1) Earnings functions, whereby the log of earnings is regressed on years of education, years of on-the-job training, and individual characteristics, have been studied for a variety of reasons. Some studies have focused on the returns to education, others on discrimination, union and non-union differentials, etc. For all these studies, a major concern has been the fact that ability should enter as a determinant of earnings, but that it is close to impossible to measure and therefore represents an omitted variable.

Assume that the coefficient on years of education is the parameter of interest. Given that education is positively correlated to ability, since, for example, more able students attract scholarships and hence receive more years of education, the OLS estimator for the returns to education could be upward-biased. To overcome this problem, various authors have used instrumental variables estimation techniques. For each of the instruments potential instruments listed below, briefly discuss instrument validity.

(a) The individual’s postal zip code.

(b) The individual’s IQ or test-score on a work-related exam.

(c) Years of education for the individual’s mother or father.

(d) Number of siblings the individual has.
2) The figure shows a plot and a fitted linear regression line of the age-earnings profile of 1,744 individuals, taken from the Current Population Survey.

(a) Describe the problems in predicting earnings using the fitted line. What would the pattern of the residuals look like for the age category under 40?

(b) What alternative functional form might fit the data better?

(c) What other variables might you want to consider in specifying the determinants of earnings?

3) Suggest a transformation in the variables that will linearize the deterministic part of the population regression functions below. Write the resulting regression function in a form that can be estimated by using OLS.

(a) \( Y_i = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} \)

(b) \( Y_i = \frac{X_i}{\beta_0 + \beta_1 X_i} \)

(c) \( Y_i = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \)

(d) \( Y_i = \beta_0 X_i^{\beta_1 e^{\beta_2 X_i}} \)
4) Your textbook gives the following example of simultaneous causality bias of a two equation system:

\[
\begin{align*}
Y_i &= \beta_0 + \beta_1 X_i + u_i \\
X_i &= \gamma_0 + \gamma_1 Y_i + v_i
\end{align*}
\]

In microeconomics, you studied the demand and supply of goods in a single market. Let the demand \( (Q^D_i) \) and supply \( (Q^S_i) \) for the \( i \)-th good be determined as follows,

\[
\begin{align*}
Q^D_i &= \beta_0 - \beta_1 P_i + u_i' \\
Q^S_i &= \gamma_0 + \gamma_1 P_i + v_i'
\end{align*}
\]

where \( P \) is the price of the good. In addition, you typically assume that the market clears.

Explain how the simultaneous causality bias applies in this situation. The textbook explained a positive correlation between \( X_i \) and \( u_i \) for \( \gamma_1 > 0 \) through an argument that started from “imagine that \( u_i \) is negative.” Repeat this exercise here.