Notes on Piecewise Continuous Regression

David M. Boore

From Section 15.4 in Press et al. Numerical Recipes (Fortran), use this equation for fitting data:

$$y(x) = \sum_{k=1}^{M} a_k F_k(x) \tag{1}$$

where the basis functions $F_k(x)$ can be highly nonlinear functions of x. Consider multisegment regression (piecewise continuous) for a set of breakpoints (also called hingepoints, as well as other things) x_j. Equation (1) describes the functions to be fit to the data within each segment, but with N_{segs} segments, equation (1) needs to be extended to

$$y(x) = \sum_{j=1}^{N_{segs}} \sum_{k=1}^{M_j} a_{j,k} F_{j,k}(x) \tag{2}$$

So in each segment j there are M_j regression coefficients. For example, for a cubic polynomial, there would be three functions: linear, quadratic, and cubic, each with a regression coefficient describing the strength of each term in the polynomial. As described below, continuity of the functions from segment to segment requires that the functions in each segment start with a value of 0.0 and that they be continuous, with no offsets.

How are the basis functions chosen such that there is continuity from segment to segment? The basis functions for each segment are defined for all values of x, but are constant on either side of the segment boundaries, as specified in these equations:

$$F_{j+1,k}(x) = 0,\ x \le x_j \tag{3}$$

(assume that breakpoint j is to the right of the jth segment, so, for example, breakpoint 1 is between segments 1 and 2),

$$F_{j+1,k}(x),\ x_j \le x \le x_{j+1} \tag{4}$$

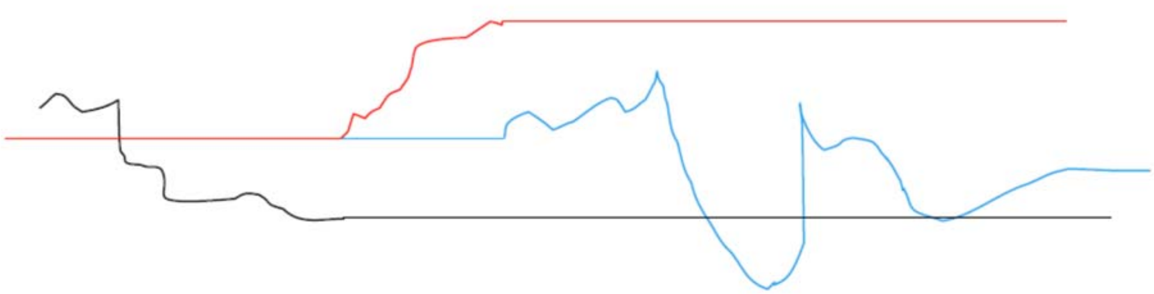
where $F_{j+1,k}$ is a specified arbitrary function that requires only one regression coefficient (not including an intercept term, which is taken care of in the call to the regression function). For example, $F_{j,1} = x - x_{j-1}$, $F_{j,2} = \left(x - x_{j-1}\right)^2$, and $F_{j,3} = \left(x - x_{j-1}\right)^3$ would be the three basis functions needed to describe a cubic polynomial in segment j.

This restriction follows from how R uses functions in regression routines. See *piecewise-continuous_regression_using_basis_functions.R.examples.16jul23.pdf* for a number of examples (NOTE: this file might have been updated; if so, "16jul23" will be replaced with the new date). To the right of breakpoint x_{j+1} the function is a constant with the value at the breakpoint at the right side of segment $j + 1$:

$$F_{j+1,k}(x) = F_{j+1,k}(x_{j+1}),\ x_{j+1} \leq x \tag{5}$$

Note that the basis function is thus 0 before breakpoint x_j and constant after breakpoint x_{j+1}. This is what guarantees continuity, as will be shown below.

Here is a rough sketch of basis functions for three segments (two breakpoints), with one basis function in each segment (in this made-up example, the amplitude, but not the shape, of each basis function will be determined in the regression).:



To demonstrate that the above formulation achieves continuity, consider a three-segment regression. According to equation (2), for any x

$$y(x) = a_1F_1(x) + a_2F_2(x) + a_3F_3(x) \tag{6}$$

where the basis-number subscript k has been dropped (because there is only one basis function in each segment in this example). Consider the value of y on either side of breakpoint x_1: just the left of the breakpoint, at $x = x_1 - \epsilon$, where ϵ is a small number, both F_2 and F_3 are 0.0, from equation (3), and so

$$y(x_1 - \epsilon) = a_1F_1(x_1 - \epsilon). \tag{7}$$

And just to the right of the breakpoint,

$$y(x_1 + \epsilon) = a_1F_1(x_1 + \epsilon) + a_2F_2(x_1 + \epsilon) \tag{8}$$

But from equation (3), $F_2(x_1) = 0$, so there is no discontinuity at the breakpoint as ϵ approaches 0.0. Now consider the same thing for breakpoint x_2. To the left of breakpoint x_2

$$y(x_2 - \epsilon) = a_1 F_1(x_2 - \epsilon) + a_2 F_2(x_2 - \epsilon) \tag{9}$$

and to the right

$$y(x_2 + \epsilon) = a_1 F_1(x_2 + \epsilon) + a_2 F_2(x_2 + \epsilon) + a_3 F_3(x_2 + \epsilon) \tag{10}$$

But $F_3(x_2) = 0$, so again there is continuity at breakpoint x_2. This is true no matter how many segments are in the regression.

As demonstrated in *piecewise-continuous_regression_using_basis_functions.R.examples.16jul23.pdf*, flat segments do not require basis functions.

Finally, what are the basis functions for a model that is flat to x=L, constrained to cross the zero axis at x = xc, and then is flat beyond x=R (where L<xc<R)? Thinking about it, there is only one basis function, even though it has breakpoints at x=L and x=R. The reason is that the slope of the line between x=L and x=R is determined by the value of the constant portion for x < L and the condition that the line crosses the zero line at x=xc. In addition, the level of the flat portion for x > R is determined by the level for x <L and by the slope, which is NOT a regression parameter. There is only one basis function, even though it has two breakpoints, and fitting that basis function to data requires only one regression parameter. The document *piecewise-continuous_regression_using_basis_functions.R.examples.16jul23.pdf* contains an example of such a function.