CENTER OF PLANNING AND ECONOMIC RESEARCH

32

SPLINE FUNCTIONS FITTED BY STANDARD REGRESSION METHODS

By

DANIEL B. SUITS, ANDREW MASON, AND LOUIS CHAN

ATHENS 1977

SPLINE FUNCTIONS FITTED BY STANDARD REGRESSION METHODS

Printed in Greece
«Graphical Arts» George Papanicolaou
58 Ippokratous Str. – Athens 114 – Tel. 3624728 - 3627413

CENTER OF PLANNING AND ECONOMIC RESEARCH

32

SPLINE FUNCTIONS FITTED BY STANDARD REGRESSION METHODS

By

DANIEL B. SUITS, ANDREW MASON, AND LOUIS CHAN

ATHENS 1977

Opinions or value judgements expressed in this book are those of the author and do not necessarily represent those of the Center of Planning and Economic Research.

CENTER OF PLANNING AND ECONOMIC RESEARCH

The Center of Planning and Economic Research (KEPE) was founded in 1961 as an autonomous public organization, under the title "Center of Economic Research", its basic objective being research into the problems of the operation, structure and development of the Greek economy. Another of its objectives was the training of young Greek economists in modern methods of economic analysis and research. For the establishment and operation of the Center considerable financial aid was provided by foreign foundations.

During 1964, the Center of Economic Research was reorganized into its present form, as the Center of Planning and Economic Research. In addition to its function as a Research and Training Institute, the Center, in its new form, was assigned the following tasks by the State: (1) the preparation of economic development plans at a national and regional level, (2) the evaluation of public investment programmes, and (3) the study of short-term developments in the Greek economy and ad-

vising on current problems of economic policy.

For the realization of these aims, the KEPE, during its first years of operation (1961-1966) collaborated with foreign scientists and foundations. The latter helped in the selection of foreign economists who joined the Center to carry out scientific research into the problems of the Greek economy and in the organization of an exchange programme, including the post-graduate training of young Greek economists at universities abroad.

The Center has also developed a broad programme of scholarships for post-graduate studies in economics. Thus, in collaboration with foreign universities and international organizations, a number of young economists from Greece are sent abroad each year to specialize in the various fields of economics. In addition, the KEPE organizes a series of training seminars and lectures, frequently given by distinguished foreign scholars invited for that purpose to Greece.

In addition to the above, the KEPE maintains contact with similar institutions abroad, and exchanges publications and information concerning developments in methods of economic research, thus contributing to the promotion of the science of economics in the country.

TABLE OF CONTENTS

The subject matter of this publication has been delivered as a lecture at the Center of Planning and Economic Research in August 1977 by Professor Daniel B. Suits.

The authors are, respectively, Professor of Economics, Michigan State University, Assistant Professor of Economics, University of Hawaii, and undergraduate student, University of Hawaii.

SPLINE FUNCTIONS FITTED BY STANDARD REGRESSION METHODS

It sometimes happens that when a new mathematical or statistical procedure is adopted from one discipline into another, it arrives complete with terminology, usage and (these days) computer software devised for specialized application to problems in the parent area. This is probably inevitable, but until the new method is more broadly perceived its application may fall considerably short of potential in the adopting discipline. A recent example is the *spline function.* Briefly put, spline functions are a device for approximating the shape of a curvilinear stochastic function without the necessity of pre-specifying the mathematical form of the function. That is, it is unnecessary to restrict the estimate to a straight line, a polynomial of pre-specified degree, an exponential, or any other particular form.

Brought over from engineering and the mathematics of interpolation, spline functions have appeared in several places in economic statistics in recent years. Application to economic problems has been made by Barth, Kraft and Kraft

(1976), McGee and Carlton (1970) and Poitier (1973, 1976). Buse and Lim (1977) have shown spline functions to be a special case of restricted least squares. Yet because the idea is still wrapped in its original packaging, it is frequently overlooked when it might be a powerful adjunct to research.

Moreover, even in some of the work where the spline function has been employed, it has not always been used to best advantage. For example, Barth and others in the article cited above, although admitting that it might improve their analysis to employ a multivariate spline function, were constrained by the fact that the software package at their disposal "unfortunately . . . permits only bivariate specification." Yet, in fact, the procedure is readily adapted to bivariate or multivariate analysis.

In this article we show that by use of appropriately defined composite variables, spline functions are easily fitted by any standard package for ordinary least squares regression. Some of the examples given below were fitted by the familiar SPSS package, others were fitted by members of an undergraduate class in econometrics at the University of Hawaii, using the TSP routine.

In the presentation, piece-wise linear regression is employed as a general introduction to the procedure. This is followed by development of the bivariate and the multivariate spline function. The procedures are then illustrated by their application to the relationship of interest rates to money supply and inflation. Once the spline function is understood as a least squares regression model, additional variations become possible. As an example, we present a modified or "truncated" spline function and apply it to the relationship of fertility to per capita income. We conclude with a few general remarks on the limitations of the method.

