Linear Models and Conjoint Analysis with Nonlinear Spline Transformations

Warren F. Kuhfeld

Mark Garratt

Abstract

Many common data analysis models are based on the general linear univariate model, including linear regression, analysis of variance, and conjoint analysis. This chapter discusses the general linear model in a framework that allows nonlinear transformations of the variables. We show how to evaluate the effect of each transformation. Applications to marketing research are presented.*

Why Use Nonlinear Transformations?

In marketing research, as in other areas of data analysis, relationships among variables are not always linear. Consider the problem of modeling product purchasing as a function of product price. Purchasing may decrease as price increases. For consumers who consider price to be an indication of quality, purchasing may increase as price increases but then start decreasing as the price gets too high. The number of purchases may be a discontinuous function of price with jumps at "round numbers" such as even dollar amounts. In any case, it is likely that purchasing behavior is not a linear function of price. Marketing researchers who model purchasing as a linear function of price may miss valuable nonlinear information in their data. A transformation regression model can be used to investigate the nonlinearities. The data analyst is not required to specify the form of the nonlinear function; the data suggest the function.

The primary purpose of this chapter is to suggest the use of linear regression models with nonlinear transformations of the variables—*transformation regression* models. It is common in marketing research to model nonlinearities by fitting a quadratic polynomial model. Polynomial models often have collinearity problems, but that can be overcome with orthogonal polynomials. The problem that polynomials cannot overcome is the fact that polynomial curves are rigid; they do not do a good job of locally fitting the data. Piecewise polynomials or *splines* are generalizations of polynomials that provide more flexibility than ordinary polynomials.

*This chapter is a revision of a paper that was presented to the American Marketing Association, Advanced Research Techniques Forum, June 14–17, 1992, Lake Tahoe, Nevada. The authors are: Warren F. Kuhfeld, Manager, Multivariate Models R&D, SAS Institute Inc., Cary NC 27513-2414. Mark Garratt was with Conway | Milliken & Associates, when this paper was presented and is now with In4mation Insights. Copies of this chapter (MR-2010J), sample code, and all of the macros are available on the Web `http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html`. All plots in this chapter are produced using ODS Graphics. For help, please contact SAS Technical Support. See page 25 for more information.

Background and History

The foundation for our work can be found mostly in the psychometric literature. Some relevant references include: Kruskal & Shepard (1974); Young, de Leeuw, & Takane (1976); de Leeuw, Young, & Takane (1976); Perreault & Young (1980); Winsberg & Ramsay (1980); Young (1981); Gifi (1981, 1990); Coolen, van Rijckevorsel, & de Leeuw (1982); van Rijckevorsel (1982); van der Burg & de Leeuw (1983); de Leeuw (1986), and many others. The transformation regression problem has also received attention in the statistical literature (Breiman & Friedman 1985, Hastie & Tibshirani 1986) under the names *ACE* and *generalized additive models*.

Our work is characterized by the following statements:

- Transformation regression is an inferential statistical technique, not a purely descriptive technique.
- We prefer smooth nonlinear spline transformations over step-function transformations.
- Transformations are found that minimize a squared-error loss function.

Many of the models discussed in this chapter can be directly fit with some data manipulations and any multiple regression or canonical correlation software; some models require specialized software. Algorithms are given by Kuhfeld (1990), de Boor (1978), and in SAS/STAT documentation.

Next, we present notation and review some fundamentals of the general linear univariate model.

The General Linear Univariate Model

A general linear univariate model has the scalar form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m + \epsilon$$

and the matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The dependent variable $\mathbf{y}$ is an $(n \times 1)$ vector of observations; $\mathbf{y}$ has expected value $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and expected variance $V(\mathbf{y}) = \sigma^2\mathbf{I}_n$. The vector $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ contains the unobservable deviations from the expected values. The assumptions on $\mathbf{y}$ imply $E(\epsilon) = \mathbf{0}$ and $V(\epsilon) = \sigma^2\mathbf{I}_n$. The columns of $\mathbf{X}$ are the independent variables; $\mathbf{X}$ is an $(n \times m)$ matrix of constants that are assumed to be known without appreciable error. The elements of the column vector $\boldsymbol{\beta}$ are the parameters. The objectives of a linear models analysis are to estimate the parameter vector $\boldsymbol{\beta}$, estimate interesting linear combinations of the elements of $\boldsymbol{\beta}$, and test hypotheses about the parameters $\boldsymbol{\beta}$ or linear combinations of $\boldsymbol{\beta}$.

We discuss fitting linear models with nonlinear spline transformations of the variables. Note that we do *not* discuss models that are nonlinear in the parameters such as

$$y = e^{x\beta} + \epsilon$$

$$y = \beta_0 x^{\beta_1} + \epsilon$$

$$y = \frac{\beta_1 x_1 + \beta_2 x_1^2}{\beta_3 x_2 + \beta_4 x_2^2} + \epsilon$$

Table 1
Cubic Polynomial
Spline Basis

1	-5	25	-125
1	-4	16	-64
1	-3	9	-27
1	-2	4	-8
1	-1	1	-1
1	0	0	0
1	1	1	1
1	2	4	8
1	3	9	27
1	4	16	64
1	5	25	125

Table 2
Cubic Polynomial
With Knots at −2, 0, 2

1	-5	25	-125	0	0	0
1	-4	16	-64	0	0	0
1	-3	9	-27	0	0	0
1	-2	4	-8	0	0	0
1	-1	1	-1	1	0	0
1	0	0	0	8	0	0
1	1	1	1	27	1	0
1	2	4	8	64	8	0
1	3	9	27	125	27	1
1	4	16	64	216	64	8
1	5	25	125	343	125	27

Table 3
Basis for a Discontinuous (at 0) Spline

1	-5	25	-125	0	0	0	0
1	-4	16	-64	0	0	0	0
1	-3	9	-27	0	0	0	0
1	-2	4	-8	0	0	0	0
1	-1	1	-1	0	0	0	0
1	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
1	2	4	8	1	2	4	8
1	3	9	27	1	3	9	27
1	4	16	64	1	4	16	64
1	5	25	125	1	5	25	125

Our nonlinear transformations are found within the framework of the general linear model.

There are numerous special cases of the general linear model that are of interest. When all of the columns of **y** and **X** are interval variables, the model is a multiple regression model. When all of the columns of **X** are indicator variables created from nominal variables, the model is a main-effects analysis of variance model, or a metric conjoint analysis model. The model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

is of special interest. It is a *linear* model because it is linear in the parameters, and it models y as a *nonlinear* function of x. It is a *cubic polynomial* regression model, which is a special case of a spline.

Polynomial Splines

Splines are curves that are typically required to be continuous and smooth. Splines are usually defined as piecewise polynomials of degree d whose function values and first $d-1$ derivatives agree at the points where they join. The abscissa values of the join points are called knots. The term spline is also used for polynomials (splines with no knots), and piecewise polynomials with more than one discontinuous derivative. Splines with more knots or more discontinuities fit the data better and use more degrees of freedom (*df*). Fewer knots and fewer discontinuities provide smoother splines that user fewer *df*. A spline of degree three is a cubic spline, degree two splines are quadratic splines, degree one splines are piecewise linear, and degree zero splines are step functions. Higher degrees are rarely used.

