Queen's University
School of Graduate Studies and Research
Department of Economics

Economics 850 | **Econometrics I** | **Fall, 2022**

Professor James MacKinnon

Final Examination

December 12, 2022. Time: 3 hours

Notes: The examination is in two parts. Please answer the only question in Part I and three (3) questions from Part II. Tables with some critical values of the χ^2 and Student's t distributions appear at the end of the examination.

Part I. Please answer the following question, which is worth 28% of the final mark.

1. Consider the linear regression model with N observations and k regressors,

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \beta_2\boldsymbol{x}_2 + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{X}_1$ is an $N \times k_1$ matrix of observations on k_1 exogenous regressors, $\boldsymbol{x}_2$ is an N-vector of observations on a single exogenous regressor, and $\boldsymbol{y}$ and $\boldsymbol{u}$ are N-vectors of observations on a dependent variable and disturbances, respectively.

a) Suppose the elements of $\boldsymbol{u}$, say u_i, are normally and independently distributed with unknown variance σ^2. How would you estimate β_2? How would you then test the hypothesis that $\beta_2 = 0.75$? Write down your test statistic explicitly as a function of $\boldsymbol{y}$, $\boldsymbol{X}_1$, and $\boldsymbol{x}_2$ (or quantities that depend on them). How is this test statistic distributed when $N = 37$ and $k_1 = 3$?

b) Suppose the u_i are independently distributed with unknown variances σ_i^2 that may be related to the regressors. How would you estimate β_2? Write down the test statistic you would use to test the hypothesis that $\beta_2 = 0.75$ as a function of $\boldsymbol{y}$, $\boldsymbol{X}_1$, and $\boldsymbol{x}_2$ (or quantities that depend on them). What can you say about the distribution of this test statistic when $N = 37$ and $k_1 = 3$? What can you say about it when $N = 4{,}758$ and $k_1 = 44$?

c) Suppose the data used to estimate (1) fall into 13 clusters, indexed by g, for $g = 1, \ldots, 13$. Let $\boldsymbol{u}_g$ denote the disturbance vector for the g^{th} cluster. The $\boldsymbol{u}_g$ are assumed to be independent across clusters but to have unknown variances and covariances. There are 4,758 observations, with cluster sizes ranging from 43 to 984, and $k_1 = 44$. How could you test the hypothesis that $\beta_2 = 0.75$ at the .05 level without estimating the model more than once? You do not need to write down your test statistic explicitly, but it should look similar to the test statistic for part b). What distribution will you pretend that it follows?

Continued on next page . . .

d) Suppose that $\hat{\beta}_2 = 0.934$ and that the test statistic of part c) is 2.194. Would you feel confident in rejecting the null hypothesis at the .05 level? Briefly explain how you could test the hypothesis that $\beta_2 = 0.75$ using an alternative procedure that involves 14 OLS regressions. Would you expect the resulting test statistic to be larger or smaller than 2.194? Why?

Part II. Please answer three (3) of the following five (5) questions. Each question has four (4) parts and is worth 24% of the final mark.

2. Consider the nonlinear regression model

$$y_i = \beta_1 + \beta_2 x_{2i}^{\beta_3} x_{3i}^{1-\beta_3} + u_i, \quad u_i \sim \text{IID}(0, \sigma^2), \tag{2}$$

where the regressors x_{2i} and x_{3i} are assumed to be exogenous, and there are 76 observations.

a) When you estimate regression (2) by nonlinear least squares, the SSR is 142.85. When you impose the restriction that $\beta_3 = 0.5$, the SSR is 151.26. If you assume that $u_i \sim \text{IID}(0, \sigma^2)$, can you reject the null hypothesis that $\beta_3 = 0.5$ at the .05 level using an asymptotic test?

b) Explain how to estimate (2) subject to the restriction that $\beta_3 = 0.5$. Then explain how you would test the hypothesis that $\beta_3 = 0.5$ using a Gauss-Newton regression, or GNR, without doing any nonlinear estimation. If you are not sure what the needed derivatives are, do not waste time on them.

c) Suppose you relaxed the assumption that $u_i \sim \text{IID}(0, \sigma^2)$ and instead assumed that $\text{E}(u_i^2 \,|\, \boldsymbol{X}_i) = \sigma_i^2$, with the σ_i^2 unknown. Here $\boldsymbol{X}_i$ denotes the row vector containing 1, x_{2i}, and x_{3i}. Explain how you could use the GNR of part b) to test the hypothesis that $\beta_3 = 0.5$ under this weaker assumption.

d) Suppose you estimate all three parameters by NLS. Discuss how you would obtain a 99% asymptotic confidence interval for β_3 under the assumptions of part c). Explain how you would compute the needed standard error for $\hat{\beta}_3$.

3. Consider the linear regression model

$$y_{gi} = \beta_1 + \beta_2 x_{gi} + u_{gi}, \quad g = 1, \ldots, G, \quad i = 1, \ldots, N_g, \tag{3}$$

which is to be estimated using a sample of $N = 18{,}674$ observations divided into $G = 16$ clusters. The regressor x_{gi} is assumed to be exogenous. You are interested in the parameter $\gamma \equiv \exp(\beta_2)$.

a) Explain how you would obtain an estimate $\hat{\gamma}$ and an asymptotically valid standard error $s(\hat{\gamma})$ analytically under the assumption that the disturbances in (3) are independent across clusters but may be correlated and/or heteroskedastic within each cluster. Then show how to construct two 95% asymptotic confidence intervals for γ, one symmetric and one asymmetric.

