ESSAY. Write your answer in the space provided.

1) Sir Francis Galton, a cousin of James Darwin, examined the relationship between the height of children and their parents towards the end of the 19th century. It is from this study that the name "regression" originated. You decide to update his findings by collecting data from 110 college students, and estimate the following relationship:

\[ \text{Student}_{\text{h}} = 19.6 + 0.73 \times \text{Midparh}, \quad R^2 = 0.45, \quad \text{SER} = 2.0 \]

\[ (7.2) (0.10) \]

where \( \text{Student}_{\text{h}} \) is the height of students in inches, and \( \text{Midparh} \) is the average of the parental heights. Values in parentheses are heteroskedasticity robust standard errors. (Following Galton's methodology, both variables were adjusted so that the average female height was equal to the average male height.)

(a) Interpret the estimated coefficients.

(b) If children, on average, were expected to be of the same height as their parents, then this would imply two hypotheses, one for the slope and one for the intercept.

(i) What should the null hypothesis be for the intercept? Calculate the relevant \( t \)-statistic and carry out the hypothesis test at the 1% level.
(ii) What should the null hypothesis be for the slope? Calculate the relevant $t$-statistic and carry out the hypothesis test at the 5% level.

c) What is the prediction for the height of a child whose parents have an average height of 70.06 inches?

d) Construct a 95% confidence interval for a one inch increase in the average of parental height.

(e) Given the positive intercept and the fact that the slope lies between zero and one, what can you say about the height of students who have quite tall parents? Who have quite short parents?

(f) Galton was concerned about the height of the English aristocracy and referred to the above result as "regression towards mediocrity." Can you figure out what his concern was? Why do you think that we refer to this result today as "Galton's Fallacy?"
2) Consider the following model:

\[ X_t = \beta_1 X_{t-1} + u_t. \]

Derive the OLS estimator for \( \beta_1 \).

3) The neoclassical growth model predicts that for identical savings rates and population growth rates, countries should converge to the per capita income level. This is referred to as the convergence hypothesis. One way to test for the presence of convergence is to compare the growth rates over time to the initial starting level.

(a) If you regressed the average growth rate over a time period (1960–1990) on the initial level of per capita income, what would the sign of the slope have to be to indicate this type of convergence? Explain. Would this result confirm or reject the prediction of the neoclassical growth model?
(b) The results of the regression for 104 countries were as follows:

\[
\hat{g}_{6090} = 0.019 - 0.0006 \times \text{RelProd}_{60}, \quad R^2 = 0.00007, \quad \text{SER} = 0.016,
\]

\[
(0.004) \quad (0.0073)
\]

where \(g_{6090}\) is the average annual growth rate of GDP per worker for the 1960–1990 sample period, and \(\text{RelProd}_{60}\) is GDP per worker relative to the United States in 1960. Interpret the results. Is there any evidence of unconditional convergence between the countries of the world? Is this result surprising? What other concept could you think about to test for convergence between countries?

(c) Using the OLS estimator with homoskedasticity-only standard errors, the results changed as follows:

\[
\hat{g}_{6090} = 0.019 - 0.0006 \times \text{RelProd}_{60}, \quad R^2 = 0.00007, \quad \text{SER} = 0.016
\]

\[
(0.002) \quad (0.0068)
\]

Why didn’t the estimated coefficients change? Given that the standard error of the slope is now smaller, can you reject the null hypothesis of no beta convergence? Are the results in (c) more reliable than the results in (b)? Explain.

(d) You decide to restrict yourself to the 24 OECD countries in the sample. This changes your regression output as follows:
\( \bar{g}_{6090} = 0.048 - 0.0404 \, \text{RelProd}_{60}, \, R^2 = 0.82, \, \text{SER} = 0.0046 \) 
(0.004) (0.0063)

How does this result affect your conclusions from above? When you test for convergence, should you worry about the relatively small sample size?
4) The effect of decreasing the student-teacher ratio by one is estimated to result in an improvement of the districtwide score by 2.28 with a standard error of 0.52. Construct a 90% and 99% confidence interval for the size of the slope coefficient and the corresponding predicted effect of changing the student-teacher ratio by one. What is the intuition on why the 99% confidence interval is wider than the 90% confidence interval?
5) Prove that the regression $R^2$ is identical to the square of the correlation coefficient between two variables $Y$ and $X$. Regression functions are written in a form that suggests causation running from $X$ to $Y$. Given your proof, does a high regression $R^2$ present supportive evidence of a causal relationship? Can you think of some regression examples where the direction of causality is not clear? Is without a doubt?
EViews Exercise [50 pts]

The data for this exercise is already in EViews form and can be found in the file “cal_scores.wf1.” A brief explanation of the contents of this file can be found in the document “cal_scores.doc.” Make sure to become familiar with the description of the data, how it was collected and any possible issues that may help you interpret your regression output. These data are used throughout chapters 4-7. Here, we will replicate some of the examples in the book. This will help you check that you are doing things correctly. I am more interested in your interpretation of the results than in your ability to copy output in EViews onto a Word document.

1. Display the scatter plot of the average test score variable against the student-teacher ratio. Be sure to display the graph that includes the regression line.

2. Estimate the OLS regression of average test score on student-teacher ratio. Make sure that you are able to replicate the estimate results reported in page 99-101 in your book.

3. Estimate the OLS regression of average reading score on student-teacher ratio and then another regression where the dependent variable is the math score instead. Compare these results with those obtained in part 2 above and comment on their differences and similarities. In particular, given that the U.S. ranks low in math but high in reading, what do the regression results tell you about the relative merits of reducing class size in each field?

4. Going back to the original regression in part 2, do a hypothesis test on the slope coefficient to determine whether class size has any impact on test scores. Repeat on the regressions in part 3. Should this be a one-sided or a two-sided type of test? Be sure to do the hypothesis test that makes the most common sense. For all three regressions, calculate a 95% confidence level and comment on how much do the confidence intervals overlap with each other. What does this tell you, if anything? Note: be sure to do all these calculations using heteroskedasticity robust standard errors.

5. Subsample your data according to whether or not there are English learners in the school and repeat the regression in part 2 for each of these two groups (i.e., 0% English learners versus > 0% English learners in school). Comment on the results of this exercise – what do the results tell you from a policy point of view?