Generalized additive (mixed) models

29 Nov 2024

```
library(mgcv)
library(gratia)
library(tidyverse)
```

Additive models

- generally a way to specify more complex (smooth) terms based on *individual* covariates: $\mu = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots$
- lots of ways to generate $f_i(x_i)$: kernel estimators, locally weighted polynomials, … see T. J. Hastie and Tibshirani (1990), T. Hastie, Tibshirani, and Friedman (2009) (*backfitting algorithm* etc.)
- we will focus on the approach of Simon N. Wood (2017), which is in some ways more restricted (everything is done explicitly via bases + latent Gaussian variables)

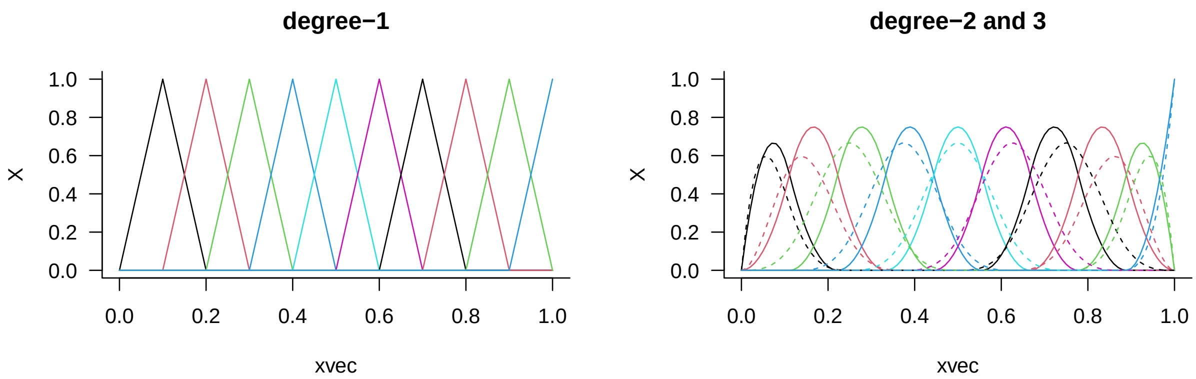
Basis expansions

- (theoretically) infinitely expandable
- e.g. polynomials ('regular'/raw, orthogonal, Legendre, Hermite)
- wavelet, Fourier
- splines: **piecewise polynomial** with continuity/smoothness constraints

Spline degree

```
xvec <- seq(0, 1, length.out = 101)
sfun <- function(d = 3, type = c("bs", "ns", "rcs"), off = 1e-5, lty = 1, ...) {
    type <- match.arg(type)
    X <- switch(type,
```

```
                 bs = splines::bs(xvec, df = 10, degree = d),
                 ns = splines::ns(xvec, df = 10), ## only cubic
                 rcs = rms::rcs(xvec, 10) + off
                 )
    par(bty = "l", las = 1)
    matplot(xvec, X, type = "l", lty = lty, ...)
}
par(mfrow=c(1,2))
sfun(d = 1, main = "degree-1")
sfun(d = 2, main = "degree-2 and 3")
sfun(d = 3, add = TRUE, lty = 2)
```



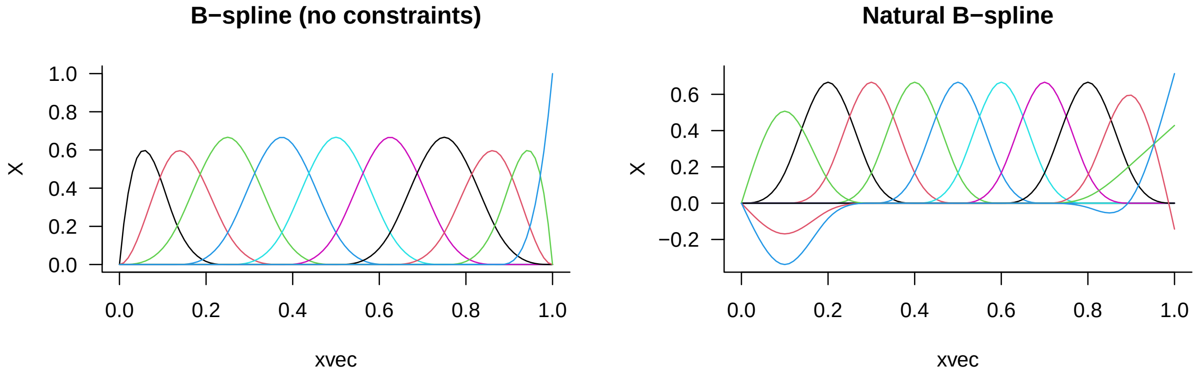
spline terminology

- **knots**: breakpoints (boundary, interior)
- order-M (ESL): continuous derivatives up to order $M-2$ (cubic, $M=4$)
- typically $M=1,2,4$
- number of knots = df (degrees of freedom) -1 -intercept

