THE UNIVERSITY OF CHICAGO
Graduate School of Business
Business 41912, Spring Quarter 2004, Mr. Ruey S. Tsay

Midterm

Notes:

1. Open book and notes. The exam time is 90 minutes.
2. Write your answers in a bluebook. Mark the solution clearly.
3. You are required to pledge to uphold the honor code of GSB in the exam.

1. (21 points) Consider the multiple linear regression model

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

where the dimension of $\boldsymbol{Y}$ is $n \times 1$ and those of $\boldsymbol{Z}$, $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ are $n \times (r+1)$, $(r+1) \times 1$ and $n \times 1$, respectively. Assume that $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$, and $n > r+1$. Define the residuals as $\hat{\boldsymbol{\epsilon}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$, where $\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ is the Hat matrix.

   (a) Assume further that $\boldsymbol{\epsilon}$ is distributed as $N_n(\mathbf{0}, \boldsymbol{\Sigma})$. What is the distribution of the i-th residual, i.e. the i-th element of $\hat{\boldsymbol{\epsilon}}$? Briefly justify your answer.

   (b) Consider the i-th and j-th residuals $\hat{\epsilon}_i$ and $\hat{\epsilon}_j$. Are they independent? Why?

   (c) Show that $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{\epsilon}} = 0$.

2. (28 points) Let $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)'$ be a 4-dimensional Gaussian random variable with mean $\boldsymbol{\mu} = (1, 2, 3, 4)'$ and covariance matrix $\boldsymbol{I}$, the identity matrix. Define $\boldsymbol{X} = (X_1, X_2)' = (Z_1 + Z_2, Z_1 + Z_4)'$. Answer the following questions briefly.

   (a) What is the distribution of $\boldsymbol{X}$?

   (b) What is the conditional distribution of $\boldsymbol{X}$ given that $Z_3 = 3$?

   (c) What is the conditional distribution of X_1 given $Z_3 = 4$?

   (d) What is the distribution of $Y = X_1 + X_2 - Z_3$?

3. (21 points) Consider a bivariate time series $\boldsymbol{X}_t = \boldsymbol{c} + \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1}$, where $\boldsymbol{c}$ is a 2-dimensional constant vector, and $\{\boldsymbol{a}_t\}$ are independent and identically distributed Gaussian random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma} > 0$. Suppose that we have a realization of the series, say $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n\}$. Let $\bar{\boldsymbol{X}}_n$ be the sample mean, and define the lag-k autocovariance matrix of $\boldsymbol{X}_t$ as $\boldsymbol{\Gamma}_k = \text{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-k})$, where k is an integer. Answer the following questions.

(a) What is $E(\bar{\boldsymbol{X}}_n)$?

(b) Derive $\text{Cov}(\sqrt{n}\bar{\boldsymbol{X}}_n)$.

(c) What is the asymptotic covariance matrix of $\sqrt{n}\bar{\boldsymbol{X}}_n$ as $n \to \infty$?

4. (30 points) Consider the biting flies data in Table 6.15 and Exercise 6.28, p. 347, of the textbook. There are 35 flies from each specy. Computer outputs from Splus and Matlab, including the needed critical value of F-distribution, are attached for your information. Answer the following questions:

   (a) Write down the sample mean vectors for the two species.

   (b) Do the data indicate any difference in the mean vectors of the two species, using 5% significance test?

   (c) If the null hypothesis of equal mean vectors is rejected, determine the linear combinations of the mean vectors most responsible for rejecting H_o.

   (d) Construct the simultaneous 95% confidence intervals for the components of the difference of mean vectors.

   (e) What are the assumptions, if any, used in the above analyses?

```
***
*** Computer output for problem 4 of Mid-term Exam, 2004.

*** Splus  ****

*** Multivariate Analysis of Variance Model ***

Short Output:
Call:
   manova(formula = cbind(x1, x2, x3, x4, x5, x6, x7) ~ fac,
data = T6.15, na.action = na.exclude)

Terms:
                fac Residuals
Deg. of Freedom   1        68

5 out of 7 effects not estimable
Estimated effects may be unbalanced

Analysis of Variance Table:
          Df Hotelling-Lawley approx. F num df
      fac 1   1.56081         13.82428         7
Residuals 68

          den df     P-value
      fac       62         0
Residuals

**** Matlab ****

x: data for the L. torrens
y: data for L. carteri

>> [xbar1,cov1]=mdesc(x);

xbar1 =

   96.4571   42.9143   35.3714   14.5143   25.6286    9.5714    9.7143

cov1 =

   40.7261   11.7168    2.3252    2.1992    6.2630    1.8782    1.6639
```

```
   11.7168    7.4924    1.8269    1.8395    3.2613    0.3151    0.2983
    2.3252    1.8269    4.8286   -0.7849    0.6126   -0.3655    0.1681
    2.1992    1.8395   -0.7849    3.3748    1.6966    0.1681    0.0042
    6.2630    3.2613    0.6126    1.6966    6.2403   -0.0168   -0.0504
    1.8782    0.3151   -0.3655    0.1681   -0.0168    0.8403    0.6387
    1.6639    0.2983    0.1681    0.0042   -0.0504    0.6387    0.7983

>> [xbar2,cov2]=mdesc(y);

xbar2 =

   99.3429   43.7429   39.3143   14.6571   30.0000    9.6571    9.3714

cov2 =

   31.2908   17.4731    9.8303    5.1504   12.8824    2.9739    3.6336
   17.4731   25.7849    3.7008    4.1445    8.9412    1.7916    1.5689
    9.8303    3.7008    8.0454    2.1697    2.6176    0.7874    1.1739
    5.1504    4.1445    2.1697    2.7025    3.1176    0.3790    0.4546
   12.8824    8.9412    2.6176    3.1176   21.2941    1.1471    1.3235
    2.9739    1.7916    0.7874    0.3790    1.1471    1.5849    1.1899
    3.6336    1.5689    1.1739    0.4546    1.3235    1.1899    1.1815

>> Spool=0.5*cov1+0.5*cov2

Spool =

   36.0084   14.5950    6.0777    3.6748    9.5727    2.4261    2.6487
   14.5950   16.6387    2.7639    2.9920    6.1013    1.0534    0.9336
    6.0777    2.7639    6.4370    0.6924    1.6151    0.2109    0.6710
    3.6748    2.9920    0.6924    3.0387    2.4071    0.2735    0.2294
    9.5727    6.1013    1.6151    2.4071   13.7672    0.5651    0.6366
    2.4261    1.0534    0.2109    0.2735    0.5651    1.2126    0.9143
    2.6487    0.9336    0.6710    0.2294    0.6366    0.9143    0.9899

>> dif=xbar1-xbar2

dif =

   -2.8857   -0.8286   -3.9429   -0.1429   -4.3714   -0.0857    0.3429

>> a=inv(Spool)*dif'
```

```
a =

    0.0062
    0.1506
   -0.8538
    0.2676
   -0.3828
   -2.1873
    2.9707

>>
>> dmx=dif*a

dmx =

    6.0648

>> finv(0.95,7,62)

ans =

    2.1613
```