STATS 413 PROBLEM SET 8

This problem set is due at **noon ET** on **Nov 23, 2021**. Please upload your solutions to Canvas in a PDF file. You are encouraged to collaborate on problem sets with your classmates, but the final write-up (including any code) **must be your own**.

1. Power and sample size. Consider the problem of inferring whether a coin is biased or not. We formalize the problem as that of inferring whether the probability of success of a Bernoulli random variable is $\frac{1}{2}$. You observe n independent realizations of the Bernoulli random variable.

(a) What is an *exact* 0.05-level test of the hypothesis $H_0 : p = \frac{1}{2}$. By exact, we mean a test that controls the Type I error rate at the nominal level *exactly*. All the tests we derived in class are inexact because they are based on asymptotic approximations to the actual distribution of the test statistic.
Solution: The rejection region is when $\sum_{i=0}^{L} \binom{n}{i}(0.5^n) \approx .025$ and $\sum_{i=U}^{n} \binom{n}{i}(0.5^n) \approx .025$. Thus, reject when $S_n \in \{0, ..., L, U, ..., n\}$, where S_n is the number of successes.

(b) What is an asymptotic 0.05-level test of H_0 based on the Gaussian approximation of the distribution of the sample mean.
Solution: Reject when $1.96 < |2\sqrt{n}(\bar{x} - .5)|$

(c) What is the asymptotic approximation of the power of the test if $n = 100$?
Solution: Power: $P(X \in R|\theta \in \Omega_0^c)$. Where R is the rejection region, θ is the true parameter value, and Ω_0^c is the alternative hypothesis space. In our case, the power is $P(\mathbf{z} > z_{.025} - \mu_n) + (\mathbf{z} < -z_{.025} - \mu_n)$. Where $\mathbf{z} \sim N(0, 1)$ and $\mu_n = 2\sqrt{100}\delta$

(d) Assume $p = 0.51$, so the coin is slightly biased. What is the approximate sample size n required for the asymptotic test to have 80% power.
Solution: The power is $P(\mathbf{z} > z_{.025} - 2\sqrt{n}(.01)) + P(\mathbf{z} < -z_{.025} - 2\sqrt{n}(.01))$. $n = 19,623$ give us a power of 80%.

2. Testing linear hypotheses. Recall

$$\begin{aligned}\mathbf{w}_n &= n(C\widehat{\boldsymbol{\beta}}_n - b)^T \left(C\mathbf{s}_n^2\left(\tfrac{1}{n}\mathbf{X}^T\mathbf{X}\right)^{-1}C^T\right)^{-1}(C\widehat{\boldsymbol{\beta}}_n - b)\\ &= \mathbf{s}_n^{-2}(C\widehat{\boldsymbol{\beta}}_n - b)^T(C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}}_n - b)\end{aligned}$$

is the Wald statistic for testing $H_0 : C\beta^* = b$ under conditional homoskedasticity. As we shall see, the Wald statistic is r times the F**-statistic** (r is the dimension of b; the number of constraints/restrictions imposed by H_0)

$$\mathbf{F} = \frac{\text{SSR}(\widetilde{\boldsymbol{\beta}}) - \text{SSR}(\widehat{\boldsymbol{\beta}})}{r \cdot \mathbf{s}_n^2},$$

where

$$\widetilde{\boldsymbol{\beta}} = \arg\min_{\beta\in\mathbf{R}^d}\text{SSR}(\beta) \text{ s.t. } C\beta = b.$$

is the **restricted OLS estimator** and r is the number of constraints imposed by the linear hypothesis. Under Conditions I.1 and I.2 and H_0, it is possible to show that the F-statistic has a $F_{r,n-d}$ distribution, which is the basis of the F-test.

