Assignment 4

DUE DATE: Tuesday 6^{th}, November 2018 before 15.30 pm

1. Individual Wages(50 points)

Using a sample of 545 full-time workers in the USA, a researcher is interested in the question whether women are systematically underpaid compared to men. First, she estimates the average hourly wages in the sample for men and women, which are \$5.91 and \$5.09, respectively.

a. Do these numbers give an answer to the question of interest? Why not? How could one (at least partially) correct for this?

The researcher also runs a simple regression of an individual's wage on a male dummy, equal to 1 for males and 0 for females. This gives the results reported in Table 2.7.

Table 2.7 Hourly wages explained from gender: OLS results

Variable	Estimate	Standard error	*t*-ratio
constant	5.09	0.58	8.78
male	0.82	0.15	5.47

$N = 545 \quad s = 2.17 \quad R^2 = 0.26$

b. How can you interpret the coefficient estimate of 0.82? How do you interpret the estimated intercept of 5.09?

c. How do you interpret the R^2 of 0.26?

d. Explain the relationship between the coefficient estimates in the table and the average wage rates of males and females.

e. A student is unhappy with this model as 'a female dummy is omitted from the model'. Comment upon this criticism.

f. Test, using the above results, the hypothesis that men and women have, on average, the same wage rate, against the *one-sided* alternative that women earn less. State the assumptions required for this test to be valid.

g. Construct a 95% confidence interval for the average wage differential between males and females in the population.

Subsequently, the above 'model' is extended to include differences in age and education, by including the variables *age* (age in years) and *educ* (education level, from 1 to 5). Simultaneously, the endogenous variable is adjusted to be the *natural logarithm* of the hourly wage rate. The results are reported in Table 2.8.

h. How do you interpret the coefficients of 0.13 for the male dummy, and 0.09 for age?

i. Test the joint hypothesis that gender, age and education do not affect a person's wage.

j. A student is unhappy with this model as 'the effect of education is rather restrictive'. Can you explain this criticism? How could the model be extended or changed to meet the above criticism? How can you test whether the extension has been useful?

The researcher re-estimates the above model including age^2 as an additional regressor. The t-value on this new variable becomes -1.14, while $R^2 = 0.699$ and $\bar{R}^2$ increases to 0.683.

k. Could you give a reason why the inclusion of age^2 might be appropriate?

l. Would you retain this new variable given the R^2 and the $\bar{R}^2$ measures? Would you retain age^2 given its t-value? Explain this apparent conflict in conclusions.

Table 2.8 Log hourly wages explained from gender, age and education level: OLS results

Variable	Coefficient	Standard error	t-ratio
constant	-1.09	0.38	2.88
male	0.13	0.03	4.47
age	0.09	0.02	4.38
educ	0.18	0.05	3.66

$N = 545 \quad s = 0.24 \quad R^2 = 0.691 \quad \bar{R}^2 = 0.682$

2. Asset Pricing-Empirical (50 points)

In the recent finance literature it is suggested that asset prices are fairly well described by a so-called factor model, where excess returns are linearly explained from excess returns on a number of 'factor portfolios'. As in the CAPM, the intercept term should be zero, just like the coefficient for any other variable included in the model the value of which is known in advance (e.g. a January dummy). The data set ASSETS2 contains excess returns on four factor portfolios for January 1960 to December 2010 [26]

rmrf:	excess return on a value-weighted market proxy
smb:	return on a small-stock portfolio minus the return on a large-stock portfolio (Small minus Big)
hml:	return on a value-stock portfolio minus the return on a growth-stock portfolio (High minus Low)
umd:	return on a high prior return portfolio minus the return on a low prior return portfolio (Up minus Down)

All data are for the USA. Each of the last three variables denotes the difference in returns on two hypothetical portfolios of stocks. These portfolios are re-formed each month on the basis of the most recent available information on firm size, book-to-market value of equity and historical returns, respectively. The *hml* factor is based on the ratio of book value to market value of equity, and reflects the difference in returns between a portfolio of stocks with a high book-to-market ratio (value stocks) and a portfolio of stocks with a low book-to-market ratio (growth stocks). The factors are motivated by empirically found anomalies of the CAPM (for example, small firms appear to have higher returns than large ones, even after the CAPM risk correction).

In addition to the excess returns on these four factors, we have observations on the returns on ten different 'assets' which are ten portfolios of stocks, maintained by the Center for Research in Security Prices (CRSP). These portfolios are size-based, which means that portfolio 1 contains the 10% smallest firms listed at the New York Stock Exchange and portfolio 10 contains the 10% largest firms that are listed. Excess returns (in excess of the riskfree rate) on these portfolios are denoted by $r1$ to $r10$, respectively.

In answering the following questions use $r1$, $r10$ and the returns on two additional portfolios that you select.

a. Regress the excess returns on your four portfolios upon the excess return on the market portfolio (proxy), noting that this corresponds to the CAPM. Include a constant in these regressions.

b. Give an economic interpretation of the estimated β coefficients.

c. Give an economic and a statistical interpretation of the R^2s.

d. Test the hypothesis that $\beta_j = 1$ for each of the four portfolios. State the assumptions you need to make for the tests to be (asymptotically) valid.

e. Test the validity of the CAPM by testing whether the constant terms in the four regressions are zero.

f. Test for a January effect in each of the four regressions.

g. Next, estimate the four-factor model

$$r_{jt} = \alpha_j + \beta_{j1} rmrf_t + \beta_{j2} smb_t + \beta_{j3} hml_t + \beta_{j4} umd_t + \varepsilon_{jt}$$

by OLS. Compare the estimation results with those obtained from the one-factor (CAPM) model. Pay attention to the estimated partial slope coefficients and the R^2s.

h. Perform F-tests for the hypothesis that the coefficients for the three new factors are jointly equal to zero.

i. Test the validity of the four-factor model by testing whether the constant terms in the four regressions are zero. Compare your conclusions with those obtained from the CAPM.