Name of the Semester: II (M.Com)
Name of the subject: Managerial Economics
Paper: CC.202
Name of the teacher: Arindam Jana
Lecture No.: 1

Module- I

Demand Estimation and Forecasting

Demand Estimation

Estimating the future demand for products, either existing or new is a significant aspect of demand analysis. A manager needs to have information about likely future demand of its product to enable the firm to produce the required quantities of a product at the right time and arrange well in advance for the various inputs (like labor, raw material, machines etc.) as well as to pursue optimal pricing strategies. Demand estimation and forecasting means predicting future demand for the product under given conditions and helped the manager in making decisions with regard to production, sales, investment, expansion, employment of manpower etc., both in the short run as well as in the long run.

Consumer Interviews

To obtain information concerning the demand function for a particular product, firms frequently interview consumers and administer questionnaires concerning their buying habits, motives, and intentions. Firms may also run focus groups in an attempt to discern consumers' tastes. For example, a firm might ask a random sample of consumers how much more gasoline they would purchase if its price were reduced by 5 percent. Or, a market researcher might ask a sample of consumers whether they liked a new type of perfume better than a leading existing brand, and if so, how much more they would be willing to pay for it (than for the existing brand).

Unfortunately, consumer surveys of this sort have many well-known limitations. The direct approach of simply asking people how much they would buy of a particular commodity at particular prices often does not seem to work very well. Frequently, the answers provided by consumers to such a hypothetical question are not very accurate. However, more subtle approaches can be useful. Interviews indicated that most buyers of a particular baby food selected it on their doctor's recommendation and that most of them knew very little about prices of substitutes. This information, together with other data, suggested that the price elasticity of demand was quite low in absolute value.

Despite the limitations of consumer interviews and questionnaires, many managers believe that such surveys can reveal a great deal about how their firms can serve the market better. For example, the Campbell Soup Company's researchers contacted close to 110,000 people to talk about the taste, preparation, and nutritional value of food. On the basis of these interviews, Campbell changed the seasonings in five Le Menu dinners and introduced a line of low-salt soups (called Special Request). Some of the factors influencing the quality of survey results can be quite subtle. For example, according to research findings, there are sometimes advantages in respondents' keypunching answers, rather than verbalizing them, because the respondents tend to answer emotional questions more honestly this way.

Market Experiments

Another method of estimating the demand curve for a particular commodity is to carry out direct market experiments. The idea is to vary the price of the product while attempting to keep other market conditions fairly stable (or to take changes in other market conditions into account). For example, a manufacturer of ink conducted an experiment some years ago to determine the price elasticity of demand for its product. It raised the price from 15 cents to 25 cents in four cities and found that demand was quite inelastic. Attempts were made to estimate the cross elasticity of demand with other brands as well.

Controlled laboratory experiments can sometimes be carried out. Consumers are given money and told to shop in a simulated store. The experimenter can vary the prices, packaging, and location of particular products, and see the effects on the consumers' purchasing decisions. While this technique is useful, it suffers from the fact that consumers participating in such an experiment know that their actions are being monitored. For that reason, their behavior may depart from what it normally would be.

Before carrying out a market experiment, weigh the costs against the benefits. Direct experimentation can be expensive or risky because customers may be lost and profits cut by the experiment. For example, if the price of a product is raised as part of an experiment, potential buyers may be driven away. Also, since they are seldom really controlled experiments and since they are often of relatively brief duration and the number of observations is small, experiments often cannot produce all the information that is needed. Nonetheless, market experiments can be of considerable value, as illustrated by the following actual case.

Regression Analysis

Although consumer interviews and direct market experiments are important sources of information concerning demand functions, they are not used as often as regression analysis. Suppose that a firm's demand function is

$$Y = A + B_1 X + B_2 P + B_3 I + B_4 P_r \quad ....(1)$$

where *Y* is the quantity demanded of the firm's product, *X* is the selling expense (such as advertising) of the firm, *P* is the price of its product, *I* is the disposable income of consumers, and P_r is the price of the competing product sold by its rival. What we want are estimates of the values of A, B_1, B_2, B_3, and B_4.

