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

Economics 850 **Fall, 2024**

Professor James MacKinnon

Answers to Midterm Examination

October 24, 2024. Time Limit: 80 minutes

Please answer all 3 questions. The marks for each question are shown in square brackets. All answers should be as short as it is reasonably possible to make them.

1. You are given an N-vector $\boldsymbol{y}$ and an $N \times k$ matrix $\boldsymbol{X}$, with $N > k$. You regress $\boldsymbol{y}$ on $\boldsymbol{X}$ to obtain $\hat{\boldsymbol{\beta}}$, a residual vector $\hat{\boldsymbol{u}}$, and a vector of fitted values $\hat{\boldsymbol{y}}$.

a) Suppose you regress $\boldsymbol{y}$ on $\hat{\boldsymbol{y}}$. What can you say about the coefficient estimate? How will the SSR from this regression compare with the one from the original regression?

b) Suppose you regress $\hat{\boldsymbol{u}}$ on $\boldsymbol{X}$. What can you say about the coefficient estimates? How will the SSR from this regression compare with the one from the original regression?

c) Suppose you create a dummy variable $\boldsymbol{d}$ equal to 1 for observation 1 and equal to 0 for all other observations. If you regress $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{d}$, how will the SSR from this regression compare with the one from the original regression? How will it compare with the original sum of squared residuals for observations 2 through N only?

d) For the regression of part c), how is the coefficient on $\boldsymbol{d}$ related to the first element of $\hat{\boldsymbol{u}}$? Could you have calculated this coefficient without running this regression?

[32 marks]

NOTE: Parts of answers in square brackets are just explanations, in case students don't understand. The correct answers are not expected to include them.

ANSWER [8 marks for each part]

a) The estimated coefficient will be 1. The SSR will be identical to the one from the original regression. [The coefficient must be 1, because $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ minimizes the SSR. For any other value, the SSR would be larger. This also shows that the SSR must be identical to the one from the original regression.]

b) The coefficient eatimates will all be 0. The SSR will be identical to the one from the original regression. [We know the coefficients are all 0, because $\hat{\boldsymbol{u}}$ is orthogonal to $\boldsymbol{X}$. Since $\hat{\boldsymbol{u}}$ is just the vector of residuals, its sum of squares is unchanged by regressing it on $\boldsymbol{X}$.]

c) The SSR will be smaller than the original one, because observation 1 will now fit perfectly. It will also be smaller than the original SSR for observations 2 through N only, [because the estimate of $\boldsymbol{\beta}$ will now minimize the SSR for those observations.]

d) The coefficent on $\boldsymbol{d}$ must be larger in absolute value than the first element of $\hat{\boldsymbol{u}}$, because it is equal to $\hat{u}_1/(1-h_1)$, and $0 < h_1 < 1$. This also answers the second part of the question. We just compute the coefficient on $\boldsymbol{d}$ as $\hat{u}_1/(1-h_1)$.

2. Consider the model

$$y_i = \beta_1 + \beta_2 d_i + u_i, \quad u_i \sim \text{IID}(0, \sigma^2), \quad i = 1, \ldots, N, \tag{1}$$

where d_i is a binary variable that equals either 0 or 1.

a) Under the stated assumptions, explain how to construct a 95% confidence interval for β_2. Is this interval exact in finite samples? Why or why not?

b) Consider two different samples for the model (1). One has $\sigma = 1$ and $N = 50$, and the other has $\sigma = 2$ and $N = 200$. Under what circumstances, if any, would you expect the 95% confidence intervals for β_2 from the two samples to be roughly the same length? Explain.

c) Suppose the probability that $d_i = 1$ is δ and may vary across samples. If you have two independent samples with $N = 100$ and the same value of σ, but one has $\delta = 0.2$ and one has $\delta = 0.25$, which sample will probably yield the shortest confidence interval. Will this be true for every such pair of samples? Explain.

[30 marks]

ANSWER [10 marks for each part]

a) This is the usual confidence interval based on the $t(N-2)$ distribution and the OLS standard error for $\hat{\beta}_2$. [Students do not need to write it down explicitly.] The interval is not exact in finite samples, because the u_i are not assumed to be normally distributed.

b) They will be roughly (but not exactly) the same length if the proportion of 1s in $\boldsymbol{d}$ were the same in the two samples.

