

The American Statistician

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/utas20>

Nearly Parallel Lines in Residual Plots

J. A. Nelder^a

^a Department of Mathematics, Imperial College, London, SW7 2BZ, England

To cite this article: J. A. Nelder (1990) Nearly Parallel Lines in Residual Plots, The American Statistician, 44:3, 221-222

To link to this article: <http://dx.doi.org/10.1080/00031305.1990.10475725>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

TEACHER'S CORNER

In this department *The American Statistician* publishes articles, reviews, and notes of interest to teachers of the first mathematical statistics course and of applied statistics courses. The department includes the Accent on Teaching Materials section; suitable contents for the section are described

under the section heading. Articles and notes for the department, but not intended specifically for the section, should be useful to a substantial number of teachers of the indicated types of courses or should have the potential for fundamentally affecting the way in which a course is taught.

Nearly Parallel Lines in Residual Plots

J. A. NELDER*

When the residuals from a generalized linear model are to be plotted against the fitted values, the plot can be made to resemble that for a normal error model by transforming the fitted values to the constant-information scale.

KEY WORDS: Constant-information scale; Generalized linear model.

In this journal Searle (1988) recently pointed out that if the residuals $y - \hat{\mu}$ from a model are plotted against the fitted values $\hat{\mu}$, then the contours for constant y are parallel lines with slope -1 . The lines show up most clearly when the y values are discrete and repeated.

Such plots are commonly made to check for homoscedasticity (constant variance) of errors or presence of nonlinearity in the systematic part of the model. To check for constancy of variance we look at variation in the range of residuals for varying values of $\hat{\mu}$, that is, we examine vertical cross-sections. To check for nonlinearity in the systematic part we look for a systematic trend in the mean. If the data are discrete then for fixed $\hat{\mu}$, points will appear grouped at values $y - \hat{\mu}$, and the positions will change as we change $\hat{\mu}$. If the data are, for example, counts and so nonnegative, the contour for $y = 0$ is a boundary below which no points can lie.

The plot of $y - \hat{\mu}$ against $\hat{\mu}$ is mainly used in checking regression models with normal errors. For data from such a model we expect the vertical cross-sections to look like a sample from the normal distribution with variance independent of $\hat{\mu}$. A practical problem of interpretation arises if the local density of points varies widely with $\hat{\mu}$. For the values of $\hat{\mu}$ with high local density will on average show a wider range of residuals than values with low local density, and the eye may find this hard to allow for.

The graph of residuals against fitted values has much wider uses than checking the fit of regression models. This note deals with the corresponding graph for the more general class of generalized linear models (McCullagh and Nelder 1989), the aim being to make the graph in the general case look as much like that for the normal error case as possible.

A generalized linear model (GLM) extends classical regression models in two ways. First, the assumption of normal errors is widened to that of errors from an exponential family. This allows, for example, Poisson, binomial, gamma, and inverse Gaussian errors as alternatives. Second, the assumption that the mean μ is linear in the explanatory variables is replaced by the assumption that some (monotonic) function of μ is linear. Familiar quantities in regression analysis have their analogs in GLM analysis; for instance, the residual sum of squares, which measures the discrepancy of the data with respect to the model, is replaced by the deviance $D = \sum d_i$, where

$$d_i = 2 \int_{\mu_i}^{y_i} \frac{y_i - u}{V(u)} du.$$

In this formula $V(\cdot)$ is the variance function of the distribution assumed for the errors; thus for the Poisson, $V(\mu) = \mu$. Because, in general, the variance changes with the mean, we must amend the definition of residuals to allow for this. Two kinds of generalized residual are common. The first is the Pearson residual $r_p = (y - \hat{\mu})/\sqrt{V(\hat{\mu})}$, where $V(\cdot)$ is the variance function of the error distribution; the other is the deviance residual $r_D = \text{sgn}(y - \hat{\mu}) \sqrt{d}$, where d is defined as before. In general, deviance residuals look more normal than Pearson residuals, so it is better to use r_D . If we look at the plot of r_D against $\hat{\mu}$ we find that the contours for constant y are curves, unless $V(\mu) = 1$. Can these curves be made approximately straight and parallel by transforming the $\hat{\mu}$ scale? Suppose that we look for a transformation that makes the slope of the curve at $r = 0$ equal to -1 for all y as in the normal case [we cannot hope to make it -1 everywhere, unless $V(\mu) = 1$]. We show that the required transformation is the constant-information scale, which, for the one-parameter exponential families that underlie GLMs, is given by $\int V^{-1/2}(\mu) d\mu$. The scales for the standard GLM distributions are thus as follows.

