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

Economics 850 **Fall, 2024**

Professor James MacKinnon

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]

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,

Continued overleaf...

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]

3. Consider the linear regression model

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

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 $\boldsymbol{\beta}_2 = \mathbf{0}$ 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 $\boldsymbol{\beta}_2 = \mathbf{0}$ 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 $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0 \neq \mathbf{0}$, the Wald statistic must instead be $O_p(N^b)$. Exactly what are the values of a and b here?

[38 marks]