8.3 In a similar industry, firms relate their *intended* stocks of finished goods, Y^*, to their expected annual sales, X^e, according to a linear relationship

$$Y^* = \beta_1 + \beta_2 X^e.$$

Actual sales, X, differ from expected sales by a random quantity u, which is distributed with zero mean and constant variance:

$$X = X^e + u$$

where u is distributed independently of X^e. Since unexpected sales lead to a reduction in stocks, actual stocks are given by

$$Y = Y^* - u.$$

An investigator has data on Y and X (but not on Y^* or X^e) for a cross-section of firms in the industry. Describe analytically the problems that would be encountered if OLS were used to estimate β_1 and β_2, regressing Y on X. [*Note:* You are warned that the standard expression for measurement error bias is not valid in this case.]

Answer: This model is slightly more complex than the basic errors-in-variables model since the measurement error in the (measured) dependent variable is correlated with the (measured) explanatory variable.

Given the definitions,

$$Y + u = \beta_1 + \beta_2(X - u)$$

so

$$Y = \beta_1 + \beta_2 X - (1 + \beta_2)\, u = \beta_1 + \beta_2 X + v$$

where

$$v = -(1 + \beta_2)u$$

If you use OLS to fit the equation, $b_2 = \dfrac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2}$ can be decomposed as

$$b_2 = \beta_2 + \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(v_i - \overline{v}\right)}{\sum\left(X_i - \overline{X}\right)^2} = \beta_2 + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(v_i - \overline{v}\right)}{\frac{1}{n}\sum\left(X_i - \overline{X}\right)^2}$$

The numerator and denominator have both been divided by i to assure that they have probability limits. Hence

$$\operatorname{plim} b_2 = \beta_2 + \frac{\operatorname{cov}(X, v)}{\operatorname{var}(X)} = \beta_2 + \frac{\operatorname{cov}\left([X^e + u], -(1+\beta_2)u\right)}{\sigma_X^2}$$

$$= \beta_2 - \frac{(1+\beta_2)\sigma_u^2}{\sigma_X^2}$$

assuming there is some variation in X. Hence, in large samples, the slope estimate is downward biased.

The intercept is calculated as:

$$\beta_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

which, as the sample goes to infinity, becomes:

$$\beta_1 = \mu_y - (\text{plim}\, \hat{\beta}_2)\mu_x$$

since plim $\hat{\beta}_2$ does not equal β_2, the intercept is also estimated inconsistently.