Spline choices

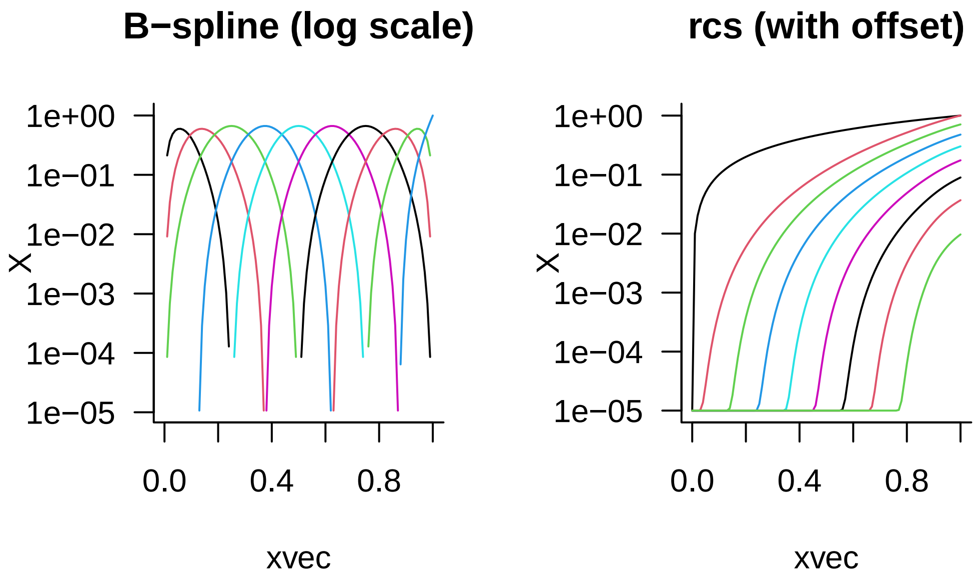
- continuous derivatives up to $d-1$
- truncated polynomial basis (simple)
- B-splines: complex, but *minimal support*/maximum sparsity
- *natural* splines: extra constraint, derivatives >1 vanish at boundaries

```
par(mfrow = c(1,2))
sfun(main = "B-spline (no constraints)")
sfun(type = "ns", main = "Natural B-spline")
```



Truncated polynomial vs B-spline

```
par(mfrow=c(1,2))
sfun(main = "B-spline (log scale)", log = "y")
sfun(type = "rcs", log = "y", main = "rcs (with offset)")
```



choosing knot locations

- generally not that important: evenly spaced, *or* evenly spaced based on quantiles

choosing basis dimension

- in principle could expand dimension to match total number of points (*interpolation spline*)
- ... but that would overfit
- AIC, adjusted R^2, cross-validation ...

smoothing splines

- as many knots as data points
- plus squared-second-derivative ("wiggliness") penalty

$$\text{RSS} + \lambda \int (f''(t))^2 \, dt$$

- defined on an infinite-dimensional space
- minimizer is a (natural?) cubic spline with knots at x_i

$$(\mathbf{y} - \mathbf{Zb})^\top (\mathbf{y} - \mathbf{Zb}) + \lambda \mathbf{b}^\top \Omega \mathbf{b}$$

with $\{\Omega\}_{jk} = \int \mathbf{Z}_j''(t) \mathbf{Z}_k''(t) \, dt$

- **generalized** ridge regression: penalize by $\lambda \Omega_N$ rather than λI
- could use same data augmentation methods as before except that now we use $\sqrt{\lambda} C$ where C is a matrix, and the "square root" (Cholesky factor) of Ω_N

See Simon N. Wood (2017), Perperoglou et al. (2019)

generalized cross-validation score

- very close to AIC
- Larsen (2015), Golub, Heath, and Wahba (1979)
- minimize $\text{RSS}/(\text{Tr}(\mathbf{I} - \mathbf{S}(\lambda)))^2$, where S is "a rotation-invariant version of PRESS [predicted residual error sum of squares]" ($\sum (e_i/(1-h_{ii}))^2$)
- replace RSS with approximation of deviance,

$$||\sqrt{\mathbf{W}}(\mathbf{z} - \mathbf{X}\beta)||^2$$

  for generalized (non-Gaussian) models

connection to mixed models

- note that $\lambda \mathbf{b}^\top \Omega \mathbf{b}$ is equivalent to $(1/\sigma^2)\mathbf{b}^\top \Sigma'^{-1} \mathbf{b}$; if Σ' is a *scaled* covariance matrix (i.e. $\Sigma = \sigma^2 \Sigma$), then this is the core of the MVN log-likelihood $\log L(\mathbf{b}|\Sigma)$ (all we're missing is a factor of $\text{Det}(\Sigma)^{-1/2}$ and a normalization constant)
- So we can fit this with any of the mixed model machinery, provided we can set up the correct covariance matrix

