assume $\hat{\hat{\boldsymbol{\beta}}}$ is the constrained least squares estimator subject to the constraint $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{o}$, and $\hat{\boldsymbol{\beta}}$ is the unconstrained least squares estimator.*

• **a.** *1 point With the usual notation $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\hat{\hat{\boldsymbol{y}}} = \boldsymbol{X}\hat{\hat{\boldsymbol{\beta}}}$, show that*

$$\boldsymbol{y} = \hat{\hat{\boldsymbol{y}}} + (\hat{\boldsymbol{y}} - \hat{\hat{\boldsymbol{y}}}) + \hat{\boldsymbol{\varepsilon}} \tag{17}$$

Point out these vectors in the `reggeom` simulation.

Answer. In the `reggeom`-simulation, $\boldsymbol{y}$ is the purple line; $\boldsymbol{X}\hat{\hat{\boldsymbol{\beta}}}$ is the red line starting at the origin, one could also call it $\hat{\hat{\boldsymbol{y}}}$; $\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\hat{\boldsymbol{\beta}}}) = \hat{\boldsymbol{y}} - \hat{\hat{\boldsymbol{y}}}$ is the light blue line, and $\hat{\boldsymbol{\varepsilon}}$ is the green line which does not start at the origin. In other words: if one projects $\boldsymbol{y}$ on a plane, and also on a line in that plane, and then connects the footpoints of these two projections, one obtains a zig-zag line with two right angles. □

• **b.** *4 points Show that in* (17) *the three vectors* $\hat{\hat{\boldsymbol{y}}}$, $\hat{\boldsymbol{y}} - \hat{\hat{\boldsymbol{y}}}$, *and* $\hat{\boldsymbol{\varepsilon}}$ *are orthogonal. You are allowed to use, without proof, the following formula for* $\hat{\hat{\boldsymbol{\beta}}}$*:*

$$\hat{\hat{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top\big)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{u}). \tag{18}$$

Answer. One has to verify that the scalar products of the three vectors on the right hand side of (17) are zero. $\hat{\hat{\boldsymbol{y}}}^\top\hat{\boldsymbol{\varepsilon}} = \hat{\hat{\boldsymbol{\beta}}}^\top\boldsymbol{X}^\top\hat{\boldsymbol{\varepsilon}} = 0$ and $(\hat{\boldsymbol{y}} - \hat{\hat{\boldsymbol{y}}})^\top\hat{\boldsymbol{\varepsilon}} = (\hat{\boldsymbol{\beta}} - \hat{\hat{\boldsymbol{\beta}}})^\top\boldsymbol{X}^\top\hat{\boldsymbol{\varepsilon}} = 0$ follow from $\boldsymbol{X}^\top\hat{\boldsymbol{\varepsilon}} = \boldsymbol{o}$; geometrically on can simply say that $\hat{\boldsymbol{y}}$ and $\hat{\hat{\boldsymbol{y}}}$ are in the space spanned by the columns of $\boldsymbol{X}$, and $\hat{\boldsymbol{\varepsilon}}$ is orthogonal to that space. Finally, using (18) for $\hat{\boldsymbol{\beta}} - \hat{\hat{\boldsymbol{\beta}}}$,

$$\begin{aligned}
\hat{\hat{\boldsymbol{y}}}^\top(\hat{\boldsymbol{y}} - \hat{\hat{\boldsymbol{y}}}) &= \hat{\hat{\boldsymbol{\beta}}}^\top\boldsymbol{X}^\top\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\hat{\boldsymbol{\beta}}}) = \\
&= \hat{\hat{\boldsymbol{\beta}}}^\top\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top\big(\boldsymbol{R}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top\big)^{-1}\boldsymbol{R}\hat{\boldsymbol{\beta}} = \\
&= \hat{\hat{\boldsymbol{\beta}}}^\top\boldsymbol{R}^\top\big(\boldsymbol{R}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top\big)^{-1}\boldsymbol{R}\hat{\boldsymbol{\beta}} = 0
\end{aligned}$$

because $\hat{\hat{\boldsymbol{\beta}}}$ satisfies the constraint $\boldsymbol{R}\hat{\hat{\boldsymbol{\beta}}} = \boldsymbol{o}$, hence $\hat{\hat{\boldsymbol{\beta}}}^\top\boldsymbol{R}^\top = \boldsymbol{o}^\top$. □

Problem 14. *In the intermediate econometrics textbook [WW79], the following regression line is estimated:*

$$b_t = 0.13 + .068 y_t + 0.23 w_t + \hat{\varepsilon}_t, \tag{19}$$

where b_t is the public purchase of Canadian government bonds (in billion \$), y_t is the national income, and w_t is a dummy variable with the value $w_t = 1$ for the war years 1940–45, and zero otherwise.

