Identifying Influential Observations in Nonlinear Regression

a focus on parameter estimates and the score test

Karin Stål

Academic dissertation for the Degree of Doctor of Philosophy in Statistics at Stockholm University to be publicly defended on Tuesday 14 April 2015 at 10.00 in De Geersalen, Geovetenskapens hus, Svante Arrhenius väg 14.

Abstract

This thesis contributes to influence analysis in nonlinear regression and in particular the detection of influential observations. The focus is on a regression model with a known mean function, which is nonlinear in its parameters and where the function is chosen according to the knowledge about the process generating the data. The error term in the regression model is assumed to be additive.

The main goal of this thesis is to work out diagnostic measures for assessing the influence of observations on various results from a nonlinear regression analysis. The obtained results comprise diagnostic tools for detecting observations that, individually or jointly with some other observations, are influential on the parameter estimates. Moreover, assessing conditional influence, i.e. the influence of an observation conditional on the deletion of another observation, is of interest. This can help to identify influential observations which could be missed due to complex relationships among the observations. Novelties of the proposed diagnostic tools include the possibility to assess influence of observations on a specific parameter estimate and to assess influence of multiple observations.

A further emphasis of this thesis is on the observations' influence on the outcome of a hypothesis testing procedure based on Rao's score test. An innovative solution to the problem of visual identification of influential observations regarding the score test statistic obtained in this thesis is the so called added parameter plot. As a complement to the added parameter plot, new diagnostic measures are derived for assessing the influence of single and multiple observations on the score test statistic.

Keywords: *Added parameter plot, differentiation approach, influential observation, nonlinear regression, score test.*

Stockholm 2015
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-114324

ISBN 978-91-7649-115-7

Department of Statistics

Stockholm University, 106 91 Stockholm

Identifying Influential Observations
in Nonlinear Regression – a focus on parameter estimates and the score test

Karin Stål

Identifying Influential Observations in Nonlinear Regression

a focus on parameter estimates and the score test

Karin Stål

Abstract

This thesis contributes to influence analysis in nonlinear regression and in particular the detection of influential observations. The focus is on a regression model with a known mean function, which is nonlinear in its parameters and where the function is chosen according to the knowledge about the process generating the data. The error term in the regression model is assumed to be additive.

The main goal of this thesis is to work out diagnostic measures for assessing the influence of observations on various results from a nonlinear regression analysis. The obtained results comprise diagnostic tools for detecting observations that, individually or jointly with some other observations, are influential on the parameter estimates. Moreover, assessing conditional influence, i.e. the influence of an observation conditional on the deletion of another observation, is of interest. This can help to identify influential observations which could be missed due to complex relationships among the observations. Novelties of the proposed diagnostic tools include the possibility to assess influence of observations on a specific parameter estimate and to assess influence of multiple observations.

A further emphasis of this thesis is on the observations' influence on the outcome of a hypothesis testing procedure based on Rao's score test. An innovative solution to the problem of visual identification of influential observations regarding the score test statistic obtained in this thesis is the so called added parameter plot. As a complement to the added parameter plot, new diagnostic measures are derived for assessing the influence of single and multiple observations on the score test statistic.

Keywords: Added parameter plot, differentiation approach, influential observation, nonlinear regression, score test

ISBN 978-91-7649-115-7

Printed in Sweden by Publit, Stockholm 2015

Distributor: Department of Statistics, Stockholm University

This thesis is dedicated to my loving family.

Contents

List of Figures

List of Tables

Acknowledgments

Finishing this thesis was a very stressful task, where dreams in the night about theorems and proofs were haunting me and the day didn't seem to have enough hours. However, being a Ph.D. student has been a true experience, and it warms my heart to think about the people who have supported me and been there for me, in good times and bad.

First and foremost, my deepest gratitude goes to my supervisor, Associate Professor Tatjana von Rosen. You have been a solid ground with your guidance, infinite knowledge and positive attitude. Your constant willingness to help is remarkable and you are never too tired to make an effort. Moreover, I really appreciate your sense of humor and I would like to thank you for the laughs we had.

To my assistant supervisor Professor Dietrich von Rosen I would like to say thank you for all the interesting discussions and for your insightful comments. During the final stage, your help has been invaluable and I appreciate that you always find time to read and answer any questions.

Ellinor Fackle-Fornius, my assistant supervisor, thank you for all your help with my research and for the fantastic times we have had in the past nine years. Some are truly memorable, for instance the way we finished our UPC-course.

Thank you, all fellow Ph.D. students, former and present, at the Department of Statistics. Together, we have had a lot of fun and I take good memories with me. I would like to send a special "thank you" to Olivia and Yuli. Your support has been invaluable and coming to work is much more fun when you are there.

Moreover, thanks to all my colleagues at the Department of Statistics, and especially to Dan Hedlin, Richard Hager and Michael Carlson for being helpful and willing to listen when problems arise. I would also like to thank Professor Hans Nyquist for introducing me to the topic and for providing me with good ideas.

To my wonderful parents, mamma Berit and pappa Erik: what would I do without you? Better parents and supporters cannot be found. Whenever I need

you, you are there for me and I count myself lucky having you.

Daniel Bruce, I am sincere when I say that this thesis would not have been written if it wasn't for you. Your dedication to our family and the way you prioritized me and me finalizing my thesis is extraordinary. You are a wonderful, wonderful man and I love you.

The best things in my life are my sons, Elliott and Elmer. With you, there is never a dull moment. Thank you for being you and for being part of my life.

Karin Stål
Stockholm, March 9, 2015

1. Introduction

1.1 Francis Galton, linear regression and correlation

In 1889 the English polymath Francis Galton was taking a country walk at Naworth Castle near Carlisle, Northern England. A rainstorm was sweeping the country, and as Galton took shelter from the rainstorm, an idea flashed across him, namely the idea of correlation analysis. This was the beginning of correlation and regression analysis, according to Barnes (1998), where an amusing story about the history of statistics in general, and regression analysis in particular, is told. Whether this story is true or not is debated, see for instance Stigler (1986). However, assuming it is true, the idea of correlation that flashed across Galton that day in the rain did not appear in a vacuum. It was a concluding step in a 20-year research project. He first observed reversion towards the mean in the late 1870's when he conducted experiments on seed size in successive generations of sweet peas. In the 1880's Galton was investigating the heights of parents and their offspring, see Bulmer (2003). Galton found that tall parents, or taller than mediocrity as he called it, had children who were shorter than themselves and that parents who were shorter than mediocrity had children taller than themselves. This led him to call the phenomena "regression toward mediocrity." According to Sen and Srivastava (1990), the phenomenon regression did not start with Galton. There were other mathematicians that were doing what we could call regression prior to Galton. What was interesting with Galton's work was that he connected regression and correlation. According to Stigler (1989), in the late 1880's Galton was simultaneously pursuing two unrelated investigations, one in anthropology and one in forensic science. In anthropology, the question was as follows: If a single thigh bone is recovered from an ancient grave and measured, what can the measurement of the bone tell us about the total height of the individual to whom it belonged? The other question was related: For the purpose of criminal identification, what can be said about the relationship between measurements taken from different parts of the same person? What dawned on Galton was that these new problems were identical to the old one on kinship and that all three of them were no more than special cases of a much more general problem, namely that of correlation. Not only did he describe the relationships between variables through regression

toward mediocrity, but he also found a way to measure the strength of this relationship through the correlation coefficient. Moreover, Galton realized that the variation of one variable around the regression line could be divided into two parts, one part that could be explained by the other variable and one that could not.

Galton's ideas about correlation and regression are not far from our conception of them. Correlation is used to study linear association between two variables. The correlation coefficient for assessing the strength of this association ranges between -1 and 1, where the sign indicates the direction of the association. In linear regression the linear association between the variables is described using a linear function. Moreover, as Galton realized, the dependent variable depends on some unobservable error, often assumed to be a normally distributed random variable with expectation zero and constant variance. When the unknown parameters in the linear function is estimated, a fitted linear regression model is obtained. Correlation and regression are connected as the estimate of the slope parameter in the linear function and the correlation coefficient are functionally related.

Galton rightly foresaw that the methods of regression and correlation would have a prominent place in many applications, see Bulmer (2003). The linear regression model is widely used in e.g. business, the social and behavioral sciences, the biological sciences and many other disciplines.

1.2 Introduction to influence analysis in linear regression

It is well understood that not all observations in the data set play an equal role when fitting a regression model. Some observations might have more impact on, for instance, the estimation process than others. Observations that significantly influence certain results from the regression analysis are called influential observations. The study of the data and how different parts of it influence the inference is called influence analysis. Influence analysis of the inference in linear regression models is a well-established area of research. Andrews and Pregibon (1978) highlighted that we need to find the outliers that matter. What is meant by this is that not all outliers need to be harmful in the way that they have an undue influence on, for instance, the estimation of the parameters in the regression model. If not all outliers matter, examining the residuals alone might not lead us to the detection of aberrant or unusual observations. Thus, other ways for finding influential observations are needed. Hoaglin and Welsch (1978) discussed the importance of the projection ma-

trix in linear regression, where the projection matrix is the matrix that projects onto the regression space. They argued that the diagonal elements of the projection matrix are important ingredients in influence analysis. The diagonal elements are referred to as leverages, since they can be thought of as the amount of leverage concerning the response value on the corresponding predicted response value. Perhaps the most well-known influence measure was proposed by Cook (1977), referred to as Cook's distance. Cook's distance is an influence measure used for assessing the influence of the observations on the estimated parameter vector in the linear regression model. Cook's distance is widely used by practitioners for detecting influential observations, and it is included in most statistical computer programs. There exists a wide range of other influence measures to use in linear regression analysis for assessing the influence of the observations on various results of the regression analysis. For example, Andrews and Pregibon (1978) derived a measure of the influence of an observation on the estimated parameters. This measure, the Andrews-Pregibon statistic, is based on the change in volume of confidence ellipsoids with and without a particular observation. Moreover, Belsley *et al.* (1980) suggested an influence measure for assessing the influence of an observation on the variance of the estimated parameters in the linear regression model, known as COVRATIO. Besides the influence measures mentioned here there exist many more, see e.g. Chatterjee and Hadi (1986) and Hadi (1992) for excellent overviews of influence measures.

Graphical investigation of data is a powerful tool in explorative analysis. It can be used to examine relationships between variables and discover observations deviating form other. Hence, influential observations can also be detected using graphical tools. Mosteller and Tukey (1977) introduced the added variable plot, which is used for graphically detecting observations that have a large influence on the parameter estimates. For details concerning the added variable plot, such as construction and properties, see e.g. Belsley *et al.* (1980), where the plot is referred to as the partial regression leverage plot, and Cook and Weisberg (1982). Other results on graphical tools in influence analysis are provided by e.g. Atkinson (1982) and Johnson and McCulloch (1987). It is important to note that the graphical tools used in influence analysis are not conclusive, but rather suggestive.

From the previous paragraphs we can see that the 1970's and the 1980's were the decades when most research results on influence analysis in linear regression came to see the light. However, influence analysis in linear regression is still an active research area. Nurunnabi *et al.* (2014) proposed a modification of Cook's distance. This modification enables the identification of multiple

influential observations. Furthermore, Beyaztas and Alin (2014) used a combined Bootstrap and Jackknife algorithm to detect influential observations.

In applied data analysis, there is an increasing availability of data sets containing a large number of variables. When such data is in the hands of the researcher sparse regression can be implemented, which is another field of research active today. In sparse regression, a penalty term on the regression parameters is added which shrinks the number of parameters. Common approaches to estimate the parameter in sparse regression are, however, sensitive to influential observations and new methods are needed. Alfons *et al.* (2013) and Park *et al.* (2014) proposed robust estimation methods, where influential observations are not harmful to the resulting estimates.

1.3 Introduction to influence analysis in nonlinear regression

In this thesis, new tools for conducting influence analysis in nonlinear regression are proposed. The nonlinear regression model referred to in this thesis is a model where the relationship between the variables is a function that is nonlinear in its parameters. We assume that the error term enters the model linearly. The motivation for using the nonlinear regression model arises from the need to describe real-life phenomena with a meaningful and realistic model. This meaning might be biological, chemical or physical (Bates and Watts, 1988). Thus, the function used to describe the relationship between the variables is often known and it is chosen due to the knowledge about the process generating the data. The Michaelis-Menten model (Michaelis and Menten, 1913) will be frequently used as an example of a nonlinear regression model throughout this thesis. The model is used, for instance, in studying enzymatic-catalyzed reactions, called enzyme kinetics. The motivation for using it is that the behavior of the enzymatic reaction's velocity (dependent variable) when adding different substrate concentrations (independent variable) to the process is known to be well described by the Michaelis-Menten model. Moreover, the parameters in the Michaelis-Menten model have chemically meaningful interpretations. A more detailed discussion of the Michaelis-Menten model will be given in Chapter 2 where the estimation of the parameters is also discussed.

The existing literature on influence analysis in nonlinear regression is not as extensive as for linear regression. One reason for this can be that there do not generally exist closed form estimators for the parameters in the nonlinear re-

gression model. Detection of influential observations on the fit of the nonlinear regression model is discussed by Cook and Weisberg (1982) and St. Laurent and Cook (1993). Cook and Weisberg (1982) developed a nonlinear version of Cook's distance, and St. Laurent and Cook (1993) proposed an approach for assessing the influence of the observations on the fitted values and on the estimate of the variance in a nonlinear regression model. These diagnostic tools will be more thoroughly discussed in Chapter 3. For more recent research results, see Galea *et al.* (2005) and Vanegas and Cysneiros (2010). Moreover, for a discussion of influence analysis concerning a specific nonlinear regression model, see Lemonte and Patriota (2011) and Vanegas *et al.* (2012).

Two graphical tools for identifying observations that are influential on the parameter estimates in a nonlinear regression model are presented in Cook (1987). One of these plots is referred to as the first-order extension of an added variable plot. This plot will be discussed in detail in Chapter 4.

1.4 Influence analysis regarding a test statistic

Testing of hypotheses is an important part in regression analysis. There are several testing procedures available for linear and nonlinear regression. In this thesis, the focus is on Rao's score test (Rao, 1948).

Several authors have presented work on the sensitivity of the score test statistic. Lee *et al.* (2004) used the score test to test for zero-inflation in count data. The null hypothesis under consideration was that the Poisson distribution fits the observed data well. However, for some applications there might be a large number of zeros in the data. In this case a more appropriate model could be a zero-inflated Poisson model. The alternative hypothesis is thus that the data follows a zero-inflated Poisson distribution. Another score test is also of interest, namely to test the null hypothesis that the data follows a zero-inflated Poisson model with the alternative that the zero-inflated negative binomial is a better model. When deriving the influence diagnostic, Lee *et al.* (2004) used the local influence approach, proposed by Cook (1986), which will be discussed in Chapter 3.

Lustbader and Moolgavkar (1985) derive an expression for the change in the score test statistic when deleting observations. This expression is derived for linear regression models, but is discussed in detail for deletion of entire risk sets in matched case-control studies and survival studies. Matched case-control studies are retrospective, observational studies where one seeks to determine the relationship between a risk factor and, for instance, a disease, using a par-

ticular matching variable to produce groups. With case-control data, it is natural to consider change in the score test for deletion of entire risk sets, i.e. the number of subjects at risk of experiencing a certain event. In survival analysis it is more desirable to compute the diagnostic for each individual.

Chen (1985) discussed the robustness of score tests for generalized linear regression models. The robustness referred to here is the robustness against the functional form chosen under the alternative hypothesis. Chen also discussed how the score test statistic can be made more robust against possible extreme observations. Moreover, Li (2001) discussed the sensitivity of the score test, the Wald test and the likelihood ratio test in relation to nuisance parameters, i.e. how the corresponding test statistics are affected by changes in the values of the nuisance parameters. Furthermore, see Vanegas *et al.* (2012, 2013) for a discussion of influence analysis concerning the F-test.

1.5 Aims of the dissertation

The general purpose of this thesis is to develop diagnostic tools for nonlinear regression models with additive error terms. More specifically, five aims can be outlined.

The first aim of this dissertation is to propose a new approach for detecting single observations with high influence on the parameter estimates in a nonlinear regression model. There is a lack of existing approaches for finding observations with high influence on a specific parameter estimate in nonlinear regression models and an aim is that our new approach should have this property.

In real-life studies data sets seldom contain only one influential observation, and therefore methods for finding multiple influential observations are needed. Multiple influential observations have not yet been discussed in the literature on influence analysis in nonlinear regression. A second aim of this thesis is therefore to extend the approach for finding single influential observations for detecting multiple influential observations.

The third aim of the thesis is to study the conditional influence, i.e. the influence of an observation on the parameter estimates given that another observation is deleted first. By using the conditional influence approach, influential observations can be revealed, observations that might go unnoticed when "unconditional" methods are used. Moreover, the use of the conditional influence approach can provide a more intimate knowledge about the data, since hidden

dependence among certain observations in the data set can be revealed.

A further aim of the present thesis is to evaluate how results from testing hypotheses about the parameters in a nonlinear regression model are affected by individual observations. The focus is on a particular test, namely Rao's score test. This test has the advantage, over other tests such as the likelihood ratio test, that only quantities evaluated for the parameter estimates under the null hypothesis need to be considered when constructing the test statistic. Hence, the derivation of diagnostic tools might be less complicated compared to tests where parameter estimates under both the null and the alternative were to be considered.

Graphical exploration of the data is of the utmost importance. Furthermore, graphical inspection of the observations' contribution to a test statistic is also of great interest. Hence, the fourth aim is to construct a plot that allows for visual identification of influential observations on the score test statistic. However, a graphical tool is used for explorative purposes and does not quantify the influence of the observations. To add more information to the influence analysis concerning the test statistic, a fifth aim is to propose influence measures that can be used to assess the influence of observations on the score test statistic.

To summarize, the aims of this thesis are as follows:

- To propose a new approach to assessing the influence of a single observation on the parameter estimates in nonlinear regression models.
- To extend the influence approach concerning single observations to assessing the influence of multiple observations.
- To propose an approach for assessing conditional influence, i.e. influence of an observation conditional on the deletion of another observation.
- To develop a graphical tool for explorative data analysis, where observations with high influence on the score test statistic can be identified.
- To propose influence measures for assessing the influence of observations on the score test statistic.

Chapters 1-4 presents the appropriate background for a discussion of the above listed aims whereas Chapters 5-7 include a more explicit discussion of the aims.

2. Regression models

This section gives mainly a brief overview of existing results concerning estimation and testing in regression models. In particular we focus on least squares estimation in nonlinear regression models, which are central for this thesis. Another focus in this chapter is Rao's score test (Rao, 1948), since one of the new results obtained in this thesis is closely connected to the score test.

2.1 Linear regression models and least squares estimation

In regression, the relationship between a response variable and explanatory variables is often represented by a functional relationship, f, and an additive error term. The function, f, is called the expectation function. When f is linear in its parameters, we may write the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{y} : n \times 1$ is a response vector, $\boldsymbol{X} : n \times p$ is the matrix of p explanatory variables, $\boldsymbol{\beta} : p \times 1$ is a vector of unknown regression parameters and $\boldsymbol{\varepsilon} : n \times 1$ is random error, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$. Here $\mathbf{0}_n : n \times 1$ is a vector of zeros and $\mathbf{I}_n : n \times n$ is the identity matrix.

The parameters in (2.1) are often estimated by the method of least squares or maximum likelihood. Both methods yield the following estimator of $\boldsymbol{\beta}$ in (2.1)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y},$$

assuming that the rank of $\boldsymbol{X}$ equals p and where T denotes the transpose of the matrix. Moreover, the maximum likelihood estimator of σ^2 is

$$n\widehat{\sigma}^2 = \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right).$$

Linear regression models are fairly flexible since even a nonlinear behavior of the data can be modeled by introducing nonlinear explanatory variables. An

example of a linear regression model with nonlinear explanatory variables is the polynomial regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, \ldots, n.$$

However, there is a limit to what can adequately be approximated by a linear model. Moreover, it may be difficult to interpret the results. If a linear regression model does not seem to fit the data well, an alternative solution might be to use a model that is not linear in its parameters, i.e. a nonlinear regression model.

2.2 Nonlinear regression models and estimation

In this section we introduce the nonlinear regression model, the estimation process, provide some examples of different applications of nonlinear regression models and give a briefly discuss the geometry in nonlinear regression.

Nonlinear regression models are widely used in many areas, such as economics, agriculture and biology. The decision to use a nonlinear regression model can be made on the basis of the theoretical knowledge about the problem at hand and the process generating the data. The function f is usually entirely known except for the parameters in the model. The parameters are often meaningful to the researcher or scientist, where the meaning can be for example graphical, physical, biological or chemical.

In this thesis we assume that a nonlinear regression model has a known f, and that this function is chosen due to the knowledge about the process generating the data. The regression model is not linear in its parameters (can be partly linear) and the error term is assumed to be additive. The general form of the model is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{2.2}$$

where $\boldsymbol{y} : n \times 1$ is a response vector, $\boldsymbol{X} : n \times p$ is the matrix of explanatory variables, $\boldsymbol{\theta} : q \times 1$ is a vector of unknown parameters and $\boldsymbol{\varepsilon} : n \times 1$ is the random error, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$.

Nonlinear models have many applications to real life problems and next we consider two examples of nonlinear regression models.

Example 2.1. The Michaelis-Menten model in enzyme kinetics

Consider a scientist who will study enzyme-catalyzed reactions. The scientist knows that the initial velocity of an enzymatic reaction follows Michaelis-Menten kinetics. That is, the relationship between y, the initial velocity of the enzymatic reaction, and x, substrate concentration, is modeled by the Michaelis-Menten equation

$$f(x) = \frac{V_{max}x}{K_{max}+x}, \tag{2.3}$$

where V_{max} and K_{max} are unknown parameters, which will be explained later. In this case, due to knowledge about chemical reactions, the function f is known to the scientist and there is no need to search for the correct functional relationship between y and x. For more details on the theoretical basis of the Michaelis-Menten equation see Briggs and Haldane (1925).

The Michaelis-Menten equation (2.3) can be used to formulate a nonlinear regression model by assuming an additive error term

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.4}$$

where $\theta_1 = V_{max}$ and $\theta_2 = K_{max}$ and $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Interested readers may see Richie and Prvan (1996), Pasaribu (1999) and Dette and Kunert (2014) for statistical analysis of enzyme kinetics data using the Michaelis-Menten equation.

It can be seen from (2.4) that the parameter θ_1 enters the model linearly but the parameter θ_2 enters nonlinearly, and thus the relationship between y and x is nonlinear. In the model (2.4), the parameters θ_1 and θ_2 have physical interpretations. The parameter θ_1 is the maximum initial velocity, which is theoretically attained when the enzyme has been saturated with respect to concentration of a substrate. The second parameter, θ_2, is the Michaelis parameter, which equals the concentration of substrate for "half-maximum" initial velocity. When the parameters in the model are estimated they are dependent, since a change in θ_1 results in a change in θ_2 as well. The Michaelis-Menten curve, with $\theta_1 = 0.9$ and $\theta_2 = 0.2$, is given in Figure 2.1.

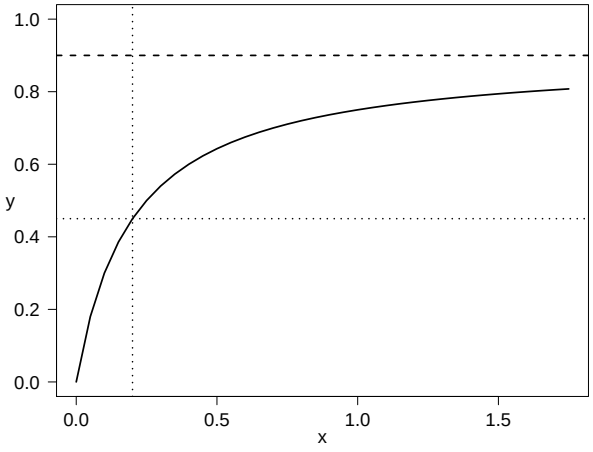


Figure 2.1: The Michaelis-Menten curve where y is the initial velocity and x is the substrate concentration. The parameter values are $\theta_1 = 0.9$ and $\theta_2 = 0.2$. The dashed line represents the value of θ_1, the dotted horizontal line represents the value of y that is half of θ_1, and the dotted vertical line represents the value of θ_2.

In the next example we will describe another interesting nonlinear regression model.

Example 2.2. The Gompertz Growth Curve Model

In microbiology, models are often used to describe the behavior or growth of microorganisms under different physical or chemical conditions. In order to build these models, growth is measured and modeled. For this purpose, it is common to use a type of nonlinear models called growth curve models. The Gompertz growth curve model is of particular interest, and it is defined as

$$f(t) = k\exp(-\exp(a-bt)), \tag{2.5}$$

where $f(t)$ is the size of the population at time t, and k, a and b are unknown parameters. See e.g. Zweitering *et al.* (1990) for a discussion of modeling bacterial growth with growth curve models and Chakraborty *et al.* (2014) for statistical analysis of the Gompertz growth curve model.

A common feature of the Gompertz growth curve models are that they have two asymptotes: the curve approaches zero as $x \to -\infty$, a positive constant as $x \to \infty$, and accelerates to a maximum value, after which the growth rate declines. The point (x, f) where the growth rate is maximum is called point of

inflection. For the Gompertz growth curve model (2.5), the curve approaches k as $x \to \infty$. Another property of the Gompertz growth curve model is that it is not symmetric around the point of inflection.

Zwietering *et al.* (1990) analyzed growth data of Lactobacillus plantarum. Bacterial growth often shows a phase in which the specific growth rate starts at a value of zero. The growth rate then accelerates to a maximal value in a certain period of time, resulting in a so-called lag time. Thereafter, the growth curve enters a final phase in which the growth rate decreases and finally reaches zero. The size of the population of bacteria is then approaching an asymptote.

Different growth curve models can be used to describe the behavior of bacterial growth. However, the most suitable model should contain parameters that are micro-biologically relevant. One of the candidate models that has been used is the modified Gompertz growth curve model, which is given by

$$f(t) = A \exp\left[-\exp\left[(\mu_m e/A)(\lambda - t) + 1\right]\right],$$

where A is the asymptote as $x \to \infty$, λ is the lag time and μ_m is the maximum growth rate in a certain period of time. An alternative definition of λ and μ_m is given by considering a tangent line at the point of inflection. The parameter μ_m is defined as the slope of the tangent line and the parameter λ is the intercept of the tangent line. An example of a growth curve is given in Figure 2.2, where $A = 20$, $\lambda = 5/3$ and $\mu_m = 12/e$.

By assuming an additive error term, the following nonlinear regression model can be formulated

$$y_i = A \exp\{-\exp\left[(\mu_m e/A)(\lambda - t_i) + 1\right]\} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$.

To estimate parameters in nonlinear regression models, least squares or maximum likelihood methods are often used. These methods of estimation yield the same mean parameter estimates when the errors in the nonlinear regression model are independent, normally distributed and have constant variance, i.e. $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$. In contrast to linear regression, analytical solutions for the least squares and maximum likelihood estimators can generally not be found. Instead, numerical algorithms are required. The perhaps most well known algorithms are the Gauss-Newton (GN) algorithm and the Newton (N) algorithm. The GN algorithm is a modification of the N algorithm, proposed by Gauss in 1809. Though the theory behind these algorithms is old, they are

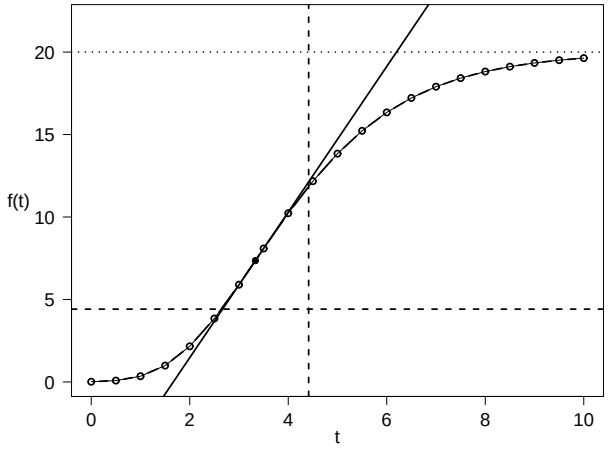


Figure 2.2: A growth curve where $f(t)$ is the size of the population and t is time. The solid line represents the tangent line at the point of inflection, represented by the filled circle. The slope of the tangent line is equal to $\mu_m = 12/e$. The lag time, $\lambda = 5/3$, is the intercept of the tangent line. The dotted line represents the asymptote, $A = 20$.

still very useful. However, nowadays there are numerous modifications that can make them more reliable. Examples of such modifications are the quasi-Newton method, Hartley's method and the Levenberg-Marquardt method. See for instance Nocedal and Wright (2006) for a thorough discussion about numerical optimization, and Seber and Wild (2003) for a detailed description of the algorithms mentioned above and others. Next the unmodified GN algorithm will be illustrated, since this algorithm forms the basis of a number of least squares problems.

The problem is to find $\boldsymbol{\theta}$ that minimizes

$$(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}))^T (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})), \tag{2.6}$$

in (2.2). In the GN algorithm the expansion of $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ is utilized in a Taylor expansion around an initial value, $\boldsymbol{\theta}^{(0)}$, called starting value. The expansion of $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ by the Taylor expansion around the starting value results in a linear

model given by

$$\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) \approx \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}^{(0)}) + \left.\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|^{T}_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(0)}} \left(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)}\right), \tag{2.7}$$

where the derivative is defined in Appendix A. Using the approximation (2.7) in (2.6), the minimization problem is converted to a linear least squares problem, namely

$$\left(\mathbf{r}^{(0)} - \boldsymbol{F}^{(0)\,T}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})\right)^{T}\left(\mathbf{r}^{(0)} - \boldsymbol{F}^{(0)\,T}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})\right), \tag{2.8}$$

where $\boldsymbol{F}^{(0)} = \left.\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(0)}}$ and $\mathbf{r}^{(0)} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}^{(0)})$.

Now, minimizing (2.8) yields

$$\left(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)}\right) = \left(\boldsymbol{F}^{(0)}\boldsymbol{F}^{(0)\,T}\right)^{-1}\boldsymbol{F}^{(0)}\boldsymbol{r}^{(0)},$$

assuming that the matrix inverse exists, leading to the GN algorithm

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)},$$

where $\boldsymbol{\delta}^{(0)} = \left(\boldsymbol{F}^{(0)}\boldsymbol{F}^{(0)\,T}\right)^{-1}\boldsymbol{F}^{(0)}\boldsymbol{r}^{(0)}$ is referred to as the Gauss increment.

The process of updating $\boldsymbol{\theta}$ is repeated until the increment, $\boldsymbol{\delta}$, is so small that there is no useful change in the elements of the parameter vector and the process results in the final estimate, $\widehat{\boldsymbol{\theta}}$. The GN algorithm is convergent, i.e. the iterated values tend to the least squares estimate of $\boldsymbol{\theta}$, as the number of iterations tends to infinity, provided that the starting value is close enough to the true $\boldsymbol{\theta}$. Moreover, there are some restrictions that need to be fulfilled in order for the algorithm to provide the least squares estimate, see Seber and Wild (2003).

In order to ensure a successful nonlinear regression analysis, one should prioritize the task of obtaining good starting values. One approach for finding starting values is to interpret the behavior of the expectation function in terms of the parameters, analytically or graphically. Other approaches are to use information available from previous, or related, experiments, or to transform the expectation function into a form that can be easily estimated. Examples of how starting values can be obtained are listed in Bates and Watts (1988).

2.2.1 Geometry of nonlinear regression

To get a flavor of the challenges that come when working with nonlinear regression models we briefly discuss the geometry of nonlinear least squares.

Let Θ denote the subset of $\mathbb{R}^q$ consisting of all possible parameter values $\boldsymbol{\theta}$. Moreover, define M to be the surface in $\mathbb{R}^n$, such that it contains $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ in the parameter space, i.e.

$$M = \{\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\} \subset \mathbb{R}^n.$$

The set M is called the expectation surface. If the function $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, the expectation surface is a plane, i.e. a linear surface. However, for nonlinear regression models the expectation surface is not a linear surface. The nonlinearity of the expectation surface results in challenges when analyzing nonlinear regression models and techniques used for linear regression models must be extended, which introduces considerable complexity.

To overcome these challenges, one idea is to use a linear approximation of the expectation surface through the tangent plane. The tangent plane of the expectation surface at the point $\widehat{\boldsymbol{\theta}}$ is given by the equations

$$\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) = \boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}) + \boldsymbol{F}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}), \tag{2.9}$$

where

$$\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) = \left(\boldsymbol{F}_1(\widehat{\boldsymbol{\theta}}),\ldots,\boldsymbol{F}_n(\widehat{\boldsymbol{\theta}})\right) = \frac{d}{d\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

The tangent plane, defined by the equations (2.9), is the space spanned by the columns of $\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})$ and it is a linear, local approximation to the expectation surface in a neighborhood of $\widehat{\boldsymbol{\theta}}$. This approximation will be appropriate if $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ is reasonably flat in the region near $\widehat{\boldsymbol{\theta}}$. There exist techniques to evaluate if $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ is reasonably flat, but the discussion of these techniques will be omitted here and we refer interested readers to e.g. Bates and Watts (1988) and Seber and Wild (2003).

In this thesis, we will utilize the linear approximation to the expectation surface via the tangent plane repeatedly. For instance, we will make use of the matrix $\boldsymbol{P}_{\boldsymbol{F}}$, defined to be $\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}))^{-1}\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$, which is a matrix that projects onto the tangent plane. The projection matrix $\boldsymbol{P}_{\boldsymbol{F}}$ will be more thoroughly described in Chapter 3. In Chapter 5 we will demonstrate the role of $\boldsymbol{P}_{\boldsymbol{F}}$ in influence analysis.

2.3 Score testing in regression analysis

There are several techniques available for testing hypotheses about the parameters in a nonlinear regression model, discussed by, for instance, Gallant (1987) and Seber and Wild (2003). Hypothesis testing procedures addressed in most text-books on nonlinear regression models comprise three classical tests: the likelihood ratio test (Neyman and Pearson, 1928), the Wald test (Wald, 1943) and Rao's score test (Rao, 1948). Various modifications of them are also discussed, such as the efficient score test considered by Hamilton (1986) and a re-scaling of the score test considered by Gallant (1987). Other examples of modifications of these tests are given in Hamilton and Wiens (1987), where corrections of the likelihood ratio test and the efficient score test are made due to the nonlinearity of the expectation surface. Moreover, Markatou and Manos (1996) discussed robust tests in nonlinear regression and the extensions of the Wald test and the score test in particular. A comparison in power between the three classical tests is done by Gallant (1987), where it is found that the likelihood ratio test has slightly better power than other tests. However, the Wald test and the score test only require the estimates of the parameters under the alternative and the null hypothesis, respectively, and they are therefore less computationally demanding than the likelihood ratio test. The focus in this thesis is on the score test, since one of our new results obtained in this thesis is closely connected to this test.

2.3.1 The score test in linear regression

In this section the score test will be derived for linear regression models, which will be followed by the derivation of the score test for nonlinear regression models in Section 2.3.2.

Consider the linear regression model (2.1) and without loss of generality, and consider testing a single parameter in the model. For the derivation of the score test, where restrictions on several parameters are specified in the null hypothesis, see Chen (1983). We let $\boldsymbol{\Psi} = \left(\boldsymbol{\beta}^T, \sigma^2\right)^T$ be the parameter space and the null hypothesis of interest

$$H_0 : \boldsymbol{\Psi} = \boldsymbol{\Psi}^0, \tag{2.10}$$

where $\boldsymbol{\Psi}^0 = \left(\beta_0, \ldots, \beta_{p-1}, 0, \sigma^2\right)^T$.

The score test is based on the score function, which is the partial derivative of the log likelihood function, with respect to the parameters. The likelihood

function for $\boldsymbol{y}$ in (2.1) is

$$L(\boldsymbol{\Psi},\boldsymbol{y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\}, \tag{2.11}$$

the log likelihood function, $\ell = \ln L(\boldsymbol{\Psi},\boldsymbol{y})$, is the following

$$\ell = -\frac{2}{n}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}), \tag{2.12}$$

and the score vector is defined as

$$\boldsymbol{U}(\boldsymbol{\Psi}) = \frac{d\ell}{d\boldsymbol{\Psi}}.$$

Now, let $\widetilde{\boldsymbol{\Psi}} = \left(\widetilde{\beta}_0,\ldots,\widetilde{\beta}_{p-1},\,0,\,\widetilde{\sigma}^2\right)^T$ denote the maximum likelihood estimate of $\boldsymbol{\Psi}$ under the null hypothesis (2.10). The score test statistic for the hypothesis in (2.10) is given by

$$S(\widetilde{\boldsymbol{\Psi}}) = \mathbf{U}^T(\widetilde{\boldsymbol{\Psi}})\mathbf{I}^{-1}(\widetilde{\boldsymbol{\Psi}})\mathbf{U}(\widetilde{\boldsymbol{\Psi}}), \tag{2.13}$$

where $\mathbf{U}(\widetilde{\boldsymbol{\Psi}})$ and $\mathbf{I}(\widetilde{\boldsymbol{\Psi}})$ are the score vector and the Fisher information matrix, respectively, both evaluated for the parameter estimates under the null hypothesis. The Fisher information matrix is defined to be

$$\begin{aligned}
\mathbf{I}(\widetilde{\boldsymbol{\Psi}}) &= E\left[\mathbf{U}(\boldsymbol{\Psi})\mathbf{U}^T(\boldsymbol{\Psi})\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} \\
&= \begin{pmatrix} E\left[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}^T(\boldsymbol{\beta})\right] & E\left[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}(\sigma^2)\right] \\ E\left[\mathbf{U}(\sigma^2)\mathbf{U}^T(\boldsymbol{\beta})\right] & E\left[\mathbf{U}(\sigma^2)\mathbf{U}(\sigma^2)\right] \end{pmatrix}_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} \\
&= \begin{pmatrix} E\left[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}^T(\boldsymbol{\beta})\right] & \mathbf{0}_p \\ \mathbf{0}_p^T & E\left[\mathbf{U}(\sigma^2)\mathbf{U}(\sigma^2)\right] \end{pmatrix}_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}},
\end{aligned} \tag{2.14}$$

since the first central moment of the normal distribution is zero.

If the score vector is evaluated using the estimates under the null hypothesis we get

$$\mathbf{U}(\widetilde{\boldsymbol{\Psi}}) = \begin{pmatrix} \mathbf{U}(\widetilde{\boldsymbol{\beta}}) \\ \mathbf{U}(\widetilde{\sigma}^2) \end{pmatrix}$$

where

$$\mathbf{U}(\widetilde{\boldsymbol{\beta}}) = \left.\frac{d\ell}{d\boldsymbol{\beta}}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{X}^T(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}),$$

and

$$\mathbf{U}(\widetilde{\sigma}^2) = \left.\frac{d\ell}{d\sigma^2}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} = -\frac{n}{2\widetilde{\sigma}^2} + \frac{1}{2\widetilde{\sigma}^4}(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}) = 0,$$

since $\widetilde{\sigma}^2 = (1/n)(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})$ is the maximum likelihood estimate of σ^2. Therefore

$$\mathbf{U}(\widetilde{\boldsymbol{\Psi}}) = \begin{pmatrix} \mathbf{U}(\widetilde{\boldsymbol{\beta}}) \\ 0 \end{pmatrix}. \tag{2.15}$$

Inserting (2.14) and (2.15) in (2.13) we get

$$S(\widetilde{\boldsymbol{\Psi}}) = \mathbf{U}^T(\widetilde{\boldsymbol{\beta}})\mathbf{I}^{-1}_{\widetilde{\beta}\widetilde{\beta}}\mathbf{U}(\widetilde{\boldsymbol{\beta}}),$$

where

$$\mathbf{I}_{\widetilde{\beta}\widetilde{\beta}} = E\left[\mathbf{U}(\boldsymbol{\beta})\mathbf{U}^T(\boldsymbol{\beta})\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{X}^T\boldsymbol{X}.$$

Using the results above, the score test statistic in (2.13) can be simplified, resulting in the explicit expression given by

$$S(\widetilde{\boldsymbol{\beta}}) = \frac{1}{\widetilde{\sigma}^2}\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right). \tag{2.16}$$

We will now show that, under the null hypothesis (2.10), the score test statistic (2.16) has asymptotically a χ^2-distribution with one degree of freedom.

Use the partition $\boldsymbol{X} = (\boldsymbol{X}_1 \vdots \boldsymbol{x}_p)$, where $\boldsymbol{X}_1 : n \times (p-1)$ with rank $p-1$ and $\boldsymbol{x}_p : n \times 1$. Observe that

$$\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}} = (\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y},$$

where $\boldsymbol{P}_{\boldsymbol{X}_1} = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$ and that

$$\begin{aligned} S(\widetilde{\boldsymbol{\beta}}) &= \frac{1}{\widetilde{\sigma}^2}\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right) \\ &= \frac{1}{\widetilde{\sigma}^2}\boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{P}_{\boldsymbol{X}}(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} \\ &= \frac{1}{\widetilde{\sigma}^2}\boldsymbol{y}^T(\boldsymbol{P}_{\boldsymbol{X}}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}. \end{aligned}$$

In the following we need to utilize the following property of the projection matrix $\boldsymbol{P}_{\boldsymbol{X}}$.