Regression analysis describes the way in which one variable is related to another. Regression analysis derives an equation that can be used to estimate the unknown value of one variable on the basis of the known value of the other variable. For example, suppose that the Miller Pharmaceutical Company is scheduled to spend

Table-1: Selling Expense and Sales, Miller Pharmaceutical Company

Selling expense (millions of dollars)	Sales (millions of units)
1	4
2	6
4	8
8	14
6	12
5	10
8	16
9	16
7	12

$4 million next year on selling expense (for promotion, advertising, and related marketing activities) and it wants to estimate its next-year's sales, on the basis of the data in Table 5.1 regarding its sales and selling expense in the previous nine years. In this case, although the firm's selling expense next year is known, its next year's sales are unknown. Regression analysis describes the way in which the firm's sales are historically related to its selling expense.

Simple Regression Model

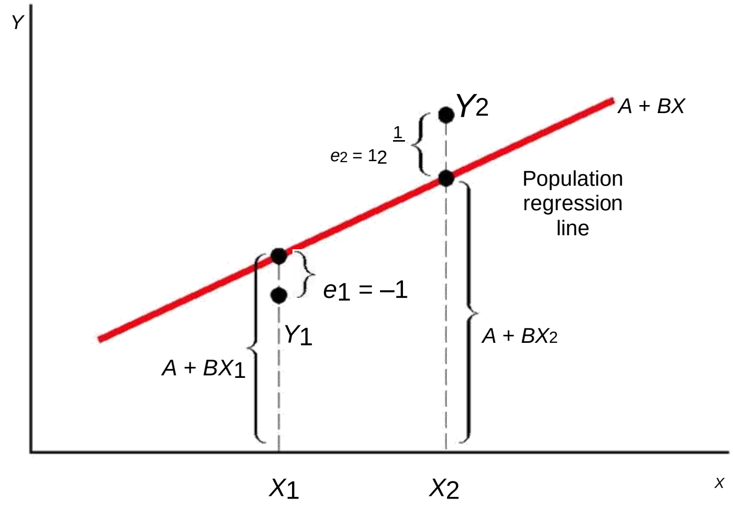
As you recall from Chapter 1, a **model** is a simplified or idealized representation of the real world. In this section, we describe the model—that is, the set of simplifying assumptions—on which regression analysis is based. We begin by visualizing a population of all relevant pairs of observations of the indepen-dent and dependent variables. For instance, in the case of the Miller Pharmaceutical Company, we visualize a population of pairs of observations concerning sales and selling expense. This population includes all the levels of sales cor-responding to all the levels of selling expense in the history of the firm.

The mean of a variable equals the sum of its values divided by their num-ber. Therefore, the mean of a variable that assumes four values, 3, 2, 1, and 0, is $(3 + 2 + 1 + 0)/4$, or 1.5. Regression analysis assumes that the mean value of Y, given the value of X, is a linear function of X. In other words, the mean value of the dependent variable is assumed to be a linear function of the inde-pendent variable, the equation of this being $A + BX$, as shown in Figure 5.4.

FIGURE 5.4

Regression Model

The mean value of *Y*, given the value of *X*, falls on the population regression line.



This straight line is called the **population regression line** or the **true regression line.**

Put differently, regression analysis assumes that

$$Y_i = A + BX_i + e_i \text{....(2)}$$

where Y_i is the ith observed value of the dependent variable and X_i is the ith observed value of the independent variable. Essentially, e_i is an **error term**, that is, a random amount that is added to $A + BX_i$ (or subtracted from it if e_i is negative). Because of the presence of this error term, the observed values of Y_i fall around the population regression line, not on it. Hence, as shown in Figure 5.4, if e_1 (the value of the error term for the first observation) is -1, Y_1 lies 1 below the population regression line. And if e_2 (the value of the error term for the second observation) is -1.50, Y_2 lies 1.50 above the population regression line. Regression analysis assumes that the values of e_i are independent and their mean value equals zero.

.

Although the assumptions underlying regression analysis are unlikely to be met completely, they are close enough to the truth in a sufficiently large num-ber of cases that regression analysis is a powerful technique. Nonetheless, it is important to recognize at the start that, if these assumptions are not at least approximately valid, the results of a regression analysis can be misleading.

