UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: **ECON3150 – Introductory Econometrics**

Date of exam: Wednesday, August 8, 2018

Time for exam: 09:00 a.m. – 12:00 noon

The problem set covers 6 pages (incl. cover sheet)

Resources allowed:

- All written and printed resources – as well as two alternative calculators - are allowed

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Exam ECON3150/4150: Introductory Econometrics.

This is an open book examination where all printed and written resources, in addition to a calculator, are allowed. If you are asked to derive something, give all intermediate steps. Do not answer questions with a "yes" or "no" only, but carefully motivate your answer.

In this exercise, we will use a data set collected from a sample of US individuals. The data are described at the end of the exercises.

Start by considering the bivariate OLS-regression that relates individual's wages to the number of years of tenure in the current employment relationship

$$\texttt{ln\_wage}_i = \alpha + \beta \texttt{tenure}_i + \epsilon_i.$$

1. (10 points) Figure 1 at the end of the exercise provides a scatter plot of `ln_wage` against `tenure`.
   (a) Explain in words how OLS finds the regression line.
   (b) We say that the OLS-estimator is *'unbiased'* and *'consistent'*. Explain the difference between these two concepts.
   (c) The Gauss-Markov theorem tells us that the OLS-estimator is also *'efficient'*. Explain what is meant by this and what assumptions are necessary for it to be true.

2. (10 points) Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the mean of the variable X. Show that the OLS-estimators $\hat{\beta}$ and $\hat{\alpha}$ of β and α in the regression above are

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left(\texttt{ln\_wage}_i - \overline{\texttt{ln\_wage}}\right)\texttt{tenure}_i}{\sum_{i=1}^{n}\left(\texttt{tenure}_i - \overline{\texttt{tenure}}\right)^2}$$
$$\hat{\alpha} = \overline{\texttt{ln\_wage}} - \hat{\beta}\cdot\overline{\texttt{tenure}}$$

   (Hint: Consider the first-order conditions of the OLS objective function and note that $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\bar{X} = \bar{X}\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = 0$*. You may use the simplified notation* $Y = \texttt{ln\_wage}$ *and* $X = \texttt{tenure}$ *in your derivations.)*

3. (10 points) Using the output below, calculate the OLS-estimate of β and α. *(Hint: Note that* $\sum_{i=1}^{n} X_i = n\cdot\bar{X}$*.)*

```
. gen tenure_sq = tenure*tenure

. gen ln_wage_sq = ln_wage*ln_wage

. gen tenure_lwage = tenure*ln_wage

. sum ln_wage tenure tenure_sq ln_wage_sq tenure_lwage
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ln_wage	2,231	1.872672	.573017	.0049396	3.707372
tenure	2,231	5.97785	5.510331	0	25.91667
tenure_sq	2,231	66.08483	102.5389	0	671.6736
ln_wage_sq	2,231	3.835101	2.34625	.0000244	13.74461
tenure_lwage	2,231	12.1415	12.62416	0	75.7233

```
. reg ln_wage tenure
```

Source	SS	df	MS		
Model	65.9164523	1	65.9164523	Number of obs	= 2,231
Residual	666.300616	2,229	.29892356	F(1, 2229)	= 220.51
				Prob > F	= 0.0000
				R-squared	= 0.0900
				Adj R-squared	= 0.0896
Total	732.217068	2,230	.328348461	Root MSE	= .54674

ln_wage	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
tenure	[redacted]	.0021011	[redacted]	0.000	[redacted]	[redacted]
_cons		.0170805		0.000		

4. (5 points) Say that we omit the constant term from the OLS regression.
   (a) Make a rough sketch of the scatter plot in Figure 1 and include an approximation to the OLS fitted line through the scatter plot with and without a constant term.
   (b) Consider that in the true model, $\alpha = 0$, but that you included the constant term in the regression. How would you expect this to affect your estimates?

5. (5 points)
   (a) Explain what we mean by heteroskedasticity.
   (b) Considering Figure 1, would you be concerned about heteroskedasticity in this case? Why or why not?
   (c) How would heteroskedasticity affect the OLS-estimates of our model? How would you account for this in your estimation?

6. We may be concerned that the bivariate model is inappropriate. Consider the extended multivariate regression

$$\texttt{ln\_wage}_\texttt{i} = \alpha + \beta\texttt{tenure}_\texttt{i} + \gamma_1\texttt{BlueCollar}_i + \gamma_2\texttt{WhiteCollar}_i + \epsilon_i$$

where `BlueCollar` and `WhiteCollar` are mutually exclusive dummy variables equal to one if the individual's occupation is classified as blue collar and white collar, respectively. The estimation output from this regression is included below.