Proposition 2.3.1. *The projection matrix $\boldsymbol{P_X}$ can be written as a sum of projection matrices such that*

$$\begin{aligned} \boldsymbol{P_X} &= \boldsymbol{P_{X_1}} + \boldsymbol{P_{x^*}} \\ &= \boldsymbol{P_{X_1}} + \frac{(\mathbf{I} - \boldsymbol{P_{X_1}})\boldsymbol{x}_p \boldsymbol{x}_p^T (\mathbf{I} - \boldsymbol{P_{X_1}})}{\boldsymbol{x}_p^T (\mathbf{I} - \boldsymbol{P_{X_1}})\boldsymbol{x}_p}. \end{aligned} \tag{2.17}$$

Proof. We want to prove that $\boldsymbol{P_X} = \boldsymbol{P_{X_1}} + \boldsymbol{P_{x^*}}$, where $\boldsymbol{x}^* = (\mathbf{I} - \boldsymbol{P}_{X_1})\boldsymbol{x}_p$.

Using the partitioned form of $\boldsymbol{X} = (\boldsymbol{X}_1 \vdots \boldsymbol{x}_p)$ in the expression of $\boldsymbol{P_X}$ yields

$$\boldsymbol{P_X} = \left(\boldsymbol{X}_1 \vdots \boldsymbol{x}_p\right) \begin{pmatrix} \boldsymbol{X}_1^T \boldsymbol{X}_1 & \boldsymbol{X}_1^T \boldsymbol{x}_p \\ \boldsymbol{x}_p^T \boldsymbol{X}_1 & \boldsymbol{x}_p^T \boldsymbol{x}_p \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}_1^T \\ \boldsymbol{x}_p^T \end{pmatrix}. \tag{2.18}$$

In the continuation of the proof we are using well known rules of inversion of a partitioned matrix, see e.g. Chatterjee and Hadi (1988, p. 15). Let $\boldsymbol{M} : q \times q$ be partitioned into a block form

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^T & c \end{pmatrix},$$

where $\boldsymbol{A} : (q-1) \times (q-1)$ is an invertible matrix, $\boldsymbol{b} : (q-1) \times 1$ and c is a scalar. Observe that when $\boldsymbol{X}$ is partitioned, the matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ can be written in the same form as $\boldsymbol{M}$. The inverse of $\boldsymbol{M}$ equals

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} + \frac{1}{k}\boldsymbol{A}^{-1}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{A}^{-1} & -\frac{1}{k}\boldsymbol{A}^{-1}\boldsymbol{b} \\ -\frac{1}{k}\boldsymbol{b}^T\boldsymbol{A}^{-1} & \frac{1}{k} \end{pmatrix}, \tag{2.19}$$

where $k = c - \boldsymbol{b}^T\boldsymbol{A}^{-1}\boldsymbol{b}$.

Applying the rules of inversion of a partitioned matrix defined in (2.19) the matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ can be expressed as

$$\left(\begin{array}{c|c} \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1} + \frac{1}{k}\left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{x}_p\boldsymbol{x}_p^T\boldsymbol{X}_1\left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1} & -\frac{1}{k}\left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{x}_p \\ -\frac{1}{k}\boldsymbol{x}_p^T\boldsymbol{X}_1\left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1} & \frac{1}{k} \end{array}\right) \tag{2.20}$$

where $k = \boldsymbol{x}_p^T\boldsymbol{x}_p - \boldsymbol{x}_p^T\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{x}_p = \boldsymbol{x}_p^T(\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p$.

Inserting (2.20) in (2.18) yields

$$\begin{aligned} \boldsymbol{P}_{\boldsymbol{X}} &= \boldsymbol{P}_{\boldsymbol{X}_1} + k^{-1}\left(\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{x}_p\boldsymbol{x}_p^T\boldsymbol{P}_{\boldsymbol{X}_1} - \boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{x}_p\boldsymbol{x}_p^T - \boldsymbol{x}_p\boldsymbol{x}_p^T\boldsymbol{P}_{\boldsymbol{X}_1} + \boldsymbol{x}_p\boldsymbol{x}_p^T\right) \\ &= \boldsymbol{P}_{\boldsymbol{X}_1} + k^{-1}\left((\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\right). \end{aligned}$$

Let us evaluate the second term on the right hand side of the expression above. Since k is a scalar we can write

$$k^{-1}\left((\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\right) = (\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\left(\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\right)^{-1}\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1}),$$

and since $(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})$ is idempotent, $\left(\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\right)$ can be written as $\left(\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\right)$. Hence, we can identify

$$(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\left(\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\right)^{-1}\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1}),$$

to be a projection matrix for $(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p$, which equals $\boldsymbol{x}^*$.

Hence,

$$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}_1} + \boldsymbol{P}_{\boldsymbol{x}^*},$$

and the proof is complete.

■

Using (2.17) the score test statistic can be written

$$S(\widetilde{\boldsymbol{\beta}}) = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{y}^T(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{y}^T\boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{y}. \tag{2.21}$$

Let us look at the distribution of $\boldsymbol{y}^T\boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{y}$. Under the null hypothesis we have that

$$\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}_1^T, 0)^T,$$

where $\widetilde{\boldsymbol{\beta}}_1 = (\widetilde{\beta}_1, \ldots, \widetilde{\beta}_{p-1})^T$. The expected value of $\boldsymbol{P}_{x^*}\boldsymbol{y}$ is

$$E\left[\boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{y}\right] = \boldsymbol{P}_{\boldsymbol{x}^*}E\left[\boldsymbol{y}\right] = \boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{X}_1\boldsymbol{\beta}_1 = \frac{(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{X}_1\boldsymbol{\beta}_1}{\boldsymbol{x}_p^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p} = \mathbf{0},$$

since $(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{X}_1\boldsymbol{\beta}_1 = \mathbf{0}$.

Therefore, since $\boldsymbol{P}_{\boldsymbol{x}^*}$ is idempotent and $rank(\boldsymbol{x}^*) = 1$,

$$\frac{1}{\sigma^2}\boldsymbol{y}^T\boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{y} \sim \chi^2_{(1)}.$$

Since $\widetilde{\sigma}^2$ converges to σ^2, according to Cramér's theorem (see Gut, 1995),

$$\frac{1}{\widetilde{\sigma}^2}\boldsymbol{y}^T\boldsymbol{P}_{\boldsymbol{x}^*}\boldsymbol{y} \to \chi^2_{(1)}.$$

Indeed one may note that (2.21) is exact F-distributed.

2.3.2 The score test in nonlinear regression

We will now derive the explicit expression of the score test statistic when testing a hypothesis about a parameter in a nonlinear regression model. Consider the nonlinear model (2.2) and let $\boldsymbol{\Psi} = \left(\boldsymbol{\theta}^T, \sigma^2\right)^T$, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$, be the parameter space. Without loss of generality, consider the following hypothesis about a single parameter

$$H_0 : \boldsymbol{\Psi} = \boldsymbol{\Psi}^0, \tag{2.22}$$

where $\boldsymbol{\Psi}^0 = \left(\theta_1, \ldots, \theta_{q-1}, 0, \sigma^2\right)^T$. Let $\widetilde{\boldsymbol{\Psi}} = (\widetilde{\boldsymbol{\theta}}^T, \widetilde{\sigma}^2)^T$ be the maximum likelihood estimate of $\boldsymbol{\Psi}$ under the null hypothesis (2.22).

The score test statistic for testing (2.22) is equal to the expression in (2.13), i.e

$$S(\widetilde{\boldsymbol{\Psi}}) = \mathbf{U}^T(\widetilde{\boldsymbol{\Psi}})\mathbf{I}^{-1}(\widetilde{\boldsymbol{\Psi}})\mathbf{U}(\widetilde{\boldsymbol{\Psi}}).$$

The likelihood function and the log likelihood function for the nonlinear regression model is equal to (2.11) and (2.12), respectively, with the function $\boldsymbol{X\beta}$ replaced with $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$. Therefore, the score vector for $\widetilde{\boldsymbol{\theta}}$ is equal to

$$\mathbf{U}(\widetilde{\boldsymbol{\theta}}) = \left.\frac{d\ell}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}},$$

where

$$\widetilde{\mathbf{r}} = (\widetilde{r}_k) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}) \tag{2.23}$$

are the residuals under the null hypothesis, and $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) : q \times n$ is the matrix such that

$$\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) = \left(\boldsymbol{F}_1(\widetilde{\boldsymbol{\theta}}), \ldots, \boldsymbol{F}_n(\widetilde{\boldsymbol{\theta}})\right) = \left.\frac{d}{d\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}. \tag{2.24}$$

As in the linear regression case, $\mathbf{U}(\widetilde{\sigma}^2) = 0$, since

$$\widetilde{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}))^T(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}))$$

is the maximum likelihood estimate of σ^2. Hence, we have that

$$\mathbf{U}(\widetilde{\boldsymbol{\Psi}}) = \begin{pmatrix} \mathbf{U}(\widetilde{\boldsymbol{\theta}}) \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\widetilde{\sigma}^2}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}} \\ 0 \end{pmatrix}. \tag{2.25}$$

The information matrix, $\mathbf{I}(\widetilde{\boldsymbol{\Psi}})$, is given by

$$\mathbf{I}(\widetilde{\boldsymbol{\Psi}}) = \begin{pmatrix} E\left[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}^T(\boldsymbol{\theta})\right] & \mathbf{0}_q \\ \mathbf{0}_q^T & E\left[\mathbf{U}(\sigma^2)\mathbf{U}^T(\sigma^2)\right] \end{pmatrix}_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}}, \tag{2.26}$$

for the same reason as in the linear regression case. If (2.25) and (2.26) are inserted in the expression for $S(\widetilde{\boldsymbol{\Psi}})$, given in (2.13), we get

$$S(\widetilde{\boldsymbol{\Psi}}) = \mathbf{U}^T(\widetilde{\boldsymbol{\theta}})\mathbf{I}^{-1}_{\widetilde{\theta}\widetilde{\theta}}\mathbf{U}(\widetilde{\boldsymbol{\theta}}),$$

where

$$\begin{aligned} \mathbf{I}_{\widetilde{\theta}\widetilde{\theta}} &= E\left[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}^T(\boldsymbol{\theta})\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} \\ &= \left(\frac{1}{\sigma^4}\boldsymbol{F}(\boldsymbol{\theta})E\left[(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))^T\right]\boldsymbol{F}^T(\boldsymbol{\theta})\right)_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}} \\ &= \frac{1}{\widetilde{\sigma}^2}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}}). \end{aligned}$$

Using the results above we find that the score test statistic for testing the hypothesis (2.22) can be written

$$S(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\widetilde{\sigma}^2}\widetilde{\mathbf{r}}^T\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)^{-1}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}}, \tag{2.27}$$

where $\widetilde{\boldsymbol{r}}$ and $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ are defined in (2.23) and (2.24), respectively.

Under the null hypothesis (2.22), the score test statistic in (2.27) has asymptotically a χ^2 distribution with 1 degree of freedom, see Seber and Wild (2003).

3. Influence analysis in regression

It is well understood that not all observations in a data set play an equal role for inference about a statistical model. For instance, in regression analysis the character of the regression line may be determined by only a few observations while most of the data is somewhat ignored. Such observations, that substantially influence the results of the inference and/or the data analysis, are called influential observations. The study of the effect they have on the inference is called influence analysis or sensitivity analysis.

Detection of influential observations is an important part of the statistical paradigm. Examination of the data, and the ability to find influential observations, can be beneficial in several ways. It can

- help reveal spurious observations that might be a result of errors during the collection or the processing of the data. Examples of such errors are measurement errors and keypunching errors.
- make the researcher aware of the possibility that some part of the data might come from another regime, or subpopulation, that have very different features compared to the population of study.
- give a hint of what properties data should have, if additional data collection is relevant, for instance in order to produce a more stable model and insensitive estimates.
- give the researcher an increased confidence in the results and intimate knowledge about the data.

Influence analysis in regression has been a very active area of research. Nowadays, there are many strategies available for detecting influential observations, for instance graphical displays. A popular graphical tool for detection of observations with a substantial influence on the parameter estimates in linear regression models is the added variable plot, proposed by Mosteller and Tukey (1977). In Cook (1987), a similar plot is proposed for use in nonlinear regression. A detailed discussion of diagnostic plots is given in Chapter 4.

Other approaches to identifying influential observations can be to use influence measures that enable a quantification of the observations' influence on various aspects of the regression analysis. The most well-known influence measure is Cook's distance, proposed by Cook (1977) and widely used in linear regression. This influence measure is used for assessing the influence of the individual observations on the vector of estimated regression parameters. The explicit expression of Cook's distance for model (2.1) is given by

$$C_k = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(k)})^T \boldsymbol{X}^T \boldsymbol{X} (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(k)})}{p\widehat{\sigma}^2}, \quad k = 1, \ldots, n, \tag{3.1}$$

where p is the number of explanatory variables in the model and $\widehat{\boldsymbol{\beta}}_{(k)}$ is the estimate of $\boldsymbol{\beta}$ when the kth observation is excluded from the calculations.

A version of Cook's distance for assessing the influence of the observations on the vector of estimated parameters in the nonlinear regression model (2.2) is proposed by Cook and Weisberg (1982). The explicit expression of this measure is given by

$$\frac{\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(k)}\right)^T \boldsymbol{F}(\widehat{\boldsymbol{\theta}}) \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) \left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(k)}\right)}{q\widehat{\sigma}^2}, \quad k = 1, \ldots, n, \tag{3.2}$$

where q is the number of parameters in the model, $\widehat{\boldsymbol{\theta}}_{(k)}$ is the estimate of $\boldsymbol{\theta}$ when the kth observation is excluded from the calculations, and $\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) : q \times n$ is a matrix of derivatives such that

$$\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) = \left(\boldsymbol{F}_1(\widehat{\boldsymbol{\theta}}), \ldots, \boldsymbol{F}_n(\widehat{\boldsymbol{\theta}})\right) = \frac{d}{d\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}, \tag{3.3}$$

where the derivative is defined in Appendix A.

Some other influence measures available for use in linear and nonlinear regression analysis will be further discussed in Chapter 5. Moreover, in Chapter 5, the new influence measures proposed in this thesis will be derived in detail.

The idea behind the construction of the new influence measure, proposed in this thesis, is to perform small perturbations in the model formulation. When the perturbations are imposed, the resulting model is referred to as the perturbed model. The perturbed model used in this thesis is defined to be

$$\boldsymbol{y}_\omega = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_\omega, \tag{3.4}$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2\mathbf{W}^{-1}(\omega))$ and $\mathbf{W}(\omega) : n \times n$ is a diagonal weight matrix. The expectation function $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ can be both linear and nonlinear in its parameters.

Using the diagonal weight matrix, $\mathbf{W}(\omega)$, we perturb the error variance in the regression model. We can choose to perturb the error variance of the kth observation by introducing the weight, $0 < \omega_k \leq 1$, as the kth diagonal element in $\mathbf{W}$, the other diagonal elements being equal to one. We can also choose to perturb the error variance for multiple observations simultaneously. If we perturb the error variance for all n observations in the data set we let the diagonal elements of $\mathbf{W}$ be equal to the vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$, where $0 < \omega_k \leq 1$, $k = 1, \ldots, n$.

The perturbed model (3.4) is used for e.g. estimation of the parameters. The estimates of the parameters are functions of the imposed perturbation weight, denoted $\widehat{\boldsymbol{\theta}}(\omega)$. The idea is to study the rate of change in the estimates as the weight approaches one. This approach to influence analysis is called the differentiation approach, see e.g. Chatterjee and Hadi (1988) and will be discussed in detail in Chapter 5. Moreover, we assume that there exists some null perturbation weight ω_0, so that $\widehat{\boldsymbol{\theta}}(\omega_0) = \widehat{\boldsymbol{\theta}}$ is the estimate from the unperturbed model, i.e. the nonlinear regression model (2.2). In our case $\omega_0 = 1$.

The structure of the imposed perturbations is called perturbation scheme. There exist other perturbation schemes than the one described above, where the error variance is perturbed.

We will now describe the case-weighted perturbation scheme. Let $0 \leq \omega_k \leq 1$ and let $\mathbf{W}(\omega_k) = diag(1, \ldots, 1, \omega_k, 1, \ldots, 1)$. For the linear regression model (2.1), define the perturbed model as

$$\boldsymbol{y}_\omega = \boldsymbol{X}_\omega\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}_\omega^T = \boldsymbol{y}^T\mathbf{W}(\omega_k)$ and $\boldsymbol{X}_\omega^T = \boldsymbol{X}^T\mathbf{W}(\omega_k)$. The estimator for $\boldsymbol{\beta}$ in the perturbed model is a function of ω_k and is equal to

$$\widehat{\boldsymbol{\beta}}(\omega_k) = \left(\boldsymbol{X}_\omega^T\boldsymbol{X}_\omega\right)^{-1}\boldsymbol{X}_\omega^T\boldsymbol{y}_\omega.$$

If $\omega_k = 1$ then we have that $\widehat{\boldsymbol{\beta}}(\omega_k) = \widehat{\boldsymbol{\beta}}$, the estimator of $\boldsymbol{\beta}$ in the unperturbed linear regression model (2.1). Moreover, if $\omega_k = 0$, then we have that $\widehat{\boldsymbol{\beta}}(\omega_k) = \widehat{\boldsymbol{\beta}}_{(k)}$, the estimator of $\boldsymbol{\beta}$ in the unperturbed linear regression model (2.1) when the kth observation is excluded from the calculations. Using the

case-weighted perturbation scheme with $\omega_k = 0$ is also referred to as the case-deletion approach. An example of diagnostic measures constructed using case-deletion is Cook's distance defined in (3.1) and the nonlinear version defined in (3.2). The approach of using case-deletion is also referred to as global influence.

According to Ross (1987), rather than using the case-weighted perturbation scheme, an alternative way of studying case-deletion in nonlinear regression is to define the case-deletion model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{d}_i \gamma + \boldsymbol{\varepsilon},$$

where $\boldsymbol{d}_i$ is the ith column of the identity matrix of size n and γ is an unknown parameter. Adding $\boldsymbol{d}_i \gamma$ to the model deletes the ith observation when the model is fitted. This approach can also be used for linear regression models, see e.g. Chatterjee and Hadi (1988).

Besides case-deletion, another approach widely used in influence analysis is the local influence approach. This approach was proposed by Cook (1986) and has had, and still has, a great impact on the research area of influence analysis. In the local influence approach the weights are not restricted to be zero, as in the case-deletion approach. Rather, they can vary between zero and one. This approach relies on a well-behaved likelihood since a central concept is to use the likelihood displacement, *LD*, in an influence graph.

The *LD* measures the amount that the maximum likelihood estimates, MLE's, of the parameters from the perturbed model are displaced from the MLE's of the parameters from the unperturbed model. The *LD* for the perturbed model (3.4) using a single perturbation weight, ω_k, is defined as

$$LD(\omega_k) = 2\left(\ln L(\widehat{\boldsymbol{\theta}}) - \ln L(\widehat{\boldsymbol{\theta}}(\omega_k))\right),$$

where $\ln L(\widehat{\boldsymbol{\theta}})$ and $\ln L(\widehat{\boldsymbol{\theta}}(\omega_k))$ are the log-likelihood functions for the unperturbed model and the perturbed model, respectively.

An influence graph is the graph of a statistic, which is a function of the perturbation weight, versus the perturbation weight. As an example, a graph of $LD(\omega_k)$ versus $0 < \omega_k \leq 1$ is an influence graph in $\mathbb{R}^2$. If we decide to use perturbation weights for all n observations, the resulting influence graph is a graph in $\mathbb{R}^{n+1}$.

A central influence measure in the local influence approach is the curvature C, of the influence graph in the neighborhood of null perturbation weight, $\omega_0 = 1$. If n observations are perturbed, another central diagnostic is the vector $\boldsymbol{\ell}$ in $\mathbb{R}^n$ that describes the direction of perturbation. Let $L(\boldsymbol{\theta})$ be the likelihood of an unperturbed model, not necessarily the nonlinear regression model (2.2). Let $\boldsymbol{\theta} : q \times 1$ be a vector of unknown parameters and $\boldsymbol{\omega} : n \times 1$ be the perturbation weights introduced to the unperturbed model, for instance the perturbations of the error variance as described above. Now, denote the curvature of the influence graph at the null perturbation by $C_{\boldsymbol{\ell}}$. Then,

$$C_{\boldsymbol{\ell}} = 2\boldsymbol{\ell}^T \Delta^T \ddot{\mathbf{L}}^{-1} \Delta \boldsymbol{\ell},$$

where

$$\begin{aligned} \ddot{\mathbf{L}} &= \frac{d}{d\boldsymbol{\theta}} \left(\frac{d \ln L(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\ \Delta &= \frac{d}{d\boldsymbol{\theta}} \left(\frac{d \ln L(\boldsymbol{\theta}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}} \right) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \end{aligned}$$

and where the matrix derivative is defined in Appendix A.

Cook (1986) suggested the use of $\boldsymbol{\ell}_{max}$, the direction causing the maximum curvature, C_{max}, as an influence measure. The vector $\boldsymbol{\ell}_{max}$ indicates how to perturb the model in order to obtain the greatest local change in the likelihood displacement. For instance, assume perturbing all observations in the data set simultaneously and suppose that the kth element of $\boldsymbol{\ell}_{max}$ is found to be relatively large. This indicates that perturbations in the weight ω_k of the kth observation may lead to substantial changes in the results of the analysis and that the kth observation is relatively influential.

The approach of local influence has several benefits. It is appealing as it allows for measuring the influence of a single observation as well as the assessment of the influence of multiple observations, which was a new idea at the time the article was written. It is further discussed in Cook (1986) how to assess the influence of the observations on subsets of parameters in the linear regression model (2.1). Moreover, the local influence approach is not restricted to linear regression models: it can be used for a variety of problems. In Cook (1986) the local influence approach is discussed, not only for linear regression models, but also for generalized linear models.

The local influence approach is extended to nonlinear regression models by St. Laurent and Cook (1993). The interest is in assessing the influence of the

observations on the fitted values and the estimate of the error variance when all n observations in the data set are perturbed. They also discussed the opportunity to assess the influence of a single observation on the fitted values and the estimate of the error variance.

Influential observations are closely connected to high-leverage observations and outliers. A deeper understanding of the diagnostic measures used to detect influential observations can be achieved when they are analyzed in terms of high-leverage observations and outliers. Therefore, we will devote a few paragraphs to define and discuss high-leverage observations and outliers.

According to Hoaglin and Welsch (1978), a high-leverage observation in linear regression analysis is an outlying observation in the $X-$space. The kth diagonal element of the projection matrix, defined as

$$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \tag{3.5}$$

,
is a measure of leverage for the kth observation, and it is denoted

$$p_{kk} = \boldsymbol{x}_k^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{x}_k.$$

The diagonal elements of the matrix $\boldsymbol{P_X}$ are called leverages since they can be thought of as the amount of leverage of the response value on the corresponding predicted value, i.e.

$$\boldsymbol{P_X y} = \widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}. \tag{3.6}$$

The matrix $\boldsymbol{P_X}$ is also known as the hat matrix or the prediction matrix. This is due to the fact that when $\boldsymbol{P_X}$ is post-multiplied by $\boldsymbol{y}$ it "puts a hat" on $\boldsymbol{y}$ and creates the predictions, as seen from (3.6).

It is worth noting that the values of the diagonal elements of $\boldsymbol{P_X}$ are between zero and one, i.e. $0 \leq p_{kk} \leq 1$. Moreover, $rank(\boldsymbol{P_X}) = trace(\boldsymbol{P_X}) = p$, where p is the numbers of columns of $\boldsymbol{X}$. As a consequence, the average of the diagonal elements in $\boldsymbol{P_X}$ is (p/n). Experience suggests that a reasonable rule of thumb for large values of p_{kk} is $p_{kk} > (2p/n)$. To read more about the projection matrix in linear regression, see for instance Hoaglin and Welsch (1978).

Deriving measures of leverage for observations in a nonlinear regression model is more complex than in the linear regression case, since the expectation surface is not a linear subspace. St. Laurent and Cook (1992) define two types

of leverages. One type of leverage is the tangent plane leverage. Consider the matrix of derivatives $\boldsymbol{F} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}})$ defined in (3.3). When $\boldsymbol{F}$ is of full column rank we make use of the matrix that projects onto the tangent plane, i.e. $\boldsymbol{P}_{\boldsymbol{F}} = \boldsymbol{F}^T(\boldsymbol{F}\boldsymbol{F}^T)^{-1}\boldsymbol{F}$. A measure of the tangent plane leverage for the kth observation is the kth diagonal elements of $\boldsymbol{P}_{\boldsymbol{F}}$ given by

$$p_{kk} = \boldsymbol{F}_k^T(\boldsymbol{F}\boldsymbol{F}^T)^{-1}\boldsymbol{F}_k, \tag{3.7}$$

where $\boldsymbol{F}_k^T$ is the kth row of $\boldsymbol{F}^T$. The matrix $\boldsymbol{P}_{\boldsymbol{F}}$ is referred to as the tangent plane leverage matrix (St. Laurent and Cook, 1992).

Another measure of leverage in nonlinear regression models is referred to by St. Laurent and Cook (1992) as the Jacobian leverage. The Jacobian leverage matrix is given by

$$\boldsymbol{J} = \boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T - \boldsymbol{G}(\boldsymbol{r}\otimes\mathbf{I}_q)\right)^{-1}\boldsymbol{F},$$

where $\boldsymbol{r} = (r_k) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}})$ is the $n-$vector of residuals, $\boldsymbol{F}$ is defined in (3.3) and $\boldsymbol{G} = \boldsymbol{G}(\widehat{\boldsymbol{\theta}})$ is a $q \times nq$ matrix of derivatives such that

$$\boldsymbol{G}(\widehat{\boldsymbol{\theta}}) = \frac{d}{d\boldsymbol{\theta}}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{d\boldsymbol{F}(\widehat{\boldsymbol{\theta}})}{d\widehat{\boldsymbol{\theta}}}.$$

St. Laurent and Cook (1992) argued that $\boldsymbol{P}_{\boldsymbol{F}}$ and $\boldsymbol{J}$ can differ dramatically if the tangent plane is not an adequate approximation of the nonlinear expectation surface (near $\widehat{\boldsymbol{\theta}}$). By inspection of $\boldsymbol{P}_{\boldsymbol{F}}$ and $\boldsymbol{J}$, we see that they will be different if the components of $\boldsymbol{G}(\boldsymbol{r}\otimes\mathbf{I}_q)$ dramatically differ from zero. However, St. Laurent and Cook (1992) suggested using $\boldsymbol{P}_{\boldsymbol{F}}$ whenever possible since it is easier to conduct computations and interpretation of the results is similar to linear regression. For example, the diagonal elements of $\boldsymbol{P}_{\boldsymbol{F}}$ have the following properties: $0 \leq p_{kk} \leq 1$ and $\sum_{k=1}^{n} p_{kk} = q$, where q is the number columns of $\boldsymbol{F}^T$. These properties do generally not hold for the diagonal elements of the Jacobian leverage matrix, see St. Laurent and Cook (1992). In accordance with the properties of the projection matrix $\boldsymbol{P}_{\boldsymbol{X}}$ for the linear regression model, we can in this thesis define a large value of the tangent plane leverage as $p_{kk} > (2q/n)$.

We often define an outlier to be an observation for which the residual is large in magnitude compared to the other observations. Outliers can be detected by analyzing the ordinary residuals $\boldsymbol{r} = (r_k) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}})$, $k = 1,\ldots,n$, and where the function $\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}})$ can be either linear or nonlinear in its parameters.

It is important to note the following: Outliers do not need to be influential observations, and influential observations do not need to be outliers. For examples see Andrews and Pregibon (1978) and Chatterjee and Hadi (1986). The

same applies for high-leverage observations, which do not need to be influential observations. However, plots of residuals and the diagonal elements of the projection matrix provide a good basis for influence measures. It can provide more understanding concerning why an observation is influential. To illustrate the relationship between an influence measure and high-leverage observations and outliers, rewrite Cook's distance, defined in (3.1), as follows

$$C_k = \frac{1}{p}\frac{p_{kk}}{1-p_{kk}} r_{k,stud},$$

where $r_{k,stud} = r_k/(\widehat{\sigma}\sqrt{1-p_{kk}})$ is the kth studentized residual and $\widehat{\sigma}^2 = (\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})/(n-p)$, see Chatterjee and Hadi (1988). From the equation above, we see that a high-leverage observation and/or an outlier will increase the value of Cooks' distance through p_{kk} and $r_{k,stud}$, respectively.

What one should do when having identified an influential observation is a question with no clear answer. However, Cook and Weisberg (1982) provide some advice. If unusual and influential observations are a consequence of, for instance, a mistake in the data collecting process, these points could simply be removed or if possible corrected. Collecting more data or reporting the results of separate analyses, with and without the observations in question, are two additional possibilities. Moreover, if predictions are important, it may be possible to partially circumvent the effects of the influential observations by isolating stable regions, or regions where the influence is minimal or unimportant. The emphasis of this thesis is, however, on identifying influential observations rather than how to address them once they are found.

4. Graphical displays

Graphical displays are widely used as diagnostic methods in linear regression and have a long history. We can go back almost a decade and find methods that are still in use today: for instance, the partial residual plot proposed by Ezekiel (1924). However, a significant amount of work has been done to improve existing graphical diagnostic tools and to develop new. A few achievements in the area of regression graphics worth mentioning is the evolution in the 1970's of the use of ordinary residuals in various scatter plots towards the use of different standardized residuals, see e.g. Behnken and Draper (1972) and Andrews and Pregibon (1978). Mosteller and Tukey (1977) proposed the added variable plot, a diagnostic tool that can be used in multiple linear regression. This plot will be described more thoroughly in the next section. For more references concerning regression graphics, see Cook (1998).

Research on graphical tools in nonlinear regression has not been as extensive as in the linear regression case. New thoughts and innovative ideas have been introduced into the area by Cook (1987), where the plot similar to the added variable plot has been proposed for use in nonlinear regression. We will provide a deeper discussion of Cook's results on the plot in Section 4.2.

This chapter will provide an overview of existing graphical methods together with new results obtained in this thesis, which contributes to influence analysis in nonlinear regression. The chapter is divided into two parts. Section 4.1 gives an overview of existing graphical methods for the linear regression model and the added variable plot is described in detail. Section 4.2 is devoted to graphical displays in nonlinear regression. In this section we will describe the added parameter plot, which is one of the main results obtained in this thesis. In Section 4.2.2, the construction and interpretation of the added parameter plot will be illustrated with a numerical example.

4.1 Graphical displays in linear regression

It is well known that various scatter plots of residuals are fundamental for valid interpretation of the results obtained from regression analysis. Residual plots are of utmost importance for validating the model assumptions. We can, for instance, construct a quantile-quantile plot of the standardized residuals to check the assumption of normality, and we can plot the residuals versus the fitted values to examine if the variance of the error terms seem to be homoscedastic.

Diagnostic plots involving the residuals, such as the added variable plot, the partial residuals plot and the augmented partial residuals plot can be used to assess the effect of an additional explanatory variable in the regression model. Moreover, the added variable plot can also be used for finding observations with high influence on the parameter estimates. These plots, and several others, are described by Chatterjee and Hadi (1988) where the partial residuals plot is referred to as the components-plus-residuals plot. In the next section the added variable plot will be described in detail.

4.1.1 The added variable plot

The added variable plot, AVP, is a diagnostic tool that is used in multiple linear regression. It displays the effect of including an extra explanatory variable to the regression, when the other explanatory variables are already taken into account. The plot is helpful for detecting influential observations (see e.g. Belsley *et al.*, 1980) and is referred to as the partial-regression leverage plot by Belsley *et al.* (1980).

For the linear regression model (2.1), the AVP for the explanatory variable X_p is constructed in two steps. Firstly, partition $\boldsymbol{X} : n \times p = (\boldsymbol{X}_1 \vdots \boldsymbol{x}_p)$ and $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T \vdots \beta_p)$. The matrix $\boldsymbol{X}_1$ is given by $\boldsymbol{X}_1 = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1})$ and the vector $\boldsymbol{\beta}_1$ is given by $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_{p-1})^T$.

Using the partitioned $\boldsymbol{X}$ and $\boldsymbol{\beta}$, model (2.1) can be expressed as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{x}_p\beta_p + \boldsymbol{\varepsilon}, \tag{4.1}$$

where $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are vectors of the response variable and the error term, respectively. It is assumed that $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}_n, \sigma^2\mathbf{I}_n)$.

Secondly, define the projection matrix $\boldsymbol{P}_{\boldsymbol{X}_1} = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$. The AVP for explanatory variable X_p is defined to be the scatter plot of

$$\widetilde{\boldsymbol{y}} = (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} \tag{4.2}$$

against

$$\widetilde{\boldsymbol{x}} = (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p, \tag{4.3}$$

along with their estimated regression line.

As a motivation for using the residuals $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ in the AVP (see e.g. Chatterjee and Hadi, 1986), let us pre-multiply (4.1) by $(\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})$ and take expectation. We obtain

$$E\left((\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}\right) = (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{X}_1\boldsymbol{\beta}_1 + (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p\beta_p. \tag{4.4}$$

Observe that $(\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{X}_1\boldsymbol{\beta}_1 = \mathbf{0}_n$, so that (4.4) becomes

$$E\left(\widetilde{\boldsymbol{y}}\right) = \widetilde{\boldsymbol{x}}\alpha, \tag{4.5}$$

where $\widetilde{\boldsymbol{y}} = (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}$ and $\widetilde{\boldsymbol{x}} = (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{x}_p$ are both $n-$vectors and $\alpha = \beta_p$. This suggests that the residuals $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ in the AVP display the effect of introducing the variable X_p to the regression model $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$. Furthermore, a linear trend in the AVP indicates that the explanatory variable X_p should be included in the model.

Belsley *et al.* (1980) refer to the same plot as the partial-regression leverage plot for the parameter estimate $\widehat{\beta}_p$, and not the explanatory variable X_p. This is due to the fact that the slope of the regression line in the plot, i.e. $\widehat{\alpha}$ in model (4.5), is equal to the estimate of β_p in (4.1). From this point of view, we can say that a strong linear trend in the plot indicates that the parameter β_p might significantly differ from zero.

The AVP is also used to detect influential observations on the parameter estimate in question, and Example 4.1 illustrates the utilization of the AVP for this purpose.

Example 4.1. Added variable plot in linear regression

An example of an AVP is given in Figure 4.1 where we use data from Stanley and Miller (1979). A detailed analysis of the data is provided by Cook

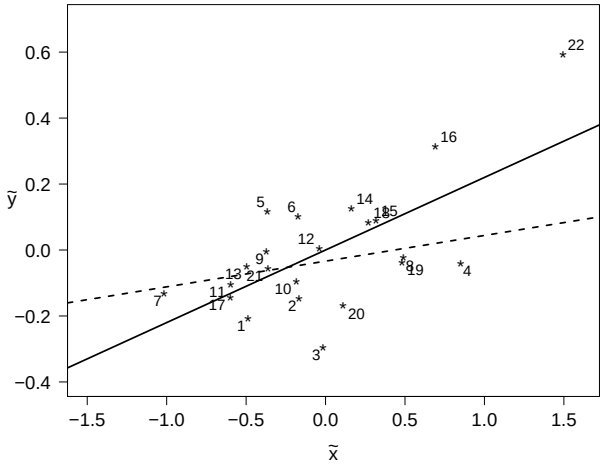


Figure 4.1: Added variable plot for explanatory variable "RGF" using the data presented in Cook and Weisberg (1982) on jet fighters.

and Weisberg (1982). By inspection of the observations in the plot observation numbers 16 and 22 attract our attention since they are separated from the rest. The solid line in Figure 4.1 represents the estimated regression line when all observations are included in the analysis. The dashed line in Figure 4.1 represents the estimated regression line when the observations 16 and 22 are deleted. We can clearly see that the presence of these observations pulls the regression line upwards. Thus, these observations might be influential and further analysis is needed.

Belsley *et al.* (1980) present different features of the plot. Firstly, which is already mentioned above, the slope of the regression line through the origin in the AVP is equal to $\widehat{\beta}_p$ in the regression of $\boldsymbol{y}$ on $\boldsymbol{X}$. Secondly, the residuals that result from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ are equal to the residuals from the regression of $\boldsymbol{y}$ on $\boldsymbol{X}$, and third, the correlation between $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ is equal to the partial correlation between $\boldsymbol{y}$ and $\boldsymbol{x}_p$ in the multiple regression of $\boldsymbol{y}$ on $\boldsymbol{X}$.

In the reminder of this section, we want to highlight another interesting feature of the AVP. We will demonstrate that the AVP is connected to the score test statistic for testing $H_0 : \beta_p = 0$ in model (2.1).

The score test statistic was defined in (2.16) and given by

$$S(\widetilde{\boldsymbol{\beta}}) = \frac{1}{\widetilde{\sigma}^2}\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right), \tag{4.6}$$

where $\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\beta}_1,\ldots,\widetilde{\beta}_{p-1},0\right)$ and $\widetilde{\sigma}^2$ are the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2 under the null hypothesis, i.e. $\widetilde{\sigma}^2 = \frac{1}{n}(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}})$.

Now, let *SSR* denote the sum of squares due to regression in the regression of $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$. In Proposition 4.1.1 we will show in detail that *SSR* is proportional to the numerator in the score test statistic.

Proposition 4.1.1. *The score test statistic, given in (4.6), for testing $H_0 : \beta_p = 0$ in model (2.1) is proportional to SSR in the regression of $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ defined in (4.2) and (4.3), respectively.*

Proof. Observe that *SSR* in the regression of $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ can be written $\widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}}$, where $\boldsymbol{P}_{\widetilde{\boldsymbol{x}}} = \widetilde{\boldsymbol{x}}(\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{x}})^{-1}\widetilde{\boldsymbol{x}}^T$. Using the definition of the residuals in (4.2) we get

$$\widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}} = \boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}. \tag{4.7}$$

Since $\boldsymbol{X}$ is a partitioned matrix, the projection matrix $\boldsymbol{P}_{\boldsymbol{X}}$ can be decomposed into a sum of projection matrices, see the proof of Proposition 2.3.1. In fact,

$$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}_1} + \boldsymbol{P}_{\widetilde{\boldsymbol{x}}}, \tag{4.8}$$

Inserting (4.8) in (4.7) results in

$$\begin{aligned}
\widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}} &= \boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})(\boldsymbol{P}_{\boldsymbol{X}}-\boldsymbol{P}_{\boldsymbol{X}_1})(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} \\
&= \boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})(\boldsymbol{P}_{\boldsymbol{X}}-\boldsymbol{P}_{\boldsymbol{X}_1}-\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_1}+\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} \\
&= \boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})(\boldsymbol{P}_{\boldsymbol{X}}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y} \\
&= \boldsymbol{y}^T(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{P}_{\boldsymbol{X}}(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y},
\end{aligned}$$

since $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_1} = \boldsymbol{P}_{\boldsymbol{X}_1}$ and $\boldsymbol{P}_{\boldsymbol{X}_1}$ is idempotent.

Now, observe that the score test statistic in (4.6) can be written as

$$S(\widetilde{\boldsymbol{\beta}}) = \frac{1}{\widetilde{\sigma}^2}\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T\boldsymbol{P}_{\boldsymbol{X}}\left(\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right),$$

and that $\boldsymbol{y}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ are the residuals under the null hypothesis, i.e. $(\mathbf{I}-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}$.

Thus,

$$\begin{aligned} SSR &= \widetilde{\boldsymbol{y}}^T \boldsymbol{P}_{\widetilde{\boldsymbol{x}}} \widetilde{\boldsymbol{y}} = \boldsymbol{y}^T (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1}) \boldsymbol{P}_{\boldsymbol{X}} (\mathbf{I} - \boldsymbol{P}_{\boldsymbol{X}_1}) \boldsymbol{y} \\ &= \left(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} \right)^T \boldsymbol{P}_{\boldsymbol{X}} \left(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} \right) \propto S(\widetilde{\boldsymbol{\beta}}), \end{aligned}$$

and this completes the proof.

■

Due to Proposition 4.1.1, the AVP can be viewed as a graphical representation of the score test for testing whether the parameter corresponding to the added variable is zero. An apparent linearity in the AVP for the explanatory variable X_p suggests a high value of the *SSR*, and thus, a high value of the score test statistic for the testing $H_0 : \beta_p = 0$ in model (2.1). This property makes the AVP useful for detecting observations that are not only influential on the parameter estimate but also on the score test statistic.

4.2 Graphical displays in nonlinear regression

The ideas behind the AVP described in Section 4.1.1, can be extended to nonlinear regression models. Cook (1987) described a plot similar to the AVP which is referred to as a first-order extension of an AVP. A further discussion about this plot will be given in the next section, where we also derive one of the main results of this thesis, the added parameter plot, APP, along with its features that separates the APP from the first-order extension of an AVP.

