Text Version

Expand Print

Modeling with Splines

This demonstration shows the basics of incorporating complex nonlinear relationships into a model by using splines. Working with splines is complex, and gaining a full understanding of working with splines is beyond the scope of this course. For additional details about working with splines, see the SAS documentation.

In the exploratory analysis of the cars data, we used PROC SGPLOT to fit a penalized B-spline curve to a scatterplot of **FuelTank** versus **Price**. This plot, which is shown here, indicates that the variable **FuelTank** has a nonlinear relationship with **Price**. Previous models incorporated **FuelTank** using a cubic polynomial. However, a spline might provide a better approximation of the relationship between **FuelTank** and **Price**.

This code compares two models--a cubic polynomial model and a spline model. Each model has only one effect and selection is turned off by specifying SELECTION=NONE.

In the first PROC GLMSELECT step, the EFFECT statement defines the third-degree polynomial effect **P_Fuel**, which is based on **FuelTank**. To deal with any possible multicollinearity, this effect is centered, as shown in previous demonstrations.

In the second PROC GLMSELECT step, the EFFECT statement defines the spline effect **Sp_Fuel**, which is also based on **FuelTank**. The only option specified in this statement is DETAILS, which controls the output. However, other options are available for spline effects, as shown in SAS documentation.

We run the code.

```
title 'Spline Effect with Cars Dataset';

proc glmselect data=mydata.cars2;
   title2 'Cubic polynomial for Fueltank';
   effect p_fuel = polynomial(fueltank /degree=3
            standardize(method=moments)=center);
   model price = p_fuel / selection=none;
run;

proc glmselect data=mydata.cars2;
   title2 'Spline for Fueltank';
   effect sp_fuel = spline(fueltank / details);
   model price = sp_fuel / selection=none;
run;
```

In the results for the cubic polynomial model, we look at the values in the last column of the Parameter Estimates table. The model with a cubic polynomial effect for **FuelTank** has three predictor terms—of degrees 1, 2, and 3, respectively. The cubic term is significant. The squared term is not nearly as significant, but with a significant cubic term, that's fine. Above the Parameter Estimates table, we see the model selection statistics. The adjusted R square is 0.6000.

In the results for the spline model, the Knots for Spline Effect Sp_Fuel table lists nine knots. Remember that knots are the places where the piecewise polynomials are joined. The Least Squares Summary table shows the SBC for **Sp_Fuel**. You could use this value to compare this model to other spline models. In the last column of the Analysis of Variance table, you can see that this is a significant model. Shown in the Parameter Estimates table, the parameter estimates for a spline effect are not easy to interpret. For one thing, spline effects are prone to overfitting. Using options to specify the number and location of knots can improve your results.

This spline model has an adjusted R square of 0.6320, which is slightly higher than the R square of the cubic polynomial model, which was 0.6000. So, the spline model appears to be a slightly better model.

Close