Research Report

Department of Statistics

Stockholms universitet

No. 2013:4

Marginal and Joint Influence in Nonlinear Regression Analysis

Karin Stål

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Marginal and Joint Influence in Nonlinear Regression Analysis

Karin Stål

Abstract

To study the influence of observations in nonlinear regression models, joint and marginal influence measures are proposed. These measures allow assessing the influence of observations on estimates of all parameters simultaneously and on an estimate of a particular parameter of interest, respectively. Moreover, a joint influence measure is provided for the score test statistic.

Keywords: Empirical influence curve, influential observation, score test statistic.

1. Introduction

Regression analysis explores the relationships between variables. The parameters in the regression model are estimated but some observations have a greater impact on the estimates than others. Observations that have a substantial effect on the obtained results are called influential. The study of the effect they have on conclusions and inferences is called influence analysis, or sensitivity analysis. The basic idea in influence analysis is to perform small perturbations in the problem formulation and it is studied how these perturbations change the result of the analysis.

Detection of influential observations is an important part of the statistical paradigm. The ability to find influential observations can benefit the analyst in several ways. For instance, it can uncover technical errors in the model, identify critical regions in the space of the input variables, simplify models and defend against falsifications of the analysis. If small changes in the data has a large effect on the analysis it is a warning that some observations might compromise the reliability of the conclusions and the inference. These observations should be given more attention.

There is a large number of measures used to quantify the influence of the observations. Some of the measures are based on the idea of the Influence Curve, IC, introduced in Hampel (1974). The IC is a theoretical concept. It measures how a statistic, expressed as a functional of a distribution function, changes as the distribution is perturbed. Approximations of the IC for finite samples are used in empirical studies. In this article one finite sample version of the IC, the Empirical Influence Curve, EIC, is of particular interest and discussed in the context of nonlinear regression estimation and hypothesis testing. In the setting of this article, the EIC will be defined so that it measures the change after some perturbation has taken place.

For linear regression models Cook's distance, proposed by Cook (1977), is the most well-established method for assessing the influence of individual observations on the estimated regression coefficients of fitted values. Cook's distance is related to a sample version of the IC as it is a normalized version of the sample influence curve, SIC. Other examples of influence measures are given in Chatterjee and Hadi (1988) and Cook and Weisberg (1982) including explicit expressions of the EIC and the SIC of the parameter estimates in linear regression. The previous mentioned authors also demonstrate that the EIC and the SIC for parameter estimates in linear regression can be constructed without referring to an IC through the differentiation approach.

A number of different methods have been proposed to assess influence for models that are nonlinear. A discussion of influence on the parameter estimates in the nonlinear regression model is given by Cook and Weisberg (1982). One of their suggestions is to use an elliptical norm of the SIC. Castillo *et al.* (2008) deals with the problem of local sensitivity analysis for estimation methods expressed as a nonlinear programming problem. Laurent and Cook (1993) discuss the connection between leverage and local influence for nonlinear regression models with an additive error term. High-leverage points are those that are isolated from the rest of the points in the $X-$space. For more discussion of local influence, see Cook (1986). Tang *et al.* (2002) propose a robust influence measure for detection of influential observations for nonlinear reproductive dispersion models.

As for parameter estimates, test statistics for a hypothesis can also be more or less sensitive to individual observations in the data set. In fact, the decision to reject, or not reject, the null hypothesis can be entirely based on the influence of a single observation. It is of interest to study this sensitivity and to gather more information to help to decide whether the null hypothesis should be rejected or not.

In this article the aim is to derive explicit expressions for the EIC of the parameter estimates and the score test statistic in the nonlinear regression case. Ideas are borrowed from linear regression and influence diagnostics are constructed without referring to an IC. The EIC can be used to quantify the influence the observations have on the parameter estimates and the score test statistic. This makes a comparison of the observations in terms of influence possible and influential observations can be given more attention. The EIC is appealing as an influence measure when using the differentiation method since the definition of the EIC is intuitively easy to understand and its concept is easily transferable to the parameter estimates and the score test statistic for a nonlinear regression model.