Before embarking on the detailed discussions of the APP, the following remark is important. Recall that the AVP is designed to display information available for assessing the significance of a specific explanatory variable for the linear regression model. Moreover, the AVP is used to detect observations that have a substantial influence on the parameter estimate corresponding to the added variable. The relation between the added variable X_p and the parameter estimate $\widehat{\beta}_p$ is possible due to the one-to-one correspondence between the variables and the parameters in the model. However, it is worth noticing that this one-to-one correspondence does not necessarily exist in nonlinear regression models. Thus, for the plots in nonlinear regression similar to the AVP, the fundamental objective is to display information relevant for inference about a selected parameter rather than a selected variable.

4.2.1 The added parameter plot

When testing the significance of parameters in a nonlinear regression model, the score test described in Section 2.3 can be used. From the point of view of explorative data analysis, it would be helpful to be able to graphically display data points that lead to a high value of the score test statistic. Previously we have shown that the AVP, used in multiple linear regression, can be considered as a graphical representation of the score test and can help visualizing observations with high influence on the score test statistic. A similar plot would certainly be desirable for the nonlinear regression model (2.2).

The goal of this section is to construct a new plot, the added parameter plot, APP, and to show that the APP has the property of being a graphical representation of the score test statistic for testing the hypothesis

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^0, \tag{4.9}$$

described in Section 2.3, where $\boldsymbol{\theta}^0 = (\theta_1, \ldots, \theta_{q-1}, 0)^T$. In deriving the plot we will utilize the results on the AVP obtained in linear regression and the results of Cook (1987), where a plot similar to the APP has been proposed.

Cook (1987) describes a plot, referred to as the first-order extension of an AVP. It is created in the same way as the AVP, letting the derivative of the expectation function with respect to the parameters take the role of the matrix $\boldsymbol{X}$ in model (2.1). Let us illustrate the construction of this plot for the parameter estimate of θ_q in the nonlinear regression model (2.2).

Let $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ and consider the linear expansion of $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ around the maximum likelihood estimate, $\widehat{\boldsymbol{\theta}}$,

$$\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) \approx \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}) + \boldsymbol{F}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}),$$

where $\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) : q \times n$ is a matrix of derivatives such that

$$\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) = \Big(\boldsymbol{F}_1(\widehat{\boldsymbol{\theta}}), \ldots, \boldsymbol{F}_n(\widehat{\boldsymbol{\theta}})\Big) = \frac{d}{d\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}},$$

and where the derivative is defined in Appendix A.

Rewriting the model $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, defined in (2.2), by replacing $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ with its linear expansion around $\widehat{\boldsymbol{\theta}}$ and rearranging terms suggests the linear model

$$\boldsymbol{r} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{\delta} + \boldsymbol{\varepsilon}^*, \tag{4.10}$$

where $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}})$ is an $n-$vector of ordinary residuals, $\boldsymbol{\delta} = (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})$ is a $q-$vector and $\boldsymbol{\varepsilon}^*$ represents the error.

Now, partition $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$ so that $\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) = (\widehat{\boldsymbol{F}}_1 \vdots \widehat{\boldsymbol{F}}_2)$ and $\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1^T \vdots \delta_2)$, where δ_2 contains the parameter of interest, θ_q. The partitioned model becomes

$$\boldsymbol{r} = \widehat{\boldsymbol{F}}_1\boldsymbol{\delta}_1 + \widehat{\boldsymbol{F}}_2\delta_2 + \boldsymbol{\varepsilon}^*. \tag{4.11}$$

Define $\boldsymbol{P}_{\widehat{\boldsymbol{F}}_1} = \widehat{\boldsymbol{F}}_1 \left(\widehat{\boldsymbol{F}}_1^T\widehat{\boldsymbol{F}}_1\right)^{-1} \widehat{\boldsymbol{F}}_1^T$, pre-multiply (4.11) by $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})$ which yields

$$(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\widehat{\boldsymbol{F}}_1\boldsymbol{\delta}_1 + (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\widehat{\boldsymbol{F}}_2\delta_2 + (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{\varepsilon}^*, \tag{4.12}$$

and observe that $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\widehat{\boldsymbol{F}}_1\boldsymbol{\delta}_1 = \mathbf{0}$, i.e. the effect of $\widehat{\boldsymbol{F}}_1$ is removed from the model.

In order to arrive at the first-order extension of an AVP, we want to show that $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r}$ in (4.12) is equal to $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}})$, the ordinary residuals when estimating parameters in (2.2) via the maximum likelihood approach.

To show that $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r} = \boldsymbol{r}$ we will start by showing that $\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}})\boldsymbol{r}$, where $\boldsymbol{P}_{\widehat{\boldsymbol{F}}} = \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}))^{-1}\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$.

Observe that $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{r} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}})(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}})) = \mathbf{0}$, since $\widehat{\boldsymbol{\theta}}$ are found by using the normal equations, given by $\boldsymbol{F}(\boldsymbol{\theta})(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})) = \mathbf{0}$. Thus,

$$(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}})\boldsymbol{r} = (\mathbf{I} - \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}))^{-1}\boldsymbol{F}(\widehat{\boldsymbol{\theta}}))\boldsymbol{r} = \boldsymbol{r}. \tag{4.13}$$

Inserting $\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}})\boldsymbol{r}$, obtained in (4.13), into $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r}$ yields

$$(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}})\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}})\boldsymbol{r} = \boldsymbol{r},$$

since $\boldsymbol{P}_{\widehat{\boldsymbol{F}}}\boldsymbol{P}_{\widehat{\boldsymbol{F}}_1} = \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1}$.

We have now shown that $(\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{r}$ in (4.12) is equal to $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}})$ and (4.12) can thus be written as

$$\boldsymbol{r} = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\widehat{\boldsymbol{F}}_2\delta_2 + (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\boldsymbol{\varepsilon}^*. \tag{4.14}$$

Therefore, the first order extension of an AVP is defined as the scatter plot of

$$\boldsymbol{y}^* = \boldsymbol{r} \text{ against } \boldsymbol{x}^* = (\mathbf{I} - \boldsymbol{P}_{\widehat{\boldsymbol{F}}_1})\widehat{\boldsymbol{F}}_2. \tag{4.15}$$

It is suggested that plotting $\boldsymbol{y}^*$ against $\boldsymbol{x}^*$ would display the effects that contribute to the estimate of δ_2, since it depends on the linear constructed model defined in (4.10). If the linear constructed model in (4.10) is a valid approximation of (2.2), the plot of the residuals $\boldsymbol{y}^*$ against $\boldsymbol{x}^*$ is a valuable diagnostic tool for assessing the influence of the observations on the parameter estimate for θ_q, see Cook (1987).

The novel idea behind the APP proposed in this thesis, is to modify the first-order extension of an AVP so that it visualizes the score test. The same approach is used as described in Cook (1987), and $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ is used as $\boldsymbol{X}$, where $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) : q \times n$ is a matrix defined as

$$\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) = \left(\boldsymbol{F}_1(\widetilde{\boldsymbol{\theta}}), \ldots, \boldsymbol{F}_n(\widetilde{\boldsymbol{\theta}})\right) = \frac{d}{d\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}, \tag{4.16}$$

and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_{q-1}, 0)$ is the maximum likelihood estimates of $\boldsymbol{\theta}$ under the null hypothesis (4.9). Hereafter, $\widetilde{\boldsymbol{F}}$ is used to denote $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$.

The derivation of the plot that is a graphical representation of the score test consists of two steps. Firstly, we partition the matrix $\widetilde{\boldsymbol{F}} : q \times n$ as $\widetilde{\boldsymbol{F}}^T = (\widetilde{\boldsymbol{F}}_1 \vdots \widetilde{\boldsymbol{F}}_2)$, where $\widetilde{\boldsymbol{F}}_2$ is the partial derivative of $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ with respect to θ_q evaluated at $\widetilde{\boldsymbol{\theta}}$ and define $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} = \widetilde{\boldsymbol{F}}_1 (\widetilde{\boldsymbol{F}}_1^T \widetilde{\boldsymbol{F}}_1)^{-1} \widetilde{\boldsymbol{F}}_1^T$. Secondly, we construct two sets of residuals using the partition of $\widetilde{\boldsymbol{F}}^T$. The APP is defined as the following.

Definition 4.2.1. *The scatter plot of*

$$\widetilde{\boldsymbol{y}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} \text{ and } \widetilde{\boldsymbol{x}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_2, \tag{4.17}$$

along with the least squares estimated regression line resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$, are defined to be the APP for $\widetilde{\theta}_q$.

As a motivation for using the residuals $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ in the APP, replace $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ in the nonlinear regression model (2.2) with its linear expansion around $\widetilde{\boldsymbol{\theta}}$ and rearrange the terms. This yields the model

$$\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}) = \widetilde{\boldsymbol{F}}^T \left(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\right) + \boldsymbol{\varepsilon}^*. \tag{4.18}$$

Let $\boldsymbol{\delta} = (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})$ and partition the model (4.18) so that it can be written

$$\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}) = \widetilde{\boldsymbol{F}}_1 \boldsymbol{\delta}_1 + \widetilde{\boldsymbol{F}}_2 \delta_2 + \boldsymbol{\varepsilon}^*, \tag{4.19}$$

where $\delta_2 = (\theta_q - \widetilde{\theta}_q)$ displays the parameter of interest. Applying the method described in Section 4.1.1 to the partitioned model (4.19) gives

$$(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{r}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_1 \boldsymbol{\delta}_1 + (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_2 \delta_2 + (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{\varepsilon}, \tag{4.20}$$

where $\widetilde{\boldsymbol{r}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}})$.

Observe that in (4.20), $(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_1 \boldsymbol{\delta}_1 = \mathbf{0}$ and the effect of $\widetilde{\boldsymbol{F}}_1$ is removed from the model. Moreover, observe that, if the constructed linear model (4.18) is a valid approximation of the nonlinear regression model under the null hypothesis (4.9), then $\widetilde{\boldsymbol{r}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}$. Thus, the vector of responses in (4.20) becomes

$$(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{r}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}.$$

It is worth noting that the residuals $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ are created in the same way as the residuals $\boldsymbol{y}^*$ and $\boldsymbol{x}^*$ defined in (4.15). However, an important distinction is that the sets of residuals in (4.17) are created when the matrix of derivatives $\boldsymbol{F}$ is evaluated for the parameter estimates under the null hypothesis, $H_0 : \theta_q = 0$.

The APP is based on similar ideas as the AVP in linear regression, and it has properties similar to the properties of the AVP. From Section 4.1.1 we know that the slope of the regression line in the AVP is equal to the estimate of the parameter corresponding to the added variable, when all other explanatory variables are included in the regression model. The following theorem illustrates that a similar property holds for the APP.

Theorem 4.2.1. *The least squares estimate, $\widehat{\alpha}$, of the slope, α, resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ is equal to the updated parameter estimate, $\widetilde{\theta}_q^{(1)}$, 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 resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ is given by

$$\widehat{\alpha} \quad = \quad (\widetilde{\boldsymbol{x}}^T \widetilde{\boldsymbol{x}})^{-1} \widetilde{\boldsymbol{x}}^T \widetilde{\boldsymbol{y}}.$$

When using the definition of $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{x}}$ in (4.17) we get

$$\begin{aligned} \widehat{\alpha} &= \left(\widetilde{\boldsymbol{F}}_2^T (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_2\right)^{-1} \widetilde{\boldsymbol{F}}_2^T (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} \\ &= \left(\widetilde{\boldsymbol{F}}_2^T \widetilde{\boldsymbol{F}}_2 - \widetilde{\boldsymbol{F}}_2^T \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} \widetilde{\boldsymbol{F}}_2\right)^{-1} \widetilde{\boldsymbol{F}}_2^T (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}. \end{aligned}$$

We will now show that the updated parameter estimate, $\widetilde{\theta}_q^{(1)}$, after one iteration of the Gauss-Newton algorithm when $\widetilde{\boldsymbol{\theta}}$ is used as starting value is equal to $\left(\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2\right)^{-1}\widetilde{\boldsymbol{F}}_2^T(\mathbf{I}-\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}$.

The Gauss Newton algorithm was discussed in Chapter 2. Recall that when using the Gauss-Newton method we rewrite the nonlinear regression model (2.2), utilizing the linear expansion of $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ around the starting value, here $\widetilde{\boldsymbol{\theta}}$. Rearranging terms results in the following linear model

$$\widetilde{\boldsymbol{r}}=\widetilde{\boldsymbol{F}}\boldsymbol{\delta}+\boldsymbol{\varepsilon},$$

where $\boldsymbol{\delta}=\left(\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\right)$ is a $q-$vector.

Minimizing the sum of squared residuals for the constructed linear model yields the least squares estimator for $\widetilde{\boldsymbol{\delta}}=(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T)^{-1}\widetilde{\boldsymbol{F}}^T\widetilde{\boldsymbol{r}}$ and an update of the starting value, $\widetilde{\boldsymbol{\theta}}$, is given by

$$\widetilde{\boldsymbol{\theta}}^{(1)}=\widetilde{\boldsymbol{\theta}}+(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T)^{-1}\widetilde{\boldsymbol{F}}^T\widetilde{\boldsymbol{r}}.$$

Using the partition of $\widetilde{\boldsymbol{F}}^T$ presented in (4.19) we obtain

$$\widetilde{\boldsymbol{\theta}}^{(1)}=\widetilde{\boldsymbol{\theta}}+\begin{pmatrix}\widetilde{\boldsymbol{F}}_1^T\widetilde{\boldsymbol{F}}_1 & \widetilde{\boldsymbol{F}}_1^T\widetilde{\boldsymbol{F}}_2\\ \widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_1 & \widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{\boldsymbol{F}}_1^T\\ \widetilde{\boldsymbol{F}}_2^T\end{pmatrix}\widetilde{\boldsymbol{r}}.$$

In the rest of the proof the rules of inversion of a partitioned matrix are used, which are presented in Chapter 2, Section 2.3.1.

Since we are interested in finding the explicit expression of $\widetilde{\theta}_q^{(1)}$, we can explicitly focus on the last element in $\widetilde{\boldsymbol{\theta}}$ and similarly on the last row of the matrix $(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T)^{-1}$. Moreover, observe that the last element of $\widetilde{\boldsymbol{\theta}}$, $\widetilde{\theta}_q$, equals zero, due to the null hypothesis.

Applying (2.19) to the second row of $(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T)^{-1}$, using $\boldsymbol{A}=\widetilde{\boldsymbol{F}}_1^T\widetilde{\boldsymbol{F}}_1$, $\boldsymbol{b}=\widetilde{\boldsymbol{F}}_1^T\widetilde{\boldsymbol{F}}_2$, $c=\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2$ and $k=\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2$ we obtain

$$\begin{aligned}\widetilde{\theta}_q^{(1)} &= \begin{pmatrix}-\frac{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_1(\widetilde{\boldsymbol{F}}_1^T\widetilde{\boldsymbol{F}}_1)^{-1}}{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2} & \frac{1}{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2}\end{pmatrix}\begin{pmatrix}\widetilde{\boldsymbol{F}}_1^T\\ \widetilde{\boldsymbol{F}}_2^T\end{pmatrix}\widetilde{\boldsymbol{r}}\\ &= \left(-\frac{\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}}{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2}+\frac{\widetilde{\boldsymbol{F}}_2^T}{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2-\widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2}\right)\widetilde{\boldsymbol{r}}.\end{aligned}$$

Continuing with writing $\widetilde{\boldsymbol{r}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}$ yields

$$\begin{aligned}\widetilde{\theta}_q^{(1)} &= \frac{\widetilde{\boldsymbol{F}}_2^T}{\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2 - \widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2}(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} \\ &= \left(\widetilde{\boldsymbol{F}}_2^T\widetilde{\boldsymbol{F}}_2 - \widetilde{\boldsymbol{F}}_2^T\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\widetilde{\boldsymbol{F}}_2\right)^{-1}\widetilde{\boldsymbol{F}}_2^T(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y},\end{aligned}$$

and this completes the proof. ■

The results obtained in Theorem 4.2.1 give us an idea of how the APP can be used for assessing the effect of introducing the parameter θ_q to the model. A steep slope of the regression line in the APP indicates that the updated estimate of $\boldsymbol{\theta}$ differs much from the starting values, $\widetilde{\boldsymbol{\theta}}$. This in turn indicates that the parameter θ_q might have a contributing effect on the regression model. If there is no linear trend in the APP, the slope is close to zero, and hence there is no change in the updated estimate of $\boldsymbol{\theta}$ when $\widetilde{\boldsymbol{\theta}}$ is used as a starting value.

The next lemma provides the results that will help when proving Theorem 4.2.2 and Theorem 4.2.3.

Lemma 4.2.1. *The projection matrix $\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}$ is equal to $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}$.*

Proof. We want to prove that $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} + \boldsymbol{P}_{\widetilde{\boldsymbol{x}}}$.

Observe that $\widetilde{\boldsymbol{F}}^T = (\widetilde{\boldsymbol{F}}_1 \vdots \widetilde{\boldsymbol{F}}_2)$ is a partitioned matrix and that $\widetilde{\boldsymbol{x}} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\widetilde{\boldsymbol{F}}_2$ by Definition 4.2.1. We can now use the proof for Proposition 2.3.1, letting $\widetilde{\boldsymbol{F}}^T = \boldsymbol{X}$, $\widetilde{\boldsymbol{F}}_1 = \boldsymbol{X}_1$ and $\widetilde{\boldsymbol{x}} = \boldsymbol{x}^*$. From the proof it follows that $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} + \boldsymbol{P}_{\widetilde{\boldsymbol{x}}}$.

■

In Section 4.1.1 it has been outlined, that the residuals obtained from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ are equal to the residuals obtained from regressing $\boldsymbol{y}$ on $\boldsymbol{X}$, i.e. all the variables contained in $\boldsymbol{X}$. The APP has a similar feature, as stated in the next theorem.

Theorem 4.2.2. *The residual vector $\boldsymbol{u} = \widetilde{\boldsymbol{y}} - \widehat{\alpha}\widetilde{\boldsymbol{x}}$, resulting from estimating the regression line in the added parameter plot is equal to $(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}})\boldsymbol{y}$, i.e. the residuals when $\boldsymbol{y}$ is regressed on all the columns of $\widetilde{\boldsymbol{F}}^T$.*

Proof. The residuals obtained from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ are equal to $(\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{x}}})\widetilde{\boldsymbol{y}}$.

Since $\widetilde{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}$, the residuals can be written

$$\boldsymbol{u} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{x}}})\widetilde{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{x}}})(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}.$$

From the result in Lemma 4.2.1, we know that $\boldsymbol{P}_{\widetilde{\boldsymbol{x}}} = \boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}$. Using this property we get

$$\boldsymbol{u} = (\mathbf{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{x}}})\widetilde{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}} + \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}})\boldsymbol{y},$$

since $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} = \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}$. This completes the proof.

■

In the next theorem it will be shown that, similar to the AVP, the APP can be considered to be a graphical representation of the score test.

Theorem 4.2.3. *When regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$, the resulting $SSR = \widetilde{\boldsymbol{y}}^T \boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}}$ is proportional to the score test statistic for testing the hypothesis*

$$H_0 : \theta_q = 0,$$

where the score test statistic is defined in (2.27) and given by

$$S(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\widetilde{\sigma}^2}\widetilde{\mathbf{r}}^T\widetilde{\boldsymbol{F}}^T\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T\right)^{-1}\widetilde{\boldsymbol{F}}\widetilde{\mathbf{r}}, \tag{4.21}$$

and $\widetilde{\mathbf{r}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{F}}$ is given by (4.16).

Proof. We want to show that $\widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}} = \widetilde{\sigma}^2 S(\widetilde{\boldsymbol{\theta}})$, where $S(\widetilde{\boldsymbol{\theta}})$ is defined in (4.21).

Observe that

$$SSR = \widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}} = \left((\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}\right)^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\left((\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}\right).$$

Using the result in Lemma 4.2.1

$$\begin{aligned} SSR &= \boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\left(\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\right)(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} \\ &= \boldsymbol{y}^T\left(\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} + \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\right)(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}, \end{aligned}$$

and using the fact that $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1} = \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}$ we get

$$SSR = \boldsymbol{y}^T\left(\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\right)(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y} = \boldsymbol{y}^T\left(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1}\right)\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}.$$

Since $(\boldsymbol{I} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}_1})\boldsymbol{y}$ are the residuals under the null hypothesis, $SSR = \widetilde{\boldsymbol{y}}^T\boldsymbol{P}_{\widetilde{\boldsymbol{x}}}\widetilde{\boldsymbol{y}}$ can

be written as

$$\begin{aligned} SSR &= \left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}})\right)^T \widetilde{\boldsymbol{F}}^T \left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T\right)^{-1} \widetilde{\boldsymbol{F}} \left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}})\right) = \widetilde{\boldsymbol{r}}^T \widetilde{\boldsymbol{F}}^T \left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T\right)^{-1} \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{r}} \\ &= \widetilde{\sigma}^2 S(\widetilde{\boldsymbol{\theta}}). \end{aligned}$$

Thus, $SSR = \widetilde{\boldsymbol{y}}^T \boldsymbol{P}_{\widetilde{\boldsymbol{x}}} \widetilde{\boldsymbol{y}} = \widetilde{\sigma}^2 S(\widetilde{\boldsymbol{\theta}})$. The proof is complete. ■

From the proof of Theorem 4.2.3 it follows that the score test statistic is proportional to *SSR* resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$. Thus, plotting $\widetilde{\boldsymbol{y}}$ against $\widetilde{\boldsymbol{x}}$ and observing the data points' location can contribute to the knowledge of which observations strongly influence the value of the score test statistic.

In the next section a numerical example will be used to illustrate the construction of an APP. The example also contains a discussion of how the individual observations contribute to the value of the score test statistic.

4.2.2 Numerical example

Data from Bates and Watts (1988), given in Table 4.1 will be used to fit the Michaelis-Menten model (2.4) given by

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

where y is the initial velocity of the enzymatic reaction and x is substrate concentration. The parameter θ_1 is the maximum initial velocity that is theoretically attained when the enzyme has been saturated by an infinite concentration of substrate. The second parameter, θ_2, is the Michaelis parameter which is numerically equal to the concentration of substrate for "half-maximum" initial velocity.

In the example presented in Bates 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 so that $x_2 = 1$ if the enzyme is treated and

$$f(x,\theta) = \frac{(\theta_1 + \theta_3 x_2) x_1}{\theta_2 + x_1}. \tag{4.22}$$

The Michaelis-Menten model is thus modified, now including θ_3 to account for the effect of puromycin on the asymptotic velocity, θ_1.

y	x_1	x_2	y	x_1	x_2
76.00	0.02	1.00	159.00	0.22	1.00
67.00	0.02	0.00	131.00	0.22	0.00
47.00	0.02	1.00	152.00	0.22	1.00
51.00	0.02	0.00	124.00	0.22	0.00
97.00	0.06	1.00	191.00	0.56	1.00
84.00	0.06	0.00	144.00	0.56	0.00
107.00	0.06	1.00	201.00	0.56	1.00
86.00	0.06	0.00	158.00	0.56	0.00
123.00	0.11	1.00	207.00	1.10	1.00
98.00	0.11	0.00	160.00	1.10	0.00
139.00	0.11	1.00	200.00	1.10	1.00
115.00	0.11	0.00			

Table 4.1: Data from Bates and Watts (1988), used to fit the Michaelis-Menten model with expectation functions (4.22) and (4.23).

A second modification of the model entails including θ_4, a parameter for the potential effect of puromycin on θ_2. The expectation function is now written as

$$f(x,\theta) = \frac{(\theta_1 + \theta_3 x_2)x_1}{(\theta_2 + \theta_4 x_2) + x_1}. \tag{4.23}$$

The score test can be used to test if the model should include different half-velocity parameters depending on whether the enzyme is treated or not. In this case the hypotheses are

$$\begin{aligned} H_0 &: \theta_4 = 0, \\ H_A &: \theta_4 \neq 0. \end{aligned} \tag{4.24}$$

To conduct the score test, first fit the model with $f(x,\theta)$ defined in (4.22) to the data in order to retrieve the estimates under the null hypothesis. This yields $\widetilde{\boldsymbol{\theta}} = (166.60,\, 0.06,\, 42.03,\, 0)^T$ and $\widetilde{\sigma}^2 = 97.43$. The score test statistic for testing (4.24) is given by

$$S(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\widetilde{\sigma}^2}\widetilde{\mathbf{r}}^T\widetilde{\boldsymbol{F}}^T\left(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^T\right)^{-1}\widetilde{\boldsymbol{F}}\widetilde{\mathbf{r}},$$

and using the estimates under the null hypothesis, we get that $S(\widetilde{\boldsymbol{\theta}}) = 1.67$. The $p-$value for this test is 0.20 and the null hypothesis that the half-velocity

parameter is unchanged by the puromycin treatment cannot be rejected.

To visualize the score test, an APP for $\widetilde{\theta}_4$ is constructed. First let the columns in $\widetilde{\boldsymbol{F}}^T$ act as independent variables, where

$$\widetilde{\boldsymbol{F}}^T = \left(\boldsymbol{F}(\widetilde{\theta}_1), \boldsymbol{F}(\widetilde{\theta}_2), \boldsymbol{F}(\widetilde{\theta}_3), \boldsymbol{F}(\widetilde{\theta}_4)\right) = \left(\frac{d\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})}{d\theta_1}\bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}, \ldots, \frac{d\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})}{d\theta_4}\bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right).$$

The first three columns in $\widetilde{\boldsymbol{F}}^T$ form the matrix $\widetilde{\boldsymbol{F}}_1$ and the last column forms the vector $\widetilde{\boldsymbol{F}}_2$. Next, $\widetilde{\boldsymbol{y}}$ is constructed as the residuals in the regression when $\boldsymbol{y}$ is regressed on $\boldsymbol{F}_1$ and $\widetilde{\boldsymbol{x}}$ is constructed as the residuals when $\widetilde{\boldsymbol{F}}_2$ is regressed on $\widetilde{\boldsymbol{F}}_1$. Now, plotting $\widetilde{\boldsymbol{y}}$ against $\widetilde{\boldsymbol{x}}$ and estimating the regression line resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$ yields the APP for $\widetilde{\theta}_4$, which is given in Figure 4.2.

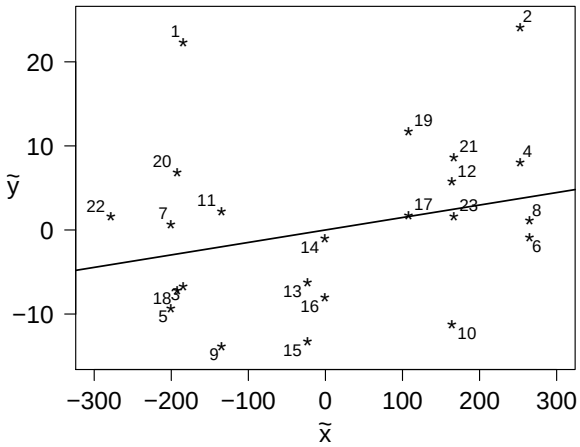


Figure 4.2: The added parameter plot for $\widetilde{\theta}_4$ consisting of the scatter plot of $\widetilde{\boldsymbol{y}}$, the residuals resulting from regressing $\boldsymbol{y}$ on $\widetilde{\boldsymbol{F}}_1$, against $\widetilde{\boldsymbol{x}}$, the residuals resulting from regressing $\widetilde{\boldsymbol{F}}_2$ on $\widetilde{\boldsymbol{F}}_1$, and the estimated regression line with slope $\widehat{\alpha}$.

The estimate of the slope of the regression line in Figure 4.2 is equal to $\widehat{\alpha} = 0.02$. Moreover, the updated estimate of $\widetilde{\boldsymbol{\theta}}$ using a single iteration of the Gauss-Newton method is

$$\widetilde{\boldsymbol{\theta}}^{(1)} = (160.90, 0.05, 51.30, 0.02)^T, \tag{4.25}$$

Thus, we see that $\widehat{\alpha} = 0.02 = \widetilde{\theta}_4^{(1)}$, and this illustrates Theorem 4.2.1.

According to Theorem 4.2.3, the value of the score test statistic is equal to the ratio of *SSR*, resulting from regressing $\widetilde{\boldsymbol{y}}$ on $\widetilde{\boldsymbol{x}}$, and $\widetilde{\sigma}^2$. Here, $SSR = 162.28$, $\widetilde{\sigma}^2 = 97.43$ and $S(\widetilde{\boldsymbol{\theta}}) = \frac{162.28}{97.43} = 1.67$, which corresponds to the value of the score test statistic obtained above.

The APP in Figure 4.2 can be studied more thoroughly, searching for observations with substantial influence on the score test statistic. Firstly, we note that there is no strong linear trend in the scatter of the observations in the plot, which is consistent with not rejecting the null hypothesis. Secondly, we note that the 1st and 2nd observations are separated from the rest of the data points and could be influential observations. A deeper analysis of the data yields that when the 1st observation is removed from the calculations, *SSR* is increased together with the value of the score test statistic. Moreover, the slope of the regression line also changes. The new values of *SSR*, the score test statistic and estimated slope of the regression line are 329.80, 4.32 and 0.02, respectively. Thus, observation 1 is influencing the score test statistic, decreasing its value. The $p-$value, corresponding to a value of 4.32 for the score test statistic, is 0.04. In fact, when the 1st observation is excluded from the data the null hypothesis would be rejected on a 5 percent significance level. When observation 2 is removed from the calculations the new values of *SSR*, the score test statistic and slope are 29.91, 0.41 and 0.01 respectively. Thus, the presence of observation 2 is increasing the score test statistic.

5. Assessment of influence on parameter estimates

Assessment of the influence of the observations on the parameter estimate is an important part of influence analysis, and there are many challenging issues to consider. For instance,

- parameter estimates can be highly influenced by single observations.
- multiple observations can also have a large influence on the parameter estimates:
  - several observations can simultaneously influence parameter estimates, hence their joint influence should be assessed.
  - due to hidden, general, dependence among observations, there can be observations that influence parameter estimates only when one or several observations are removed from the data set.

In Section 5.1, we will present another important result of this thesis, an influence measure that is used to assess the influence of a single observation on the parameter estimates in a nonlinear regression model. The section will also contain a brief description of the corresponding influence measure in linear regression.

Section 5.2 is devoted to assessing the influence of multiple observations on the parameter estimates. The section is divided into two main parts, where one part concerns the simultaneous influence of several observations on the parameter estimates. This type of influence will be referred to as joint influence. The other part treats the influence that the kth observation has on the parameter estimates *after* another observation, say observation i, has been deleted. The type of influence that the kth observation has on the parameter estimates after the deletion of the ith observation is called conditional influence. Joint and conditional influence will be discussed for both linear and nonlinear regression models, and two new diagnostic measures are worked out for use in nonlinear regression.

5.1 Assessment of influence of a single observation

The 1970's and the 1980's were the decades when a significant amount of research on influence analysis in linear regression was conducted. Statisticians were not content with using the regression model and simply accepting the data at hand as given. They were now seeking to investigate the data quality. The pioneering work of Cook (1977) in this area resulted in the diagnostic measure referred to as Cook's distance. Cook's distance is given in (3.1) and is used to measure the influence of a single observation on the parameter estimates in linear regression models. Earlier attempts to protect against influential or outlying observations came through the concept of robustness and robust regression. The use of robust regression was motivated by the fact that the least squares estimator was not robust, but rather sensitive, against outlying observations. This means that a single observation, being extremely influential, could cause the least squares estimation to produce an incorrect result. Interested readers are referred to Huber (1972) and Hampel (1974).

Belsley *et al.* (1980) came out with a book on regression diagnostics and identification of influential observations in particular. New diagnostic tools, e.g. DFFIT and DFBETA, were proposed as summaries of parameter changes by deletion of an observation. These two diagnostic tools, together with Cook's distance, are the most commonly used diagnostic measures for conducting influence analysis in linear regression, and they are implemented in most statistical software packages. Later, Cook and Weisberg (1982) and Chatterjee and Hadi (1988) discussed the role of different residuals and diagnostic measures in influence analysis.

Deleting observations and studying the change in the parameter estimates due to deletion is a popular approach to influence analysis. This approach is known as case-deletion. Cook's distance, DFFIT and DFBETA are all examples of measures where this strategy is adopted. Cook and Weisberg (1982) extended the ideas of case deletion and proposed a diagnostic measure, similar to Cook's distance, for assessing the influence of observations on parameter estimates in the nonlinear regression model. Later, Ross (1987) studied the geometry of case deletion in nonlinear regression models and discussed the adequacy of using diagnostic measures based on case-deletion in nonlinear regression.

In 1986, Cook proposed a new approach for influence analysis, which is referred to as local influence. In this approach weights are introduced to the model, which does not necessarily need to be a linear regression model, by attaching them to the observations. One novelty of the local influence approach

was that the weights were allowed to vary between zero and one, rather than being zero, as in the case deletion approach. The article by Cook (1986) contained new diagnostic measures for conducting influence analysis about the parameters in the linear regression model. However, the proposal to use local influence approach stimulated the research of influence analysis in nonlinear regression since St. Laurent and Cook (1993) discussed the relation between leverage and local influence in nonlinear regression.

Influence analysis in nonlinear regression is not widely explored and the results obtained in this thesis will make a certain contribution to this research area. One of the main results is the new influence measure for assessing the influence of single observations on the parameter estimates in a nonlinear regression model. This measure is denoted $DIM_{\widehat{\boldsymbol{\theta}},k}$. The abbreviation stands for *Differentiation approach & Influence Measure*, since for deriving the influence measure the differentiation approach is used, described by Belsley *et al.* (1980), Chatterjee and Hadi (1988) and Cook and Weisberg (1982). The differentiation approach is used in linear regression to construct the influence measure $EIC_{\widehat{\boldsymbol{\beta}},k}$, where the *EIC* stands for *Empirical Influence Curve*. Originally this measure was derived from the influence curve, a theoretical concept introduced by Hampel (1974). In Section 5.1.1 we will derive $EIC_{\widehat{\boldsymbol{\beta}},k}$ from the differentiation approach to demonstrate the idea behind the construction of it. Then, borrowing ideas from linear regression, we derive the nonlinear influence measure, $DIM_{\widehat{\boldsymbol{\theta}},k}$ in Section 5.1.2.

5.1.1 The influence measure *EIC* in linear regression, derived via the differentiation approach

An approach for assessing the influence of an observation on the parameter estimates in the linear regression model (2.1), described in Belsley *et al.* (1980), Chatterjee and Hadi (1988) and Cook and Weisberg (1982), is called the differentiation approach. The resulting influence measure is denoted $EIC_{\widehat{\boldsymbol{\beta}},k}$ and next we will demonstrate the derivation of the $EIC_{\widehat{\boldsymbol{\beta}},k}$ in linear regression. The idea behind it will later be extended to nonlinear regression models as well.

Let us consider the following perturbed linear regression model

$$\boldsymbol{y}_\omega = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\omega, \tag{5.1}$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2\mathbf{W}^{-1}(\omega_k))$, ω_k is the weight such that $0 < \omega_k \leq 1$ and the weight matrix $\mathbf{W}(\omega_k)$ is the diagonal matrix

$$\mathbf{W}(\omega_k) = diag\,(1,\ldots,\omega_k,\ldots,1)\,.$$

The weighted least squares estimator for $\boldsymbol{\beta}$ in (5.1) is given by

$$\widehat{\boldsymbol{\beta}}(\omega_k) = \left(\boldsymbol{X}^T \mathbf{W}(\omega_k)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \mathbf{W}(\omega_k)\boldsymbol{y}.$$

Definition 5.1.1. *The influence measure for assessing the influence of the kth observation on $\widehat{\boldsymbol{\beta}}$, denoted $EIC_{\widehat{\boldsymbol{\beta}},k}$, is defined as the derivative of $\widehat{\boldsymbol{\beta}}(\omega_k)$ with respect to ω_k evaluated at $\omega_k = 1$:*

$$EIC_{\widehat{\boldsymbol{\beta}},k} = \frac{d}{d\omega_k}\widehat{\boldsymbol{\beta}}(\omega_k)\bigg|_{\omega_k=1}.$$

The influence measure $EIC_{\widehat{\boldsymbol{\beta}},k}$ in Definition 5.1.1 describes how the calculated estimate changes in the area near $\omega_k = 1$, i.e as the kth observation is given full weight. For instance, a value of $EIC_{\widehat{\boldsymbol{\beta}},k}$ close to zero corresponds to no change, or a minor change, in the estimate if the kth observation is included in the calculations of $\widehat{\boldsymbol{\beta}}$. In this case, the kth observation is not an influential observation. On the other hand, a value of $EIC_{\widehat{\boldsymbol{\beta}},k}$ being substantially different from zero means that the inclusion of the kth observation in the calculations of $\widehat{\boldsymbol{\beta}}$ substantially changes the result of the estimation. A diagnostic measure similar to $EIC_{\widehat{\boldsymbol{\beta}},k}$ is given by taking the derivative of $\widehat{\boldsymbol{\beta}}(\omega_k)$, with respect to ω_k, evaluated as $\omega_k \to 0$. It is denoted $EIC_{\widehat{\boldsymbol{\beta}},(k)}$ by Chatterjee and Hadi (1988) and it measures how the estimate of $\boldsymbol{\beta}$ changes when the kth observation is deleted from the data.

In the following theorem we will derive the expression of $EIC_{\widehat{\boldsymbol{\beta}},k}$ in Definition 5.1.1, using the differentiation approach.

Theorem 5.1.1. *Let $EIC_{\widehat{\boldsymbol{\beta}},k}$ be given in Definition 5.1.1. Then,*

$$EIC_{\widehat{\boldsymbol{\beta}},k} = r_k \boldsymbol{x}_k^T (\boldsymbol{X}^T\boldsymbol{X})^{-1}, \tag{5.2}$$

where r_k is the kth component of the vector of residuals $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{x}_k^T$ is the kth row of the matrix $\boldsymbol{X}$.

Proof. Let us evaluate the derivative in Definition 5.1.1 using the product rule, (see Appendix A for details of how to apply the derivative)

$$\begin{aligned}
\frac{d}{d\omega_k}\widehat{\boldsymbol{\beta}}(\omega_k) &= \frac{d\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{W}\boldsymbol{y}}{d\omega_k} \\
&= \frac{d\mathbf{W}}{d\omega_k}(\boldsymbol{y}\otimes\boldsymbol{X})\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1} \\
&\quad -\frac{d\mathbf{W}}{d\omega_k}(\boldsymbol{X}\otimes\boldsymbol{X})\left(\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\otimes\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\right)\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{y}\otimes\boldsymbol{I}_p\right),
\end{aligned}$$

where $\otimes$ is the Kronecker product (see Kollo and von Rosen, 2010), defined as follows:

Let $\boldsymbol{A} = (a_{ij})$ be a $p \times q$ matrix and $\boldsymbol{B} = (b_{ij})$ be an $r \times s$ matrix. Then the $pr \times qs$ matrix $\boldsymbol{A} \otimes \boldsymbol{B}$ is a Kronecker product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ if

$$\boldsymbol{A} \otimes \boldsymbol{B} = [a_{ij}\boldsymbol{B}], \;\; i = 1, \ldots, p; \;\; j = 1, \ldots, q,$$

where

$$a_{ij}\boldsymbol{B} = \begin{pmatrix} a_{ij}b_{11} & \ldots & a_{ij}b_{1s} \\ \vdots & . & \vdots \\ a_{ij}b_{r1} & \ldots & a_{ij}b_{rs} \end{pmatrix}.$$

Due to linearity of $\mathbf{W}$ the following expression is obtained:

$$\frac{d\mathbf{W}}{d\omega_k} = \boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the kth column of the identity matrix of size n. Evaluating the expression above at $\omega_k = 1$ we get

$$\begin{aligned} \frac{d}{d\omega_k}\widehat{\boldsymbol{\beta}}(\omega_k)\bigg|_{\omega_k=1} &= (\boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T)\Big[(\boldsymbol{y} \otimes \boldsymbol{X})\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \\ &\qquad - (\boldsymbol{X} \otimes \boldsymbol{X})\left(\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \otimes \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)(\boldsymbol{X}^T\boldsymbol{y} \otimes \boldsymbol{I}_p)\Big] \\ &= (\boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T)\Big[(\boldsymbol{y} \otimes \boldsymbol{X})\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}} \otimes \boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\Big] \\ &= (\boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T)\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \otimes \boldsymbol{X}\right)\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \\ &= r_k\boldsymbol{x}_k^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}. \end{aligned}$$

Thus, the final expression for $EIC_{\widehat{\boldsymbol{\beta}},k}$, derived using the differentiation approach, is

$$EIC_{\widehat{\boldsymbol{\beta}},k} = r_k\boldsymbol{x}_k^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},$$

and the proof is complete. ■

In the next section, we extend the ideas of using the differentiation approach for measuring the influence of an observation on the parameter estimate for nonlinear regression models. The influence measure denoted $DIM_{\widehat{\boldsymbol{\theta}},k}$, is derived for assessing the influence of the kth observation on the parameter estimates in the nonlinear regression model (2.2).

5.1.2 The influence measure *DIM*, for use in nonlinear regression

In this section two new influence measures for the parameter estimates in the nonlinear regression model (2.2) will be derived: the $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$. The first diagnostic measure, $DIM_{\widehat{\boldsymbol{\theta}},k}$, is used to assess the influence of a single observation on all parameter estimates in the model, simultaneously. It is constructed when all parameters are estimated from a perturbed model, presented in (5.3) later on, and it is referred to as the joint-parameter influence measure.