```
. gen WhiteCollar = (occupation == 1)

. gen Managerial = (occupation == 2)

. gen BlueCollar = (occupation == 3)

. reg ln_wage tenure BlueCollar WhiteCollar
```

Source	SS	df	MS	Number of obs	=	2,231
				F(3, 2227)	=	127.78
Model	107.530402	3	35.8434674	Prob > F	=	0.0000
Residual	624.686666	2,227	.280505912	R-squared	=	0.1469
				Adj R-squared	=	0.1457
Total	732.217068	2,230	.328348461	Root MSE	=	.52963

ln_wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tenure	.0299802	.0020387	14.71	0.000	.0259822	.0339782
BlueCollar	-.1886312	.0260202	-7.25	0.000	-.2396575	-.1376048
WhiteCollar	-.3471735	.0288962	-12.01	0.000	-.4038399	-.2905072
_cons	1.852852	.0225853	82.04	0.000	1.808561	1.897142

(a) (5 points) Give an interpretation of the estimate on `tenure`.

(b) (5 points) Calculate a 90 percent confidence interval for β. Give an interpretation of this interval.

(c) (5 points) Test the hypothesis $H_0 : \beta = 0.05$ on the 5 percent-level.

(d) (5 points) Assuming that residuals are homoskedastic, test the hypothesis $\gamma_1 = \gamma_2 = 0$. *(Hint: Output from the previous regression will be necessary for this test.)*

(e) (5 points) Draw a sketch of the estimated regression lines for the three occupation groups: Managers, blue collar workers and white collar workers. Give an interpretation of γ_1 and γ_2. Discuss whether these estimates seem reasonable or if you believe that there may be some important omitted variables or that the functional form may be misspecified.

(f) (5 points) Suppose that you included the variable `Managerial` *instead of* the variable `BlueCollar`? What would be the values of the coefficients in this regression?

(g) (5 points) What would happen to your estimates if you included the variable `Managerial` *in addition to* the variables `BlueCollar` and `WhiteCollar`?

7. (5 points each) Discuss whether each of the following statements is correct or not. Note that these *do not* relate to the regression model we studied above.

(a) With municipality fixed effects in the regression, we cannot include the distance from the municipality to the closest city in our regression model.

(b) The causal effect of a treatment D_i is given by the difference between the observed outcome of the treated ($D_i = 1$) and the observed outcome of the untreated ($D_i = 0$) .

(c) In a study of the impact of education on wages, the education of parents is a good instrument for the education of their child.

(d) Excluding a covariate that explains the outcome will cause estimates on all included covariates to be biased.

Additional material

```
. describe

Contains data from /Applications/Stata/ado/base/n/nlsw88.dta
  obs:         2,231                          NLSW, 1988 extract
 vars:             4                          1 May 2016 22:52
 size:        35,696                          (_dta has notes)
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
wage            float   %9.0g                 hourly wage
tenure          float   %9.0g                 job tenure (years)
ln_wage         float   %9.0g                 ln(wage)
occupation      long    %12.0g     occupation
-------------------------------------------------------------------------------
Sorted by:
     Note: Dataset has changed since last saved.

. label list _all
occupation:
           1 Blue collar
           2 Managerial
           3 White collar

. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |      2,231    7.792448    5.764505   1.004952   40.74659
      tenure |      2,231     5.97785    5.510331          0   25.91667
     ln_wage |      2,231    1.872672     .573017   .0049396   3.707372
  occupation |      2,231    2.120574    .7851096          1          3
```

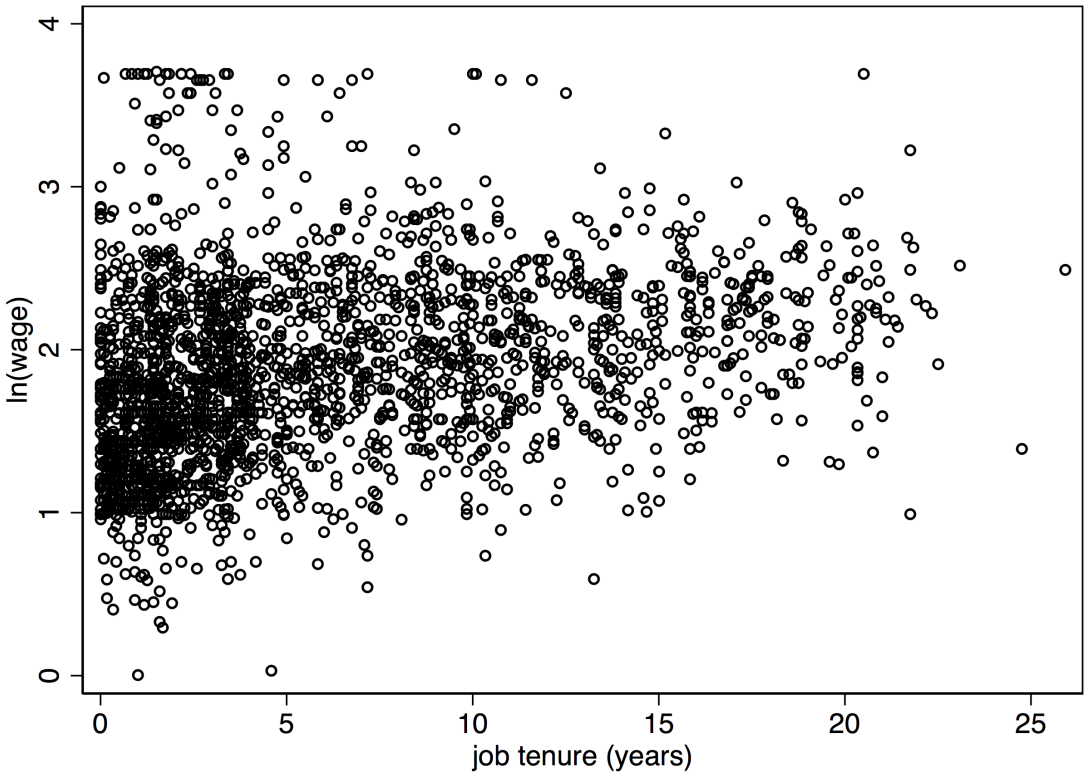

Figure 1: Log hourly wages and tenure