The score test is of particular interest when dealing with nonlinear models since inferences about the parameters must be done by relying on large sample theory. There are three large sample tests that are asymptotically equivalent under the null hypothesis, these are the score test, the Wald test and the likelihood ratio test, LRT. One can read more about the score test, the Wald test and the LRT in Rao (1973) and Lehmann and Romano (2005). Even if the tests are asymptotically equivalent under the null hypothesis, they do differ in convenience. The score test is beneficial over the other two tests since it does not require calculations of the parameter estimate under the alternative hypothesis. Also, the score test is locally most powerful, see e.g. Lehmann and Romano (2005). Another important advantage of the score test is that it has a graphical representation, the added parameter plot, APP. This plot makes it possible to visually identify observations that influence the score test, see Stål (2011).

Moreover, work on measures of influence on the score test statistic has been done by Wei (1995). He discusses the local influence of the score test statistic in nonlinear regression. Wei proposes two diagnostic statistics that measure the maximum local influence of the score test statistic, using two different perturbation schemes. The results presented in this article are different from Wei's since the two perturbation schemes are here merged into one. Also, Lustbader and Moolgavkar (1985), derive an exact expression for the change in the score test when deleting observations. This expression is computationally expensive, therefore they also suggest an approximation that is computationally simple. Deletions of individuals and of entire risk sets from matched case-control and survival studies are discussed in detail. Further, Lee *et al.* (2004) derive a diagnostic measure to assess sensitivity of score tests for zero-inflation in count data.

The outline of this article is as follows. In Section 2 the nonlinear model and the notation are presented. In Section 3 and 4 the influence diagnostics for the parameter estimates and the score test statistic, for the nonlinear model, are derived. The last section contains concluding remarks.

2. The nonlinear regression model

Nonlinear regression is an important statistical tool for investigating complex relationships among variables. In nonlinear models at least one of the parameters appear nonlinearly. To read more about nonlinear regression models see e.g. Gallant (1987), Seber and Wild (1989) and Pázman (1993).

The regression model considered in this article is a nonlinear model with an additive error term

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

where

$$\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) = \left(f(\boldsymbol{X}_1, \boldsymbol{\theta}), \ldots, f(\boldsymbol{X}_n, \boldsymbol{\theta})\right)^T = \left(f_1(\boldsymbol{\theta}), \ldots, f_n(\boldsymbol{\theta})\right)^T,$$

and where $\boldsymbol{X}$ is a $n \times p-$matrix of known explanatory variables, $\boldsymbol{y}$ is the n-vector of responses, $\boldsymbol{\theta}$ is a q-vector of unknown parameters. The vector of random errors $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n)$, where $\boldsymbol{0}_n$ and $\boldsymbol{I}_n$ denote the n-vector with all elements equal to zero and the identity matrix of size n, respectively. The function f is assumed to be twice differentiable in $\boldsymbol{\theta}$.

A well known example of a nonlinear model is the Michaelis-Menten model

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

which is used in enzyme kinetics. It relates the initial velocity of an enzymatic reaction, y, to the substrate concentration, x. The parameter θ_1 is the maximum velocity of the enzymatic reaction, representing the asymptotic value of y as $x \to \infty$; θ_2 is the parameter related to the half-velocity, representing the value of x when the velocity of the reaction reaches one-half of its ultimate value. When model (2) is used in enzyme kinetics, θ_1 is referred to as $V_{\max}$ and θ_2 is designated the Michaelis-Menten constant, K_M. For an examination of the theoretical basis of the Michaelis-Menten equation see Briggs and Haldane (1925). Interested readers may also see Cleland *et al.* (1967), for a thorough review of the statistical analysis of enzyme kinetic data or Ritchie and Prvan (1996) for current statistical methods for estimating the parameters of the Michaelis-Menten equation.