The $DIM_{\widehat{\theta}_j,k}$, on the other hand, is used to assess the influence of a single observation on the jth parameter estimate in the model. When constructing $DIM_{\widehat{\theta}_j,k}$, only the jth parameter is estimated from the perturbed model, later defined in (5.3): the other parameters are estimated from an unperturbed model and regarded to be known. The $DIM_{\widehat{\theta}_j,k}$ is referred to as the marginal-parameter influence measure.

We will now start with the definition of $DIM_{\widehat{\boldsymbol{\theta}},k}$. Consider the following perturbed nonlinear model

$$\boldsymbol{y}_\omega = \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_\omega, \tag{5.3}$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2\mathbf{W}^{-1}(\omega_k))$, ω_k is the weight such that $0 < \omega_k \leq 1$ and the weight matrix $\mathbf{W}(\omega_k) = diag(1,\ldots,\omega_k,\ldots,1)$.

Definition 5.1.2. *The influence measure for assessing the influence of the kth observation on $\widehat{\boldsymbol{\theta}}$ is defined as the following derivative*

$$DIM_{\widehat{\boldsymbol{\theta}},k} = \left.\frac{d}{d\omega_k}\widehat{\boldsymbol{\theta}}(\omega_k)\right|_{\omega_k=1}, \tag{5.4}$$

where $\widehat{\boldsymbol{\theta}}(\omega_k)$ is the weighted least squares estimate of $\boldsymbol{\theta}$ in the perturbed model (5.3).

Observe that, in Definition 5.1.2, if $\omega_k \to 1$, then $\widehat{\boldsymbol{\theta}}(\omega_k) \to \widehat{\boldsymbol{\theta}}$, the unweighted least squares estimate.

To calculate the $DIM_{\widehat{\boldsymbol{\theta}},k}$ in (5.4) we need an estimator for $\boldsymbol{\theta}$ in the perturbed model (5.3). Using the method of weighted least squares, which is equivalent to the maximum likelihood approach, we have to find $\widehat{\boldsymbol{\theta}}$ that minimizes the following

$$Q(\omega_k) = (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))^T \mathbf{W}(\omega_k) (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})).$$

Differentiating $Q(\omega_k)$ with respect to $\boldsymbol{\theta}$ one gets the following normal equations

$$\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\mathbf{W}(\omega_k)\left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\right)=\mathbf{0}, \tag{5.5}$$

where the derivative, $\frac{d\boldsymbol{f}}{d\boldsymbol{\theta}}$, is defined in Appendix A.

The normal equations in (5.5) are solved for $\boldsymbol{\theta}$ using iterative methods such as the Gauss-Newton method. The obtained least squares estimate of $\boldsymbol{\theta}$ is a function of ω_k.

In the next theorem, the explicit expression of $DIM_{\widehat{\boldsymbol{\theta}},k}$, defined in (5.4), will be presented.

Theorem 5.1.2. *Let $DIM_{\widehat{\boldsymbol{\theta}},k}$ be given in Definition 5.1.2. Then*

$$DIM_{\widehat{\boldsymbol{\theta}},k}=r_k\boldsymbol{F}_k^T(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})-\boldsymbol{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\mathbf{I}_q\right)\right)^{-1},$$

provided that the inverse exists, where

$$\boldsymbol{r}=(r_k)=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}), \tag{5.6}$$

$$\boldsymbol{F}(\widehat{\boldsymbol{\theta}})=(\boldsymbol{F}_1(\widehat{\boldsymbol{\theta}}),\ldots,\boldsymbol{F}_n(\widehat{\boldsymbol{\theta}}))=\left.\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\quad q\times n. \tag{5.7}$$

and

$$\mathbf{G}(\widehat{\boldsymbol{\theta}})=\left(\frac{d}{d\boldsymbol{\theta}}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}=\frac{d\boldsymbol{F}(\widehat{\boldsymbol{\theta}})}{d\widehat{\boldsymbol{\theta}}},\quad q\times nq. \tag{5.8}$$

Proof. Consider inserting the weighted least squares estimate, $\widehat{\boldsymbol{\theta}}(\omega_k)$, in the normal equations (5.5)

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}(\omega_k)\left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}(\omega_k))\right)=\mathbf{0}, \tag{5.9}$$

and letting $\boldsymbol{F}=\boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\omega_k))$, $\mathbf{W}=\mathbf{W}(\omega_k)$ and $\widehat{\boldsymbol{e}}=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X})=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}(\omega_k))$. The influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ can be obtained by differentiation of (5.9) with respect to ω_k on both sides, i.e.

$$\frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}}=\mathbf{0}. \tag{5.10}$$

The product rule, defined in Appendix A, shows that (5.10) equals

$$\frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} = \frac{d\boldsymbol{F}}{d\omega_k}(\mathbf{W}\widehat{\boldsymbol{e}}\otimes\mathbf{I}_q)+\frac{d\mathbf{W}}{d\omega_k}\left(\widehat{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right)+\frac{d\widehat{\boldsymbol{e}}}{d\omega_k}\mathbf{W}\boldsymbol{F}^T. \tag{5.11}$$

Now

$$\frac{d\widehat{\boldsymbol{e}}}{d\omega_k} = -\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\omega_k},$$

and

$$\frac{d\mathbf{W}}{d\omega_k} = \boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the kth column of the identity matrix of size n. Moreover, applying the chain rule, see Appendix A, to (5.11) implies that (5.10) is identical to

$$\frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\omega_k)}(\mathbf{W}\widehat{\boldsymbol{e}}\otimes\mathbf{I}_q)+\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T-\frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}\boldsymbol{F}^T=\mathbf{0},$$

which after rearrangement of terms yields

$$\left(\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T\right) = \frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}\boldsymbol{F}^T-\frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\omega_k)}(\mathbf{W}\widehat{\boldsymbol{e}}\otimes\mathbf{I}_q)\right). \tag{5.12}$$

Evaluating the derivatives in (5.12) at $\omega_k = 1$ together with (5.6)-(5.8) and Definition 5.1.2 implies

$$\mathbf{d}_k^T\boldsymbol{r}\otimes\mathbf{d}_k^T\widehat{\boldsymbol{F}}^T(\widehat{\boldsymbol{\theta}}) \;=\; \frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\bigg|_{\omega_k=1}\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})-\mathbf{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r}\otimes\mathbf{I}_q)\right).$$

Thus,

$$r_k\boldsymbol{F}_k^T(\widehat{\boldsymbol{\theta}}) \;=\; DIM_{\widehat{\boldsymbol{\theta}},k}\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})-\boldsymbol{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r}\otimes\mathbf{I}_q)\right),$$

and

$$DIM_{\widehat{\boldsymbol{\theta}},k} = r_k\boldsymbol{F}_k^T(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})-\boldsymbol{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r}\otimes\mathbf{I}_q)\right)^{-1}.$$

This completes the proof.

■

The $DIM_{\widehat{\boldsymbol{\theta}},k}$ derived in Theorem 5.1.2 measures the influence of the kth observation on all the parameter estimates in model (2.2) simultaneously. Therefore, $DIM_{\widehat{\boldsymbol{\theta}},k}$ is regarded to be a joint-parameter influence measure. However, it can

be useful to measure the influence of the kth observation on a particular parameter estimate of the model. In order to assess the influence of the k observation on the jth parameter estimate, $\widehat{\theta}_j$, a marginal-parameter influence measure will be defined and its explicit expression will be derived.

Consider the perturbed model (5.3). Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j)$ be a vector of parameter estimates, where $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\theta}_1, \ldots, \widehat{\theta}_{j-1}, \widehat{\theta}_{j+1}, \ldots, \widehat{\theta}_q)^T$, are the maximum likelihood estimates in the unperturbed model (2.2) and $\widehat{\theta}_j$ is estimated from the perturbed model (5.3), with parameter estimates $\widehat{\boldsymbol{\theta}}_1$ inserted and regarded as known.

Definition 5.1.3. *The marginal influence measure for assessing the influence of the kth observation on the parameter estimate $\widehat{\theta}_j$ is defined as the following derivative*

$$DIM_{\widehat{\theta}_j,k} = \left. \frac{d}{d\omega_k} \widehat{\theta}_j(\omega_k) \right|_{\omega_k=1}, \tag{5.13}$$

where $\widehat{\theta}_j(\omega_k)$ is the weighted least squares estimate of θ_j, given $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\theta}_1, \ldots, \widehat{\theta}_{j-1}, \widehat{\theta}_{j+1}, \ldots, \widehat{\theta}_q)^T$.

Observe that, in Definition 5.1.3, if $\omega_k \to 1$, then $\widehat{\theta}_j(\omega_k) \to \widehat{\theta}_j$, the unweighted least squares estimate.

The weighted least squares criterion and the normal equation for the case when a single parameter is estimated from the perturbed model are given by

$$Q(\omega_k) = \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right)^T \mathbf{W}(\omega_k) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right),$$

and

$$\frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)}{d\theta_j} \mathbf{W}(\omega_k) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j(\omega_k))\right) = 0. \tag{5.14}$$

In the next theorem, an explicit expression of the marginal-parameter influence diagnostic $DIM_{\widehat{\theta}_j,k}$ defined in (5.13) will be provided.

Theorem 5.1.3. *Let $DIM_{\widehat{\theta}_j,k}$ be given in Definition 5.1.3. Then*

$$DIM_{\widehat{\theta}_j,k} = r_k F_k(\widehat{\theta}_j) \left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r}\right)^{-1},$$

provided that the inverse exists, where $\boldsymbol{r}=(r_k)=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\widehat{\theta}_j)$,

$$\boldsymbol{F}(\widehat{\theta}_j)=\left(\boldsymbol{F}_1(\widehat{\theta}_j),\ldots,\boldsymbol{F}_n(\widehat{\theta}_j)\right)=\left.\frac{d\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\theta_j)}{d\theta_j}\right|_{\theta_j=\widehat{\theta}_j},\quad 1\times n, \tag{5.15}$$

$$\mathbf{G}(\widehat{\theta}_j)=\left.\frac{d\boldsymbol{F}(\theta_j)}{d\theta_j}\right|_{\theta_j=\widehat{\theta}_j}=\left.\frac{d^2\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\theta_j)}{d\theta_j^2}\right|_{\theta_j=\widehat{\theta}_j},\quad 1\times n. \tag{5.16}$$

Proof. The proof is very similar to the proof of Theorem 5.1.2. Consider inserting the weighted least squares estimate of θ_j in the normal equation (5.14)

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\theta_j)}{d\theta_j}\right|_{\theta_j=\widehat{\theta}_j(\omega_k)}\mathbf{W}(\omega_k)\left(\mathbf{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\widehat{\theta}_j(\omega_k))\right)=0, \tag{5.17}$$

and letting $\boldsymbol{F}=\boldsymbol{F}(\widehat{\theta}_j(\omega_k))$, $\mathbf{W}=\mathbf{W}(\omega_k)$ and $\widehat{\boldsymbol{e}}=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X})=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\widehat{\theta}_j(\omega_k))$. For the kth observation, the $DIM_{\widehat{\theta}_{j,k}}$ can be obtained by differentiating (5.17) on both sides with respect to ω_k, i.e.

$$\frac{d}{d\omega_k}\mathbf{F}\mathbf{W}\widehat{\mathbf{e}}=0. \tag{5.18}$$

Now, the product rule, defined in Appendix A, is used to calculate the derivative in (5.18):

$$\frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} \;=\; \frac{d\boldsymbol{F}}{d\omega_k}\mathbf{W}\widehat{\boldsymbol{e}}+\frac{d\mathbf{W}}{d\omega_k}\left(\widehat{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right)+\frac{d\widehat{\boldsymbol{e}}}{d\omega_k}\mathbf{W}\boldsymbol{F}^T. \tag{5.19}$$

Moreover, applying the chain rule, see Appendix A, to (5.19) gives

$$\frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k}\frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\omega_k)}\mathbf{W}\widehat{\mathbf{e}}+\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T-\frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\omega_k)}\mathbf{W}\boldsymbol{F}^T=0,$$

and then, rearranging terms yields

$$\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T \;=\; \frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\omega_k)}\mathbf{W}\boldsymbol{F}^T-\frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\omega_k)}\left(\mathbf{W}\widehat{\boldsymbol{e}}\right)\right).$$

As previously mentioned, evaluating the derivative at $\omega_k = 1$ gives $\widehat{\theta}_j = \widehat{\theta}_j(\omega_k = 1)$, and denoting $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j(\omega_k = 1))$ implies

$$\begin{aligned} \boldsymbol{d}_k^T \boldsymbol{r} \otimes \boldsymbol{d}_k^T \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}_j) &= \left. \frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k} \right|_{\omega_k=1} \left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r} \right), \\ r_k F_k(\widehat{\theta}_j) &= DIM_{\widehat{\theta}_j,k} \left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r} \right). \end{aligned}$$

Thus, the final expression for $DIM_{\widehat{\theta}_j,k}$ is

$$DIM_{\widehat{\theta}_j,k} = r_k F_k(\widehat{\theta}_j) \left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r} \right)^{-1}.$$

The proof is complete. ■

5.1.3 A note on $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$

When deriving the influence measures and studying the single observations' influence on the parameter estimates we observe some interesting aspects of influence analysis in nonlinear regression.

A benefit of using the differentiation approach, where we compute derivatives of various quantities with respect to ω_k and evaluate the derivatives at $\omega_k = 1$, is that no additional iterations for computing the parameter estimates are needed. As was discussed in Section 5.1.1, an alternative way of using the differentiation approach is to evaluate the same derivatives as $\omega_k \to 0$. If this approach were to be used instead, the explicit expressions of $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$ would be functions of the parameter estimates with weights attached. As an example, consider the following derivative

$$\left. \frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\omega_k))}{d\omega_k} \right|_{\omega_k \to 0} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\omega_k)),$$

where $\omega_k \to 0$. This means that we would need to compute a parameter estimate for each k and additional iterations are needed. On the contrary, with the new proposed method in this thesis

$$\left. \frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\omega_k))}{d\omega_k} \right|_{\omega_k = 1} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}),$$

which is the derivative of the expectation function from the unperturbed model (2.2) and hence, no additional iterations are needed.

We can further make a comparison between the proposed measure, $DIM_{\widehat{\boldsymbol{\theta}},k}$, and the nonlinear version of Cook's distance, discussed in Chapter 3, and given by

$$\frac{\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_{(k)}\right)^T \boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\left(\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_{(k)}\right)}{q\widehat{\sigma}^2}, \qquad k=1,\ldots,n,$$

where q is the number of parameters in the model, $\widehat{\boldsymbol{\theta}}_{(k)}$ is the estimate of $\boldsymbol{\theta}$ when the kth observation is excluded from the calculations and $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$ is defined in (5.7). The nonlinear version of Cook's distance is based on case-deletion. A consequence of this is that re-estimation of the parameters is needed for every observation we are interested in. Thus, the nonlinear version of Cook's distance demands additional iterations when estimating the parameters, which is avoided using our measure $DIM_{\widehat{\boldsymbol{\theta}},k}$.

The joint-parameter influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ is a $1\times q-$vector. The computation of $DIM_{\widehat{\boldsymbol{\theta}},k}$ will result in $q-$values, one value for each parameter estimate, which will indicate whether the kth observation is influential on the specific parameter estimate. However, it is worth noting that the influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ is affected by the dependencies among the estimated parameters due to the fact that they are estimated jointly. For instance, for the modified Gompertz growth curve model (2.6) described in Chapter 2, the parameter μ_m is defined as the slope of the tangent line at the point of inflection and the parameter λ is defined as the intercept of the tangent line. There is a dependence between λ and μ_m, when estimated values are used. If an observation has a strong influencing effect on $\widehat{\mu}_m$, this effect will be partly transmitted to the value of the influence measure regarding $\widehat{\lambda}$ as well. In the case of using the Michaelis-Menten regression model (2.4) for enzyme kinetics, the parameter θ_1 is defined as the maximum initial velocity, which is attained when the enzyme has been saturated by an infinite concentration of substrate. The parameter θ_2 is defined as the value of substrate corresponding to half the maximum velocity. For inference we remark that observations that are highly influential on $\widehat{\theta}_1$ will probably show impact on $\widehat{\theta}_2$ as well.

If one wants to be "certain" of what effect the observations have on a particular parameter estimate one should use $DIM_{\widehat{\theta}_j,k}$. This measure is constructed when only the jth parameter is estimated from the perturbed model and the other parameters in the model are assumed to be known, i.e. their estimates from the

unperturbed model are regarded to be the true parameter values. The fact that we are able to assess influence of observations on a specific parameter estimate is clearly beneficial over the already proposed approaches to influence analysis in nonlinear regression. The nonlinear version of Cook's distance can be used when assessing the influence of an observation on the whole vector of parameter estimates is of interest. The extension of the local influence approach, from linear regression models to nonlinear regression models, is considered by St. Laurent and Cook (1993). Their results concern the assessment of influence of the observations on the fitted values, and there is no suggestion of how the local influence approach can be extended to assessing the influence on a specific parameter estimate.

It is worth noting that the values of $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$ can be positive or negative. A positive value of the influence measure for a given observation means that the presence of this observation increases the value of the corresponding parameter estimate. In a similar way, a negative value for a given observation means that the presence of that observation decreases the parameter estimate.

Nonlinear regression models can differ greatly. It is important to know the shape of the expectation function used in the real-life problem. Some observations might be more influential if they are located in the area, which is important for the estimation of the parameters. These areas are of course different for different nonlinear regression models. Both the Michaelis-Menten model and the modified Gompertz growth curve model contain a parameter representing an asymptotic value. Observations located in the area near this asymptote are expected to be more important in the estimation process. These observations will thus be more influential on the parameter estimates than other observations not located in this area. This fact certainly provides more information to the analysis but it does not necessarily mean that an observation with a high absolute value of the diagnostic measure is influential in the sense that the observation is spurious. This aspect of influence analysis in nonlinear regression will be discussed further in the numerical example in Section 5.1.4.

Inspecting the explicit expression of the influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ given in Theorem 5.1.2, we observe that it is a function of the kth residual and that it is related to the kth diagonal element of the tangent plane leverage matrix discussed in Chapter 3. To see this, first consider the result from the inverse binomial theorem (see e.g. Kollo and von Rosen, 2010).

$$(\boldsymbol{A}+\boldsymbol{UBV})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{UB}\left(\boldsymbol{B}+\boldsymbol{BVA}^{-1}\boldsymbol{UB}\right)^{-1}\boldsymbol{BVA}^{-1},$$

provided that $\boldsymbol{A}$ and $\boldsymbol{B}+\boldsymbol{BVA}^{-1}\boldsymbol{UB}$ are nonsingular.

If we want to invert $(\boldsymbol{A}-\boldsymbol{UBV})$ we have that

$$(\boldsymbol{A}-\boldsymbol{UBV})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{UB}\left(\boldsymbol{BVA}^{-1}\boldsymbol{UB}-\boldsymbol{B}\right)^{-1}\boldsymbol{BVA}^{-1}. \tag{5.20}$$

In the expression of $DIM_{\widehat{\boldsymbol{\theta}},k}$ let $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})=\boldsymbol{F}$ and $\boldsymbol{G}(\widehat{\boldsymbol{\theta}})=\boldsymbol{G}$. If applying (5.20) then

$$\begin{aligned} DIM_{\widehat{\boldsymbol{\theta}},k} &= r_k\boldsymbol{F}_k^T\left(\boldsymbol{FF}^T-\boldsymbol{G}(\boldsymbol{r}\otimes\mathbf{I}_q)\right)^{-1} \\ &= r_k\boldsymbol{F}_k^T\left(\boldsymbol{FF}^T\right)^{-1} \\ &\quad -r_k\boldsymbol{F}_k^T\Big((\boldsymbol{FF}^T)^{-1}\boldsymbol{G}\left((\boldsymbol{r}\otimes\mathbf{I})(\boldsymbol{FF}^T)^{-1}\boldsymbol{G}-\mathbf{I}\right)^{-1} \\ &\quad \times(\boldsymbol{r}\otimes\mathbf{I})(\boldsymbol{FF}^T)^{-1}\Big). \end{aligned} \tag{5.21}$$

Now, the kth diagonal element of the tangent plane leverage matrix, see (3.7), is given by

$$\boldsymbol{F}_k^T\left(\boldsymbol{FF}^T\right)^{-1}\boldsymbol{F}_k,$$

and we see that the first term of the expression on the right hand side of (5.21) is a function of the kth residual and related to the kth diagonal element of the tangent plane leverage matrix. A high value of the residual and/or a high leverage value for the kth observation might result in an influential observation. Thus, the investigation of the residuals and the leverages of the observations can contribute to a deeper understanding of why an observation is influential or not.

In a similar manner, we observe that the explicit expression of the marginal influence measure, $DIM_{\widehat{\theta}_j,k}$, given in Theorem 5.1.3, is a function of

$$\boldsymbol{F}_k^T(\widehat{\theta}_j)\left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j)\right)^{-1}\boldsymbol{F}_k(\widehat{\theta}_j). \tag{5.22}$$

The quantity in (5.22) is the kth diagonal element of the (projection) matrix

$$\boldsymbol{F}^T(\widehat{\theta}_j)\left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j)\right)^{-1}\boldsymbol{F}(\widehat{\theta}_j). \tag{5.23}$$

Since the derivative of $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ with respect to θ_j is considered exclusively in (5.23), we denote the quantity in (5.22) as the marginal leverage. When studying the influence of single observations on the specific parameter estimate $\widehat{\theta}_j$, an investigation of the residuals and the marginal leverages forms a good basis.

The discussion about leverages have led us to suggest a modification of the influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$. Consider post-multiplying $\boldsymbol{F}_k$ to $DIM_{\widehat{\boldsymbol{\theta}},k}$

$$DIM^*_{\widehat{\boldsymbol{\theta}},k} = r_k \boldsymbol{F}_k^T \left(\boldsymbol{F}\boldsymbol{F}^T - \boldsymbol{G}(\boldsymbol{r}\otimes \mathbf{I}_q)\right)^{-1} \boldsymbol{F}_k.$$

In applying (5.21)

$$\begin{aligned} DIM^*_{\widehat{\boldsymbol{\theta}},k} &= r_k \boldsymbol{F}_k^T \left(\boldsymbol{F}\boldsymbol{F}^T - \boldsymbol{G}(\boldsymbol{r}\otimes \mathbf{I}_q)\right)^{-1} \\ &= r_k \boldsymbol{F}_k^T \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1} \boldsymbol{F}_k \\ &\quad - r_k \boldsymbol{F}_k^T \Big((\boldsymbol{F}\boldsymbol{F}^T)^{-1} \boldsymbol{G} \left((\boldsymbol{r}\otimes \mathbf{I})(\boldsymbol{F}\boldsymbol{F}^T)^{-1} \boldsymbol{G} - \mathbf{I} \right)^{-1} \\ &\quad \times (\boldsymbol{r}\otimes \mathbf{I})(\boldsymbol{F}\boldsymbol{F}^T)^{-1} \Big) \boldsymbol{F}_k, \end{aligned}$$

and we observe that the first term of $DIM^*_{\widehat{\boldsymbol{\theta}},k}$ consists of the kth residual and the leverage of the kth observation. Moreover, $DIM^*_{\widehat{\boldsymbol{\theta}},k}$ is a scalar and this measure can be regarded as a collective influence measure for all parameters in the model.

Next, the components of the diagnostic measures $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$ will be illustrated.

Example 5.1 Illustration of the structures of $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$
Consider the Michaelis-Menten model,

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i,\ i = 1,2,3,$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ and $\boldsymbol{\varepsilon} \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I})$. We observe that in a practical situation more than three observations are needed in order to estimate the model. However, since this is merely a demonstration of the explicit expressions of $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$ we use three observations in order to simplify the expressions.

Let us assume that we want to assess the influence of the 2nd observation on the vector of parameter estimates, $\widehat{\boldsymbol{\theta}}$. The diagnostic measure to use is

$$DIM_{\widehat{\boldsymbol{\theta}},2} = r_2 F_2^T(\widehat{\boldsymbol{\theta}}) \left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}(\widehat{\boldsymbol{\theta}}) (\boldsymbol{r}\otimes \mathbf{I}_2) \right)^{-1}, \tag{5.24}$$

where

$$\begin{aligned}
\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) &= \frac{d\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}})}{d\widehat{\boldsymbol{\theta}}} \\
&= \begin{pmatrix} \frac{df_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} & \frac{df_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} & \frac{df_3(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} \\ \frac{df_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} & \frac{df_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} & \frac{df_3(\widehat{\boldsymbol{\theta}}(\omega_2))}{d\widehat{\theta}_2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{x_1}{\widehat{\theta}_2+x_1} & \frac{x_2}{\widehat{\theta}_2+x_2} & \frac{x_3}{\widehat{\theta}_2+x_3} \\ \frac{-\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^2} & \frac{-\widehat{\theta}_1 x_2}{(\widehat{\theta}_2+x_2)^2} & \frac{-\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^2} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{G}(\widehat{\boldsymbol{\theta}}) &= \frac{d\boldsymbol{F}(\widehat{\boldsymbol{\theta}})}{d\widehat{\boldsymbol{\theta}}} \\
&= \begin{pmatrix} \frac{d^2 f_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1^2} & \frac{d^2 f_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1 d\widehat{\theta}_2} & \frac{d^2 f_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1^2} & \frac{d^2 f_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1 d\widehat{\theta}_2} & \frac{d^2 f_3(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1^2} & \frac{d^2 f_3(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1 d\widehat{\theta}_2} \\ \frac{d^2 f_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2 d\widehat{\theta}_1} & \frac{d^2 f_1(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2^2} & \frac{d^2 f_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2 d\widehat{\theta}_1} & \frac{d^2 f_2(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2^2} & \frac{d^2 f_3(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2 d\widehat{\theta}_1} & \frac{d^2 f_3(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{d}{d\widehat{\theta}_1}\frac{x_1}{\widehat{\theta}_2+x_1} & \frac{d}{d\widehat{\theta}_1}\frac{-\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^2} & \cdots & \frac{d}{d\widehat{\theta}_1}\frac{-\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^2} \\ \frac{d}{d\widehat{\theta}_2}\frac{x_1}{\widehat{\theta}_2+x_1} & \frac{d}{d\widehat{\theta}_2}\frac{-\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^2} & \cdots & \frac{d}{d\widehat{\theta}_2}\frac{-\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^2} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{-x_1}{(\widehat{\theta}_2+x_1)^2} & 0 & \frac{-x_2}{(\widehat{\theta}_2+x_2)^2} & 0 & \frac{-x_3}{(\widehat{\theta}_2+x_3)^2} \\ \frac{-x_1}{(\widehat{\theta}_2+x_1)^2} & \frac{2\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^3} & \frac{-x_2}{(\widehat{\theta}_2+x_2)^2} & \frac{2\widehat{\theta}_1 x_2}{(\widehat{\theta}_2+x_2)^3} & \frac{-x_3}{(\widehat{\theta}_2+x_3)^2} & \frac{2\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^3} \end{pmatrix}.
\end{aligned}$$

Now, we investigate the matrix

$$\left(\mathbf{F}(\widehat{\boldsymbol{\theta}})\mathbf{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\mathbf{I}_2\right)\right),$$

whose inverse should be used for the calculation of $DIM_{\widehat{\boldsymbol{\theta}},2}$ in (5.24). First,

$$\begin{aligned}
\mathbf{F}(\widehat{\boldsymbol{\theta}})\mathbf{F}^T(\widehat{\boldsymbol{\theta}}) &= \begin{pmatrix} \sum_{i=1}^{3}\left(\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1}\right)^2 & \sum_{i=1}^{3}\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1}\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} \\ \sum_{i=1}^{3}\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1}\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} & \sum_{i=1}^{3}\left(\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2}\right)^2 \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{3}\left(\frac{x_i}{\widehat{\theta}_2+x_i}\right)^2 & -\sum_{i=1}^{3}\frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2+x_i)^3} \\ -\sum_{i=1}^{3}\frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2+x_i)^3} & \sum_{i=1}^{3}\left(\frac{-\widehat{\theta}_1 x_i}{(\widehat{\theta}_2+x_i)^2}\right)^2 \end{pmatrix}.
\end{aligned}$$

Secondly, we calculate

$$\mathbf{G}(\widehat{\boldsymbol{\theta}})\,(\mathbf{r}\otimes\mathbf{I}_2) = \begin{pmatrix} \sum_{i=1}^{3}\frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}})r_i}{d\widehat{\theta}_1^2} & \sum_{i=1}^{3}\frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}})r_i}{d\widehat{\theta}_1 d\widehat{\theta}_2} \\ \sum_{i=1}^{3}\frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}})r_i}{d\widehat{\theta}_2 d\theta_1} & \sum_{i=1}^{3}\frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}})r_i}{d\widehat{\theta}_2^2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \sum_{i=1}^{3}\frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} \\ \sum_{i=1}^{3}\frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} & \sum_{i=1}^{3}\frac{2\widehat{\theta}_1 x_i r_i}{(\widehat{\theta}_2+x_i)^3} \end{pmatrix}.$$

Thus,

$$DIM_{\widehat{\boldsymbol{\theta}},2} = \begin{pmatrix} \frac{x_2 r_2}{\widehat{\theta}_2+x_2} & \frac{-\widehat{\theta}_1 x_2 r_2}{(\widehat{\theta}_2+x_2)^2} \end{pmatrix}$$

$$\times \left(\begin{pmatrix} \sum_{i=1}^{3}\left(\frac{x_i}{\widehat{\theta}_2+x_i}\right)^2 & -\sum_{i=1}^{3}\frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2+x_i)^3} \\ -\sum_{i=1}^{3}\frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2+x_i)^3} & \sum_{i=1}^{3}\left(\frac{-\widehat{\theta}_1 x_i}{(\widehat{\theta}_2+x_i)^2}\right)^2 \end{pmatrix} - \begin{pmatrix} 0 & \sum_{i=1}^{3}\frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} \\ \sum_{i=1}^{3}\frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} & \sum_{i=1}^{3}\frac{2\widehat{\theta}_1 x_i r_i}{(\widehat{\theta}_2+x_i)^3} \end{pmatrix} \right)^{-1}.$$

Next, the marginal influence of the 2nd observation on the parameter estimate $\widehat{\theta}_2$ will be studied, and the diagnostic measure to use is

$$DIM_{\widehat{\theta}_2,2} = r_2 F_2(\widehat{\theta}_2)\left(\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2) - \boldsymbol{G}(\widehat{\theta}_2)\boldsymbol{r}\right)^{-1}. \tag{5.25}$$

In (5.25) the vector of the first derivatives of $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$ is given by

$$\boldsymbol{F}(\widehat{\theta}_2) = \begin{pmatrix} \frac{-\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^2} & \frac{-\widehat{\theta}_1 x_2}{(\widehat{\theta}_2+x_2)^2} & \frac{-\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^2} \end{pmatrix},$$

and the vector of second derivatives is the following

$$\mathbf{G}(\widehat{\theta}_2) = \begin{pmatrix} \frac{2\widehat{\theta}_1 x_1}{(\widehat{\theta}_2+x_1)^3} & \frac{2\widehat{\theta}_1 x_2}{(\widehat{\theta}_2+x_2)^3} & \frac{2\widehat{\theta}_1 x_3}{(\widehat{\theta}_2+x_3)^3} \end{pmatrix}.$$

The matrix $\left(\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2) - \boldsymbol{G}(\widehat{\theta}_2)\boldsymbol{r}\right)$ is of interest, because its inverse will be used for the calculation of $DIM_{\widehat{\theta}_2,2}$. Observe that,

$$\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2) = \sum_{i=1}^{3}\left(\frac{df_i(\widehat{\theta}_2)}{d\widehat{\theta}_2}\right)^2 = \sum_{i=1}^{3}\left(\frac{-\widehat{\theta}_1 x_i}{\left(\widehat{\theta}_2+x_i\right)^2}\right)^2 = \sum_{i=1}^{3}\frac{(\widehat{\theta}_1 x_i)^2}{(\widehat{\theta}_2+x_i)^4},$$

and

$$\begin{aligned}\boldsymbol{G}(\widehat{\theta}_2)\boldsymbol{r} &= \sum_{i=1}^{3}\frac{df_i(\widehat{\theta}_2)}{d\widehat{\theta}_2}r_i = \sum_{i=1}^{3}\left(\frac{2\widehat{\theta}_1 x_i}{(\widehat{\theta}_2+x_i)^3}\left(y_i-\frac{\widehat{\theta}_1 x_i}{\widehat{\theta}_2+x_i}\right)\right)\\ &= \sum_{i=1}^{3}\left(\frac{2\widehat{\theta}_1 x_i y_i}{(\widehat{\theta}_2+x_i)^3}-\frac{2(\widehat{\theta}_1 x_i)^2}{(\widehat{\theta}_2+x_i)^4}\right).\end{aligned}$$

Since $r_2F_2 = \frac{-\widehat{\theta}_1 x_2 r_2}{(\widehat{\theta}_2+x_2)^2}$, we have that

$$DIM_{\widehat{\theta}_2,2} = \frac{-\widehat{\theta}_1 x_2 r_2}{(\widehat{\theta}_2+x_2)^2}\left(\sum_{i=1}^{3}\frac{(\widehat{\theta}_1 x_i)^2}{(\widehat{\theta}_2+x_i)^4}-\sum_{i=1}^{3}\left(\frac{2\widehat{\theta}_1 x_i y_i}{(\widehat{\theta}_2+x_i)^3}-\frac{2(\widehat{\theta}_1 x_i)^2}{(\widehat{\theta}_2+x_i)^4}\right)\right)^{-1}.$$

As an illustration of how $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$ can be used in a practical situation, we will present two numerical examples using simulated data, in Section 5.1.4 and 5.1.5.

5.1.4 Numerical example: Influence analysis using $DIM_{\widehat{\boldsymbol{\theta}},k}$

The Michaelis-Menten model is used for studying enzyme kinetics, and it relates the initial velocity, y, of an enzymatic reaction to the substrate concentration, x, through the equation

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

In this numerical example we fit the Michaelis-Menten model using simulated data. The data is simulated using a similar approach as Atkins and Nimmo (1975). First, a set of 'perfect', i.e. error free, data is formed with $\theta_1 = 1$ and $\theta_2 = 1$. The values of substrate concentration are

$$x = (0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75).$$

We make replicates of each $x-$value and a data set of 49 observations is created. Then the $y-$values are simulated from the perfect set using normally distributed errors with a mean of zero and a standard deviation equal to 0.1. Thus, we let

$$y_{ij} = \frac{\theta_1 x_i}{\theta_2+x_i} + \varepsilon_{ij}, \tag{5.26}$$

where the $\varepsilon \overset{i.i.d.}{\sim} N(0, 0.1^2)$ for $i, j = 1, \ldots, 7$.

In order to verify whether our suggested influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ can detect an influential observation we contaminate the 40th observation by increasing its $y-$value. The reason for choosing observation 40 is because this observation is located in the area, which is expected to be important for the estimation of θ_1. By increasing its $y-$value we expect that this observation will be declared the most influential when using $DIM_{\widehat{\boldsymbol{\theta}},k}$ as influence measure. Moreover, there is a strong dependence between $\widehat{\theta}_1$ and $\widehat{\theta}_2$ and we expect that this observation will stand out as influential on $\widehat{\theta}_2$ as well. The data is presented and plotted in Table 5.1 and Figure 5.1, respectively.

y_{ij}	x_i	y_{ij}	x_i	y_{ij}	x_i
0.27	0.25	0.46	0.75	0.37	1.25
0.08	0.25	0.46	0.75	0.75	1.25
0.37	0.25	0.59	0.75	0.42	1.50
0.18	0.25	0.63	0.75	0.60	1.50
0.21	0.25	0.56	1.00	0.74	1.50
0.25	0.25	0.48	1.00	0.59	1.50
0.36	0.25	0.36	1.00	1.20	1.50
0.37	0.50	0.48	1.00	0.62	1.50
0.52	0.50	0.58	1.00	0.68	1.50
0.12	0.50	0.44	1.00	0.49	1.75
0.42	0.50	0.43	1.00	0.64	1.75
0.32	0.50	0.58	1.25	0.68	1.75
0.43	0.50	0.64	1.25	0.81	1.75
0.29	0.50	0.53	1.25	0.32	1.75
0.35	0.75	0.57	1.25	0.75	1.75
0.44	0.75	0.70	1.25	0.64	1.75
0.47	0.75				

Table 5.1: Simulated data according to the model given in (5.26)

The calculated $DIM_{\widehat{\boldsymbol{\theta}},k}$, $k = 1,\ldots,49$, are presented graphically in two figures, where the values of the influence measure corresponding to $\widehat{\theta}_1$ are given in Figure 5.2a and the values of the influence measure corresponding to $\widehat{\theta}_2$ are given in Figure 5.2b.

In Figure 5.2a we can see that $DIM_{\widehat{\boldsymbol{\theta}},k}$ identifies the 40th observation as the most influential observation on $\widehat{\theta}_1$ and the value of the influence measure is $DIM_{\widehat{\theta}_1,40} = 0.09$. The second largest value of the influence measure, in magnitude, corresponds to observation 47, where $DIM_{\widehat{\theta}_1,47} = -0.08$. 75 percent of

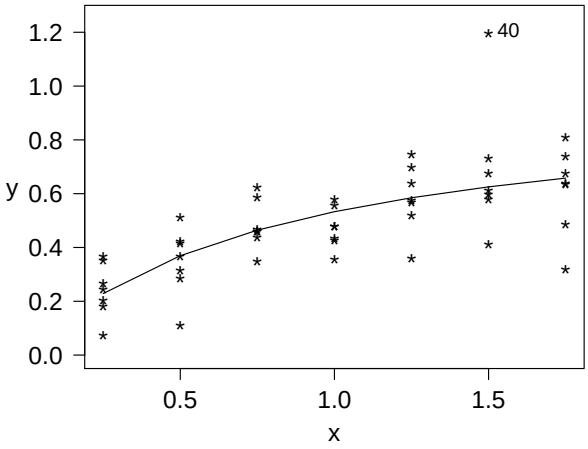
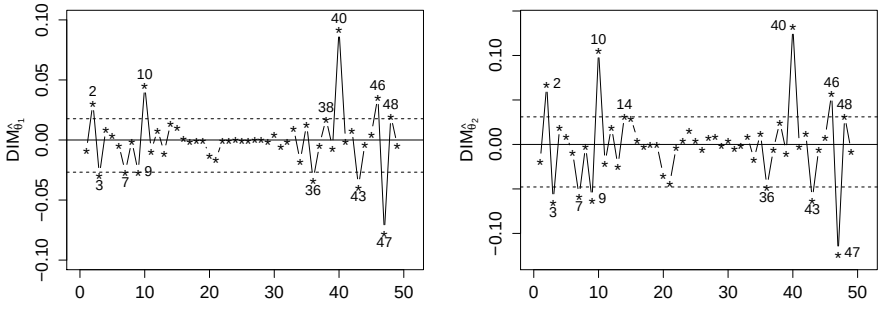


Figure 5.1: Plot of the data given in Table 5.1, where $y =$ initial velocity and $x =$ substrate concentration, together with the estimated curve. Observation 40 is contaminated.



(a) $DIM_{\widehat{\theta}_1,k}$ **(b)** $DIM_{\widehat{\theta}_2,k}$

Figure 5.2: The joint-parameter influence measure $DIM_{\widehat{\boldsymbol{\theta}},k}$ defined in (5.4), for each observation in Table 5.1. Observations within the dashed lines represents 75 percent of the data. Observe that $DIM_{\widehat{\boldsymbol{\theta}},k} = (DIM_{\widehat{\theta}_1,k}, DIM_{\widehat{\theta}_2,k})$.

the observations lies within the dashed lines. Observations 40 and 47 are well separated from these 75 percent.

The 40th observation is the most influential observation on $\widehat{\theta}_2$ as well, as can be seen in Figure 5.2b where $DIM_{\widehat{\theta}_2,40} = 0.13$. The second largest absolute value corresponds to observation 47 where $DIM_{\widehat{\theta}_2,47} = -0.12$. 75 percent of the data lies within the dashed lines and observations 40 and 47 are isolated from the most common 75 percent of the data. Moreover, the 10th observation has a large value of the influence measure and is separated from the rest of the data.

We can present the results of the calculation of $DIM_{\widehat{\boldsymbol{\theta}},k}$ in one figure by plotting $DIM_{\widehat{\theta}_1,k}$ against $DIM_{\widehat{\theta}_2,k}$, see Figure 5.3. From this figure it is clear that the 40th and 47th observations are the most influential on both $\widehat{\theta}_1$ and $\widehat{\theta}_2$.