ML criterion, REML criterion

- treat spline smoothing as a *mixed model* problem
- spline (penalized) parameters are $\mathbf{b}$
- $y|u \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \sigma^2 \mathbf{I})$; $\mathbf{b} \sim N(0, (\sigma^2/\lambda)\mathbf{W}^{-1})$ where the $\mathbf{W}$ is the penalty matrix
- corresponds to minimizing $||\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}||^2 + \lambda \mathbf{b}^\top \mathbf{W} \mathbf{b}$
- REML: "fixed effects are viewed as random effects with improper uniform priors and are integrated out" (Wood 2011)
- Laplace approximation
- slower but generally preferred now

practical stuff

- Simon Wood is insanely smart, and `mgcv` is insanely powerful and flexible
- gratia package (named after Grace Wahba
- available 'smooths' (bases + penalty terms): look for strings of the form `smooth.construct.*.smooth.spe`
- although you can *theoretically* have as many knots as data points, fewer is often good enough/computationally efficient

Available bases (using `apropos("smooth.construct")`):

```
 [1] "ad"     "bs"     "cc"     "cp"     "cr"     "cs"     "ds"     "gp"
 [9] "mrf"    "ps"     "re"     "sf"     "so"     "sos"    "sw"     "t2"
[17] "tensor" "tp"     "ts"
```

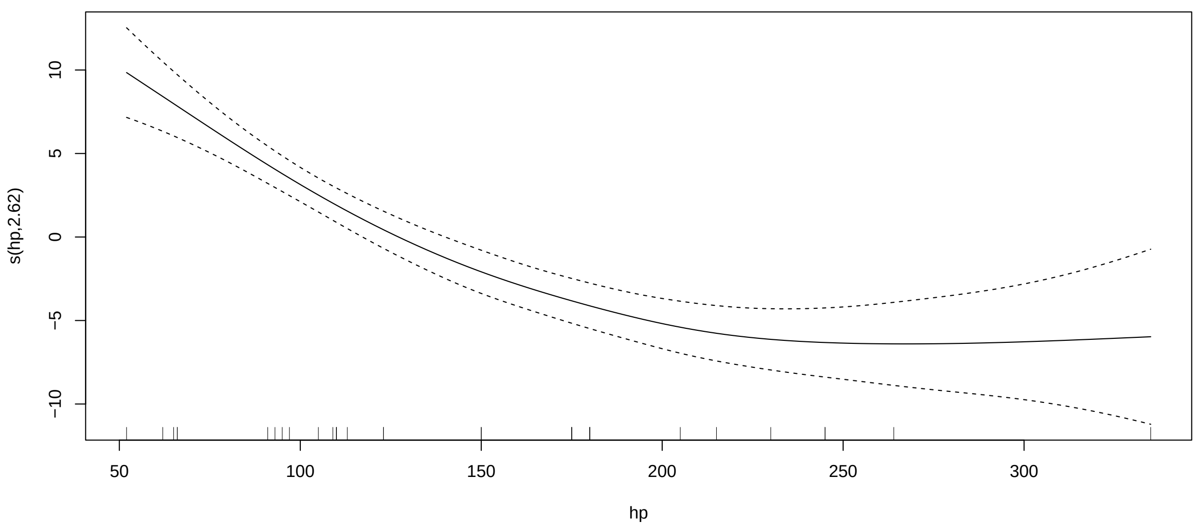
```r
g1 <- gam(mpg ~ s(hp), data = mtcars)
summary(g1)
```

```
Family: gaussian
Link function: identity
```

```
Formula:
mpg ~ s(hp)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.0906     0.5487   36.62   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
        edf Ref.df     F p-value
s(hp) 2.618  3.263 26.26  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.735   Deviance explained = 75.7%
GCV = 10.862  Scale est. = 9.6335    n = 32
```

