KEY

Stat 511 Exam 1

February 23, 2004 Prof. Vardeman

1. Consider an instance of the linear model for $n = 5$ observations,

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$$

6pts **a)** This is a full rank model. One way to easily see this is to argue that rows 1, 2, and 3 of the model matrix are linearly independent (and so the rank is at least 3). Put these three rows into a 3×3 matrix $\mathbf{M}$ and show this matrix is non-singular by arguing that $\mathbf{Mc} = \mathbf{0}$ implies that $\mathbf{c} = \mathbf{0}$.

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \underset{\sim}{0} \quad \text{means} \quad \begin{matrix} c_1 = 0 & (1) \\ c_1 - c_2 - c_3 = 0 & (2) \\ c_1 + c_2 - c_3 = 0 & (3) \end{matrix}$$

Plugging (1) into (3), $c_2 = c_3$. So plugging this (and (1)) into (2) we have $-2c_2 = 0$ i.e. $c_2 = 0$! Thus $c_1 = c_2 = c_3 = 0$ i.e. M is nonsingular, so X has rank at least 3 (and no more than 3 since it has only 3 columns).

Notice that the columns of $\mathbf{X}$ are perpendicular, so that $\mathbf{X'X}$ is diagonal.

10pts **b)** In a Gauss-Markov version of this model, which of the parameters β_1, β_2, or β_3 can be estimated with the greatest precision? Explain carefully.

$$X'X = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} \quad \text{so} \quad \mathrm{Var}\, \underset{\sim}{b}_{OLS} = \sigma^2 \begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$$

and $\hat{\beta}_1$ has the smallest variance (of $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, the entries of $\underset{\sim}{b}_{OLS}$).

8pts **c)** Compute a matrix $\mathbf{P_X}$ that projects any element of $\Re^5$ onto $C(\mathbf{X})$ (in a perpendicular fashion).

$$P_X = X(X'X)^{-1}X' = X \begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & -1 & 1 \\ 0 & -1 & -1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{14}{20} & \frac{1}{5} & \frac{1}{5} & -\frac{6}{20} \\ \frac{1}{5} & \frac{1}{5} & \frac{14}{20} & -\frac{6}{20} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & -\frac{6}{20} & \frac{14}{20} & \frac{1}{5} \\ \frac{1}{5} & -\frac{6}{20} & \frac{1}{5} & \frac{1}{5} & \frac{14}{20} \end{pmatrix}$$

8pts **d)** In a Gauss-Markov version of this model, which row of the $\mathbf{X}$ matrix represents a set of conditions under which $\mathrm{E}y$ can be estimated with the best precision? Explain carefully.

$\hat{Y} = P_X Y$ has covariance matrix $\mathrm{Var}\hat{Y} = \sigma^2 P_X$. The smallest diagonal entry of P_X is the 1st. So the 1st row of the X matrix represents a set of conditions under which Ey can be estimated with the greatest precision.

For the next two parts of this question (parts **e)** and **f)**), suppose that $\mathbf{Y}$ is such that $SSE = 3$ and $\mathbf{b}'_{\text{OLS}} = (5,6,2)$. Consider an analysis under the normal version of the Gauss Markov model.

10pts **e)** In the future, two new observations, $y_{\text{new}1}$ and $y_{\text{new}2}$ are going to be observed under the conditions described respectively by the 1[st] and 2[nd] rows of the $\mathbf{X}$ matrix. Give 95% two-sided prediction limits for $y_{\text{new}1} - y_{\text{new}2}$. (Plug correct numbers into correct formulas, but do not take time to do arithmetic.)