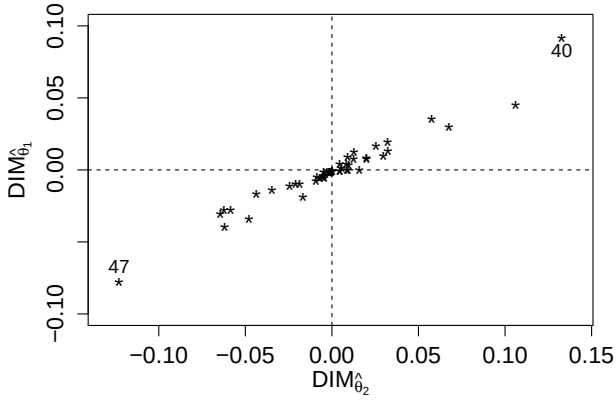


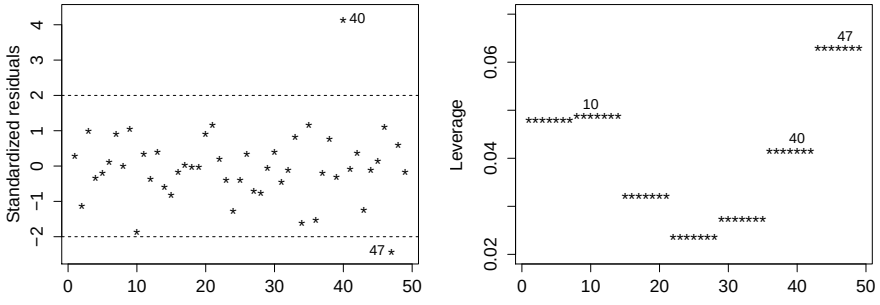
Figure 5.3: The influence measures $DIM_{\widehat{\theta}_1,k}$ and $DIM_{\widehat{\theta}_2,k}$ calculated for each observation in Table 5.1.

Summarizing, $DIM_{\widehat{\boldsymbol{\theta}},k}$ successfully identifies the 40th observation as an influential observation. Observation 47, and in some sense observation 10, have high influence on the parameter estimates $\widehat{\boldsymbol{\theta}}$ but are not contaminated. The reason for their large values of $DIM_{\widehat{\boldsymbol{\theta}},k}$ will be commented below.

In Section 5.1.3 we saw that $DIM_{\widehat{\boldsymbol{\theta}},k}$ is a function of the residuals and is closely related to the tangent plane leverages. Therefore, to get a deeper understanding of the influence of the observations, we investigate the standardized residuals and the leverages, presented in Figure 5.4a and Figure 5.4b, respectively. From the figures it can be see that the 40th observation has a large standardized residual, which is outside the normal range of -2 to 2, and that the leverage is

medium high. Thus, observation 40 has both a relatively large value of leverage and a large standardized residual, which explains the high value of $DIM_{\widehat{\boldsymbol{\theta}},k}$.

Inspecting the figures, we see that observation 47 has a large standardized residual, in magnitude. This is expected by chance and this observation is not spurious, since it is generated from the correct model. Moreover, the 47th observation belongs to the group of observations with the largest values of leverage and it is located in the area, which is expected to be important for estimation of θ_1. A large standardized residual and the location of the observation are the reasons for the large value, in magnitude, of $DIM_{\widehat{\boldsymbol{\theta}},47}$.



(a) Standardized residuals **(b)** Leverage for each observation

Figure 5.4: Standardized residuals and leverages computed using the data in Table 5.1.

From Figure 5.4a and Figure 5.4b we see that the standardized residual corresponding to the 10th observation is within the normal range and that the leverage for observation 10 is the second largest value. Moreover, the 10th observation is located in the area, which is expected to be important for estimation of θ_2. The large value of leverage and the location of the 10th observation is the explanation for the high value of $DIM_{\widehat{\boldsymbol{\theta}},k}$.

5.1.5 Numerical example: Influence analysis using $DIM_{\widehat{\theta}_j,k}$

In order to verify whether the marginal influence measure $DIM_{\widehat{\theta}_j,k}$ can successfully identify an influential observation, we will use the same simulated data set as in the previous example in Section 5.1.4. Now, instead of contam-

inating the 40th observation, we contaminate the 9th observation, increasing its $y-$value, see Figure 5.5. Since observation 9 is located in the area which is expected to be important for estimation of θ_2 we expect that the marginal influence of this observation on $\widehat{\theta}_2$ will be larger than the marginal influence on $\widehat{\theta}_1$.

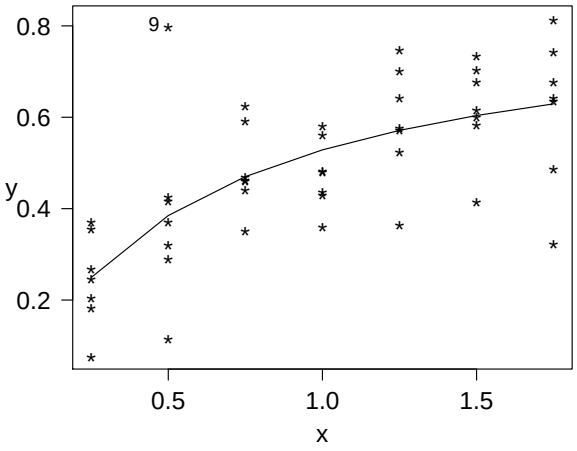
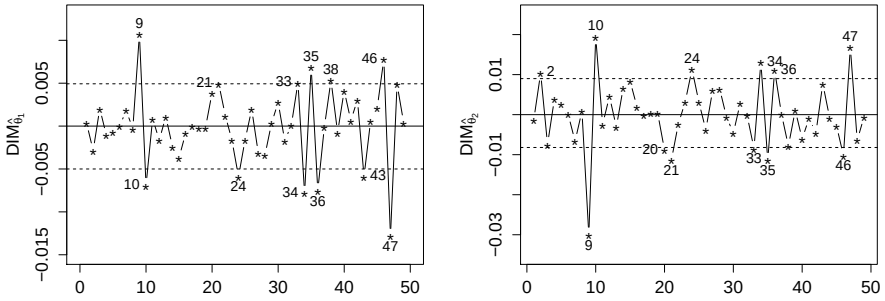


Figure 5.5: Plot of the data given in Table 5.1, where observation 9 is contaminated and observation 40 is uncontaminated.

The results from the calculations of the marginal influence of the observations on $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are presented in Figure 5.6a and Figure 5.6b, respectively. As expected, the 9th observation has the largest marginal influence on $\widehat{\theta}_2$, where $DIM_{\widehat{\theta}_2,9} = -0.03$. However, $DIM_{\widehat{\theta}_1,9} = 0.011$ is not the largest influence measure in magnitude, since $DIM_{\widehat{\theta}_1,47} = -0.013$ and the 47th observation has more influence on $\widehat{\theta}_1$ than the 9th observation.

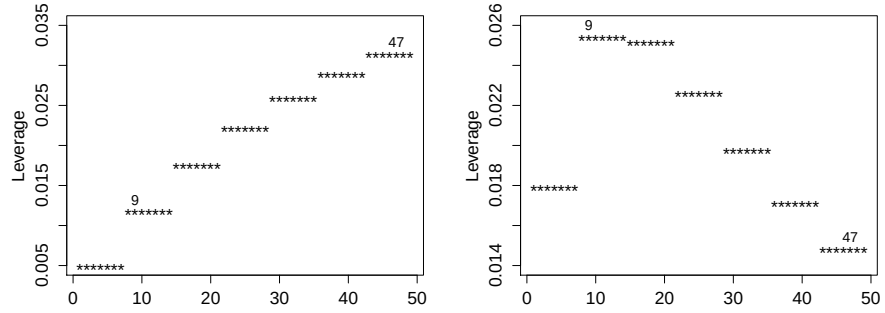
Besides studying $DIM_{\widehat{\theta}_j,k}$ we can also analyze the residuals and the marginal leverages, defined in (5.22). The standardized residuals for the 9th and 47th observations are outside the -2 to 2 range, where the standardized residual for the 9th observation is the largest in magnitude with a value of 3.33. The figure of standardized residuals is not shown here. In Figure 5.7a and Figure 5.7b the values of the marginal leverages for $\widehat{\theta}_1$ and for $\widehat{\theta}_2$ are plotted, respectively.

In Figure 5.7a we observe that the marginal leverage for the 47th observation



(a) $DIM_{\widehat{\theta}_1,k}$. **(b)** $DIM_{\widehat{\theta}_2,k}$.

Figure 5.6: The marginal influence measure, $DIM_{\widehat{\theta}_j,k}$, for $j = 1,2$ and $k = 1,\ldots,49$, when the 9th observation is contaminated. 75 percent of the data are within the dashed lines.



(a) $F_k(\widehat{\theta}_1)\left(\boldsymbol{F}(\widehat{\theta}_1)\boldsymbol{F}^T(\widehat{\theta}_1)\right)^{-1}F_k^T(\widehat{\theta}_1)$ **(b)** $F_k(\widehat{\theta}_2)\left(\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2)\right)^{-1}F_k^T(\widehat{\theta}_2)$

Figure 5.7: Marginal leverages of observations $k = 1,\ldots,49$ when the 9th observation is contaminated. (a) describes the marginal leverages when $\widehat{\theta}_1$ is under consideration and (b) describes the marginal leverages when $\widehat{\theta}_2$ is under consideration.

is much larger than for the 9th observation. In fact, observation 47 belongs to the group of observations with the largest values of marginal leverage. On the other hand, in inspecting Figure 5.7b shows that the marginal leverage of the 9th observation is much larger than the marginal leverage of the 47th observation. Thus, the explanation for our result, that observation 9 is the most

influential on $\widehat{\theta}_2$ and observation 47 the most influential on $\widehat{\theta}_1$, is found when studying the marginal leverages for these two observations.

5.2 Assessment of influence of multiple observations

Thus far, we have discussed the differentiation approach to the detection of single influential observations. However, in practice it is likely that a data set contains more than one influential observation. Influence analysis concerning multiple observations is a more challenging problem since multiple influential observations can be more difficult to detect.

Chatterjee and Hadi (1988) discussed the influence of a subset of observations on the estimated regression parameters, they argued that the problem is threefold. Firstly, it can be difficult to determine the size of the subset of observations whose influence on parameter estimates should be investigated, i.e. should we investigate all pairs or triplets of observations? If the size of the subset of interest is unknown, sequential methods can be used. For example, Belsley *et al.* (1980) suggested a procedure where one starts with a subset of two observations and analyzes every pair of observations in the data set. At the next step, one continues with examining every group of three and four and so on. The challenges with a sequential method concern how to identify a meaningful stopping rule.

Secondly, there can be computational problems when searching for multiple influential observations. For example, if we are interested in examining every pair of observations in a data set consisting of 50 observations, there will be 1225 combinations to examine. If we are also interested in examining every triplet of observations then we need to add another 19 600 combinations. For practitioners, this can be an overwhelming task. Moreover, the graphical identification of multiple influential observations is more complicated than for single influential observations. For more discussions concerning the identification of multiple influential observations we refer to, for instance, Atkinson (1986) and Peña and Yohai (1995).

Thirdly, multiple influential observations can cause so-called masking and swamping effects. Swamping occurs when observations without substantial influence on the parameter estimates are identified as influential observations due to the presence of another observation, which is highly influential. In the statistical literature several ways of defining masking has been discussed. Lawrence (1995) argued that two approaches to defining masking have emerged. He

illustrates these two approaches by using four quotations. Atkinson (1985):

> ...this structure would not be revealed by the calculation of single deletion diagnostic measures for each observation in turn, although it might well be detected by multiple deletion measures. This effect, which has been called 'masking'...

Chatterjee and Hadi (1988):

> There may exist situations in which observations are jointly but not individually influential, or the other way around...This situation is sometimes referred to as the masking effect...

Both quotations focus on a joint aspect of masking, where the masking effect concerns all the observations in one group, simultaneously. The observations in a group or subset are masking each other.

There are alternative ways for defining masking, e.g. Rousseeuw and Leroy (1987) and Atkinson (1985) argued that:

> ...masking effect means that, after the deletion of one or more influential points, another observation may emerge as extremely influential, which was not visible at first...

> ...the importance of a particular observation may not be apparent until some other observation has been deleted...In the presence of such masking effect...

These two quotations elucidate a conditional nature of masking, i.e. the observation being masked cannot be identified as an influential observation unless the observation masking it is being deleted from the data.

The two different sides of masking resulted in two distinct influence measures (Lawrence, 1995), which are both based on Cook's distance, the joint and the conditional influence measures. These influence measures will be presented in Section 5.2.1 and Section 5.2.3, respectively. In line with Lawrence (1995), we will use the terms joint and conditional influence when deriving measures for assessing the influence of multiple observations on the parameter estimates in a nonlinear regression model.

This section will be divided into several parts. In the first two parts we discuss the joint influence of multiple observations in linear and nonlinear regression analysis. In Section 5.2.1, we will give a brief overview of the discussion of

this topic given in Belsley *et al.* (1980). Then, borrowing ideas from Belsley *et al.* (1980) we derive the influence measure based on the differentiation approach to assess the influence of multiple observations simultaneously on the parameter estimates in a nonlinear regression.

The other two parts concern conditional influence of observations in linear and nonlinear regression. In Section 5.2.3, we exemplify the assessment of the conditional influence of observations on parameter estimates in a linear regression model. In Section 5.2.4, we will propose the diagnostic measure based on the differentiation approach for assessing the conditional influence of the observations on the parameter estimates in a nonlinear regression model. This diagnostic measure will be referred to as $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ and will assess the influence of the kth observation given that the ith observation is deleted.

5.2.1 Joint influence in linear regression

One approach for simultaneously assessing the influence of several observations on the parameter estimates in the linear regression model is to use an extended version of Cook's distance. Cook's distance, when the kth and lth observations are deleted, is given by

$$C_{kl} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(k,l)})^T \boldsymbol{X}^T \boldsymbol{X} (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(k,l)})}{p\widehat{\sigma}^2},$$

where $\widehat{\boldsymbol{\beta}}_{(k,l)}$ is the estimate of $\boldsymbol{\beta}$ when the kth and lth observations are excluded from calculations. This diagnostic measure is discussed more generally for a group of $m > 2$ observations by Cook and Weisberg (1980, 1982).

Belsley *et al.* (1980) suggested an idea to perturb several observations simultaneously using weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$, where $0 < \omega_k \leq 1$, $k = 1, \ldots, n$. The vector $\boldsymbol{\omega}$ is then used as the diagonal in the weight matrix $\mathbf{W}$. Consider the following perturbed model

$$\boldsymbol{y}_\omega = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\omega,$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2 \mathbf{W}^{-1}(\boldsymbol{\omega}))$ and $\mathbf{W}(\boldsymbol{\omega})$ is the diagonal weight matrix with diagonal elements equal to $\boldsymbol{\omega}$. Let K be a subset containing the indices of the observations whose influence on the parameter estimates we want to evaluate.

Belsley *et al.* (1980) also suggested to use the following "directional derivative"

$$\boldsymbol{\ell}^T \left. \frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{1}_n}, \tag{5.27}$$

to identify subsets of observations with a significant influence on $\widehat{\boldsymbol{\beta}}$. Here, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ is the weighted least squares estimate of $\boldsymbol{\beta}$ which is a function of the weight $\boldsymbol{\omega}$. Notice that if $\boldsymbol{\omega} = \mathbf{1}_n$ then $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}) = \widehat{\boldsymbol{\beta}}$, the unweighted least squares estimate of $\widehat{\boldsymbol{\beta}}$. In (5.27), $\boldsymbol{\ell} : n \times 1$ is a vector with nonzero components in the rows corresponding to indices in the subset K.

Using the derivative (5.27) we are interested in the rate of change of the function $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ as the weights simultaneously vary. However, there are several ways to allow for this. If we are interested in letting the weights vary equally fast we could use the direction defined by the vector with ones in the rows with indices in the set K and zeros elsewhere. However, there are many vectors pointing in this direction. We could replace the ones with twos, and this vector would point in the same direction. Therefore, often the unit vector is used to give the desired direction, hence $\|\boldsymbol{\ell}\| = \sqrt{\boldsymbol{\ell}^T\boldsymbol{\ell}} = 1$.

We will now borrow the idea of using the "directional" derivative and define the influence measure $DIM_{\widehat{\boldsymbol{\beta}},K}$ for assessing the influence of multiple observations on $\widehat{\boldsymbol{\beta}}$.

Definition 5.2.1. *The diagnostic measure for assessing the influence of the observations with indices specified in the subset K on* $\widehat{\boldsymbol{\beta}}$*, is defined as*

$$DIM_{\widehat{\boldsymbol{\beta}},K} = \boldsymbol{\ell}^T \left. \frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{1}}, \tag{5.28}$$

where $\boldsymbol{\ell} : n \times 1$ *is a vector with nonzero entries in the rows corresponding to indices in K and* $\boldsymbol{\ell}^T\boldsymbol{\ell} = 1$.

In the following proposition we present the explicit expression of $DIM_{\widehat{\boldsymbol{\beta}},K}$.

Proposition 5.2.1. *Let the* $DIM_{\widehat{\boldsymbol{\beta}},K}$ *be given in Definition 5.2.1. Then*

$$DIM_{\widehat{\boldsymbol{\beta}},K} = \boldsymbol{\ell}^T \boldsymbol{D}_r \boldsymbol{X} \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},$$

where $\boldsymbol{D}_r : n \times n$ *is a diagonal matrix with diagonal elements* $r_1, \ldots, r_n$ *and* $r_i = y_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}$, $i = 1, \ldots, n$.

Proof. Let $\mathbf{W} = \mathbf{W}(\boldsymbol{\omega})$. The derivative

$$\left.\frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \left.\frac{d(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{W}\boldsymbol{y}}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}$$

is calculated using the product rule as well as the chain rule and the expression for the derivative of an inverse (see Appendix A)

$$\begin{aligned}\frac{d\widehat{\beta}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} &= \frac{d\mathbf{W}}{d\boldsymbol{\omega}}(\boldsymbol{y}\otimes\boldsymbol{X})\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1} - \frac{d\mathbf{W}}{d\boldsymbol{\omega}}(\boldsymbol{X}\otimes\boldsymbol{X}) \\ &\quad \times\left(\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\otimes\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\right)\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{y}\otimes\boldsymbol{I}_p\right). \quad (5.29)\end{aligned}$$

In the expression above

$$\frac{d\mathbf{W}}{d\boldsymbol{\omega}} = \boldsymbol{U}^* = (\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_n)^T,$$

where

$$\boldsymbol{u}_i = \boldsymbol{d}_i\otimes\boldsymbol{d}_i$$

and $\boldsymbol{d}_i$ is the ith column of the identity matrix of size n.

Evaluating (5.29) at $\boldsymbol{\omega} = \mathbf{1}_n$ implies that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega} = \mathbf{1}_n) = \widehat{\boldsymbol{\beta}}$, i.e. the estimate of $\boldsymbol{\beta}$ from the unperturbed model (2.1) and $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega} = \mathbf{1}_n)$ is denoted $\boldsymbol{r}$, the residuals from the unperturbed model. Now

$$\begin{aligned}&\left.\frac{d\widehat{\beta}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \\ &= \boldsymbol{U}^*\left[(\boldsymbol{y}\otimes\boldsymbol{X})\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} - (\boldsymbol{X}\otimes\boldsymbol{X})\left(\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\otimes\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)\left(\boldsymbol{X}^T\boldsymbol{y}\otimes\boldsymbol{I}_p\right)\right] \\ &= \boldsymbol{U}^*\left[\left(\boldsymbol{y}\otimes\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right) - \left(\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}\otimes\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)\right] \\ &= \boldsymbol{U}^*\left[\left(\boldsymbol{y}\otimes\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right) - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\otimes\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)\right] \\ &= \boldsymbol{U}^*\left[(\boldsymbol{y}\otimes\boldsymbol{I}_n) - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\otimes\boldsymbol{I}_n\right)\right]\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \\ &= \boldsymbol{U}^*\left[\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\otimes\boldsymbol{I}_n\right)\right]\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},\end{aligned}$$

where $\boldsymbol{U}^*\left[(\boldsymbol{y}\otimes\boldsymbol{I}_n) - \left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}\otimes\boldsymbol{I}_n\right)\right] = \boldsymbol{D}_r$ and where $\boldsymbol{D}_r$ is defined in the statement of the proposition. Thus, we get

$$\boldsymbol{\ell}^T\left.\frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \boldsymbol{\ell}^T\boldsymbol{D}_r\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},$$

and the proposition is established.

■

Corollary 5.2.1. *The joint influence measure $DIM_{\widehat{\boldsymbol{\beta}},K}$ is a linear combination of the influence measure $EIC_{\widehat{\boldsymbol{\beta}},k}$, for $k \in K$, defined in (5.2).*

Proof. Observe that

$$\begin{aligned}\left.\frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} &= \boldsymbol{D}_r\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \\ &= \begin{pmatrix} r_1x_{11} & \dots & r_1x_{1p} \\ \vdots & \ddots & \vdots \\ r_nx_{n1} & \dots & r_nx_{np} \end{pmatrix}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},\end{aligned}$$

is a matrix of n partial derivatives. The kth row of this matrix can be written

$$r_k\boldsymbol{x}_k^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1},$$

which is equal to $EIC_{\widehat{\boldsymbol{\beta}},k}$ defined in (5.2). Therefore

$$\boldsymbol{\ell}^T\left.\frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \boldsymbol{\ell}^T\left(EIC_{\widehat{\boldsymbol{\beta}},1}^T,\dots,EIC_{\widehat{\boldsymbol{\beta}},n}^T\right)^T,$$

which shows the corollary. ■

As an example of the above corollary, let the observations contained in the subset K have indices 1 and 2. Then, $\boldsymbol{\ell}^T = (\ell_1,\ell_2,0,\dots,0)$ and

$$\boldsymbol{\ell}^T\left.\frac{d\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \ell_1EIC_{\widehat{\boldsymbol{\beta}},1} + \ell_2EIC_{\widehat{\boldsymbol{\beta}},2}.$$

It is natural to think of (5.28) as a summary measure used to assess joint influence, i.e. influence of multiple observations simultaneously on the parameter estimates.

In the next section we will extend these ideas and derive a joint influence measure for assing the influence of multiple observations on the parameter estimates in nonlinear regression models.

5.2.2 Joint influence in nonlinear regression

In this section we present another new result of the thesis, i.e. a diagnostic measure for assessing the influence of multiple observations simultaneously on the parameter estimates for a nonlinear regression model.

Consider the following perturbed nonlinear model

$$\boldsymbol{y}_{\omega} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_{\omega}, \tag{5.30}$$

where $\boldsymbol{\varepsilon}_{\omega} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{W}^{-1}(\boldsymbol{\omega}))$, $\mathbf{W}(\boldsymbol{\omega}) : n \times n$ is a diagonal weight matrix, with diagonal elements $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ and where $0 < \omega_k \leq 1$, for $k = 1, \ldots, n$. Also, let K be the subset containing the indices of the observations for which we would like to assess influence.

In correspondence with Definition 5.2.1 we present the following definition

Definition 5.2.2. *The diagnostic measure for assessing the influence of the observations with indices specified in the subset K, on the parameter estimate $\widehat{\boldsymbol{\theta}}$, is defined as the following derivative*

$$DIM_{\widehat{\boldsymbol{\theta}},K} = \boldsymbol{\ell}^T \left. \frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{1}_n}, \tag{5.31}$$

where $\boldsymbol{\ell} : n \times 1$ is a vector with nonzero entries in the rows with indices in K, where $\|\boldsymbol{\ell}\| = \sqrt{\boldsymbol{\ell}^T \boldsymbol{\ell}} = 1$ and where $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the weighted least squares estimate of $\boldsymbol{\theta}$, which is a function of the weight $\boldsymbol{\omega}$.

If $\boldsymbol{\omega} \to \mathbf{1}_n$, then $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}) \to \widehat{\boldsymbol{\theta}}$, the unweighted least squares estimate.

To derive the $DIM_{\widehat{\boldsymbol{\theta}},K}$ for assessing the influence of multiple observations simultaneously on the parameter estimates, we need the weighted least squares estimate of $\boldsymbol{\theta}$ in (5.30). The weighted least squares criterion, which should be minimized, is given by

$$Q(\boldsymbol{\omega}) = (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}))^T \mathbf{W}(\boldsymbol{\omega}) (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})).$$

The normal equations are then given by

$$\left(\frac{d\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right) \mathbf{W}(\boldsymbol{\omega}) (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})) = \mathbf{0}. \tag{5.32}$$

Since $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})$ is a nonlinear function, there is generally no explicit solution to the normal equations and iterative methods are used to find an estimate. The obtained estimate of $\boldsymbol{\theta}$ is a function of the weights, $\boldsymbol{\omega}$, and is denoted $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$.

The next theorem provides an explicit expression of the $DIM_{\widehat{\boldsymbol{\theta}},K}$ defined in (5.31).

Theorem 5.2.1. *Let $DIM_{\widehat{\boldsymbol{\theta}},K}$ be given in Definition 5.2.2. Then*

$$DIM_{\widehat{\boldsymbol{\theta}},K} = \boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) \right) \left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}}) \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}(\widehat{\boldsymbol{\theta}}) (\boldsymbol{r} \otimes \mathbf{I}_q) \right)^{-1},$$

provided that the inverse exists.

In the expression above, $\boldsymbol{U}^* : n \times n^2$ is a matrix with row vectors $\boldsymbol{u}_i^T$,

$$\boldsymbol{u}_i \quad = \quad \boldsymbol{d}_i \otimes \boldsymbol{d}_i, \text{ for } i = 1, \ldots, n, \tag{5.33}$$

where $\boldsymbol{d}_i$ is the ith column of the identity matrix of size n. The quantities $\boldsymbol{r}$, $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{G}(\widehat{\boldsymbol{\theta}})$ are defined in (5.6), (5.7) and (5.8), respectively.

Proof. Consider inserting $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ in the normal equations (5.32)

$$\left. \frac{d\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})} \mathbf{W}(\boldsymbol{\omega}) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})) \right) = \mathbf{0}, \tag{5.34}$$

and letting $\boldsymbol{F} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))$, $\mathbf{W} = \mathbf{W}(\boldsymbol{\omega})$ and $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))$.

To study the influence of multiple observations on $\widehat{\boldsymbol{\theta}}$, differentiate $\boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} = \mathbf{0}$, given in (5.34), on both sides, with respect to $\boldsymbol{\omega}$

$$\frac{d}{d\boldsymbol{\omega}} \boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} = \mathbf{0}. \tag{5.35}$$

To calculate the derivative in (5.35), the product rule, defined in Appendix A, is applied

$$\frac{d}{d\boldsymbol{\omega}} \boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} = \frac{d\boldsymbol{F}}{d\boldsymbol{\omega}} (\mathbf{W}\widehat{\boldsymbol{e}} \otimes \mathbf{I}_q) + \frac{d\mathbf{W}}{d\boldsymbol{\omega}} \left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T \right) + \frac{d\widehat{\boldsymbol{e}}}{d\boldsymbol{\omega}} \mathbf{W}\boldsymbol{F}^T. \tag{5.36}$$

In the expression above

$$\frac{d\widehat{\boldsymbol{e}}}{d\boldsymbol{\omega}} = -\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\boldsymbol{\omega}}, \quad \frac{d\mathbf{W}}{d\boldsymbol{\omega}} = \boldsymbol{U}^*.$$

Applying the chain rule, see Appendix A, to (5.36) gives

$$\frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})} (\mathbf{W}\widehat{\boldsymbol{e}} \otimes \mathbf{I}_q) + \boldsymbol{U}^* \left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T \right) - \frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})} \mathbf{W}\boldsymbol{F}^T = \mathbf{0},$$

which after rearrangement of terms yields

$$\boldsymbol{U}^* \left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T\right) = \frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})} \mathbf{W} \boldsymbol{F}^T - \frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})} (\mathbf{W}\widehat{\boldsymbol{e}} \otimes \mathbf{I}_q) \right). \tag{5.37}$$

Evaluating the derivative in (5.37) at $\boldsymbol{\omega} = \mathbf{1}_n$ implies that $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega} = \mathbf{1}_n) = \widehat{\boldsymbol{\theta}}$, the estimate of $\boldsymbol{\theta}$ for the unperturbed model (2.2) and $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega} = \mathbf{1}_n)) = \boldsymbol{r}$, the residuals for the unperturbed model. Further,

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \left.\frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))\Big|_{\boldsymbol{\omega}=\mathbf{1}_n} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}),$$

i.e. the matrix of derivatives of the expectation function from the unperturbed model. Moreover,

$$\left.\frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \left.\frac{d\boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \boldsymbol{G}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))\Big|_{\boldsymbol{\omega}=\mathbf{1}_n} = \boldsymbol{G}(\widehat{\boldsymbol{\theta}}),$$

i.e. the matrix of second derivatives of the expectation function from the unperturbed model. Thus, (5.37) becomes

$$\boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\right) = \left.\frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} \left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r} \otimes \mathbf{I}_q)\right).$$

Rearranging terms yields

$$\left.\frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\right) \left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r} \otimes \mathbf{I}_q)\right)^{-1}, \tag{5.38}$$

and

$$\boldsymbol{\ell}^T \left.\frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\right) \left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})(\boldsymbol{r} \otimes \mathbf{I}_q)\right)^{-1}.$$

The proof is complete. ■

Corollary 5.2.2. *The joint influence measure $DIM_{\widehat{\boldsymbol{\theta}},K}$ is a linear combination of the influence measures $DIM_{\widehat{\boldsymbol{\theta}},k}$, for $k \in K$, defined in (5.4), where $DIM_{\widehat{\boldsymbol{\theta}},k}$ measures the influence of a single observation on $\widehat{\boldsymbol{\theta}}$.*

Proof. Observe that $\boldsymbol{U}^*(\boldsymbol{r}\otimes\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}))$ in (5.38) equals

$$\boldsymbol{U}^*\left(\boldsymbol{r}\otimes\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\right) = (r_1\boldsymbol{F}_1, r_2\boldsymbol{F}_2, \ldots, r_n\boldsymbol{F}_n)^T.$$

It follows that

$$\left.\frac{d\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}} = \begin{pmatrix} r_1\boldsymbol{F}_1^T \\ r_2\boldsymbol{F}_2^T \\ \vdots \\ r_n\boldsymbol{F}_n^T \end{pmatrix}\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\mathbf{I}_q\right)\right)^{-1} = \begin{pmatrix} DIM_{\widehat{\boldsymbol{\theta}},1} \\ \vdots \\ DIM_{\widehat{\boldsymbol{\theta}},n} \end{pmatrix},$$

where

$$DIM_{\widehat{\boldsymbol{\theta}},k} = \left(DIM_{\widehat{\theta}_1,k}, \ldots, DIM_{\widehat{\theta}_q,k}\right) = r_k\boldsymbol{F}_k^T(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\mathbf{I}_q\right)\right)^{-1}.$$

Thus,

$$\begin{aligned} DIM_{\widehat{\boldsymbol{\theta}},K} &= \boldsymbol{\ell}^T\boldsymbol{U}^*\left(\boldsymbol{r}\otimes\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}})\right)\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\mathbf{I}_q\right)\right)^{-1} \\ &= \boldsymbol{\ell}^T\left(\begin{array}{ccc} DIM_{\widehat{\boldsymbol{\theta}},1}^T, & \ldots, & DIM_{\widehat{\boldsymbol{\theta}},n}^T \end{array}\right)^T, \end{aligned}$$

which is a linear combination of $DIM_{\widehat{\boldsymbol{\theta}},k}$ for all $k \in K$ and which establish the corollary.

■

The $DIM_{\widehat{\boldsymbol{\theta}},K}$ is a diagnostic measure for assessing the simultaneous influence of several observations on the parameter estimates, $\widehat{\boldsymbol{\theta}}$. Since *all* parameters in the model are estimated from the perturbed model, $DIM_{\widehat{\boldsymbol{\theta}},K}$ is regarded to be a joint-parameter influence measure. It can be of interest to assess the influence of multiple observations on a particular parameter estimate, $\widehat{\theta}_j$, in model (2.2). If this is the case, we use the same methodology as above and obtain a marginal-parameter influence measure.

Let $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j\right)$ be a vector of parameter estimates, where

$$\widehat{\boldsymbol{\theta}}_1 = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_{j-1}, \widehat{\theta}_{j+1}, \ldots, \widehat{\theta}_q\right),$$

are the maximum likelihood estimates from the unperturbed model (2.2), and $\widehat{\theta}_j$ is estimated from the perturbed model (5.30).

Definition 5.2.3. *The marginal influence measure for assessing the influence of the observations with indices specified in K, on the parameter estimate $\widehat{\theta}_j$, is defined as the following derivative*

$$DIM_{\widehat{\theta}_j,K} = \boldsymbol{\ell}^T \left. \frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{1}_n}, \tag{5.39}$$

where $\boldsymbol{\ell} : n \times 1$ is a vector that has nonzero entries in rows with indices in K, $\boldsymbol{\ell}^T\boldsymbol{\ell} = 1$ and $\widehat{\theta}_j(\boldsymbol{\omega})$ is the weighted least squares estimate of θ_j, which is a function of the weight $\boldsymbol{\omega}$.

To get a practically applicable expression for (5.39), we need to look at the weighted least squares estimate of $\boldsymbol{\theta}$ in model (5.30). The weighted least square criterion

$$Q(\boldsymbol{\omega}) = \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right)^T \mathbf{W}(\boldsymbol{\omega}) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right),$$

is minimized via the normal equation

$$\left(\frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)}{d\theta_j}\right) \mathbf{W}(\boldsymbol{\omega}) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right) = 0. \tag{5.40}$$

The solution for (5.40) is the weighted least squares estimate, $\widehat{\theta}_j(\boldsymbol{\omega})$.

The next theorem provides an explicit expression of the marginal-parameter influence measure $DIM_{\widehat{\theta}_j,K}$ defined in (5.39).

Theorem 5.2.2. *Let $DIM_{\widehat{\theta}_j,K}$ be given in Definition 5.2.3. Then*

$$DIM_{\widehat{\theta}_j,K} = \boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\theta}_j)\right) \left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r}\right)^{-1}, \tag{5.41}$$

provided that the inverse exists.

In (5.41), $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j)$, $\boldsymbol{U}^* : n \times n^2$ is defined in (5.33), $\boldsymbol{F}^T(\widehat{\theta}_j) : n \times 1$ and $\mathbf{G}^T(\widehat{\theta}_j) : n \times 1$ are defined in (5.15) and (5.16), respectively.

Proof. Consider inserting the weighted least squares estimate of θ_j in the normal equation (5.40)

$$\left. \frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)}{d\theta_j} \right|_{\theta_j=\widehat{\theta}_j(\boldsymbol{\omega})} \mathbf{W}(\boldsymbol{\omega}) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j(\boldsymbol{\omega}))\right) = 0. \tag{5.42}$$

and letting $\boldsymbol{F} = \boldsymbol{F}(\widehat{\theta}_j(\boldsymbol{\omega}))$, $\mathbf{W} = \mathbf{W}(\boldsymbol{\omega})$ and $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j(\boldsymbol{\omega}))$.

For the subset of observations with indices in K, $DIM_{\widehat{\theta}_j,K}$ can be obtained by differentiation of (5.42) with respect to $\boldsymbol{\omega}$ on both sides, i.e.

$$\frac{d}{d\boldsymbol{\omega}}\mathbf{F}\mathbf{W}\widehat{\mathbf{e}} = 0. \tag{5.43}$$

Using the product rule (see Appendix A) to calculate the derivative in (5.43) we get

$$\frac{d}{d\boldsymbol{\omega}}\boldsymbol{F}\mathbf{W}\widehat{\boldsymbol{e}} = \frac{d\boldsymbol{F}}{d\boldsymbol{\omega}}\mathbf{W}\widehat{\boldsymbol{e}} + \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T\right) + \frac{d\widehat{\boldsymbol{e}}}{d\boldsymbol{\omega}}\mathbf{W}\boldsymbol{F}^T. \tag{5.44}$$

Recall from Theorem 5.2.1 that

$$\frac{d\widehat{\boldsymbol{e}}}{d\boldsymbol{\omega}} = -\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\boldsymbol{\omega}}, \qquad \frac{d\mathbf{W}}{d\boldsymbol{\omega}} = \boldsymbol{U}^*.$$

Next, applying the chain rule, defined in Appendix A, to (5.44) gives

$$\frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\boldsymbol{\omega})}\mathbf{W}\widehat{\mathbf{e}} + \boldsymbol{U}^*\left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T\right) - \frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\boldsymbol{\omega})}\mathbf{W}\boldsymbol{F}^T = 0,$$

and a rearrangement of terms yields

$$\boldsymbol{U}^*\left(\widehat{\boldsymbol{e}} \otimes \boldsymbol{F}^T\right) = \frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\boldsymbol{\omega})}\mathbf{W}\boldsymbol{F}^T - \frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\boldsymbol{\omega})}\mathbf{W}\widehat{\boldsymbol{e}}\right).$$

As previously mentioned, evaluating the derivative at $\boldsymbol{\omega} = \mathbf{1}_n$ gives

$$\boldsymbol{U}^*\left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\theta}_j)\right) = \left.\frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}\left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r}\right)$$

where $\widehat{\theta}_j(\boldsymbol{\omega} = \mathbf{1}_n) = \widehat{\theta}_j$, the estimate of θ_j from the unperturbed model and where $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j(\boldsymbol{\omega} = \mathbf{1}_n)) = \boldsymbol{r}$, the residuals from the unperturbed model.

Again, rearranging terms yields

$$\left.\frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \boldsymbol{U}^*\left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\theta}_j)\right)\left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r}\right)^{-1},$$

and this completes the proof. ■

Corollary 5.2.3. *The influence measure $DIM_{\widehat{\theta}_j,K}$ in (5.39), for assessing the influence of multiple observations with indices specified in K is a linear combination of the individual influence measures $DIM_{\widehat{\theta}_j,k}$ for all k specified in K.*

Proof. When only the jth parameter is estimated from the perturbed model, we have that

$$\begin{aligned}\boldsymbol{\ell}^T \frac{d\widehat{\theta}_j(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\bigg|_{\boldsymbol{\omega}=\mathbf{1}_n} &= \boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{r} \otimes \boldsymbol{F}^T(\widehat{\theta}_j)\right)\left(\boldsymbol{F}(\widehat{\theta}_j)\boldsymbol{F}^T(\widehat{\theta}_j) - \boldsymbol{G}(\widehat{\theta}_j)\boldsymbol{r}\right)^{-1} \\ &= \boldsymbol{\ell}^T \left(DIM_{\widehat{\theta}_j,1}, \ldots, DIM_{\widehat{\theta}_j,n}\right)^T,\end{aligned}$$

where $DIM_{\widehat{\theta}_j,k}$ is defined in (5.13) and which establish the corollary.

■

As an illustration of Corollary 5.2.3, let us assume that we are interested in assessing the joint influence of the 1st and 2nd observations, i.e. $K = \{1,2\}$ and $\boldsymbol{\ell}^T = (\ell_1, \ell_2, 0, \ldots, 0)$. What values to assign to ℓ_1 and ℓ_2 depends on which direction we want to use. If we want to find the rate of change of $\widehat{\theta}_j(\boldsymbol{\omega})$ we let ω_1 and ω_2 vary equally fast, e.g. we can use the direction where $\ell_1 = \ell_2 = \frac{1}{\sqrt{2}}$. In this case the simultaneous influence of the 1st and 2nd observation on $\widehat{\theta}_j$ are simply the weighted sum of the individual influence measures, i.e. $\frac{1}{\sqrt{2}}(DIM_{\widehat{\theta}_j,1} + DIM_{\widehat{\theta}_j,2})$.

Note that, since $DIM_{\widehat{\boldsymbol{\theta}},K}$ and $DIM_{\widehat{\theta}_j,K}$ are linear functions of the individual diagnostic measures, $DIM_{\widehat{\boldsymbol{\theta}},k}$ and $DIM_{\widehat{\theta}_j,k}$, the discussion in Section 5.1.2 also applies here. For instance, the $DIM_{\widehat{\boldsymbol{\theta}},K}$ is affected by the dependence between the estimated parameters in the model, due to the fact that they are estimated simultaneously. If one wants to be "certain" of how the observations in the subset K are influencing a particular parameter estimate, the marginal diagnostic measure $DIM_{\widehat{\theta}_j,K}$ should be used. Since this diagnostic measure is constructed when only the jth parameter is estimated from the perturbed model, this diagnostic measure is used to assess the influence of the observations on the particular parameter estimate. Moreover, the individual influence measures can be positive or negative. For example, the values of $DIM_{\widehat{\boldsymbol{\theta}},K}$ and $DIM_{\widehat{\theta}_j,K}$ will be close to zero if the values of the individual influence measures of the observations in the subset K are of opposite signs and similar magnitude. Noteworthy is that the joint influence of the observations in the subset K will be large in magnitude if the values of the individual influence measures of these observations are of the same signs.

Moreover, no suggestions have yet been made in the literature regarding how to study the influence of multiple observations on the parameter estimates. In the article about local influence, Cook (1986) discussed the perturbation of a subset of q observations in the linear regression case, where $1 < q < n$. A similar perturbation scheme, where subsets of observations are under consideration, might be possible in the nonlinear regression case. Using the perturbation scheme with q perturbation weights would allow for local influence assessment of, for instance, pairs or triplets of observations. However, this opportunity is not discussed explicitly by St. Laurent and Cook (1993), where the extension of the local influence approach from linear to nonlinear regression was discussed.

5.2.3 Conditional influence in linear regression

In practice, there might be situations when an observation is not identified as an influential one unless another observation is deleted first. It can also be the opposite, an observation labeled as influential does not appear to be after the deletion of another observation. To be able to handle such situations when analyzing data, we need tools to evaluate the influence of the observations conditionally on the deletion of another observation in the data set.