Piece-wise Linear Regression

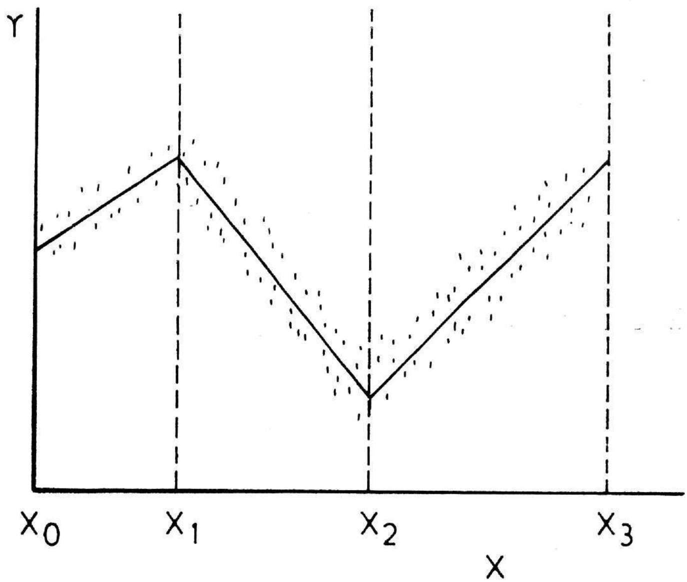
This procedure is already widely known and is reviewed here only to facilitate exposition of the nature of the spline function. The problem is illustrated in Figure I. Given the scatter of observations, a linear relationship would be a poor fit, nor is it clear that any readily available polynomial would be much of an improvement. As an alternative, one can fit a series of linear regressions, one to each of the segments marked off on the axis. The desired relationship would be:

$$(1) \quad Y = [a_1 + b_1(X - X_0)]D_1 + [a_2 + (X - X_1)]D_2 + [a_3 + (X - X_2)]D_3 + u$$

where D_i is a dummy variable whose value is 1 for all observations such that $X_{i-1} \leq X < X_i$ and is otherwise zero.

FIGURE 1

PIECE-WISE LINEAR REGRESSION



Unfortunately, a regression in form (1) will, in general, be discontinuous at X_1 and X_2, but the discontinuity can be obviated if the values of the coefficients are constrained so that

(2)
$$a_2 = a_1 + b_1 (X_1 - X_0)$$
$$a_3 = a_2 + b_2 (X_2 - X_1)$$

Substituting (2) into (1) and combining terms with like coefficients, we get

$$(3)\quad \begin{aligned} Y = a_1 &+ b_1 [(X - X_0)D_1 + (X_1 - X_0)D_2 + (X_2 - X_1)D_3] \\ &+ b_2 [(X - X_1)D_2 + (X_2 - X_1)D_3] \\ &+ b_3 [(X - X_2)D_3] + u \end{aligned}$$

(The alert reader will note that further simplification of (3) is possible, but it is adequate for our purpose as it stands.)

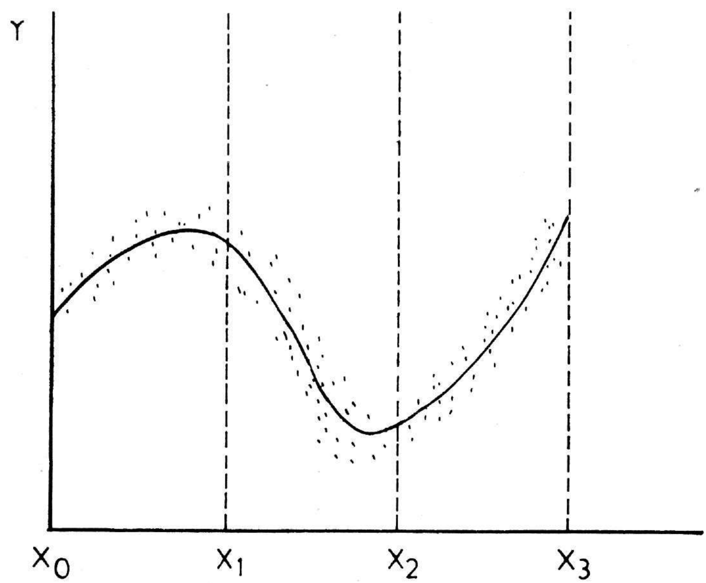
Formulation (3) converts the piece-wise linear approximation into a multiple regression in which the dependent variable Y is regressed on three composite variables whose values are constructed from the data for X, from the values of the X_i at which the function is to bend, together with the widths of the respective intervals and the three dummy variables. The constructed variables are readily generated by transformation programs included in most software packages for multiple regression, but could be calculated by hand if necessary.