$y^* = y_{new1} - y_{new2}$ has $Ey^* = (0\ 1\ 1)\boldsymbol{\beta}$ and $\mathrm{Var}\, y^* = 2\sigma^2$

Note that $n - \mathrm{rank}(X) = 5 - 3 = 2$, so 95% prediction limits are

$$\widehat{\underset{\sim}{c}'\beta}_{OLS} \pm t^* \sqrt{MSE}\sqrt{\gamma + \underset{\sim}{c}'(X'X)^{-1}\underset{\sim}{c}}$$

$$\text{i.e. } (0\ 1\ 1)\begin{pmatrix}5\\6\\2\end{pmatrix} \pm 4.303\sqrt{\frac{3}{2}}\sqrt{2 + (0\ 1\ 1)\begin{pmatrix}\frac{1}{5} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{4}\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix}}$$

10pts **f)** Write the hypothesis $\mathrm{H}_0 : \mathrm{E}\,y_1 = \mathrm{E}\,y_2$ and $\mathrm{E}\,y_1 = \mathrm{E}\,y_3$ in testable form $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ for an appropriate matrix $\mathbf{C}$ (write out such a matrix) and compute an F statistic for testing this (you need not do the arithmetic, but plug correct numbers into a correct formula).

$$Ey_1 - Ey_2 = \beta_1 - (\beta_1 - \beta_2 - \beta_3) = \beta_2 + \beta_3$$

$$Ey_1 - Ey_3 = \beta_1 - (\beta_1 + \beta_2 - \beta_3) = -\beta_2 + \beta_3$$

$$H_0: \begin{pmatrix}0 & 1 & 1\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

$$F = \frac{\left[\left(\begin{pmatrix}0 & 1 & 1\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}5\\6\\2\end{pmatrix}\right)'\left(\begin{pmatrix}0 & 1 & 1\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{5} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{4}\end{pmatrix}\begin{pmatrix}0 & 0\\ 1 & -1\\ 1 & 1\end{pmatrix}\right)^{-1}\left(\begin{pmatrix}0 & 1 & 1\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}5\\6\\2\end{pmatrix}\right)\right]/2}{3/2}$$

$\beta_1 = 7$ $\beta_2+\beta_3=3$

6pts g) Consider an Aitken version of the model on page 1, where $\mathbf{V} = \text{diag}(1,\delta,1,1,\delta)$ for δ small. Generalized least squares estimation of $\boldsymbol{\beta}$ under these circumstances will essentially force $\hat{y}_2 = \beta_1 - \beta_2 - \beta_3 \approx y_2$ and $\hat{y}_5 = \beta_1 + \beta_2 + \beta_3 \approx y_5$. This is $\beta_1 \approx (y_2 + y_5)/2$ and $\beta_2 + \beta_3 \approx (y_5 - y_2)/2$. Take these approximations as given and find estimates of β_2 and β_3 if $\mathbf{Y}' = (15,4,6,8,10)$.

Generalized least squares is minimization of $(Y-\hat{Y})'V^{-1}(Y-\hat{Y})$.
Under these approximations, this is minimization of

$$(15-7)^2 + (6-(7+\beta_2-\beta_3))^2 + (8-(7-\beta_2+\beta_3))^2$$

Call $\beta_2-\beta_3$ by a. This is minimization of

$$(-1-a)^2 + (1+a)^2$$

and we want $a=-1$. This requires $2\beta_2 = (\beta_2+\beta_3)+(\beta_2-\beta_3)$
$= 3+(-1) = 2$

and thus $\beta_2 = 1$ and $\beta_3 = 2$.

2. Attached to this exam is a printout of an R session for a time series analysis (via an ordinary linear model) of 6 years worth of quarterly retail sales data (for the JC Penney Company). For consecutive 3-month periods that we will simply label as $t = 1,2,\ldots,24$ we'll model

$$y_t = \text{sales in period } t$$

as roughly linearly increasing in t, but with different "effects" for the 4 quarters of the year. That is, with

$$q_i(t) = \begin{cases} 1 & \text{if period } t \text{ is from the } i\text{th quarter of the year} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1,2,3,4$ we consider a model

$$y_t = \beta_0 + \beta_1 t + \gamma_1 q_1(t) + \gamma_2 q_2(t) + \gamma_3 q_3(t) + \gamma_4 q_4(t) + \varepsilon_t$$

for $t = 1,2,\ldots,24$ the values $\beta_0, \beta_1, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ unknown constants and the $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{24}$ iid normal $(0,\sigma^2)$. (Period $t=1$ is a first quarter period.) You may use the printout to answer the following questions. Refer very carefully to where you find anything you take from the printout (give page and location on the page).

