Prof. James MacKinnon *QUEEN'S UNIVERSITY* *Department of Economics*

Economics 850 Fall, 2024

Assignment 4

Due: November 28, 2024

1. Suppose that θ_0 is the true value of a parameter, and $\hat{\theta}$ is a root-N consistent estimator of that parameter. Let $g(\theta)$ be a nonlinear function of θ, which is assumed to be twice differentiable and monotonically increasing.

a) Write down a second-order Taylor series expansion of $g(\hat{\theta})$ around θ_0.

b) Show that the last term in the Taylor expansion of part a) is asymptotically negligible relative to the other terms.

[18]

2. You estimate a model and obtain a parameter estimate $\hat{\beta} = 1.2641$ with a standard error of 0.1642. However, what you are really interested in is the parameter $\gamma \equiv \exp(\beta)$.

a) Calculate $\hat{\gamma}$ and its standard error using the delta method.

b) Construct a .95 confidence interval for γ based on the standard error from part a), assuming that the test statistic follows the standard normal distribution. Is this interval symmetric around $\hat{\gamma}$? Explain.

c) Construct a .95 confidence interval for γ based on inverting a t-statistic for β, assuming that the test statistic follows the standard normal distribution. Is this interval symmetric around $\hat{\gamma}$? Explain.

d) Suppose that $\hat{\beta}$ is actually the sample mean of 83 independent observations that follow a symmetric distribution. In this case, which of the two confidence intervals derived in parts b) and c) is more likely to provide you with reliable inferences? Will either of them be exact? Explain.

[32]

3. The file `trdata.csv` contains 2400 observations for a simulated dataset. The regressand is `y`, and the regressor of particular interest is a treatment dummy called `treat`. The observations fall naturally into 16 groups given by the variable `grp`.

a) Regress `y` on all the other variables. Obtain three different 95% confidence intervals for the coefficient on `treat` under the assumption that the disturbances are independent. Do the disturbances appear to be homoskedastic?

b) For the same regression, obtain confidence intervals based on CV_1, CV_2 and CV_3 covariance matrices. Do the disturbances appear to be independent? Does it matter which covariance matrix you use? Which confidence interval seems most believable?

c) Compute measures of leverage and influence for the variable `treat` in this regression. Do the observations in any of the clusters seem to be particularly influential? Is this in any way related to the results of part b)?

d) Calculate bootstrap P values for the coefficient on `treat` to equal 0, along with studentized bootstrap confidence intervals, using two versions of the unrestricted wild cluster bootstrap. Which of these comes closest to the results based on CV_3? Why does this make sense?

e) Imposing the restriction that the coefficient on `treat` is 0, calculate bootstrap P values using two versions of the restricted wild cluster bootstrap. Also calculate restricted bootstrap confidence intervals based on the same bootstrap methods. Compared with the results in part d), were these results roughly what you expected to see? If not, could this be due to simulation randomness? [50]