Abstract

An added variable plot is a commonly used plot in linear regression diagnostics. The plot provides information about the addition of a further regressor to the model. The plot can lead to the identification of nonlinearity in the selected regressor, and outliers and influential observations that may seriously impact the least-squares estimate of the parameter that corresponds to the selected regressor. In this paper added variable plots are derived for a nonlinear regression model with an additive error term. The added variable plot for this nonlinear regression model is different from the plot in the linear regression case. The plot is not created for a specific explanatory variable, but for a parameter. Thus, the plot can be called an added parameter plot, since it provides information about the modification of the model by adding a parameter. The plot also gives a more formal tool to decide the importance of the parameter, since it is closely connected to the score test of the null-hypothesis that the added parameter is zero. It is proved that the value of the score test statistic is equal to SSR of the regression through the origin in the added parameter plot, divided by the estimated variance under the null hypothesis.

1 Introduction

Model selection including sensitivity analysis is always an important part of the data analysis, and the problem of deciding which model is more appropriate for a given set of data is a problem with no simple solution. If there are scientific reasons for preferring one model over the others, strong weight should be given to the researcher's reason. However, if there still are various models to consider one might need some kind of diagnostic aid.

In this paper, one such aid is presented - the Added variable plot. One application of this plot is to provide information to help decide whether an additional candidate variable should be included into a model. In addition, the plot is also used for detecting high leverage points and influential data.

The added variable plot is a plot of two sets of residuals. In the case of a linear regression model, the first set of residuals consists of those from the regression where the response variable is regressed on all explanatory variables, except for the variable under consideration. The second set of residuals is computed from regressing the variable under consideration on all the other explanatory variables. The general scatter of the points gives an overall impression of the strength of the relationship. A stronger linear relationship indicates the need for the inclusion of the considered variable as an explanatory variable in the model. In fact, the added variable plot will show the correct strength of the linear relationship between the response variable and the added explanatory variable when the other regressors are present in the model.

The plot helps detecting leverage (potentially influential) points and influential points that might not be leverage points. Individual points that are well separated from the rest of the data give heuristic information about the effects of outlying points on individual coefficients, and may suggest cases for special study. Belsley et al. (1980) give several examples illustrating the use of these plots for detecting influential observations. Added variable plots for the linear regression model have also been discussed by, for example, Cook and Weisberg (1982), Chatterjee and Hadi (1988), Ryan (1997) and Imon (2003).

Added variable plots can also be applied to nonlinear models. One example is the Cox's regression model discussed by Lindkvist (2000). The purpose of this article is to show how an added variable plot can be created for the nonlinear regression model with an additive error term. For such a model

there is not always an one-to-one correspondence between the parameters and the variables. As a result, the variables used to construct an added variable plot are not explanatory variables in the same sense as in the linear regression model. Here, the variable in mind is a constructed variable. This variable is constructed by taking the derivative of the response function with respect to the parameter of interest. An added variable plot in the nonlinear regression case is therefore created only for parameters and the added variable plot for the nonlinear regression model will from now on be referred to as the added parameter plot.

For a discussion of other graphical diagnostics in nonlinear regression, see the article by Cook (1987). He proposes graphical methods for displaying relevant information on a selected parameter from a normal nonlinear regression model. In the article by Cook, as in this paper, the linear extension of the nonlinear model is used to create a parameter plot. However, the plot discussed by Cook is slightly different from what is presented in this paper, since he is using another set of residuals as dependent variables in the plot. In his article, Cook points out that this plot may fail in revealing important information when the linear approximation is inadequate.

This article also discusses how the score test is connected to the added parameter plot. It is shown, in addition to carrying out the score test through the likelihood function, that the value of the score test statistic can be found in the estimated model for the added parameter plot. This gives the advantage that the added parameter plot can not only graphically tell us if the additional parameter should be included, but also gives a value of a formal test statistic. Further, the added parameter plot can be used to investigate which observations are influencing the score test.

It is also shown in this paper that the added parameter plot for the nonlinear case has properties similar to the added variable plot for the linear case, properties that are useful when conducting the score test via the added parameter plot. These properties are identified and proved in the paper.

