FINAL REVIEW PROBLEMS 1

1. Vera knows what the normal linear model is, but she does not know the theory that leads to an exact confidence interval based on the t-distribution. Instead, she tries to use Monte Carlo simulation to provide a confidence interval. With

$$Y = x\beta + \sigma V, \quad V \mid X = x \sim \mathcal{N}(0, I_n),$$

and

$$b = (x'x)^{-1}x'Y, \quad e = Y - xb, \quad \text{SSR} = e'e,$$

she argues that the distribution of

$$(b - \beta)/\sqrt{\text{SSR}},$$

conditional on $X = x$, does not depend upon β or σ (although it does depend upon x).

(a) Is Vera correct so far? Explain.

(b) Now Vera uses Matlab to generate J matrices $Y^{(j)}$, each of which is $n \times 1$ and contains independent draws from the standard normal (N(0,1)) distribution. She calculates

$$b^{(j)} = (x'x)^{-1}x'Y^{(j)}, \quad e^{(j)} = Y^{(j)} - xb^{(j)}, \quad \text{SSR}^{(j)} = e^{(j)\prime}e^{(j)} \qquad (j = 1, ..., J),$$

forms the absolute values

$$|l'b^{(j)}/\sqrt{\text{SSR}^{(j)}}| \qquad (j = 1, ..., J),$$

and sorts them to find the 95$^{\text{th}}$ percentile (.95 quantile) of their empirical distribution; i.e., she finds a value, which she labels CRIT, such that $.95 \cdot J$ of the absolute values are below CRIT and $.05 \cdot J$ of them are above CRIT. Then she argues that

$$[(l'b - \text{CRIT} \cdot \sqrt{\text{SSR}}),\ (l'b + \text{CRIT} \cdot \sqrt{\text{SSR}})]$$

provides an approximation to an exact .95 confidence interval for $l'\beta$, with an approximation error that goes to zero as $J \to \infty$. Is Vera correct? Explain.

2. Provide the Matlab code for question 9 in Computer Note 3. (Check the answer against the Stata output in Note 3.)

3. (a) Computer Note 4 uses Monte Carlo simulation to approximate the probability that the robust confidence interval covers the true value. The simulation uses $J = 1000$ draws. Explain the sense in which the approximation error goes to zero as the number of draws increases.

(b) Explain how the central limit theorem can be used to assess the accuracy of the approximation when $J = 1000$.

4. Computer Note 4 uses Monte Carlo simulation to evaluate the finite sample properties of the robust confidence interval (from Note 8) when the population generating the data in fact satisfies the normal linear model. Now consider a finite sample evaluation when there is heteroskedasticity. Your data are in the $n \times 1$ matrix y and the $n \times K$ matrix x. One possibility is to start by calculating

$$b = (x'x)^{-1}x'y, \quad e = y - xb,$$

and using a nonlinear least-squares program to obtain

$$\hat{\gamma} = \arg\min_{\gamma} \sum_{i=1}^{n} [e_i^2 - \exp(x_i'\gamma)]^2.$$

Explain how to complete the analysis; provide motivation for your procedure.