A simple special case of a spline is the line,

$$\beta_0 + \beta_1 x$$

from the simple regression model

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

A line is continuous and completely smooth. However, there is little to be gained by thinking of a line as a spline. A special case of greater interest was mentioned previously. The polynomial

$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

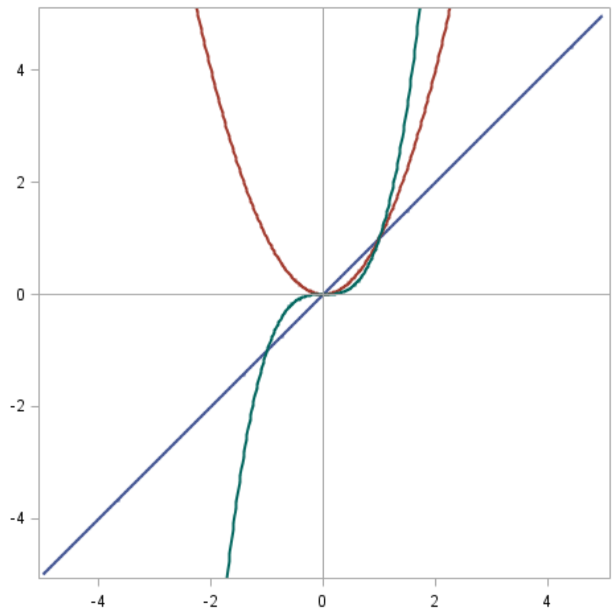

Figure 1. Linear, Quadratic, and Cubic Curves

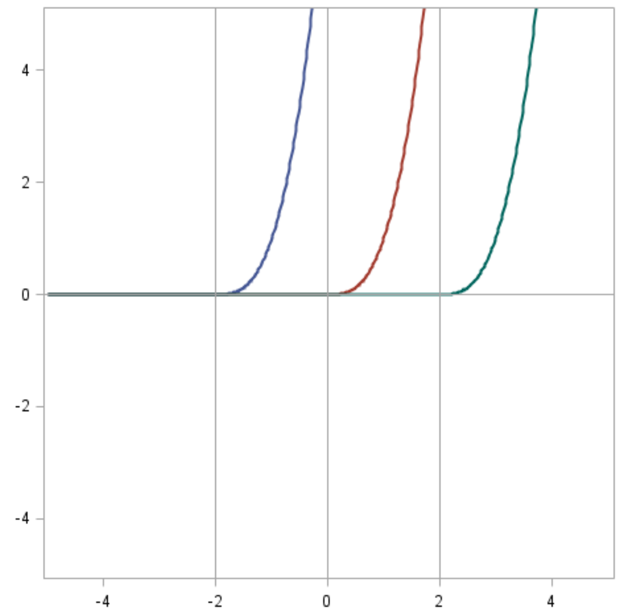

Figure 2. Curves For Knots at −2, 0, 2

from the linear model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

is a cubic spline with no knots. This equation models y as a *nonlinear* function of x, but does so with a *linear* regression model; y is a linear function of x, x^2, and x^3. Table 1 shows the $\mathbf{X}$ matrix, $(\mathbf{1} \ \ \mathbf{x} \ \ \mathbf{x}^2 \ \ \mathbf{x}^3)$, for a cubic polynomial, where $x = -5, -4, ..., 5$. Figure 1 plots the polynomial terms (except the intercept). See Smith (1979) for an excellent introduction to splines.

Splines with Knots

Here is an example of a polynomial spline model with three knots at t_1, t_2, and t_3.

$$\begin{aligned} y \;=\; & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ & \beta_4 (x > t_1)(x - t_1)^3 + \\ & \beta_5 (x > t_2)(x - t_2)^3 + \\ & \beta_6 (x > t_3)(x - t_3)^3 + \epsilon \end{aligned}$$

The Boolean expression $(x > t_1)$ is 1 if $x > t_1$, and 0 otherwise. The term

$$\beta_4 (x > t_1)(x - t_1)^3$$

is zero when $x \le t_1$ and becomes nonzero, letting the curve change, as x becomes greater than knot t_1. This spline uses more *df* and is less smooth than the polynomial model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

Assume knots at -2, 0, and 2; the spline model is:

$$y \;=\; \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 +$$

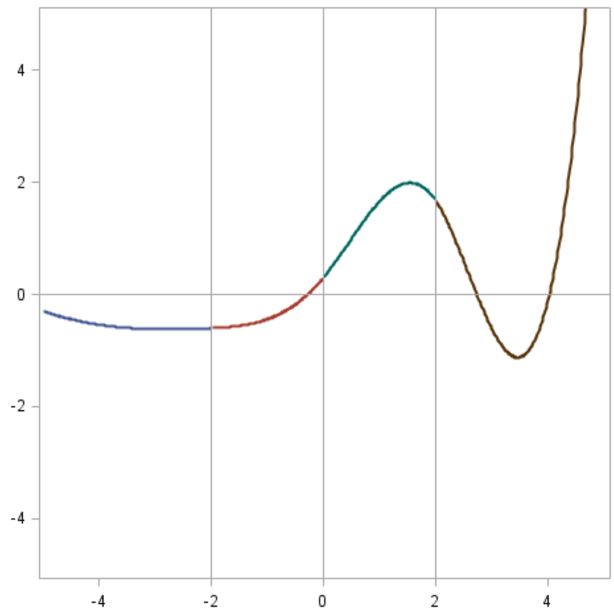

Figure 3. A Spline Curve With Knots at −2, 0, 2

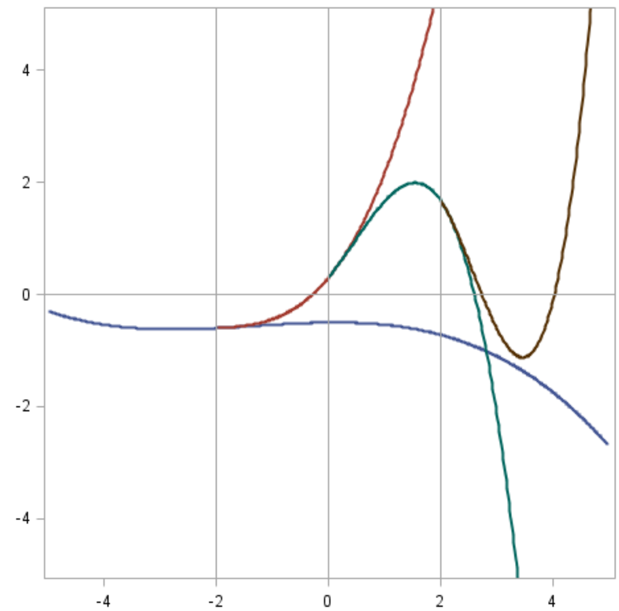

Figure 4. The Components of the Spline

$$\begin{aligned}&\beta_4(x > -2)(x - -2)^3 + \\ &\beta_5(x > \quad 0)(x - \quad 0)^3 + \\ &\beta_6(x > \quad 2)(x - \quad 2)^3 + \epsilon\end{aligned}$$

Table 2 shows an **X** matrix for this model, Figure 1 plots the polynomial terms, and Figure 2 plots the knot terms.