To create an added parameter plot and to perform the score test for the nonlinear regression model with an additive error term, it is first necessary to estimate the model, using some numerical search procedure. The procedure used here is the Gauss-Newton method. A thorough description of the Gauss-Newton method is given by e.g. Neter et al. (1996). This procedure results in a matrix with values of the derivatives of the response function with respect to the parameters. These derivatives act as constructed explanatory

3

variables used in the added parameter plot. The constructed variables are also used when performing the score test. Together with the score test, detailed examples of other large-sample tests about the parameters in nonlinear regression with an additive error term are given by Gallant (1987).

In section 2 the nonlinear model is presented. A short presentation of the Gauss-Newton method is given and the score test is outlined. Added parameter plots in nonlinear regression is discussed in section 3 together with the properties of the plot and proofs. A numerical example is presented in section 4. The example illustrates some of the properties and a score test is conducted. The last section contains concluding remarks.

2 The nonlinear model

The model considered is the nonlinear regression model with an additive error term

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{x}$ represents a $n \times p-$matrix with p known explanatory variables. $\mathbf{y}$ is a $n-$vector of responses and $\boldsymbol{\theta}$ is a $q-$vector of unknown parameters. The errors are assumed to be independent, identically distributed normal random variables with mean 0 and variance σ^2. The function f is assumed to be twice continuously differentiable in θ.

For this nonlinear model there is not necessarily a one-to-one correspondence between parameters and variables, thus it is seldom the case that $p = q$. An example of a model where $q \neq p$ is the Michaelis-Menten model

$$y = \frac{\theta_1 x}{\theta_2 + x} + \varepsilon, \tag{2}$$

where $p = 1$ and $q = 2$.

2.1 Estimation

Estimation of nonlinear models involves an iterative numerical search procedure that minimizes the residual sum of squares, $SSE(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - f(\mathbf{x}_i, \boldsymbol{\theta}))^2$. One technique that can be used to accomplish this is the Gauss-Newton

method. This entails using a Taylor series expansion of $f(\mathbf{x}_i, \boldsymbol{\theta})$ about an initial parameter vector, denoted $\boldsymbol{\theta}^{(0)}$, using only the first term in the expansion.

$$f(\mathbf{x}_i, \boldsymbol{\theta}) \approx f(\mathbf{x}_i, \boldsymbol{\theta}^{(0)}) + \sum_{j=1}^{q} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_j^{(0)} \right) \frac{\partial}{\partial \theta_j} f(\mathbf{x}_i, \boldsymbol{\theta}) \mid_{\theta = \theta^{(0)}} . \tag{3}$$

This yields a linear model in which $\boldsymbol{\theta}$ is estimated using least squares. The initial value, $\boldsymbol{\theta}^{(0)}$, is then replaced by the convergence. Assuming that convergence occurs at the k-th iteration and rewriting (3) gives the final linear approximation of the model

$$y_i^{(k)} = y_i - f(\mathbf{x}_i, \boldsymbol{\theta}^{(k)}) = \sum_{j=1}^{q} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_j^{(k)} \right) \frac{\partial}{\partial \theta_j} f(\mathbf{x}_i, \boldsymbol{\theta}) \mid_{\theta = \theta^{(k)}} . \tag{4}$$

According to Charnes et al.(1976) $\boldsymbol{\theta}^{(k)}$ is identical to the maximum likelihood estimator of θ.

2.2 The score test

Exact inference procedures about the regression parameters are available for linear regression models with normal error terms for any sample size. Unfortunately, this is not the case for non-linear regression models with normal errors, where the least squares and maximum likelihood estimators for any given sample size are not guaranteed to be normally distributed, are not unbiased and do not have minimum variance. Consequently, inferences are usually based on large-sample theory (see e.g. Neter, Kutner, Nachtsheim & Wasserman (1996)).

