10

Demand Forecasting with Regression Models

I have seen the future and it is very much like the present, only longer

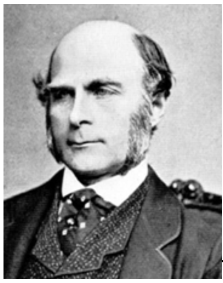
Kehlog Albran, *The Profit*

In this chapter, we introduce the concept of a linear regression model and use it for describing and creating causal models for demand forecasting purposes. You will learn about preparing data to identify a regression model, and assumptions needed to validate model relationships involving appropriate drivers of demand.

Transformations are important in the preliminary identification and diagnostic checking (residual analysis) stages of the model building process. Transformations of data should be performed to improve your understanding of what the data are trying to tell you; can make the assumptions of least-squares theory with normality assumptions valid, can have significant value outside the realm of model building, too, that is, in the evaluation and interpretation of data; and can also be applied effectively in presenting management with the results of a forecasting analysis.

The diagnostic checking process, designed to reveal departures from underlying assumptions in a statistical forecasting model, is an important phase for demand forecasters to learn about. It can be a powerful visual tool for assessing the potential effectiveness of a forecasting model, isolating unusual values, identifying hidden patterns, and understanding the nature of randomness in historical data.

What Are Regression Models?



The term regression has a rather curious origin in studies of inheritance in biology, which showed that whereas tall (or short) fathers had tall (or short) sons, the sons were *on average* not as tall (or as short) as their fathers. This phenomenon was discovered by Sir Francis Galton, (1822-1911), and is called *regression toward the mean*. Thus, Galton observed that the average height of the sons tended to move toward the average height of the overall population of fathers rather than toward reproducing the height of the parents. It has since been observed in a wide spectrum of settings from economic behavior and athletic performance to demand forecasting. Regression analysis is the principal method of causal modeling in demand forecasting.

Demand forecasters begin a regression analysis by identifying the factors or drivers of demand (Chapter 1), called **independent**, causal or explanatory variables – that they believe have influenced and will continue to influence the variable to be forecast (the **dependent** variable).

It is useful to categorize causal variables as internal or external. **Internal variables**, also called policy variables, can be controlled to a substantial degree by managerial decisions, and their future values can be set as a matter of company policy. Examples include product prices, promotion outlays, and methods of distribution. **External** or environmental variables are those whose level and influence are outside organizational control. Included here may be variables that measure weather and holidays, demographics such as the age and gender of consumers in the market area, decisions made by competing enterprises, and the state of the macro economy as measured by rates of economic growth, inflation, and unemployment. Normally a regression analysis will attempt to account for both internal and external causal variables.

The forecaster's beliefs about the way a dependent variable responds to changes in each of the independent variables is expressed as an equation, or series of equations, called a regression model. A regression model also includes an explicit expression of an error variable to describe the role of chance or underlying uncertainty.

A model with a single independent variable is called a simple regression model. Multiple regression refers to a model with one dependent and two or more independent variables.

A successful regression analysis provides useful estimates of how previous changes in each of the independent variables have affected the dependent variable. In addition, assuming that the underlying structure is stable, forecasts of the dependent variable can then be conditioned on assumptions or projections of the future behavior of the independent variables.

For example, suppose that a regression analysis of the demand for a product or service indicates that price increases in the past, holding other things constant, have been associated with less than proportional reductions in sales volumes (i.e., demand has been price inelastic – ***see* Chapter 11**). This knowledge may be useful both for forecasting future demand and for adjusting product-pricing policy.

The inelastic demand suggests that price increases might be improving profitability. Demand forecasts would then be made in light of the price changes that the company plans to institute.

Similar feedback can be obtained through regression analysis of the influence of external economic variables such as the Gross Domestic Product (GDP). Although the firm cannot control the rates of economic growth, projections of GDP growth can be translated via regression analysis into forecasts of product sales growth.

As a forecasting approach, regression analysis has the potential to provide not only demand forecasts of the dependent variable but useful managerial information for adapting to the forces and events that cause the dependent variable to change.

Indeed, a regression analysis may be motivated as much or more by the need for policy information as by the interest in demand forecasting. It is important to note that no extrapolative forecasting method can supply policy information, such as how product sales respond to price and macroeconomic variables. When such information is desired, explanations are required, not merely extrapolations. The chapter-opening quotation suggests that Einstein would have preferred the empirical approach to demand forecasting, which analyzes the data ("experience") in terms of a model ("knowledge of reality") that relates the dependent and independent variables.

The Regression Curve

A regression curve can be used to describe a relationship between a variable of interest and one or more related variables that are assumed to have a bearing on the demand forecasting problem. If data are plentiful, a curve passing through the bulk of the data represents the regression curve. The data are such that there is no functional relationship describing exactly one variable Y as a function of X; for a given value of the independent variable X, there is a *distribution* of values of Y. This relationship may be approximated by determining the average (or median) value of Y for small intervals of values of X.

A regression curve (in a two-variable case) is defined as that "typical" curve that goes through the mean value of the dependent variable Y for each fixed value of the independent variable X.

In many practical situations, there are not enough values to "even pretend that the resulting curve has the shape of the regression curve that would arise if we had unlimited data", according to Mosteller and Tukey in their *Data Analysis and Regression* book (1977, p. 266). Instead, the values result in an approximation. With only limited data, a shape for the regression curve (e.g., linear, or exponential) is assumed and the curve is fitted to the data by using a statistical method such as the **method of least squares.** This method is explained shortly.