The β_0, $\beta_1 x$, $\beta_2 x^2$, and $\beta_3 x^3$ terms contribute to the overall shape of the curve. The

$$\beta_4(x > -2)(x - -2)^3$$

term has no effect on the curve before $x = -2$, and allows the curve to change at $x = -2$. The $\beta_4(x > -2)(x - -2)^3$ term is exactly zero at $x = -2$ and increases as x becomes greater than -2. The $\beta_4(x > -2)(x - -2)^3$ term contributes to the shape of the curve even beyond the next knot at $x = 0$, but at $x = 0$,

$$\beta_5(x > 0)(x - 0)^3$$

allows the curve to change again. Finally, the last term

$$\beta_6(x > 2)(x - 2)^3$$

allows one more change. For example, consider the curve in Figure 3. It is the spline

$$\begin{aligned}y \quad = \quad & -0.5 + 0.01x + -0.04x^2 + -0.01x^3 + \\ & 0.1(x > -2)(x - -2)^3 + \\ & -0.5(x > \quad 0)(x - \quad 0)^3 + \\ & 1.5(x > \quad 2)(x - \quad 2)^3\end{aligned}$$

It is constructed from the curves in Figure 4. At $x = -2.0$ there is a branch;

$$y = -0.5 + 0.01x + -0.04x^2 + -0.01x^3$$

continues over and down while the curve of interest,

$$\begin{aligned} y &= -0.5 + 0.01x + -0.04x^2 + -0.01x^3 + \\ &\quad 0.1(x > -2)(x - -2)^3 \end{aligned}$$

starts heading upwards. At $x = 0$, the addition of

$$-0.5(x > 0)(x - 0)^3$$

slows the ascent until the curve starts decreasing again. Finally, the addition of

$$1.5(x > 2)(x - 2)^3$$

produces the final change. Notice that the curves do not immediately diverge at the knots. The function and its first two derivatives are continuous, so the function is smooth everywhere.

Derivatives of a Polynomial Spline

The next equations show a cubic spline model with a knot at t_1 and its first three derivatives with respect to x.

$$\begin{aligned} y &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ &\quad \beta_4 (x > t_1)(x - t_1)^3 + \epsilon \\ \frac{dy}{dx} &= \beta_1 + 2\beta_2 x + 3\beta_3 x^2 + \\ &\quad 3\beta_4 (x > t_1)(x - t_1)^2 \\ \frac{d^2y}{dx^2} &= 2\beta_2 + 6\beta_3 x + \\ &\quad 6\beta_4 (x > t_1)(x - t_1) \\ \frac{d^3y}{dx^3} &= 6\beta_3 + 6\beta_4 (x > t_1) \end{aligned}$$

The first two derivatives are continuous functions of x at the knots. This is what gives the spline function its smoothness at the knots. In the vicinity of the knots, the curve is continuous, the slope of the curve is a continuous function, and the rate of change of the slope function is a continuous function. The third derivative is discontinuous at the knots. It is the horizontal line $6\beta_3$ when $x \leq t_1$ and jumps to the horizontal line $6\beta_3 + 6\beta_4$ when $x > t_1$. In other words, the cubic part of the curve changes at the knots, but the linear and quadratic parts do not change.

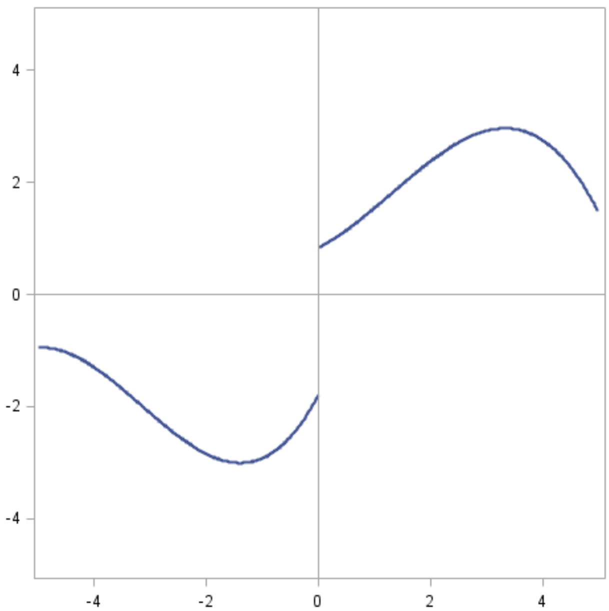

Figure 5. A Discontinuous Spline Function

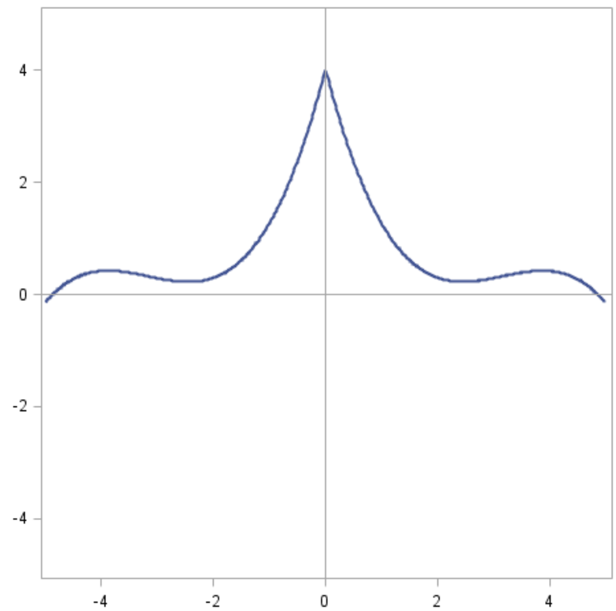

Figure 6. A Spline With a Discontinuous Slope

Discontinuous Spline Functions

Here is an example of a spline model that is discontinuous at $x = t_1$.

$$\begin{aligned} y \;=\; & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ & \beta_4 (x > t_1) + \\ & \beta_5 (x > t_1)(x - t_1) + \\ & \beta_6 (x > t_1)(x - t_1)^2 + \\ & \beta_7 (x > t_1)(x - t_1)^3 + \epsilon \end{aligned}$$

Figure 5 shows an example, and Table 3 shows a design matrix for this model with $t_1 = 0$. The fifth column is a binary (zero/one) vector that allows the discontinuity. It is a change in the intercept parameter. Note that the sixth through eighth columns are necessary if the spine is to consist of two independent polynomials. Without them, there is a jump at $t_1 = 0$, but both curves are based on the same polynomial. For $x \leq t_1$, the spline is