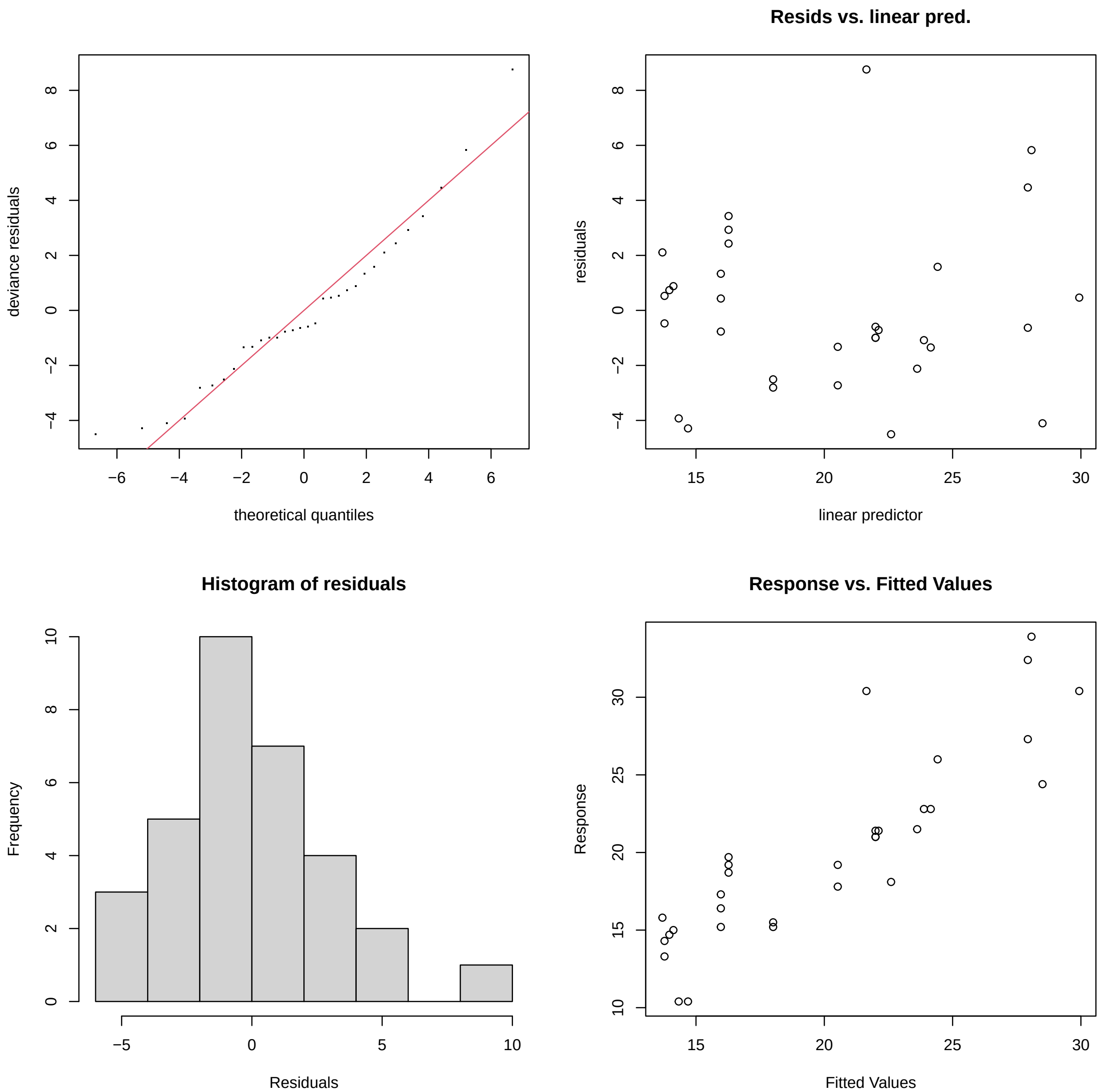
Plot:

```
plot(g1)
```



Check:

```
gam.check(g1)
```

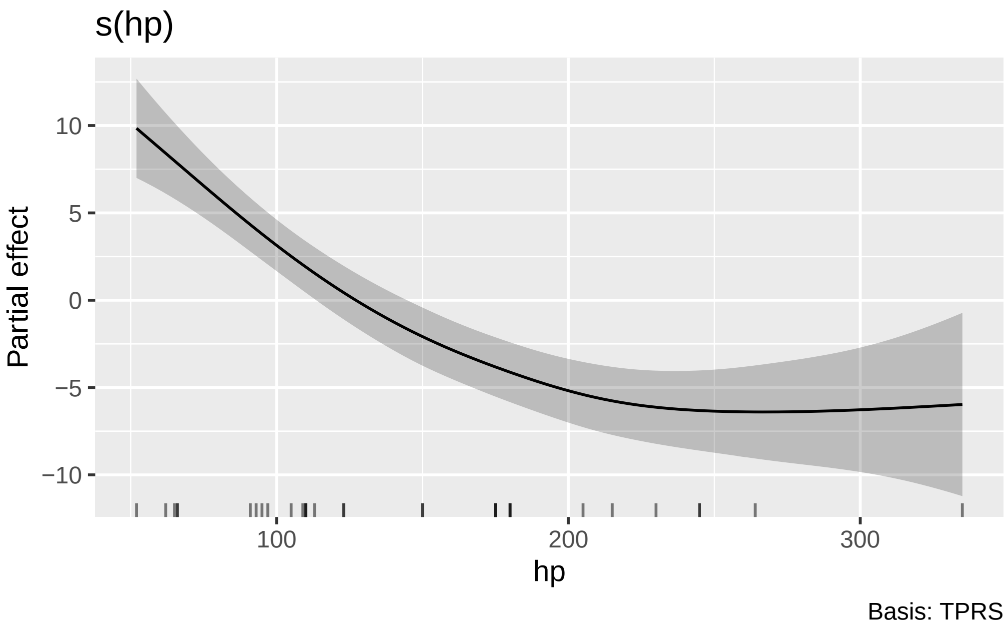



```
Method: GCV   Optimizer: magic
Smoothing parameter selection converged after 4 iterations.
The RMS GCV score gradient at convergence was 4.290111e-05 .
The Hessian was positive definite.
Model rank =  10 / 10

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.
```