Spline Functions

Piece-wise linear regression suffers from two obvious shortcomings. First, although the function itself is continuous, its derivatives are not. Discontinuity of derivatives at the kinks can prove a serious disadvantage in many economic applications where the result would be discontinuous—and probably spurious—shifts in elasticities, marginals, or other relationships that would becloud analysis. Secondly, a curvilinear relationship may provide a significantly better fit to the data than is obtained from linear segments. This consideration is especially important when we are confronted by a complicated curve without obvious critical positions to which linear segments could be fitted. If the X_i are to be located arbitrarily, we had better not rely on linear approximation to map out the function between them.

Spline functions overcome these disadvantages by replacing the linear approximations of (1) by a system of piece-wise polynomial approximations. Any degree of polynomial could be employed, but the cubic is convenient for most purposes. The nature of the cubic spline is illustrated in Figure 2. The X axis has been divided into three segments by the points X_0, X_1, X_2 and X_3. In spline theory, the points chosen are called "knots". This is as good a term as any and will be employed hereafter. For convenience, the intervals between the knots have been taken as equal. This is not an essential part of the procedure, but equality of interval is generally advisable unless there is important reason to do otherwise. More than four knots, and correspondingly more than three intervals can be used, but, as we will see below, the more intervals there are, the greater the number of composite variables required to fit the curve and the greater the loss of degrees of freedom.

FIGURE 2
PIECE-WISE POLYNOMIAL REGRESSION



It is now proposed to fit a regression in the form

$$(4)\ Y = [a_1 + b_1 (X - X_0) + c_1 (X - X_0)^2 + d_1 (X - X_0)^3]D_1 + [a_2 + b_2 (X - X_1) + c_2 (X - X_1)^2 + d_2 (X - X_1)^3]D_2 + [a_3 + b_3 (X - X_2) + c_3 (X - X_2)^2 + d_3 (X - X_2)^3]D_3 + u$$

Again, D_i is a dummy variable defined by the i-th interval.

In general, of course, (4) is discontinuous at the knots, as are its derivatives, but application of appropriate constraints to the coefficients not only makes the function continuous, but guarantees continuity of its first and second derivatives. The constraints required for this purpose are:

$$
\begin{aligned}
(5)\quad a_2 &= a_1 + b_1 (X_1 - X_0) + c_1 (X_1 - X_0)^2 + d_1 (X_1 - X_0)^3 \\
b_2 &= b_1 + 2c_1 (X_1 - X_0) + 3d_1 (X_1 - X_0)^2 \\
c_2 &= c_1 + 3d_1 (X_1 - X_0) \\
a_3 &= a_2 + b_2(X_2 - X_1) + c_2 (X_2 - X_1)^2 + d_2(X_2 - X_1)^3 \\
b_3 &= b_2 + 2c_2 (X_2 - X_1) + 3d_2 (X_2 - X_1)^2 \\
c_3 &= c_2 + 3d_2(X_2 - X_1)
\end{aligned}
$$

The constraints on the a_i equate the values of the left and right branches of the function at the knots. The constraints on the b_i equate the slopes of the right and left branches at the knots, while constraint of the c_i does the same for the second derivatives.

At this point it facilitates the exposition to consider the case of equal intervals, so let $(X_1-X_0)=(X_2-X_1)=(X_3-X_2)=w$. Substitution of w into (5), and the result into (4) yields, after collecting terms with the same coefficient:

$$
\begin{aligned}
(6)\, Y = {} & a_1 + b_1(X-X_0) \\
& + c_1\{(X-X_0)^2D_1 + (w+X-X_1)^2D_2 \\
& \qquad + (2w + X-X_2)^2D_3\} \\
& + d_1\{(X-X_0)^3D_1 + [w^3 + 3w(X-X_1)^2 + 3w(X-X_1)^3]D_2 \\
& \qquad + [7w^3 + 9w^2 (X-X_2) + 3w(X-X_2)^2]D_3\} \\
& + d_2\{(X-X_1)^3D_2 + [w^3 + w^2 (X-X_2) \\
& \qquad + 3w(X-X_2)^2]D_3\} \\
& + d_3(X-X_2)^3D_3
\end{aligned}
$$

The task of fitting a spline function is now seen as a straightforward multiple regression of Y on a set of five composite variables. Once the data have been transformed to yield the required five variables, the regression directly yields the desired values of a_1, b_1, c_1, d_1, d_2 and d_3. The remaining coefficients of (4) are readily obtained by substituting these values into (5).