• **a.** *1 point This equation represents two regression lines, one for peace and one for war, both of which have the same slope, but which have different intercepts. What is the intercept of the peace time regression, and what is that of the war time regression line?*

Answer. In peace, $w_t = 0$, therefore the regression reads $b_t = 0.13 + .068 y_t + \hat{\varepsilon}_t$, therefore the intercept is .13. In war, $w_t = 1$, therefore $b_t = 0.13 + .068 y_t + 0.23 + \hat{\varepsilon}_t$, therefore the intercept is $.13 + .23 = .36$. □

• **b.** *1 point What would the estimated equation have been if, instead of w_t, they had used a variable p_t with the values $p_t = 0$ during the war years, and $p_t = 1$ otherwise? (Hint: the coefficient for p_t will be negative, because the intercept in peace times is below the intercept in war times).*

Answer. Now the intercept of the whole equation is the intercept of the war regression line, which is .36, and the coefficient of p_t is the difference between peace and war intercepts, which is -.23.

$$b_t = .36 + .068y_t - .23p_t + \hat{\varepsilon}_t. \tag{20}$$

□

• **c.** *1 point What would the estimated equation have been if they had thrown in both w_t and p_t, but left out the intercept term?*

Answer. Now the coefficient of w_t is the intercept in the war years, which is .36, and the coefficient of p_t is the intercept in the peace years, which is .13.

$$b_t = .36w_t + .13p_t + .068y_t + \hat{\varepsilon}_t? \tag{21}$$

□

• **d.** *2 points What would the estimated equation have been, if bond sales and income had been measured in millions of dollars instead of billions of dollars? (1 billion = 1000 million.)*

Answer. From $b_t = 0.13 + .068y_t + 0.23w_t + \hat{\varepsilon}_t$ follows $1000b_t = 130 + .068 \cdot 1000y_t + 230w_t + 1000\hat{\varepsilon}_t$, or

$$b_t^{(m)} = 130 + .068y_t^{(m)} + 230w_t + \hat{\varepsilon}_t^{(m)}, \tag{22}$$

where $b_t^{(m)}$ is bond sales in millions (i.e., $b_t^{(m)} = 1000b_t$), and $y_t^{(m)}$ is national income in millions (i.e., $y_t^{(m)} = 1000y_t$). □

-

Problem 15. *Regression models incorporate seasonality often by the assumption that the intercept of the regression is different in every season, while the slopes remain the same. Assuming* $\boldsymbol{X}$ *contains quarterly data (but the constant term is not incorporated in* $\boldsymbol{X}$*), this can be achieved in several different ways: You may write your model as*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{23}$$

Alternatively, you may write your model in the form

$$\boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \qquad \boldsymbol{K} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{24}$$

In R *this is the default method to generate dummy variables from a seasonal factor variable. (*`Splus` *has a different default.) This is also the procedure shown in* [Gre97,

p. 383]. *But the following third alternative is often preferrable:*

$$(25)\qquad \boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \qquad \boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

In R *one gets these dummy variables from a seasonal factor variable if one specifies* `contrast="contr.sum"`.

3 points What is the meaning of the seasonal dummies δ_1, δ_2, δ_3, and of the constant term α or the fourth seasonal dummy δ_4, in models (23), (24), and (25)?

Answer. Clearly, in model (23), δ_i is the intercept in the ith season. For (24) and (25), it is best to write the regression equation for each season separately, filling in the values the dummies take for these seasons, in order to see the meaning of these dummies. Assuming $\boldsymbol{X}$ consists of one column only, (24) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \alpha + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ \vdots \end{bmatrix} \beta + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \vdots \end{bmatrix}$$

or, written element by element

$$\begin{aligned} y_1 &= 1 \cdot \alpha + x_1 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_1 && \text{winter} \\ y_2 &= 1 \cdot \alpha + x_2 \cdot \beta + 1 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_2 && \text{spring} \\ y_3 &= 1 \cdot \alpha + x_3 \cdot \beta + 0 \cdot \delta_1 + 1 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_3 && \text{summer} \\ y_4 &= 1 \cdot \alpha + x_4 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 1 \cdot \delta_3 + \varepsilon_4 && \text{autumn} \end{aligned}$$

therefore the overall intercept α is the intercept of the first quarter (winter); δ_1 is the difference between the spring intercept and the winter intercept, etc.

(25) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \alpha + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ \vdots \end{bmatrix} \beta + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \vdots \end{bmatrix}$$

or, written element by element

$$\begin{aligned} y_1 &= 1 \cdot \alpha + x_1 \cdot \beta + 1 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_1 && \text{winter} \\ y_2 &= 1 \cdot \alpha + x_2 \cdot \beta + 0 \cdot \delta_1 + 1 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_2 && \text{spring} \\ y_3 &= 1 \cdot \alpha + x_3 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 1 \cdot \delta_3 + \varepsilon_3 && \text{summer} \\ y_4 &= 1 \cdot \alpha + x_4 \cdot \beta - 1 \cdot \delta_1 - 1 \cdot \delta_2 - 1 \cdot \delta_3 + \varepsilon_4 && \text{autumn} \end{aligned}$$

Here the winter intercept is $\alpha + \delta_1$, the spring intercept $\alpha + \delta_2$, summer $\alpha + \delta_3$, and autmumn $\alpha - \delta_1 - \delta_2 - \delta_3$. Summing this and dividing by 4 shows that the constant term α is the arithmetic mean of all intercepts, therefore δ_1 is the difference between the winter intercept and the arithmetic mean of all intercepts, etc. □

Maximum number of points: 52.

References

[Gre97] William H. Greene, *Econometric analysis*, third ed., Prentice Hall, Upper Saddle River, NJ, 1997. 16

[WW79] Ronald J. Wonnacott and Thomas H. Wonnacott, *Econometrics*, second ed., Wiley, New York, 1979. 12

Economics Department, University of Utah, 1645 Campus Center Drive, Salt Lake City, UT 84112-9300, U.S.A

E-mail address: `ehrbar@econ.utah.edu`

URL: `http://www.econ.utah.edu/ehrbar`