   [Quadrupling the sample size should halve the standard error on $\hat{\beta}_2$, but doubling σ should double it. Thus, if the matrix $\boldsymbol{X}^\top \boldsymbol{X}/N$ were unchanged and s/σ were unchanged, the confidence interval would be unchanged. The matrix will be unchanged if the proportion of 1s in $\boldsymbol{d}$ were the same, but s will almost certainly be different, so the CIs still won't be exactly the same length.]

[Of course, for any actual samples, the two $\boldsymbol{X}^\top\boldsymbol{X}/N$ matrices will probably differ, and the two values of s will almowt certainly also differ. Thus either interval could be longer, and with such small samples the differences could be substantial.]

c) The actual proportion of 1s in $\boldsymbol{d}$ matters greatly. The closer it is to 0.5, the more efficiently β_2 will be estimated. Thus we would expect the CI to be shorter, most of the time, for $\delta = 0.25$ than for $\delta = 0.2$. But this will not be true all the time, because the actual proportion of 1s may differ substantially from δ. The sample variance of the disturbances also matters. The larger it is, the longer will be the CI.

[It is the actual proportion of 1s in the sample that matters, not the expected proportion. Given the small sample size, there are bound to be some samples with $\delta = 0.20$ that have more 1s than the other sample with $\delta = 0.25$. Similarly, it is the actual value of s that matters, not σ.]

3. Consider the linear regression model

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

where there are N observations and $k = k_1 + k_2$ regressors, with the regressors in $\boldsymbol{X}_1$ exogenous and the ones in $\boldsymbol{X}_2$ predetermined but not exogenous.

a) Write the F statistic for $\boldsymbol{\beta}_2 = \boldsymbol{0}$ explicitly as a function of N, k, k_2, $\boldsymbol{y}$, $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$. It will probably be convenient to use projection matrices that project onto or off the subspaces spanned by $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and/or $\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2]$.

b) To what distribution should you compare the F statistic of part a)? Will this F statistic be exactly distributed according to this distribution? Will it follow this distribution asymptotically? Explain briefly.

c) Write down the Wald statistic for $\boldsymbol{\beta}_2 = \boldsymbol{0}$ in the model (2). How is it related to the F statistic of part a)? How is it distributed asymptotically under the null hypothesis?

d) When the null hypothesis is true, the Wald statistic of part c) must be $O_p(N^a)$. When $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0 \neq \boldsymbol{0}$, the Wald statistic must instead be $O_p(N^b)$. Exactly what are the values of a and b here?

[38 marks]

ANSWER [10, 8, 10, 10]

a) The FWL regression is

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \text{resids}.$$

The numerator of the F statistic is the ESS from this regression, divided by k_2. The denominator is the SSR for either the original regression or the FWL regression, divided by k. Thus

$$F = \frac{\boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y} / k_2}{\boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y} / (n-k)}.$$

b) Compare the F statistic to the $F(k_2, N-k)$ distribution. The test is not exact, because $\boldsymbol{X}_2$ is not exogenous. But it will be asymptotically valid, because the denominator tends to 1 and k_2 times the numerator tends to $\chi^2(k_2)$.

c) The Wald statistic uses the facts that

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}$$

and

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2) = s^2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1}.$$

Thus, because $\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2^{-1} (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} = \mathbf{I}$, the Wald statistic is

$$\frac{1}{s^2} \boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}.$$

This is equal to k_2 times the F statistic. It is asymptotically distributed as $\chi^2(k_2)$.

[It is *not* acceptable just to write down the general formula for a Wald test of the restrictions that $\boldsymbol{\beta} = \mathbf{0}$ or the general linear restrictions $\boldsymbol{R\beta} = \boldsymbol{r}$, because doing this does not show the relationship to the F statistic.]

d) Clearly, $a = 0$ and $b = 1$. [Under the null hypothesis, any asymptotically valid test statistic must be $O_p(1)$; hence $a = 0$. Under the alternative, the noncentrality parameter

$$\frac{1}{\sigma^2} \boldsymbol{\beta}_2^\top \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2 \boldsymbol{\beta}_2$$

is evidently $O_p(n)$. This is the leading-order term, and so the test statistic itself must also be $O_p(n)$; hence $b = 1$.]