Ec240a, Fall 2024

Professor Bryan Graham

Problem Set 3

Due: November 27th, 2024

Problem sets are due at 5PM. The GSI will provide instructions on how to turn in your problem set. You may work in groups, but each student should turn in their own write-up (including a "printout" of a narrated/commented and executed Jupyter Notebook if applicable). Please also e-mail a copy of any Jupyter Notebook to the GSI (if applicable).

1 Bayesian Bootstrap

For this part of the problem set you may find the article by Chamberlain & Imbens (2003) helpful. David Card's Fisher-Schultz lecture is a useful overview of the literature on estimating the return to schooling (Card, 2001).

The file `nlsy97ss.csv` is in the problem sets folder on GitHub. This is a comma delimited text file It includes a measure of average annual earnings (`avg_earn_2014_to_2018`), years of schooling (`hgc_ever`), 'AFQT' score (`asvab`), a female dummy and two ethnicity dummies for a sub-sample of respondents in the National Longitudinal Survey of Youth 1997 cohort. Earnings equals average annual earnings over the 2014, 2016 and 2018 calendar years in 2012 prices. Define `LogEarn` to be the natural logarithm of Earnings.

1. Construct a sub-sample of non-black, non-hispanic, non-female respondents with positive earnings. Construct the `LogEarn` variable. Create a table of summary statistics for `avg_earn_2014_to_2018, LogEarn, hgc_ever` and `asvab` for this sub-sample.

2. Compute the least squares fit of `LogEarn` onto a constant and `hgc_ever`. Report the point estimate on the schooling variable as well as its heteroscedastic robust asymptotic standard error (you may use the `StatsModels` implementation of OLS to do this; later in the course we will construct our own program for these calculations).

3. Compute the least squares fit of `LogEarn` on a constant, `hgc_ever` and `asvab`. Does the estimate coefficient on `hgc_ever` change?

4. Estimate the parameters of the following linear regression model by the method of least squares

$$\mathbb{E}^*[\texttt{LogEarn}|\,\texttt{X}] = \alpha_0 + \beta_0\texttt{hgc\_ever} + \gamma_0\texttt{hgc\_ever} \times (\texttt{asvab} - 50) + \delta_0\texttt{asvab}$$

   where $X = (\texttt{hgc\_ever}, \texttt{hgc\_ever} \times (\texttt{asvab} - 50), \texttt{asvab})'$.

   (a) Provide a semi-elasticity interpretation of β_0.

   (b) Provide a semi-elasticity interpretation of $\beta_0 + \gamma_0 (\texttt{asvab} - 50)$.

5. Construct a plot with the OLS estimate of $\beta_0 + \gamma_0 (\texttt{asvab} - 50)$ on the y-axis and a grid of `asvab` values on the x-axis.

6. Using the Bayes' Bootstrap to approximate a posterior distribution for $\beta_0 + \gamma_0\,(\texttt{asvab} - 50)$ at each value of `asvab` shown in your plot. Add (estimates of) the 0.025 and 0.975 quantiles, as well as the mean, of the posterior distribution of $\beta_0 + \gamma_0\,(\texttt{asvab} - 50)$ to your plot.

7. Summarize what you have learned about the relationship between earnings, schooling and AFQT among white male millennials?

8. Repeat your analysis for another demographic group of your choice and discuss your findings.

References

Card, D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica*, 69(5), 1127 – 1160.

Chamberlain, G. & Imbens, G. W. (2003). Nonparametric applications of bayesian inference. *Journal of Business and Economic Statistics*, 21(1), 12 – 18.