One large-sample test in common use is the score test (or Lagrange multiplier test). This test uses the score function, which is the partial derivative of the log likelihood function, with respect to the parameters, $\boldsymbol{\theta}$. For normally distributed y's the log likelihood function is defined as

$$\ell = -\frac{n}{2} \ln(2\sigma^2 \pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2 .$$

The score function, U, is a $q-$ vector of scores

$$U(\boldsymbol{\theta}) = \left(\frac{d\ell}{d\theta_1}, \frac{d\ell}{d\theta_2}, ..., \frac{d\ell}{d\theta_q} \right)^T .$$

5

Asymptotically, $U(\boldsymbol{\theta})$ is $N(0, \mathbf{I}(\boldsymbol{\theta}))$ under some regularity assumptions (see for example Fahrmeir & Tutz (2001)), where $\mathbf{I}(\boldsymbol{\theta})$ is the information matrix for $\boldsymbol{\theta}$, defined as

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[U(\boldsymbol{\theta})U(\boldsymbol{\theta})^T\right].$$

For some hypothesized value, $\boldsymbol{\theta}_0$, the test statistic for the score test is

$$S(\boldsymbol{\theta}_0) = U(\boldsymbol{\theta}_0)^T \left[\mathbf{I}(\boldsymbol{\theta}_0)\right]^{-1} U(\boldsymbol{\theta}_0).$$

The score statistic, S, will asymptotically have a χ^2-distribution, with the number of restrictions under the null hypothesis as the number of degrees of freedom (Buse (1982)).

In order to test if a parameter is significant the score test can be used. Assume that $\boldsymbol{\theta}$ is a $(q+1)-$vector, $\left[\begin{array}{cccc} \theta_1 & \theta_2 & \cdots & \theta_{q+1} \end{array}\right]^T$. To conduct the score test of $H_0 : \theta_{q+1} = 0$, the model is first estimated under H_0, yielding the estimate $\widetilde{\boldsymbol{\theta}} = \left[\begin{array}{ccccc} \widetilde{\theta}_1 & \widetilde{\theta}_2 & \cdots & \widetilde{\theta}_q & 0 \end{array}\right]^T$. Next, the vector $\widetilde{\boldsymbol{\theta}}$ is used as starting value for a single iteration with the Gauss-Newton algorithm for the $(q+1)-$parameter model. This iteration gives a $n \times (q+1)$ matrix, $\mathbf{F}$, with the ith row

$$\mathbf{F}_i = \frac{\partial}{\partial \theta} f(\mathbf{x}_i, \boldsymbol{\theta})\mid_{\theta=\widetilde{\theta}}.$$

The matrix $\mathbf{F}$ is assumed to have full column rank.

Let $f(\widetilde{\boldsymbol{\theta}})$ denote the $n-$vector with elements $f(\mathbf{x}_i, \widetilde{\boldsymbol{\theta}})$, $i = 1, ..., n$, and $\mathbf{y}$ is the $n-$vector with responses. Then the test statistic for the score test of $H_0 : \theta_{q+1} = 0$ may then be written as

$$S(\widetilde{\boldsymbol{\theta}}) = U(\widetilde{\boldsymbol{\theta}})^T \left[I(\widetilde{\boldsymbol{\theta}})\right]^{-1} U(\widetilde{\boldsymbol{\theta}}).$$

$U(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\sigma^2}\mathbf{F}^T\left(\mathbf{y} - f(\widetilde{\boldsymbol{\theta}})\right)$ and $I(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\sigma^2}\mathbf{F}^T\mathbf{F}$. Thus $S(\widetilde{\boldsymbol{\theta}})$ can be written

$$S(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\sigma^2}\left(\mathbf{y} - f(\widetilde{\boldsymbol{\theta}})\right)^T \mathbf{F}\left[\mathbf{F}^T\mathbf{F}\right]^{-1}\mathbf{F}^T\left(\mathbf{y} - f(\widetilde{\boldsymbol{\theta}})\right). \tag{5}$$

Since σ^2 is unknown, it is estimated with $s^2 = MSE$ under the null hypothesis.

3   Added parameter plot

There are various diagnostic plots in regression analysis. One such plot is the added variable plot, that is used to study the effect of an additional explanatory variable in a regression model. Consider a linear regression model with regressors $X_1, ..., X_p$. An added variable plot for regressor X_p is constructed as follows. The residuals that result from regressing y on $X_1, ..., X_{p-1}$, are computed, as are the residuals that result from regressing X_p on $X_1, ..., X_{p-1}$. The added variable plot is defined as the scatter plot of the first set of residuals against the second set.