Lawrence (1995) was the first to introduce the term conditional influence and suggested calculating Cook's distance for an observation before and after the deletion of another observation. The conditional influence measure defined as Cook's distance of the kth observation after the deletion of the ith observation is given by

$$C_{k,(i)} = \frac{(\widehat{\boldsymbol{\beta}}_{(k,i)} - \widehat{\boldsymbol{\beta}}_{(i)})^T \boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)} (\widehat{\boldsymbol{\beta}}_{(k,i)} - \widehat{\boldsymbol{\beta}}_{(i)})}{p\widehat{\sigma}^2},$$

where $\widehat{\boldsymbol{\beta}}_{(i)}$ and $\widehat{\boldsymbol{\beta}}_{(k,i)}$ are the estimates of $\boldsymbol{\beta}$ when the ith observation is excluded from calculations and when both the kth and the ith observations are excluded, respectively. For other references concerning conditional influence, see e.g. Wang and Critchley (2000) and Poon and Poon (2001).

In this section we will derive an influence measure for assessing the influence of the kth observation on the estimate of the parameter vector in the linear regression model (2.1) conditional on the deletion of the ith observation. The ideas for deriving the conditional influence measure will then be extended to nonlinear regression models in Section 5.2.4.

Let us consider the perturbed linear model

$$\boldsymbol{y}_\omega = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{d}_i\gamma + \boldsymbol{\varepsilon}_\omega, \tag{5.45}$$

where $\boldsymbol{d}_i : n \times 1$ is the ith column of the identity matrix of size n, γ is an unknown parameter, $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2\mathbf{W}(\omega_k))$ and $\mathbf{W}(\omega_k) = diag(1,\ldots,1,\omega_k,1,\ldots,1)$. Adding the component $\boldsymbol{d}_i\gamma$ and fitting the model deletes the ith observation in the estimates (Chatterjee and Hadi, 1988). Thus, using model (5.45) we perturb the error variance of the kth observation and when the model is fitted, the ith observation is deleted. In the next definition the perturbed linear model will be utilized.

Definition 5.2.4. *The influence measure for assessing the influence of the kth observation on $\widehat{\boldsymbol{\beta}}$, conditional on the deletion of the ith observation, is defined as*

$$DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k} = \left.\frac{d\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k)}{d\omega_k}\right|_{\omega_k=1},$$

for $i,k = 1,\ldots,n$ and $i \neq k$, where $\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k)$, is the weighted least squares estimate of $\boldsymbol{\beta}$ in the perturbed model (5.45), i.e the estimate when the ith observation is excluded from the calculations.

In Definition 5.2.4 the weighted least squares estimator of $\boldsymbol{\beta}$ is needed. In order to derive the estimator we start by estimating γ in the perturbed model (5.45). Differentiating

$$Q_{(i)}(\omega_k) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{d}_i\gamma)^T \mathbf{W}(\omega_k)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{d}_i\gamma), \tag{5.46}$$

yields the following normal equation for γ

$$\frac{dQ_i(\omega_k)}{d\gamma} = -2\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \gamma\right) = 0. \tag{5.47}$$

Utilizing (5.47), the estimator is given by

$$\widehat{\gamma} = y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}.$$

Now, inserting $\widehat{\gamma}$ in (5.46) deletes the ith observation from the expression and (5.46) results in

$$Q_{(i)}(\omega_k) = \left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\boldsymbol{\beta}\right)^T \mathbf{W}_{(i)}\left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\boldsymbol{\beta}\right), \tag{5.48}$$

where $\boldsymbol{X}_{(i)} : (n-1) \times p$ is the matrix with explanatory variables with the ith row omitted, $\boldsymbol{y}_{(i)} : (n-1) \times 1$ is the response vector with the ith response omitted and $\mathbf{W}_{(i)} = \mathbf{W}_{(i)}(\omega_k)$ is the weight matrix of order $(n-1)$ with the ith row and the ith column omitted.

Minimizing (5.48), the following normal equations for $\boldsymbol{\beta}$ are obtained

$$\frac{dQ_i(\omega_k)}{d\boldsymbol{\beta}} = -2\boldsymbol{X}_{(i)}^T \mathbf{W}_{(i)} \left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\boldsymbol{\beta}\right) = \mathbf{0}. \tag{5.49}$$

Utilizing (5.49), the weighted least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k) = \left(\boldsymbol{X}_{(i)}^T \mathbf{W}_{(i)} \boldsymbol{X}_{(i)}\right)^{-1} \boldsymbol{X}_{(i)}^T \mathbf{W}_{(i)} \boldsymbol{y}_{(i)}. \tag{5.50}$$

In the next theorem we obtain the $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ utilizing (5.50). Observe that the theorem contains two explicit expressions of $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$, one for the case where $i < k$ and one for the case where $i > k$. When $i < k$, the position of the kth observation, in the response vector and the matrix of explanatory variables, is affected by the deletion of the ith observation. The kth observation will in this case be denoted $k-1$. On the other hand, when $i > k$ the position of the kth observation will not be affected by the deletion of the ith observation.

Theorem 5.2.3. *Let $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ be given in Definition 5.2.4. Then if $i > k$*

$$DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k} = \left(\boldsymbol{y}_{k,(i)} - \boldsymbol{x}_{k,(i)}^T \widehat{\boldsymbol{\beta}}_{(i)}\right) \boldsymbol{x}_{k,(i)}^T \left(\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)}\right)^{-1}.$$

Moreover, if $i < k$

$$DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k} = \left(\boldsymbol{y}_{k-1,(i)} - \boldsymbol{x}_{k-1,(i)}^T \widehat{\boldsymbol{\beta}}_{(i)}\right) \boldsymbol{x}_{k-1,(i)}^T \left(\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)}\right)^{-1}.$$

In the expressions above, $\boldsymbol{y}_{(i)} = (y_{1,(i)}, \ldots, y_{k-1,(i)}, y_{k,(i)}, \ldots, y_{n-1,(i)})^T$ is the vector of responses with the ith observation excluded. The matrix $\boldsymbol{X}_{(i)} : (n-1) \times p$ is the matrix with explanatory variables, with the ith row excluded and $\boldsymbol{x}_{k,(i)}^T$ is the kth row of $\boldsymbol{X}_{(i)}$.

Proof. The explicit expression of $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ is found by first differentiating the estimator of $\boldsymbol{\beta}$ in (5.45) and then evaluating the derivative at $\omega_k = 1$.

Using the product rule and the rule of differentiating a matrix inverse, defined in Appendix A, the derivative equals

$$\frac{d\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k)}{d\omega_k} = \frac{d\left(\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{X}_{(i)}\right)^{-1}\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{y}_{(i)}}{d\omega_k} \tag{5.51}$$

$$\begin{aligned} = \quad & \frac{d\mathbf{W}_{(i)}}{d\omega_k}\left(\boldsymbol{y}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{X}_{(i)}\right)^{-1} - \frac{d\mathbf{W}_{(i)}}{d\omega_k}\left(\boldsymbol{X}_{(i)}\otimes\boldsymbol{X}_{(i)}\right) \\ & \times\left(\left(\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{X}_{(i)}\right)^{-1}\otimes\left(\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{X}_{(i)}\right)^{-1}\right)\left(\boldsymbol{X}_{(i)}^T\mathbf{W}_{(i)}\boldsymbol{y}_{(i)}\otimes\boldsymbol{I}_p\right). \end{aligned}$$

Due to linearity of $\mathbf{W}_{(i)}$ we obtain, if $i > k$,

$$\frac{d\mathbf{W}_{(i)}}{d\omega_k} = \boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the $k-$th column of the identity matrix of size $n-1$.

If $i < k$, we obtain

$$\frac{d\mathbf{W}_{(i)}}{d\omega_k} = \boldsymbol{d}_{(k-1)}^T\otimes\boldsymbol{d}_{(k-1)}^T,$$

where $\boldsymbol{d}_{(k-1)}$ is the $(k-1)$th column of the identity matrix of size $n-1$.

Now, if $i > k$, evaluating the derived expression (5.51) at $\omega_k = 1$ we get

$$\begin{aligned} & \frac{d}{d\omega_k}\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k)\Big|_{\omega_k=1} = \left(\boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T\right)\left[\left(\boldsymbol{y}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}\right. \\ & \left. - \left(\boldsymbol{X}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}\otimes\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{y}_{(i)}\otimes\boldsymbol{I}_p\right)\right] \\ = \quad & \left(\boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T\right)\left[\left(\boldsymbol{y}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1} - \left(\boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}\right] \\ = \quad & \left(\boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T\right)\left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)}\otimes\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1} \\ = \quad & \left(\boldsymbol{d}_k^T\left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)}\right)\otimes\boldsymbol{d}_k^T\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1} \\ = \quad & \left(y_{k,(i)} - \boldsymbol{x}_{k,(i)}^T\widehat{\boldsymbol{\beta}}_{(i)}\right)\boldsymbol{x}_{k,(i)}^T\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}, \end{aligned} \tag{5.52}$$

and hence, the final expression for $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ is

$$DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k} = \left(y_{k,(i)} - \boldsymbol{x}_{k,(i)}^T\widehat{\boldsymbol{\beta}}_{(i)}\right)\boldsymbol{x}_{k,(i)}^T\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}.$$

On the other hand, if $i < k$, $\boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T$ is replaced with $\boldsymbol{d}_{(k-i)}^T \otimes \boldsymbol{d}_{(k-i)}^T$ in (5.52), resulting in

$$\begin{aligned} \frac{d}{d\omega_k}\widehat{\boldsymbol{\beta}}_{(i)}(\omega_k)\Big|_{\omega_k=1} &= \left(\boldsymbol{d}_{k-1}^T\left(\boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)}\right) \otimes \boldsymbol{d}_{k-1}^T\boldsymbol{X}_{(i)}\right)\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1} \quad (5.53) \\ &= \left(y_{k-1,(i)} - \boldsymbol{x}_{k-1,(i)}^T\widehat{\boldsymbol{\beta}}_{(i)}\right)\boldsymbol{x}_{k-1,(i)}^T\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1}, \end{aligned}$$

Thus, the final expression for $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ is

$$DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k} = \left(y_{k-1,(i)} - \boldsymbol{x}_{k-1,(i)}^T\widehat{\boldsymbol{\beta}}_{(i)}\right)\boldsymbol{x}_{k-1,(i)}^T\left(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)}\right)^{-1},$$

and the proof is complete. ■

To clarify why two expressions of the $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ are needed, we present a small example.

Example 5.2. Illustration of the structure of $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$

Let the linear regression model be

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} : 4 \times 1$, $\boldsymbol{X} : 4 \times 2$, $\boldsymbol{\beta} : 2 \times 1$ and $\boldsymbol{\varepsilon} : 4 \times 1$. Assume that we want to study the influence of the 2nd observation conditional on the deletion of the 1st observation. In this case $i < k$.

In the proof of Theorem 5.2.3 we see that we need to consider the derivative of $\mathbf{W}_{(i)}$ with respect to ω_k. First, observe that $\mathbf{W}$ equals

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Secondly, $\mathbf{W}_{(i)}$ is in this example denoted $\mathbf{W}_{(1)}$ and it equals

$$\mathbf{W}_{(1)} = \begin{pmatrix} \omega_2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with derivative

$$\frac{d\mathbf{W}_{(1)}}{d\omega_k} = (1,0,\ldots,0) = \boldsymbol{d}_{2-1}^T \otimes \boldsymbol{d}_{2-1}^T,$$

where $\boldsymbol{d}_{2-1} = \boldsymbol{d}_1$ is the first column of the identity matrix of size 3.

Consider (5.53) in the proof of Theorem 5.2.3. For this simple example we have that

$$\boldsymbol{y}_{(1)} = \begin{pmatrix} y_2 \\ y_3 \\ y_4 \end{pmatrix}, \quad \boldsymbol{X}_{(1)} = \begin{pmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{pmatrix}, \quad \widehat{\boldsymbol{\beta}}_{(1)} = \begin{pmatrix} \widehat{\beta}_{1,(1)} \\ \widehat{\beta}_{2,(1)} \end{pmatrix},$$

so that $y_{k-1,(i)}$ in (5.53) equals y_2, the first component of the vector $\boldsymbol{y}_{(1)}$. In a similar way, we observe that $\boldsymbol{x}_{k-1,(i)}^T$ in (5.53) equals $\boldsymbol{x}_2^T$, the first row of the matrix $\boldsymbol{X}_{(1)}$ and finally, the expression of $DIM_{\widehat{\boldsymbol{\beta}}_{(1)},2}$ is

$$DIM_{\widehat{\boldsymbol{\beta}}_{(1)},2} = \left(y_2 - \boldsymbol{x}_2^T \widehat{\boldsymbol{\beta}}_{(1)}\right) \boldsymbol{x}_2^T \left(\boldsymbol{X}_{(1)}^T \boldsymbol{X}_{(1)}\right)^{-1}.$$

If, on the other hand, $i = 2$ and $k = 1$, we have that $i > k$ and the position of the weight, ω_k, is not effected by the deletion of the ith observation. To see this we consider the weight matrix, $\mathbf{W}$, and the weight matrix, $\mathbf{W}_{(2)}$

$$\mathbf{W} = \begin{pmatrix} \omega_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{W}_{(2)} = \begin{pmatrix} \omega_1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence, in the case where $i > k$, the position of ω_k is unchanged in the matrix $\mathbf{W}_{(i)}$ and it follows that k can remain unchanged in the explicit expression of $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ and throughout the proof.

In the next section the ideas from this section will be used in order to derive an influence measure for assessing the conditional influence of the observations on the parameter estimates in the nonlinear regression model (2.2).

5.2.4 Conditional influence in nonlinear regression

In this section, we will define and derive an influence measure for use in nonlinear regression analysis, similar to $DIM_{\widehat{\boldsymbol{\beta}}_{(i)},k}$ discussed in the previous section. This measure is denoted $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ and is used for assessing the influence

on the kth observation on the parameter estimate of $\boldsymbol{\theta}$ in (2.2), conditional on the deletion of the ith observation.

Now, consider the following perturbed nonlinear model

$$\boldsymbol{y}_\omega = \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) + \boldsymbol{d}_i\gamma + \boldsymbol{\varepsilon}_\omega, \tag{5.54}$$

where $\boldsymbol{d}_i$ is the ith column of the identity matrix of size n and γ is an unknown parameter. Adding $\boldsymbol{d}_i\gamma$ to the perturbed model deletes the ith observation when the model is fitted, see Ross (1987).

Moreover, in (5.54) $\boldsymbol{\varepsilon}_\omega \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{W}^{-1}(\omega_k)\right)$ where the weight matrix $\mathbf{W}(\omega_k)$ equals

$$\mathbf{W}(\omega_k) = diag\,(1,\ldots,1,\omega_k,1\ldots,1)\,.$$

Thus, in (5.54) we perturb the error variance for the kth observation and when the model is fitted, the ith observation is deleted. In the next definition the perturbed nonlinear model will be utilized.

Definition 5.2.5. *The influence measure for assessing the influence of the kth observation on the parameter estimates in the nonlinear regression model (2.2), conditional on the deletion of the ith observation, is defined as*

$$DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k} = \left.\frac{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}{d\omega_k}\right|_{\omega_k=1},$$

where $\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)$, for $i,k = 1,\ldots,n$ and $i \neq k$, is the weighted least squares estimate of $\boldsymbol{\theta}$ from the perturbed model (5.54), i.e. the estimate when the ith observation is excluded from the calculations.

Observe that if $\omega_k \to 1$, then $\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k) \to \widehat{\boldsymbol{\theta}}_{(i)}$, the unweighted least squares estimate of $\boldsymbol{\theta}$ calculated without the ith observation.

To estimate $\boldsymbol{\theta}$ in the perturbed nonlinear model (5.54) the method of weighted least squares will be used. As in Section 5.2.3 we start by finding the estimator for γ, which for any $\boldsymbol{\theta}$ is given by

$$\widehat{\gamma} = y_i - f_i(\boldsymbol{X},\boldsymbol{\theta}),$$

where $f_i(\boldsymbol{X},\boldsymbol{\theta})$ is the ith entry of the vector $\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})$.

Using $\widehat{\gamma}$ in the estimation process deletes the ith observation and the normal equations obtained by minimizing the weighted sum of squares is the following

$$\left(\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)\mathbf{W}_{(i)}\left(\boldsymbol{y}_{(i)}-\boldsymbol{f}(\boldsymbol{X}_{(i)},\boldsymbol{\theta})\right)=\mathbf{0}. \tag{5.55}$$

where $\boldsymbol{X}_{(i)}$ and $\boldsymbol{y}_{(i)}$ is the design matrix and the response vector, respectively, with the ith row excluded. Moreover, $\mathbf{W}_{(i)}=\mathbf{W}_{(i)}(\omega_k)$ is the weight matrix of order $(n-1)$ with the ith row and the ith column excluded.

The normal equations in (5.55) are solved for $\boldsymbol{\theta}$ using iterative methods with the ith observation excluded from the calculations. The obtained weighted least squares estimate is denoted $\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)$.

In the next theorem, an explicit expression of $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ is presented. Observe that, as in Theorem 5.2.3, the next theorem will contain two expressions of $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$, one corresponding to the case where $i>k$ and one corresponding to the case where $i<k$, for the same reason as for the linear regression model; see the previous section.

Theorem 5.2.4. *Let $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ be given in Definition 5.2.5. Then, if $i>k$*

$$DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}=r_{k,(i)}\boldsymbol{F}_{k,(i)}^T\left(\boldsymbol{F}_{(i)}\boldsymbol{F}_{(i)}^T-\boldsymbol{G}_{(i)}\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right)^{-1},$$

provided that the matrix inverse exists.

Moreover, if $i<k$

$$DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}=r_{k-1,(i)}\boldsymbol{F}_{k-1,(i)}^T\left(\boldsymbol{F}_{(i)}\boldsymbol{F}_{(i)}^T-\boldsymbol{G}_{(i)}\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right)^{-1},$$

provided that the matrix inverse exists.

In the expressions above, $\boldsymbol{F}_{(i)}=\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})$ is a matrix of derivatives such that

$$\boldsymbol{F}_{(i)}=\left.\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{(i)}},\quad q\times(n-1)$$

where $\boldsymbol{F}_{k-1,(i)}^T$ and $\boldsymbol{F}_{k,(i)}^T$ are the $(k-1)$th row and kth row, respectively.

The matrix $\boldsymbol{G}_{(i)}=\boldsymbol{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})$ is the following matrix of derivatives

$$\boldsymbol{G}_{(i)}=\frac{d\boldsymbol{F}_{(i)}}{d\widehat{\boldsymbol{\theta}}_{(i)}},\quad q\times q(n-1).$$

Moreover, $\boldsymbol{r}_{(i)}=(r_{1,(i)},\ldots,r_{k-1,(i)},r_{k,(i)},\ldots,r_{n-1,(i)})^T=\boldsymbol{y}_{(i)}-\boldsymbol{f}(\boldsymbol{X}_{(i)},\widehat{\boldsymbol{\theta}}_{(i)})$.

Proof. Consider inserting $\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)$ in (5.55)

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)} \mathbf{W}_{(i)}\left(\boldsymbol{y}_{(i)}-\boldsymbol{f}(\boldsymbol{X}_{(i)},\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k))\right)=\mathbf{0}, \tag{5.56}$$

and letting $\widehat{\boldsymbol{e}}_{(i)}=\boldsymbol{y}_{(i)}-\boldsymbol{f}(\boldsymbol{X}_{(i)})=\boldsymbol{y}_{(i)}-\boldsymbol{f}(\boldsymbol{X}_{(i)},\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k))$, $\mathbf{W}_{(i)}=\mathbf{W}_{(i)}(\omega_k)$ and $\boldsymbol{F}_{(i)}=\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k))$. To derive the explicit expression of $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ the following derivative of the normal equations (5.56) will be utilized

$$\frac{d}{d\omega_k}\boldsymbol{F}_{(i)}\mathbf{W}_{(i)}\widehat{\boldsymbol{e}}_{(i)}=\mathbf{0}. \tag{5.57}$$

To calculate the derivative in (5.57), the product rule defined in Appendix A is applied

$$\begin{aligned}\frac{d}{d\omega_k}\left(\boldsymbol{F}_{(i)}\mathbf{W}_{(i)}\widehat{\boldsymbol{e}}_{(i)}\right)=&\frac{d\boldsymbol{F}_{(i)}}{d\omega_k}\left(\mathbf{W}_{(i)}\widehat{\boldsymbol{e}}_{(i)}\otimes\mathbf{I}_q\right)+\frac{d\mathbf{W}_{(i)}}{d\omega_k}\left(\widehat{\boldsymbol{e}}_{(i)}\otimes\boldsymbol{F}_{(i)}^T\right)\\&+\frac{d\widehat{\boldsymbol{e}}_{(i)}}{d\omega_k}\mathbf{W}_{(i)}\boldsymbol{F}_{(i)}^T.\end{aligned} \tag{5.58}$$

In (5.58)

$$\frac{d\widehat{\boldsymbol{e}}_{(i)}}{d\omega_k}=-\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)})}{d\omega_k},$$

and due to linearity of $\mathbf{W}_{(i)}$ the following expressions are obtained

$$\frac{d\mathbf{W}_{(i)}}{d\omega_k}=\boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T,\quad i>k,\quad \frac{d\mathbf{W}_{(i)}}{d\omega_k}=\boldsymbol{d}_{k-1}^T\otimes\boldsymbol{d}_{k-1}^T,\quad i<k.$$

where $\boldsymbol{d}_k$ and $\boldsymbol{d}_{k-1}$ are the kth and the $(k-1)$th column of the identity matrix of size $n-1$, respectively.

If $i>k$, continuing from (5.57) to (5.58), applying the chain rule, defined in Appendix A, to (5.58) and rearranging terms give

$$\begin{aligned}&\frac{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)})}{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}\mathbf{W}_{(i)}\boldsymbol{F}_{(i)}^T-\frac{d\boldsymbol{F}_{(i)}}{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}\left(\mathbf{W}_{(i)}\widehat{\boldsymbol{e}}_{(i)}\otimes\mathbf{I}_q\right)\right)\\&\quad=\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}_{(i)}\otimes\boldsymbol{d}_k^T\boldsymbol{F}_{(i)}^T.\end{aligned} \tag{5.59}$$

Evaluating the derivative (5.59) at $\omega_k = 1$ yields

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)})}{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}\right|_{\omega_k=1} = \frac{d\boldsymbol{f}(\boldsymbol{X}_{(i)},\widehat{\boldsymbol{\theta}}_{(i)})}{d\widehat{\boldsymbol{\theta}}_{(i)}} = \boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)}),$$

$$\left.\frac{d\boldsymbol{F}_{(i)}}{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}\right|_{\omega_k=1} = \frac{d\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})}{d\widehat{\boldsymbol{\theta}}_{(i)}} = \boldsymbol{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)}),$$

and

$$\mathbf{d}_k^T\boldsymbol{r}_{(i)}\otimes\boldsymbol{d}_k^T\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) = \left.\frac{d\widehat{\boldsymbol{\theta}}_{(i)}(\omega_k)}{d\omega_k}\right|_{\omega_k=1}\left(\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) - \mathbf{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right)$$

$$r_{k,(i)}\boldsymbol{F}_{k,(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) \;=\; DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}\left(\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) - \boldsymbol{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right).$$

Hence, the final expression of $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ is given by

$$DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k} = r_{k,(i)}\boldsymbol{F}_{k,(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) - \boldsymbol{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right)^{-1}.$$

On the other hand, if $i < k$ we have that

$$\mathbf{d}_{k-1}^T\boldsymbol{r}_{(i)}\otimes\boldsymbol{d}_{k-1}^T\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) = r_{k-1,(i)}\boldsymbol{F}_{k-1,(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}),$$

resulting in

$$DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k} = r_{k-1,(i)}\boldsymbol{F}_{k-1,(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{F}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\boldsymbol{F}_{(i)}^T(\widehat{\boldsymbol{\theta}}_{(i)}) - \boldsymbol{G}_{(i)}(\widehat{\boldsymbol{\theta}}_{(i)})\left(\boldsymbol{r}_{(i)}\otimes\mathbf{I}_q\right)\right)^{-1},$$

provided that the inverse exists, and the proof is complete.

■

We say that hidden dependencies among the observations can be revealed when studying the conditional influence. To clarify this statement and to explain why conditional influence might be of interest, consider the following: Some observations are expected to be more important than others when it comes to estimating the parameters. This is due to the functional form of the particular expectation function. For instance, when studying the Michaelis-Menten model (2.4) we know that observations with high substrate concentration (i.e. high values on X) are important for estimation of θ_1. It is expected that some of the observations with high substrate concentration will stand out from the rest when studying the influence on $\widehat{\theta}_1$. If such an observation, located in the

area important for estimation and with a high value of the corresponding influence measure were to be deleted, it would not be surprising at all if another observation located in the same area would emerge as influential. However, if an observation outside the area that is important for estimation of θ_1 would emerge as influential when deleting the influential observation with high substrate concentration, we would become suspicious. This would indicate some kind of dependence between the deleted observation and the influential observation that we were not aware of. Further investigation of these two observations are thus needed if one wants to understand the reason for this dependence.

Using the conditional influence approach can also give the researcher information about the observations that are deleted. If one observation is conditionally influential by the deletion of another observation, we get information about the deleted observation as well. For instance, we know that the observation being deleted exerts considerable influence on the other observations since it has strong enough influence to hide another influential observation.

Studying the conditional influence between observations can certainly be interesting and useful. However, one should have in mind that this approach involves case-deletion, which leads to the fact that additional iterations for parameter estimation are needed. For every i we study, i.e. for every "deleted observation", we need to re-estimate the parameters iteratively.

5.3 Summary of Chapter 5

This chapter contains numerous new influence measures to use in various situations. We will now make a summary of the proposed measures together with a short description of when to use them.

- $DIM_{\widehat{\boldsymbol{\theta}},k}$ - an influence measure to be used when we are interested in assessing the influence of the kth observation on all parameter estimates in the model. Thus, all parameter estimates in the model are of equal interest, see Section 5.1.2.

- $DIM_{\widehat{\theta}_j,k}$ - a marginal influence measure for assessing the influence of the kth observation, on a specific parameter estimate, $\widehat{\theta}_j$. The other parameters in the model are regarded as known, see Section 5.1.2.

- $DIM_{\widehat{\boldsymbol{\theta}},K}$ - a measure to be used when we are interested in assessing joint influence of observations, i.e. influence of observations considered simultaneously on the vector of parameter estimates. The indices of the

observations, for which we want to assess influence, are contained in the subset K, see Section 5.2.2.

- $DIM_{\widehat{\theta}_{j,K}}$ - a marginal influence measure for assessing the influence of multiple observations, jointly, on a specific parameter estimate, $\widehat{\theta}_j$. As above, the indices of the observations for which we want to assess influence are contained in the subset K, see Section 5.2.2.

- $DIM_{\widehat{\boldsymbol{\theta}}_{(i),k}}$ - a measure to be used when we are interested in assessing conditional influence of observations, i.e. the influence of the kth observation, conditional on the deletion of the ith observation, on the vector of parameter estimates, see Section 5.2.4.

6. Assessment of influence on the score test statistic

The existing influence measures in regression analysis are constructed to measure the impact of observations on the parameter estimates or the fitted values. However, it is of interest to assess the observations' influence on other aspects of the statistical inference as well, for instance testing a hypothesis. Some observations have a stronger impact on the outcome of hypothesis testing than others. In fact, the result of the hypothesis testing procedure can become significant or non-significant due to the influence of a single observation. If the data contains such influential observation, it is beneficial for the analyst to be able to detect it, since this observation may carry a lot of additional information.

In Chapter 4 we showed that the added variable plot can be considered as a graphical representation of the score test and we derived a nonlinear analogue, the added parameter plot, that has the same feature. Both of these plots can be used for data examination and search for influential observations on the score test statistic. In this chapter, we will continue the work on assessing the influence of the observations on the score test statistic. In Section 6.1, we will derive a diagnostic measure for assessing the influence of single observations on the score test statistic, both in linear and nonlinear regression. This diagnostic measure is derived using the differentiation approach and it is referred to as $DIMS_k$, which is an abbreviation for *Differentiation approach, Influence Measure, Score test*. In Section 6.2, $DIMS_K$, that measures the joint influence of multiple observations on the score test statistic in nonlinear regression, is proposed.

6.1 Assessment of influence of a single observation

In this section, an important aspect of the influence analysis is in focus, namely; how do individual observations contribute to the decision making when testing an hypothesis. One possibility to approach this question is through the explo-

rative analysis using the added parameter plot (see Section 4.2.1). In addition to the graphical tool, we propose in this section a formal influence measure for the score test statistic. This measure can be used to quantify the influence of the individual observations on the score test statistic, in order to pinpoint the influential observations and add more information to the analysis.

We will start by deriving the expression of the influence measure for the linear regression model in Section 6.1.1, and continue with the nonlinear regression model in Section 6.1.2.

6.1.1 Linear regression

In this section we will derive a diagnostic measure for assessing the influence of the observations on the score test statistic for testing the hypothesis $H_0 : \beta_p = 0$, where β_p is a regression parameter in the linear regression model (2.1). Moreover, when deriving the measure the same ideas as presented in Chapter 5 are used.

Consider the perturbed linear regression model discussed in Section 5.1.1

$$\boldsymbol{y}_\omega = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\omega, \tag{6.1}$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2\mathbf{W}^{-1}(\omega_k))$ and the weight matrix $\mathbf{W}(\omega_k)$ is the following diagonal matrix $\mathbf{W}(\omega_k) = diag\,(1,\ldots,\omega_k,\ldots,1)$.

Suppose that we want to use the score test to test

$$\begin{aligned} H_0 &: \beta_p = 0, \\ H_A &: \beta_p \neq 0. \end{aligned} \tag{6.2}$$

We will now derive the score test statistic, for testing (6.2), when the parameter estimates from the perturbed model are used. As a first step we will describe the parameter estimates from the perturbed model under the restriction that $\beta_p = 0$. Partitioning the perturbed model (6.1) yields

$$\boldsymbol{y}_\omega = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{\varepsilon}_\omega,$$

so that $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T$, $\beta_2 = \beta_p$ and $\boldsymbol{X} = (\boldsymbol{X}_1 \vdots \boldsymbol{X}_2)$ is a corresponding partition of $\boldsymbol{X}$. Let $\mathbf{W} = \mathbf{W}(\omega_k)$.

The estimators for $\boldsymbol{\beta}$ and σ^2 in the perturbed model, with the restriction that $\beta_p = 0$, are given by

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}(\omega_k) &= \left(((\boldsymbol{X}_1^T\mathbf{W}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\mathbf{W}\boldsymbol{y})^T, 0\right)^T \\
&= \left(\widetilde{\beta}_0(\omega_k), \widetilde{\beta}_1(\omega_k), \ldots, \widetilde{\beta}_{p-1}(\omega_k), 0\right)^T, \\
\widetilde{\sigma}^2(\omega_k) &= \frac{1}{n}(\boldsymbol{y}_\omega - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k))^T\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)).
\end{aligned}$$

In Section 2.3 it was seen that the score test statistic for testing (6.2) is a function of the score vector and the information matrix, both evaluated with parameter estimates under the null hypothesis. Here we have that the score test statistic equals

$$S(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \boldsymbol{U}^T(\widetilde{\boldsymbol{\Psi}}(\omega_k))\mathbf{I}^{-1}(\widetilde{\boldsymbol{\Psi}}(\omega_k))\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k)), \tag{6.3}$$

where $\widetilde{\boldsymbol{\Psi}}(\omega_k) = (\widetilde{\boldsymbol{\beta}}^T(\omega_k), \widetilde{\sigma}^2(\omega_k))^T$ is the vector of parameter estimates from the perturbed model under the restriction that $\beta_p = 0$. Moreover, $\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k))$ and $\mathbf{I}(\widetilde{\boldsymbol{\Psi}}(\omega_k))$ are the score vector and the information matrix, both evaluated for the parameter estimates under the null hypothesis.

Now, as a second step in deriving the expression of the score test statistic, corresponding to testing (6.2), when the parameter estimates from the perturbed model are used, we will evaluate $\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k))$ and $\mathbf{I}(\widetilde{\boldsymbol{\Psi}}(\omega_k))$. The score vector is given by

$$\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \begin{pmatrix} \boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k)) \\ \boldsymbol{U}(\widetilde{\sigma}^2(\omega_k)) \end{pmatrix},$$

where

$$\boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k)) = \left.\frac{d\ell_\omega}{d\boldsymbol{\beta}}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} = \frac{1}{\widetilde{\sigma}^2(\omega_k)}\boldsymbol{X}^T\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)), \tag{6.4}$$

and where the derivative is defined in Appendix A. In (6.4) the log-likelihood, ℓ_ω, equals

$$\ell_\omega = -\frac{2}{n}\ln(2\pi\sigma^2) + \frac{1}{2}\ln|\mathbf{W}| - \frac{1}{2\sigma^2}(\boldsymbol{y}_\omega - \boldsymbol{X}\boldsymbol{\beta})^T\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{X}\boldsymbol{\beta}),$$

where $|\boldsymbol{.}|$ denotes the determinant. Observe that $\boldsymbol{U}(\widetilde{\sigma}^2(\omega_k)) = 0$ since

$$\boldsymbol{U}(\widetilde{\sigma}^2(\omega_k)) = \left.\frac{d\ell_\omega}{d\sigma^2}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} = 0.$$

Thus, we have that

$$\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \begin{pmatrix} \boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k)) \\ 0 \end{pmatrix}. \tag{6.5}$$

From Section 2.3 we know that the information matrix evaluated with parameter estimates under the null hypothesis is block diagonal, i.e.

$$\mathbf{I}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \begin{pmatrix} \mathbf{I}(\widetilde{\boldsymbol{\beta}}(\omega_k)) & \mathbf{0}_p \\ \mathbf{0}_p^T & \mathbf{I}(\widetilde{\sigma}^2(\omega_k)) \end{pmatrix}. \tag{6.6}$$

Using (6.5) and (6.6) in (6.3) we get

$$S(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \boldsymbol{U}^T(\widetilde{\boldsymbol{\beta}}(\omega_k))\mathbf{I}^{-1}(\widetilde{\boldsymbol{\beta}}(\omega_k))\boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k)).$$

The information matrix evaluated for the parameter estimates from the perturbed model, under the restriction that $\beta_p = 0$, equals

$$\begin{aligned}
\mathbf{I}(\widetilde{\boldsymbol{\beta}}(\omega_k)) &= E\left[\boldsymbol{U}(\boldsymbol{\beta})\boldsymbol{U}^T(\boldsymbol{\beta})\right]_{\Psi=\widetilde{\Psi}(\omega_k)} \\
&= E\left[\frac{1}{\sigma^4}\boldsymbol{X}^T\mathbf{W}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T\mathbf{W}\boldsymbol{X}\right]_{\Psi=\widetilde{\Psi}(\omega_k)} \\
&= \left(\frac{1}{\sigma^4}\boldsymbol{X}^T\mathbf{W}E\left[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T\right]\mathbf{W}\boldsymbol{X}\right)_{\Psi=\widetilde{\Psi}(\omega_k)} \\
&= \left(\frac{1}{\sigma^4}\boldsymbol{X}^T\mathbf{W}\sigma^2\mathbf{W}^{-1}\mathbf{W}\boldsymbol{X}\right)_{\Psi=\widetilde{\Psi}(\omega_k)} \\
&= \frac{1}{\widetilde{\sigma}^2(\omega_k)}\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}.
\end{aligned}$$

Thus, the score test statistic for testing $H_0 : \beta_p = 0$, using the estimates of the parameters in the perturbed linear regression model, is given by

$$\begin{aligned}
&S(\widetilde{\boldsymbol{\beta}}(\omega_k)) = \boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k))^T\mathbf{I}^{-1}(\widetilde{\boldsymbol{\beta}}(\omega_k))\boldsymbol{U}(\widetilde{\boldsymbol{\beta}}(\omega_k)) \\
&= \frac{1}{\widetilde{\sigma}^2(\omega_k)}(\boldsymbol{y}_\omega - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k))^T\mathbf{W}\boldsymbol{X}\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)).
\end{aligned} \tag{6.7}$$

The score test statistic (6.7) will now be used to define the influence measure $DIMS_k$.

Definition 6.1.1. *The diagnostic measure $DIMS_k$ for assessing the influence of the kth observation on the score test statistic is defined as*

$$DIMS_k = \left.\frac{dS(\widetilde{\boldsymbol{\beta}}(\omega_k))}{d\omega_k}\right|_{\omega_k=1},$$

where $S(\widetilde{\boldsymbol{\beta}}(\omega_k))$ is defined in (6.7).

In Definition 6.1.1, observe that when $\omega_k \to 1$, $S(\widetilde{\boldsymbol{\beta}}(\omega_k)) \to S(\widetilde{\boldsymbol{\beta}})$, i.e. the score test statistic using the parameter estimates from the unperturbed linear regression model (2.1) under the restriction that $\beta_p = 0$.

The next theorem provides an explicit expression of $DIMS_k$.

Theorem 6.1.1. *Let $DIMS_k$ be given in Definition 6.1.1. Then,*

$$DIMS_k = \frac{1}{\widetilde{\sigma}^2}\left[2\widetilde{r}_k \boldsymbol{x}_k^T \mathbf{g} - \left(\boldsymbol{x}_k^T \mathbf{g}\right)^2 - \frac{S(\widetilde{\boldsymbol{\beta}})\widetilde{r}_k^2}{n}\right],$$

where $\widetilde{\boldsymbol{r}} = (\widetilde{r}_k) = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$, $\boldsymbol{x}_k^T : 1 \times p$ is the kth row of $\boldsymbol{X}$, $S(\widetilde{\boldsymbol{\beta}})$ is the score test statistic and $\mathbf{g} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\widetilde{\boldsymbol{r}} : p \times 1$.

