Specifying a Natural Cubic Spline

Christian A. Maino Vieytes

May 9, 2023

Background

Using splines in regression models is a popular method for flexibly modeling exposure-outcome relationships in epidemiological studies.[1,2] Splines in a regression context can be conceptualized as a series of local polynomials fit over the domain of the regression function. It provides a nice alternative to a categorical analysis of dose response, which has its limitations. One of those key limitations is the assumption that the response is uniform over all observations within a category for which a parameter estimate is made, resulting in the appearance of a step-wise function for the relationship between exposure and response.[3] In contrast, fitting a model with spline terms is a parametric technique that generates a smooth curve, allowing the user to visualize the dose-response relationship. The parameter estimates are rarely of interest, owing to their lack of interpretability.[3] However, given that a model specifying the exposure as a continuous variable is nested in a model with an expansion of spline terms for that variable, the departure from a linear relationship can be formally tested with a Likelihood Ratio Test.[2]

Splines

Fitting spline models involves applying a set of transformations to the exposure variable, X. In the regression model, we replace the original variable with the set of transformations (which we term *basis functions*) and estimate parameters for each of those transformations:

$$Y = \sum_{m=1}^{M} \beta_m h_m(X) + \gamma V$$

which is a linear basis expansion in X, where h_m is the mth basis function (of which there are M), and γV is a term for an additional covariate we may want to adjust for (no basis expansion on this term).[4] We generalize this approach to the multivariable setting where we desire basis expansions in several variables:

$$Y = \sum_{j=1}^{J} \sum_{m=1}^{M_j} \beta_{jm} h_{jm}(X_j) + \gamma V$$

where j is the index of variables. The set of fixed basis functions is of interest as are the *knots* and *knot locations*, which are the locations along the domain where separate local piecewise polynomials are fit. In order to achieve smoothness in the spline curve across these disjoint regions, a continuity constraint is applied. Using a truncated power basis, we achieve the continuity constraint:

$$(x - \xi_\ell)_+^g = \begin{cases} (x - \xi_\ell)^g, & \text{if } x > \xi_\ell \\ 0, & otherwise \end{cases}$$

where ξ_ℓ is the ℓth knot and g is the order of the polynomial we choose to fit. We arrive at the number of basis functions by beginning with basis functions for the degree of the polynomial we desire and then include one truncated power basis function for for each knot we specify (an additional basis function where we apply a constant–i.e., 1–can also be included for the intercept).[5] For a cubic spline with two interior knot, we use the following basis functions:

$$h_1 = 1, h_2 = x, h_3 = x^2, h_4 = x^3, h_5 = (x - \xi_1)_+^3, h_6 = (x - \xi_2)_+^3$$

It is shown that for a degree g polynomial, the basis functions will results in a curve that is continuous (i.e., smooth) up to the $(g - 1)$ derivative.[5] For the cubic spline basis representation above, we have that the function and the first and second derivatives will be continuous at the knot boundaries.

```r
#Calculate CE for each counterparty
Value.A <- data.frame() #MTM value of each contract within cp A
# ...
for(i in 1:length(foo)){
  if(isTrue(as.character(portfolio_data[i,1])=="A")==TRUE){
    # ...
  }
}
```

References

1. Greenland, S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. eng. *Epidemiology (Cambridge, Mass.)* **6,** 356–365. doi:10.1097/00001648-199507000-00005 (July 1995).

2. Witte, J. S. & Greenland, S. A nested approach to evaluating dose-response and trend. en. *Annals of Epidemiology* **7,** 188–193. doi:10.1016/S1047-2797(96)00159-7 (Apr. 1997).

3. Steenland, K. & Deddens, J. A. A Practical Guide to Dose-Response Analyses and Risk Assessment in Occupational Epidemiology. *Epidemiology* **15.** Publisher: Lippincott Williams & Wilkins, 63–70 (2004).

4. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* ISBN: 978-0-387-84857-0 978-0-387-84858-7. doi:10.1007/978-0-387-84858-7 (Springer New York, New York, NY, 2009).

5. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning* ISBN: 978-1-4614-7137-0 978-1-4614-7138-7. doi:10.1007/978-1-4614-7138-7 (Springer New York, New York, NY, 2013).