Continued on next page . . .

b) There are two natural ways to generate bootstrap samples for (3) under the assumptions of part a) without imposing any restrictions. Briefly explain how each of them would work in this case. Will both of them generate bootstrap samples with 18,674 observations? Explain.

c) Using whichever of the methods from part b) you prefer, explain how you could obtain a bootstrap standard error $s^*(\hat{\gamma})$ and how you would you use that standard error to construct a 95% confidence interval for γ. Would this interval be symmetric around $\hat{\gamma}$? Explain.

d) Explain how you would construct a 95% studentized bootstrap confidence interval for γ using the standard error $s(\hat{\gamma})$ from part a) and the bootstrap DGP from part c). Would this interval be symmetric around $\hat{\gamma}$? Explain.

4. This question deals with the linear regression model

$$y_{i1} = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 y_{i2} + u_i, \tag{4}$$

which is to be estimated using a dataset with 516 observations. The matrix $\boldsymbol{X}$ has typical row $[1 \quad z_{i1} \quad z_{i2} \quad y_{i2}]$, where the z_{ij} are predetermined and y_{i2} may be endogenous. It is assumed that both $\boldsymbol{X}^\top\boldsymbol{X}$ and the matrix that $1/N$ times it tends to asymptotically have full rank, and that the u_i are homoskedastic and independent. Three predetermined variables, w_{i1}, w_{i2}, and w_{i3}, are also observed. They are believed to be uncorrelated with u_i but correlated with y_{i2}. They are also assumed to satisfy standard regularity conditions.

a) Explain how you would test the null hypothesis that the OLS estimates of the coefficients in (4) are consistent. What would you conclude if the null hypothesis were rejected?

b) How would you obtain consistent estimates of all the coefficients in (4) using an IV estimator? Write down the covariance matrix of the IV estimates that you would report. Would you expect the standard error of $\hat{\beta}_{3\text{IV}}$ to be larger or smaller than the standard error of $\hat{\beta}_{3\text{OLS}}$? Explain.

c) The IV estimator of part b) involves a first-stage regression. Just what is this regression? Investigators often report the value of a certain test statistic associated with this regression. What is this test statistic, how many restrictions is it testing, and how is it distributed asymptotically? What would you conclude if the P value associated with this test statistic were 0.00013? Explain.

d) Suppose the u_i in (4) are assumed to be heteroskedastic. What is the covariance matrix of the IV estimates that you would report now? If the t-statistic for $\beta_3 = 0$ based on the appropriate diagonal element of this matrix were 2.027, would you be comfortable rejecting the hypothesis that $\beta_3 = 0$ at the .05 level? Explain why or why not.

Continued on next page . . .

5. Suppose you are given a sample of 2,365 observations on the incomes of corporate lawyers in 2019. The largest income is \$14,688,455 and the second-largest is \$2,371,346. The mean is \$725,234, and the median is about 2/3 of the mean. The rest of the distribution is more or less as you would expect it to be given these values.

a) If you were to plot the empirical distribution function for this sample, what would it look like? Would it have any interesting features? For example, how would the distance between the α quantile and the median be related to the distance between the median and the $1 - \alpha$ quantile for $\alpha = 0.10$?

b) Explain how you would estimate the first quartile, the median, and the third quartile of the population distribution using this sample. Then explain how you would construct standard errors for these estimates using the bootstrap. Which of the three bootstrap standard errors would you expect to be the largest? Explain.

c) For the sample mean, you could easily construct a sample standard error. Then you could use the bootstrap to form a studentized bootstrap confidence interval at the 0.95 level. Explain how you would do this. Would this interval be symmetric around \$725,234? What would it look like?

d) Suppose that, in addition to the original sample of 2,365 incomes for corporate lawyers, you are given a sample of 1,465 incomes for tax lawyers. Discuss how you could use bootstrap methods to test the hypothesis that the income distributions for corporate lawyers and tax lawyers are the same.

6. Consider the linear regression model

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \tag{5}$$

where the regressors are assumed to be exogenous and the error terms are assumed to be independent and identically distributed with mean zero and variance σ^2.

a) Suppose you have 270 observations, which naturally divide into two subsamples, the first with 160 observations and the second with 110. The sums of squared residuals from OLS estimation of (5) over the whole sample and each of the two subsamples are 23.61, 13.33, and 9.45, respectively. Can you reject the null hypothesis that all the parameters are the same for both subsamples using an asymptotic test at the .01 level? Explain.

b) Explain precisely how you would perform a bootstrap test of the hypothesis that all the parameters are the same for both subsamples using no more than 10^4 bootstrap samples. Be sure to specify the bootstrap DGP and explain how you would decide whether or not to reject the null hypothesis at the .01 level.

Continued on next page ...

c) A more restrictive alternative hypothesis is that

$$y_i = \gamma_1 d_{1i} + \gamma_2(1 - d_{1i}) + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \qquad (6)$$

where d_{1i} is a dummy variable that is equal to 1 if observation i belongs to the first subsample and equal to 0 otherwise. Suppose the SSR from OLS estimation of (6) were 23.18. Using an asymptotic test, would you reject (5) against (6) at the .01 level?

d) Because of the IID assumption, the tests you have done so far must have assumed that the variance of the error terms is the same for both subsamples. Suppose you want to relax this assumption by allowing each of the u_i to have its own variance σ_i^2. Explain how you would generate 9999 bootstrap samples assuming that (5) holds under this weaker assumption. Then explain how you would use these bootstrap samples to test (5) against (6). Be sure to explain why you would, or would not, use the same test statistic as in part b).

Table 1. Upper-Tail Critical Values of the χ^2 Distribution

D.F. / Level	.10	.05	.025	.01
1	2.706	3.841	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086
6	10.645	12.592	14.449	16.812

Table 2. Two-Tail Critical Values of the Student's t Distribution

D.F. / Level	.10	.05	.02	.01
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921