Econometrics Honours
Economics 467

Wednesday December 13, 2006, 14.00–17.00

Examiner: R. Davidson Associate Examiner: V. Zinde-Walsh

Instructions:

- Calculators are allowed
- NO notes or texts allowed
- Answer in exam book(s)
- Language Dictionaries are allowed
- You may keep the exam
- This exam comprises 7 pages, including the cover page

As for the midterm exam, do not be upset if this exam seems too long for you. Do not waste time on questions for which you do not see how to obtain the answer. Rather, answer as much as you can. Everything you do will be taken into account. However, it should be worth your while to take the time to read over the entire exam before plunging in.

1. Let $\{z_t\}_{t=1}^{\infty}$ be a sequence of independent random variables. Denote the expectation of z_t by m_t, and the variance by v_t. Clearly describe the consequences of (i) the law of large numbers, and (ii) the central limit theorem, with respect to the sequence $\{z_t\}$.

Consider the results given below for the following three linear regressions, run on a sample of size 50,

$$\boldsymbol{y} = \alpha + \beta_1 \boldsymbol{x}_1 + \boldsymbol{u};$$
$$\boldsymbol{y} = \alpha + \beta_2 \boldsymbol{x}_2 + \boldsymbol{u}; \text{ and}$$
$$\boldsymbol{y} = \alpha + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \boldsymbol{u}.$$

Variable	Parameter estimate	Standard error	T statistic
constant	3.975865	0.540199	7.360000
x1	7.186191	0.457228	15.716866

Sum of squared residuals = 683.799704

Variable	Parameter estimate	Standard error	T statistic
constant	3.975822	0.540195	7.359970
x2	7.186230	0.457226	15.717006

Sum of squared residuals = 683.789493

Variable	Parameter estimate	Standard error	T statistic
constant	3.968922	0.546318	7.264861
x1	-1200.157677	4961.991625	-0.241870
x2	1207.348754	4962.011683	0.243318

Sum of squared residuals = 682.939435

Estimated covariance matrix:

0.298463	141.546902	-141.586210
141.546902	24621360.882225	-24621460.306808
-141.586210	-24621460.306808	24621559.945032

What can one say about the explanatory power of the exogenous variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ for the dependent variable $\boldsymbol{y}$? In the case of the model with both explanatory variables, is the hypothesis $\beta_1 = 0$ compatible with the data? What about the hypothesis $\beta_2 = 0$? Or the joint hypothesis $\beta_1 = 0$, $\beta_2 = 0$? What is the phenomenon behind your results?

2. If F is a strictly increasing CDF defined on an interval $[a, b]$ of the real line, where either or both of a and b may be infinite, then the inverse function F^{-1} is a well-defined mapping from $[0, 1]$ on to $[a, b]$. Show that, if the random variable X is a drawing from the U(0, 1) distribution, then $F^{-1}(X)$ is a drawing from the distribution of which F is the CDF.

You estimate a parameter θ by least squares. The parameter estimate is $\hat{\theta} = 2.5762$, and its standard error is 0.4654. You then generate 999 bootstrap samples and from each of them calculate t_j^*, the t statistic for the hypothesis that $\theta = \hat{\theta}$. When the t_j^* are sorted from smallest to largest, number 25 is equal to -2.2214, and number 975 is equal to 1.7628. Find a .95 studentized bootstrap confidence interval for θ, and compare it with the usual asymptotic confidence interval.

3. A square matrix $\boldsymbol{U}$ is said to be **orthogonal** if its transpose is equal to its inverse: $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^\top = \mathbf{I}$. Show that the product of two orthogonal matrices of the same dimensions is also an orthogonal matrix.

Let $\boldsymbol{P_X}$ be the orthogonal projection matrix of which the image is the linear span of the k columns of the $n \times k$ matrix $\boldsymbol{X}$, where $n > k$. What are the defining properties of any orthogonal projection matrix? Give an explicit expression for the matrix $\boldsymbol{P_X}$ in terms of $\boldsymbol{X}$.

Now let $\boldsymbol{M_X} \equiv \mathbf{I} - \boldsymbol{P_X}$ be the complementary projection matrix. Show that the product $\boldsymbol{P_X}(\mathbf{I} - \boldsymbol{P_X})$ is a zero matrix *without* making use of the fact that $\boldsymbol{P_X}$ is an orthogonal projection matrix rather than any projection matrix. If $\boldsymbol{P_X}$ is in fact an orthogonal projection matrix, prove that $\boldsymbol{M_X}$ is so also, making use in your proof of the defining properties.

