Worksheet week # 7

1. Consider the following model describing the corporate investment behavior

$$I_t = \beta_0 + \beta_1 F_{t-1} + \beta_2 K_{t-1} + \varepsilon_t \quad , \quad t = 1935, ..., 1953 \quad ,$$

where I_t = current gross investment, F_{t-1} = end-of-period value of outstanding shares, and K_{t-1} = end-of-period capital stock. The regression was estimated from the data for General Motors Corporation (standard errors in parentheses):

$$\hat{I}_t = \underset{(97.436)}{-\ 109.8} + \underset{(0.0235)}{0.1142}\ F_{t-1} + \underset{(0.0394)}{0.3261}\ K_{t-1} \quad , \quad R^2 = 0.8907 \quad .$$

   (a) Test that the amount of capital stock affects positively future investment behavior.

   (b) Explain how you would test the hypothesis that no relevant explanatory variables have been omitted using the RESET test.

2. Imagine that you want to estimate the race-specific crime rates. Given the data on over 2000 criminals (*crime2.gdt*), estimate the relationship between the race and the number of crimes committed. The dataset contains the following variables:

crime86 - number of crimes committed in 1986
race - race (=1 if black, =2 if Hispanic, =0 otherwise)
tottime - total number of months spent in prison since 18 years old
pcnv - proportion of prior convictions
qemp86 - number of quarters employed in 1986
inc86 - legal income in 1986, $100s

   (a) Estimate the baseline model of the impact of race on number of crimes committed in 1986:

$$crime86_i = \alpha_0 + \alpha_1 race_i + \varepsilon_i \quad .$$

   (b) Interpret the results. Do you believe that the coefficient α_1 is correctly estimated? Under what assumptions would it be?

   (c) Create two dummy variables for black and Hispanic individuals. Estimate the equation again with these two variables. Interpret the results.

$$crime86_i = \beta_0 + \beta_1 black_i + \beta_2 hispanic_i + v_i \quad .$$

(d) Is there anything that could still create a bias in this equation? If yes, how would you solve for this problem? What direction of bias do you expect?

(e) Re-estimate the equation with variables controlling for crime history of a person:

$$crime86_i = \gamma_0 + \gamma_1 black_i + \gamma_2 hispanic_i + \gamma_3 tottime_i + \gamma_4 pcnv_i + e_i \quad .$$

(f) Control further for a current employment status and income of an individual:

$$crime86_i = \delta_0 + \delta_1 black_i + \delta_2 hispanic_i + \delta_3 tottime_i + \delta_4 pcnv_i + \delta_5 qemp86_i + \delta_6 inc86_i + u_i \quad .$$

(g) Interpret the results from part e and f (in comparison with c). How did the coefficients of *black* and *hispanic* change? Did you expect this direction of potential bias? Would you conclude that the additional variables indeed belong to the model?

(h) Test the hypothesis that no relevant explanatory variables have been omitted using the RESET test in Gretl (test for the model from part f).

3. Use data *wage.gdt* to estimate the returns to education equation.

(a) Estimate the baseline model of the impact of education and experience on wages:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \varepsilon_i \quad .$$

Interpret the meaning of the coefficient β_1.

(b) Reestimate the model using robust standard errors, comment on the differences.

(c) Test for heteroskedasticity in the model in part (a). Is it necessary to use robust standard errors in this case?

(d) Estimate the model with quadratic specification of experience:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \varepsilon_i \quad .$$

Comment on how and why the coefficient β_2 changed with respect to part (a). Did the coefficient β_1 change as well? Why or why not?

(e) Do you believe that the coefficient β_1 is correctly estimated? Is there anything that could create a bias in this equation? If yes, how would you solve for this problem?

(f) Include in the model the education of the mother and of the father of the observed individuals:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 motheduc_i + \beta_5 fatheduc_i + \varepsilon_i \quad .$$

i. Is there an impact on the coefficient β_1? Does this signal there was a bias in the model from part (d)? Comment on the sign of this bias.

ii. Are both *motheduc* and *fatheduc* individually significant? Are they jointly significant?

iii. What happens if you exclude one these variables from the regression? Which one would you keep?

(g) Instead of the education of parents, include the variable measuring ability in the model:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper^2 + \beta_4 abil + \varepsilon_i \quad .$$

Is there an impact on the coefficient β_1? Does this signal there was a bias in the model from part (d)? Comment on the sign of this bias.