Exercise 1: Introduction to Stata

This exercise is partially based on problem 4.1. of Wooldridge (2010) „Econometric analysis of cross section and panel data“.

Before starting with this exercise you want to have a first look at my introductory slides on Stata. You can download the NLS80.RAW data set using the following link `https://www.stata.com/data/jwooldridge/eacsap/nls80.dta`.[1]

(a) Load the data into Stata and familiarize yourself with the structure and the variables using different descriptive statistics.

(b) Have a closer look at the wage variable (monthly earnings in \$). How does the mean wage compare to the median? Plot the wage distribution. What is the expected wage of a working man if he is married?

(c) Consider the following equation for men under the assumption that all explanatory variables are exogenous:[2]

$$wage = \beta_0 + \beta_1 married + \beta_2 high\_iq + \beta_3 medium\_educ + \beta_4 high\_educ + z\gamma + u$$

$$E(u|married, iq, educ, z) = 0$$

where z contains further observed factors other than marital status, education, and intelligence quotient which are also included in the data set and can can affect wage.

How do you interpret β_1? How does the constant interpret?

(d) Run the regression and interpret the regression coefficients. Include all education dummies to the regression model. Why does Stata drop one of the dummies? How does the interpretation of the constant change if you change the included education dummies?

(e) Why would you like to interact the $married-$dummy with the $high\_iq-$ dummy? Run the regression including the interaction term and interpret the coefficient.

(f) Consider now the a standard log($wage$)-equation:

$$log(wage) = \beta_0 + \beta_1 married + \beta_2 educ + z\gamma + u$$

When β_1 is small, $100 \cdot \beta_1$ is approximately the ceteris paribus percentage difference

[1] NLS data on wages of working men.

[2] Use the iq-variable to generate the *high_iq*-dummy that indicates if the intelligence quotient is above the median. Similarly, define the *medium_educ*-dummy (*high_educ*-, *low_educ*-dummy) if the years of education are eaqual to 13 years (more than 13 years, below 13 years).

in wages between married and unmarried men.[3] When β_1 is large, it is preferable to use the exact percentage difference in $E(wage|married, educ, z)$. Call this θ_1. A natural, consistent, estimator of θ_1 is $\hat{\theta_1} = 100 \cdot [exp(\hat{\beta_1}) - 1]$ where $\hat{\beta_1}$ is the OLS estimator from the previous equation. Show that, if u is independent of all explanatory variables in the equation, then $\theta_1 = 100 \cdot [exp(\beta_1) - 1]$.

(g) Assume now a log-log specification. How does β_2 interpret? Derive the interpretation of β_2 analytically.

$log(wage) = \beta_0 + \beta_1 married + \beta_2 log(educ) + z\gamma + u$

(h) Assume now a level-log specification. How does β_2 interpret? Derive the interpretation of β_2 analytically.

$wage = \beta_0 + \beta_1 married + \beta_2 log(educ) + z\gamma + u$

Exercise 2: Simultaneous Equations

Consider the system of two simultaneous equations.

$$y_{i1} = y_{i2}\alpha_1 + x_i\beta_1 + u_{i1}$$
$$y_{i2} = y_{i1}\alpha_2 + x_i\beta_2 + u_{i2}$$

where y_{i1} and y_{i2} are endogenous variables and x_i is a single exogenous regressors. There are four scalar structural parameters α_1, α_2, β_1 and β_2. In matrix notation the structural equations can be re-written as

$$y_i'\Gamma = x_i\Delta + u_i'$$

where $y_i = (y_{i1}, y_{i2})'$, $u_i = (u_{i1}, u_{i2})'$, Γ is a 2×2 matrix and Δ is a 1×2 vector. The reduced form model can be written as

$$y_i' = x_i\Pi + \epsilon_i'$$

where $\Pi = \Delta\Gamma^{-1} = (\pi_1, \pi_2)$ is the 1×2 vector of reduced form parameters and $\epsilon_i' = u_i'\Gamma^{-1}$.

(a) Express Γ and Δ in terms of α_1, α_2, β_1 and β_2.

(b) Calculate $\det(\Gamma)$. Under what condition on α_1, α_2 is Γ invertible? In the following we assume that $\det(\Gamma)$ is invertible.

(c) Express Π in terms of α_1, α_2, β_1 and β_2.

[3] Interpretation is approximately true for values $-0.1 < \beta_1 < 0.1$

(d) Assume that $\beta_1 = 0$ and $\pi_2 \neq 0$. Show that α_1 is identified and find an expression for α_1 in terms of π_1 and π_2.

(e) Assume that $\alpha_1 = -\alpha_2$, $\beta_1 = \beta_2$, and $\pi_1 + \pi_2 \neq 0$. Show that all structural parameters are identified and find an expression for α_1, α_2, β_1 and β_2 in terms of π_1 and π_2.