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

and for $x > t_1$, the spline is

$$\begin{aligned} y \;=\; & \beta_0 + \beta_4 + \\ & \beta_1 x + \beta_5 (x - t_1) + \\ & \beta_2 x^2 + \beta_6 (x - t_1)^2 + \\ & \beta_3 x^3 + \beta_7 (x - t_1)^3 + \epsilon \end{aligned}$$

The discontinuities are as follows:

$$\beta_7 (x > t_1)(x - t_1)^3$$

specifies a discontinuity in the third derivative of the spline function at t_1,

$$\beta_6 (x > t_1)(x - t_1)^2$$

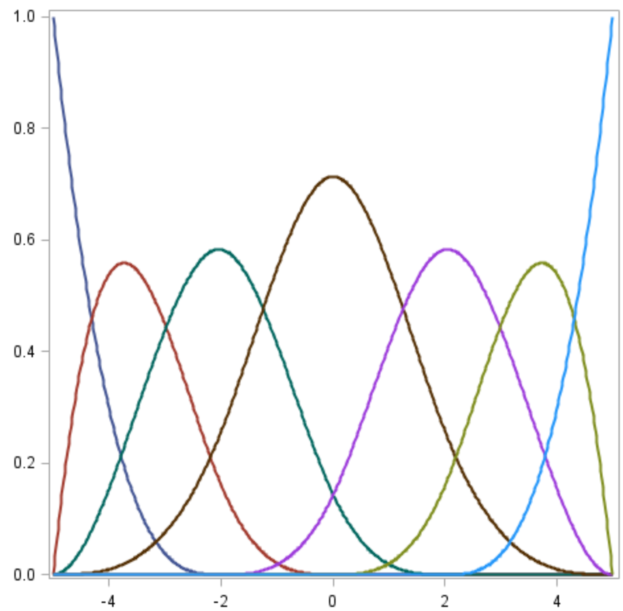

Figure 7. B-Splines With Knots at −2, 0, 2

Table 4
Cubic B-Spline With Knots at −2, 0, 2

1.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30	0.54	0.15	0.01	0.00	0.00	0.00
0.04	0.45	0.44	0.08	0.00	0.00	0.00
0.00	0.16	0.58	0.26	0.00	0.00	0.00
0.00	0.02	0.41	0.55	0.02	0.00	0.00
0.00	0.00	0.14	0.71	0.14	0.00	0.00
0.00	0.00	0.02	0.55	0.41	0.02	0.00
0.00	0.00	0.00	0.26	0.58	0.16	0.00
0.00	0.00	0.00	0.08	0.44	0.45	0.04
0.00	0.00	0.00	0.01	0.15	0.54	0.30
0.00	0.00	0.00	0.00	0.00	0.00	1.00

specifies a discontinuity in the second derivative at t_1,

$$\beta_5(x > t_1)(x - t_1)$$

specifies a discontinuity in the derivative at t_1, and

$$\beta_4(x > t_1)$$

specifies a discontinuity in the function at t_1. The function consists of two separate polynomial curves, one for $-\infty < x \leq t_1$ and the other for $t_1 < x < \infty$. This kind of spline can be used to model a discontinuity in price.

Here is an example of a spline model that is continuous at $x = t_1$ but does not have $d - 1$ continuous derivatives.

$$\begin{aligned} y &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ &\quad \beta_4(x > t_1)(x - t_1) + \\ &\quad \beta_5(x > t_1)(x - t_1)^2 + \\ &\quad \beta_6(x > t_1)(x - t_1)^3 + \epsilon \end{aligned}$$

$$\begin{aligned} \frac{dy}{dx} &= \beta_1 + 2\beta_2 x + 3\beta_3 x^2 + \\ &\quad \beta_4(x > t_1) + \\ &\quad 2\beta_5(x > t_1)(x - t_1) + \\ &\quad 3\beta_6(x > t_1)(x - t_1)^2 \end{aligned}$$

Since the first derivative is not continuous at $t_1 = x$, the slope of the spline is not continuous at $t_1 = x$. Figure 6 contains an example with $t_1 = 0$. Notice that the slope of the curve is indeterminate at zero.

Table 5

Polynomial and B-Spline Eigenvalues

B-Spline Basis

Eigenvalue	Proportion	Cumulative
0.107872	0.358718	0.35872
0.096710	0.321599	0.68032
0.046290	0.153933	0.83425
0.030391	0.101062	0.93531
0.012894	0.042878	0.97819
0.006559	0.021810	1.00000

Polynomial Spline Basis

Eigenvalue	Proportion	Cumulative
10749.8	0.941206	0.94121
631.8	0.055317	0.99652
37.7	0.003300	0.99982
1.7	0.000148	0.99997
0.3	0.000029	1.00000
0.0	0.000000	1.00000

Monotone Splines and B-Splines

An increasing *monotone spline* never decreases; its slope is always positive or zero. Decreasing monotone splines, with slopes that are always negative or zero, are also possible. Monotone splines cannot be conveniently created from polynomial splines. A different basis, the *B-spline* basis, is used instead. Polynomial splines provide a convenient way to describe splines, but B-splines provide a better way to fit spline models.

The columns of the B-spline basis are easily constructed with a recursive algorithm (de Boor 1978, pages 134–135). A basis for a vector space is a linearly independent set of vectors; every vector in the space has a unique representation as a linear combination of a given basis. Table 4 shows the B-spline $\mathbf{X}$ matrix for a model with knots at -2, 0, and 2. Figure 7 shows the B-spline curves. The columns of the matrix in Table 4 can all be constructed by taking linear combinations of the columns of the polynomial spline in Table 2. Both matrices form a basis for the same vector space.

The numbers in the B-spline basis are all between zero and one, and the marginal sums across columns are all ones. The matrix has a diagonally banded structure, such that the band moves one position to the right at each knot. The matrix has many more zeros than the matrix of polynomials and much smaller numbers. The columns of the matrix are not orthogonal like a matrix of orthogonal polynomials, but collinearity is not a problem with the B-spline basis like it is with a polynomial spline. The B-spline basis is very stable numerically.

To illustrate, 1000 evenly spaced observations were generated over the range -5 to 5. Then a B-spline basis and polynomial spline basis were constructed with knots at -2, 0, and 2. The eigenvalues for the centered $\mathbf{X}'\mathbf{X}$ matrices, excluding the last structural zero eigenvalue, are given in Table 5. In the polynomial splines, the first two components already account for more than 99% of the variation of the points. In the B-splines, the cumulative proportion does not pass 99% until the last term. The eigenvalues show that the B-spline basis is better conditioned than the piecewise polynomial basis. B-splines are used instead of orthogonal polynomials or Box-Cox transformations because B-splines allow knots and more general curves. B-splines also allow monotonicity constraints.

A transformation of x is monotonically increasing if the coefficients that are used to combine the columns of the B-spline basis are monotonically increasing. Models with splines can be fit directly using ordinary least squares (OLS). OLS does not work for monotone splines because OLS has no method of ensuring monotonicity in the coefficients. When there are monotonicity constraints, an alternating least square (ALS) algorithm is used. Both OLS and ALS attempt to minimize a squared error loss function. See Kuhfeld (1990) for a description of the iterative algorithm that fits monotone splines. See Ramsay (1988) for some applications and a different approach to ensuring monotonicity.

