RSS ordinary meeting 15 December 1998
— contribution to discussion of paper by Verbyla, et al.

Peter Green (University of Bristol). I welcome this paper for encouraging the use of splines in statistical modelling; there is little justification for the suspicion about splines shown elsewhere in this discussion. Maybe it helps to remark that cubic smoothing splines provide a continuum of models from a trend linear in t, to treating t as a factor (obtained as $\lambda \to \infty$ and 0 respectively) — nothing very alarming there!

However, I do have misgivings about some of the methodology proposed. Formulating spline smoothing as a mixed model is simply a mathematical device; the suggested logical distinction between the *fixed* linear trend and the *random* smooth variation about it is artificial. Thus I would not freely adopt random effects methodology in this context.

It is of interest and pedagogical value to recall that the mixed model formulation applies much more widely, including to more traditional models not involving splines.

A rather general linear model for data y in terms of treatment effects τ of interest can be written

$$\mathrm{E}(y) = D\tau + X\beta \quad ; \quad \mathrm{var}(y) = \sigma^2 V(\lambda)$$

where β quantifies appropriate linear trends, blocks, etc., and $V(\lambda)$ models one or more variance components.

Compare this to the additive decomposition

$$y = D\tau + g + e$$

(cf. equation (5)), where $g \sim (X\beta, \sigma^2(V - I))$ and $e \sim (0, \sigma^2 I)$. Evidently these define the same mean and variance for y, hence the same GLS estimates $\widehat{\tau}$, solving

$$D^T(I - S)(y - D\widehat{\tau}) = 0,$$

where

$$S = I - V^{-1} + V^{-1}X(X^TV^{-1}X)^{-}X^TV^{-1}.$$

Even without mention of splines, S has characteristics of a smoothing matrix — eigenvalues in $[0, 1]$, but not generally a projection. For a traditional example, in a resolvable incomplete blocks design S is a convex combination of the replicate and block projection operators.

GLS estimates of τ and g also arise as the solution of the intuitively plausible simultaneous equations

$$\begin{aligned} \hat{\tau} &= (D^T D)^{-1} D^T (y - \hat{g}) \\ \hat{g} &= S(y - D\hat{\tau}), \end{aligned}$$

and indeed iterating between these defines the backfitting algorithm. (This reminds us that the S-plus function `gam()` is capable of fitting the models in this paper.)

When does this smoothing interpretation derive from penalised least squares, as it does for the authors? For any variance components setup

$$V(\lambda) = I + \sum_r \lambda_r^{-1} Z_r G_r Z_r^T,$$

the same GLS estimates result from minimising

$$e^T e + \sum_r \lambda_r u_r^T G_r^{-1} u_r$$

subject to

$$y = D\tau + X\beta + \sum_r Z_r u_r + e,$$

(cf. equation (10)). Here $\{u_r\}$ are multiple random effects with $\text{var}(u_r) = \sigma^2 \lambda_r^{-1} G_r$.

Most of this, which extends to general R, not just $R = I$ as assumed above, can be found in Green (1985); this also uncovers intimate relationships between methods for estimating the variance ratios λ — including GCV, REML and Yates' recovery of interblock information.

Additional reference

Green, P. J. (1985) Linear models for field trials, smoothing and cross-validation. *Biometrika*, **72**, 523–537.