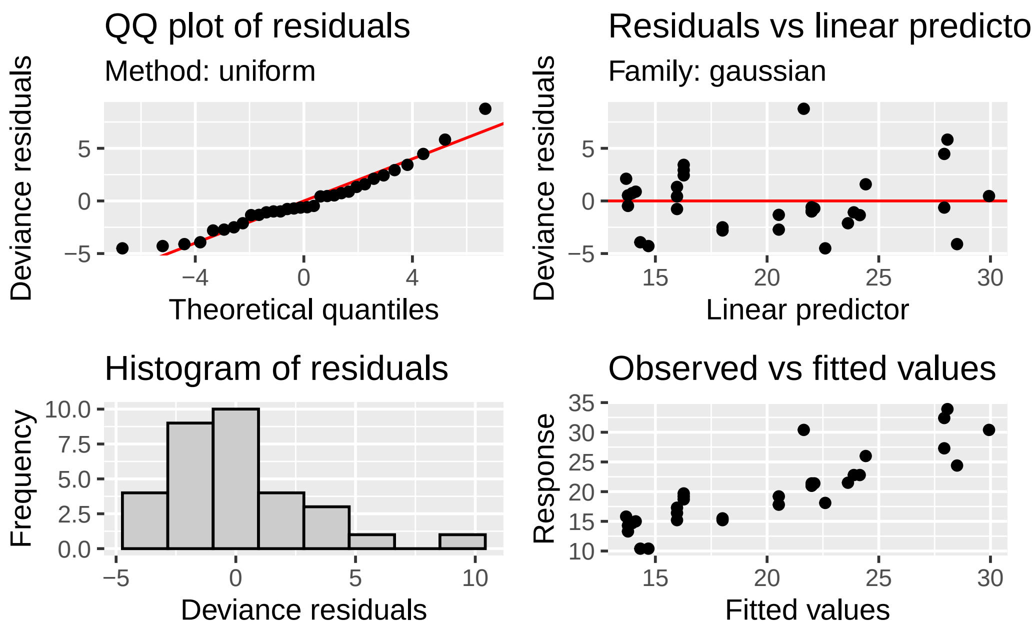
```
        k'   edf k-index p-value
s(hp) 9.00 2.62    0.87    0.18
```