Transformation Regression

If the dependent variable is not transformed and if there are no monotonicity constraints on the independent variable transformations, the transformation regression model is the same as the OLS regression model. When only the independent variables are transformed, the transformation regression model is nothing more than a different rendition of an OLS regression. All of the spline models presented up to this point can be reformulated as

$$y = \beta_0 + \Phi(x) + \epsilon$$

The nonlinear transformation of x is $\Phi(x)$; it is solved for by fitting a spline model such as

$$\begin{aligned} y \;=\; & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ & \beta_4 (x > t_1)(x - t_1)^3 + \\ & \beta_5 (x > t_2)(x - t_2)^3 + \\ & \beta_6 (x > t_3)(x - t_3)^3 + \epsilon \end{aligned}$$

where

$$\begin{aligned} \Phi(x) \;=\; & \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \\ & \beta_4 (x > t_1)(x - t_1)^3 + \\ & \beta_5 (x > t_2)(x - t_2)^3 + \\ & \beta_6 (x > t_3)(x - t_3)^3 \end{aligned}$$

Consider a model with two polynomials:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \beta_4 x_2 + \beta_5 x_2^2 + \beta_6 x_2^3 + \epsilon$$

It is the same as a transformation regression model

$$y = \beta_0 + \Phi_1(x_1) + \Phi_2(x_2) + \epsilon$$

where $\Phi_\bullet(\bullet)$ designates cubic spline transformations with no knots. The actual transformations in this case are

$$\widehat{\Phi}_1(x_1) = \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_1^2 + \widehat{\beta}_3 x_1^3$$

and

$$\widehat{\Phi}_2(x_2) = \widehat{\beta}_4 x_2 + \widehat{\beta}_5 x_2^2 + \widehat{\beta}_6 x_2^3$$

There are six model *df.* The test for the effect of the transformation $\Phi_1(x_1)$ is the test of the linear hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$, and the $\Phi_2(x_2)$ transformation test is a test that $\beta_4 = \beta_5 = \beta_6 = 0$. Both tests are *F*-tests with three numerator *df.* When there are monotone transformations, constrained least-squares estimates of the parameters are obtained.

Degrees of Freedom

In an ordinary general linear model, there is one parameter for each independent variable. In the transformation regression model, many of the variables are used internally in the bases for the transformations. Each linearly independent basis column has one parameter and one model *df.* If a variable is not transformed, it has one parameter. Nominal classification variables with c categories have $c-1$ parameters. For degree d splines with k knots and $d-1$ continuous derivatives, there are $d+k$ parameters.

When there are monotonicity constraints, counting the number of scoring parameters is less precise. One way of handling a monotone spline transformation is to treat it as if it were simply a spline transformation with $d+k$ parameters. However, there are typically fewer than $d+k$ *unique* parameter estimates since some of those $d+k$ scoring parameter estimates may be tied to impose the order constraints. Imposing ties is equivalent to fitting a model with fewer parameters. So, there are two available scoring parameter counts: $d+k$ and a potentially smaller number that is determined during the analysis. Using $d+k$ as the model *df* is *conservative* since the scoring parameter estimates are restricted. Using the smaller count is too *liberal* since the data and the model together are being used to determine the number of parameters. Our solution is to report tests using both liberal and conservative *df* to provide lower and upper bounds on the "true" p-values.

Dependent Variable Transformations

When a dependent variable is transformed, the problem becomes multivariate:

$$\Phi_0(y) = \beta_0 + \Phi_1(x_1) + \Phi_2(x_2) + \epsilon$$

Hypothesis tests are performed in the context of a multivariate linear model, with the number of dependent variables equal to the number of scoring parameters for the dependent variable transformation. Multivariate normality is assumed even though it is known that the assumption is *always* violated. This is one reason that all hypothesis tests in the presence of a dependent variable transformation should be considered approximate at best.

For the transformation regression model, we redefine three of the usual multivariate test statistics: Pillai's Trace, Wilks' Lambda, and the Hotelling-Lawley Trace. These statistics are normally computed using all of the squared canonical correlations, which are the eigenvalues of the matrix $\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}$. Here, there is only one linear combination (the transformation) and hence only one squared canonical correlation of interest, which is equal to the R^2. We use R^2 for the first eigenvalue; all other eigenvalues are set to zero since only one linear combination is used. Degrees of freedom are computed assuming that all linear combinations contribute to the Lambda and Trace statistics, so the F-tests for those statistics are conservative. In practice, the adjusted Pillai's Trace is very conservative—perhaps too conservative to be useful. Wilks' Lambda is less conservative, and the Hotelling-Lawley Trace seems to be the least conservative.

It may seem that the Roy's Greatest Root statistic, which always uses only the largest squared canonical correlation, is the only statistic of interest. Unfortunately, Roy's Greatest Root is very liberal and only provides a lower bound on the p-value. The p-values for the liberal and conservative statistics are used together to provide approximate lower and upper bounds on p.

Scales of Measurement

Early work in scaling, such as Young, de Leeuw, & Takane (1976), and Perreault & Young (1980) was concerned with fitting models with mixed nominal, ordinal, and interval scale of measurement variables. Nominal variables were optimally scored using Fisher's (1938) optimal scoring algorithm. Ordinal variables were optimally scored using the Kruskal and Shepard (1974) monotone regression algorithm. Interval and ratio scale of measurement variables were left alone nonlinearly transformed with a polynomial transformation.

In the transformation regression setting, the Fisher optimal scoring approach is equivalent to using an indicator variable representation, as long as the correct *df* are used. The optimal scores are category means. Introducing optimal scaling for nominal variables does not lead to any increased capability in the regression model.

For ordinal variables, we believe the Kruskal and Shepard monotone regression algorithm should be reserved for the situation when a variable has only a few categories, say five or fewer. When there are more levels, a monotone spline is preferred because it uses fewer model *df* and because it is less likely to capitalize on chance.

Interval and ratio scale of measurement variables can be left alone or nonlinearly transformed with splines or monotone splines. When the true model has a nonlinear function, say

$$y = \beta_0 + \beta_1 log(x) + \epsilon$$

or

$$y = \beta_0 + \beta_1 / x + \epsilon$$

the transformation regression model

$$y = \beta_0 + \Phi(x) + \epsilon$$

can be used to hunt for parametric transformations. Plots of $\widehat{\Phi}(x)$ may suggest log or reciprocal transformations.

Conjoint Analysis

Green & Srinivasan (1990) discuss some of the problems that can be handled with a transformation regression model, particularly the problem of degrees of freedom. Consider a conjoint analysis design where a factor with $c > 3$ levels has an inherent ordering. By finding a quadratic monotone spline transformation with no knots, that variable will use only two *df* instead of the larger $c - 1$. The model *df* in a spline transformation model are determined by the data analyst, not by the number of categories in the variables. Furthermore, a "*quasi-metric*" conjoint analysis can be performed by finding a monotone spline transformation of the dependent variable. This model has fewer restrictions than a metric analysis, but will still typically have error *df*, unlike the nonmetric analysis.

