Economics 662

December 10-12, 2024 **R. Davidson**

Final Examination

Your completed exam should be sent to our TA Raphaël Langevin by 13.00 on December 12 – <**raphael.langevin@mail.mcgill.ca**> *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, but you must not ask for or receive any help from any other person.*

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 seven questions in this exam. Note that the different questions have quite different values in terms of marks.

Faites tous les sept exercices de cet examen. Les différents exercices ont des valeurs assez différentes.

1. A random variable Z follows the **lognormal distribution** if $\log Z$ is normally distributed. Let $\log Z \sim \mathrm{N}(m, \sigma^2)$. What is the cumulative distribution function (CDF) of Z? What are the first four moments of Z? The variance of Z? [A useful fact is that, if $W \sim \mathrm{N}(0,1)$, then $\mathrm{E}(\exp \sigma W) = \exp(\sigma^2/2)$.]

2. Consider the following linear regression:

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

where $\boldsymbol{y}$ is $n \times 1$, $\boldsymbol{X}_1$ is $n \times k_1$, and $\boldsymbol{X}_2$ is $n \times k_2$. Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ be the OLS parameter estimates from running this regression.

Now consider the following regressions, all to be estimated by OLS:

(a) $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(b) $\boldsymbol{P}_1\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(c) $\boldsymbol{P}_1\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(d) $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(e) $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(f) $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(g) $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(h) $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(i) $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$;

(j) $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$.

Here $\boldsymbol{P}_1$ projects orthogonally on to the span of $\boldsymbol{X}_1$, and $\boldsymbol{M}_1 = \mathbf{I} - \boldsymbol{P}_1$. For which of the above regressions are the estimates of $\boldsymbol{\beta}_2$ the same as for the original regression? Why? For which are the residuals the same? Why?

Prove that the Frisch-Waugh-Lovell theorem applies to the heteroskedasticity-consistent covariance estimator (HCCME) for the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \quad E(\boldsymbol{u}\boldsymbol{u}^\top) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2). \tag{1}$$

where $\boldsymbol{y}$ is $n \times 1$, $\boldsymbol{X}_1$ is $n \times k_1$, and $\boldsymbol{X}_2$ is $n \times k_2$. More specifically, what you are asked to prove is that the HCCME for the OLS estimates $\hat{\boldsymbol{\beta}}_2$, calculated from the results of regression (1), is the same as the HCCME calculated from the FWL regression

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{M}_1\boldsymbol{u}$$

3. Some of the results from running the linear regression

$$\boldsymbol{y} = \beta_0\boldsymbol{\iota} + \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \beta_3\boldsymbol{x}_3 + \boldsymbol{u} \tag{2}$$

are given below:

```
Ordinary Least Squares:

Variable    Parameter estimate  Standard error  T statistic

constant       131.986603          33.702976        3.916171
x1               9.497008           0.302840       31.359806
x2               0.050977           0.262696        0.194052
x3               0.456077           0.217973        2.092356

Number of observations = 100   Number of estimated parameters = 4
```

Obtain an equal-tailed confidence interval for the coefficient β_1 at confidence level 95%, using the standard normal distribution.

For a bootstrap confidence interval, the following bootstrap DGP was used:

$$\boldsymbol{y}^* = \hat\beta_0\boldsymbol{\iota} + \hat\beta_1\boldsymbol{x}_1 + \hat\beta_2\boldsymbol{x}_2 + \hat\beta_3\boldsymbol{x}_3 + \boldsymbol{u}^*,$$

the $\hat\beta_i$, $i = 0, 1, 2, 3$ being the OLS estimates from running the regression (2). The bootstrap disturbances $\boldsymbol{u}^*$ were obtained by resampling the residuals $\hat{\boldsymbol{u}}$ from (2). The bootstrap statistics were computed as

$$\tau_b^* = (\beta_1^* - \hat\beta_1)/\text{se}_1^*, \quad b = 1, \ldots, 199,$$

where β_1^* and se_1^* are respectively the OLS estimate and standard error of β_1 from regressing $\boldsymbol{y}^*$ on the regressors in (2). Explain why and how the (empirical) distibution of the τ_b^* can be used to construct a bootstrap confidence interval for β_1.

