



Regional residual plots for assessing the fit of linear regression models

E. Deschepper*, O. Thas, J.P. Ottoy

Ghent University, Department of Applied Mathematics, Biometrics and Process Control, Coupure Links 653, B-9000 Gent, Belgium

Received 21 March 2005; received in revised form 22 March 2005; accepted 22 March 2005

Available online 29 April 2005

Abstract

An intuitively appealing lack-of-fit test to assess the adequacy of a regression model is introduced together with a graphical diagnostic tool. The graphical method itself includes a formal testing procedure, and, it is particularly useful to detect the location of lack-of-fit. The procedure is based on regional residuals, using subsets of the space of the independent variables. A simulation study shows that, the proposed procedures in simple linear regression have similar power as those of some popular classical lack-of-fit tests. In case of local departures from the hypothesized regression model, the new tests are shown to be more powerful. Therefore, when it becomes difficult to discriminate between systematic deviations and noise, regional residual plots are very helpful in formally locating areas of lack-of-fit in the predictor space. Data examples illustrate the ability of the new methods to detect and to locate lack-of-fit.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Diagnostic tool; Graphical method; Lack-of-fit; Multiple regression

1. Introduction

An important problem in applied statistics is the examination of the adequacy of parametric regression models to fit the observed data. Residuals are highly informative for this purpose and are widely used in both statistical tests as well as graphical diagnostic tools.

* Corresponding author. Tel.: +32 92645987; fax: +32 92646220.

E-mail address: Ellen.Deschepper@UGent.be (E. Deschepper).

Among the graphical diagnostic tools, the classical residual plot is probably the best known. It is often used as a descriptive method to assess lack-of-fit in a regression analysis. In general, graphical methods allow visualization of possible discrepancies between the fitted model and the data. Nevertheless, judging whether the observed discrepancies are really present or not is often a major problem and systematic departures smaller than the noise level can often not be observed. So far, most graphics introduced to assess the adequacy of regression models are illustrative and indicative, and the results depend on the data analyst. [Kuchibhatla and Hart \(1996\)](#) proposed an approach to lack-of-fit testing wherein graphs of smoothers play both a descriptive and inferential role. This approach is an attempt to obtain a graphical diagnostic tool, including a lack-of-fit test, but a major disadvantage of this procedure is its dependency on the choice of the smoother and its bandwidth.

In this paper, a formal procedure is introduced based on so-called regional residuals, which are defined on subsets of the sample space of the independent variables. A graphical diagnostic tool and a corresponding statistical test are proposed to check how well a parametric linear model for the mean fits to a set of observed data. The graphical method itself includes a formal testing procedure to assess the adequacy of a regression model. In particular, it allows the localization of lack-of-fit in the predictor space. The test statistic is simple and intuitively appealing, and is closely related to the lack-of-fit tests of [Lin et al. \(2002\)](#), who use cumulative and moving sums of residuals, or to earlier work of [Stute \(1997\)](#) or the cusum-based test discussed by [Buckley \(1991\)](#), who consider only cumulative sums of residuals.

Other popular, classical lack-of-fit tests are constructed from nonparametric smoothers. [Eubank et al. \(1993\)](#) and [Hart \(1997\)](#) provide a number of references of smoothing-based tests. [Eubank and Hart \(1993\)](#) concluded that the cusum test is the most powerful for very smooth departures from the no-effect hypothesis, whereas the smoothing-based tests are clearly superior when the alternative is high frequency. In more recent papers, e.g. [Kuchibhatla and Hart \(1996\)](#), [Aerts et al. \(1999, 2000\)](#), omnibus lack-of-fit tests that use data-driven selection criteria are proposed.

In Section 2, the regional residuals are defined on subsets of the sample space of the independent variables. The proposed tests and graphical tools are described in the linear regression context, and a bootstrap approach is proposed for approximating critical values and p -values. Also in Section 2, the link with the cumulative and moving sums of residuals of [Lin et al. \(2002\)](#) is discussed in more detail. In Section 3 the results of a simulation study are presented. The behavior of the new method is also investigated in case the model assumption of homoscedasticity is violated. Section 4 provides an illustration of the methodology on two data examples, and, finally, the concluding remarks are summarized in Section 5.

2. Plots and tests based on regional residuals

2.1. Regional residuals

In this section the discussion is restricted to the linear regression setting, assuming identically and independently distributed error terms. Given the independent observations

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات