STAT40810 — Stochastic Models

Brendan Murphy

Week 5

Generalized Additive Models: Extensions

Extending the GAM

- The generalized addtive model (GAM) has the form

$$Y_i = \alpha + g_1(X_{i1}) + g_2(X_{i2}) + \cdots + g_K(X_{iK}) + \epsilon_i,$$

  where Y_i is the response variable and the covariates are $X_i = (X_{i1}, X_{i2}, \ldots, X_{iK})$.
- The model is somewhat restricted because the effect of each covariate is independent of the values of the others.
- This is analagous to fitting a linear regression model with no interactions.

Thin Plate Splines

- We can extend the generalized addtive model (GAM) to allow for terms that depend on two (or more covariates), that is, we can add extra terms that are of the form

$$g_{kl}(X_{ik}, X_{il}), \quad g_{klm}(X_{ik}, X_{il}, X_{im}), \quad \text{etc}$$

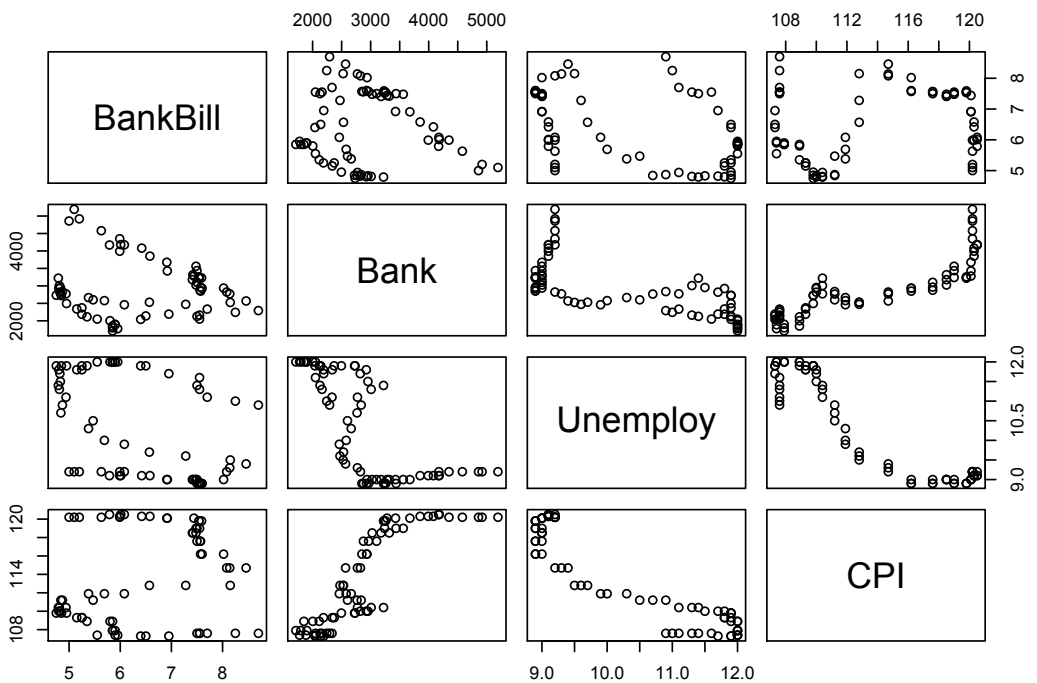
  so that terms in the model can interact.
- The form of the function $g_{kl}(\cdot, \cdot)$ is no longer a cubic spline but something called a *thin plate spline*.
  We won't go into these in detail, but they are essentially an analogue of splines in two or more dimensions.
- The name *thin plate spine* comes from the fact that the functions also arise in physics when modeling the deformation of a thin plate of metal.

Example: 90 Day Bank Bill Rate

- Data from a bank were collected that record:

BankBill	90 Day Bank Bill Rate
Bank	Bank Share Price
Unemploy	Unemployment Rate
CPI	Consumer Price Index

- We wish to study the effect of the covariates on BankBill.



Code: 90 Day Bank Bill Rate

```
# Read in the data
# (you might need to set the working directory)

bankbill <- read.table("bankbill.txt",header=TRUE)

# Fit a generalized additive model
fit0 <- gam(BankBill~s(Bank)+s(Unemploy)+s(CPI),data=bankbill)

# Add interaction term between CPI and Unemploy
fit1 <- gam(BankBill~s(Bank)+s(Unemploy)+s(CPI)+s(CPI,Unemploy),data=bankbill)

# Compare the models
anova(fit0,fit1)
```

Results

- The model without interactions is doing very well

```
Formula:
BankBill ~ s(Bank) + s(Unemploy) + s(CPI)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.4182     0.0305   210.4   <2e-16

(Intercept) ***
---
Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Approximate significance of smooth terms:
              edf Ref.df     F  p-value
s(Bank)     2.906  3.708 15.28 2.11e-07 ***
s(Unemploy) 6.978  7.971 20.04  < 2e-16 ***
s(CPI)      6.576  7.565 15.60 2.57e-14 ***
---
Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

R-sq.(adj) =  0.951   Deviance explained = 96.3%
GCV = 0.087605  Scale est. = 0.066061  n = 71
```

Extension

- We can see that the model doesn't improve much when the interaction is added.

```
Analysis of Deviance Table

Model 1: BankBill ~ s(Bank) + s(Unemploy) + s(CPI)
Model 2: BankBill ~ s(Bank) + s(Unemploy) + s(CPI) + s(CPI, Unemploy)
  Resid. Df Resid. Dev     Df Deviance
1    50.756     3.5369
2    38.012     0.9588 12.744    2.578
```

- It is worth noting that the interaction term uses requires a large number of effective parameters (and uses up many degrees of freedom).