Proof. To find $DIMS_k$ we want to take the derivative of $S(\widetilde{\boldsymbol{\beta}}(\omega_k))$ with respect to ω_k. Rewrite $S(\widetilde{\boldsymbol{\beta}}(\omega_k))$ as $\frac{a(\omega_k)}{b(\omega_k)}$, where $b(\omega_k) = \widetilde{\sigma}^2$ and $a(\omega_k) = b(\omega_k)DIMS_k$. When differentiating, the quotient rule will be used. Hence,

$$\begin{aligned}\frac{dS(\widetilde{\boldsymbol{\beta}}(\omega_k))}{d\omega_k} &= \frac{a'(\omega_k)b(\omega_k) - a(\omega_k)b'(\omega_k)}{b^2(\omega_k)} \\ &= \frac{a'(\omega_k) - S(\widetilde{\boldsymbol{\beta}}(\omega_k))b'(\omega_k)}{b(\omega_k)}.\end{aligned}$$

First, the derivative of $a(\omega_k)$ is considered. Let $\mathbf{D} = \boldsymbol{X}^T\mathbf{W}\boldsymbol{X}$ and $\mathbf{C} = \boldsymbol{X}^T\mathbf{W}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)) = \boldsymbol{X}^T\mathbf{W}\widetilde{\mathbf{e}}$. Then,

$$\begin{aligned}\frac{da(\omega_k)}{d\omega_k} &= \frac{d\mathbf{C}^T}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C} + \frac{d\mathbf{D}^{-1}}{d(\omega_k)}(\mathbf{C}\otimes\mathbf{C}) + \frac{d\mathbf{C}}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C} \\ &= 2\left(\frac{d\mathbf{C}}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C}\right) + \frac{d\mathbf{D}^{-1}}{d(\omega_k)}(\mathbf{C}\otimes\mathbf{C}).\end{aligned}$$

The derivative of $\mathbf{C}$ with respect to ω_k equals

$$\frac{d\mathbf{C}}{d\omega_k} = \frac{d}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\widetilde{\mathbf{e}} = \frac{d\mathbf{W}}{d\omega_k}(\widetilde{\mathbf{e}}\otimes\boldsymbol{X}) + \frac{d\widetilde{\mathbf{e}}}{d\omega_k}\mathbf{W}\boldsymbol{X}. \tag{6.8}$$

Applying the chain rule, defined in Appendix A, to (6.8) yields

$$\frac{d\mathbf{C}}{d\omega_k} = \boldsymbol{d}_k^T\widetilde{\boldsymbol{e}} \otimes \boldsymbol{d}_k^T\boldsymbol{X} - \frac{d\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}, \tag{6.9}$$

since

$$\frac{d\mathbf{W}}{d\omega_k} = \boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the $k-$th column of the identity matrix of size n, and

$$\frac{d\widetilde{\boldsymbol{e}}}{d\omega_k} = -\frac{d\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\widetilde{\boldsymbol{\beta}}(\omega_k)}\mathbf{W}\boldsymbol{X} = -\frac{\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}.$$

Next, the derivative of $\mathbf{D}$ with respect to ω_k is considered. Using the rule for differentiation of a matrix inverse (see Appendix A) the derivative of $\mathbf{D}^{-1}$ with respect to ω_k is given by

$$\frac{d\mathbf{D}^{-1}}{d\omega_k} = -\frac{d\mathbf{D}}{d\omega_k}\left(\mathbf{D}^{-1}\otimes\mathbf{D}^{-1}\right).$$

Now,

$$\frac{d\mathbf{D}}{d\omega_k} = \frac{d}{d\omega_k}\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right) = \frac{d\mathbf{W}}{d\omega_k}\left(\boldsymbol{X}\otimes\boldsymbol{X}\right),$$

and

$$\begin{aligned}\frac{d\mathbf{D}^{-1}}{d\omega_k} &= -\frac{d\mathbf{W}}{d\omega_k}\left(\boldsymbol{X}\otimes\boldsymbol{X}\right)\left(\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\otimes\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\right)\\ &= -\left(\boldsymbol{d}_k^T\boldsymbol{X}\otimes\boldsymbol{d}_k^T\boldsymbol{X}\right)\left(\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\otimes\left(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)^{-1}\right).\end{aligned} \tag{6.10}$$

Finally, $\widetilde{\boldsymbol{\beta}}(\omega_k = 1) = \widetilde{\boldsymbol{\beta}}$ and let $\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k = 1) = \widetilde{\boldsymbol{r}}$. The derivative in (6.9), evaluated at $\omega_k = 1$, becomes

$$\begin{aligned}\frac{d}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\widetilde{\mathbf{e}}\Big|_{\omega_k=1} &= \left(\boldsymbol{d}_k^T\widetilde{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\boldsymbol{X} - \frac{\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\boldsymbol{X}\right)_{\omega_k=1}\\ &= \widetilde{r}_k\boldsymbol{x}_k^T - \frac{\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\Big|_{\omega_k=1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)\\ &= \widetilde{r}_k\boldsymbol{x}_k^T - EIC_{\widetilde{\boldsymbol{\beta}},k}\left(\boldsymbol{X}^T\boldsymbol{X}\right),\end{aligned}$$

since we define

$$EIC_{\widetilde{\boldsymbol{\beta}},k} = \left.\frac{\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\right|_{\omega_k=1} = \left(\left.\frac{d\widetilde{\beta}_1(\omega_k)}{d\omega_k}\right|_{\omega_k=1}, \quad \ldots, \quad \left.\frac{d\widetilde{\beta}_{p-1}(\omega_k)}{d\omega_k}\right|_{\omega_k=1}, \quad 0 \right).$$

The derivative in (6.10), evaluated at $\omega_k = 1$, can be written

$$\begin{aligned} \left.\frac{d(\boldsymbol{X}^T\mathbf{W}\boldsymbol{X})^{-1}}{d\omega_k}\right|_{\omega_k=1} &= -\left(\boldsymbol{d}_k^T\boldsymbol{X} \otimes \boldsymbol{d}_k^T\boldsymbol{X}\right)\left(\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \otimes \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right) \\ &= -\left(\boldsymbol{x}_k^T \otimes \boldsymbol{x}_k^T\right)\left(\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \otimes \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right). \end{aligned}$$

Let $\mathbf{g} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\widetilde{\mathbf{r}}$. Then,

$$\begin{aligned} \left.\frac{da(\omega_k)}{d\omega_k}\right|_{\omega_k=1} &= 2\left(\widetilde{r}_k\boldsymbol{x}_k^T - EIC_{\widetilde{\boldsymbol{\beta}},k}\left(\boldsymbol{X}^T\boldsymbol{X}\right)\right)\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\widetilde{\boldsymbol{r}} \\ &\quad - \left(\boldsymbol{x}_k^T \otimes \boldsymbol{x}_k^T\right)\left(\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \otimes \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)\left(\boldsymbol{X}^T\widetilde{\mathbf{r}} \otimes \boldsymbol{X}^T\widetilde{\mathbf{r}}\right) \\ &= 2\left(\widetilde{r}_k\boldsymbol{x}_k^T - EIC_{\widetilde{\boldsymbol{\beta}},k}\left(\boldsymbol{X}^T\boldsymbol{X}\right)\right)\mathbf{g} - \left(\boldsymbol{x}_k^T \otimes \boldsymbol{x}_k^T\right)(\mathbf{g} \otimes \mathbf{g}) \\ &= 2\left(\widetilde{r}_k\boldsymbol{x}_k^T\mathbf{g} - EIC_{\widetilde{\boldsymbol{\beta}},k}\boldsymbol{X}^T\widetilde{\boldsymbol{r}}\right) - \left(\boldsymbol{x}_k^T\mathbf{g}\right)^2. \end{aligned} \tag{6.11}$$

The expression above can be simplified. Observe that the first $p-1$ elements of $\boldsymbol{X}^T\widetilde{\boldsymbol{r}}$ equals zero due to the normal equations. Moreover, observe that the last element of $EIC_{\widetilde{\boldsymbol{\beta}},k}$ is equal to zero and we get that $EIC_{\widetilde{\boldsymbol{\beta}},k}\boldsymbol{X}^T\widetilde{\boldsymbol{r}} = 0$. Now, (6.11) equals

$$\left.\frac{da(\omega_k)}{d\omega_k}\right|_{\omega_k=1} = 2\widetilde{r}_k\boldsymbol{x}_k^T\mathbf{g} - \left(\boldsymbol{x}_k^T\mathbf{g}\right)^2.$$

Moreover, the derivative of the variance term needs to be calculated. The maximum likelihood estimator of $\sigma^2(\omega_k)$ under the null hypothesis of (6.2) satisfies

$$\widetilde{\sigma}^2(\omega_k) = \frac{1}{n}\left(\widetilde{\mathbf{e}}^T\mathbf{W}\widetilde{\mathbf{e}}\right).$$

Using the product and the chain rule, defined in Appendix A, the derivative of $\widetilde{\sigma}^2(\omega_k)$ with respect to ω_k is the following

$$\begin{aligned}\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k} &= \frac{1}{n}\frac{d}{d\omega_k}\widetilde{\mathbf{e}}^T\mathbf{W}\widetilde{\mathbf{e}} = \frac{1}{n}\left(2\frac{d\widetilde{\mathbf{e}}}{d\omega_k}\mathbf{W}\widetilde{\mathbf{e}} + \frac{d\mathbf{W}}{d\omega_k}(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}})\right)\\ &= \frac{1}{n}\left(\frac{d\mathbf{W}}{d\omega_k}(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}) - 2\frac{d\boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\mathbf{W}\widetilde{\mathbf{e}}\right)\\ &= \frac{1}{n}\left(\left(\mathbf{d}_k^T\widetilde{\mathbf{e}}\otimes\mathbf{d}_k^T\widetilde{\mathbf{e}}\right) - 2\frac{d\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\widetilde{\mathbf{e}}\right)\\ &= \frac{1}{n}\left(\widetilde{e}_k^2 - 2\frac{d\widetilde{\boldsymbol{\beta}}(\omega_k)}{d\omega_k}\boldsymbol{X}^T\mathbf{W}\widetilde{\mathbf{e}}\right).\end{aligned}$$

Evaluated at $\omega_k = 1$, $\widetilde{\boldsymbol{\beta}}(\omega_k = 1) = \widetilde{\boldsymbol{\beta}}$, $\widetilde{\mathbf{e}} = \widetilde{\mathbf{r}}$ and hence

$$\begin{aligned}\left.\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k}\right|_{\omega_k=1} &= \frac{1}{n}\left(\widetilde{r}_k^2 - 2EIC_{\widetilde{\boldsymbol{\beta}},k}\boldsymbol{X}^T\widetilde{\mathbf{r}}\right)\\ &= \frac{1}{n}\widetilde{r}_k^2.\end{aligned}$$

The expression above is simplified since $EIC_{\widetilde{\boldsymbol{\beta}},k}\boldsymbol{X}^T\widetilde{\boldsymbol{r}} = 0$.

Finally, the expression for $DIMS_k$ is given by

$$DIMS_k = \frac{1}{\widetilde{\sigma}^2}\left(2\widetilde{r}_k\boldsymbol{x}_k^T\mathbf{g} - \left(\boldsymbol{x}_k^T\mathbf{g}\right)^2 - \frac{S(\widetilde{\boldsymbol{\beta}})\widetilde{r}_k^2}{n}\right),$$

and the proof is complete. ■

Remark

In Theorem (6.1.1) we can observe that $\boldsymbol{x}_k^T\boldsymbol{g} = \sum_{j=1}^n p_{kj}\widetilde{r}_j$ and $(\boldsymbol{x}_k^T\mathbf{g})^2 = \left(\sum_{j=1}^n p_{kj}\widetilde{r}_j\right)^2$, where p_{kj} denotes the element in the kth row and the jth column of the projection matrix $\mathbf{P}_{\boldsymbol{X}}$ defined in (3.5). Hence, the $DIMS_k$ can be rewritten as

$$DIMS_k = \frac{1}{\widetilde{\sigma}^2}\left[2\widetilde{r}_k\sum_{j=1}^n p_{kj}\widetilde{r}_j - \left(\sum_{j=1}^n p_{kj}\widetilde{r}_j\right)^2 - \frac{S(\widetilde{\boldsymbol{\beta}})\widetilde{r}_k^2}{n}\right]. \tag{6.12}$$

We see from (6.12) that the $DIMS_k$ is a function of the residuals under the null hypothesis, the score test statistic and the leverages for the kth observation

since the leverage, p_{kk}, is a part of $\sum_{j=1}^{n} p_{kj}$.

In the next section we will derive one of the main results in this thesis, an explicit expression of the influence measure $DIMS_k$ for use in nonlinear regression analysis.

6.1.2 Nonlinear regression

Similar ideas and techniques from the previous section, Section 6.1.1, will be utilized when deriving the influence measure, $DIMS_k$, for assessing the influence of the observations on the score test statistic, given in (2.27), when testing

$$\begin{aligned} H_0 &: \theta_q = 0 \\ H_A &: \theta_q \neq 0, \end{aligned} \tag{6.13}$$

where θ_q is the last element of the vector of parameters for the nonlinear regression model (2.2).

First, we consider the perturbed nonlinear model, discussed in Section 5.1.2, defined as

$$\boldsymbol{y}_\omega = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_\omega, \tag{6.14}$$

where $\boldsymbol{\varepsilon}_\omega \sim N_n(\mathbf{0}, \sigma^2 \mathbf{W}^{-1}(\omega_k))$ and the weight matrix $\mathbf{W}(\omega_k)$ is the following diagonal matrix

$$\mathbf{W}(\omega_k) = diag\,(1, \ldots, \omega_k, \ldots, 1)\,.$$

We will now derive the score test statistic when testing (6.13) using the parameter estimates from the perturbed model (6.14).

Let $\boldsymbol{\Psi} = (\boldsymbol{\theta}^T, \sigma^2)^T$ be the vector of parameters and let $\widetilde{\boldsymbol{\Psi}}(\omega_k) = (\widetilde{\boldsymbol{\theta}}^T(\omega_k), \widetilde{\sigma}^2(\omega_k))^T$ be the maximum likelihood estimates from the perturbed model, under the restriction that $\theta_q = 0$. Recall from Section 6.1.1 that the score test statistic is a function of the score vector and the information matrix, both evaluated with the plug-in parameter estimates under the null hypothesis, i.e.

$$S(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k))^T \mathbf{I}^{-1}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) \boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k)).$$

The score vector is given by

$$\boldsymbol{U}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \begin{pmatrix} \boldsymbol{U}(\widetilde{\boldsymbol{\theta}}(\omega_k)) \\ \boldsymbol{U}(\widetilde{\sigma}^2(\omega_k)) \end{pmatrix}.$$

As in the linear regression case $\boldsymbol{U}(\widetilde{\sigma}^2(\omega_k)) = 0$ since $\widetilde{\sigma}^2(\omega_k)$ is the maximum likelihood estimate of σ^2 and

$$\begin{aligned} \boldsymbol{U}(\widetilde{\boldsymbol{\theta}}(\omega_k)) &= \left.\frac{d\ell_\omega}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} \\ &= \frac{1}{\widetilde{\sigma}^2(\omega_k)}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))), \end{aligned}$$

where

$$\ell_\omega = -\frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\ln|\mathbf{W}| - \frac{1}{2\sigma^2}(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))^T\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})),$$

$\mathbf{W} = \mathbf{W}(\omega_k)$ and $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k)) : q \times n$ is the matrix such that

$$\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \left(\boldsymbol{F}_1(\widetilde{\boldsymbol{\theta}}(\omega_k)),\ldots,\boldsymbol{F}_n(\widetilde{\boldsymbol{\theta}}(\omega_k))\right) = \left.\frac{d}{d\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}(\omega_k)}. \quad (6.15)$$

Moreover, the information matrix is block diagonal, see Section 2.3 for details, such that

$$\mathbf{I}(\widetilde{\boldsymbol{\Psi}}(\omega_k)) = \begin{pmatrix} \mathbf{I}(\widetilde{\boldsymbol{\theta}}(\omega_k)) & \mathbf{0}_q \\ \mathbf{0}_q^T & \mathbf{I}(\widetilde{\sigma}^2(\omega_k)) \end{pmatrix}.$$

Using the results from deriving the score vector and using the fact that the information matrix is block diagonal, the score test statistic equals

$$S(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \boldsymbol{U}^T(\widetilde{\boldsymbol{\theta}}(\omega_k))\mathbf{I}^{-1}(\widetilde{\boldsymbol{\theta}}(\omega_k))\boldsymbol{U}(\widetilde{\boldsymbol{\theta}}(\omega_k)). \quad (6.16)$$

The information matrix in (6.16) is defined as

$$\begin{aligned} \mathbf{I}(\widetilde{\boldsymbol{\theta}}(\omega_k)) &= E\left[\boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{U}^T(\boldsymbol{\theta})\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} = E\left[\frac{d\ell_\omega}{d\boldsymbol{\theta}}\left(\frac{d\ell_\omega}{d\boldsymbol{\theta}}\right)^T\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} \\ &= E\left[\frac{1}{\sigma^4}\boldsymbol{F}(\boldsymbol{\theta})\mathbf{W}(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))^T\mathbf{W}\boldsymbol{F}^T(\boldsymbol{\theta})\right]_{\boldsymbol{\Psi}=\widetilde{\boldsymbol{\Psi}}(\omega_k)} \\ &= \frac{1}{\widetilde{\sigma}^2(\omega_k)}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))\mathbf{W}\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}}(\omega_k)), \end{aligned}$$

where in the second row it was used that

$$E[(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))(\boldsymbol{y}_\omega - \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}))^T] = \sigma^2 \mathbf{W}^{-1}.$$

Thus, the score test statistic for testing (6.13), derived from the perturbed nonlinear model (6.14), is as follows

$$S(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \frac{1}{\widetilde{\sigma}^2(\omega_k)} \widetilde{\mathbf{e}}^T \mathbf{W} \boldsymbol{F}^T \left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1} \boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}, \tag{6.17}$$

where $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))$ and $\widetilde{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\omega_k))$.

The score test statistic in (6.17) will now be used in the following definition of the influence measure $DIMS_k$.

Definition 6.1.2. *The diagnostic measure $DIMS_k$ for assessing the influence of the kth observation on the score test statistic is defined as*

$$DIMS_k = \left. \frac{dS(\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k} \right|_{\omega_k = 1},$$

where $S(\widetilde{\boldsymbol{\theta}}(\omega_k))$ is defined in (6.17).

Note that, when $\omega_k \to 1$ we observe that $S(\widetilde{\boldsymbol{\theta}}(\omega_k)) \to S(\widetilde{\boldsymbol{\theta}})$, i.e. the score test statistic using the parameter estimates from the unperturbed nonlinear regression model (2.2) under the restriction that $\theta_q = 0$.

Before presenting the explicit expression of the $DIMS_k$ in a theorem, we will state the definition of the influence measure $DIM_{\widetilde{\boldsymbol{\theta}},k}$, similar to the influence measure presented in Definition 5.1.2, since $DIMS_k$ is a function of $DIM_{\widetilde{\boldsymbol{\theta}},k}$.

Definition 6.1.3. *Let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_{q-1}, 0)^T$ be the parameter estimates under the null hypothesis, $H_0 : \theta_q = 0$. The diagnostic measure for assessing the influence of the kth observation on $\widetilde{\boldsymbol{\theta}}$ is defined as*

$$DIM_{\widetilde{\boldsymbol{\theta}},k} \;=\; \left(DIM_{\widehat{\boldsymbol{\theta}},k},\, 0\right),$$

where $DIM_{\widehat{\boldsymbol{\theta}},k}$ is given in Definition 5.1.2, and $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_{q-1})^T$ are the parameter estimates for the restricted model, i.e. the model under the null hypothesis.

The next theorem provides an explicit expression of the $DIMS_k$ for use in nonlinear regression analysis.

Theorem 6.1.2. *Let $DIMS_k$ be given in Definition 6.1.2. Then*

$$\begin{aligned} DIMS_k = \frac{1}{\widetilde{\sigma}^2}\Big[&2\left(\widetilde{r}_k \boldsymbol{F}_k^T \mathbf{g} + DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}(\widetilde{\mathbf{r}}\otimes\mathbf{g})\right) \\ &-DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\mathbf{g}\otimes\mathbf{g}\right)+\mathbf{G}^*\left(\mathbf{g}\otimes\boldsymbol{F}^T\mathbf{g}\right)\right) \\ &-\left(\mathbf{g}^T\boldsymbol{F}_k\boldsymbol{F}_k^T\mathbf{g}\right) - S(\widetilde{\boldsymbol{\theta}})\frac{\widetilde{r}_k^2}{n}\Big], \end{aligned}$$

where $\widetilde{\mathbf{r}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}})$ and $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ is defined (6.15). Moreover, $\mathbf{G}$ and $\mathbf{G}^*$ are defined as

$$\mathbf{G} = \mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \frac{d\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}}, \qquad \mathbf{G}^* = \mathbf{G}^*(\widetilde{\boldsymbol{\theta}}) = \frac{d\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}}, \tag{6.18}$$

respectively, and $\mathbf{g} = \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}$.

Proof. Let $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))$, $\mathbf{W} = \mathbf{W}(\omega_k)$ and $\widetilde{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))$. In (6.17), let

$$a(\omega_k) = \widetilde{\mathbf{e}}^T\mathbf{W}\boldsymbol{F}^T\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}$$

and

$$b(\omega_k) = \widetilde{\sigma}^2(\omega_k).$$

When differentiating $S(\widetilde{\boldsymbol{\theta}}(\omega_k))$ the quotient rule is used. Hence,

$$\begin{aligned} \frac{dS(\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k} &= \frac{a'(\omega_k)b(\omega_k) - a(\omega_k)b'(\omega_k)}{b^2(\omega_k)} \\ &= \frac{a'(\omega_k) - S(\widetilde{\boldsymbol{\theta}}(\omega_k))b'(\omega_k)}{b(\omega_k)}, \end{aligned}$$

where

$$a'(\omega_k) = \frac{da(\omega_k)}{d\omega_k}, \qquad b'(\omega_k) = \frac{db(\omega_k)}{d\omega_k}.$$

First, the derivative of $a(\omega_k)$ is considered. Let $\mathbf{C} = \boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}$ and $\mathbf{D} = \boldsymbol{F}\mathbf{W}\boldsymbol{F}^T$, then

$$\begin{aligned} \frac{da(\omega_k)}{d\omega_k} &= \frac{d\mathbf{C}^T}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C} + \frac{d\mathbf{D}^{-1}}{d(\omega_k)}(\mathbf{C}\otimes\mathbf{C}) + \frac{d\mathbf{C}}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C} \\ &= 2\left(\frac{d\mathbf{C}}{d(\omega_k)}\mathbf{D}^{-1}\mathbf{C}\right) + \frac{d\mathbf{D}^{-1}}{d(\omega_k)}(\mathbf{C}\otimes\mathbf{C}). \end{aligned}$$

The derivative of $\mathbf{C}$ with respect to ω_k is

$$\begin{aligned}\frac{d\mathbf{C}}{d\omega_k} &= \frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}} \\ &= \frac{d\boldsymbol{F}}{d\omega_k}\left(\mathbf{W}\widetilde{\mathbf{e}}\otimes\mathbf{I}_q\right)+\frac{d\mathbf{W}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\mathbf{F}^T\right)+\frac{d\widetilde{\mathbf{e}}}{d\omega_k}\mathbf{W}\boldsymbol{F}^T. \qquad (6.19)\end{aligned}$$

Applying the chain rule, defined in Appendix A, to (6.19) gives

$$\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{W}\widetilde{\mathbf{e}}\otimes\mathbf{I}_q\right)-\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}\boldsymbol{F}^T\right)+\boldsymbol{d}_k^T\widetilde{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\boldsymbol{F}^T \qquad (6.20)$$

since

$$\frac{d\mathbf{W}}{d\omega_k}=\boldsymbol{d}_k^T\otimes\boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the kth column of the identity matrix of size n.

Next, the derivative of $\mathbf{D}$ with respect to ω_k is considered. Using both the chain rule and the rule for differentiation of a matrix inverse (see Appendix A), the derivative of $\mathbf{D}^{-1}$ with respect to ω_k is

$$\frac{d\mathbf{D}^{-1}}{d\omega_k}=-\frac{d\mathbf{D}}{d\omega_k}\left(\mathbf{D}^{-1}\otimes\mathbf{D}^{-1}\right).$$

Now,

$$\begin{aligned}\frac{d\mathbf{D}}{d\omega_k} &= \frac{d}{d\omega_k}\left(\mathbf{F}\mathbf{W}\boldsymbol{F}^T\right) \\ &= \frac{d\boldsymbol{F}}{d\omega_k}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\frac{d\mathbf{W}}{d\omega_k}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)+\frac{d\boldsymbol{F}^T}{d\omega_k}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right) \\ &= \frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right)\right) \\ &\quad+\frac{d\mathbf{W}}{d\omega_k}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right),\end{aligned}$$

and

$$\begin{aligned}\frac{d\mathbf{D}^{-1}}{d\omega_k}=-\Bigg[&\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right)\right) \qquad (6.21)\\ &+\frac{d\mathbf{W}}{d\omega_k}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\Bigg]\left(\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1}\right).\end{aligned}$$

Consider evaluation at $\omega_k = 1$. We get that

$$\widetilde{\boldsymbol{\theta}}(\omega_k = 1) = \widetilde{\boldsymbol{\theta}}, \quad \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\omega_k = 1)) = \widetilde{\boldsymbol{r}},$$

the parameter estimates and the residuals from the unperturbed model (2.2), respectively. Moreover,

$$\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k = 1)) = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}), \quad \boldsymbol{G} = \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}(\omega_k = 1)) = \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}),$$
$$\boldsymbol{G}^* = \boldsymbol{G}^*(\widetilde{\boldsymbol{\theta}}(\omega_k = 1)) = \boldsymbol{G}^*(\widetilde{\boldsymbol{\theta}}),$$

are the matrices of derivatives evaluated for the parameter estimates from the unperturbed model (2.2).

The derivatives in (6.20) evaluated at $\omega_k = 1$ equals

$$\left.\frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}^{-1}\widetilde{\mathbf{e}}\right|_{\omega_k=1} = \widetilde{r}_k\boldsymbol{F}_k^T - DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\boldsymbol{F}\boldsymbol{F}^T - \mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right),$$

since

$$\left.\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\right|_{\omega_k=1} = DIM_{\widetilde{\boldsymbol{\theta}},k},$$

and the derivative in (6.21) evaluated at $\omega_k = 1$ equals

$$\begin{aligned}\left.\frac{d(\boldsymbol{F}\mathbf{W}^{-1}\boldsymbol{F}^T)^{-1}}{d\omega_k}\right|_{\omega_k=1} = & -\Big[DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)\\ & +\boldsymbol{d}_k^T\boldsymbol{F}^T\otimes\boldsymbol{d}_k^T\boldsymbol{F}^T\Big]\left(\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\right),\end{aligned}$$

since

$$\left.\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\right|_{\omega_k=1} = \boldsymbol{G}, \quad \left.\frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\right|_{\omega_k=1} = \boldsymbol{G}^*.$$

Let $\mathbf{g} = \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}$, then

$$\begin{aligned}\left.\frac{da(\omega_k)}{d\omega_k}\right|_{\omega_k=1} = & \ 2\left(\widetilde{r}_k\boldsymbol{F}_k^T - DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\boldsymbol{F}\boldsymbol{F}^T - \mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\right)\mathbf{g}\\ & -\Big[DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)\\ & +\boldsymbol{d}_k^T\boldsymbol{F}^T\otimes\boldsymbol{d}_k^T\boldsymbol{F}^T\Big]\left(\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\right)\left(\boldsymbol{F}\widetilde{\mathbf{r}}\otimes\boldsymbol{F}\widetilde{\mathbf{r}}\right)\\ = & \ 2\left(\widetilde{r}_k\boldsymbol{F}_k^T + DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\mathbf{g}\\ & -\Big[DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes I_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)\\ & +\boldsymbol{d}_k^T\boldsymbol{F}^T\otimes\boldsymbol{d}_k^T\boldsymbol{F}^T\Big]\left(\mathbf{g}\otimes\mathbf{g}\right).\end{aligned}$$

In the expression above, $DIM_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\boldsymbol{F}^T\mathbf{g} = 0$. This is due to the fact that the normal equations for estimating $\boldsymbol{\theta}$ under the restriction that $\theta_q = 0$ is set to zero, the last element in $DIM_{\widetilde{\boldsymbol{\theta}},k}$ is equal to zero and $DIM_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}} = DIM_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\widetilde{\mathbf{r}} = 0$.

Now the derivative of the variance term needs to be calculated. The maximum likelihood estimator of $\sigma^2(\omega_k)$ under the restriction that $\theta_q = 0$ is

$$\widetilde{\sigma}^2(\omega_k) = \frac{1}{n}\left(\widetilde{\mathbf{e}}^T\mathbf{W}\widetilde{\mathbf{e}}\right).$$

Using the product rule and the chain rule (see Appendix A), the derivative of $\widetilde{\sigma}^2(\omega_k)$ with respect to ω_k is the following

$$\begin{aligned}
\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k} &= \frac{1}{n}\frac{d}{d\omega_k}\widetilde{\mathbf{e}}^T\mathbf{W}\widetilde{\mathbf{e}} = \frac{1}{n}\left(2\frac{d\widetilde{\mathbf{e}}}{d\omega_k}\mathbf{W}\widetilde{\mathbf{e}} + \frac{d\mathbf{W}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&= \frac{1}{n}\left(\frac{d\mathbf{W}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right) - 2\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k}\mathbf{W}\widetilde{\mathbf{e}}\right) \\
&= \frac{1}{n}\left(\left(\mathbf{d}_k^T\otimes\mathbf{d}_k^T\right)\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&\quad -\frac{1}{n}\left(2\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}\widetilde{\mathbf{e}}\right)\right) \\
&= \frac{1}{n}\left(\widetilde{e}_k^2 - 2\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\mathbf{W}\widetilde{\mathbf{e}}\right)\right). \qquad (6.22)
\end{aligned}$$

Evaluating (6.22) at $\omega_k = 1$ we get

$$\left.\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k}\right|_{\omega_k=1} = \frac{1}{n}\left(\widetilde{r}_k^2 - 2DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{F}\widetilde{\mathbf{r}}\right) = \frac{\widetilde{r}_k^2}{n},$$

since $DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{F}\widetilde{\mathbf{r}} = 0$.

Finally, the expression for $DIMS_k$ is given by

$$\begin{aligned} DIMS_k &= \frac{1}{\widetilde{\sigma}^2}\Bigg[2\left(\widetilde{r}_k \boldsymbol{F}_k^T + DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\mathbf{g} \\ &\quad - \left(DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)+\boldsymbol{d}_k^T\boldsymbol{F}^T\otimes\boldsymbol{d}_k^T\boldsymbol{F}^T\right) \\ &\quad \times(\mathbf{g}\otimes\mathbf{g}) - S(\widetilde{\boldsymbol{\theta}})\frac{\widetilde{r}_k^2}{n}\Bigg] \\ &= \frac{1}{\widetilde{\sigma}^2}\Bigg[2\left(\widetilde{r}_k\boldsymbol{F}_k^T\mathbf{g} + DIM_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{g}\right)\right) \\ &\quad - DIM_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\mathbf{g}\otimes\mathbf{g}\right)+\mathbf{G}^*\left(\mathbf{g}\otimes\boldsymbol{F}^T\mathbf{g}\right)\right) - \mathbf{g}^T\boldsymbol{F}_k\boldsymbol{F}_k^T\mathbf{g} - S(\widetilde{\boldsymbol{\theta}})\frac{\widetilde{r}_k^2}{n}\Bigg]. \end{aligned}$$

■

In Theorem 6.1.2 we observe that $\boldsymbol{F}_k^T\mathbf{g} = \sum_{j=1}^n p_{kj}\widetilde{r}_j$ and that $\boldsymbol{F}_k^T\mathbf{g}\otimes\boldsymbol{F}_k^T\mathbf{g} = \left(\sum_{j=1}^n p_{kj}\widetilde{r}_j\right)^2$, where p_{kj} is the kth row and the jth column of the tangent plane projection matrix $\mathbf{P}_{\boldsymbol{F}} = \boldsymbol{F}^T(\boldsymbol{F}\boldsymbol{F}^T)^{-1}\boldsymbol{F}$, and where p_{kk} is defined in (3.7). As in the linear regression case, the $DIMS_k$ is a function of the residuals under the null hypothesis, the score test statistic and the leverages of the kth observation, since the leverage of the kth observation is p_{kk}. However, the expression is more complicated due to the fact that we have to consider the second derivative of the expectation function. Also, the influence measure (DIM) for the parameter estimates under the null hypothesis has a more complicated expression in the nonlinear regression case, compared to the linear regression case.

As discussed in Section 5.1.3, an apparent benefit of the approach used to construct $DIMS_k$ is that when differentiating various quantities with respect to ω_k we evaluate at $\omega_k = 1$. As a consequence, the resulting quantities in the expression of $DIMS_k$, first obtained from the perturbed model, is now independent of the weight, ω_k, and equals the quantities for the unperturbed model. If we would evaluate these derivatives at any value other than one, e.g at $\omega_k \to 0$, the expression of $DIMS_k$ would become more complicated. Moreover, for each k of interest we would need to re-estimate the model, since then the parameter estimate would be a function of the weight, e.g. $\widetilde{\boldsymbol{\theta}}(\omega_k \to 0)$.

The signs of the values of the $DIMS_k$ provide important information. A positive value of the $DIMS_k$ means that the kth observation has positive influence on the score test statistic, i.e. the presence of this observation is increasing

the value of the score test statistic. Similarly, the kth observation exercises a negative influence on the score test statistic if the value of $DIMS_k$ is negative. This means that the presence of the kth observation is reducing the score test statistic.

To illustrate the components of $DIMS_k$ given in Theorem 6.1.2 we will give a small technical example.

Example 6.1: An illustration the components of the $DIMS_k$

Consider the same model as in Example 5.1, where

$$\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) = \left(\begin{array}{ccc} \frac{\theta_1 x_1}{\theta_2+x_1}, & \frac{\theta_1 x_2}{\theta_2+x_2}, & \frac{\theta_1 x_3}{\theta_2+x_3} \end{array} \right)^T.$$

Let the hypothesis of interest be $H_0 : \theta_2 = 0$ and the vector of estimated parameters under the null hypothesis be $\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\theta}_1, 0\right)^T$. Of course, there is no practical interest in testing the hypothesis $H_0 : \theta_2 = 0$, since under the null hypothesis the expectation function equals a constant. However, this example is constructed to display the components of $DIMS_k$, and for this purpose the example works well.

From Theorem 6.1.2 we can see that the components of $DIMS_k$ are $\boldsymbol{F}$, $\widetilde{\mathbf{r}}$, $\mathbf{G}$, $\mathbf{G}^*$ and $DIM_{\widetilde{\boldsymbol{\theta}},k}$. Now, let us describe these matrices.

For this particular test and model, the vector of residuals that results from estimating the model $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ under the null hypothesis is given by

$$\widetilde{\mathbf{r}} = \boldsymbol{y} - \widetilde{\theta}_1 \mathbf{1}_3.$$

The first row of the 2×3 matrix $\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ is the following

$$\left.\frac{d}{d\theta_1}\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \mathbf{1}_3^T,$$

its second row equals

$$\left.\frac{d}{d\theta_2}\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left(\begin{array}{ccc} -\frac{\widetilde{\theta}_1}{x_1} & -\frac{\widetilde{\theta}_1}{x_2} & -\frac{\widetilde{\theta}_1}{x_3} \end{array} \right),$$

and

$$\boldsymbol{F}_k^T(\widetilde{\boldsymbol{\theta}}) = \left(\begin{array}{cc} \frac{df_k(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}, & \frac{df_k(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2} \end{array} \right) = \left(\begin{array}{cc} 1, & -\frac{\widetilde{\theta}_1}{x_k} \end{array} \right).$$

The influence measure $DIM_{\widetilde{\boldsymbol{\theta}},k}$ for the parameter estimates under the null hypothesis is given by

$$DIM_{\widetilde{\boldsymbol{\theta}},k} = \left(\begin{array}{cc} DIM_{\widetilde{\theta}_1,k}, & DIM_{\widetilde{\theta}_2,k} \end{array} \right) = \left(\begin{array}{cc} DIM_{\widehat{\theta}_1,k}, & 0 \end{array} \right).$$

The $q \times nq$ matrix $\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \left. \frac{d\boldsymbol{F}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$ is here

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \left(\begin{array}{cccccc} \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} \\ \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} \end{array} \right).$$

In the matrix $\mathbf{G}(\widetilde{\boldsymbol{\theta}})$

$$\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} = 0, \quad \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} = \frac{2\widetilde{\theta}_1}{x_i^2}$$

and

$$\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} = \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} = -\frac{x_i}{(\widetilde{\theta}_2 + x_i)^2} = -\frac{1}{x_i},$$

so that

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \left(\begin{array}{cccccc} 0 & -\frac{1}{x_1} & 0 & -\frac{1}{x_2} & 0 & -\frac{1}{x_3} \\ -\frac{1}{x_1} & \frac{2\widetilde{\theta}_1}{x_1^2} & -\frac{1}{x_2} & \frac{2\widetilde{\theta}_1}{x_2^2} & -\frac{1}{x_3} & \frac{2\widetilde{\theta}_1}{x_3^2} \end{array} \right).$$

Similarly, the $q \times nq$ matrix $\mathbf{G}^*(\widetilde{\boldsymbol{\theta}}) = \left. \frac{d\boldsymbol{F}^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$ and thus

$$\begin{aligned} \mathbf{G}^*(\widetilde{\boldsymbol{\theta}}) &= \frac{d\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}} \\ &= \left(\begin{array}{cccccc} \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} \\ \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} \end{array} \right) \\ &= \left(\begin{array}{cccccc} 0 & 0 & 0 & -\frac{1}{x_1} & -\frac{1}{x_2} & -\frac{1}{x_3} \\ -\frac{1}{x_1} & -\frac{1}{x_2} & -\frac{1}{x_3} & \frac{2\widetilde{\theta}_1}{x_1^2} & \frac{2\widetilde{\theta}_1}{x_2^2} & \frac{2\widetilde{\theta}_1}{x_3^2} \end{array} \right). \end{aligned}$$

In the next section a continuation of the numerical example in Section 4.2.2 will be given. In this example we illustrate how the influence diagnostic, $DIMS_k$, can be used together with the added parameter plot.

6.1.3 Numerical example

This numerical example illustrates how the influence diagnostic $DIMS_k$ can be used in a practical situation. We continue the numerical example in Section 4.2.2, where we fitted the modified Michaelis-Menten model (4.23) under the null hypothesis, $H_0 : \theta_4 = 0$. The added parameter plot for $\widetilde{\theta}_4$ is presented in Figure 4.2. By inspection of the scatter in the plot we concluded that the 1st and 2nd observation were a bit far from the rest and that these observations could be influential observations. We will now assess the influence of all the observations in the data set by using the influence measure $DIMS_k$.

The values of $DIMS_k$ are computed for $k = 1, \ldots, 23$ and the results are presented in Figure 6.1.

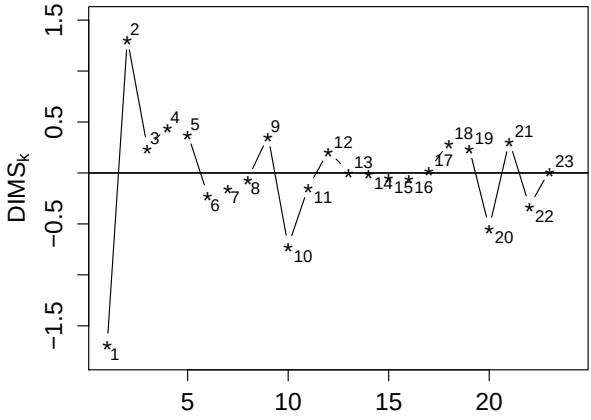


Figure 6.1: A plot of $DIMS_k$ against the observation number, where $DIMS_k$ is the diagnostic measure for assessing the influence of the observations on the score test statistic, given in Definition 6.1.2. The data used are presented in Table 4.1.

The 1st and 2nd observation have the largest absolute values of the $DIMS_k$. All observations, the 1st and 2nd observations excluded, have values of the influence measure within +/-0.72, whereas $DIMS_1 = -1.68$ and $DIMS_2 = 1.32$. Relative to the other observations, the 1st and 2nd observations clearly have more influence on the outcome of the score testing procedure.

The signs of the values of $DIMS_k$ can also give us some additional information. A negative value of the $DIMS_k$ tells us that the presence of the kth observation is decreasing the value of the score test statistic. Recall that we noted in the numerical example of Section 4.2.2, that the score test statistic was equal to 1.67 with a corresponding $p-$value of 0.20 when all observations were included in the analysis. When the 1st observation is removed, the value increases to 4.32 with accompanying $p-$value of 0.04. Thus the presence of the 1st is decreasing the value of the score test statistic. This is also depicted by a negative value of the $DIMS_k$. For this particular example, the 1st observation is very influential on the score test statistic. If this observation was not present in the analysis we would actually reject the null hypothesis of $H_0 : \theta_4 = 0$ on a 5 percent level of significance.

A positive value of the $DIMS_k$ means that the presence of the kth observation is increasing the value of the score test statistic. Observation 2 is thus contributing to the value of the score test statistic making it larger. If the 2nd observation were to be removed from the analysis, the value of the score test statistic decreases to 0.41.

The results of this numerical example will be further discussed in the next section, where we will assess the influence of multiple observations on the score test statistic.

6.2 Assessment of influence of multiple observations

In this section we present the obtained results concerning the assessment of influence from multiple observations on the score test statistic, given in (2.27). A diagnostic measure is proposed, which is a generalization of the measure $DIMS_k$, derived in the previous section, Section 6.1.2.

Let us assume that K is the subset containing the indices of the observations for which we would like to assess influence. In order to derive the measure $DIMS_K$, for assessing the influence of multiple observations on the score test statistic, consider the nonlinear regression model (2.2) and the same null hypothesis (6.13) as in Section 6.1.2, i.e. $H_0 : \theta_q = 0$.

Moreover, consider the perturbed nonlinear model, also given in (5.30),

$$\boldsymbol{y}_{\omega} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_{\omega}, \tag{6.23}$$

where $\boldsymbol{\varepsilon}_{\omega} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{W}^{-1}(\boldsymbol{\omega}))$, $\mathbf{W}(\boldsymbol{\omega}) : n \times n$ is a diagonal weight matrix, with diagonal elements $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^T$ and where $0 < \omega_k \leq 1$, for $k = 1, \ldots, n$.

Similar to the previous section, we utilize the score test statistic evaluated for the estimates from the perturbed model (6.23) and define the $DIMS_K$ as follows

Definition 6.2.1. *The $DIMS_K$, that measures the influence of the observations with indices in the subset K, on the score test statistic, is defined as the following derivative*

$$DIMS_K = \boldsymbol{\ell}^T \left. \frac{dS(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{1}_n},$$

where $\boldsymbol{\ell} : n \times 1$ is a vector with nonzero components in the rows with indices in K and where $S(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$ is the score test statistic evaluated for the estimates from the perturbed model (6.23) under the restriction $\theta_q = 0$.

With the same reasoning as in the previous section, replacing ω_k with $\boldsymbol{\omega}$, the score test statistic, evaluated for the parameter estimates from the perturbed nonlinear model (6.23), equals

$$S(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})) = \frac{1}{\widetilde{\sigma}^2(\boldsymbol{\omega})} \widetilde{\boldsymbol{e}}^T \mathbf{W} \boldsymbol{F}^T \left(\boldsymbol{F} \mathbf{W} \boldsymbol{F}^T \right)^{-1} \boldsymbol{F} \mathbf{W} \widetilde{\boldsymbol{e}}, \tag{6.24}$$

where $\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the estimate of $\boldsymbol{\theta}$ from the perturbed model (6.23) under the restriction that $\theta_q = 0$. Also, $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$, $\mathbf{W} = \mathbf{W}(\boldsymbol{\omega})$, $\widetilde{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$ and $\widetilde{\sigma}^2(\boldsymbol{\omega}) = \frac{1}{n} \left(\widetilde{\mathbf{e}}^T \mathbf{W} \widetilde{\mathbf{e}} \right)$.

The next theorem provides an explicit expression of the $DIMS_K$, which characterizes the influence of multiple observations on the score test statistic.