By some additional manipulation*, however, (6) may be further simplified to yield an expression that not only is easier to use in practice, but which readily lends itself to generalization to a larger number of intervals. For this purpose, we define a new set of dummy variables, D_1^* and D_2^* with the property that $D_i^*=1$, if and only if $X \geq X_i$, otherwise, $D_i^*=0$. That is, D_1^* has value 0 until X reaches X_1, and is thereafter equal to 1. D_2^* is equal to 0 until X reaches X_2 and is thereafter equal to 1. Employing this notation, it can be shown that (6) is equivalent to

$$\begin{aligned}(7)\quad Y = a_1 &+ b_1(X-X_0) + c_1(X-X_0)^2 + d_1(X-X_0)^3 \\ &+ (d_2-d_1)(X-X_1)^3D_1^* + (d_3-d_2)(X-X_2)^3D_2^*\end{aligned}$$

Like (6), expression (7) is a multiple regression on five composite variables. The regression

---

* We acknowledge the contribution of Mr. James Coulter of Michigan State University in pointing out how equation (6) could be simplified to (7).

coefficients directly yield a_1, b_1, c_1 and d_1, and values of d_2 and d_3 are readily calculated from the two final coefficients.

The regression procedure itself is readily carried out by any standard least-squares regression package. Moreover, goodness of fit, significance tests, and other related statistics for the spline function are those obtained in the usual fashion from the multiple regression program.

Spline Functions Defined on More than Three Intervals

When more than four knots and three intervals are to be employed in fitting the spline function, the procedure is merely extended to incorporate additional branches of the function, related to each other by additional constraints. The process of deriving the required regression equation requires only a measure of patience, and the result is a generalization of (7).

A spline function fitted to k intervals defined by knots placed at $X_0, X_1, \ldots X_{k+1}$, with corresponding dummy variables $D_1^*, D_2^*, \ldots D_k^*$, is given by a multiple regression in the form

$$(8) \quad Y = a_1 + b_1(X-X_0) + c_1(X-X_0)^2 + d_1(X-X_0)^3 + \sum_{i=1}^{k} (d_{i+1}-d_i)\,(X-X_i)^3 D_i^*$$

Each additional interval used to fit the function involves an additional variable in the regression equation and the loss of an additional degree of freedom in the residual.

Spline Functions with More than One Independent Variable

Variables in addition to X are easily incorporated in the analysis. We may consider two situations. In the first case, an additional variable Z is to be incorporated linearly in the regression. That is, if we represent the spline function by $Y = S(X)$, then we want to incorporate Z additively in the form $Y = S(X) + kZ$. This is accomplished merely by regressing Y on Z in addition to the five constructed variables in (7) following standard multiple regression procedure.

In the second case, to allow for curvature in the Z as well as in the X dimension, we may incorporate Z in the relationship as a spline function $T(Z)$ to obtain $Y = S(X) + T(Z)$. To arrive at this result, knots Z_j are selected on the Z axis. Again, the intervals thus defined are generally, but not necessarily equal, but there is no special reason why there should be the same number of intervals on the Z as on the X axis. Nor, for that matter is it required that the polynomials pieced together on the Z dimension be of the same order as those employed in the X direction although the cubic is the most generally useful. In any event, continuity of the function in the Z dimension is assured by application of constraints exactly as in the case of X. For the sake of illustration, suppose the Z axis has been divided into equal segments by knots Z_0, Z_1 Z_2 and Z_3, defining three intervals of uniform width equal to v. As before, we now splice together three cubic polynomials in Z to add to those in X to obtain:

$$
\begin{aligned}
(9)\quad Y = S(X) &+ f_1(Z - Z_0) + g_1(Z - Z_0)^2 \\
&+ h_1(Z - Z_0)^3 + (h_2 - h_1)(Z - Z_1)^3 E_1^* \\
&\quad + (h_3 - h_2)(Z - Z_2)E_2^*
\end{aligned}
$$

As in (7), E_i^* is a dummy variable with value zero until Z reaches Z_i, and value 1 thereafter. S(X) represents (7) above, so (9) is a multiple regression of Y on a set of 10 composite variables. From there on, the regression procedure is straight forward.

The scheme is plainly generalizable. Additional variables can be included additively along with as many spline functions as desired, up to the limits of data availability and theoretical meaning.

Measures of Partial Correlation Significance Tests and Other Statistics

In what follows it is convenient to adopt a special notational convention: $Y = S(X) + u_1$ and $Y = T(Z) + u_2$ will represent regressions in which Y is approximated by, respectively, a spline function of X alone, and a spline function of Z alone, whereas $Y = S^*(X) + T^*(Z) + u_{12}$ will represent a regression like (10) in which Y is approximated by additive splines in X and Z.

Coefficients of partial determination associated with $S^*(X)$ and $T^*(Z)$ respectively can be calculated and their significance tested by comparing residual sums of squares as in an analysis of variance table:

	Sum of squares	d.f.	Mean squares
Residual from S(X)	Σu_1^2	$N-6$	
Residual from $S^*(X)+T^*(Z)$	Σu_{12}^2	$N-11$	$\Sigma u_{12}^2/(N-11)$
Contribution of $T^*(Z)$	$\Sigma u_1^2 - \Sigma u_{12}^2$	5	$(\Sigma u_1^2 - \Sigma u_{12}^2)/5$

Partial R^2 associated with $T^*(Z)$ is calculated as the ratio of the contribution of $T^*(Z)$ to the residual from S(X). Its significance is tested by the F ratio of the mean square of the contribution of $T^*(Z)$ to the mean square residual from $S^*(X) + T^*(Z)$. To examine the contribution of $S^*(X)$, merely interchange $S(X)$ and $T(Z)$ in the table.

