ECON 3150/4150: INTRODUCTORY ECONOMETRICS

PROBLEM SET, EXAM SPRING 2012

Sensorveiledning/assessment guidance in italics

Question 1

In order to analyze how the wage rate of an individual depends on age, education, work experience, a measure of intellectual ability etc., a set of observations from 935, rather young, persons (U.S.A., 1980) has been collected for the variables:

```
wage       monthly earnings, U.S. dollars
lwage      log(wage)
age        age in years
educ       no. of years of education
exper      no. of years of work experience
iq         Intelligence Quotient (IQ) score
married    = 1 if married, = 0 if unmarried
educlong   = 1 if educ > 12, = 0 if educ < or = 12
```

All the following results are from slightly edited Stata printouts.

The means, standard deviations, min and max values are:

Variable	Mean	Std. Dev.	Min	Max
lwage	6.779004	0.42114	4.7449	8.0320
wage	957.9455	404.3608	115	3078
age	33.08021	3.10780	28	38
educ	13.46845	2.19665	9	18
exper	11.56364	4.37458	1	23
iq	101.2824	15.0526	50	145
married	0.893048	0.30921	0	1
educlong	0.4855615	0.50006	0	1

The correlation coefficients of the variables are

	lwage	wage	age	educ	exper	iq	married	educlong
lwage	1.0000							
wage	0.9531	1.0000						
age	0.1618	0.1567	1.0000					
educ	0.3121	0.3271	-0.0123	1.0000				
exper	0.0206	0.0022	0.4953	-0.4556	1.0000			
iq	0.3148	0.3091	-0.0437	0.5157	-0.2249	1.0000		
married	0.1500	0.1366	0.1070	-0.0586	0.1063	-0.0147	1.0000	
educlong	0.2730	0.2839	-0.0795	0.8278	-0.4551	0.4631	-0.0377	1.0000

(1.a): Regression 1-1 below shows the results of an OLS regression of `lwage` on `age, educ, exper, iq`.

(i) Give the interpretation of the coefficient of `educ`, construct a 95 % confidence interval for this coefficient and explain precisely the interpretation of this interval.

(ii) What is the implied estimated effect on the wage rate of increasing the education period by one year while reducing the work experience period by one year?

(i) Conf.interval: $(0.0510 \pm 1.96 \times 0.0076)$. *(ii) 0.0510-0.0132=0.0378, approx. 3.8 % wage increase.*

REGRESSION 1-1

Source	SS	df	MS
Model	28.1848336	4	7.0462084
Residual	137.471461	930	0.147818775
Total	165.656294	934	0.177362199

Number of obs	= 935
F(4, 930)	= 47.67
Prob > F	= 0.0000
R-squared	= 0.1701
Root MSE	= 0.38447

lwage	Coef.	Std. Err.	t	P>|t|
age	0.0144485	0.0048571	2.97	0.003
educ	0.0509697	0.0076026	6.70	0.000
exper	0.0131736	0.0038725	3.40	0.001
iq	0.0059629	0.0009774	6.10	0.000
_cons	4.858288	0.1664242	29.19	0.000

(1.b): Regressions 1-2 and 1-3 show similar results when, respectively, `iq` and both `iq` and `educ` have been omitted as explanatory variables. In Regression 1-2 the estimated coefficient of `educ` is increased substantially. In Regression 1-3, the estimated coefficient of `exper` changes from being positive to being negative. The R^2 (`R-squared`) for Regression 1-3 is less than 1/5 of that for Regression 1-1. State briefly your interpretation of these findings. You may well perform an F-test as part of your answer.

R1-1 versus R1-2: Omitted variable, iq, positively correlated with included, educ. R1-2 versus R1-3: Omitted variable, educ, negatively correlated with included, exper. The candidate should refer to the relevant elements in the correlation matrix. Both the omitted variables problem and F-testing of composite hypotheses have been discussed rather thoroughly in the lectures and the seminars. See e.g. Lecture note no. 8 (EB) on the web site.

REGRESSION 1-2

Source	SS	df	MS
Model	22.6827547	3	7.56091825
Residual	142.97354	931	0.15356986
Total	165.656294	934	0.177362199

Number of obs	= 935
F(3, 931)	= 49.23
Prob > F	= 0.0000
R-squared	= 0.1369
Root MSE	= 0.39188

lwage	Coef.	Std. Err.	t	P>|t|
age	0.0126416	0.0049414	2.56	0.011
educ	0.0729655	0.0068223	10.70	0.000
exper	0.0142265	0.0039432	3.61	0.000
_cons	5.213573	0.158907	32.81	0.000

REGRESSION 1-3

Source	SS	df	MS
Model	5.11650552	2	2.55825276
Residual	160.539789	932	0.172252992
Total	165.656294	934	0.177362199

Number of obs	= 935
F(2, 932)	= 14.85
Prob > F	= 0.0000
R-squared	= 0.0309
Root MSE	= 0.41503

lwage	Coef.	Std. Err.	t	P>|t|
age	0.0272259	0.0050302	5.41	0.000
exper	-0.0075973	0.0035736	-2.13	0.034
_cons	5.966217	0.1508931	39.54	0.000

(1.c): The result of an OLS estimation where the (level of the) wage rate, `wage`, is regressed on `educlong` and `married`, as well as their product, is given in Regression 1-4. Are the following hypotheses supported by these results:

(i) A married person gets a higher wage than an unmarried person?