Estimation of the parameters of the nonlinear regression model is usually carried out by the method of least squares, LS, or the method of maximum likelihood, ML. Unlike in linear regression, it is usually not possible to find explicit expressions for the LS and ML estimators for nonlinear regression models. Instead some iterative method is required, such as the Gauss-Newton method or the Levenberg-Marquardt method. See for instance Bates and Watts (1988) or Nocedal and Wright (2006) for a detailed description of these and other methods for solving optimization problems.

3. Influence diagnostics in nonlinear regression

It is well understood that not all observations in a data set play an equal role in determining estimates, tests, and other statistics. For instance, in regression analysis the character of the regression 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. The study of the effect they have on conclusions and inferences is called influence analysis or sensitivity analysis. To read more about influence analysis, see e.g. Saltelli *et al.* (2008). They give a thorough introduction to sensitivity analysis in general and then focus on global influence instead of local influence. They propose different methods for conducting sensitivity analysis, such as the variance-based method and the elementary effects method.

A popular approach used to construct influence diagnostics is through a perturbation scheme where weights are attached to individuals. Mostly, the case-weights are restricted to be either 0 or 1 so that a case is either deleted from the analysis or retained in the analysis with full weight. The scheme where the case-weight is 0 is called the case-deletion scheme. The approach used to construct influence diagnostics with the case-weight scheme is by Chatterjee and Hadi (1988) referred to as the omission approach.

One diagnostic for the parameter estimates in linear regression constructed with the omission approach is the EIC. As mentioned before, the EIC is an empirical version of the influence curve, IC, introduced by Hampel (1974). For definitions of the IC and the EIC see e.g. Cook and Weisberg (1982) and Chatterjee and Hadi (1988). The previously mentioned authors give an alternative derivation for the EIC, which is not connected to an IC. The approach for this alternative derivation is by Chatterjee and Hadi (1988) referred to as the differentiation approach. The differentiation approach gives another interpretation to the EIC, namely, it measures the rate of change in the vector of parameter estimates as the case weights are either zero or one.

In this article ideas are borrowed from linear inference and without referring to an $IC-$curve the EIC for the parameter estimates in model (1) is defined as follows.

Definition 3.1. *Consider the following nonlinear model* $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, *where* $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{W}(\omega_k))$, *and the weight matrix* $\boldsymbol{W}(\omega_k)$ *equals*

$$\boldsymbol{W}(\omega_k) = diag(1, \ldots, \omega_k^{-1}, \ldots, 1).$$

The EIC for $\widehat{\boldsymbol{\theta}}$ *is defined as*

$$\frac{d}{d\omega_k} \widehat{\boldsymbol{\theta}}(\omega_k) \bigg|_{\omega_k = 1}, \tag{3}$$

where $\widehat{\boldsymbol{\theta}}(\omega_k)$ *is the weighted least squares estimate of* $\boldsymbol{\theta}$ *which is a function of the weight* ω_k, $k = 1, \ldots, n$. *If* $\omega_k \to 1$, *then* $\widehat{\boldsymbol{\theta}}(\omega_k) \to \widehat{\boldsymbol{\theta}}$, *the unweighted least squares estimator.*

Because of Definition 3.1, to estimate $\boldsymbol{\theta}$ in model (1), the method of weighted least squares will be used, i.e.

$$Q(\omega_k) = (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}))^T \boldsymbol{W}^{-1}(\omega_k) (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})) \tag{4}$$

should be minimized. Differentiating $Q(\omega_k)$ with respect to $\boldsymbol{\theta}$, yields the following normal equations:

$$\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)\boldsymbol{W}^{-1}(\omega_k)\left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})\right)=\mathbf{0}, \tag{5}$$

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

The normal equations in (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 and will be denoted by $\widehat{\boldsymbol{\theta}}(\omega_k)$. Now, inserting $\widehat{\boldsymbol{\theta}}(\omega_k)$ in (5) gives

$$\left.\frac{\partial\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\omega_k)}\boldsymbol{W}^{-1}(\omega_k)\left(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}(\omega_k))\right)=\mathbf{0}. \tag{6}$$

In the next theorem it will be shown how to utilize (6) in order to obtain the EIC defined in (3). This diagnostic will be referred to as a joint influence measure, since it measures the influence of a single observation on all parameter estimates in the model simultaneously.