Use of standard regression procedures to fit spline functions also makes it possible to apply Durbin-Watson tests for auto correlation of residuals in the usual way.

Example: Interest Rate as a Function of Money Supply and Inflation

As an example of application of the procedures, we explore the three-month Treasury bill rate (R) as a function of two variables: the ratio of money supply to GNP (M/GNP) both expressed in current dollars, and the rate of inflation (I), using quarterly data drawn from the period 1952-1970. Money supply is defined in the narrow sense as the sum of demand deposits and currency outside banks. Inflation rate was measured as the annual percentage increase in GNP deflator (P_t) taken over the preceding four quarters. That is, $I_t = 100(P_t - P_{t-4}) / P_{t-4}$. Since rising prices directly affect the dollar value of nominal GNP, the variable M/GNP already contains one aspect of inflation. In addition, however, past inflation rates influence expected future inflation rates and hence exert an independent influence over interest rates. To capture this effect, inflation rate is entered into the equation with a 1-quarter lag.

Because of the definition of I_t and its 1-quarter lag, we are left with 71 quarterly observations extending from the second quarter of 1953 through the last quarter of 1970. Over this period, M/GNP varied from a high of slightly over .36 to a low of nearly .20. This range was divided into three equal segments by knots established at M/GNP=.2011, .2551, .3091 and .3631, forming three intervals of uniform width equal to .0540. The range of I_{-1} was formed into intervals by knots established at I_{-1} = .273, 2.089, 3.905 and 5.721, with uniform interval equal to 1.816.

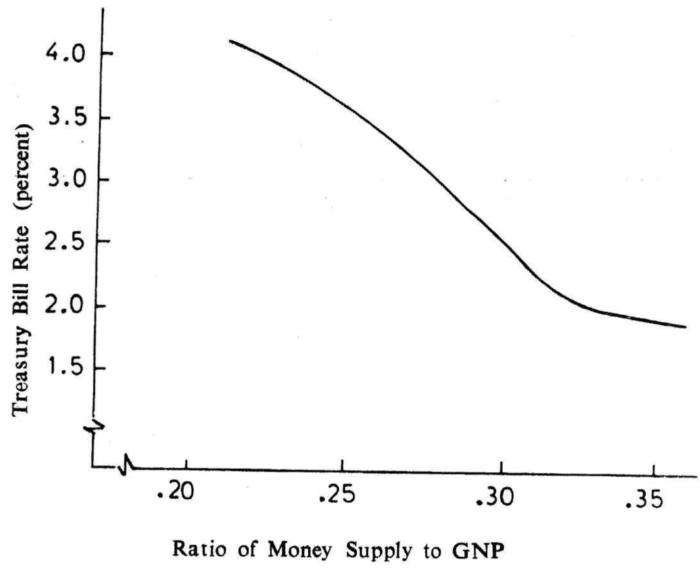
Additive spline functions in M/GNP and I_{-1} were formulated into a multiple regression as in (9). Composite variables were generated and the regression was fitted by TSP. If we represent the composite variables involving M/GNP by $x_1,\dots,x_5$, and those embodying I_{-1} by $z_1,\dots,z_5$, we can express the result as

$$(10)\quad R = 4.020 - 11.594x_1 + 68.846x_2 - 1684.86x_3 + 4180.47x_4 - 4224.56x_5 - 1.665z_1 + 1.611z_2 - .3454z_3 + .4327z_4 - .0516z_5.$$

Total R^2=.92 with F(10.65)=78.8, and DW=1.15. Partial R^2 for S(M/GNP) was .72; for $T(I_{-1})$, partial R^2 = .53; both were highly significant.

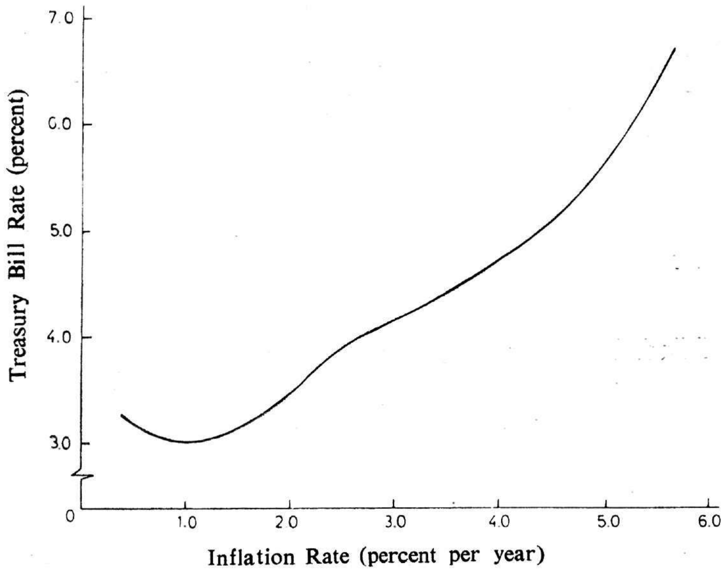
Calculated curves are depicted in Figures 3 and 4. Given the rate of inflation, the relationship of bill rate to M/GNP appears as in Figure 3. Figure 4 shows the bill rate as increasingly responsive to inflationary expectations as in-