The `gratia` package has prettier versions:

```r
draw(g1)
```

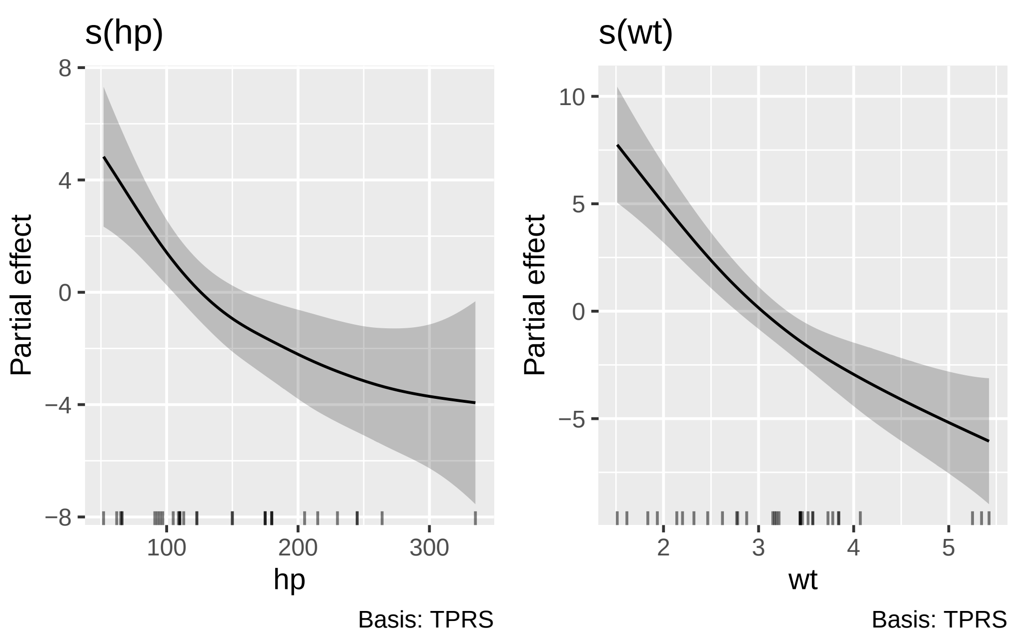



```r
appraise(g1)
```



concurvity (analogous to "collinearity"): [CV question](), Ramsay, Burnett, and Krewski (2003); rule of thumb is that a value of (0.3? 0.5? 0.8?) suggests trouble …

```r
g3 <- gam(mpg ~ s(hp) + s(wt), data = mtcars)
draw(g3)
```

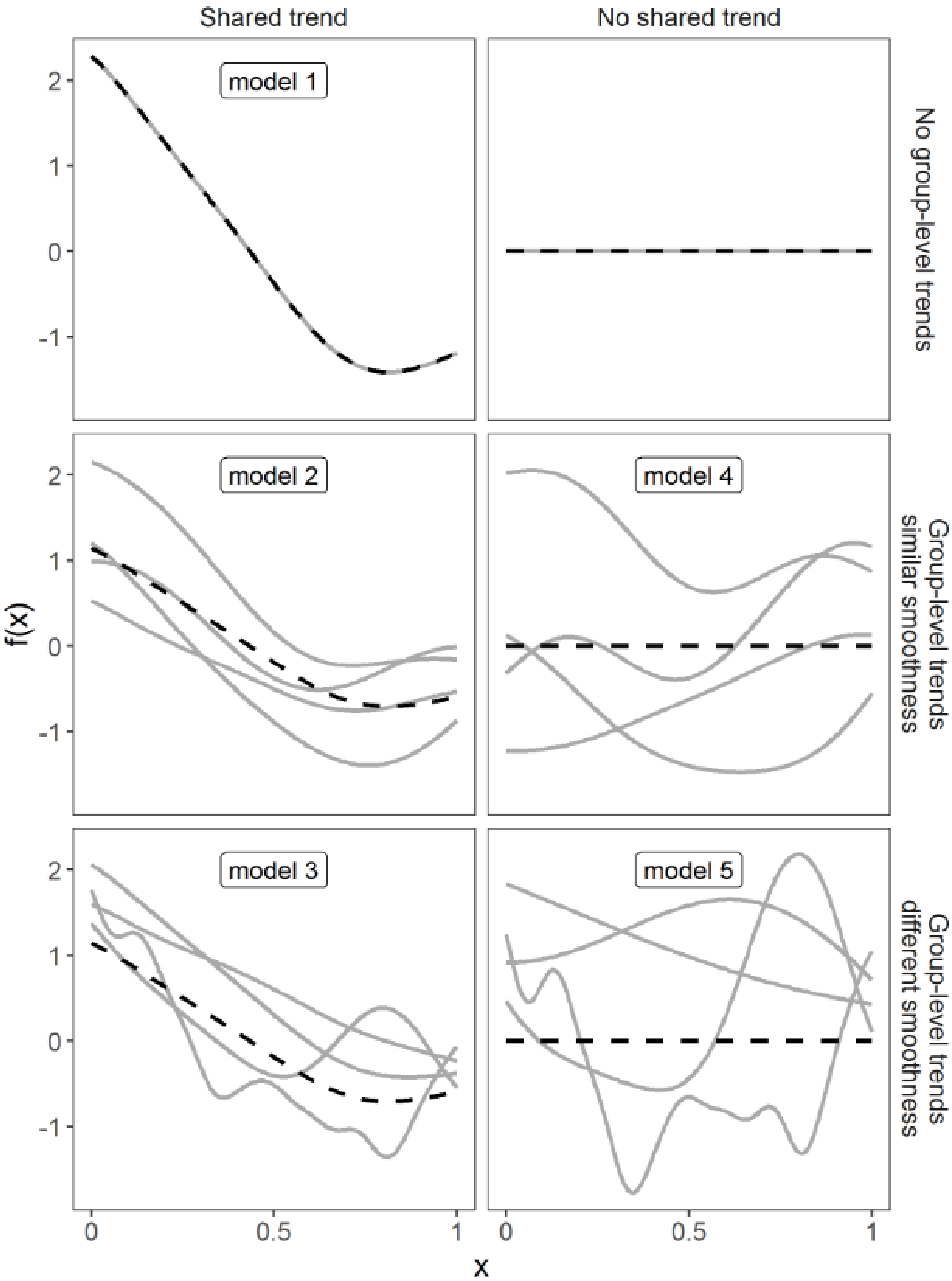


```r
concurvity(g3)
```

```
                 para     s(hp)     s(wt)
worst    2.534415e-18 0.9575001 0.9575001
observed 2.534415e-18 0.6883189 0.6968782
estimate 2.534415e-18 0.4784240 0.7978191
```