In linear regression there is a one-to-one correspondence between the parameters and the variables. In nonlinear regression this is not always the case. As a result, the variables used to construct an added variable plot are not explanatory variables in the same sense as in the linear regression case. Here, the variable in mind is a constructed variable. This variable is constructed by taking the derivative of the response function with respect to the parameter of interest. An added variable plot in the nonlinear regression case is therefore created only for parameters.

Using the notation in section 3, an added parameter plot for the constructed variable, $\frac{\partial}{\partial \theta_{q+1}} f(\mathbf{x}_i, \boldsymbol{\theta}) \mid_{\theta = \widetilde{\theta}}$, corresponding to the parameter θ_{q+1}, can be created. This plot examines the effect of modifying the q-parameter model by adding the parameter θ_{q+1} to the regression. Consider partitioning the matrix, $\mathbf{F}$, defined above, so that $\mathbf{F} = [\mathbf{F}_1, \ \mathbf{F}_2]$. Now, $\mathbf{F}_1$ contains the first q columns of $\mathbf{F}$ and $\mathbf{F}_2$ contains the last column of $\mathbf{F}$ corresponding to the parameter of interest, θ_{q+1}. Also, define the projection matrices $\mathbf{P}_F = \mathbf{F}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ and $\mathbf{P}_{F_1}$ and $\mathbf{P}_{F_2}$ in the same manner.

The added parameter plot for θ_{q+1} is a plot of

$$\mathbf{y}^* = (\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y} \text{ versus } \mathbf{x}^* = (\mathbf{I} - \mathbf{P}_{F_1})\mathbf{F}_2. \tag{6}$$

Here $\mathbf{y}^*$ is the vector of residuals when estimating the model without θ_{q+1}, and $\mathbf{x}^*$ is the vector of residuals from the least-squares regression of $\mathbf{F}_2$ on $\mathbf{F}_1$. In the regression through the origin of $\mathbf{y}^*$ on $\mathbf{x}^*$, $\mathbf{P}_{x^*}$ is defined as $\mathbf{x}^*(\mathbf{x}^{*T}\mathbf{x}^*)^{-1}\mathbf{x}^{*T}$.

The resulting plot has the following properties.

Property 1 The least squares estimate $\widehat{\alpha}$, of the slope α, in the simple linear regression through the origin in the added variable plot (6), is equal

to the updated parameter estimate $\widehat{\theta}_{q+1}$ in the model after one iteration of the Gauss-Newton algorithm when $\widetilde{\boldsymbol{\theta}}$ is used as starting value.

Proof The least squares estimate of the slope α is

$$
\begin{aligned}
\widehat{\alpha} &= (\mathbf{x}^{*T}\mathbf{x}^*)^{-1}\mathbf{x}^{*T}\mathbf{y}^* \\
&= \left[\mathbf{F}_2^T\mathbf{F}_2 - \mathbf{F}_2^T\mathbf{P}_{F_1}\mathbf{F}_2\right]^{-1}\mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y}.
\end{aligned}
$$

Consider the linearized model (4) and use $\mathbf{F}$ as the independent variable and the residuals under H_0 as the dependent variable, the starting values with the Gauss-Newton algorithm can be revised.

$$
\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \left[\mathbf{F}^T\mathbf{F}\right]^{-1} \times \left[\mathbf{F}^T\right] \times \left[\mathbf{y} - f(\mathbf{X}, \boldsymbol{\theta}^{(0)})\right].
$$

Using $\widetilde{\boldsymbol{\theta}}$ as starting value, the partition of $\mathbf{F}$ yields

$$
\boldsymbol{\theta}^{(1)} = \widetilde{\boldsymbol{\theta}} + \left[\begin{array}{cc} \mathbf{F}_1^T\mathbf{F}_1 & \mathbf{F}_1^T\mathbf{F}_2 \\ \mathbf{F}_2^T\mathbf{F}_1 & \mathbf{F}_2^T\mathbf{F}_2 \end{array}\right]^{-1} \times \left[\begin{array}{c} \mathbf{F}_1^T \\ \mathbf{F}_2^T \end{array}\right] \times \left[f(\widetilde{\boldsymbol{\theta}})\right].
$$

The estimate of θ_{q+1} is obtained as

$$
\left[\mathbf{F}_2^T\mathbf{F}_2 - \mathbf{F}_2^T\mathbf{P}_{F_1}\mathbf{F}_2\right]^{-1}\mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y},
$$

where $f(\widetilde{\boldsymbol{\theta}})$ is replaced by $(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y}$ since both is an expression for the residuals under the null hypothesis. ∎

Property 2 The projection matrix $\mathbf{P}_{x^*}$ for the regression through the origin in the added variable plot (5) is equal to $\mathbf{P}_F - \mathbf{P}_{F_1}$.