Curve Fitting Applications

With a simple regression model, you can fit a line through a $y \times x$ scatter plot. With a transformation regression model, you can fit a curve through the scatter plot. The y-axis coordinates of the curve are

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\Phi}(x)$$

from the model

$$y = \beta_0 + \Phi(x) + \epsilon$$

With more than one group of observations and a multiple regression model, you can fit multiple lines, lines with the same slope but different intercepts, and lines with common intercepts but different slopes. With the transformation regression model, you can fit multiple curves through a scatter plot. The curves can be monotone or not, constrained to be parallel, or constrained to have the same intercept. Consider the problem of modeling the number of product purchases as a function of price. Separate curves can be simultaneously fit for two groups who may behave differently, for example those who are making a planned purchase and those who are buying impulsively. Later in this chapter, there is an example of plotting brand by price interactions.

Figure 8 contains an artificial example of two separate spline functions; the shapes of the two curves are independent of each other, and $R^2 = 0.87$. In Figure 9, the splines are constrained to be parallel, and $R^2 = 0.72$. The parallel curve model is more restrictive and fits the data less well than the unconstrained model. In Figure 8, each curve follows its swarm of data. In Figure 9, the curves find paths through the data that are best on the average considering both swarms together. In the vicinity of $x = -2$, the top curve is high and the bottom curve is low. In the vicinity of $x = 1$, the top curve is low and the bottom curve is high.

Figure 10 contains the same data and two monotonic spline functions; the shapes of the two curves are independent of each other, and $R^2 = 0.71$. The top curve is monotonically decreasing, whereas the bottom curve is monotonically increasing. The curves in Figure 10 flatten where there is nonmonotonicity in Figure 8.

Parallel curves are very easy to model. If there are two groups and the variable g is a binary variable indicating group membership, fit the model

$$y = \beta_0 + \beta_1 g + \Phi_1(x) + \epsilon$$

where $\Phi_1(x)$ is a linear, spline, or monotone spline transformation. Plot $\widehat{y}$ as a function of x to see the two curves. Separate curves are almost as easy; the model is

$$y = \beta_0 + \beta_1 g + \Phi_1(x \times (1 - g)) + \Phi_2(x \times g) + \epsilon$$

When $x \times (1 - g)$ is zero, $x \times g$ is x, and vice versa.

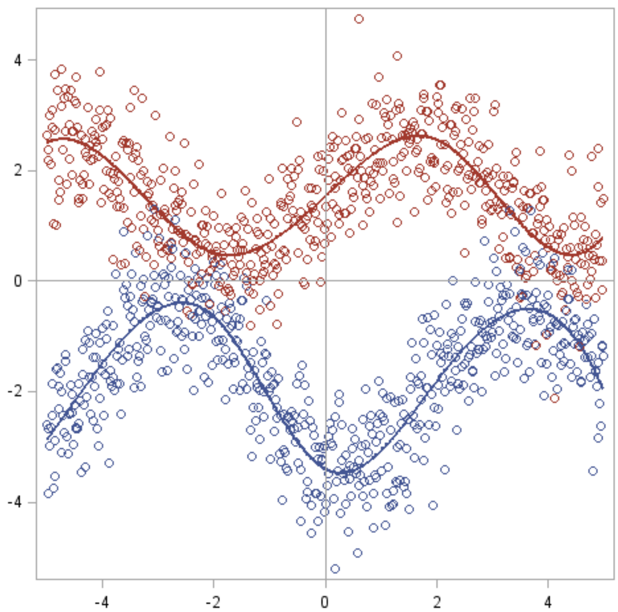

Figure 8. Separate Spline Functions, Two Groups

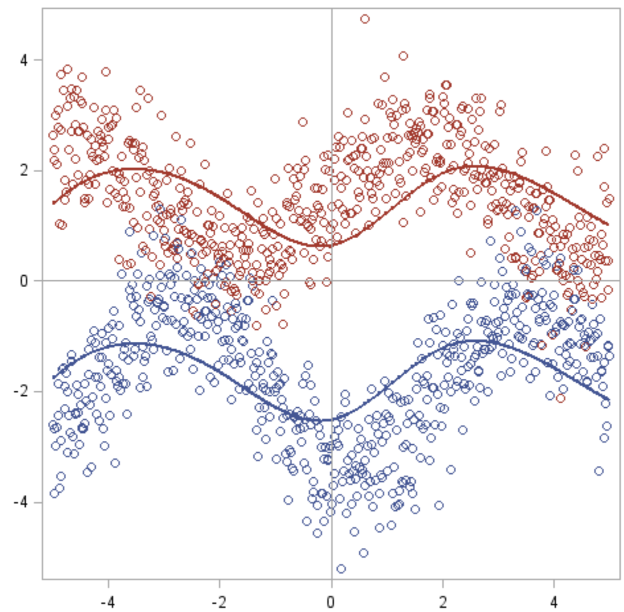

Figure 9. Parallel Spline Functions, Two Groups

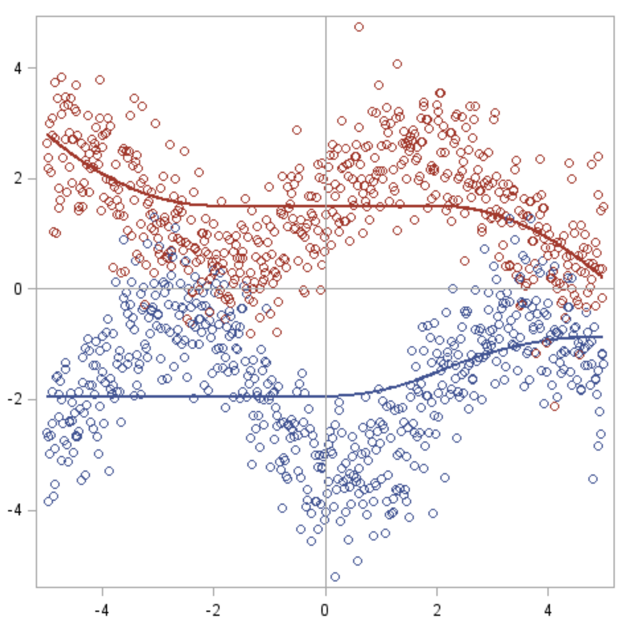

Figure 10. Monotone Spline Functions, Two Groups

Spline Functions of Price

This section illustrates splines with an artificial data set. Imagine that subjects were asked to rate their interest in purchasing various types of spaghetti sauces on a one to nine scale, where nine indicated definitely will buy and one indicated definitely will *not* buy. Prices were chosen from typical retail trade prices, such as $1.49, $1.99, $2.49, and $2.99; and one penny more than a typical price, $1.00, $1.50, $2.00, and $2.50. Between each "round" number price, such as $1.00, and each typical price, such as $1.49, three additional prices were chosen, such as $1.15, $1.25, and $1.35. The goal is to allow a model with a separate spline for each of the four ranges: $1.00 — $1.49, $1.50 — $1.99, $2.00 — $2.49, and $2.50 — $2.99. For each range, a spline with zero or one knot can be fit.