Theorem 3.1. *The joint influence of the k-th observation on the parameter estimate $\widehat{\boldsymbol{\theta}}$, obtained by minimizing $Q(\omega_k)$ in (4) equals*

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

where $\boldsymbol{r}=(r_k)=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}})$, $\boldsymbol{F}(\widehat{\boldsymbol{\theta}})$ is the $q\times n$-matrix such that

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

and $\boldsymbol{G}(\widehat{\boldsymbol{\theta}})$ is a $q\times nq$ matrix derivative

$$\mathbf{G}(\widehat{\boldsymbol{\theta}})=\frac{d\boldsymbol{F}(\widehat{\boldsymbol{\theta}})}{d\widehat{\boldsymbol{\theta}}},$$

where the derivative is defined in the Appendix.

Proof. For simplicity, let $\boldsymbol{F}=\boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\omega_k))$, $\boldsymbol{W}^{-1}=\boldsymbol{W}^{-1}(\omega_k)$ and $\widehat{\boldsymbol{e}}=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X})=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}(\omega_k))$. Using this notation, the normal equations in (6) can be written as

$$\boldsymbol{F}\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}=\mathbf{0}. \tag{7}$$

To study the influence of the k-th observation, differentiate (7) on both sides with respect to ω_k,

$$\frac{d}{d\omega_k}\left(\boldsymbol{F}\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\right) = \mathbf{0}. \tag{8}$$

To calculate the derivative in (8), the product rule (see Appendix) is applied:

$$\frac{d}{d\omega_k}\left(\boldsymbol{F}\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\right) = \frac{d\boldsymbol{F}}{d\omega_k}\left(\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\otimes\boldsymbol{I}_q\right) + \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\widehat{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right) + \frac{d\widehat{\boldsymbol{e}}}{d\omega_k}\boldsymbol{W}^{-1}\boldsymbol{F}^T. \tag{9}$$

In the expression above

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

and due to linearity of $\boldsymbol{W}^{-1}$ the following nice expression is obtained

$$\frac{d\boldsymbol{W}^{-1}}{d\omega_k} = \boldsymbol{d}_k^T \otimes \boldsymbol{d}_k^T,$$

where $\boldsymbol{d}_k$ is the $k-$th column of the identity matrix of size n. Applying the chain rule (see Appendix) to (9) gives

$$\frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\left(\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\otimes\boldsymbol{I}_q\right) + \boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T - \frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\boldsymbol{W}^{-1}\boldsymbol{F}^T = \mathbf{0},$$

which after rearrangement of terms yields

$$\left(\boldsymbol{d}_k^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T\right) = \frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\boldsymbol{W}^{-1}\boldsymbol{F}^T - \frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\omega_k)}\left(\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\otimes\boldsymbol{I}_q\right)\right). \tag{10}$$

Evaluating the derivatives in (10) at $\omega_k = 1$ implies that $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\omega_k = 1)$, $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\omega_k = 1))$ is denoted $\boldsymbol{r}$, and

$$\begin{aligned}
\mathbf{d}_k^T\boldsymbol{r}\otimes\mathbf{d}_k^T\widehat{\boldsymbol{F}}^T(\widehat{\boldsymbol{\theta}}) &= \left.\frac{d\widehat{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\right|_{\omega_k=1}\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \mathbf{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\boldsymbol{I}_q\right)\right) \\
r_k\boldsymbol{F}_k^T(\widehat{\boldsymbol{\theta}}) &= EIC_k^J\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\boldsymbol{I}_q\right)\right).
\end{aligned}$$

