Question 1. Let us consider the following cross-sectional model for household consumption C

$$C = \beta_0 + \beta_1 Y + \beta_2 Y^2 + u, \qquad \textbf{(1)}$$

where Y denotes household income both measured in british pounds £, $E(u) = 0$ and $E(u^2) = \sigma^2$. We will assume that the error is independent of Y (and therefore Y^2 as well). You are provided with a random sample $\{(C_i, Y_i)\}_{i=1}^n$.

(a) (13 marks) □ Explain the economic meaning of including a quadratic term in a regression equation and discuss the expected values of the regression coefficients. What are the advantages and disadvantages of model (1) compared to the corresponding linear model?

Solution:
For majority of commodities it is natural to assume that there will be a diminishing effect of income on consumption which would require both $\beta_1 > 0$ and $\beta_2 < 0$.
However, for certain categories of goods, such as luxury goods, positive value of β_2 may be observed.
In any case, the value β_2 is rather small in absolute value.
Quadratic regression has advantage to capture important economic property – diminishing effect of income (if it really takes place) that increases quality of regression (a significant increase in R-squared, the significance of the coefficients).
On the other hand in some cases correlation between Y and Y^2 may lead to multicollinearity.

□ How can you test that the dependence is really quadratic and cannot be satisfactorily approximated using a linear regression of the form $C = \beta_0 + \beta_1 Y + u$? Discuss the possibility of using t and F tests.

Solution:

Standard t-test for β_2 is applicable here $t = \dfrac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)}$ to test pair of hypothesis $H_0 : \beta_2 = 0$, $H_a : \beta_2 < 0$ taking into account considerations above.

Alternatively F-test can be conducted: run (1) to get R_U^2, then run restricted version of (1) with $\beta_2 = 0$ to get R_R^2 and then apply $F(1, n-3) = \dfrac{(R_U^2 - R_R^2)}{1 - R_U^2} / (n-3)$. But using F test it is impossible to test $H_0 : \beta_2 = 0$ against $H_a : \beta_2 < 0$. So t test is preferrable here.

□ Obtain the expression for the change in consumption as response to 1 pound increase in income using both direct calculation according to the regression equation (1) and the marginal income effect calculation. Comment.

Solution:
Direct calculation

$$\Delta C = \beta_0 + \beta_1 (Y+1) + \beta_2 (Y+1)^2 - \beta_0 - \beta_1 Y - \beta_2 Y^2 = \beta_1 + 2\beta_2 Y + \beta_2$$

Using marginal effect

$$\frac{dC}{dY} = (\beta_0 + \beta_1 Y + \beta_2 Y^2)' = \beta_1 + 2\beta_2 Y,$$

so here effect is rather overestimated (we assume $\beta_2 < 0$).

Since β_2 is rather small for meaningful reasons, both methods give approximately the same results.

□ What will happen to the OLS parameter estimates when we measure consumption and income in £1,000 instead of £? Explain your answer.

Solution:

The parameter β_1 remains unchanged, but β_0 and β_2 need to be rescaled. Let $C^* = \dfrac{C}{1{,}000}$ and $Y^* = \dfrac{Y}{1{,}000}$ denote our new variables. Let's substitute them:

$$1{,}000C^* = \beta_0 + \beta_1 1{,}000Y^* + \beta_2(1{,}000Y^*)^2 + u,$$

$$C^* = \frac{\beta_0}{1{,}000} + \beta_1 Y^* + 1{,}000\beta_2 (Y^*)^2 + u.$$

The intercept will be divided by 1,000 whereas β_2 will be multiplied by 1,000.

(b) (12 marks) □ Let the model (1) represent the true relationship between consumption and income. Discuss the effect of omitting the quadratic term on the OLS estimator of β_1. Will the bias be upward or downward? Explain your answer, clearly indicating necessary assumptions.

Solution:

If we estimate:

$$C = \beta_0 + \beta_1 Y + e, \text{ (short equation)}$$

while correct equation is $C = \beta_0 + \beta_1 Y + \beta_2 Y^2 + u$

we will get the classical omitted variavle bias problem.

$$\hat{\beta}_1 = \frac{\text{Cov}(Y,C)}{\text{Var}(Y)} = \frac{\text{Cov}(Y;\beta_0 + \beta_1 Y + \beta_2 Y^2 + u)}{\text{Var}(Y)} = \beta_1 \frac{\text{Cov}(Y,Y)}{\text{Var}(Y)} + \beta_2 \frac{\text{Cov}(Y,Y^2)}{\text{Var}(Y)} + \frac{\text{Cov}(Y,u)}{\text{Var}(Y)}$$

and (assuming Y_i not stochastic and using $\text{E}(u) = 0$):

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\text{Cov}(Y,Y^2)}{\text{Var}(Y)} + \frac{\text{Cov}(Y,Eu)}{\text{Var}(Y)} = \beta_1 + \beta_2 \frac{\text{Cov}(Y,Y^2)}{\text{Var}(Y)}.$$

As we expect that $\text{Cov}(Y_i, Y_i^2) > 0$ and in typical case $\beta_2 < 0$; the bias will be negative.

Solution:

If we estimate:

$$C = \beta_0 + \beta_1 Y = e, \text{ (short equation)}$$

we will get the classical OVB problem. With:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(C_i - \overline{C})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(Y_i^2 - \overline{Y}^2)}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} + \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(u_i - \overline{u})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

and (assuming Y_i is not stochastic and using $\mathrm{E}(u) = 0$):

$$E(\hat{\beta}_1) = \beta_1 + \underbrace{\beta_2 \frac{\mathrm{SampleCov}(Y_i, Y_i^2)}{\mathrm{SampleVar}(Y_i)}}_{\mathrm{OVB}}.$$

As we expect that $\mathrm{SampleCov}(Y_i, Y_i^2) > 0$ and $\beta_2 < 0$; the bias will be negative.

□ How this bias will affect the estimate of the constant β_0? (*you may use a graphic illustration with a brief commentary*).

Solution:

Since the graph is located in the first quarter (average income and average consumption are both positive), and the quadratic function is convex upwards in the assumption of diminishing effect, then omittig of quadratic term leads to an increase of intercept.