Show that the matrix $\boldsymbol{M_X} - \boldsymbol{P_X}$ is an orthogonal matrix. What is its inverse?

4. The following model:

$$\boldsymbol{y} = \alpha + \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \boldsymbol{u}$$

was estimated by ordinary least squares. Here are the results:

```
ols y c x1 x2
Ordinary Least Squares:
Variable  Parameter estimate  Standard error  T statistic
constant     -5.645143           1.327595      -4.252157
x1            0.038164           0.240304       0.158817
x2           14.340118           1.994437       7.190059

Number of observations = 100 Number of regressors = 3
Sum of squared residuals = 1844.976846
Explained sum of squares = 1912.406935

Estimate of residual variance
 (with d.f. correction) = 19.020380
R squared (uncentred) = 0.508973
Estimated covariance matrix:
  1.762509   -0.117984   -2.044940
```

```
   -0.117984     0.057746    -0.100414
   -2.044940    -0.100414     3.977778
```

In order to test the hypothesis that $\beta_2 = 1/\beta_1$, an econometrician formulated a Gauss-Newton regression, in which she used the estimates $\hat{\alpha}$ and $\hat{\beta}_1$ of the above table. If the GNR is written in the form:

$$\boldsymbol{r}(\hat{\alpha}, \hat{\beta}_1) = \alpha + \boldsymbol{R}_\beta(\hat{\alpha}, \hat{\beta}_1)b_\beta + \text{residuals},$$

give explicit formulas for the artificial variables $\boldsymbol{r}(\hat{\alpha}, \hat{\beta}_1)$ and $\boldsymbol{R}_\beta(\hat{\alpha}, \hat{\beta}_1)$.
Here are the results of the GNR:

```
ols r c Rb

Ordinary Least Squares:

Variable  Parameter estimate  Standard error  T statistic

constant     -0.036371            1.221706      -0.029771
Rb            0.017319            0.002833       6.113372

Number of observations = 100 Number of regressors = 2
Sum of squared residuals = 1845.075658
Explained sum of squares = 5629.131123
Estimate of residual variance (with d.f. correction) = 18.827303
R squared (uncentred) = 0.753141

Estimated covariance matrix:

   1.492566   0.003235
   0.003235   0.000008
```

Test the hypothesis that $\beta_2 = 1/\beta_1$. (By "test" I mean: calculate the numerical value of a suitable test statistic, give the (asymptotic) probability distribution of the statistic under the null hypothesis, and, if possible, say whether or not the null hypothesis can be rejected.)

5. Give the formal definition of a positive definite matrix. How is this definition different from that of a positive semi-definite matrix? If $\boldsymbol{A}$ is positive definite, show that $\boldsymbol{A}^{-1}$ is so also. Why does the same result not hold in general for positive semi-definite matrices?

Suppose that we wish to carry out a two-tailed test using a test statistic τ which has a distribution that is not symmetric about the origin. Let the CDF of the statistic τ be denoted as F, where $F(-x) \neq 1 - F(x)$ for general x. Suppose that, for any significance level α, the critical values c_α^- and c_α^+ are defined by the equations

$$F(c_\alpha^-) = \alpha/2 \quad \text{and} \quad F(c_\alpha^+) = 1 - \alpha/2.$$

Show that the marginal significance level (P value) associated with a realized statistic $\hat{\tau}$ is $2\min\big(F(\hat{\tau}), 1 - F(\hat{\tau})\big)$.

