BOSTON COLLEGE
Department of Economics
EC771: Econometrics
Spring 2004
Prof. Baum, Ms. Uysal

Solution Key for Problem Set 3

1. For the classical normal regression model $\mathbf{y} = \mathbf{x}\beta + \epsilon$ with no constant term and K regressors, what is plim $F[K, n-K] = plim \frac{R^2/K}{(1-R^2)/(n-K)}$, assuming that the true value of β is zero?

The F ratio is computed as $\frac{\mathbf{b'X'Xb}/K}{[\mathbf{e'e}/(n-K)]}$. We substitute $\mathbf{e} = \mathbf{M}\epsilon$, and $\mathbf{b} = \beta + (\mathbf{X'X})^{-1}\mathbf{X}'\epsilon = (\mathbf{X'X})^{-1}\mathbf{X}'\epsilon$. Then,

$$\begin{aligned} F &= [\epsilon' X'(X'X)^{-1}X'X(X'X)^{-1}X'\epsilon/K]/[\epsilon'\mathbf{M}\epsilon/(n-K)] \\ &= [\epsilon'(\mathbf{I}-\mathbf{M})'\epsilon/K]/[\epsilon'\mathbf{M}\epsilon/(n-K)] \end{aligned}$$

The denominator converges to σ^2. The numerator is an idempotent quadratic form in a normal vector. The trace of $(\mathbf{I}-\mathbf{M})$ is K regardless of the sample size, so the numerator is always distributed as σ^2 times a chi-squared variable with K degrees of freedom. Therefore the numerator of F does not converge to a constant, it converges to σ^2/K times a chi-squared variable with K degrees of freedom.Since the denominator of F converges to a constant, σ^2, the statistic converges to a random variable, *(1/K)* times a chi-squared variable with K degrees of freedom.

2. Let e_i be the ith residual in the ordinary least squares regression of $\mathbf{y}$ on $\mathbf{X}$ in the classical regression model, and let ϵ_i be the corresponding true disturbance. Prove that plim$(e_i - \epsilon_i)$=0.

We can write e_i as $e_i = y_i - \mathbf{b}'\mathbf{x_i} = (\beta'\mathbf{x_i} + \epsilon_i) - \mathbf{b}'\mathbf{x_i} = \epsilon_i + (\mathbf{b}-\beta)'\mathbf{x_i}$. We know that plim $\mathbf{b} = \beta$, and $\mathbf{x_i}$ is unchanged and as n increses, so as $n \to \infty$, e_i is arbitrarily close to ϵ_i.

3. For simple regression model $y_i = \mu + \epsilon_i$, $\epsilon_i \sim N[0, \sigma^2]$, prove that the sample mean is consistent and asymptotically normally distributed. Now, consider the alternative estimator $\hat{\mu} = \sum_i w_i y_i$, where $w_i = \frac{i}{(n(n+1)/2)} = \frac{i}{\sum_i i}$. Note that $\sum_i w_i = 1$. Prove that this is a consistent estimator of μ and obtain its asymptotic variance. [Hint: $\sum_i i^2 = n(n+1)(2n+1)/6$.]

The estimator is $\bar{y} = (1/n)\sum_i y_i = (1/n)\sum_i(\mu+\epsilon_i) = \mu + (1/n)\sum_i \epsilon_i$. Then, $E[\bar{y}] = \mu + (1/n)\sum_i E[\epsilon_i] = \mu$ and var$[\bar{y}] = (1/n^2)\sum_i\sum_j cov[\epsilon_i, \epsilon_j] = \sigma^2/n$. Since the mean equals μ and the variance vanishes as $n \to \infty$, $\bar{y}$ is consistent. In addition, since $\bar{y}$ is a linear combination of normally distributed variables, $\bar{y}$ has a normal distribution with the mean and variance given above in every sample. Suppose that ϵ_i were not normally distributed. Then, $\sqrt{n}(\bar{y} - \mu) = (1/\sqrt{n})(\sum_i \epsilon_i)$ satisfies the requirements for the central limit theorem. Thus, the asymptotic normal distribution applies whether or not the disturbances have a normal distribution.

