COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 18, 2006, 9:00 A.M.–1:00 P.M.

STATISTICS 174 QUESTIONS

Answer all parts. Closed book, calculators allowed. It is important to show all working, especially with numerical calculations. Statistical tables are provided. You may freely quote results from the course notes or text without proof, but to the extent that it is feasible to do so, state precisely the result you are quoting.

Consider the following bivariate linear regression model (BLRM):

$$y_{i,1} = \sum_{j=1}^{p} x_{i,j}\beta_{j,1} + \epsilon_{i,1}, \tag{1}$$

$$y_{i,2} = \sum_{j=1}^{p} x_{i,j}\beta_{j,2} + \epsilon_{i,2}, \tag{2}$$

where $i = 1, 2, \ldots, n$, $\epsilon_i = (\epsilon_{i,1}, \epsilon_{i,2})$ has i.i.d. bivariate normal distribution with mean zero, $\text{Var}(\epsilon_{i,k}) = \sigma_k^2$, and $\text{Cov}(\epsilon_{i,1}, \epsilon_{i,2}) = \sigma_{12}$.

1. (10 pts) Let $\boldsymbol{\beta}_k = (\beta_{1,k}, \ldots, \beta_{p,k})'$, $k = 1, 2$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$, $\mathbf{x}_i = (x_{i,1}, \ldots, \mathbf{x}_{i,p})'$, and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$. Write down the BLRM in matrix form.

2. (25 pts) Suppose $\sigma_1^2, \sigma_2^2, \sigma_{12}$ are known, derive the generalized least squares (GLS) estimator $\widehat{\boldsymbol{\beta}}_{GLS}$. An economist wants to fit the above BLRM to his data, but he only knows how to run ordinary LS (OLS) regression. So he estimated β_1 and β_2 seperately by applying OLS procedures to (1) and (2) respectively. Will his estimate in general be the same as $\widehat{\boldsymbol{\beta}}_{GLS}$? If your answer is "yes", prove your claim. If your answer is "No", give a counter example.

3. (15 pts) Assuming that $\mathbf{X}$ is full rank, give unbiased estimators of σ_1^2 and σ_{12}, prove that they are indeed unbiased, and give the distribution of $\widehat{\sigma_1^2}$.

4. (5 pts) Derive the variance-covariance matrix for $\widehat{\boldsymbol{\beta}}_{GLS}$.

The price and consumption per capita of beef and pork annually from 1925 to 1941 together with food consumption per capita index are given below.

1. PBE = Price of beef (cents/lb)
2. CBE = Consumption of beef per capita (lbs)
3. PPO = Price of pork (cents/lb)
4. CPO = Consumption of pork per capita (lbs)
5. CFO = Food consumption per capita index (1947-1949 = 100)

YEAR	PBE	CBE	PPO	CPO	CFO
1925	59.7	58.6	60.5	65.8	90.9
1926	59.7	59.4	63.3	63.3	92.1
1927	63.0	53.7	59.9	66.8	90.9
1928	71.0	48.1	56.3	69.9	90.9
1929	71.0	49.0	55.0	68.7	91.1
1930	74.2	48.2	59.6	66.1	90.7
1931	72.1	47.9	57.0	67.4	90.0
1932	79.0	46.0	49.5	69.7	87.8
1933	73.1	50.8	47.3	68.7	88.0
1934	70.2	55.2	56.6	62.2	89.1
1935	82.2	52.2	73.9	47.7	87.3
1936	68.4	57.3	64.4	54.4	90.5
1937	73.0	54.4	62.2	55.0	90.4
1938	70.2	53.6	59.9	57.4	90.6
1939	67.8	53.9	51.0	63.9	93.8
1940	63.4	54.2	41.5	72.4	95.5
1941	56.0	60.0	43.9	67.4	97.5

The Pork Producers Association (PPA) would like to understand the relationship between pork/beef price and their per capita consumption. A BLRM is fitted to the data, and following are the R outputs:

```
> bp.lm <- lm(cbind(CBE,CPO)~YEAR+PBE+PPO+CFO)
> summary(bp.lm)

Response CBE :

Call:
lm(formula = CBE ~ YEAR + PBE + PPO + CFO)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.5413 -0.3821  0.3516  0.6650  1.6315

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -886.50268  173.34403  -5.114 0.000256 ***
YEAR           0.55177    0.09957   5.541 0.000127 ***
PBE           -0.84239    0.10464  -8.050 3.53e-06 ***
PPO            0.24590    0.05475   4.491 0.000738 ***
CFO           -0.90892    0.31591  -2.877 0.013905 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.457 on 12 degrees of freedom
Multiple R-Squared: 0.9131,Adjusted R-squared: 0.8841
F-statistic: 31.51 on 4 and 12 DF,  p-value: 2.795e-06

Response CPO :

Call:
lm(formula = CPO ~ YEAR + PBE + PPO + CFO)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0415 -0.8768 -0.5417  0.6569  1.9128

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1940.60799  129.45688  14.990 3.92e-09 ***
YEAR          -0.99782    0.07436 -13.419 1.38e-08 ***
PBE            0.25411    0.07815   3.252  0.00694 **
PPO           -0.79008    0.04089 -19.323 2.08e-10 ***
CFO            0.87094    0.23593   3.691  0.00308 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.088 on 12 degrees of freedom
Multiple R-Squared: 0.9799,Adjusted R-squared: 0.9732
F-statistic: 146.2 on 4 and 12 DF,  p-value: 4.548e-10
```