Proof The projection matrix in the added variable plot is

$$
\begin{aligned}
\mathbf{P}_{x^*} &= (\mathbf{I} - \mathbf{P}_{F_1})\mathbf{F}_2 \left[\mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{F}_2\right]^{-1}\mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1}) \\
&= \frac{\mathbf{e}_{F_2 \cdot F_1}\mathbf{e}_{F_2 \cdot F_1}^T}{\mathbf{e}_{F_2 \cdot F_1}^T\mathbf{e}_{F_2 \cdot F_1}},
\end{aligned}
$$

where $\mathbf{e}_{F_2 \cdot F_1}$ is the residual vector when $\mathbf{F}_2$ is regressed on $\mathbf{F}_1$. Since $\mathbf{F}$ is a partitioned matrix, $\mathbf{P}_F$ can be decomposed into a sum of projection matrices (Chatterjee & Hadi (1988)),

$$
\mathbf{P}_F = \mathbf{P}_1 + \mathbf{P}_2,
$$

8

and $\mathbf{P}_F$ can be written

$$
\begin{aligned}
\mathbf{P}_F &= \mathbf{P}_{F_1} + (\mathbf{I} - \mathbf{P}_{F_1})\mathbf{F}_2 \left[\mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{F}_2\right]^{-1} \mathbf{F}_2^T(\mathbf{I} - \mathbf{P}_{F_1}) \\
&= \mathbf{P}_{F_1} + \frac{\mathbf{e}_{F_2 \cdot F_1}\mathbf{e}_{F_2 \cdot F_1}^T}{\mathbf{e}_{F_2 \cdot F_1}^T\mathbf{e}_{F_2 \cdot F_1}} \\
&\Rightarrow \mathbf{P}_F - \mathbf{P}_{F_1} = \mathbf{P}_{x^*} \ \blacksquare
\end{aligned}
$$

The matrix $\mathbf{P}$ can also be referred to as the weight- or the leverage matrix, whose diagonal elements are termed as leverages. Influential observations are those whose presence or absence can make a huge impact on the fitting of the model and hence the resulting analyses. Observations possessing excessively large values on the diagonal of $\mathbf{P}$ are termed as high leverage points and are potentially influential. The leverages in $\mathbf{P}_{x^*}$ tells us about the influence on the parameter estimate for the additional variable, given that all other variables are included in the regression.

Property 3 The residual vector, $\mathbf{e}_{y^* \cdot x^*}$, for the regression through the origin in the added variable plot is $(\mathbf{I} - \mathbf{P}_F)\mathbf{y}$, that is the residuals when $\mathbf{y}$ is regressed on $\mathbf{F}$.

Proof The residuals are

$$
\mathbf{e}_{y^* \cdot x^*} = (\mathbf{I} - \mathbf{P}_{x^*})\mathbf{y}^*.
$$

Using property 2 and

$$
\begin{aligned}
(\mathbf{I} - \mathbf{P}_{x^*})\mathbf{y}^* &= (\mathbf{I} - \mathbf{P}_{x^*})(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y} \\
&= (\mathbf{I} - \mathbf{P}_F + \mathbf{P}_{F_1})(\mathbf{I} - \mathbf{P}_{F_1})\mathbf{y} \\
&= \mathbf{I} - \mathbf{P}_F + \mathbf{P}_{F_1} - \mathbf{P}_{F_1} + \mathbf{P}_F\mathbf{P}_{F_1} - \mathbf{P}_{F_1} \\
&= (\mathbf{I} - \mathbf{P}_F)\mathbf{y},
\end{aligned}
$$

The $\mathbf{P}_{F_1}$ terms on line 3 is cancelled out because $\mathbf{P}_F\mathbf{F}_1 = \mathbf{F}_1$ and thus $\mathbf{P}_F\mathbf{P}_{F_1} = \mathbf{P}_{F_1}$. $\blacksquare$

Property 3 states that the residual vector, $\mathbf{e}_{y^* \cdot x^*}$, is the same whichever variable in the model we are examining in the added variable plot (Cook & Weisberg (1982)).