For, the alternative estimator, $\hat{\mu} = \sum_i w_i y_i$, so $E[\hat{\mu}] = \sum_i w_i E[y_i] = \sum_i w_i \mu = \mu \sum_i w_i = \mu$ and $\text{var}[\hat{\mu}] = \sum_i w_i^2 \sigma^2 = \sigma^2 \sum_i w_i^2$. The sum of squares of the weights is $\sum_i w_i^2 = \sum_i i^2/[\sum_i i]^2 = [n(n+1)(2n+1)/6]/[n(n+1)/2]^2 = [2(n^2 + 3n/2 + 1/2)]/[1.5n(n^2 + 2n + 1)]$. As $n \to \infty$, the fraction will be dominated by the term $(1/n)$ and will tend to zero. This establishes the consistency of this estimator. The last expression also provides the asymptotic variance. The large sample can be found as $\text{Asy.var}[\hat{\mu}] = (1/n) \lim_{n\to\infty} var[\sqrt{n}(\hat{\mu} - \mu)]$. For the estimator above, we can use $\text{Asy.var}[\hat{\mu}] = (1/n) \lim_{n\to\infty} nvar[\hat{\mu} - \mu] = (1/n) \lim_{n\to\infty} \sigma^2[2(n^2 + 3n/2 + 1/2)]/[1.5n(n^2 + 2n + 1)] = 1.333\sigma^2$. Notice that this is unambiguously larger than the variance of the sample mean, which is the ordinary least squares estimator.

4. For the model in (5-25) and (5-26), prove that when only x^* is measured with error, the squared correlation between y and x is less than between y^* and x^*. (Note the assumption that $y^* = y$). Does the same hold true if y^* is also measured with error?

Using the notation in the text, $\text{var}[x^*] = Q^*$ so, if $y = \beta x^* + \epsilon$,

$$Corr^2[y, x^*] = (\beta Q^*)^2/[(\beta^2 Q^* + \sigma_\epsilon^2)Q^*] = \beta^2 Q^*/[(\beta^2 Q^* + \sigma_\epsilon^2)]$$

In terms of the erroneously measured variables,

$$\begin{aligned} cov[y, x] &= cov[\beta x^* + \epsilon, x^* + \mu] = \beta Q^* \\ Corr^2[y, x] &= (\beta Q^*)^2/[(\beta^2 Q^* + \sigma_\epsilon^2)(Q^* + \sigma_u^2)] \\ &= [Q^*/(Q^* + \sigma_u^2)]Corr^2[y, x^*] \end{aligned}$$

If y^* is also measured with error, the attenuation in the correlation is made even worse. The numerator of the squared correlation is unchanged, but the term $(\beta^2 Q^* + \sigma_\epsilon^2)$ in the denominator is replaced with $(\beta^2 Q^* + \sigma_\epsilon^2 + \sigma_v^2)$ which reduces the squared correlation yet further.

6. A multiple regression of y on a constant, x_1 and x_2 produces the following results: $\hat{y} = 4 + 0.4x_1 + 0.9x_2$, $R^2 = 8/60$, $\mathbf{e}'\mathbf{e} = 520$, $n = 29$,

$$\begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}$$

Test the hypothesis that two slopes sum to 1.

The estimated covariance matrix for the least squares estimates is

$$s^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{20}{3900}\begin{bmatrix} 3900/29 & 0 & 0 \\ 0 & 80 & -10 \\ 0 & -10 & 50 \end{bmatrix} = \begin{bmatrix} .69 & 0 & 0 \\ 0 & .40 & -.051 \\ 0 & -.051 & .256 \end{bmatrix}$$

where $s^2 = 520/(29 - 3) = 20$. Then, the test may be based on $t = (.4 + .9 - 1)/[.410 + .256 - 2(.051)]^{1/2} = .399$. This is smaller than the critical value of 2.056, so we would not reject the hypothesis.