

Exploring data sets using partial residual plots based on robust fits

Joseph W. McKean

Western Michigan University, USA

Simon J. Sheather

University of New South Wales, Australia

Abstract: Partial residual plots are one of the most useful graphical procedures in the exploratory fitting of data sets. They are frequently used in the identification of unknown functions, g , of predictor variables. Traditionally these plots have been based on least squares (LS) fitting. It is well known that LS estimates are sensitive to outlying observations. The examples and sensitivity study discussed in this paper show that this vulnerability to outliers carries over to the LS based partial residual plots. A few outliers in the data set can distort the LS partial residual plot making the identification of g impossible. Furthermore, if g is non-linear, good data points may act as outliers and cause distortion in the plot. Partial residual plots based on highly efficient robust estimates are presented. In the simulated data sets explored in this paper, the robust based partial residual plots are insensitive to the outlying observations leading to a much easier identification of the unknown functions than their LS counterparts. In the sensitivity study presented, these robust based partial residual plots do not become distorted in the presence of outliers but they maintain their focus, enabling the identification of g .

Key words: Linear model, M-estimates, outlier, regression diagnostics, R-estimates, rank based methods.

AMS subject classification: 62J20, 62G35.

1 Introduction

Partial residual plots are one of the most useful graphical procedures in the