In addtion, we know that $\sum_i e_{i,1}e_{i,2} = -14.869$, where $e_{i,k}$ are the residuals from the BLRM, and $(\mathbf{X}'\mathbf{X})^{-1}$ is given by

	(Intercept)	YEAR	PBE	PPO	CFO
(Intercept)	14151.76	-8.015584	4.061781	-0.869267	12.20913
YEAR	-8.02	0.004669	-0.002907	0.000283	-0.00907
PBE	4.06	-0.002907	0.005157	0.000192	0.01309
PPO	-0.87	0.000283	0.000192	0.001412	0.00252
CFO	12.21	-0.009071	0.013089	0.002522	0.04700

5. (10 pts) Write a report of no more than 200 words explaining the fitted model to someone from PPA who don't know much about statistics.
6. (15 pts) If for the year 1942 the CFO remains the same as 1941, and the PBE is expected to decrease by 6 cents/lb. In order to maintain the pork per capita consumption at the same level as in 1941, how much should the PPO decrease? Give a point estimator and its 95% confidence interval.
7. (10 pts) Suppose for the year 1942, CFO=95, PBE=50, and PPO=40. Estimate the difference between per capita consumption of beef and pork for 1942, and compute its standard error.
8. (10 pts) Outline in no more than 200 words any further analysis you can to do before writing a final report for PPA on relationship between PBE, CBE, PPO, and CPO.

Solutions, 2006 CWE 174

1. Let $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2')'$, $\boldsymbol{\epsilon} = ((\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2')'$, and $\mathbf{Z} = \text{diag}(\mathbf{X}, \mathbf{X})$, the BLRM can be written as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

2. $\widehat{\boldsymbol{\beta}}_{GLS} = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}$, where

$$\mathbf{V} = \text{Var}(\mathbf{Y}) = \begin{pmatrix} \sigma_1^2\mathbf{I} & \sigma_{12}\mathbf{I} \\ \sigma_{12}\mathbf{I} & \sigma_2^2\mathbf{I} \end{pmatrix}.$$

   Matrix computation leads to $\widehat{\boldsymbol{\beta}}_{GLS} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1 \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_2 \end{pmatrix}$, i.e., the GLS and OLS esitmater is the same, and for the rest we drop the subscript GLS.

3. Let $\mathbf{e} = \mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}} = (\mathbf{e_1}', \mathbf{e_2}')'$, $\widehat{\sigma_1^2} = \mathbf{e_1}'\mathbf{e_1}/(n-p) \sim \chi^2_{n-p}$, and $\widehat{\sigma}_{12} = \mathbf{e_1}'\mathbf{e_2}/(\mathbf{n}-\mathbf{p})$.

4. $\text{Var}(\widehat{\boldsymbol{\beta}}) = \begin{pmatrix} \sigma_1^2(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{12}(\mathbf{X}'\mathbf{X})^{-1} \\ \sigma_{12}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_2^2(\mathbf{X}'\mathbf{X})^{-1} \end{pmatrix}$.

5. Interpretation of the coefficients.

6. Let x_0 be the price decrease for PPO, then

$$\beta_{22} - 6\beta_{32} - x_0\beta_{42} = 0.$$

   Point estimator: $\widehat{x_0} = (-0.99782 - 6 * 0.25411)/(-0.79008) = 3.19$.

   CI can be obtained using Feiller's method: solve for x_0 the following inequality

$$(\beta_{22} - 6\beta_{32} - x_0\beta_{42})^2/\text{Var}(\beta_{22} - 6\beta_{32} - x_0\beta_{42}) \le t^2_{12,0.975},$$

   and we have $x_0 \in (2.205, 4.225)$.

7. Let $\mathbf{x}_h = (1, 1942, 50, 40, 95)'$, the difference can be written as $(\mathbf{x}_h', -\mathbf{x}_h')\boldsymbol{\beta}$.

   Point estimator: $(\mathbf{x}_h', -\mathbf{x}_h')\widehat{\boldsymbol{\beta}} = -0.279$.

   Variance: $(\mathbf{x}_h', -\mathbf{x}_h')\text{Var}(\widehat{\boldsymbol{\beta}})(\mathbf{x}_h', -\mathbf{x}_h')' = (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})\mathbf{x}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h = 1.247$

8. Diagnostics, check for temporal correlation, colinearity, etc.