Economics 662

October 29, 2024 **R. Davidson**

Midterm Examination

Please send your completed exam to Raphaël Langevin <**raphael.langevin@mcgill.ca**> *by 12.00 (noon) on Thursday October 31. As for the assignments, please submit two files per student: one, which should be a PDF file, with your written answers, and another, which may or may not be a simple text file, with your computer code. These files must be all your own work. You may make use of whatever non-human resources you wish, provided that you acknowledge them with a suitable citation, but you must not ask for or receive any help from any other person. Note that the questions all have different weights in the marking scheme.*

All students in this course have the right to submit in English or in French any written work that is to be graded.

Tou(te)s les étudiant(e)s qui suivent ce cours on le droit de soumettre tout travail écrit en français ou en anglais.

Academic Integrity statement [approved by Senate on 29 January 2003]:

McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures.

L'université McGill attache une haute importance à l'honnêteté académique. Il incombe par conséquent à tou(te)s les étudiant(e)s de comprendre ce que l'on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l'étudiant(e) et des procédures disciplinaires.

Answer all five questions in this exam.

Faites tous les cinq exercices de cet examen.

1. If $\boldsymbol{X}$ and $\boldsymbol{W}$ are both $n \times k$ matrices, $n > k$, with full column rank and such that $\boldsymbol{X}^\top \boldsymbol{W}$ is nonsingular, it is known that the $n \times n$ matrix

$$\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{W}^\top \boldsymbol{X})^{-1}\boldsymbol{W}^\top$$

is in general an oblique projection matrix.

(i) Under what conditions is $\boldsymbol{P}$ and *orthogonal* projection matrix?

(ii) Suppose $\boldsymbol{X}$ is replaced by another $n \times k$ matrix, $\boldsymbol{X}'$, with full column rank, the columns of which are linear combinations of those of $\boldsymbol{X}$. Let $\boldsymbol{P}' = \boldsymbol{X}'(\boldsymbol{W}^\top \boldsymbol{X}')^{-1}\boldsymbol{W}^\top$. What is the relation between $\boldsymbol{P}$ and $\boldsymbol{P}'$?

(iii) Now leave $\boldsymbol{X}$ unchanged, and replace $\boldsymbol{W}$ by $\boldsymbol{W}'$, where $\boldsymbol{W}'$ bears the same relation to $\boldsymbol{W}$ as does $\boldsymbol{X}'$ to $\boldsymbol{X}$. How is $\boldsymbol{P}$ affected by this replacement?

(iv) Finally, let $\boldsymbol{X}$ be replaced by $\boldsymbol{X}'$ and $\boldsymbol{W}$ by $\boldsymbol{W}'$. What becomes of $\boldsymbol{P}$ in this case?

2. Consider the following autoregressive model:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t, \quad t = 1, \ldots, n.$$

Run a series of simulation experiments with u_t centred and normally distributed, $\rho_2 = 0$, and $n = 10, 20, 30, 50, 100$, and $\rho_1 = 0.3, 0.5, 0.9$, for a total of $5 \times 3 = 15$ experiments. For each replication, run the regression of y_t on its first two lags. The aim is to estimate the bias of the estimator $\hat{\rho}_1$. How does the bias seem to vary with n and ρ_1?

3. A variable represented by an n-vector $\boldsymbol{y}$ is said to be **centred** if the sum of its elements is zero:

$$\sum_{t=1}^{n} y_t = 0.$$

A non-centred variable $\boldsymbol{y}$ can be centred by replacing each element y_t by $y_t - \bar{y}$, where $\bar{y}$ is the mean of the elements of $\boldsymbol{y}$.