(a) By the method of Lagrange multipliers, the restricted OLS estimator is the solution to

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \boldsymbol{\lambda} \end{bmatrix} - \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ b \end{bmatrix} = 0.$$

Solve the linear system to obtain

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\beta}} - (\mathbf{X}^T\mathbf{X})^{-1}C^T(C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}} - b) \\ \boldsymbol{\lambda} &= (C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}} - b), \end{aligned}$$

where $\widehat{\boldsymbol{\beta}}$ is the (unrestricted) OLS estimator.
Solution: After multiplying out we get two equations:

$$\mathbf{X}^T\mathbf{X}\tilde{\beta} + C^T\lambda - \mathbf{X}^T y = 0 \quad (1)$$

and

$$C\tilde{\beta} - b = 0 \quad (2)$$

This give us:

$$\begin{aligned} \tilde{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y - (\mathbf{X}^T\mathbf{X})^{-1}C^T\lambda \\ \tilde{\beta} = \hat{\beta} - (\mathbf{X}^T\mathbf{X})^{-1}C^T\lambda \end{aligned}$$

and

$$\begin{aligned} C\tilde{\beta} = C\hat{\beta} - C(\mathbf{X}^T\mathbf{X})^{-1}C^T\lambda \\ C(\mathbf{X}^T\mathbf{X})^{-1}C^T\lambda = C\hat{\beta} - b \\ \lambda = (C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\hat{\beta} - b) \end{aligned}$$

(b) Show that

$$\begin{aligned} \text{SSR}(\widetilde{\boldsymbol{\beta}}) - \text{SSR}(\widehat{\boldsymbol{\beta}}) &= (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\ &= (C\widehat{\boldsymbol{\beta}}_n - b)^T(C(\mathbf{X}^T\mathbf{X})^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}}_n - b). \end{aligned}$$

Solution: Note:

$$\begin{aligned} \text{SSR}(\beta) = (y - X\beta)^T(y - X\beta) \\ y^T y - 2(X\beta)^T y + (X\beta)^T(X\beta) \end{aligned}$$

This gives us:

$$
\begin{aligned}
&\mathrm{SSR}(\widetilde{\boldsymbol{\beta}}) - \mathrm{SSR}(\widehat{\boldsymbol{\beta}}) \\
&= y^T y - 2(X\widetilde{\boldsymbol{\beta}})^T y + (X\widetilde{\boldsymbol{\beta}})^T(X\widetilde{\boldsymbol{\beta}}) - y^T y + 2(X\widehat{\boldsymbol{\beta}})^T y - (X\widehat{\boldsymbol{\beta}})^T(X\widehat{\boldsymbol{\beta}}) \\
&= 2\widehat{\boldsymbol{\beta}}^T X^T y - 2\widetilde{\boldsymbol{\beta}}^T X^T y + \widetilde{\boldsymbol{\beta}}^T X^T X\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^T X^T X\widehat{\boldsymbol{\beta}} \\
&= 2(\widehat{\boldsymbol{\beta}}^T - \widetilde{\boldsymbol{\beta}}^T)X^T y + \widetilde{\boldsymbol{\beta}}^T X^T X\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^T X^T X\widehat{\boldsymbol{\beta}} \\
&= Tr(2(\widehat{\boldsymbol{\beta}}^T - \widetilde{\boldsymbol{\beta}}^T)X^T X\widehat{\boldsymbol{\beta}}) + Tr(\widetilde{\boldsymbol{\beta}}^T X^T X\widetilde{\boldsymbol{\beta}}) - Tr(\widehat{\boldsymbol{\beta}}^T X^T X\widehat{\boldsymbol{\beta}}) \\
&= Tr(2\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^T - \widetilde{\boldsymbol{\beta}}^T)X^T X) + Tr(\widetilde{\boldsymbol{\beta}}\widetilde{\boldsymbol{\beta}}^T X^T X) - Tr(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^T X^T X) \\
&= Tr((2\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^T - \widetilde{\boldsymbol{\beta}}^T) + \widetilde{\boldsymbol{\beta}}\widetilde{\boldsymbol{\beta}}^T - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^T)X^T X) \\
&= Tr((\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^T - 2\widehat{\boldsymbol{\beta}}\widetilde{\boldsymbol{\beta}}^T + \widetilde{\boldsymbol{\beta}}\widetilde{\boldsymbol{\beta}}^T)X^T X) \\
&= Tr((\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T X^T X) \\
&= Tr((\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T X^T X(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})) \\
&= (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T X^T X(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\
&= (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T X^T X(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\
&= ((X^T X)^{-1}C^T(C(X^T X)^{-1}C^T)^{-1}(C\hat{\beta} - b))^T(X^T X)(X^T X)^{-1}C^T(C(X^T X)^{-1}C^T)^{-1}(C\hat{\beta} - b) \\
&= (C\hat{\beta} - b)^T(C(X^T X)^{-1}C^T)^{-1}C(X^T X)^{-1}C^T(C(X^T X)^{-1}C^T)^{-1}(C\hat{\beta} - b) \\
&= (C\widehat{\boldsymbol{\beta}} - b)^T(C(X^T X)^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}} - b)
\end{aligned}
$$