FIGURE 3
TREASURY BILL RATE AS A FUNCTION OF M/GNP (INFLATION RATE=2 PERCENT)



flation rates rise, but what do we make of the little hook at the lower end of the function? The perverse response may, of course, be an accident of the data, and, in fact, when we tested the improvement of $R = S^*(M/GNP) + T^*(I_{-1})$ over the form $R = S(M/GNP) + kI_{-1}$, in which

FIGURE 4
TREASURY BILL RATE AS A FUNCTION
OF INFLATION (M/GNP=.25)



inflation is treated as an additive linear variable, we found the difference was not statistically significant. In any event, serious study of the relationship would entail much more careful representation of the dynamics of inflationary expectations than have been attempted here. In particular, it should be recognized that the relationship approximated by (10) is only one equation in a complex system in which interest rates, inflation rates, GNP, unemployment, and even—to a degree—the money supply itself are simultaneously determined. Nothing about the use of spline functions alters, in any way, the familiar problems of estimation that arise under these circumstances. In a careful research project, (10) would probably be fitted by two-stage least-squares or some other consistent method. This is readily done, for once the spline function is recognized as a regression the procedure is seen to be amenable to all the standard variations.

Modified Spline Functions

There is nothing about either the theory or the practice of piece-wise regression that requires that all individual segments be fitted by polynomials of the same degree, and for certain purposes it is useful to employ different degrees in different segments. An interesting example occurs in the relationship of fertility to per capita income as revealed by data from a cross-section of nations. Although there is good reason to suppose that fertility declines with rising income among very poor developing nations, it is also clear that once nations have reached a certain stage of development, fertility is likely to be little affected by further income increases. Among developed nations differences in fertility are primarily, if not entirely, the result of other factors, and any effort to relate them to income will yield spurious results.

To take advantage of the flexibility of the spline function at low income levels, and yet to avoid spurious relationships at high income, a spline function can be estimated under the additional constraint that its value be constant for incomes above a specified level. For this purpose, knots are established at x_0, x_1, x_2, and x_3 and the function is fitted in the form

$$
\begin{aligned}
(11)\quad Y = {} & [a_1 + b_1 (X - X_0) + c_1 (X - X_0)^2 \\
& + d_1 (X - X_0)^3]\, D_1 \\
& \quad + [a_2 + b_2(X - X_1) + c_2 (X - X_1)^2 \\
& + d_2 (X - X_1)^3\,]D_2 \\
& \quad + [a_3 + b_3(X - X_2) + c_3 (X - X_2)^2 \\
& + d_3 (X - X_2)^3\,]D_3 \\
& + a_4\, D_4
\end{aligned}
$$

where the X_i are equally spaced knots on the GNP per capita axis, and the D_i are defined as before, to have value 1 in the i-th interval and zero everywhere else. The value of D_4 is 1 for all values of GNP per capita that exceed X_3.

Although proper restriction of the coefficients of (11) would produce a continuous function with continuous first and second derivatives, it would impose an unnecessarily severe constraint on the spline function to force the second derivative of the cubic polynomial to zero at X_3. We therefore restrict the values of the coefficients of (11) to produce a continuous function with continuous first derivative throughout the range, but whose second derivative may be discontinuous at X_3. The result is still a smooth function, although not so "deeply" smooth as otherwise.

If, as before, we represent the uniform interval by w, the necessary constraints are

$$
(12) \qquad \begin{aligned}
a_1 + wb_1 + w^2c_1 + w^3d_1 &= a_2 \\
b_1 + 2wc_1 + 3w^2d_1 &= b_2 \\
c_1 + 3wd_1 &= c_2 \\
a_2 + wb_2 + w^2c_2 + w^3d_2 &= a_3 \\
b_2 + 2wc_2 + 3w^2d_2 &= b_3 \\
c_2 + 3wd_2 &= c_3 \\
-(1/3w^2)\,(b_3 + 2wc_3) &= d_3 \\
a_3 + wb_3 + w^2c_3 + w^3d_3 &= a_4
\end{aligned}
$$