Sample Regression Line

The purpose of a regression analysis is to obtain the mathematical equation for a line that describes the average relationship between the dependent and inde-pendent variables. This line is calculated from the sample observations and is called the **sample** or **estimated regression line**. It should not be confused with the population regression line discussed in the previous section. Whereas the population regression line is based on the entire population, the sample regression line is based on only the sample.

The general expression for the sample regression line is

$$\hat{Y} = a + bX$$

where *Y* is the value of the dependent variable predicted by the regression line, and *a* and *b* are estimators of *A* and *B*, respectively. (An estimator is a function of the sample observations used to estimate an unknown parameter. For example, the sample mean is an estimator often used to estimate the popula-tion mean.) The term *a* is often called the *Y* **intercept** of the regression line. And *b*, which clearly is the slope of the line, measures the change in the predicted value of *Y* associated with a one-unit increase in *X*.

Figure 5.5 shows the estimated regression line for the data concerning sales and selling expense of the Miller Pharmaceutical Company. The equation for this regression line is

$$Y = 2.536 + 1.504X$$

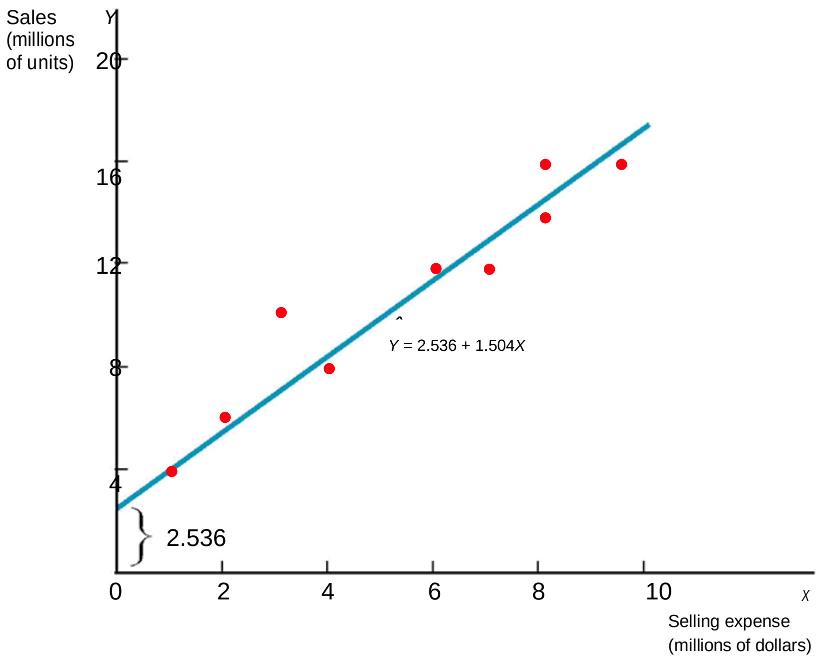
where *Y* is sales in millions of units and *X* is selling expense in millions of dollars. What is 2.536? It is the value of *a*, the estimator of *A*. What is 1.504? It is the value of *b*, the estimator of *B*. For the moment, we are not interested in how this equation was determined; what we want to consider is how it should be interpreted.

At the outset, note the difference between *Y* and *Y*. Whereas *Y* denotes an *observed* value of sales, *Y* denotes the *computed* or *estimated* value of sales,

FIGURE 5.5

Sample Regression Line

This line is an estimate of the population regression line.



based on the regression line. For example, the first row of Table -1 shows that, in the first year, the actual value of sales was 4 million units when selling expense was $1 million. Therefore, Y = 4.0 millions of units when $X = 1$. In contrast, the regression line indicates that $\hat{Y} = 2.536 + 1.504(1)$, or 4.039 mil-lions of units when $X = 1$. In other words, while the regression line predicts that sales will equal 4.039 millions of units when selling expense is $1 million, the actual sales figure under these circumstances (in the first year) was 4 mil-lion units.

It is essential to be able to identify and interpret the *Y* intercept and slope of a regression line. What is the *Y* intercept of the regression line in the case of the Miller Pharmaceutical Company? It is 2.536 millions of units. This means that, if the firm's selling expense is zero, the estimated sales would be 2.536 millions of units. (As shown in Figure 5.5, 2.536 millions of units is the value of the dependent variable at which the regression line intersects the vertical axis.) What is the slope of the regression line in this case? It is 1.504. This means that the estimated sales go up by 1.504 millions of units when selling expense increases by $1 million.