Thus, the final expression for EIC_k^J is

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

provided that the inverse exists. □

The EIC_k^J derived in Theorem 3.1 measures the influence of the k-th observation on all the parameter estimates in model (1) simultaneously. Therefore, EIC_k^J is regarded to be a joint influence measure. However, it can be useful to measure the influence of the k-th observation on a particular parameter estimate of the model. Thus, in order to assess the influence of the k-th observation on the j-th parameter, $\widehat{\theta}_j$, a marginal influence measure will be defined. Thus Definition 3.1 is somewhat reformulated and thereafter a theorem corresponding to Theorem 3.1 is presented.

Definition 3.2. *Consider the following nonlinear model* $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, *where* $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{W}(\omega_k))$, *and the weight matrix* $\boldsymbol{W}(\omega_k)$ *is the following*

$$\boldsymbol{W}(\omega_k) = diag(1, \ldots, \omega_k^{-1}, \ldots, 1).$$

Let $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j\right)$ *be a vector of parameter estimates, where* $\widehat{\boldsymbol{\theta}}_1 = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_q\right)$, *are the maximum likelihood estimates from the unperturbed model,* $\widehat{\theta}_h \neq \widehat{\theta}_j$ *and* $\widehat{\theta}_j$ *is estimated from the perturbed model.*

The EIC for $\widehat{\theta}_j$ *is defined as*

$$EIC_{j,k}^M = \frac{d}{d\omega_k} \widehat{\theta}_j(\omega_k) \bigg|_{\omega_k = 1}, \tag{11}$$

where $\widehat{\theta}_j(\omega_k)$ *is the weighted least squares estimate of* θ_j, *which is a function of the weight* ω_k, $k = 1, \ldots, n$. *If* $\omega_k \to 1$, *then* $\widehat{\theta}_j(\omega_k) \to \widehat{\theta}_j$, *the unweighted least squares estimator.*

The weighted least squares criterion and the normal equation for the single parameter case is

$$Q(\omega_k) = \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right)^T \boldsymbol{W}^{-1}(\omega_k) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)\right), \tag{12}$$

$$\left(\frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)}{d\theta_j}\right) \boldsymbol{W}^{-1}(\omega_k) \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j(\omega_k))\right) = 0. \tag{13}$$

Inserting the weighted least squares estimate of θ_j in (13) gives

$$\frac{d\boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \theta_j)}{d\theta_j}\bigg|_{\theta_j = \widehat{\theta}_j(\omega_k)} \boldsymbol{W}^{-1}(\omega_k) \left(\mathbf{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}_1, \widehat{\theta}_j(\omega_k))\right) = 0. \tag{14}$$

In the next theorem (14) is used to provide an explicit expression of the marginal influence measure $EIC^M_{j,k}$ defined in (11).

Theorem 3.2. *The marginal influence of the k-th observation on a single parameter estimate, $\widehat{\theta}_j$, obtained by minimizing (12) equals*

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

where $\boldsymbol{r}=(r_k)=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}_1,\widehat{\theta}_j)$, $\boldsymbol{F}^T(\widehat{\theta}_j)$ is the n-vector such that

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

and $\mathbf{G}^T(\widehat{\theta}_j)$ is the $n-$vector of second derivatives such that

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

Proof. For simplicity, let $\boldsymbol{F} = \boldsymbol{F}(\widehat{\theta}_j(\omega_k))$, $\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}(\omega_k)$ 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 k-th observation, the EIC for $\widehat{\theta}_j$ can be obtained by differentiation of (14) on both sides with respect to ω_k,

$$\frac{d}{d\omega_k}\left(\mathbf{F}\mathbf{W}^{-1}\widehat{\mathbf{e}}\right) = 0. \tag{15}$$