When (12) is inserted into (11) and terms with like coefficients collected, we have, after some simplification

$$
\begin{aligned}
(13)\ Y = a_1 &+ b_1\{(X - X_0)(1 - D_4) - (1/3w^2)(X - X_2)^3 D_3 \\
&\quad + (8/3)wD_4\} \\
&+ c_1\{(X - X_0)^2 (1 - D_4) + (2/w)(X - X_2)^3 D_3 + 7w^2 D_4\} \\
&+ d_1\{(X - X_0)^3 (1 - D_4) - 9(X - X_2)^3 D_3 + 18w^3 D_4\} \\
&\quad + (d_2 - d_1)\{(X - X_1)^3 (D_2 + D_3) - 4(X - X_2)^3 D_3 \\
&\quad + 4wD_4\}
\end{aligned}
$$

As expressed in (13), the desired relationship becomes a multiple regression of Y on four composite variables. It is interesting to note that fewer composite variables are required to estimate the modified spline function than were needed without the modification. The reason is that although there is one more parameter to be estimated in (11) than in (4), the additional parameter is accompanied by two additional constraints, which reduces by one the number of coefficients to be estimated.

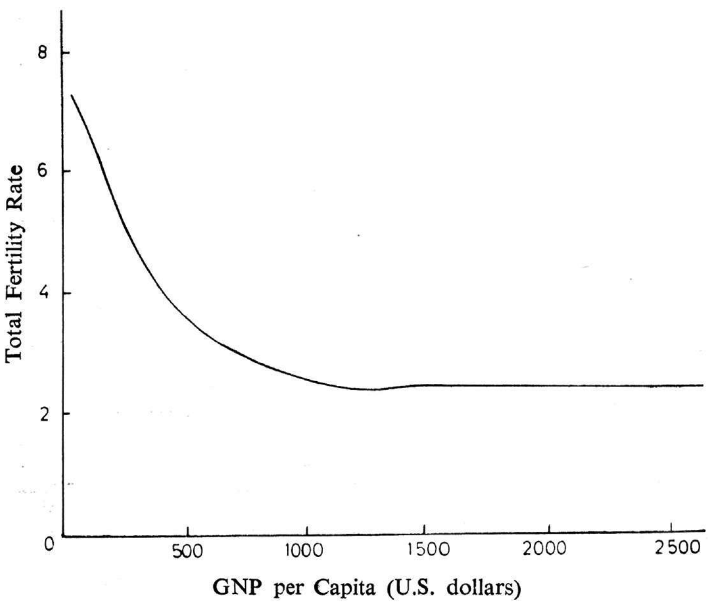
To estimate the coefficients of (13), we employed total fertility rates (TFR) for a sample of 59 nations. The fertility data were those collected for the project "Regression Estimation of Fertility" (U.S. National Institute for Child Health and Human Development Grant No. HD—09051) and were supplied to us by the principal investigator, Dr. James A. Palmore. Data for GNP per capita were taken from Suits (1973). Knots were established at values of GNP per capita equal to 0, \$500, \$1,000 and \$1,500, and four composite variables $x_1, \ldots, x_4$ were constructed as required.

Because of the strong influence of religion and culture on fertility, however, two additional variables were included in the regression: a dummy variable ISLAM that assumed the value 1 for all Islamic nations and zero for all others, and a second dummy variable SP that assumed the value 1 for all Spanish or Portuguese speaking nations. The composite variables were constructed and the regression was fitted using the SPSS package. The regression result was:

(14) $TFR = 8.395 - 21.624x_1 + 36.944x_2 - 24.880x_3 + 28.374x_4 + .834\,ISLAM + 1.287SP$. Total $R^2 = .72$, with $F(6,51) = 22.3$. Partial R^2s were for S(GNP/N), .56, for ISLAM, .05, for SP, .18. All were highly significant.

Calculated fertility rates for nations that were neither Islamic nor Spanish or Portuguese speaking are plotted in Figure 5. The chart

FIGURE 5
TOTAL FERTILITY RATE AS A FUNCTION OF GNP PER CAPITA



shows a sharp fertility decline among developing nations as incomes rise, but the rate of decline steadily diminishes, approaching zero near $1,000. It is interesting to note that fertility appears to become constant at income levels well below those at which the constraint becomes effective.

Concluding Remarks

It has been our purpose in this article to demonstrate how spline functions can be fitted with standard regression procedures. Although our examples have been limited to functions fitted to only three intervals, the same principles apply to any desired number.

Although it is a procedure of great promise, the spline function has important limitations. Such functions are most useful when data are uniformly distributed throughout the observed range. They work best when the scatter of observations is uniformly dense like those depicted in Figures 1 and 2 above. This uniformity of distribution maintains a uniformly close discipline over the function. Absence of this uniformity, that is, thin patches, sizeable gaps, or isolated points, reduces the discipline and the function is free to twist and squirm through the sparce parts of the data to yield spurious curvature.