Method of Least Squares

The method used to determine the values of a and b is the so-called method of least squares. Since the deviation of the ith observed value of Y from the regression line equals $\hat{Y} - Y_i$, the sum of the squared deviations equal

$$\sum_{i=1}^{n}(Y_i - \hat{Y}) = \sum_{i=1}^{n}(Y_i - a - bX_i)^2 \ldots.(3)$$

where n is the sample size. Using the minimization technique we can find the values of a and b that minimize the expression in equation (3) by differentiating this expression with respect to a and b and set-ting these partial derivatives equal to zero:

$$\frac{\partial \sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\partial a} = -2\sum_{i=1}^{n}(Y_i - a - bX_i) = 0 \ldots.(4)$$

$$\frac{\partial \sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\partial b} = -2\sum_{i=1}^{n}X_i(Y_i - a - bX_i) = 0 \ldots.(5)$$

Solving equations (4) and .5) simultaneously and letting $\overline{X}$ equal the mean value of X in the sample and $\overline{Y}$ equal the mean value of Y, we find that

$$b = \sum (X_i - \overline{X})(Y_i - \overline{Y}) / \sum (X_i - \overline{X})^2 \ldots.(6)$$

$$b = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \ldots.(6)$$

and $a = \overline{Y} - b\overline{X} \ldots.(7)$

The value of b in equation (6) is often called the **estimated regression coefficient**.

Computation of $\sum X_i$, $\sum Y_i$, $\sum X_i^2$, $\sum Y_i^2$, and $\sum X_i Y_i$

	$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum Y_i^2$	$\sum X_i Y_i$
	1	4	1	16	4
	2	6	4	36	12
	4	8	16	64	32
	8	14	64	196	112
	6	12	36	144	72
	5	10	25	100	50
	8	16	64	256	128
	9	16	81	256	144
	7	12	49	144	84
	50	98	340	1,212	638
Total	100	196	680	2424	1276

From a computational point of view, it frequently is easier to use a somewhat different formula for *b* than the one given in equation (5.6). This alter-native formula, which yields the same answer as equation (5.6), is

$$b = \{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)\} / \{n\sum X_i^2 - (\sum X_i)^2\}$$

In the case of the Miller Pharmaceutical Company

$$b = \{9(638) - 50(98)\} / \{9(340) - (50)^2\} = 1.504$$

Therefore, the value of *b*, the least-squares estimator of *B*, is 1.504, which is the result given in the previous section. In other words, an increase in selling expense of $1 million results in an increase in estimated sales of about 1.504 millions of units.

Having calculated *b*, we can readily determine the value of *a*, the least-squares estimator of *A*. According to equation (5.7),

$$a = \overline{Y} - b\overline{X}$$

Where $\overline{Y}$ is the mean of the values of *Y*, and $\overline{X}$ is the mean of the values of *X*. Since, as shown in Table 5.2, $\overline{Y}$ = 10.889 and $\overline{X}$ =5.556, it follows that

Y a =10.889-1.504(5.556)=2.536

Therefore, the least-squares estimate of *A* is 2.536 millions of units, which is the result given in the previous section.

Having obtained *a* and *b*, it is a simple matter to specify the average relationship in the sample between sales and selling expense for the Miller Pharmaceutical Company. This relationship is

$$Y = 2.536 + 1.504X \quad ....(8)$$

Where Y is measured in millions of units and *X* is measured in millions of dollars. As we know, this line is often called the *sample regression line* or the *regression of Y on X*. It is the line presented in the previous section and plot-ted in Figure 5.5. Now, we see how this line is derived.

To illustrate how a regression line of this sort can be used, suppose that the managers of the firm want to predict the firm's sales if they decide to devote $4 million to selling expense. Using equation (5.8), they would predict that its sales would be

$$2.536+1.504(4)=8.55$$

Since sales are measured in millions of units, this means that sales would be expected to be 8.55 million units.