10pts a) Is the parametric function $\gamma_1 - \gamma_2$ estimable in this model? Argue this very carefully. (Write the $\mathbf{X}$ matrix for the first 5 periods below and use it in your argument.)

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Now $\gamma_1 - \gamma_2 = (0\ 0\ 1\ -1\ 0\ 0)\boldsymbol{\beta}$ $\leftarrow \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix}$

1st row − 2nd row $= (0\ -1\ 1\ -1\ 0\ 0)$
5th row − 1st row $= (0\ 4\ 0\ 0\ 0\ 0)$

So $\frac{1}{4}$(5th row − 1st row) + (1st row − 2nd row)
$= (0\ 0\ 1\ -1\ 0\ 0)$

That is, $(0\ 0\ 1\ -1\ 0\ 0) \in C(X')$ and $\gamma_1 - \gamma_2$ is estimable

Since the model as originally posed is not full rank, a call to R's `lm()` function introduces a restriction in order to produce a full rank version. The restriction used by R in this case is to set to 0 the coefficient for the last column of the model matrix entered in the function call. That is, R fits the model

$$y_t = \beta_0^* + \beta_1^* t + \gamma_1^* q_1(t) + \gamma_2^* q_2(t) + \gamma_3^* q_3(t) + \varepsilon_t$$

10pts **b)** In this model, find 90% two-sided confidence limits for σ. (No need to simplify after plugging in.) Use $\left(\sqrt{\frac{SSE}{\chi^2_U}}, \sqrt{\frac{SSE}{\chi^2_L}}\right)$ i.e. $\left(\sqrt{\frac{6{,}102{,}250}{30.143}}, \sqrt{\frac{6{,}102{,}250}{10.117}}\right)$

i.e. $(449.9, 776.6)$

10pts **c)** Give 95% two-sided confidence limits for $\gamma_1^* - \gamma_2^*$. (No need to simplify after plugging in.)

Use $\widehat{\underset{\sim}{c}'\beta}_{OLS} \pm t\sqrt{MSE}\sqrt{\underset{\sim}{c}'(X'X)^{-1}\underset{\sim}{c}}$

Note that vcov is $MSE(X'X)^{-1}$. So this is

$$\left(-2274.21 - (-2564.58)\right) \pm 2.093\sqrt{(1,-1)\begin{pmatrix} 109{,}638 & 55{,}249 \\ 55{,}249 & 108{,}204 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix}}$$

10pts **d)** Give 95% prediction limits for y_{28} (the retail sales in the 4th quarter of the year after the end of the data in hand) based on this model. (Plug correct numbers into a correct formula, but you need not do arithmetic.)

$y^* = y_{28}$ has mean $\beta_0^* + 28\beta_1^*$ and variance σ^2. So

use $\widehat{\underset{\sim}{c}'\beta}_{OLS} \pm t\sqrt{MSE}\sqrt{\gamma + \underset{\sim}{c}'(X'X)^{-1}\underset{\sim}{c}}$

$\widehat{\underset{\sim}{c}'\beta}_{OLS} \pm t\sqrt{\gamma(MSE) + \underset{\sim}{c}'(MSE(X'X)^{-1})\underset{\sim}{c}}$

i.e.

$$\left(1850.76 + 28(99.54)\right) \pm 2.093\sqrt{(566.7)^2 + (1\ \ 28)\begin{pmatrix} 109{,}733 & -4015 \\ -4015 & 286.8 \end{pmatrix}\begin{pmatrix} 1 \\ 28 \end{pmatrix}}$$