positive autocorrelation is still apparent. A second iteration involves computing $\hat{\rho}$ from Eq. (7.16) using the residuals of the regression of model (7.15). This produces a new y_t^* and x_t^* and hence a new β_0^* and β_1^*.

7.3 TRANSFORMATIONS TO IMPROVE FIT AND PREDICTION

Transformation of data is sometimes an effective alternative that produces a better fitting, or perhaps better predicting model. Transformation or *reexpression* of data is used in order to make an assumption more reasonable. Thus far, in this chapter, we have made use of transformations to stabilize variance and to eliminate autocorrelation in regression data. Here the motivation revolves around another assumption, that of *model form.*

In the following subsections, procedures are discussed for determining *alternative model forms*, or *transformations*, and situations under which the use of these transformations might prove successful. In some cases the need for some type of transformation is obvious. Plots of data in the single regressor case may show a definite curvilinear appearance.

TRANSFORMATION IN THE CASE OF A SINGLE REGRESSOR

Consider a regression situation with a single regressor variable. A transformation to change the model structure is normally required when the data reflect curvature. A deviation from the ordinary straight line regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

may, of course, be detected from residual plots (see Section 5.2) or from a simple plotting of the data. If a plot reflects a trend that has a specific curvilinear appearance, one may be able to alter the model to accommodate the trend.

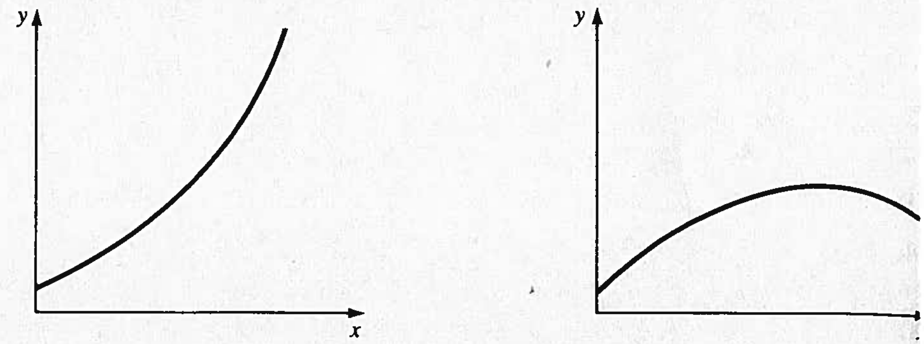
Parabola

The first model alteration we consider is more than a mere transformation but rather *the addition of a quadratic model term.* The model is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

The nature of the plot produced by data generated by the parabolic model depends on the signs and magnitudes of the coefficients β_0, β_1, and β_2. Two typical examples are given in Figure 7.3.

FIGURE 7.3 ***(a)*** **Parabola for β_0, β_1, and $\beta_2 > 0$** ***(b)*** **Parabola for $\beta_0 > 0$, $\beta_1 >$ $\beta_2 < 0$**

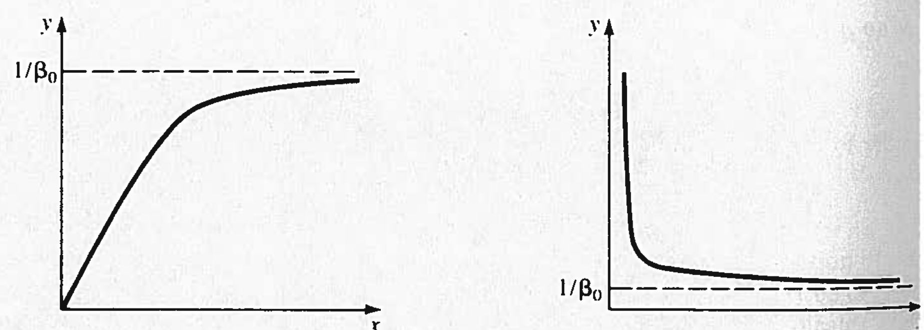


Hyperbola (Inverse Transformation on y and x)

Applications in areas of biology, economics, and certain other fields lead to the us of a hyperbolic function, which can be produced by transformations on both th response variable y and the regressor variable x. Typical plots that suggest the us of the hyperbola are given in Figure 7.4. The true functional form of the hyperbol is *nonlinear* in the model coefficients. The equation is given by $y = x/(\alpha + \beta x)$. Th linearized form involves the inverse transformation on both variables; i.e., on regresses $1/y$ against $1/x$, and thus adopts the model structure (assuming obse vations x_i, y_i, $i = 1, 2, \ldots, n$)

$$\frac{1}{y_i} = \beta_0 + \beta_1\left(\frac{1}{x_i}\right) + \varepsilon_i$$

FIGURE 7.4 ***(a)*** **Hyperbola with negative curvature** ***(b)*** **Hyperbola with positive curvature**



It is easily verified that $\beta_0 = \beta$ and $\beta_1 = \alpha$. The asymptote that may be of interest to the analyst is indicated by the dotted lines in Figure 7.4. Negative and positive curvature are, respectively, produced when $\beta_1 > 0$ and $\beta_1 < 0$.