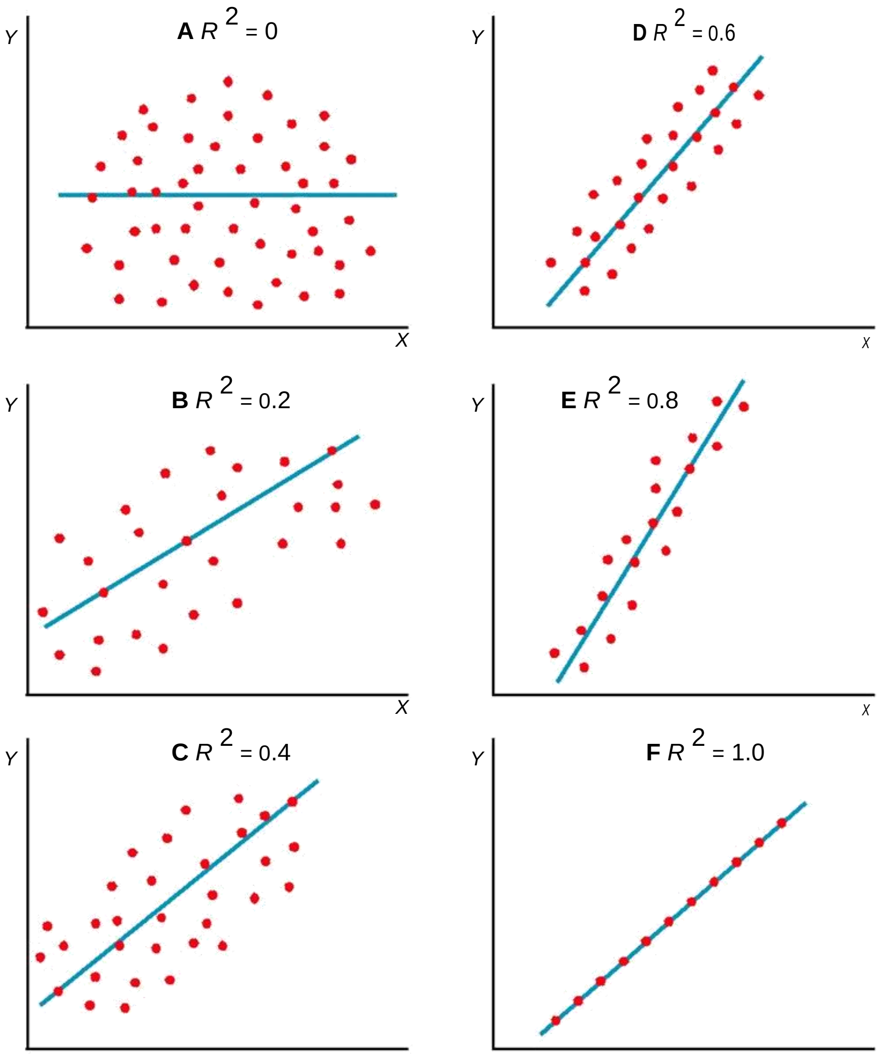
Coefficient of Determination

Once the regression line has been calculated, we want to know how well this line fits the data. There can be huge differences in how well a regression line fits a set of data, as shown in Figure 5.6. Clearly, the regression line in panel F of Figure 5.6 provides a much better fit than the regression line in panel B of the same figure. How can we measure how well a regression line fits the data?

The most commonly used measure of the goodness of fit of a regression line is the coefficient of determination. For present purposes, it is not necessary to know the formula for the coefficient of determination, because it is seldom calculated by hand. .

5.6 Six Regression Lines: Coefficient of Determination Equals 0, 0.2, 0.4, 0.6, 0.8, and 1.0

When there is only one independent variable, the coefficient of determination is often designated by r^2, rather than R^2, but computer printouts generally use R^2, regardless of the number of independent variables. We use R^2 here, even though there is only one independent variable. See footnote 8.



x x

Multiple Regression

In previous sections of this chapter, we discussed regression techniques in the case in which there is only one independent variable. In practical applications of regression techniques, it frequently is necessary and desirable to include two or more independent variables. Now, we extend the treatment of regression to the case in which there is more than one independent variable.

Whereas a **simple regression** includes only one independent variable, a **multiple regression** includes two or more independent variables. The first step in multiple regression analysis is to identify the independent variables and specify the mathematical form of the equation relating the mean value of the dependent variable to these indepen-dent variables.

