Universidad Católica del Norte, Economic Department

Econometrics: Problem Set 2

Professor: Mauricio Sarrias

First Semester 2018

This homework is due on 10th, April at 18:00 hrs. Please, write your answers clearly. You should also report the codes (R-script or do-file) for Section 2.

1 Theory

1. (Restricted regression and F ratio) We derive the F test in class. However, there is a convenient alternative formula involving two different sum of squared residuals that you probably learned in your introductory class of Econometrics : one is SSR, the minimized sum of squared residuals denoted as SSR_U, and the other is the restricted sum of squared residuals, denoted SSR_R, obtained from:

$$\min_{\boldsymbol{\beta}} SSR(\widetilde{\boldsymbol{\beta}}) \quad st \quad \boldsymbol{R}\widetilde{\boldsymbol{\beta}} = \boldsymbol{r}$$

   Finding the $\widetilde{\boldsymbol{\beta}}$ that achieves this constrained minimization is called the **restricted regression** or **restricted least squared**. In this exercise I ask you to show that the F-ratio you learned in class equals:

$$F = \frac{(SSR_R - SSR_U)/\#\boldsymbol{r}}{SSR_U/(n-K)}$$

   In the restricted least squares, the sum of squared residuals is minimized subject to the constraint implied by the null hypothesis $\boldsymbol{R}\boldsymbol{\beta}_0 = \boldsymbol{r}$. Form the Lagrangian as:

$$\mathcal{L} = \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right) + \boldsymbol{\lambda}'\left(\boldsymbol{R}\widetilde{\boldsymbol{\beta}} - \boldsymbol{r}\right),$$

where $\boldsymbol{\lambda}$ here is the $\#\boldsymbol{r}$-dimensional vector of Lagrange multipliers (recall: $\boldsymbol{R}$ is $\#\boldsymbol{r} \times K$, $\widetilde{\boldsymbol{\beta}}$ is $K \times 1$, and $\boldsymbol{r}$ is $\#\boldsymbol{r} \times 1$). Let $\widetilde{\boldsymbol{\beta}}$ be the restricted least squares estimator of $\boldsymbol{\beta}$. It is the solution to the constrained minimization problem.

a) Let $\widehat{\boldsymbol{\beta}}$ be the unrestricted OLS estimator. Show:

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\beta}} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\left[\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
\boldsymbol{\lambda} &= \left[\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)
\end{aligned}.$$

b) Let $\widetilde{\boldsymbol{\varepsilon}} \equiv \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$, the residuals from the restricted regression. Show:

$$SSR_R - SSR_U = (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\left(\boldsymbol{X}'\boldsymbol{X}\right)(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) = \widetilde{\boldsymbol{\varepsilon}}'\boldsymbol{P}\widetilde{\boldsymbol{\varepsilon}}.$$

where $\boldsymbol{P}$ is the projection matrix. (Hint: Add and subtract $\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})$ to $\widetilde{\boldsymbol{\varepsilon}} \equiv \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$).

c) Verify that you have proved in (b) that both F-ratios are equal.

2. (Chow Test) Assume that you are interested in whether the coefficients of one group of the data (women, for example) are equal to those of other group (men). For instance, consider the following regressions

$$\begin{aligned}
y_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \quad \text{for all observations} \\
y_{i1} &= \boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 + \epsilon_{i1} \quad \text{for } n_1 \text{ observations (group 1)} \\
y_{i2} &= \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 + \epsilon_{i2} \quad \text{for } n_2 \text{ observations (group 2)}
\end{aligned} \tag{1.1}$$

Explain how you would test $H_0 = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. (Hint: Use the F test you derived in previous question and think that the pooled model (model 1) is the restricted model, whereas allowing different coefficients for each group is the unrestricted model.)

2 Applications

In this section you have to work with the data set `wage_obesity.dta`. This dataset is a sample from the EPS (Encuesta de Protección Social), which is the largest and oldest panel-type survey that exists in Chile, with a sample of around 16,000 respondents distributed across all regions of the country. It includes the labor and social security history of the respondents with details information in areas as education, health, social security, job training, assets and family history. You are free to use R or STATA.
In this particular exercise, we are interested in the relationship between Body Mass Index (bmi) and hourly wages. The BMI is defined as the body mass divided by the square

of the body height, and is universally expressed in units of kg/m^2, resulting from mass in kilograms and height in meters. The BMI is an attempt to quantify the amount of tissue mass (muscle, fat, and bone) in an individual, and then categorize that person as underweight, normal weight, overweight, or obese based on that value.[1]

1. Before running any regression, a potential problem we have to deal with is that the dependent variable (`lhwage`) might have outliers. In order to minimize the problem due to outliers, please generate a new variable (`lhwage_no`) which should be equal to `lhwage` if and only if the studentized residuals from this regression

   ```
   reg lhwage bmi i.female i.female#c.bmi i.married exp c.exp#c.exp age ///
        i.year i.collar i.sector i.region i.smoker i.diab schooling
   ```

   are located at the tails of the distribution (above and below -2.5 and 2.5 respectively). Report summary statistics for both `lhwage` and `lhwage_no` and comment the results. (Hint: studentized residuals are obtained using `predict, rstudent` in Stata for example.)

2. Consider the following model:

   $$\texttt{lhwage\_no}_i = \beta_0 + \beta_1 \texttt{bmi}_i + \boldsymbol{x}_i' \gamma + \varepsilon_i, \quad i = 1, ..., n \tag{2.1}$$

   where `lhwage_no` is the natural logarithm of hourly wage without outliers (`lhwage_no`) for individual i, `bmi` is the BMI, and $\boldsymbol{x}_i'$ is a vector of controls. What sign do you expect for β_1? Why?

3. Run a OLS model for model in Equation 2.1 where the controls are: female, married, experience, experience squared, age, a dummy variable indicating whether the individual has diabetes, and a set of dummies for year, collar, sector and region. What is the predicted ceteris paribus difference in hourly wage for individuals with bmi different by one? Please state you answer in percentage.

4. One potential pitfall of previous model is that BMI's coefficient is probably capturing other health-related problems that affect wages (via productivity) and not the true effect of weight on hourly wages. Re-estimate the model for question (3), but now add the dummy variable indicating whether the individual has diabetes (`diab`). Explain what happens to the BMI's coefficient using the bias-equation. (Hint: see section 1.3.4 on class notes.)

5. Suppose that you are presenting the model estimated in (4) in a conference. Suddenly, someone tells you that your coefficient for BMI is biased since you are omitting `schooling`. Estimate the model you estimate in (4) (including `diab`) and add `schooling`. Interpret the results using again the bias-equation formulation.

[1] For a light discussion about the relationship between obesity and wage see `https://www.nytimes.com/roomfordebate/2011/11/28/should-legislation-protect-obese-people/the-obesity-wage-penalty`

6. Write a model that would allow you to test whether BMI has different effects on wages for men and women. How would you test that there are no differences in the effects of BMI for men and women?

7. Run the model you proposed in (6) and interpret the results carefully (use the same controls as in question (5)).

8. Using the estimates from your model in (7), plot the predicted values of `lhwage_no` for different values of BMI over men and women. At what value of BMI, a woman earns the same as a man, given the same levels of BMI?

9. Describe how you test the hypothesis that the expected increase of ln(`hwage`) for a 2-year of experience worker is 4. You do not need to derive the theory behind your procedure.

10. Now, assume that you suspect that the coefficients for all the variables are different for men and women. Please, perform the Chow test you described in question 2 in first section. What is your conclusion?

11. Do you think that `bmi` is exogenous? Explain.