Thus, with stochastic regressors the formulas we have derived are valid if y and x are jointly normally distributed. Otherwise, they have to be considered as valid conditional on the observed x's.

*3.12 The Regression Fallacy

In the introduction to this chapter we mentioned a study by Galton, who analyzed the relationship between the height of children and the height of parents. Let

x = mid-parent height

y = mean height (at maturity) of all children whose mid-parent height is x

Galton plotted y against x and found that the points lay close to a straight line but the slope was less than 1.0. What this means is that if the mid-parent height is 1 inch above $\bar{x}$, the childrens' height (on the average) is less than 1 inch above $\bar{y}$. There is thus a "*regression* of childrens' heights toward the average."

A phenomenon like the one observed by Galton could arise in several situations where y and x have a bivariate normal distribution and thus is a mere statistical artifact. That is why it is termed a "regression fallacy." To see this, first we have to derive the conditional distributions $f(x|y)$ and $f(y|x)$ when x and y are jointly normal. We will show that both these conditional distributions are normal.

The Bivariate Normal Distribution

Suppose that X and Y are *jointly* normally distributed with means, variances, and covariance given by

$$E(X) = m_x \qquad E(Y) = m_y \qquad V(X) = \sigma_x^2 \qquad V(Y) = \sigma_y^2$$

and

$$\text{cov}(X, Y) = \rho\sigma_x\sigma_y$$

Then the joint density of X and Y is given by

$$f(x, y) = (2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2})^{-1} \exp(Q)$$

where

$$Q = -\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x - m_x}{\sigma_x}\right)^2 - 2\rho\frac{x - m_x}{\sigma_x}\frac{y - m_y}{\sigma_y} + \left(\frac{y - m_y}{\sigma_y}\right)^2\right]$$

Now completing the square in x and simplifying, we get

$$Q = -\frac{1}{2(1 - \rho^2)}\left(\frac{x - m_x}{\sigma_x} - \rho\frac{y - m_y}{\sigma_y}\right)^2 - \frac{1}{2}\left(\frac{y - m_y}{\sigma_y}\right)^2$$

Thus

$$f(x, y) = f(x|y)f(y)$$

where

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2\sigma_y^2}(y - m_y)^2\right]$$

and

$$f(x|y) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1 - \rho^2}} \exp\left[-\frac{1}{2\sigma_x^2(1 - \rho^2)}(x - m_{x\cdot y})^2\right]$$

where

$$m_{x\cdot y} = m_x + \frac{\rho\sigma_x}{\sigma_y}(y - m_y)$$

Thus we see that the marginal distribution of y is normal with mean m_y and variance σ_y^2. The conditional distribution of x given y is also normal with

$$\text{mean} = m_x + \frac{\rho\sigma_x}{\sigma_y}(y - m_y)$$

and

$$\text{variance} = \sigma_x^2(1 - \rho^2)$$

The conditional distribution of y given x is just obtained by interchanging x and y in the foregoing relationships.

Thus for a bivariate normal distribution, both the marginal and conditional distributions are univariate normal.[13] Note that the converse need not be true, that is, if the marginal distributions of X and Y are normal, it does not necessarily follow that the joint distribution of X and Y is bivariate normal. In fact, there are many nonnormal bivariate distributions for which the marginal distributions are both normal.[14]

Galton's Result and the Regression Fallacy

Consider now the mean of Y for given value of X. We have seen that it is given by

$$E(Y|X = x) = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x)$$

[13]This result is more general. For the multivariate normal distribution we can show that all marginal and conditional distributions are also normal.

[14]C. J. Kowalski, "Non-normal Bivariate Distributions with Normal Marginals," *The American Statistician*, Vol. 27, No. 3, June 1973, pp. 103–106; K. V. Mardia, *Families of Bivariate Distributions* (London: Charles Griffin, 1970).

The slope of this line is $\rho\sigma_y/\sigma_x$ and if $\sigma_x \simeq \sigma_y$, since $\rho < 1$ we have the result that the slope is less than 1, as observed by Galton.

By the same token, if we consider $E(X \mid Y = y)$ we get

$$E(X \mid Y = y) = m_x + \frac{\rho\sigma_x}{\sigma_y}(y - m_y)$$

Since we have assumed that $\sigma_x = \sigma_y$, the slope of this line is also less than unity (note that we are taking dx/dy as the slope in this case). Thus if Galton had considered the conditional means of parents' heights for given values of offsprings' heights, he would have found a "regression" of parents' heights toward the mean. It is not clear what Galton would have labeled this regression.

Such "regression" toward average is often found when considering variables that are jointly normally distributed and that have almost the same variance. This has been a frequent finding in the case of test scores. For example, if

$$x = \text{score on the first test}$$

and

$$y = \text{score on the second test}$$

then considering the conditional means of y for given values of x shows a regression toward the mean in the second test. This does not mean that the students' abilities are converging toward the mean. This finding in the case of test scores has been named a *regression fallacy* by the psychologist Thorndike.[15]

This, then, is the story of the term "regression." The term as it is used now has no implication that the slope be less than 1.0, nor even the implication of linearity.

Summary

1. The present chapter discusses the simple linear regression model with one explained variable and one explanatory variable. The term regression literally means "backwardation," but that is not the way it is used today, although that is the way it originated in statistics. A brief history of the term is given in Section 3.1, but it is discussed in greater detail in Section 3.12 under the title "regression fallacy."

2. Two methods—the method of moments (Section 3.3) and the method of

[15]F. L. Thorndike, "Regression Fallacies in the Matched Group Experiment," *Psychometrika*, Vol. 7, No. 2, 1942, pp. 85–102.