Distribution	Transformation
Normal	μ
Poisson	$\sqrt{\mu}$
Binomial	$\arcsin(\sqrt{\mu})$
Gamma	$\log \mu$
Inverse Gaussian	$-\mu^{-1/2}$

*J. A. Nelder is Visiting Professor, Department of Mathematics, Imperial College, London SW7 2BZ, England.

The proof of the foregoing rule for transformations is as follows. First we show it for the Pearson residual r_p . We want to find $f(\mu)$ such that $dr_p/df(\mu) = -1$ at $\mu = y$; that is, $f'(\mu) = -dr_p/d\mu$ at $\mu = y$. Now $r_p = (y - \mu)/\sqrt{V(\mu)}$, so

$$[dr_p/d\mu]_{\mu=y} = -V^{-1/2}(y).$$

Thus

$$f(\mu) = \int \frac{d\mu}{V^{1/2}(\mu)}.$$

For the deviance residual, we begin with the formula for the deviance

$$D = 2 \int_{\mu}^y \frac{y - u}{V(u)} du.$$

Thus

$$dr_D/d\mu = dD^{1/2}/d\mu = -D^{1/2} \frac{y - \mu}{V(\mu)}.$$

From a Taylor expansion of D about $\mu = y$ we get

$$D = \frac{(y - \mu)^2}{V(\mu)} + O(y - \mu)^3.$$

Thus $\lim_{\mu \rightarrow y} dr_D/d\mu = -V^{-1/2}(y)$ as before.

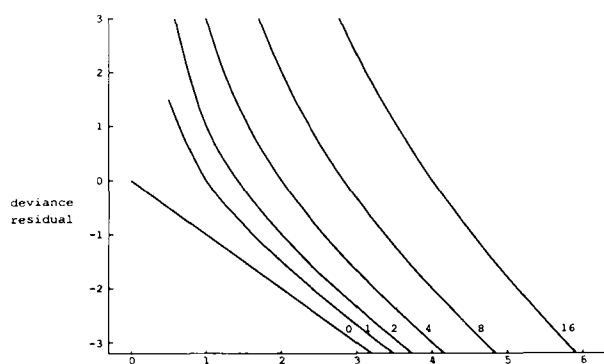


Figure 1. Contours of Constant $y = 0, 1, 2, 4, 8, 16$ for Poisson Variables When Deviance Residual Is Plotted Against Fitted Value on the Square-Root Scale.

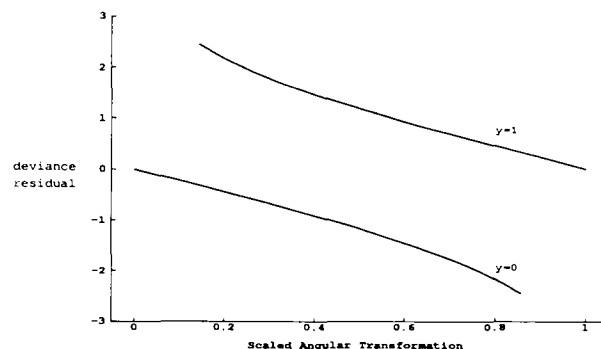


Figure 2. Deviance Residuals Versus Fitted Values on the Angular Scale for Binary Data.

The transformation produces near-linearity, and in practice one often finds that these transformations also improve the plot by spreading out the $\hat{\mu}$ s fairly uniformly; this aids interpretation, as discussed previously. Figure 1 shows the form of the contours for constant y for Poisson errors, and Figure 2 shows the two contours for the extreme case of binary data with binomial errors. In any particular plot all points will lie on these contours.

Clearly, the plot for binary data will be uninformative about heteroscedasticity. It may be used to look for nonlinearity, however, if the plot is accompanied by a smoothed line obtained, for example, from Cleveland's algorithm (Cleveland 1979). For counts it is useful to mark the $y = 0$ boundary on the plot, below which points cannot occur. This shows where, for small $\hat{\mu}$, it is unreasonable to expect the vertical cross-sections to look like a normal sample.

[Received October 1988. Revised September 1989.]

REFERENCES

- Cleveland, M. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829-836.
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall.
- Searle, S. R. (1988), "Parallel Lines in Residual Plots," *The American Statistician*, 42, 211.