6. Let s_1, s_2, s_3, and s_4 be the four seasonal variables conventionally used with quarterly data. Three more variables are defined in terms of the s_i as follows:

$$S_1 \equiv s_1 - s_4; \quad S_2 \equiv s_2 - s_4; \quad S_3 \equiv s_3 - s_4.$$

The consumption of Canadian households, C, was regressed on a constant and disposable household income, Y, with the following results:

```
Ordinary Least Squares:

Variable  Parameter estimate   Standard error   T statistic

c         0.6486616            0.2239607        2.8963195
Y         0.8417831            0.0062445        134.8032551

Number of observations = 164
Number of regressors = 2
Sum of squared residuals = 716.6373212
Explained sum of squares = 153917.8284388
Estimate of residual variance (with d.f. correction) = 4.4236872
R squared (uncentred) = 0.9953656
```

Next, the regression was run again with the variables S_1, S_2, and S_3 as extra regressors. The results:

```
Ordinary Least Squares:

Variable    Parameter estimate  Standard error  T statistic

c           0.6024778           0.1658912       3.6317649
Y           0.8436771           0.0046308       182.1895453
S1          0.0548180           0.2107650       0.2600908
S2          0.5438105           0.2105196       2.5831827
S3         -2.2508859           0.2108066      -10.6774943

Number of observations = 164
Number of regressors = 5
Sum of squared residuals = 385.1420528
Explained sum of squares = 154249.3237072
Estimate of residual variance (with d.f. correction) = 2.4222771
R squared (uncentred) = 0.9975093
```

Is there evidence of a seasonal effect? (Justify your answer!)
Three further variables were generated by the following relations:

$$YS_1 \equiv Y \times S_1; \quad YS_2 \equiv Y \times S_2; \quad YS_3 \equiv Y \times S_3,$$

and incorporated into the model, with the following results:

```
Ordinary Least Squares:

Variable    Parameter estimate  Standard error  T statistic

c           0.5899112           0.1132038       5.2110561
Y           0.8455305           0.0031791       265.9614667
S1          0.0531812           0.1952764       0.2723380
S2          0.2834064           0.1954219       1.4502286
S3         -0.6315483           0.1975918      -3.1962272
```

```
YS1        -0.0012106           0.0058144      -0.2082134
YS2         0.0096389           0.0055428       1.7390034
YS3        -0.0617163           0.0052379     -11.7827459

Number of observations = 164
Number of regressors = 8
Sum of squared residuals = 175.8096432
Explained sum of squares = 154458.6561168
Estimate of residual variance (with d.f. correction) = 1.1269849
R squared (uncentred) = 0.9988631
```

Calculate the marginal propensity to consume for the third season. What is the average marginal propensity to consume averaged over all seasons? Explain your calculations.

7. It is often natural to divide a sample into two subsamples, and ask whether a linear regression model has the same parameters for both the subsamples. An F test can be used for this purpose.

Let the sizes of the two subsamples be n_1 and n_2, with $n = n_1 + n_2$. If we separate the subsamples by partitioning the variables, we can write

$$\boldsymbol{y} \equiv \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{X} \equiv \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix},$$

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are, respectively, an n_1–vector and an n_2–vector, while $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are $n_1 \times k$ and $n_2 \times k$ matrices. Suppose that the parameter vector for the first subsample is $\boldsymbol{\beta}_1$, and for the second, $\boldsymbol{\beta}_2$. Consider the following regression model:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2 \end{bmatrix} \gamma + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}). \tag{1}$$

Show that, in the first subsample, the regression functions are the components of $\boldsymbol{X}_1\boldsymbol{\beta}$, while, in the second, they are the components of $\boldsymbol{X}_2(\boldsymbol{\beta} + \boldsymbol{\gamma})$. How do $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ relate to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$?

Show that the sum of squared residuals from regression (1) is equal to the sum of the SSRs from the two subsample regressions:

$$\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}_1, \quad \boldsymbol{u}_1 \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \text{ and}$$
$$\boldsymbol{y}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2, \quad \boldsymbol{u}_2 \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}).$$

Give the algebraic form of the F statistic that can be used to test the hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Describe carefully how to obtain a bootstrap P value for this F test.

8. Consider the following two regressions:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}, \text{ and}$$
$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}.$$

It is plain that the second regression reduces to the first if $\boldsymbol{\beta}_2 = \mathbf{0}$. Suppose that you have the results of the OLS estimation of both of these regressions. In terms of these results, how would you compute an F statistic that would let you test the hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$? Suppose that the covariance matrix of the random elements, $\boldsymbol{\Omega} \equiv \text{Var}(\boldsymbol{u})$ is a diagonal but not a scalar matrix. By this it is meant that there does not exist any σ^2 such that $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$. Explain why the F statistic computed as you have described above is no longer valid to test the hypothesis. How can you obtain a statistic that does provide an asymptotically valid test of the hypothesis?