(ii) The 'wage premium' obtained by taking a long education depends on whether the person is married or unmarried?

State briefly the reason for your answers.

*(i) Coefficient of married=0 rejected with t-test, p-value=1.1%. (ii) Interaction not significant, p-value=34%. The interpretation of the interaction dummy, educlong*married, should, preferably, be explained.*

REGRESSION 1-4

Source	SS	df	MS
Model	15753893.9	3	5251297.95
Residual	136962274	931	147113.077
Total	152716168	934	163507.675

Number of obs	= 935
F(3, 931)	= 35.70
Prob > F	= 0.0000
R-squared	= 0.1032
Root MSE	= 383.55

wage	Coef.	Std. Err.	t	P>|t|
educlong	165.1465	76.9573	2.15	0.032
married	151.746	59.46681	2.55	0.011
educlong*married	77.08907	81.41499	0.95	0.344
_cons	709.2609	56.55184	12.54	0.000

Question 2:

The variances/covariances of `lwage, iq, educ` and their correlation coefficients are given in the following matrices:

COVARIANCE MATRIX:

	lwage	iq	educ
lwage	0.177362		
iq	1.99554	226.582	
educ	0.288741	17.0517	4.82529

CORRELATION MATRIX:

	lwage	iq	educ
lwage	1.0000		
iq	0.3148	1.0000	
educ	0.3121	0.5157	1.0000

The following simplified wage equation is specified

$$\texttt{lwage} = \alpha + \beta\,\texttt{educ} + \epsilon,$$

where we believe that `educ` is correlated with the disturbance ϵ because unspecified variables affecting the wage rate can be expected to influence both `educ` and ϵ. It has been proposed to use `iq` as an instrument (instrumental variable) for `educ` in this equation. State briefly the properties which `iq` should then satisfy, and compute the estimate of β obtained by using this procedure from the data above.

(i) The very simplest case of IV estimation has been discussed in one lecture and one seminar, see e.g. Lecture note no. 9 (EB) on the web site. Discussion related to IV strength could not be expected. The IV iq should, to be valid, be correlated with educ and uncorrelated with ϵ. Good candidates should point out that what one is concerned with, is theoretical correlations. (ii) $\beta^{IV} = 1.996/17.052 = 0.1171$.

Question 3:

Let Y and X be observable stochastic variables. Assume that three unknown parameters with economic interpretations, β, α, and ϕ, can be shown to satisfy the following relations:

$$\text{(1)} \qquad \mathsf{E}(Y) = e^{\beta}\mathsf{E}(X),$$

$$\text{(2)} \qquad \mathsf{E}[\ln(Y)] = \alpha\mathsf{E}[\ln(X)],$$

$$\text{(3)} \qquad \text{var}\left(\frac{Y}{X}\right) = \phi^4.$$

(3.a): Assume that you have N observations on (Y, X): $(Y_1, X_1), \ldots, (Y_N, X_N)$. Explain which empirical moments you would compute and how you, by utilizing these statistics in combination with equations (1)–(3), would estimate, respectively, β, α and ϕ by using the method of moments. **Hint:** It may be convenient to introduce specific symbols for transformed variables, e.g., $Z = \ln(X)$, $W = Y/X$, etc.

Application of the simple MM for simple non-linear equations has been discussed in one lecture and one seminar exercise. $\widehat{\beta} = \ln(\bar{Y}) - \ln(\bar{X})$. $\widehat{\alpha} = \overline{\ln(Y)}/\overline{\ln(X)}$. $\widehat{\phi} = \sqrt{S_Z}$, *where* $S_Z =$ *empirical standard deviation of* $Z = Y/X$.

(3.b): What can you say about the properties of the estimators for β and α that you have proposed? **Hint:** Use Slutsky's Theorem.

Maybe a somewhat difficult question: Good candidates could be expected to invoke Slutsky's theorem and that the plims of first or second order moments are equal to corresponding theoretical moments. See also Lecture note no. 4 (EB).

Question 4:

Macro-economists are often interested in equations explaining a country's imports and exports. In this question we want to examine how the size of Norway's exports is related to its Gross National Product (GNP). We have annual time series data from the National Accounts for the years 1966–1997 (32 observations). The variables of interest are:

```
YEAR       Calendar year
GNP        GNP, mill. 1990 NOK
EX         Export, mill 1990 NOK
lnGNP      log(GNP)
lnEX       log(EX)
DlnGNP     First difference of lnGNP, approximately the growth rate of GNP
DlnEX      First difference of lnEX,  approximately the growth rate of EX
```

The correlation coefficients of the logarithmically transformed variables and `YEAR` are given in the following tables:

```
CORRELATION COEFFICIENTS:
             |     lnEX    lnGNP     YEAR
-------------+---------------------------
        lnEX |   1.0000
       lnGNP |   0.9886   1.0000
        YEAR |   0.9968   0.9931   1.0000
-----------------------------------------
```

```
CORRELATION COEFFICIENTS:
             |    DlnEX   DlnGNP     YEAR
-------------+---------------------------
       DlnEX |   1.0000
      DlnGNP |   0.3509   1.0000
        YEAR |   0.0737  -0.1927   1.0000
-----------------------------------------
```

Printouts from Stata for three regressions, Regressions 4-1, 4-2 and 4-3, and the Durbin-Watson (DW) statistics for the two former are given below.

