Problem Set 4

Artschil Okropiridse*

February 16, 2024

I try to be persistent, but can't promise no typos (check how its printed in Hanson when you get confused): When I write $\mathbf{X}$, I mean sample, thus a $n \times k$ matrix. When I write X, I mean population, thus a $k \times 1$ vector.

1 Hypothesis Testing

Exercise 1

(Ch. 9, ex. 9.2) You have two independent samples $(y_{1i}, \mathbf{X}_{1i})$ and $(y_{2i}, \mathbf{X}_{2i})$ both with sample sizes n which satisfy $y_1 = X_1'\beta_1 + e_1$ and $y_2 = X_2'\beta_2 + e_2$, where $E[X_1 e_1] = 0$ and $E[X_2 e_2] = 0$. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the OLS estimators of $\beta_1 \in R^k$ and $\beta_2 \in R^k$.

(a) Find the asymptotic distribution of $\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1))$ as $n \to \infty$.

(b) Find an appropriate test statistic for $H_0 : \beta_2 = \beta_1$.

(c) Find the asymptotic distribution of this statistic under H_0.

Exercise 2

(Ch. 9, ex. 9.7) Take the model $y = X\beta_1 + X^2\beta_2 + e$ with $\mathbf{e}[\mathbf{X}e] = 0$ where y is wage (dollars per hour) and X is age. Describe how you would test the hypothesis that the expected wage for a 40-year old worker is \$20 an hour.

Exercise 3

(Ch. 9, ex. 9.10) In Exercise 7.8 you showed that $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \to_d N(0, V)$ as $n \to \infty$ for some V. Let $\hat{V}$ be an estimator of V.

*many thanks to Jakob Beuschlein

(a) Using this result construct a t-statistic for $\mathrm{H}_0 : \sigma^2 = 1$ against $\mathrm{H}_1 : \sigma^2 \neq 1$.

(b) Using the Delta Method find the asymptotic distribution of $\sqrt{n}(\hat{\sigma} - \sigma)$.

(c) Use the previous result to construct a t-statistic for $\mathrm{H}_0 : \sigma = 1$ against $\mathrm{H}_1 : \sigma \neq 1$.

(d) Are the null hypothesis in a) and c) the same or are they different? Are the tests in a) and b) the same or are they different? If they are different, describe a context in which the two tests would give contradictory results.

Exercise 4

(Ch. 9, ex. 9.26) In a paper in 1963, Marc Nerlove analyzed a cost function for 145 American electric companies. Nerlov was interested in estimating a cost function: $C = f(Q, PL, PF, PK)$ where the variables are listed in the table below. His data set `Nerlove1963` is on the textbook website.

C	Total Cost
Q	Output
PL	Unit price of labor
PK	Unit price of capital
PF	Unit price of fuel

First, estimate an unrestricted Cobb-Douglas specification

$$\log(C) = \beta_1 + \beta_2 \log Q + \beta_3 \log PL + \beta_4 \log PK + \beta_5 \log PF + e.$$

Report parameter estimates and standard errors.

(a) What is the economic meaning of the restriction $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$?

(b) Estimate the model by constrained least squares imposing $\beta_3 + \beta_4 + \beta_5 = 1$. Report your parameter estimates and standard errors.

(c) Estimate the model by efficient minimum distance imposing $\beta_3 + \beta_4 + \beta_5 = 1$. Report your parameter estimates and standard errors.

(d) Test $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$ using a Wald statistic.

(e) Test $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$ using a minimum distance statistic.

Exercise 5

(Ch. 9, ex. 9.28) Using the `cps09mar` dataset and the subsample of non-Hispanic Black individuals (race code = 2) test the hypothesis that marriage status does not affect mean wages.

(a) Take the regression reported in Table 4.1. Which variables will need to be omitted to estimate a regression for this subsample?

(b) Express the hypothesis"marriage status does not affect mean wages" as a restriction on the coefficients. How many restrictions is this?

(c) Find the Wald for this hypothesis. What is the appropriate distribution for the test statistic? Calculate the p-value of the test.

(d) What do you conclude

2 Lecture 9 Resampling methods

Exercise 6

(Ch. 10, ex. 10.19) Take the model $y = \mathbf{X}'\beta + e$ with $\mathbf{e}[\mathbf{X}e] = 0$. Describe the bootstrap percentile confidence interval for $\sigma^2 = \mathbf{e}[e^2]$.

Exercise 7

(Ch 10, ex. 10.20) The model is $y = X_1\beta_1 + X_2\beta_2 + e$ with $\mathbf{e}[\mathbf{X}e] = 0$ and X_2 scalar. Describe how to test $\mathrm{H}_0 : \beta_2 = 0$ against $\mathrm{H}_1 : \beta_2 \neq 0$ using the nonparametric bootstrap.

Exercise 8

(Ch. 10, ex. 10.28) In Exercise 9.26 you estimated a cost function for 145 electric companies and tested the restriction $\theta = \beta_3 + \beta_4 + \beta_5 = 1$.

(a) Estimate the regression by unrestricted least squares and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(b) Estimate $\theta = \beta_3 + \beta_4 + \beta_5$ and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(c) Report confidence intervals for θ using the percentile and BCa methods.

Exercise 9

(Ch. 10, ex. 10.30) In Exercise 7.28 you estimated a wage regression with the `cps09mar` dataset and the subsample of white Male Hispanics. Further restrict the sample to those never-married and live in the Midwest region. (This sample has 99 observations.) As in subquestion (b) let θ be the ratio of the return to one year of education to the return of one year of experience.

(a) Estimate θ and report standard errors calculated by asymptotic, jackknife and the bootstrap.

(b) Explain the discrepancy between the standard errors.

(c) Report confidence intervals for θ using the BC percentile method.