The τ_b^* were sorted from smallest to greatest. The first 12 order statistics were

$$-2.868238, -2.548953, -2.213190, -2.066852, -2.037952, -1.993715,$$
$$-1.933711, -1.854560, -1.847543, -1.721334, -1.652404, -1.645914,$$

and the last 12 were

$$1.571601, 1.607258, 1.665539, 1.693339, 1.763482, 1.780297,$$
$$1.827205, 1.852331, 1.930293, 2.123498, 2.218978, 2.593901.$$

Determine the upper and lower limits of the equal-tailed bootstrap confidence interval at confidence level 95%.

Determine the **shortest**, not necessarily equal-tailed, bootstrap confidence interval at 95% confidence. Explain how you went about this.

4. The model with AR(1) disturbances can be written as follows:

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \rho y_{t-1} - \rho\boldsymbol{X}_{t-1}\boldsymbol{\beta} + u_t, \tag{3}$$

where we may assume that the disturbances are white noise. This nonlinear regression is a special case of the linear regression

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \rho y_{t-1} + \boldsymbol{X}_{t-1}\boldsymbol{\gamma} + u_t, \tag{4}$$

subject to the nonlinear restrictions given by $\boldsymbol{\gamma} + \rho\boldsymbol{\beta} = \boldsymbol{0}$. These restrictions are known as the **common-factor restrictions**, since they say that the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are proportional, with a common factor $-\rho$ of proportionality.

The file

https://russell-davidson.research.mcgill.ca/data/e662dec24.2.dat

contains 80 observations on variables $\boldsymbol{y}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$. With $\boldsymbol{X} = [\boldsymbol{\iota} \;\; \boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \boldsymbol{x}_3]$ and $\boldsymbol{\beta} = [\alpha \vdots \beta_1 \vdots \beta_2 \vdots \beta_3]$, estimate the unrestricted model (4) by OLS, and obtain estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$ of the parameters $\boldsymbol{\beta}$ and ρ. (**Warning:** the lagged constant is just the constant; take care to avoid collinear regressors.)

Construct the Gauss-Newton regression (GNR) corresponding to the restricted model (3), and evaluate its regressand and regressors at $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$. Obtain an asymptotic F statistic to test the common-factor restrictions, and state the degrees of freedom of its asymptotic distribution, giving your reasoning for the answer. Express the result of the test as an asymptotic P value.

Perform a bootstrap test of the same null hypothesis, namely that the common-factor restrictions hold. Give the explicit form of the bootstrap DGP, and explain why it satisfies Golden Rule 1.

If possible, estimate the restricted model (3) by nonlinear least squares, and use the result to get another version of the F statistic for the null hypothesis. How might you modify the bootstrap DGP in an attempt to satisfy Golden Rule 2?

5. If data are clustered in a linear regression model, the regression can be represented as

$$\boldsymbol{y} \equiv \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_G \end{bmatrix} = \boldsymbol{X\beta} + \boldsymbol{u} \equiv \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_G \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_G \end{bmatrix},$$

where the data are divided into G clusters, indexed by g. The g^{th} cluster has n_g observations. It is shown in the textbook that the covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} &(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &\quad = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\left(\sum_{g=1}^{G}\boldsymbol{X}_g^\top\boldsymbol{\Omega}_g\boldsymbol{X}_g\right)(\boldsymbol{X}^\top\boldsymbol{X})^{-1}, \end{aligned} \tag{5}$$

where $\boldsymbol{\Omega}_g$ is the $n_g \times n_g$ covariance matrix of the disturbances $\boldsymbol{u}_g$ in cluster g.

Consider the very special case mentioned in the textbook, where $n_g = m$ for all $g = 1, \ldots, G$, and $\boldsymbol{X}_g$ has only one column, with every element equal to x_g. If we denote by $\boldsymbol{\iota}$ the m-vector with every element equal to one, then we may write $\boldsymbol{X}_g = x_g\boldsymbol{\iota}$. The parameter vector $\boldsymbol{\beta}$ now has only one component, which we write as β, and its OLS estimator as $\hat{\beta}$.