Exponential Function; Natural Log Transformation on y

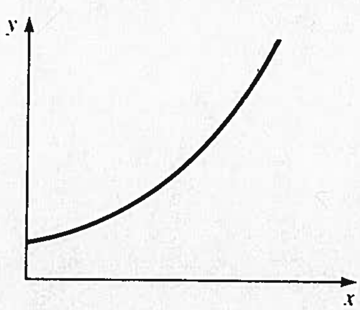
In Section 7.1 we discussed the log transformation on the response as a mechanism that may be useful for countering heterogeneous variance in certain situations. The transformation may also be useful to produce a reasonable model assumption when the data's appearance suggests curvature of a certain type. If the picture is not a straight line but, rather, looks like that depicted in either part of Figure 7.5, the true structure may be of the form $y = \alpha e^{\beta x}$. Thus a fitted model of the type

$$\ln(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$$

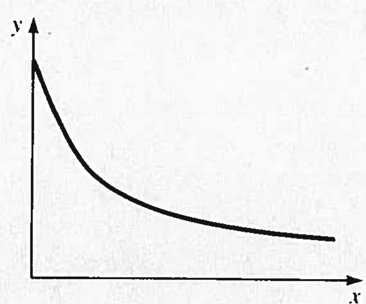
may be appropriate. Here, of course, $\beta_0 = \ln \alpha$ and $\beta_1 = \beta$. Figure 7.5(a) illustrates the situation in which $\beta > 0$. Figure 7.5(b) reveals a case in which $\beta < 0$.

FIGURE 7.5

(a) **Exponential function: use ln *y* transform**



(b) **Exponential function: use ln *y* transform**



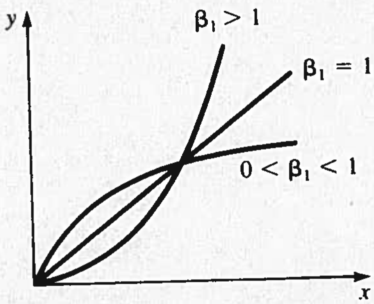
Power Functions (Natural Log Transformations on y and x)

At times, the curvature that the analyst sees in the plot of y against x suggests a mechanism of the type $y = \alpha x^{\beta}$, a power function that is clearly linearized by using the log transformation on both variables. Thus the fitted model is given by

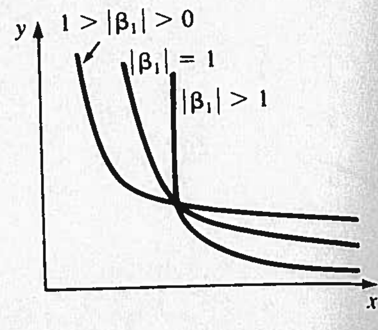
$$\ln(y_i) = \beta_0 + \beta_1(\ln x_i) + \varepsilon_i$$

with the coefficients β_0 and β_1 estimated by standard least squares procedures. The actual appearance of the power function plot depends on the sign and magnitude of the constant β_1. One set of pictures appears in Figure 7.6(a). Another set appears in Figure 7.6(b).

FIGURE 7.6



(a) Power function: use natural log transformation on y and x



(b) Power function: β_1 negative; use natural log transformation on y and x

Inverse Exponential (Natural Log Transformation on y; Inverse Transformation on x)

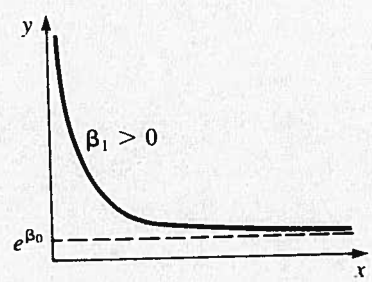
There are many scientific phenomena that are exponential in nature but do not fall into a family presented earlier. Often the exponential portion of the mechanism is proportional to the *inverse* of x instead of x; this results in pictures that do not resemble those presented in Figure 7.5. The mechanism is $y = \alpha e^{\beta/x}$. The fitted model, reflecting the transformation is given by

$$\ln(y_i) = \beta_0 + \beta_1\left(\frac{1}{x_i}\right) + \varepsilon_i$$

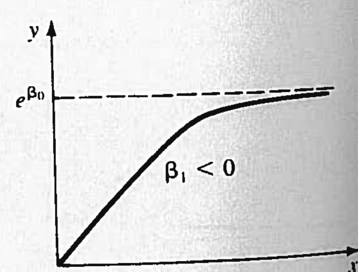
Figure 7.7 reveals the appearance of the y against x plot.

The purpose of this section is to give the user alternative models that may provide more success in fitting and predicting than that of the straight line simple

FIGURE 7.7



(a) Inverse exponential: use natural log transformation on y and inverse transformation on x



(b) Inverse exponential: use natural log transformation on y and inverse transformation on x