One rating for each price was constructed and various models were fit to the data. Figures 11 through 18 contain results. For each figure, the number of model *df* are displayed. One additional *df* for the intercept is also used. The SAS/STAT procedure TRANSREG was used to fit all of the models in this chapter.

Figure 11 shows the linear fit, Figure 12 uses a quadratic polynomial, and Figure 13 uses a cubic polynomial. The curve in Figure 13 has a slight nonmonotonicity in the tail, and since it is a polynomial, it is rigid and cannot locally fit the data values.

Figure 14 shows a monotone spline. It closely follows the data and never increases. A range for the model *df* is specified; the larger value is a conservative count and the smaller value is a liberal count.

The curves in Figures 12 through 14 are all continuous and smooth. These curves do a good job of following the data, but inspection of the data suggests that a different model may be more appropriate. There is a large drop in purchase interest when price increases from $1.49 to $1.50, a smaller drop between $1.99 and $2.00, and a still smaller drop between $2.49 and $2.50.

In Figure 15, a separate quadratic polynomial is fit for each of the four price ranges: $1.00 — $1.49, $1.50 — $1.99, $2.00 — $2.49, and $2.50 — $2.99. The functions are connected. The function over the range $1.00 — $1.49 is nearly flat; there is high purchase interest for all of these prices. Over the range $1.50 — $1.99, purchase interest drops more rapidly with a slight leveling in the low end; the slope decreases as the function increases. Over the range $2.00 — $2.49, purchase interest drops less rapidly; the slope increases as the function increases. Over the range $2.50 — $2.99, the function is nearly flat. At $1.99, $2.49, and $2.99 there is a slight increase in purchase interest.

In Figure 16, there is a knot in the middle of each range. This gives the spline more freedom to follow the data. Figure 17 uses the same model as Figure 16, but monotonicity is imposed. When monotonicity is imposed the curves touch fewer of the data values, passing in between the nonmonotonic points. In Figure 18, the means for each price are plotted and connected. This analysis uses the most model *df* and is the least smooth of the plots.

Benefits of Splines

In marketing research and conjoint analysis, the use of spline models can have several benefits. Whenever a factor has three or more levels and an inherent ordering of the levels, that factor can be modeled as a quadratic monotone spline. The *df* used by the variable is controlled at data analysis time; it is not simply the number of categories minus one. When the alternative is a model in which a factor is designated as nominal, splines can be used to fit a more restrictive model with fewer model *df*. Since the spline model has fewer *df*, it should yield more reproducible results.

