ECON 3150/4150: INTRODUCTORY ECONOMETRICS

PROBLEM SET, EXAM SPRING 2011

Sensorveiledning/Assessment Guidance in italics

We want to analyze how household consumption and savings depends on income and its composition, and to examine some hypotheses about the form of the relationship. We have observations from 50 households on the variables (data from the U.S.A., measured in 10000 Dollars):

`c` = Annual consumption
`s` = Annual saving
`r` = Annual income
`rmean` = Mean income over the last 10 years
`rexc` = Excess income, defined as Annual income minus Mean income

The observations are computed so that they satisfy exactly the relationships

```
r = c + s
r = rmean + rexc
```

All the tables below are from slightly edited Stata printouts.

Question 1: The means, standard deviations, min and max values are:

Variable	Obs	Mean	Std.Dev.	Min	Max
c	50	7.1667	1.3049	4.6246	10.2885
s	50	0.3951	0.1003	0.1812	0.6117
r	50	7.5618	1.2944	4.9197	10.6234
rmean	50	6.7393	0.5063	5.6080	8.0904
rexc	50	0.8225	1.0440	-1.1402	2.5330

The correlation coefficients of the variables are

	c	s	r	rmean	rexc
c	1.0000				
s	-0.1432	1.0000			
r	0.9971	-0.0669	1.0000		
rmean	0.6121	0.3246	0.6422	1.0000	
rexc	0.9393	-0.2404	0.9283	0.3112	1.0000

(1.a): The standard deviations of `r` and `rexc` are much larger than that of `rmean`. Can you explain why this is reasonable?

Volatility in `r` *is smoothed when taking means. Probably the* `r`*'s in different years are positively correlated, so we can expect st.dev of* `rmean` *to exceed st.dev of* `r`$/\sqrt{50}$.

(1.b): Regressing consumption on income `r` and regressing saving on income `r`, we get the result in the tables below. The estimated marginal propensities to consume and to save (i.e. the two slope coefficients) add to 1 exactly. Can you, by utilizing that `c+s=r`, show that this result follows from the application of the OLS method. (Hint: Formulate the underlying regression models.) The two estimated equations can therefore be said to 'tell the same story', as the estimated intercepts are equal with opposite signs, and the `Root MSE`s and the `Std.Err.`s are equal.

The estimates add to one because $M[c, r]/M[r, r] + M[s, r]/M[r, r] = M[r, r]/M[r, r] = 1$ *(*$M[\cdot, \cdot]$ *denoting empirical moments).*

(1.c): The R^2 (`R-squared`), however, differ widely. Could you explain this finding?

R^2 is related to the LHS variable of the regression, so the ratios between explained and total variation for the consumption function and the savings function differ, although Sum of squared residuals (SSR) is the same.

```
Number of obs =      50          ------------------------------------------------------
F(  1,    48) = 8109.03                     c |      Coef.   Std. Err.       t    P>|t|
Prob > F      =  0.0000          -------------+----------------------------------------
R-squared     =  0.9941                     r |   1.005185   .0111625    90.05   0.000
Root MSE      =  .10114                 _cons |   -.434277   .0856116    -5.07   0.000
                                 ------------------------------------------------------
*************************************************************************************
Number of obs =      50          ------------------------------------------------------
F(  1,    48) =    0.22                     s |      Coef.   Std. Err.       t    P>|t|
Prob > F      =  0.6444          -------------+----------------------------------------
R-squared     =  0.0045                     r |  -.005185    .0111625    -0.46   0.644
Root MSE      =  .10114                 _cons |   .434277    .0856116     5.07   0.000
                                 ------------------------------------------------------
************************************************************************************
```

Question 2:

The composition of income `r` into 'regular income' and 'irregular income', may, according to some economic theory, be expected to affect consumption. Let us, for simplicity, in our data set, associate the former component with `rmean` and the latter with `rexc`. An estimation result based on this idea is given below.

```
Number of obs =      50        ------------------------------------------------------
F(  2,    47) = 5169.44                   c |      Coef.   Std. Err.       t    P>|t|
Prob > F      =  0.0000        -------------+----------------------------------------
R-squared     =  0.9955               rmean |   .9123745   .0266101    34.29   0.000
Root MSE      =  .08963                rexc |   1.036325   .0129056    80.30   0.000
                                      _cons |   .1655914   .1767737     0.94   0.354
                               ------------------------------------------------------
*********************************************************************************
```

(2.a): Construct a 95 % confidence interval for the marginal propensity to consume of `rexc`. Explain precisely, in words, the interpretation of this interval.

Confidence interval is: $1.036 \pm 2.01 \cdot 0.013 = 1.036 \pm 0.026$.

(2.b): Test the null hypothesis that the marginal propensity to consume of `rmean` is one, against the hypothesis that it is less than one and formulate your conclusion.