The disturbances are characterised by an error-components model:

$$u_{gi} = v_g + \varepsilon_{gi}, \quad v_g \sim \text{IID}(0, \sigma_v^2), \quad \varepsilon_{gi} \sim \text{IID}(0, \sigma_\varepsilon^2),$$

for $i = 1, \ldots, m$, $g = 1, \ldots, G$. Here v_g is a random variable that affects every observation in cluster g and no observation in any other cluster, while ε_{gi} is an idiosyncratic shock that affects only the single observation gi. This model implies that

$$\operatorname{Var}(u_{gi}) = \sigma_v^2 + \sigma_\varepsilon^2 \quad \text{and} \quad \operatorname{cov}(u_{gi}, u_{gj}) = \sigma_v^2,$$

so that

$$\rho \equiv \frac{\operatorname{cov}(u_{gi}, u_{gj})}{\operatorname{Var}(u_{gi})} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2} \text{ for all } g \text{ and } i \neq j.$$

Thus all the intra-cluster correlations are the same and equal to ρ. Show that $\boldsymbol{\Omega}_g = \sigma_v^2 \boldsymbol{\iota\iota}^\top + \sigma_\varepsilon^2 \mathbf{I}$, where $\mathbf{I}$ is the $m \times m$ identity matrix.

If the clustering is ignored, the variance of $\hat\beta$ would be wrongly thought to be equal to $\sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$, which, for the special case considered here, is

$$(\sigma_v^2 + \sigma_\varepsilon^2)\Big[\sum_{g=1}^{G} \boldsymbol{X}_g^\top \boldsymbol{X}_g\Big]^{-1}. \tag{6}$$

The true variance is given by the formula (5). Show that this formula gives for our special case a variance equal to

$$(\sigma_\varepsilon^2 + m\sigma_v^2)\Big[m\sum_{g=1}^{G} x_g^2\Big]^{-1}, \tag{7}$$

so that the ratio of the true variance (7) to the incorrect variance (6) is, as stated in the textbook, $1 + (m-1)\rho$.

6. The one-period return on a financial asset can be modelled as a normal random variable W, with expectation zero and variance σ^2, plus a "jump" variable J, independent of W. The variable J is equal to $Q(\gamma + \delta X)$, where X and Q are mutually independent, with $X \sim \mathrm{N}(0,1)$ and Q a Bernoulli (binary) variable, with

$$Q = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

What are the expectation and variance of the return $W + J$? Linearise these in the two limits in which $p \to 1$ and $p \to 0$.

7. The file at

https://russell-davidson.research.mcgill.ca/data/e662dec24.dat

contains 60 observations on three variables, $\boldsymbol{y}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$.

Perform a simulation experiment to investigate the properties of the following test with a null of homoskedasticity against an alternative of heteroskedasticity.

The DGP under the null is

$$y_t = \alpha + x_{t1}\beta_1 + x_{t2}\beta_2 + \sigma u_t, \tag{8}$$

where $\alpha = -30$, $\beta_1 = 4$, $\beta_2 = -3$, $\sigma = 100$, and the u_t are independent standard normal disturbances. It is clear that this DGP satisfies the null. For each of N replications, generate the dependent variable $\boldsymbol{y}$ and run regression (8). Save the vector of residuals $\hat{\boldsymbol{u}}$, and run the testing regression

$$\hat{u}_t^2 = b_\delta + \boldsymbol{Z}_t\boldsymbol{\gamma} + \text{residual}, \quad t = 1, \ldots, 60,$$

(which is equation (9.23) in the textbook), with $\boldsymbol{Z}_t$ the following row matrix:

$$\boldsymbol{Z}_t = [x_{t1} \;\; x_{t2} \;\; x_{t1}^2 \;\; x_{t2}^2 \;\; x_{t1}x_{t2}], \tag{9}$$

essentially the matrix of squares and cross-products of the regressors in (8). The test statistic is n times the centred R^2 from (9) ($n = 60$). At the same time, repeat the calculation of the test statistic under a different DGP:

$$y_t = \alpha + x_{t1}\beta_1 + x_{t2}\beta_2 + \sigma x_{t2}u_t, \tag{10}$$

where there is now heteroskedasticity.

For the null DGP only, perform a bootstrap test, using a bootstrap DGP with disturbances resampled from the residuals $\hat{u}_t$.

Display your results graphically. For the asymptotic tests, graph the empirical distribution functions (EDFs) of the N realisations of the test statistics under the null and the alternative, and, in the same graph, plot the CDF of the $\chi^2(5)$ distribution (chi-squared with 5 degrees of freedom), which is the nominal asymptotic distribution of the statistic under the null. For the bootstrap test, make a separate graph, in which you plot the EDF of the bootstrap P value.

(You should choose N and B, the number of bootstrap repetitions, according to the computing power you have available.)

The nominal distribution of the statistic under the alternative DGP (10) is noncentral χ^2 with 5 degrees of freedom and some noncentrality parameter (NCP) Λ. Estimate Λ by computing the mean of the statistics realised under the alternative.