Many options: simple random effects (`bs = "re"`); *cyclic* splines (make $x(0) = x(T)$; `bs="cc"`) ; multidimensional splines (thin-plate, *tensor product* (`te()`); spherical (*Duchon*) splines (`bs = "sos"`); Markov random fields (`bs = "mrf"`); Gaussian processes (`bs = "gp"`); splines by category (`by=` argument); constrained splines (`scam` package, Pya and Wood (2015)); *soap film* splines; etc etc etc etc ...

The paper by Pedersen et al. (2019) on **hierarchical splines** is especially important.

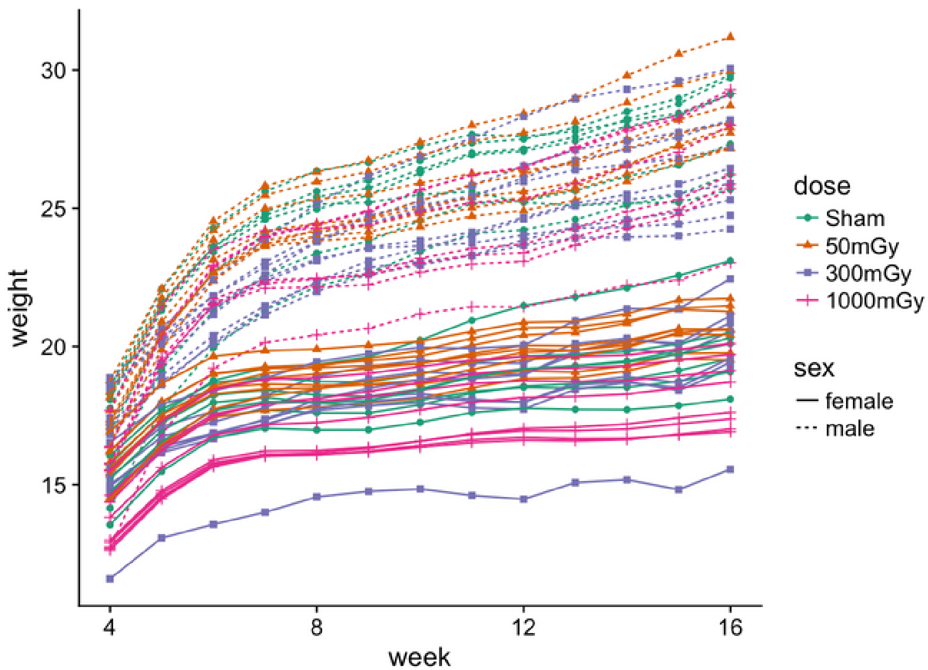


Shared trend
No shared trend
No group-level trends
Group-level trends similar smoothness
Group-level trends different smoothness
model 1
model 2
model 3
model 4
model 5
f(x)
x

example: rat birth weights

- see https://rpubs.com/bbolker/ratgrowthcurves

Most complex model (`dose*sex + s(week,by=dose_sex)+s(week,mother_id,bs='fs')`):



Duality between Z and correlation structure

- Hefley et al. (2017)
- "first-order specification": $\mathbf{y} \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \sigma_\epsilon^2 \mathbf{I})$
- "second-order specification: $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma_\epsilon^2 \mathbf{I} + \sigma_\mathbf{b}^2 \Sigma)$
- if $\mathbf{b}$ are iid Normal, integrating first-order specification shows that $\Sigma = \mathbf{Z}\mathbf{Z}^\top$
- e.g. latent-variable specification of an AR1 correlation structure
- e.g. `phyloglmm`

Penalty matrices as fixed effects

- can reparameterize latent variables to make them iid (and hence fittable with any random effects package)
- variables in the *null space* of the smooth will turn into fixed effects

- Simon N. Wood (2004)