Consider the following regression, where $\boldsymbol{\iota}$ is the vector every element of which is equal to one:

$$\boldsymbol{y} = \alpha\boldsymbol{\iota} + \beta\boldsymbol{x} + \boldsymbol{u}. \tag{1}$$

Denote by $\boldsymbol{w}$ the result of centring $\boldsymbol{y}$, and by $\boldsymbol{z}$ the result of centring $\boldsymbol{x}$, and then consider these regressions:

$$\begin{array}{cl}
\text{(i)} & \boldsymbol{w} = \beta\boldsymbol{z} + \boldsymbol{u}, \\
\text{(ii)} & \boldsymbol{w} = \alpha\boldsymbol{\iota} + \beta\boldsymbol{z} + \boldsymbol{u}, \\
\text{(iii)} & \boldsymbol{y} = \alpha\boldsymbol{\iota} + \beta\boldsymbol{z} + \boldsymbol{u}, \\
\text{(iv)} & \boldsymbol{y} = \beta\boldsymbol{z} + \boldsymbol{u}, \\
\text{(v)} & \boldsymbol{y} = \beta\boldsymbol{x} + \boldsymbol{u}.
\end{array}$$

For which of these regressions is the estimate of β the same as that given by (1)? Why? In the two cases in which α is estimated, what is the connection between these estimates of α and that given by (1)? Explain.

After estimating the equation

$$\boldsymbol{y} = \alpha + \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \beta_3\boldsymbol{x}_3 + \boldsymbol{u},$$

by OLS, a student observed that the regressors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ were linearly dependent – the package had eliminated $\boldsymbol{x}_3$ on the grounds of collinearity with the other regressors. Another attempt was made using centred versions of the four variables $\boldsymbol{y}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$, and eliminating the constant from the regression. Was there still a problem of collinearity? Explain. If there was, what was the linear dependency? If not, would there be a collinearity problem if the constant was put back into the regression?

4. Consider the following classical normal linear model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\mathbf{I})$$

where the explanatory variables in the matrix $\boldsymbol{X}$ are exogenous, in the sense that they are independent of the disturbances $\boldsymbol{u}$. An estimator of the parameter vector is provided by the solution to the following set of linear equations

$$\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{W}$ is a matrix of exogenous variables, of dimension $n \times k$, where n is the sample size, and k is the dimension of the parameter vector $\boldsymbol{\beta}$. Sketch the reasoning that shows that this estimator is consistent and asymptotically normal, and mention the basic regularity conditions for the reasoning to be valid.

Give an explicit expression for the covariance matrix of the estimator defined by equations (2). Show that this covariance matrix is optimised by choosing $\boldsymbol{W}$ equal to $\boldsymbol{X}$. In what precise sense is the covariance matrix "optimised"?

5. Prove that, for a linear regression model with a constant term, the uncentred R_u^2 is always no smaller than the centred R_c^2.

Consider a linear regression model for a dependent variable y_t that has a sample mean of 17.21. Suppose that we create a new variable $y_t' = y_t + 10$ and run the same linear regression using y_t' instead of y_t as the regressand. How will R_c^2, R_u^2, and the estimate of the constant term be related in the two regressions? What if instead $y_t' = y_t - 10$?

For the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{x}_2\beta_2 + \boldsymbol{u},$$

give the algebraic expression for the t statistic associated with the (scalar) parameter β_2 provided by OLS estimation of the regression. Show that the t statistic is $(n-k)^{1/2}$ times the cotangent of the angle between the n-vectors $\boldsymbol{M}_1\boldsymbol{y}$ and $\boldsymbol{M}_1\boldsymbol{x}_2$. Here n is sample size, k is the number of explanatory variables in the above model, and $\boldsymbol{M}_1$ is the orthogonal projection on to the orthogonal complement of the span of the regressors $\boldsymbol{X}_1$. (Note for the forgetful: $\cot\theta = \cos\theta/\sin\theta$.)

6. The file `e662_midterm24.dat` contains data from 200 observations, of which the first 100 are for Canadian women in the year 2000, and the final 100 for men in the same year. The three variables in the file are, in order, the age of an income recipient, the number of weeks in the year he or she worked, and the income earned.

In a linear regression of income on a constant, age, and weeks worked, it can be expected that the coefficients are different for men and women. Perform a Chow test of the null hypothesis that they are in fact the same for the two sexes. What is the nominal distribution under the null of the statistic you use? What is the nominal P value that corresponds to the statistic? Would you be willing to reject the null, and, if so, for what significance level, if you must choose a level of 1%, 2%, 5%, or 10%?