Exercise Sheet 5
Specification Errors

Note: Some of the exercises include estimations and references to the data files. Use these to compare them to the results you obtained with Gretl.

1. Assume the following model
$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$
complies with $E(\varepsilon|X_1) = 0$ for all X_1, but now we estimate using OLS, the regression without a constant of Y on X_1, that is, $\widetilde{Y} = \delta_1 X_1$. Is $\widehat{\delta}_1$ a consistent estimator of β_1? Explain your answer. Use your answer to evaluate the possibility to exclude the constant term from the regression.

2. Suppose that the appropriate model is
$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i2}^2 + u_i$$
but it is estimated
$$Y_i = \beta_1^* + \beta_2^* X_{i2} + \varepsilon_i$$
   a) Is the OLS estimator of β_2^*consistent for β_2?
   b) When is the bias the OLS estimator of β_2^* positive and when is negative?
   c) Would the problems be similar if the appropiate regression were
$$Y_i = \beta_1 + \beta_2 \ln X_{i2} + \beta_3 \ln X_{i2}^2 + v_i?$$

3. [Problem 9.4 in Wooldridge Textbook] The following equation describes the number of hours of television watched per week by a child as a function of his age, his education, his mother's education, his father's education, and the number of siblings:
$$tvhours^* = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 motheduc + \beta_4 fatheduc + \beta_5 sibs + u.$$
We suspect that $tvhours^*$ contains a certain error of measurement. If $tvhours$ is declared number of hours of television watched per week
   a) What are the requirements for the classic assumption on errors in variables in this context?
   b) Is it likely that the classical assumption on errors in variables holds in this case? Explain your answer.

4. The file `WAGE2.GDT` contains data used in Blackburn y Neumark (QJE n^o 107, pp. 1421-1436, 1992). It has information on monthly salaries, level of education, certain demographic variables, and IQ coefficients for 935 men in 1980. In order to avoid a bias due to the omission of ability we add IQ to the common equation of $\log(wage)$:
$$\begin{aligned}\log(wage) &= \beta_0 + \beta_1 educ + \beta_2\ exper + \beta_3 tenure + \beta_4 married + \beta_5 south \\ &\quad + \beta_6 urban + \beta_7 black + u\end{aligned}$$

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \; exper + \beta_3 tenure + \beta_4 married + \beta_5 south + \beta_6 urban + \beta_7 black + \beta_8 IQ + u$$

a) Estimate both equations and compare the estimated return of education. What sign does the correlation of education and omitted ability have and what is the sign of the bias due to the omission in the first equation?

b) Interpret the coefficients of *educ* and IQ. Compare a rise in wage due to 15 points more IQ to a rise due to one additional year of education.. If the true standard deviation of IQ were 15, what can be said about the relative importance of ability and of education for income?

c) Compare the R^2 of both models.

d) Based on the estimation of both models, what can be said about the differences between black and white men?

e) Let's now add the interaction between *educ* and IQ to the equation in order to check if the return on education is greater for individuals with higher IQ :

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \; exper + \beta_3 tenure + \beta_4 married + \beta_5 south + \beta_6 urban + \beta_7 black + \beta_8 IQ + \beta_9 educ \cdot IQ + u.$$

Does the data confirm this hypothesis? What happens to the individual significance of *educ* and IQ in our new equation? Which model is more appropriate?

f) Use KWW (test results of "knowledge of the world of work") as a proxy for ability instead of IQ. What is the return on education in this case?

g) Now use both variables IQ and KWW as proxy variables. What happens to the return on education?

h) Referring to the last section: Are variables IQ and KWW individually significant? Are they jointly significant?

5. The manager of a shopping mall wants to know the effect of income (R, in euros) and gender ($S = 1$ if the client is a man and 0 if she is a woman) on customer purchases (in hundreds of euros) (V). In order to study his customer's behavior he uses the following model:

$$E(V|\, R, S) = \beta_0 + \beta_1 \ln(R) + \beta_2 S$$

In order to estimate the above model, the manager asked 528 customers (chosen randomly) about their income each time they purchased something.

OLS, using observations 1–528

Dependent variable: V

	Coefficient	Std. Error	t-ratio	p-value
const	2,5210	0,0175	144,31	0,000
$\ln R$	0,0160	0,0014	11,25	0,000
S	−0,0640	0,0060	−10,61	0,000

Sum squared resid	14,6110	S.E. of regression	0,1670
R^2	0,2944	Adjusted R^2	0,2917

a) Assume that customers lie about their current income and given $\widehat{\beta}_1$and $\widehat{\beta}_2$ the OLS estimators for $\beta_1$and β_2, respectively, indicate which of the following statements are TRUE:

i) $\widehat{\beta}_1$ will be inconsistent and $\widehat{\beta}_2$will be consistent.

ii) $\widehat{\beta}_1$ will be consistent and $\widehat{\beta}_2$ will be inconsistent.

iii) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be consistent.

iv) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be inconsistent.

b) Indicate which of the following statements is TRUE:

i) The larger the variance of the measurement error in respect to the variance of true income, the larger the inconsistency bias of $\widehat{\beta}_1$.

ii) The larger the variance of the measurement error in respect to the variance of true income, the smaller the inconsistency bias of $\widehat{\beta}_1$.

iii) The inconsistency bias of $\widehat{\beta}_1$ does not depend on the relative variances of the measurement error and the one od true income.

iv) $\widehat{\beta}_1$ is consistent whenever the model is estimated by OLS.

c) Assume that the measurement error is a fixed fraction of income in the sense that, if the true income R^*, the relationship with the observed income can be written as $R = \delta R^*$, with $0 < \delta < 1$. Then:

i) $\widehat{\beta}_1$ will be inconsistent and $\widehat{\beta}_2$ will be consistent.

ii) $\widehat{\beta}_1$ will be consistent and $\widehat{\beta}_2$ will be inconsistent.

iii) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be consistent.

iv) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be inconsistent.

d) Assume that income is measured without any error but that purchases (V). are reported with an error. In this case, if the covariance of the measurement error and income is zero and in addition the expected mean of the measurement error is also equal to zero:

i) $\widehat{\beta}_1$ will be inconsistent and $\widehat{\beta}_2$ will be consistent.

ii) $\widehat{\beta}_1$ will be consistent and $\widehat{\beta}_2$ will be inconsistent.

iii) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be consistent.

iv) In general, $\widehat{\beta}_1$ as well as $\widehat{\beta}_2$ will be inconsistent.

e) In case there is no measurement error in income nor in documented purchases, what is the mean difference between purchases by men and women of the same income level?

i) 252 euros.

ii) 6,4 euros less for men.

iii) There is no significant difference between purchases by men and purchases by women.

iv) 6,4 euros more for men.