Computational tricks

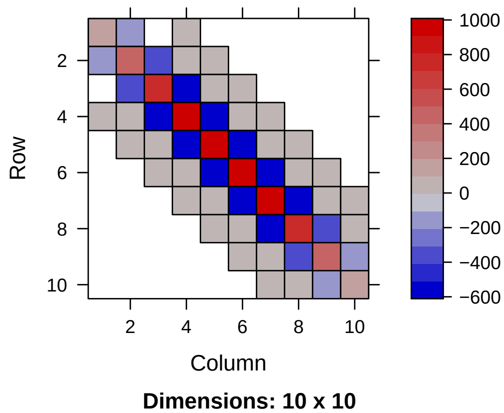
- work with *precision matrix* where possible Σ^{-1}
- for a **multivariate normal** response, $\Sigma^{-1}_{ij} = \Sigma^{-1}_{ji} = 0 \leftrightarrow x_i$ and x_j are **conditionally independent**
- e.g. precision matrix of AR1 is tridiagonal with diagonal $1+\rho^2$, first off-diagonal elements $-\rho$ (see here)
- work with *reduced-rank* forms where necessary

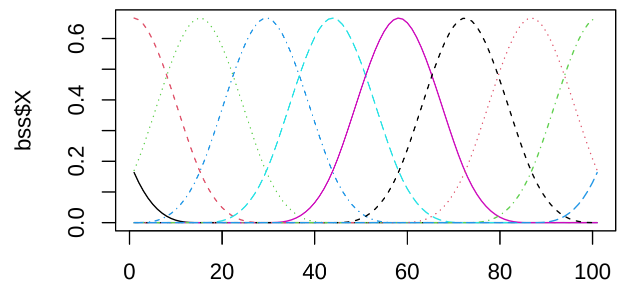
Penalty matrices

```
library(mgcv)
library(Matrix)
library(cowplot)
dd <- data.frame(x = seq(0, 1, length.out = 101))
bss <- smooth.construct.bs.smooth.spec(s(x, bs = "bs"),
                                data = dd, knots = NULL)
names(bss)
```

```
 [1] "term"           "bs.dim"         "fixed"          "dim"
 [5] "p.order"        "by"             "label"          "xt"
 [9] "id"             "sp"             "m"              "X"
[13] "knots"          "S"              "D"              "rank"
[17] "null.space.dim"
```

```
par(mfrow=c(1,2))
p1 <- image(Matrix(bss$S[[1]]))
p2 <- ~matplot(bss$X, type = "l")
plot_grid(p1, as_grob(p2))
```





effective degrees of freedom

$$\sum_i (1 + \lambda D_{ii})^{-1} = \mathrm{tr}(\tau)$$

$$\tau = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{S})^{-1} \mathbf{X}^\top \mathbf{X}$$

references

Golub, Gene H., Michael Heath, and Grace Wahba. 1979. "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter." *Technometrics* 21 (2): 215–23. https://doi.org/10.1080/00401706.1979.10489751.

Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized Additive Models*. CRC Press.

Hastie, Trevor, Robert Tibshirani, and J. H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer. https://hastie.su.domains/Papers/ESLII.pdf.

Hefley, Trevor J., Kristin M. Broms, Brian M. Brost, Frances E. Buderman, Shannon L. Kay, Henry R. Scharf, John R. Tipton, Perry J. Williams, and Mevin B. Hooten. 2017. "The Basis Function Approach for Modeling Autocorrelation in Ecological Data." *Ecology* 98 (3): 632–46. https://doi.org/10.1002/ecy.1674.

Larsen, Kim. 2015. "GAM: The Predictive Modeling Silver Bullet | Stitch Fix Technology – Multithreaded." *MultiThreaded (StitchFix)*. https://multithreaded.stitchfix.com/blog/2015/07/30/gam/.

Pedersen, Eric J., David L. Miller, Gavin L. Simpson, and Noam Ross. 2019. "Hierarchical Generalized Additive Models in Ecology: An Introduction with Mgcv." *PeerJ* 7 (May): e6876. https://doi.org/10.7717/peerj.6876.

Perperoglou, Aris, Willi Sauerbrei, Michal Abrahamowicz, and Matthias Schmid. 2019. "A Review of Spline Function Procedures in R." *BMC Medical Research Methodology* 19 (1): 46. https://doi.org/10.1186/s12874-019-0666-3.

Pya, Natalya, and Simon N. Wood. 2015. "Shape Constrained Additive Models." *Statistics and Computing* 25 (3): 543–59. https://doi.org/10.1007/s11222-013-9448-7.

Ramsay, Timothy O., Richard T. Burnett, and Daniel Krewski. 2003. "The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter." *Epidemiology* 14 (1): 18. https://journals.lww.com/epidem/Fulltext/2003/01000/The_Effect_of_Concurvity_in_Generalized_Additive.9.aspx.

Wood, Simon N. 2004. "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models." *Journal of the American Statistical Association* 99 (467): 673–86. https://doi.org/10.1198/016214504000000980.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. CRC Texts in Statistical Science. Chapman & Hall.