As a second important limitation, spline functions are particularly ill-adapted to extrapolation beyond the observed range of data. As every student is warned, regression equations in general are unreliable guides to the world beyond the observed range to which they were fitted, but spline functions are especially poor in this regard. Properly speaking, indeed, the spline function is not even defined outside the range to which it has been fitted, unless, perhaps, under constraint of theoretical restrictions such as those applied to the modified spline. Because of their close ties to the observed range, spline functions may prove to be of limited usefulness in econometric models designed for forecasting and related purposes.

On the other hand, the spline function constitutes a simple and convenient way to approximate functions with complicated curvature. At its best the cubic spline approaches the statistical ideal of a regression technique that can approximate any relationship without prespecification of its mathematical form. As such, it will prove a powerful, yet relatively inexpensive addition to the tool box of standard statistical methods.

REFERENCES

BARTH, J., A. KRAFT and J. DRAFT, "Estimation of the Liquidity Trap Using Spline Functions," *Review of Economics and Statistics*, May 1976: 218-222.

BUSE, A. and L. LIM, "Cubic Splines as a Special Case of Restricted Least Squares," *Journal of the American Statistical Association*, March 1977: 64-68.

MCGEE, V. and W. CARLTON, "Piece-wise Regression," *Journal of the American Statistical Association*, September 1970: 1109-1124.

POITIER, D. J., "Piece-wise Regression Using Cubic Splines," *Journal of the American Statistical Association*, September 1973: 515-524.

——————, *The Econometrics of Structural Change*, North Holland Publishing Co., 1976.

SUITS, D. B., *Principles of Economics*, Harper and Row, 1973.

IN THE SAME SERIES

1. ANDREAS G. PAPANDREOU, *Planning Resource Allocation for Economic Development*, (1962).
2. FREDERIC E. BALDERSTON, *The Evolution of Management Sciences*, (1962).
3. ADAM A. PEPELASIS, *Les Problèmes de la Main d'Oeuvre en Grèce dans le Cadre du Marché Commun*, (1962).
4. GEORGE COUTSOUMARIS, *The Location Pattern of Greek Industry*, (1962).
5. MILTON FRIEDMAN, *Post War Trends in Monetary Theory and Policy*, (1963).
6. JAN TINBERGEN, *The Trend Towards Integration*, (1963).
7. ANNA KOUTSOYIANNI-KOKKOVA, *Production Functions for the Greek Industry*, (1963 - in Greek only).
8. ADAM A. PEPELASIS, *Labour Shortages in Greek Agriculture, 1963-1973*, (1963).
9. DIOMEDES PSILOS, *The Choice of Financial Assets in Greece*, (1963).
10. NICOS DEVLETOGLOU, *Montesquieu and the Wealth of Nations*, (1963).
11. ROBERT GORDON, *Some Thoughts on the Recent Slow Growth of the American Economy*, (1964).
12. PAN A. YOTOPOULOS, *The Elasticity of the Labour Supply Curve: A Theory and an Evaluation for Greek Agriculture*, (1964).
13. RICHARD M. WESTEBBE, *Saving and Investment in Greece*, (1964).
14. RALPH TURVEY, *Description of the Greek Tax System*, (1964 - in Greek only).
15. G. C. ARCHIBALD, *Investment and Technical Change in Greek Manufacturing*, (1964).

16. G. C. ARCHIBALD, *Industrialisation and Capital Requirements in Greece*, (1964).
17. JOHN HENRY MERRYMAN, *Some Problems of Greek Shoreland Development*, (1965).
18. KENNETH J. ARROW, *Statistical Requirements for Greek Economic Planning*, (1965).
19. A. FATOUROS, *Legal Aspects of the Problem of Attracting and Guaranteeing Foreign Private Investment*, (1965 - in Greek only).
20. GERHARD WEISSER, *Grenzen und Probleme der Planung*, (1965).
21. ROBERT EISNER, *Some Factors in Growth Reconsidered*,(1966).
22. JEROME ROTHENBERG, *An Approach to the Welfare Analysis of Intertemporal Resource Allocation*, (1967).
23. *The General Frame of Regional Development*, (1973 - in Greek only).
24. HELMUT REICHARDT, *Optimization Problems in Planning Theory*, (1971).
25. GEORGE D. DEMOPOULOS, *Monetary Policy, Unemployment and Wage Rates: The U.S. Postwar Experience*, (1971).
26. K. P. PRODROMIDIS, *Forecasting Aggregate Consumer Expenditure in Greece: A Long-Run Analysis*, (1973).
27. CARL S. SHOUP, *The Value-Added Tax*, (1973).
28. THEODORE P. LIANOS *and* KYPRIANOS P. PRODROMIDIS, *Aspects of Income Distribution in Greece*, (1974).
29. PANAYOTIS G. KORLIRAS, *The Economics of Disorder*, (1977).
30. LAWRENCE R. KLEIN, *Project LINK*, (1977).
31. GEORGE N. PARASKEVOPOULOS, *An Econometric Analysis of International Tourism*, (1977).