Property 4 SSR in the regression for the added variable plot, divided by MSE from the model not including θ_{q+1}, equals the test statistic for the score test of

$$
H_0 : \theta_{q+1} = 0
$$

Proof SSR in the added variable plot regression is $\mathbf{y}^{*T}\mathbf{P}_{x^*}\mathbf{y}^*$.

$$
\begin{aligned}
SSR &= \mathbf{y}^{*T}\mathbf{P}_{x^*}\mathbf{y}^* \\
&= \left[(\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y}\right]^T \mathbf{P}_{x^*}\left[(\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y}\right] \\
&\quad \text{using property 2} \\
&= \mathbf{y}^T(\mathbf{I}-\mathbf{P}_{F_1})\left[\mathbf{P}_F - \mathbf{P}_{F_1}\right](\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y} \\
&= \mathbf{y}^T\left[\mathbf{P}_F - \mathbf{P}_{F_1} - \mathbf{P}_{F_1}\mathbf{P}_F + \mathbf{P}_{F_1}\right](\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y} \\
&\quad \text{using property 3} \\
&= \mathbf{y}^T\left(\mathbf{P}_F - \mathbf{P}_{F_1}\right)(\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y} \\
&= \mathbf{y}^T\left(\mathbf{I}-\mathbf{P}_{F_1}\right)\mathbf{P}_F(\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y}.
\end{aligned}
$$

$(\mathbf{I}-\mathbf{P}_{F_1})\mathbf{y}$ is the residuals under the null and thus, $\mathbf{y}^{*T}\mathbf{P}_{x^*}\mathbf{y}^*$ can be written

$$
\left(\mathbf{y} - f(\widetilde{\boldsymbol{\theta}})\right)^T \mathbf{F}\left[\mathbf{F}^T\mathbf{F}\right]^{-1}\mathbf{F}^T\left(\mathbf{y} - f(\widetilde{\boldsymbol{\theta}})\right).
$$

Thus, $\mathbf{y}^{*T}\mathbf{P}_{x^*}\mathbf{y}^* \propto S(\widetilde{\boldsymbol{\theta}})$ in (5). This result makes it easy to compute the test statistic of the score test. ∎

4   Example

The Michaelis-Menten model, (2), presented in Bates and Watts (1988), for enzyme kinetics relates the initial "velocity" of an enzymatic reaction to the substrate concentration, x, through the equation

$$
f(x,\theta) = \frac{\theta_1 x}{\theta_2 + x}.
$$

θ_1 is the asymptotic velocity of the enzymatic reaction and graphically represents the asymptotic value of f as $x \to \infty$. θ_2 represent the half-concentration, i.e. the value of x such that when the concentration reaches that value the velocity is one-half its ultimate value.

In the example presented in Baytes and Watts (1988), two blocks of experiments were run. In one block the enzyme was treated with puromycin, and in the other the enzyme was untreated. It was hypothesized that the

Puromycin should affect the maximum velocity parameter θ_1, but not the half-velocity parameter θ_2. An indicator variable, x_2 was introduced.

$$f(x, \theta) = \frac{(\theta_1 + \phi_1 x_2)\, x_1}{\theta_2 + x_1},$$

The model (2) is thus modified, now including ϕ_1 to account for the effect of puromycin on the asymptotic velocity, θ_1.

The second modification of model (2) includes ϕ_1 as well as ϕ_2, a parameter for potential effect of puromycin on "half-concentration":

$$f(x, \theta) = \frac{(\theta + \phi_1 x_2)\, x_1}{(\theta_2 + \phi_2 x_2) + x_1},$$

The score test can now be used to test if the model should include different half-velocity parameters depending on wether the enzyme is treated or not. That is, the hypothesis:

$$
\begin{aligned}
H_0 &: \quad \phi_2 = 0 \\
H_A &: \quad \phi_2 \neq 0.
\end{aligned}
\tag{7}
$$

Together with the test an added parameter plot for ϕ_2 can be created.

