SEGMENTED REGRESSION: SOME METHODS AND CASE STUDIES

Fred Potter

AgResearch Limited, New Zealand

Email: *fred.potter@agresearch.co.nz*

Some biological processes show discontinuities in their response. Methods are described for finding maximum likelihood estimates of the positions of the discontinuities. For k straight line segments we have:

$y = a + b_1x + b_2(x>p_1)(x-p_1) + \ldots b_k(x>p_{k-1})(x-p_{k-1})$; where the p parameters are discontinuities in terms of x

For two straight lines, a putative value of p is produced by the optimisation algorithm and two x variables fitted, one being the original x variable and the other being zeros for $x<p$ and $x-p$ beyond. Thus b_2 is the gradient adjustment at $x=p$. Of the parameters in this model, only p is nonlinear. Plots of the grid of values for the likelihood against p are made to check the relevance of the solution, since spurious minima may occur. With more than one discontinuity, contours of the likelihood for parameter pairs can be plotted. Case studies are given of fitting two straight lines, a line to a quadratic curve, a time lag to a straight line and a plateau to a decaying exponential curve. The experimenter specifies initial estimates and upper and lower bounds, because of the multiple solutions for p that can arise. It is hoped this will lead to the fitting of biologically plausible models. While data exploration can suggest discontinuities, the knowledge and input of the experimenter are required to ensure a relevant outcome. The importance of examining the fitted response to the data is stressed and various facets of this are discussed.