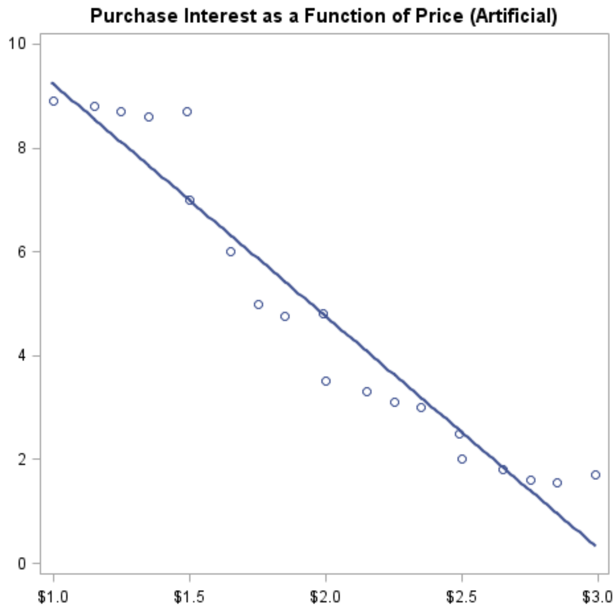

Figure 11. Linear Function, 1 df

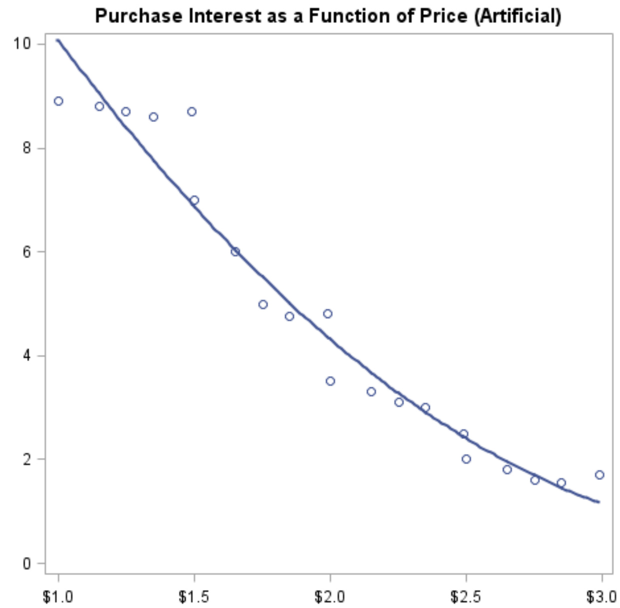

Figure 12. Quadratic Function, 2 df

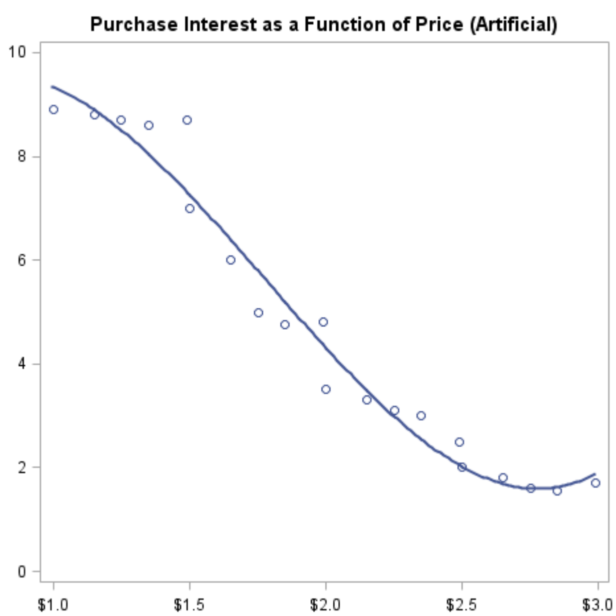

Figure 13. Cubic Function, 3 df

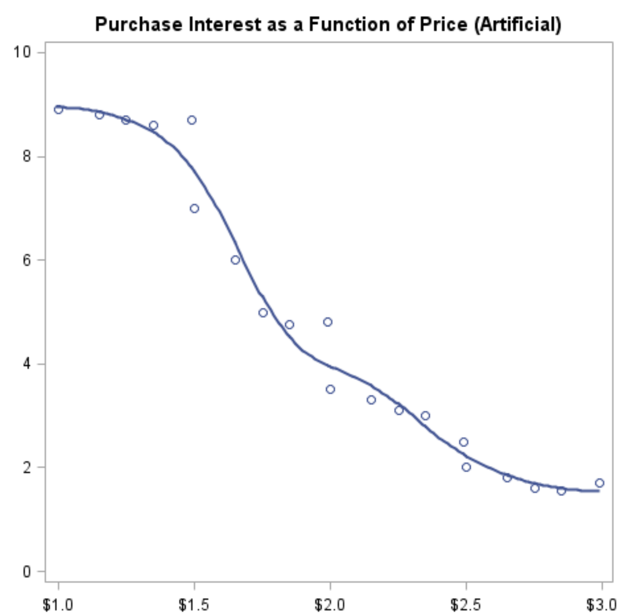

Figure 14. Monotone Function, 5–7 df

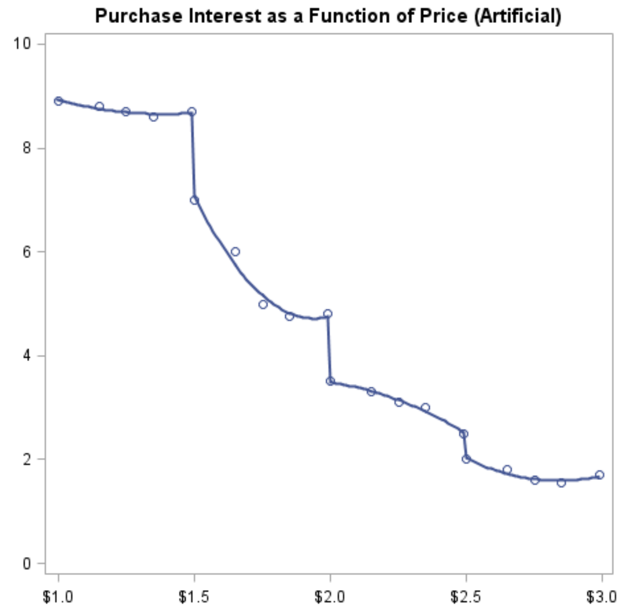

Figure 15. Discontinuous Polynomial, 11 df

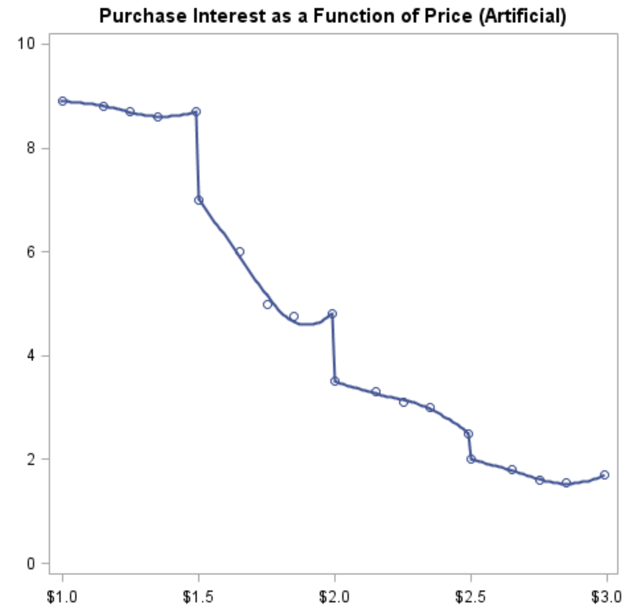

Figure 16. Discontinuous Spline Function, 15 df

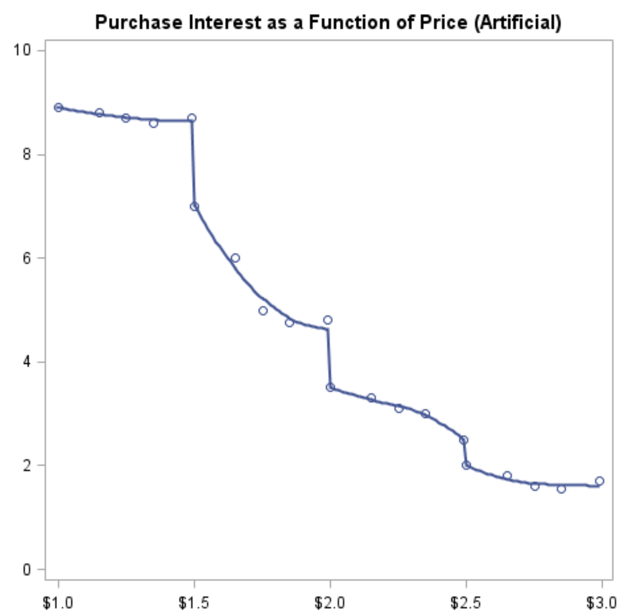

Figure 17. Discontinuous Monotone Spline, 12–15 df

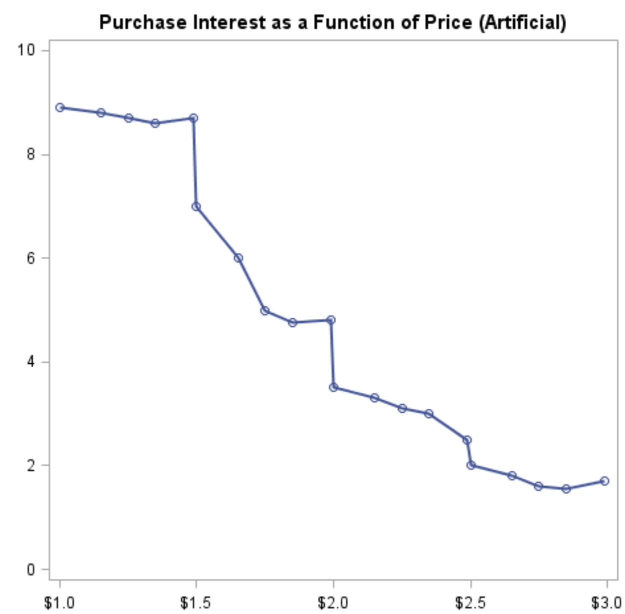

Figure 18. Cell Means, 19 df

The opposite alternative is also important. Consider a variable with many values, like price in some examples. Instead of using a restrictive single *df* linear model, splines can be used to fit a more general model with more *df*. The more general model may show information in the data that is not apparent from the ordinary linear model. This can be a benefit in conjoint analyses that focus on price, in the analysis of scanner data, and in survey research. Splines give you the ability to examine the nonlinearities that may be very important in predicting consumer behavior.

Fitting quadratic and cubic polynomial models is common in marketing research. Splines extend that capability by adding the possibility of knots and hence more general curves. Spline curves can also be restricted to be monotone. Monotone splines are less restrictive than a line and more restrictive than splines that can have both positive and negative slopes. You are no longer restricted to fitting just a line, polynomial, or a step function. Splines give you possibilities in between.

Conclusions

Splines allow you to fit curves to your data. Splines may not revolutionize the way you analyze data, but they will provide you with some new tools for your data analysis toolbox. These new tools allow you to try new methods for solving old problems and tackle new problems that could not be adequately solved with your old tools. We hope you will find these tools useful, and we hope that they will help you to better understand your marketing data.