To conduct the score test the parameter vector must be defined. Let

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \phi_1 \\ \phi_2 \end{bmatrix}.$$

Now the model under the null hypothesis is estimated to retrieve the starting values and MSE. These estimates are:

$$\widetilde{\boldsymbol{\theta}} = \begin{bmatrix} 166.6041 \\ 0.0580 \\ 42.0260 \\ 0 \end{bmatrix}, \tag{8}$$

and

$$MSE = 112.05$$

respectively. $\widetilde{\boldsymbol{\theta}}$ is now used as starting values for a single iteration of the Gauss-Newton method to retrieve the matrix of $\mathbf{F}$. After this iteration, we see that the updated estimates of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}^{(1)} = \begin{bmatrix} 160.8924 \\ 0.0485 \\ 51.2969 \\ 0.0148 \end{bmatrix} \tag{9}$$

When $s^2 = MSE = 112.05$ is used as an estimate of σ^2, the value of the test statistic is

$$S = 1.4483,$$

where S has 1 degree of freedom. The p-value for this test is 0.2280 and the null hypothesis that the half-velocity parameter is unchanged by the puromycin treatment cannot be rejected.

We can compare this result with the result from another large-sample test. The three-parameter model and the four-parameter model is compared in Bates and Watts (1988). The hypothesis (7) is tested with the likelihood ratio test. This test gives a value of the test statistic of 1.7, which has a corresponding p-value of 0.21. The p-value is close to 0.2280 and according to this result there is not enough evidence to reject the null hypothesis. The likelihood ratio test requires both the restricted and the unrestricted estimates of the parameter, therefore the likelihood ratio test is computational more demanding.

Both theses test statistics has an asymptotic χ^2−distribution. The distribution will approach the χ^2−distribution as the number of observations gets larger. In this example the number of observations is 23, which might be too small.

To visualize the score test an added parameter plot for ϕ_2 is created. In the plot, the columns of $\mathbf{F}$ acts as independent variables. Here, the first three columns of $\mathbf{F}$ form the matrix $\mathbf{F}_1$ and the last column form the vector $\mathbf{F}_2$. First, $\mathbf{y}^*$ is constructed as the residuals in the regression when $\mathbf{y}$ is regressed on $\mathbf{F}_1$. Second, $\mathbf{x}^*$ is constructed as the residuals when $\mathbf{F}_2$ is regressed on $\mathbf{F}_1$. Now, $\mathbf{y}^*$ is regressed on $\mathbf{x}^*$, and an added parameter plot for ϕ_2 is constructed. The results are displayed in Figure 1.

ystar = 0.0148xstar
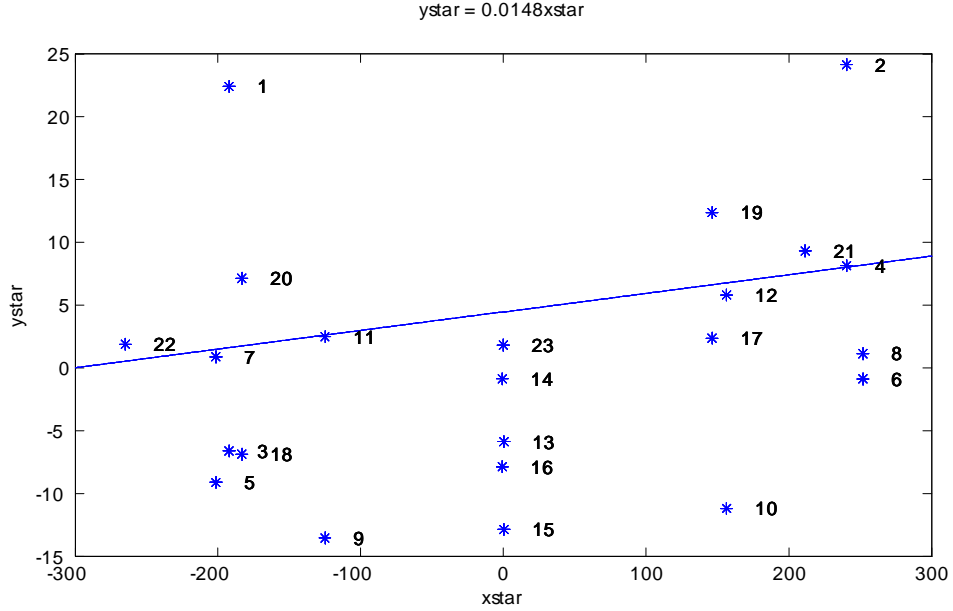
Added parameter plot of ϕ_2