Theorem 6.2.1. *Let $DIMS_K$ be given in Definition 6.2.1. Then*

$$\begin{aligned} DIMS_K \quad = \quad & \frac{1}{\widetilde{\sigma}^2} \Bigg[2 \left(\boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \boldsymbol{F}^T \right) + DIM_{\widetilde{\boldsymbol{\theta}}} \mathbf{G} \left(\widetilde{\mathbf{r}} \otimes \mathbf{I}_q \right) \right) \mathbf{g} \\ & - \left(DIM_{\widetilde{\boldsymbol{\theta}}} \left(\mathbf{G} \left(\boldsymbol{F}^T \otimes \mathbf{I}_q \right) + \mathbf{G}^* \left(\mathbf{I}_q \otimes \boldsymbol{F}^T \right) \right) + \boldsymbol{U}^* \left(\boldsymbol{F}^T \otimes \boldsymbol{F}^T \right) \right) \\ & \times (\mathbf{g} \otimes \mathbf{g}) - \frac{S(\widetilde{\boldsymbol{\theta}})}{n} \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \widetilde{\boldsymbol{r}} \right) \Bigg], \end{aligned}$$

where $\widetilde{\mathbf{r}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}})$ *and* $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ *is defined in (6.15). The matrices* $\mathbf{G}$ *and* $\mathbf{G}^*$ *are defined in (6.18) and the q-vector* $\mathbf{g} = \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}$.

Moreover, $\boldsymbol{U}^* : n \times n^2$ *has row vectors* $\boldsymbol{u}_i^T$ *such that*

$$\boldsymbol{u}_i \quad = \quad \boldsymbol{d}_i \otimes \boldsymbol{d}_i \tag{6.25}$$

for $i = 1, \ldots, n$ *and* $\boldsymbol{d}_i$ *is the ith column of the identity matrix of size n and* $DIM_{\widetilde{\boldsymbol{\theta}}} : n \times q$ *is defined as*

$$DIM_{\widetilde{\boldsymbol{\theta}}} = \begin{pmatrix} DIM_{\widetilde{\boldsymbol{\theta}},1} \\ DIM_{\widetilde{\boldsymbol{\theta}},2} \\ \vdots \\ DIM_{\widetilde{\boldsymbol{\theta}},n} \end{pmatrix},$$

where $DIM_{\widetilde{\boldsymbol{\theta}},k}$ *is given in Definition 6.1.3. The last column of* $DIM_{\widetilde{\boldsymbol{\theta}}}$ *has all elements equal to zero.*

Proof. Let $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$, $\mathbf{W} = \mathbf{W}(\boldsymbol{\omega})$ and $\widetilde{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$. In (6.24), let

$$a(\boldsymbol{\omega}) = \widetilde{\mathbf{e}}^T \mathbf{W}\boldsymbol{F}^T \left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1} \boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}$$

and

$$b(\boldsymbol{\omega}) = \widetilde{\sigma}^2(\boldsymbol{\omega}).$$

When differentiating $S(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))$, the quotient rule is used, hence

$$\begin{aligned} \frac{dS(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}} &= \frac{a'(\boldsymbol{\omega})b(\boldsymbol{\omega}) - a(\boldsymbol{\omega})b'(\boldsymbol{\omega})}{b^2(\boldsymbol{\omega})} \\ &= \frac{a'(\boldsymbol{\omega}) - S(\widetilde{\boldsymbol{\theta}})b'(\boldsymbol{\omega})}{b(\boldsymbol{\omega})} \end{aligned} \tag{6.26}$$

where

$$a'(\boldsymbol{\omega}) = \frac{da(\boldsymbol{\omega})}{d\boldsymbol{\omega}}, \quad b'(\boldsymbol{\omega}) = \frac{db(\boldsymbol{\omega})}{d\boldsymbol{\omega}}.$$

First, the derivative of $a(\boldsymbol{\omega})$ is considered. Let $\mathbf{C} = \boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}$ and $\mathbf{D} = \boldsymbol{F}\mathbf{W}\boldsymbol{F}^T$, then

$$\begin{aligned} \frac{da(\boldsymbol{\omega})}{d\boldsymbol{\omega}} &= \frac{d\mathbf{C}^T}{d(\boldsymbol{\omega})}\mathbf{D}^{-1}\mathbf{C} + \frac{d\mathbf{D}^{-1}}{d(\boldsymbol{\omega})}(\mathbf{C} \otimes \mathbf{C}) + \frac{d\mathbf{C}}{d(\boldsymbol{\omega})}\mathbf{D}^{-1}\mathbf{C} \\ &= 2\left(\frac{d\mathbf{C}}{d(\boldsymbol{\omega})}\mathbf{D}^{-1}\mathbf{C}\right) + \frac{d\mathbf{D}^{-1}}{d(\boldsymbol{\omega})}(\mathbf{C} \otimes \mathbf{C}). \end{aligned}$$

The derivative of $\mathbf{C}$ with respect to $\boldsymbol{\omega}$ equals

$$\frac{d\mathbf{C}}{d\boldsymbol{\omega}} = \frac{d}{d\boldsymbol{\omega}}\boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}} = \frac{d\boldsymbol{F}}{d\boldsymbol{\omega}}\left(\mathbf{W}\widetilde{\mathbf{e}}\otimes\mathbf{I}_q\right) + \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\widetilde{\mathbf{e}}\otimes\mathbf{F}^T\right) + \frac{d\widetilde{\mathbf{e}}}{d\boldsymbol{\omega}}\mathbf{W}\boldsymbol{F}^T. \quad (6.27)$$

Applying the chain rule, defined in Appendix A, to (6.27) gives

$$\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\left(\mathbf{W}\widetilde{\mathbf{e}}\otimes\mathbf{I}_q\right) - \frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\mathbf{W}\boldsymbol{F}^T\right) + \boldsymbol{U}^*\left(\widetilde{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right), \quad (6.28)$$

since

$$\frac{d\mathbf{W}}{d\boldsymbol{\omega}} = \boldsymbol{U}^*,$$

where $\boldsymbol{U}^*$ is defined in (6.25).

Next, the derivative of $\mathbf{D}$ with respect to $\boldsymbol{\omega}$ is considered. Using the rule for differentiation of a matrix inverse (see Appendix A), the derivative of $\mathbf{D}^{-1}$ with respect to $\boldsymbol{\omega}$ is given by

$$\frac{d\mathbf{D}^{-1}}{d\boldsymbol{\omega}} = -\frac{d\mathbf{D}}{d\boldsymbol{\omega}}\left(\mathbf{D}^{-1}\otimes\mathbf{D}^{-1}\right).$$

Now,

$$\begin{aligned}
\frac{d\mathbf{D}}{d\boldsymbol{\omega}} &= \frac{d}{d\boldsymbol{\omega}}\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right) \\
&= \frac{d\boldsymbol{F}}{d\boldsymbol{\omega}}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right) + \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right) + \frac{d\boldsymbol{F}^T}{d\boldsymbol{\omega}}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right) \\
&= \frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right) + \frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right)\right) \\
&\quad + \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\boldsymbol{F}^T\otimes\mathbf{F}^T\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\mathbf{D}^{-1}}{d\boldsymbol{\omega}} = -\Bigg[&\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\left(\mathbf{W}\boldsymbol{F}^T\otimes\mathbf{I}_q\right) + \frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\left(\mathbf{I}_q\otimes\mathbf{W}\boldsymbol{F}^T\right)\right) \\
&+ \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\left(\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T\right)^{-1}\right)\Bigg]. \quad (6.29)
\end{aligned}$$

Now, $\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}=\mathbf{1}_n)=\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}=\mathbf{1}_n))=\widetilde{\boldsymbol{r}}$, $\boldsymbol{F}=\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$, $\mathbf{G}=\mathbf{G}(\widetilde{\boldsymbol{\theta}})$ and $\mathbf{G}^*=\mathbf{G}^*(\widetilde{\boldsymbol{\theta}})$. The derivatives in (6.28) and (6.29) evaluated at $\boldsymbol{\omega}=\mathbf{1}_n$ become

$$\left.\frac{d}{d\boldsymbol{\omega}}\boldsymbol{F}\mathbf{W}\widetilde{\mathbf{e}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}=\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\boldsymbol{F}^T\right)-\left.\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}\left(\boldsymbol{F}\boldsymbol{F}^T-\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right),$$

$$\begin{aligned}\left.\frac{d(\boldsymbol{F}\mathbf{W}\boldsymbol{F}^T)^{-1}}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} &= -\left[\left.\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\mathbf{F}^T\right)\right)\right.\\ &\quad\left.+\boldsymbol{U}^*\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\right]\times\left(\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\right).\end{aligned}$$

We know from the proof of Corollary 5.2.2 that

$$\left.\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}=DIM_{\widetilde{\boldsymbol{\theta}}}=\begin{pmatrix}DIM_{\widetilde{\boldsymbol{\theta}},1}\\ DIM_{\widetilde{\boldsymbol{\theta}},2}\\ \vdots\\ DIM_{\widetilde{\boldsymbol{\theta}},n}\end{pmatrix},$$

where $DIM_{\widetilde{\boldsymbol{\theta}},k}$ is given in Definition 6.1.3 and thus, the last column of $DIM_{\widetilde{\boldsymbol{\theta}}}$ has all elements equal to zero.

Now, let $\mathbf{g}=\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}$. Then,

$$\begin{aligned}\left.\frac{da(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} &= 2\left(\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\boldsymbol{F}^T\right)-DIM_{\widetilde{\boldsymbol{\theta}}}\left(\boldsymbol{F}\boldsymbol{F}^T-\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\right)\mathbf{g}\\ &\quad-\left[DIM_{\widetilde{\boldsymbol{\theta}}}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)\right.\\ &\quad\left.+\boldsymbol{U}^*\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\right](\mathbf{g}\otimes\mathbf{g})\\ &= 2\left(\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\boldsymbol{F}^T\right)+DIM_{\widetilde{\boldsymbol{\theta}}}\left(\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\mathbf{g}\right)\\ &\quad-\left[DIM_{\widetilde{\boldsymbol{\theta}}}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right)\right.\\ &\quad\left.+\boldsymbol{U}^*\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\right](\mathbf{g}\otimes\mathbf{g}).\end{aligned}\tag{6.30}$$

In the expression above, $DIM_{\widetilde{\boldsymbol{\theta}}}\boldsymbol{F}\boldsymbol{F}^T\mathbf{g}=\mathbf{0}_q$. This is due to the fact that the normal equations for estimating $\boldsymbol{\theta}$ under the restriction that $\theta_q=0$ is set to zero and that the last column of $DIM_{\widetilde{\boldsymbol{\theta}}}$ has all elements equal to zero. Thus, $DIM_{\widetilde{\boldsymbol{\theta}}}\boldsymbol{F}\boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}=DIM_{\widetilde{\boldsymbol{\theta}}}\boldsymbol{F}\widetilde{\mathbf{r}}=\mathbf{0}_q$.

Now the derivative of the variance term needs to be calculated. The maximum likelihood estimator of $\sigma^2(\boldsymbol{\omega})$ under the null hypothesis, $H_0 : \theta_q = 0$, equals

$$\widetilde{\sigma}^2(\boldsymbol{\omega}) = \frac{1}{n}\left(\widetilde{\mathbf{e}}^T \mathbf{W}\widetilde{\mathbf{e}}\right).$$

Using the product and the chain rule, see Appendix A, the derivative of $\widetilde{\sigma}^2(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ is the following

$$\begin{aligned}
\frac{d\widetilde{\sigma}^2(\boldsymbol{\omega})}{d\boldsymbol{\omega}} &= \frac{1}{n}\frac{d}{d\boldsymbol{\omega}}\widetilde{\mathbf{e}}^T\mathbf{W}\widetilde{\mathbf{e}} \\
&= \frac{1}{n}\left(2\frac{d\widetilde{\mathbf{e}}}{d\boldsymbol{\omega}}\mathbf{W}\widetilde{\mathbf{e}} + \frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&= \frac{1}{n}\left(\frac{d\mathbf{W}}{d\boldsymbol{\omega}}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right) - 2\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}}\mathbf{W}\widetilde{\mathbf{e}}\right) \\
&= \frac{1}{n}\left(\boldsymbol{U}^*\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&- \frac{1}{n}\left(2\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\mathbf{W}\widetilde{\mathbf{e}}\right)\right) \\
&= \frac{1}{n}\left(\boldsymbol{U}^*\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right) - 2\frac{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega})}\mathbf{W}\widetilde{\mathbf{e}}\right)\right).
\end{aligned}$$

Evaluated at $\boldsymbol{\omega} = \mathbf{1}_n$, $\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega} = \mathbf{1}_n) = \widetilde{\boldsymbol{\theta}}$, $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega} = \mathbf{1})) = \widetilde{\mathbf{r}}$ and hence

$$\begin{aligned}
\left.\frac{d\widetilde{\sigma}^2(\boldsymbol{\omega})}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} &= \frac{1}{n}\left(\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\widetilde{\boldsymbol{r}}\right) - 2DIM_{\widetilde{\theta}}\mathbf{F}\widetilde{\mathbf{r}}\right) \qquad (6.31)\\
&= \frac{1}{n}\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\widetilde{\boldsymbol{r}}\right),
\end{aligned}$$

since $DIM_{\widetilde{\theta}}\mathbf{F}\widetilde{\mathbf{r}} = \mathbf{0}_q$.

Now, inserting (6.30) and (6.31) in (6.26) we get

$$\begin{aligned}
&\left.\frac{dS(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n} = \frac{1}{\widetilde{\sigma}^2}\Bigg[2\left(\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\boldsymbol{F}^T\right) + DIM_{\widetilde{\boldsymbol{\theta}}}\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_q\right)\right)\mathbf{g} \\
&- \left(DIM_{\widetilde{\boldsymbol{\theta}}}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right) + \mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right) + \boldsymbol{U}^*\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\right) \qquad (6.32)\\
&\times\left(\mathbf{g}\otimes\mathbf{g}\right) - \frac{S(\widetilde{\boldsymbol{\theta}})}{n}\boldsymbol{U}^*\left(\widetilde{\boldsymbol{r}}\otimes\widetilde{\boldsymbol{r}}\right)\Bigg],
\end{aligned}$$

and $\boldsymbol{\ell}^T \left.\frac{dS(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n}$ equals the expression in Theorem 6.2.1.
This completes the proof.

■

Corollary 6.2.1. *The influence measure $DIMS_K$ is a linear combination of the influence measures $DIMS_k$, given in Definition 6.1.2, for all k contained in the subset K.*

Proof. Now, consider

$$\boldsymbol{\ell}^T \left.\frac{dS(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\omega}))}{d\boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\mathbf{1}_n},$$

where $\boldsymbol{\ell} : n \times 1$ is a vector with nonzero entries in the rows with indices in K.

Pre-multiplying (6.32) by $\boldsymbol{\ell}^T$, we need to consider the following terms

$$\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \boldsymbol{F}^T\right), \qquad \boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}}\mathbf{G}\left(\widetilde{\mathbf{r}} \otimes \mathbf{I}_q\right),$$

$$\boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}}\mathbf{G}\left(\boldsymbol{F}^T \otimes \mathbf{I}_q\right), \qquad \boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}}\mathbf{G}^*\left(\mathbf{I}_q \otimes \boldsymbol{F}^T\right),$$

$$\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{F}^T \otimes \boldsymbol{F}^T\right), \qquad \boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \widetilde{\boldsymbol{r}}\right).$$

Firstly, we evaluate the terms containing $\boldsymbol{\ell}^T \boldsymbol{U}^* : 1 \times n^2$. We have that

$$\boldsymbol{\ell}^T \boldsymbol{U}^* = \left(\begin{array}{c|c|c|c} \boldsymbol{d}_1^T & \boldsymbol{d}_2^T & \ldots & \boldsymbol{d}_n^T \end{array} \right),$$

where $\boldsymbol{d}_i$ is the ith column of the identity matrix of size n, for all i contained in the subset K and $\boldsymbol{d}_i = \mathbf{0}_n$ for all i not contained in K.

From this, it follows that

$$\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \boldsymbol{F}^T\right) = \sum_{i \in K} \ell_i \widetilde{r}_i \boldsymbol{F}_i^T, \tag{6.33}$$

$$\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{F}^T \otimes \boldsymbol{F}^T\right) = \sum_{i \in K} \ell_i \boldsymbol{F}_i \boldsymbol{F}_i^T, \tag{6.34}$$

$$\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \widetilde{\boldsymbol{r}}\right) = \sum_{i \in K} \ell_i \widetilde{r}_i. \tag{6.35}$$

Secondly, we observe that

$$\boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}} = \boldsymbol{\ell}^T \begin{pmatrix} DIM_{\widetilde{\boldsymbol{\theta}},1} \\ DIM_{\widetilde{\boldsymbol{\theta}},2} \\ \vdots \\ DIM_{\widetilde{\boldsymbol{\theta}},n} \end{pmatrix} = \sum_{i \in K} \ell_i DIM_{\widetilde{\boldsymbol{\theta}},i}, \tag{6.36}$$

where $DIM_{\widetilde{\boldsymbol{\theta}},i} : 1 \times q$ is given in Definition 6.1.3.

Assuming, without loss of generality, that $K = \{k\}$, then we have that

$$\begin{aligned}
\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \boldsymbol{F}^T\right) &= \ell_k \widetilde{r}_k \boldsymbol{F}_k^T, \\
\boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}} \mathbf{G} \left(\widetilde{\mathbf{r}} \otimes \mathbf{I}_q\right) &= \ell_k DIM_{\widetilde{\boldsymbol{\theta}},k} \boldsymbol{G} \left(\widetilde{\mathbf{r}} \otimes \mathbf{I}_q\right), \\
\boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}} \mathbf{G} \left(\boldsymbol{F}^T \otimes \mathbf{I}_q\right) &= \ell_k DIM_{\widetilde{\boldsymbol{\theta}},k} \boldsymbol{G} \left(\boldsymbol{F}^T \otimes \mathbf{I}_q\right), \\
\boldsymbol{\ell}^T DIM_{\widetilde{\boldsymbol{\theta}}} \mathbf{G}^* \left(\mathbf{I}_q \otimes \boldsymbol{F}^T\right) &= \ell_k DIM_{\widetilde{\boldsymbol{\theta}},k} \boldsymbol{G}^* \left(\mathbf{I}_q \otimes \boldsymbol{F}^T\right), \\
\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\boldsymbol{F}^T \otimes \boldsymbol{F}^T\right) &= \ell_k \boldsymbol{F}_k \boldsymbol{F}_k^T, \\
\boldsymbol{\ell}^T \boldsymbol{U}^* \left(\widetilde{\boldsymbol{r}} \otimes \widetilde{\boldsymbol{r}}\right) &= \ell_k \widetilde{r}_k^2.
\end{aligned}$$

Inserting the equalities above in the expression for $DIMS_K$ we arrive at

$$\begin{aligned}
\boldsymbol{\ell}^T DIMS_K = \frac{1}{\widetilde{\sigma}^2} \Bigg[& 2 \left(\ell_k \widetilde{r}_k \boldsymbol{F}_k^T \mathbf{g} + \ell_k DIM_{\widetilde{\boldsymbol{\theta}},k} \boldsymbol{G} \left(\widetilde{\mathbf{r}} \otimes \mathbf{I}_q\right) \mathbf{g} \right) \\
& - \left(\ell_k DIM_{\widetilde{\boldsymbol{\theta}},k} \left(\boldsymbol{G} \left(\boldsymbol{F}^T \otimes \mathbf{I}_q\right) + \boldsymbol{G}^* \left(\mathbf{I}_q \otimes \boldsymbol{F}^T\right) \right) + \ell_k \boldsymbol{F}_k \boldsymbol{F}_k^T \right) \\
& \times (\mathbf{g} \otimes \mathbf{g}) - \frac{S}{n} \ell_k \widetilde{r}_k^2 \Bigg].
\end{aligned}$$

which, if $\ell_k = 1$, equals the expression in Theorem 6.1.2. Since the equalities in (6.33)-(6.36) are sums over all observations contained in the subset K we observe that, for a general subset K, the expression of $DIMS_K$ is a linear combination of the $DIMS_k$, given in Definition 6.1.2, for all k contained in the subset K. This completes the proof.

■

From Corollary 6.2.1 we see that the $DIMS_K$ is a linear combination of the $DIMS_k$, the influence measure used to assess the influence of single observations on the score test statistic, for all $k \in K$. Thus, in order to assess the joint

influence of multiple observations on the score test statistic we only need to consider the measure given in Theorem 6.1.2.

If observations with equal signs of the values of $DIMS_k$ will be considered together, the joint influence that these observations exercise on the score test statistic will be more extensive than when they are considered separately. For instance, two observations with positive influence on the score test statistic will result in a much larger positive influence when they are considered jointly. Also, two observations with a negative influence on the score test statistic will result in a larger negative influence when they are considered jointly. It is important to remember that two observations with unequal signs of the $DIMS_k$ will even out the value of the $DIMS_K$, resulting in a value more close to zero, i.e. no joint influence.

In the next section, a numerical example illustrates how multiple observations can influence the score test statistic.

6.2.1 Numerical example

We will continue with the numerical example given in Section 6.1.3 and we will assess the influence of multiple observations on the score test statistic using the diagnostic $DIMS_K$. We know that the $DIMS_K$ is a linear combination of the diagnostic $DIMS_k$, therefore we only need to consider Figure 6.1 in Section 6.1.3 to assess the joint influence of the observations.

In Figure 6.1 we can see that the 1st and 10th observations are the observations with the largest negative influence on the score test statistic. Mutually, they exercise quite a large negative influence on the score test statistic. The value of $DIMS_1$ is -1.68 and the value of $DIMS_{10}$ is -0.72; hence, the value of $DIMS_K = -2.40$ when $K = \{1, 10\}$. The presence of the 1st and 10th observations are decreasing the value of the score test statistic. If both these observations are removed from the analysis the score test statistic equals 5.82 with a corresponding $p-$value 0.02.

If we look at the scenario when the 1st and 2nd observation is considered jointly we can expect that the joint influence will even out since these observations have unequal signs of the $DIMS_k$. The value of the influence measure corresponding to the 2nd observation is $DIMS_2 = 1.32$ and the resulting $DIMS_K = -0.37$ when $K = \{1, 2\}$. Separately, these two observations exercise quite a large influence on the score test statistic, but when considered mutually the joint influence is almost zero. The result of the testing procedure would not

change dramatically if these two observations were removed from the analysis. In fact, the score test statistic equals 2.58 with a corresponding $p-$value of 0.11. This is a small increase from 1.67, i.e. the value of the score test statistic when all observations are present in the analysis.

7. Concluding remarks and further research

It is well known that not all observations play an equal role in determining the various results from a regression analysis. For instance, the character of the regression line may be determined by only a few observations, while most of the data is somewhat ignored. Such observations that highly influence the results of the analysis are called influential observations. It is beneficial, for many reasons, to be able to detect influential observations, see Chapter 3. For the linear regression model there is a vast collection of diagnostic tools to use for identifying influential observations. The amount of literature and research on influence analysis for nonlinear regression models is not as extensive as in the linear regression case. With this dissertation we want to make a contribution to influence analysis concerning various results of the nonlinear regression analysis. In particular, we focus on the task of identifying observations with substantial influence on the parameter estimates of a nonlinear regression model and on the score test statistic, when testing a hypothesis that a specific parameter in the nonlinear regression model equals zero.

The main contributions of this thesis are as follows:

- Two different diagnostic measures for assessing the influence of single observations on the parameter estimates in the nonlinear regression model (2.2) are proposed. The first measure, $DIM_{\widehat{\boldsymbol{\theta}},k}$, is to be used when we are interested in assessing the influence of an observation on the whole vector of parameter estimates. The explicit expression of this measure is given in Theorem 5.1.2. The second measure, for assessing the influence of an observation on a specific parameter estimate is denoted $DIM_{\widehat{\theta}_j,k}$ and the explicit expression of the measure is given in Theorem 5.1.3 (*Aim 1*).

- We extend the ideas and techniques used to assess the influence of single observations, to multiple observations, on the parameter estimates in the nonlinear regression model (2.2). In correspondence with the first contribution, we present two measures: One measure for assessing the

influence of multiple observations on the whole vector of parameter estimates, $DIM_{\widehat{\boldsymbol{\theta}},K}$, and one measure for assessing the influence of multiple observations on a specific parameter estimate, $DIM_{\widehat{\theta}_j,K}$. These measures are presented in Theorem 5.2.1 and 5.2.2, respectively. The influence that multiple observations exercise on the parameter estimates in this case are referred to as joint influence, since we consider the observations simultaneously when assessing the influence (*Aim 2*).

- As opposed to joint influence, multiple observations can exercise what we refer to as conditional influence on the parameter estimates. Conditional influence arises if an observation is not identified as influential unless another observation is deleted first. Thus, an influence measure for assessing the influence of the kth observation, given that the ith observation is deleted, is proposed. The measure is denoted $DIM_{\widehat{\boldsymbol{\theta}}_{(i)},k}$ and its explicit expression is given in Theorem 5.2.4 (*Aim 3*).

- We develop a graphical tool for visually identifying observations that are influential on the score test statistic, when testing the null hypothesis of $H_0 : \theta_q = 0$, where θ_q is a parameter in the nonlinear regression model (2.2). This graphical tool is referred to as the added parameter plot and it is presented in Definition 4.2.1 (*Aim 4*).

- The added parameter plot is for explorative purposes only. In order to quantify the influence of the observations on the score test statistic we propose two influence measures. The first measure is to be used when assessing the influence of a single observation on the score test statistic, denoted $DIMS_k$. The explicit expression of this measure is given in Theorem 6.1.2. Moreover, we propose a measure for assessing the influence of multiple observations, jointly, on the score test statistic, denoted $DIMS_K$, presented in Theorem 6.2.1 (*Aim 5*).

In general, we are proud to propose our new measures and diagnostic tools, since they add to the research of influence analysis in nonlinear regression. With this thesis, we give practitioners more approaches to chose from when conducting influence analysis, and hence more flexibility. However, we want to highlight some of the contributions that we feel particularly strong about: Firstly, the use of the proposed marginal influence measures, $DIM_{\widehat{\theta}_j,k}$ and $DIM_{\widehat{\theta}_j,K}$, provides the opportunity to assess the influence of observations on a specific parameter estimate. There exist diagnostic measures for assessing the influence of observations on a specific parameter estimate in the linear regression model, but it has not yet been done for parameter estimates in the nonlinear regression model. Secondly, estimating the parameters in nonlinear

regression models are complicated since there is generally no closed form of the estimators. Instead, iterative methods must be used to find the estimates. To make a comparison, consider adopting the case-deletion approach for assessing the influence of observations on the parameter estimates (or other statistics which are functions of the parameter estimates). With this method we need to, iteratively, find the estimates for each observation that is deleted. This can become an overwhelming task. Our proposed approach to influence analysis reduces the burden with additional iterations, since we only need to find the estimates of the parameters once. Moreover, after scrutinizing the literature on influence analysis in nonlinear regression, we have not yet seen any research results on how one can identify observations that are influential on the outcome of a hypothesis testing procedure. With the proposed results, the added parameter plot and the diagnostic measures $DIMS_k$ and $DIMS_K$, we give another view of influence analysis in nonlinear regression, since most research is focused on the parameter estimates.

Of course, we are only able to cover a fraction of all there is to discover in the area of influence analysis in nonlinear regression. There are still many things one can do to extend the work done in this thesis. The following are three examples of directions for future work:

- In this thesis we do not discuss what constitutes a substantially influential observation. We rather put the results of the computed influence measures in relation to each other and rely on the judgment of the researcher or practitioner. A further task could be to develop cut-offs, or thresholds, that determine when an observation is substantially, or significantly, influential. One idea is to use the bootstrap method to accomplish this.

- We are aware that Rao's score test (see Chapter 2) is asymptotically χ^2-distributed, so that larger samples are needed in order to get reliable p-values. An intriguing task would be to examine how the influence analysis, and the use of the proposed methods, are affected as the sample size grows.

- Since nonlinear regression models can differ greatly, a future task could be to customize the results obtained in this thesis to a specific nonlinear regression model, such as the Michaelis-Menten model.

Appendix A

Matrix derivative

In this section rules for matrix differentiation are presented.

Definition. *Let the elements of* $\mathbf{Y} \in \mathbb{R}^{r\times s}$ *be functions of* $\mathbf{X} \in \mathbb{R}^{p\times q}$*. The matrix* $\frac{d\mathbf{Y}}{d\mathbf{X}} \in \mathbb{R}^{pq\times rs}$ *is called matrix derivative of* $\mathbf{Y}$ *by* $\mathbf{X}$ *in a set A, if the partial derivative* $\frac{dy_{kl}}{dx_{ij}}$ *exist, are continues in A and*

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d}{d\mathbf{X}} vec^T \mathbf{Y},$$

where

$$\frac{d}{d\mathbf{X}} = \left(\frac{d}{dx_{11}}, \ldots, \frac{d}{dx_{p1}}, \frac{d}{dx_{12}}, \ldots, \frac{d}{dx_{p2}}, \ldots, \frac{d}{dx_{1q}}, \ldots, \frac{d}{dx_{pq}} \right)^T.$$

Properties of the matrix derivative in the definition is presented in the following table, where $\mathbf{Z} : s \times t$ is a function of $\mathbf{Y}$, and where $\mathbf{A}$ and $\mathbf{B}$ are matrices of constants and of proper size.

Differentiated function	Derivative
$\mathbf{Z} = \mathbf{Z}(\mathbf{Y}),\ \mathbf{Y} = \mathbf{Y}(\mathbf{X})$	$\frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}} \frac{d\mathbf{Z}}{d\mathbf{Y}}$
$\mathbf{Y} = \mathbf{AXB}$	$\frac{d\mathbf{Y}}{d\mathbf{X}} = \mathbf{B} \otimes \mathbf{A}^T$
$\mathbf{Z} = \mathbf{AYB}$	$\frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}} \left(\mathbf{B} \otimes \mathbf{A}^T \right)$
$\mathbf{W} = \mathbf{YZ}$	$\frac{d\mathbf{W}}{d\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}} \left(\mathbf{Z} \otimes \mathbf{I}_r \right) + \frac{d\mathbf{Z}}{d\mathbf{X}} \left(\mathbf{I}_t \otimes \mathbf{Y}^T \right)$
$\mathbf{W} = \mathbf{RYZ},\ \mathbf{R} \in \mathbb{R}^{p\times r}$	$\frac{d\mathbf{W}}{d\mathbf{X}} = \frac{d\mathbf{R}}{d\mathbf{X}} \left(\mathbf{YZ} \otimes \mathbf{I}_p \right) + \frac{d\mathbf{Y}}{d\mathbf{X}} \left(\mathbf{Z} \otimes \mathbf{R}^T \right) + \frac{d\mathbf{Z}}{d\mathbf{X}} \left(\mathbf{I}_t \otimes (\mathbf{RY})^T \right)$
$\mathbf{Y}^{-1}$	$\frac{d\mathbf{Y}^{-1}}{d\mathbf{X}} = -\frac{d\mathbf{Y}}{d\mathbf{X}} \left(\mathbf{Y}^{-1} \otimes \mathbf{Y}^{-1} \right)$

Sammanfattning

Alla observationer är inte lika viktiga för resultaten från en regressions analys. I de mest extrema fall kan en eller två observationer helt bestämma, till exempel, värdet på parameterskattningarna, medan resten av datat till stora delar ignoreras. Sådana observationer, som har stort inflytande på inferensen, kallas inflytelserika och att kunna identifiera inflytelserika observationer är av stor vikt.

Avhandlingen erbjuder metoder för att identifiera inflytelserika observationer då man arbetar med en icke-linjär regressionsmodell. Detta uppnås genom att mått konstrueras, vilka mäter inflytandet av en eller flera observationer på parameterskattningarna. Metoden som används för att konstruera dessa mått är hämtad från influensanalys inom linjär regression och kallas deriveringsmetoden.

Hypotesprövning är en viktig del av den statistiska inferensen och även resultatet från en hypotesprövning kan till stor del påverkas av en eller flera betydelsefulla observationer. En intressant aspekt av influensanalys är därför hur de individuella observationerna påverkar resultatet från en hypotesprövning. I avhandlingen ges flera metoder för att identifiera inflytelserika observationer på teststatistikan, då score-testet används med nollhypotesen att en parameter i den icke-linjär regressionsmodellen är lika med noll. Med hjälp av deriveringsmetoden konstruerar vi mått som mäter inflytandet av en eller flera observationer, och på så sätt kan inflyteserika observationer identifieras. Utöver detta konstruerar vi ett grafiskt hjälpmedel som kan användas för att visuellt identifiera observationer som har stort inflytande på teststatistikan för score-testet.

References

[1] Alfons, A., Croux, C. & Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*, **7**, 226-248.

[2] Andrews, D.F. & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society. Series B*, **40**, 85-93.

[3] Atkins, G.L. & Nimmo, I.A. (1975). A comparison of seven methods for fitting the Michaelis-Menten equation. *Biochemical Journal*, **149**, 775-777.

[4] Atkinson, A.C. (1982). Regression diagnostics, transformations and constructed variables. *Journal of the Royal Statistical Society. Series B*, **44**, 1-36.

[5] Atkinson, A.C. (1985). *Plots, Transformations and Regression*. Clarendon, Oxford.

[6] Atkinson, A.C. (1986). Masking unmasked. *Biometrika*, **73**, 533-541.

[7] Barnes, T.J. (1998). The history of regression: actors, networks, machines and numbers. *Environment and Planning A*, **30**, 203-223.

[8] Bates, D.M. & Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley, New Jersey.

[9] Behnken, D.W. & Draper, N.R. (1972). Residuals and their variance patterns. *Technometrics*, **14**, 101-111.

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

[11] Beyaztas, U. & Alin, A. (2014). Sufficient jackknife-after-bootstrap method for detection of influential observations in linear regression models. *Statistical Papers*, **55**, 1001-1018.

[12] Briggs, G.E. & Haldane, J.B.S. (1925). A note on the kinetics of enzyme action. *Biochemical Journal*, **19**, 338-339.

[13] Bulmer, M. (2003). *Francis Galton: Pioneer of Heredity and Biometry*. John Hopkins University Press, Baltimore.

[14] Chakraborty, B., Bhattacharya, S., Basu, A., Bandyopadhyay, S. & Bhattacharjee, A. (2014). Goodness-of-fit testing for the Gompertz growth curve model. *Metron*, **72**, 45-64.

[15] Chatterjee, S. & Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, **1**, 379-393.

[16] Chatterjee, S. & Hadi, A.S. (1988). *Sensitivity Analysis in Regression*. Wiley, New York.

[17] Chen, C-F. (1983). Score tests for regression models. *Journal of the American Statistical Association*, **78**, 158-161.

[18] Chen, C-F. (1985). Robustness aspects of score tests for generalized linear and partially linear regression models. *Technometrics*, **27**, 277-283.

[19] Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 16-18.

[20] Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B*, **48**, 133-169.

[21] Cook, R.D. (1987). Parameter Plots in Nonlinear Regression. *Biometrika*, **74**, 669-677.

[22] Cook, R.D. (1998). *Regression Graphics*. Wiley, New York.

[23] Cook, R.D. & Weisberg, S. (1980). Characterizations of an emprical influence function for detecting influential cases in regression. *Technometrics*, **22**, 495-508.

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

[25] Dette, H. & Kunert, J. (2014). Optimal designs for the Michaelis-Menten model with correlated observations. *Statistics: A Jounal of Theoretical and Applied Statistics*, **48**, 1254-1267.

[26] Ezekiel, M. (1924). A method for handling curvilinear correlation for any number of variables. *Journal of the American Statistical Association*, **19**, 431-453.

[27] Galea, M., Paula, G.A. & Cysneiros, F.J.A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters*, **73**, 459-467.

[28] Gallant, A.R. (1987). *Nonlinear Statistical Models*. Wiley, New York.

[29] Gut, A. (1995). *An intermediate Course in Probability*. Springer, New York.

[30] Hadi, A.S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistics and Data Analysis*, **14**, 1-27.

[31] Hamilton, D. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika*, **73**, 57-64.

[32] Hamilton, D. & Wiens, D. (1987). Correction factors for F ratios in nonlinear regression. *Biometrika*, **74**, 423-425.

[33] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383-393.

[34] Hoaglin, D.C. & Welsch, R.E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, **32**, 17-22.

[35] Huber, P.J. (1972). The 1972 Wald lecture robust statistics: A review. *The Annals of Mathematical Statics*, **43**, 1041-1067.

[36] Johnson, B.W. & McCulloch, R.E. (1987). Added-Variable plots in linear regression. *Technometrics*, **29**, 427-433.

[37] Kollo, T. & von Rosen, D. (2010). *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht.

[38] Lawrence, A.J. (1995). Deletion influence and masking in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 181-189.

[39] Lee, A.H., Xiang, L. & Fung, W.K. (2004). Sensitivity of score tests for zero-inflation in count data. *Statistics in Medicine*, **23**, 2757-2769.

[40] Lemonte, A.J. & Patriota, A.G. (2011). Influence diagnostics in Birnbaum-Saunders nonlinear regression models. *Journal of Applied Statistics*, **38**, 871-884.

[41] Li, B. (2001). Sensitivity of Rao's score test, the Wald test and the likelihood ratio test to nuisance parameters. *Journal of Statistical Planning and Inference*, **97**, 57-66.

[42] Lustbader, E.D. & Moolgavkar, S.H. (1985). A diagnostic statistic for the score test. *Journal of the American Statistical Association*, **80**, 375-379.

[43] Markatou, M. & Manos, G. (1996). Robust tests in nonlinear regression models. *Journal of Statistical Planning and Inference*, **55**, 205-217.

[44] Michaelis, L. & Menten, M.L. (1913). Die kinetik der invertinwirkung. *Biochemische Zeitschrift*, **49**, 333-369.

[45] Mosteller, F. & Tukey, J.W. (1977). *Data Analysis and Linear Regression*. Addison-Wesley, Reading.

[46] Neyman, J. & Pearson, E.S. (1928). On the use and interpretation of certain test criteria. *Biometrika*, **20A**, 175-240, 263-294.

[47] Nocedal, J. & Wright, S.J. (2006). *Numerical Optimization*. Springer, New York.

[48] Nurunnabi, A.A.M., Hadi, A.S. & Imon, A.H.M.R. (2014). Procedures for the identification of multiple influential observations in linear regression. *Journal of Applied Statistics*, **41**, 1315-1331.

[49] Park, H., Sakaori, F. & Konishi, S. (2014). Robust sparse regression and tuning parameter selection via the efficient bootstrap information criteria. *Journal of Statistical Computation and Simulation*, **84**, 1596-1607.

[50] Pasaribu, U.S. (1999). Statistical assumptions underlying the fitting of the Michaelis-Menten equation. *Journal of Applied Statistics*, **26**, 327-341.

[51] Peña, D. & Yohai, V.J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society. Series B*, **57**, 145-156.

[52] Poon, W. & Poon, Y.S. (2001). Conditional local influence in case-weights linear regression. *British Journal of Mathematical and Statistical Psychology*, **54**, 177-191.

[53] Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, 50-57.

[54] Ritchie, R.J. & Prvan, T. (1996). Current statistical methods for estimating the K_m and V_{max} of Michaelis-Menten kinetics. *Biochemical Education*, **24**, 196-206.

[55] Ross, W.H. (1987). The Geometry of case deletion and the assessment of influence in nonlinear regression. *The Canadian Journal of Statistics*, **15**, 91-103.

[56] Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

[57] Seber, G.A.F. & Wild, C.J. (2003). *Nonlinear Regression*. Wiley, New Jersey.

[58] Sen, A. & Srivastava, M. (1990). *Regression Analysis: Theory, Methods, and Applications*. Springer, New York.

[59] Stanley, W. & Miller, M. (1979). Measuring technological change in jet fighter aircraft. Report No. R-2249-AF, Rand Corp., Santa Monica, CA.

[60] Stiegler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press, Cambridge.

[61] Stiegler, S.M. (1989). Francis Galton's account of the invention of correlation. *Statistical Science*, **4**, 73-79.

[62] St. Laurent, R.T. & Cook, R.D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, **87**, 985-990.

[63] St. Laurent, R.T. & Cook, R.D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika*, **80**, 99-106.

[64] Vanegas, L.H. & Cysneiros, F.J.A. (2010). Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics and Data Analysis*, **54**, 1002-1016.

[65] Vanegas, L.H., Rondón, L.M. & Cysneiros, F.J.A. (2012). Diagnostic procedures in Birnbaum-Saunders nonlinear regression models. *Computational Statistics and Data Analysis*, **56**, 1662-1680.

[66] Vanegas, L.H., Rondón, L.M. & Cysneiros, F.J.A. (2013). Assessing robustness of inference in symmetrical nonlinear regression models. *Comunication in Statistics - Theory and Methods*, **42**, 1692-1711.

[67] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the Americal Matematical Society*, **54**, 426-482.

[68] Wang, D.Q. & Critchley, F. (2000). Multiple deletion measures and conditional influence in regression model. *Communication in Statistics - Theory and Methods*, **29**, 2391-2404.

[69] Zwietering, M.H., Jongenburger, I., Rombouts, F.M. & van't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, **56**, 1875-1881.