In the case of the Miller Pharmaceutical Company, suppose that the firm's executives feel that its sales depend on its price, as well on its selling expense. More specifically, they assume that

$$Y_i = A + B_1 X_i + B_2 P_i + e_i \text{....(10)}$$

where X_i is the selling expense (in millions of dollars) of the firm during the ith year and P_i is the price (in dollars) of the firm's product during the ith year (measured as a deviation from \$10, the current price). Of course, B_2 is assumed to be negative. This is a different model from that in equation (2). Here, we assume that Y_i (the firm's sales in the ith year) depends on two independent variables, not one. Of course, there is no reason why more independent variables cannot be added, so long as data are available concerning their values and there is good reason to expect them to affect Y_i. But, to keep matters simple, we assume that the firm's executives believe that only selling expense and price should be included as independent variables.

$Y = 2.536 + 1.304X$

The object of multiple regression analysis is to estimate the unknown constants A, B_1, and B_2 in equation (10). Just as in the case of simple regression, these constants are estimated by finding the value of each that minimizes the sum of the squared deviations of the observed values of the dependent variable from the values of the dependent variable predicted by the regression equation. Suppose that a is an estimator of A, b_1 is an estimator of B_1, and b_2 is an estimator of B_2.

Potential Problems of Regression

1) Equation Specification

Suppose that we are fitted a linear equation, but actually the relation may be of non-linear type.

2) Multicollinearity

One important problem that can arise in multiple regression studies is **multi-collinearity**, a situation in which two or more independent variables are very highly correlated. In the case of the Miller Pharmaceutical Company, suppose that there had been a perfect linear relationship in the past between the firm's selling expense and its price. In a case of this sort, it is impossible to estimate the regression coefficients of both independent variables (X and P) because the data provide no information concerning the effect of one independent variable, holding the other independent variable constant. All that can be observed is the effect of both independent variables together, given that they both move together in the way they have in previous years.

Regression analysis estimates the effect of each independent variable by seeing how much effect this one independent variable has on the dependent variable when other independent variables are held constant. If two independent variables move together in a rigid, lockstep fashion, there is no way to tell how much effect each has separately; all we can observe is the effect of both combined. If there is good reason to believe that the independent variables will continue to move in lockstep in the future as they have in the past, multicollinearity does not prevent us from using regression analysis to predict the dependent variable. Since the two independent variables are perfectly correlated, one of them in effect stands for both and we therefore need use only one in the regression analysis. However, if the independent variables cannot be counted on to continue to move in lockstep, this procedure is dangerous, since it ignores the effect of the excluded independent variable.

In reality, you seldom encounter cases in which independent variables are perfectly correlated, but you often encounter cases in which independent variables are so highly correlated that, although it is possible to estimate the regression coefficient of each variable, these regression coefficients cannot be estimated at all accurately. To cope with such situations, it sometimes is possible to alter the independent variables in such a way as to reduce multicollinearity. Suppose that a managerial economist wants to estimate a regression equation where the quantity demanded per year of a certain good is the dependent variable and the average price of this good and disposable income of U.S. consumers are the independent variables. If disposable income is measured in money terms (that is, without adjustment for changes in the price level), there may be a high correlation between the independent variables. But if disposable income is measured in real terms (that is, with adjustment for changes in the price level), this correlation may be reduced considerably. Therefore, the managerial economist may decide to measure disposable income in real rather than money terms to reduce multicollinearity.

If techniques of this sort cannot reduce multicollinearity, there may be no alternative but to acquire new data that do not contain the high correlation among the independent variables. Whether you (or your board of directors) like it or not, there may be no way to estimate accurately the regression coefficient of a particular independent variable that is very highly correlated with some other independent variable.

1) **Serial Correlation**

In addition to multicollinearity, another important problem that can occur in regression analysis is that the error terms (the values of e_i) are not indepe dent; instead, they are serially correlated. For example, Figure-10 shows a case in which, if the error term in one period is positive, the error term in the next period is almost always positive. Similarly, if the error term in one period is negative, the error term in the next period almost always is negative. In such a situation, we say that the errors are *serially correlated* (or *autocorrelated*, which is another term for the same thing). Because this violates the assumptions underlying regression analysis, it is important that we be able to detect its occurrence.