The estimated slope in the regression of $\mathbf{y}^*$ on $\mathbf{x}^*$, as shown in Figure 1, is 0.0148 which equals the parameter estimate of ϕ_2 in (9) after one iteration with the Gauss-Newton method. This illustrates property 1. According to property 4, the value of the score test statistic can be found by dividing SSR in the regression for the added variable plot with MSE from the restricted model. SSR for the added variable plot equals 162.2754. The value of the test statistic is thus

$$S = \frac{162.2754}{112.05} = 1.4483,$$

which equals the value of the score test statistic.

5   Discussion

In this article the added variable plot for the nonlinear regression model with an additive error term is created. It is shown that the plot is not created in the same manner as the added variable plot for the linear regression model. In the nonlinear regression model, there is not always a one-to-one correspondence

13

between variables and parameters. The plot cannot be created for a variable that does not correspond to a parameter, and therefore the added variable plot can be interpreted as an added parameter plot. However, it is shown that some properties of the plot in the linear case can be applied to the nonlinear case. One useful property of the added variable plot, and the added parameter plot, is that the score test can be easily carried out. The value of the score test statistic of the null-hypothesis that the added parameter is zero, can be found in the estimated model of the added parameter plot. The sum of squares due to regression in the estimated model of the added parameter plot, divided with s^2 - the estimated variance of the restricted model, is the value of the test statistic. Thus, a strong linear relationship between the two sets of residuals in the plot suggests high SSR. With the value of the variance held constant, this would lead to a high value of the test statistic, and to a rejection of the null-hypothesis.

The added variable plot gives not only a visual overview of the importance of an additional parameter. The plot can also visualize which individual observations that contributes to a high, or low, SSR. That is, the plot may indicate which observations seem to be influencing for the score test.

Further studies can be done to derive the empirical influence function, EIC, of the score test statistic.

References

Bates, D.M., and D.G. Watts (1988). *Nonlinear Regression Analysis and Its Applications.* New Yord: Wiley.

Belsley, D.A., E. Kuh, and R.E. Welsch (1980). *Regression Diagnostics: Identifying Influencial Data and Sources of Collinearity.* New York: Wiley.

Buse, A. (1982). The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note. *The American Statistician, Vol 36, No. 3, pp 153-157.*

Charnes, A., L. Frome, and P. L- Yu (1976). The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family. *Journal of the American Statistical Association, Vol 71, No. 353, pp. 214-222.*

Chatterjee, S., and A.S. Hadi (1988). *Sensitivity Analysis in Regression.* New York: Wiley.

Cook, R.D. (1987). Parameter Plots in Nonlinear Regression. *Biometrica, Vol.74, No. 4, pp. 669-677.*

Cook, R.D. and S. Weisberg (1982). *Residuals and Influence in Regression.* New York: Chapman and Hall.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd edition.* New York: Springer-Verlag.

Gallant, A.R. (1987). *Nonlinear Statistical Models.* New York: Wiley.

Imon, A.H.M. Rahmatullah (2003). Residuals from Deletion in Added Variable Plots. *Journal of Applied Statistics, Vol.30, No. 7, pp. 827-841.*

Lindkvist, M. (2000). *Added Variable Plots and Influence in Cox's Regression Model.* Department of Statistics, University of Umeå.

Neter, J.,M.H. Kutner, C.J. Nachtsheim and W. Wasserman (1983?). *Applied Linear Regression Models, 3rd edition.* Chicago: Irwin.

Ryan, T.P. (1997). *Modern Regression Methods.* New York: Wiley.