(c) Check that (a) and (b) together implies the Wald statistic is r times the F-statistic.

Solution: From above:

$$
\mathbf{F} = \frac{(C\widehat{\boldsymbol{\beta}} - b)^T(C(X^T X)^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}} - b)}{r * s_n^2}
$$

$$
\Rightarrow r * \mathbf{F} = s_n^{-2}(C\widehat{\boldsymbol{\beta}} - b)^T(C(X^T X)^{-1}C^T)^{-1}(C\widehat{\boldsymbol{\beta}} - b) = \mathbf{w}_n
$$

3. Returns to scale lab continued. In the returns to scale lab, we fitted the restricted model $\mathbf{y}_i = \mathbf{x}_i^T\beta + \boldsymbol{\epsilon}_i$, where

$$
\mathbf{y}_i = \log\frac{\mathbf{c}_i}{\mathbf{p}_{c,i}}, \quad \mathbf{x}_i = \begin{bmatrix} 1 & \log\mathbf{q}_i & \log\frac{\mathbf{p}_{l,i}}{\mathbf{p}_{c,i}} & \log\frac{\mathbf{p}_{f,i}}{\mathbf{p}_{c,i}} \end{bmatrix}^T, \tag{3.1}
$$

to data on 145 US electric utility companies in 1955. The data consists of

- $\mathbf{c}_i$: total costs (in millions of dollars)
- $\mathbf{q}_i$: output (in terawatt hours)
- $\mathbf{p}_{l,i}$: price of labor
- $\mathbf{p}_{f,i}$: price of fuel
- $\mathbf{p}_{c,i}$: price of capital

We discovered the residuals (of the restricted model) has trends that suggest the restricted model is misspecified. To address this issue, we divided the companies into 5 groups of 29 companies, ordered by output, and fit the (restricted) model separately to each group:

$$\begin{bmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(5)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} & & \\ & \ddots & \\ & & \mathbf{X}^{(5)} \end{bmatrix} \begin{bmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(5)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}^{(1)} \\ \vdots \\ \boldsymbol{\epsilon}^{(5)} \end{bmatrix},$$

where $\mathbf{y}^{(k)}, \boldsymbol{\epsilon}^{(k)} \in \mathbf{R}^{29}$ are the vectors of responses and error terms of the companies in the k-th group, $\mathbf{X}^{(k)} \in \mathbf{R}^{29\times 4}$ is the matrix of features of the companies in the k-th group, and $\beta^{(k)} \in \mathbf{R}^4$ is the vector of regression coefficients of the k-th group. Test the null hypothesis that the coefficients of the restricted model are identical across the groups:

$$H_0 : \beta_*^{(1)} = \cdots = \beta_*^{(5)}.$$

In econometrics, this test is called the **Chow test for structural change**. Report the value of the test statistic and p-value.

Solution: Done in lab. The wald statistic was 67.3118 and the pval was 2.936197e-08. Thus we reject the null.