$T = (1 - 0.912)/0.027 = -0.088/0.027 = -3.26$. *Reject.*

(2.c): Test the null hypothesis that the marginal propensities to consume of `rmean` and `rexc` are equal against the hypothesis that they are different. The estimated covariance of the two coefficient estimates is -0.00010688 (the estimate of the corresponding correlation coefficient is -0.3112)

Coefficient difference $= 0.912 - 1.036 = -0.124$.
Standard error of coefficient difference $= [0.0266^2 - 2(-0.000107) + 0.0129^2]^{1/2} = 0.033$.
$t = .0.124/0.033 = -3.76$. *Reject equality of the two marginal propensities to consume.*

Question 3: Another theory maintains that the marginal propensity to consume may depend on the income level. To examine this hypothesis we use the simple income measure `r` and extend the set of regressors by including quadratic and cubic terms, `r2=`r^2 and `r3=`r^3. The results from two regressions are given in the following two tables, supplemented by the correlation matrix of the three regressors at the bottom:

```
Number of obs =      50   ------------------------------------------------------
F(  2,    47) = 4056.76              c |      Coef.   Std. Err.      t    P>|t|
Prob > F      =  0.0000   -------------+----------------------------------------
R-squared     =  0.9942              r |   .8927904   .1118173     7.98   0.000
Root MSE      =  .10112             r2 |   .0074335   .0073583     1.01   0.318
                                 _cons |  -.0216255   .4173522    -0.05   0.959
                          ------------------------------------------------------
*********************************************************************************

Number of obs =      50   ------------------------------------------------------
F(  3,    46) = 2660.08              c |      Coef.   Std. Err.      t    P>|t|
Prob > F      =  0.0000   -------------+----------------------------------------
R-squared     =  0.9943              r |   .5303655    .769732     0.69   0.494
Root MSE      =  .10196             r2 |   .0553329   .1009065     0.55   0.586
                                    r3 |  -.0020628   .0043337    -0.48   0.636
                                 _cons |   .8701224   1.920182     0.45   0.653
                          ------------------------------------------------------
**********************************************************************************

                          |        r       r2       r3
             -------------+---------------------------
                        r |   1.0000
                       r2 |   0.9950   1.0000
                       r3 |   0.9803   0.9951   1.0000
             *****************************************
```

(3.a): Does the hypothesis that the marginal propensity to consume is income dependent find support in the data? State the reason for your answer.

No: Neither of `r2, r3` *come out with significant coefficient estimates. The joint hypothesis that both are zero could also be tested from an F-test using the two reported R^2s, 0.9942 and 0.9943.*

(3.b): In the second regression, with both the quadratic and the cubic terms included as regressors, all regressors come out with small t-values. Does this mean that income 'drops out as an explanatory variable' for consumption? Try, as part of your answer, to interpret the results `F(3,46) = 2660.08` and `Prob > F = 0.0000` in the printout and explain what the arguments 3 and 46 mean.

No: The joint hypothesis that `r, r2, r3` *all have coefficients =0 is rejected by F-test:*
`F(3,46) = 2660.08` *and p-value* = 0.0000.

Question 4: It has been objected against the above analysis that the disturbance variance may not be constant (heteroskedasticity). To examine this we go back to the model in **Question 1** and let `ehat2`= $\hat{e}_i^2$ be the squared OLS residual from the first estimation. Regressing `ehat2` on `r` gives the results in the first table below. Repeating the analysis, after having first transformed `r` and `c` to logarithms, denoted as `logr` and `logc`, respectively, gives the results in the second table below. `eloghat2` is the corresponding squared OLS residual, which is regressed on `logr`.

(4.a): Which conclusion can you draw from this?

There is no sign of significant heteroskedasticity of disturbance in first regression: For regressor `r` *in first auxiliary regression:* $t = -1.19$. *Log transformation induces significant heteroskedasticity in second regression: For regressor* `logr` *in second auxiliary regression:* $t = -2.63$. *Conclusion: Log-transformation 'over-adjusts' in this case: Log-transformation of a LHS variable may be viewed as a way of modeling multiplicative errors in the equation.*

(4.b): If there is sign of heteroskedasticity, would this destroy the property that the OLS estimators of the regression coefficients are unbiased? Explain briefly.

Under heteroskedasticity, OLS is still unbiased, but the standard error estimates are biased, as wrong formulae are used in its calculation.

```
------------------------------------------------------
       ehat2 |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------
           r |   -.001635   .0013795    -1.19   0.242
       _cons |   .0221834   .0105804     2.10   0.041
------------------------------------------------------
******************************************************

------------------------------------------------------
    eloghat2 |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------
        logr |  -.0006372   .0002425    -2.63   0.011
       _cons |   .0015007   .0004887     3.07   0.004
------------------------------------------------------
******************************************************
```