(4.a): The R^2 (`R-squared`) when the equation is estimated from logarithmic variables in levels (Regression 4-1) is much higher than when it is estimated from corresponding variables in differences (Regression 4-2). What are your comments to this?

Both variables are trending, and differencing removes log-linear trends.

(4.b): Could you interpret the finding that the estimated coefficient of `DlnGNP` is much larger when the intercept is set to zero (Regression 4-3) than when it is unrestricted (Regression 4-2)?

Maybe a somewhat difficult question: Origo regression of differenced equation mimics estimation of level equation, but 'forces' the inherent positive trend in the export equation to be zero.

(4.c): We want, by using the table of critical values for the Durbin-Watson (DW) test at the end of the problem set, to test whether the hypothesis of no autocorrelation (serial correlation) of the disturbances in the export equation can be rejected against the alternative that autocorrelation exists.

(i) Formulate your null hypothesis and perform this test when using the results for Regression 4-1.

(ii) Repeat the test for Regression 4-2, and state briefly your conclusion.

DW-test has been discussed, rather briefly, in the lectures. See also Lecture note no. 14 (EB). $H_0 : \rho = 0$. *(i) Rejection of non-autocorrelation. (ii) Non-rejection when the equation is transformed to differences (and intercept included!).*

REGRESSION 4-1

Number of obs = 32
R-squared = 0.9773
Root MSE = 0.07481
Durbin-Watson d-statistic(2,32) = 0.2106419

lnEX	Coef.	Std. Err.	t	P>|t|
lnGNP	1.476115	0.0410327	35.97	0.000
_cons	-7.395485	0.5431687	-13.62	0.000

REGRESSION 4-2

Number of obs = 32
R-squared = 0.1231
Root MSE = 0.03167
Durbin-Watson d-statistic(2,32) = 1.746448

DlnEX	Coef.	Std. Err.	t	P>|t|
DlnGNP	0.6981948	0.340209	2.05	0.049
_cons	0.0292024	0.0135301	2.16	0.039

REGRESSION 4-3

Number of obs = 32
Root MSE = 0.03348

DlnEX	Coef.	Std. Err.	t	P>|t|
DlnGNP	1.36668	0.1488356	9.18	0.000

Durbin-Watson (DW) 5 % Critical Values (dL=lower, dU=upper):

T = No. of observations. K = No. of coefficients (incl. intercept)

T	K	dL	dU	T	K	dL	dU
27	2	1.31568	1.46878	31	2	1.36298	1.49574
27	3	1.23991	1.55620	31	3	1.29685	1.57011
27	4	1.16239	1.65101	31	4	1.22915	1.65002
27	5	1.08364	1.75274	31	5	1.16021	1.73518
27	6	1.00421	1.86079	31	6	1.09040	1.82522
27	7	0.92463	1.97449	31	7	1.02008	1.91976
27	8	0.84546	2.09313	31	8	0.94962	2.01834
27	9	0.76726	2.21588	31	9	0.87940	2.12046
28	2	1.32844	1.47589	32	2	1.37340	1.50190
28	3	1.25534	1.55964	32	3	1.30932	1.57358
28	4	1.18051	1.65025	32	4	1.24371	1.65046
28	5	1.10444	1.74728	32	5	1.17688	1.73226
28	6	1.02762	1.85022	32	6	1.10916	1.81867
28	7	0.95052	1.95851	32	7	1.04088	1.90931
28	8	0.87366	2.07148	32	8	0.97239	2.00381
28	9	0.79754	2.18844	32	9	0.90401	2.10171
29	2	1.34054	1.48275	33	2	1.38335	1.50784
29	3	1.26992	1.56312	33	3	1.32119	1.57703
29	4	1.19762	1.64987	33	4	1.25756	1.65110
29	5	1.12407	1.74260	33	5	1.19272	1.72978
29	6	1.04971	1.84088	33	6	1.12698	1.81282
29	7	0.97499	1.94420	33	7	1.06065	1.89986
29	8	0.90036	2.05196	33	8	0.99402	1.99057
29	9	0.82626	2.16358	33	9	0.92743	2.08455
30	2	1.35204	1.48936	34	2	1.39285	1.51358
30	3	1.28373	1.56661	34	3	1.33251	1.58045
30	4	1.21380	1.64981	34	4	1.27074	1.65189
30	5	1.14262	1.73860	34	5	1.20779	1.72770
30	6	1.07060	1.83259	34	6	1.14393	1.80758
30	7	0.99815	1.93133	34	7	1.07944	1.89129
30	8	0.92564	2.03432	34	8	1.01462	1.97849
30	9	0.85351	2.14102	34	9	0.94973	2.06882