Practice Final

1. In a study of salary discrimination in a company, data are collected on the monthly salary and the characteristics of the employees. A regression model is specified with monthly salary $Y$ as the dependent variable and gender as an independent variable. Beside gender the only other independent variable is whether the employee does clerical work or not. Hence the independent variables in the regression model are $X_1$, a dummy variable that is 1 if the employee is female and 0 otherwise, and $X_2$, a dummy variable that is 1 if the employee does clerical work and 0 otherwise. The regression model also has an intercept.

   a. It is suggested that the treatment of male and female employees in the regression model is asymmetrical. To remedy this the researcher includes a third dummy variable $X_3$ that is 1 if the employee is male and 0 otherwise. The researcher does not succeed in estimating the model that also contains $X_3$. Why?

   b. The researcher estimates the regression model with an intercept and independent variables $X_1$ and $X_2$. An F-test of the hypothesis that the regression coefficients of both these variables are 0 is performed. The hypothesis is soundly rejected. The t-test of the hypothesis that the coefficient of $X_1$ is 0 is not rejected, as is the t-test of the hypothesis that the coefficient of $X_2$ is 0. Can you explain this result?

   c. Does the fact that in the model that only contains the dummy for clerical work $X_2$, the coefficient of the gender dummy is not significantly different from 0 means that there is no evidence of differential treatment of men and women in this company?

2. Consider the linear regression model with dependent variable log income $Y$ and independent variable years of education $X_1$. The regression model also has an intercept. The data is a cross-section of individuals in the US.

   a. The OLS estimate of the regression coefficient of $X_1$ is .016. What is the interpretation of this number (note that the dependent variable is log income)? This regression coefficient of years of education $X_1$ is often referred to as the return to education.

   b. Define two dummy variables $X_2, X_3$ that together for each individual indicate in which industry he or she works. Why do you need only two dummy variables?

   c. It is hypothesized that the return to education differs by industry. Specify a linear regression model that allows you to test this hypothesis. What is the null hypothesis for this test? Which test statistic would you use and how would you compute it?

3. Let $Y$ be aggregate investment (in 100 billions $) and let $X_1$ be GDP (in 100 billion $) and $X_2$ the interest rate (%). Using annual time-series data for 1970-1999, the OLS estimates (standard errors) are
\[ Y = 3.23 + 0.089 X_{1t} + 0.0099 X_{2t} \]

\[ R^2 = 0.89 \]

The Durbin-Watson statistic is \( d = 1.089 \).

a. Give the linear regression model with an intercept and independent variables \( X_{1t}, X_{2t} \) and with random errors that are AR(1).

b. Test the null hypothesis of no serial correlation against the AR(1) model with positive serial correlation. What does the outcome of the test imply for the omitted variables in the regression model?

c. An alternative to the Durbin-Watson test is the LM test. Describe the steps of this test for the linear regression model of this problem.

d. What is the implication of the result of b. for the OLS estimates, their standard errors and the \( R^2 \)?

e. Explain why the result in b. suggests to reformulate the regression model by including \( Y_{t-1} \) and \( X_{1t-1}, X_{2t-1} \) as additional independent variables in the regression model. Do the regression coefficients in this reformulated regression model correspond to the regression coefficients in the model in a.? What is the relation?