A Simple Linear Model

Because regression analysis seeks an algebraic relationship between a dependent variable *Y* and one or more independent variables, the deterministic (*change*) component of the model describes the mean (expected) value for Y given a specific value of X:

Deterministic component = Mean *Y*,

when Y is some function of X. In practice, there is considerable variability in Y for a given X around a mean value. This mean value is an unknown quantity that is commonly denoted by the Greek letter μ with a subscript $Y(X)$ to denote its dependence on X:

$$\mu_{Y(X)} = \beta_0 + \beta_1 X$$

One key assumption in the linear regression model is that for any value of X, the value of Y is scattered around a mean value.

The straight line may be approximately true; the difference between Y and the straight line is ascribed to a random error ("*chance*") component. Thus, the observed values of Y will not necessarily lie on a straight line in the XY plane but will differ from it by some random errors.

$$Y = \beta_0 + \beta_1 X + \text{Random errors}$$

Thus, the **simple linear regression model** (SLR) for Y can be expressed by the sum of a deterministic component $\mu_{Y(X)}$ and a random component ε:

$$Y = \mu_{Y(X)} + \varepsilon$$

$$= \beta_0 + \beta_1 X + \varepsilon$$

where the mean (expected) value of random errors ε is assumed to be zero. The intercept β_0 and slope β_1 are known as the regression parameters. The model is linear in the parameters, both β_0 and β_1 are unknown parameters to be estimated from the data. As a standard statistical convention, it is useful to designate unknown parameters in models by Greek letters, to distinguish them from the corresponding statistics b_0 and b_1 (in Roman letters) estimated from the data.

As demand forecasters, we can view the deterministic component as describing a systematic *change,* whereas the random errors depict measured *chance* in the sense of our notion of embracing **both** "change & chance" for demand forecasting best practices. In a particular application of the model, the demand forecaster has data that are assumed to have arisen as a realization of the hypothetical model. The next step is to come up with a rationale for estimating the parameters, β_0 and β_1, in the model from a given set of data.

The Least-Squares Assumption

There are many techniques around for estimating parameters from data, but the method of **ordinary least squares** (OLS) is the most common and it has a sound basis in statistical theory. This is not to say that other techniques have little merit.

In fact, weighted least-squares techniques of several kinds have been found to have increased importance in practical applications of outlier-resistant data analysis methods and robust regression (cf. C. Fred. Mosteller and John W. Tukey, *Data Analysis and Regression*, 1977).

The method of least squares is the most widely accepted criterion for estimating parameters in a model.

Consider now one reasonable criterion for estimating β_0 and β_1 from data in a simple linear regression model. OLS determines values b_0 and b_1 (Notation is important now because parameters will be estimated from data, so we “plug in” lower-case Roman letters b_0 and b_1 to replace the Greek symbols β_0 and β_1), so that the sum of squared vertical deviations (squared residuals) between the data and a fitted line is less than the sum of the squared vertical deviations from any other straight-line fit that could be drawn through the data:

$$\text{Minimum of } \Sigma(\text{Data - Fit})^2 = \Sigma(\text{Residuals})^2$$

Y_i	X	$y_i = (Y_i - \bar{Y})$	$x_i = (X_i - \bar{X})$	$y_i\, x_i$	x_i^2
3	1	−5	−2	10	4
5	2	−3	−1	3	1
7	3	−1	0	0	0
14	4	6	1	6	1
11	5	3	2	6	4
Sum 40	15			25	10

Average: $\bar{Y} = 8 \quad \bar{X} = 3$
$b_1 = 25/10$
$b_0 = \bar{Y} - b_1 \quad \bar{X} = 8 - 2.5(3) = 0.5$
Fitted equation: $\hat{Y} = 0.5 + 2.5X$

Figure10.1 Example illustrating the calculation of regression coefficients in a simple linear regression.

Recall Residual = Data - ($b_0 + b_1\, X$). A vertical deviation is the vertical distance from an observed data point to the line. Each deviation in the sample is squared and the least-squares line is defined to be the straight line that makes the sum of these squared deviations a minimum. The notation for this is as follows.

Consider the data pairs (Y_i, X_i) for (i = 1, . . . , n). Let $y_i = Y_i - \bar{Y}$ and $x_i = X_i - \ddot{X}$, where $\bar{Y} = (\Sigma Y_i)/n$, and $\ddot{X} = (\Sigma X_i)/n$. The symbol Σ denotes the summation over n values. Then $\Sigma\, D^2 = \Sigma\, (Y_i - b_0 - b_1\, X_i)^2$ is minimized. Figure 10.1 shows the calculations of b_0 and b_1 for a small set of data.

CASE: Sales and Advertising of a Weight Control Product

The VP of Sales has called you into her office to help plan for an upcoming advertising campaign. To date, much of the planning has used a seat-of-the-pants approach and not entirely satisfactory. To help improve the situation, you recommend investigating some quantitative approaches to the problem. You want to gain some familiarity with some methodologies you find in a search on the Internet, and you embark on a preliminary analysis of an existing data set. It is hoped that this investigation will lead to some insights that will help tackle your company’s data.

The complete chapter can be found in

Change & Chance Embraced

ACHIEVING AGILITY WITH DEMAND FORECASTING IN THE SUPPLY CHAIN

HANS LEVENBACH, PhD

ISBN-10:0692945989
ISBN-13:978-069294588

Cover art: Rose Tannenbaum, Berkshire TypeGraphic, www.tgo.com

Cartoon art: Don Engel, Isographic LLC

Copy editor: David Coen, Freelance copyeditor

Printed by Createspace, www.createspace.com

Professional development: Certified Professional Demand Forecaster (CPDF®), www.cpdftraining.org

Contents

: