A.1.3 Semi-parametric estimation of mean and variance

Model description An assumption underlying the ordinary regression

$$y_i = a + bx_i + \varepsilon_i'$$

is that all observations have the same variance, i.e. $\mathrm{Var}(\varepsilon_i') = \sigma^2$. This assumption does not always hold, as for the data shown in the upper panel of Figure A.2. This example is taken from Ruppert et al. (2003).

It is clear that the variance increases to the right (for large values of x). It is also clear that the mean of y is not a linear function of x. We thus fit the model

$$y_i = f(x_i) + \sigma(x_i)\varepsilon_i,$$

where $\varepsilon_i \sim N(0,1)$, and $f(x)$ and $\sigma(x)$ are modelled nonparametrically. We take f to be a penalized spline. To ensure that $\sigma(x) > 0$ we model $\log[\sigma(x)]$, rather than $\sigma(x)$, as a spline function. For f we use a cubic spline (20 knots) with a 2nd order difference penalty

$$-\lambda^2 \sum_{k=3}^{20} (u_j - 2u_{j-1} + u_{j-2})^2,$$

while we take $\log[\sigma(x)]$ to be a linear spline (20 knots) with the 1st order difference penalty (A.2).

Implementation details Details on how to implement spline components are given Example A.1.2.

- Parameter associated with f should be given 'phase 1' in ADMB, while those associated with σ should be given 'phase 2'. The reason is that in order to estimate the variation, one first needs to have fitted the mean part.

- In order to estimate the variation function, one first needs to have fitted the mean part. Parameter associated with f should thus be given 'phase 1' in ADMB, while those associated with σ should be given 'phase 2'.

Files http://otter-rsch.com/admbre/examples/lidar/lidar.html

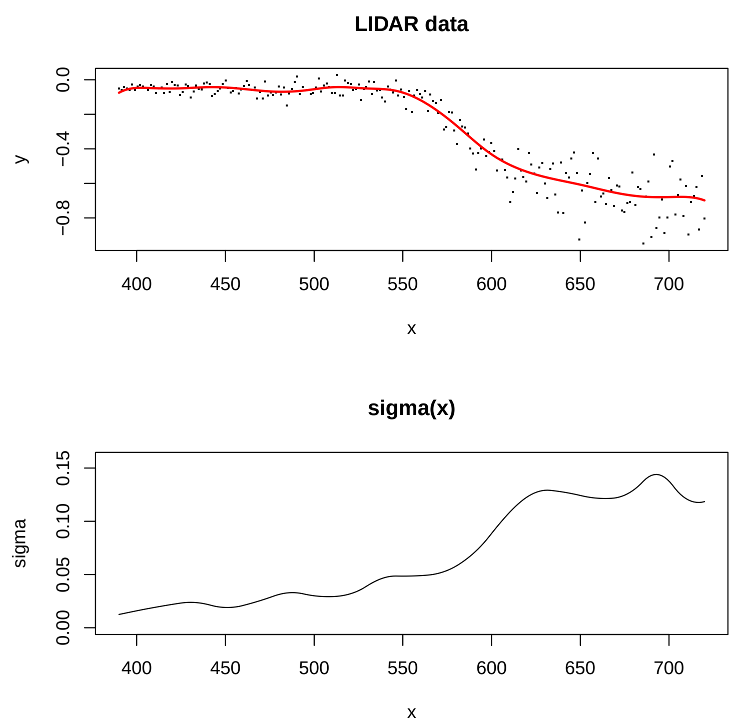

Figure A.2: LIDAR data (upper panel) used by Ruppert et al. (2003) with fitted mean. Fitted standard deviation is shown in the lower panel.

Bibliography

ADMB Development Core Team (2009), *An Introduction to AD Model Builder*, ADMB project.

ADMB Foundation (2009), 'ADMB-IDE: Easy and efficient user interface', *ADMB Foundation Newsletter* **1**, 1–2.

Eilers, P. & Marx, B. (1996), 'Flexible smoothing with B-splines and penalties', *Statistical Science* **89**, 89–121.

Harvey, A., Ruiz, E. & Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Kuk, A. Y. C. & Cheng, Y. W. (1999), 'Pointwise and functional approximations in Monte Carlo maximum likelihood estimation', *Statistics and Computing* **9**, 91–99.

Lin, X. & Zhang, D. (1999), 'Inference in generalized additive mixed models by using smoothing splines', *J. Roy. Statist. Soc. Ser. B* **61**(2), 381–400.

Pinheiro, J. C. & Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer.

Rue, H. & Held, L. (2005), *Gaussian Markov random fields: theory and applications*, Chapman & Hall/CRC.

Ruppert, D., Wand, M. & Carroll, R. (2003), *Semiparametric Regression*, Cambridge University Press.

Skaug, H. & Fournier, D. (2006), 'Automatic approximation of the marginal likelihood in non-gaussian hierarchical models', *Computational Statistics & Data Analysis* **56**, 699–709.

Zeger, S. L. (1988), 'A regression-model for time-series of counts', *Biometrika* **75**, 621–629.