Now, the product rule is used calculate the derivative in (15)

$$\begin{aligned}\frac{d}{d\omega_k}\left(\boldsymbol{F}\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\right) &= \frac{d\boldsymbol{F}}{d\omega_k}\left(\boldsymbol{W}^{-1}\widehat{\boldsymbol{e}}\right) + \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\widehat{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right) \\ &+ \frac{d\widehat{\boldsymbol{e}}}{d\omega_k}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right).\end{aligned} \tag{16}$$

Moreover, applying the chain rule to (16) gives

$$\begin{aligned}&\frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k}\frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\omega_k)}\boldsymbol{W}^{-1}\widehat{\mathbf{e}} + \boldsymbol{d}_k^T\widehat{\boldsymbol{e}} \otimes \boldsymbol{d}_k^T\widehat{\boldsymbol{F}}^T \\ &\quad -\frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\omega_k)}\boldsymbol{W}^{-1}\boldsymbol{F}^T = 0,\end{aligned}$$

and then, rearranging terms yields

$$
\begin{aligned}
& \boldsymbol{d}_k^T \widehat{\boldsymbol{e}} \otimes \boldsymbol{d}_k^T \widehat{\boldsymbol{F}}^T \\
= & \frac{d\boldsymbol{f}(\boldsymbol{X})}{d\omega_k} \boldsymbol{W}^{-1} \boldsymbol{F}^T - \frac{d\boldsymbol{F}}{d\omega_k} \left(\boldsymbol{W}^{-1} \widehat{\mathbf{e}} \right) \\
= & \frac{d\widehat{\theta}_j(\omega_k)}{d\omega_k} \left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_j(\omega_k)} \boldsymbol{W}^{-1} \boldsymbol{F}^T - \frac{d\boldsymbol{F}}{d\widehat{\theta}_j(\omega_k)} \left(\boldsymbol{W}^{-1} \widehat{\boldsymbol{e}} \right) \right).
\end{aligned}
$$

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

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

Thus, the final expression for $EIC_{j,k}^M$ is

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

□

The results obtained in Theorem 3.1 and Theorem 3.2 will be illustrated in the next example.

Example 3.1 Consider the following nonlinear model

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, $\boldsymbol{\varepsilon} \sim N_3(\mathbf{0}, \sigma^2 \boldsymbol{W})$ and $\boldsymbol{W} = diag(1, \omega_2^{-1}, 1)$.

First, the joint influence of the 2nd observation on θ_1 and θ_2 will be studied. Let $\boldsymbol{F} = \boldsymbol{F}(\widehat{\boldsymbol{\theta}}(\omega_2))$, $\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}(\omega_2)$ and $\widehat{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\omega_2))$. The derivative in (8) is as follows

$$
\begin{aligned}
\frac{d}{d\omega_2} \left(\mathbf{F} \mathbf{W}^{-1} \widehat{\mathbf{e}} \right) &= \frac{d\boldsymbol{F}}{d\omega_2} \left(\boldsymbol{W}^{-1} \widehat{\mathbf{e}} \otimes \mathbf{I}_2 \right) + \frac{d\boldsymbol{W}^{-1}}{d\omega_2} \left(\widehat{\mathbf{e}} \otimes \mathbf{F}^T \right) \qquad (17) \\
&+ \frac{d\widehat{\mathbf{e}}}{d\omega_2} \left(\boldsymbol{W}^{-1} \boldsymbol{F}^T \right).
\end{aligned}
$$

In the expression above

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

Now each term in (17) will be calculated, starting with

$$
\begin{aligned}
\frac{d\boldsymbol{W}^{-1}}{d\omega_2}\left(\widehat{\boldsymbol{e}}\otimes\boldsymbol{F}^T\right) &= \boldsymbol{d}_2^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_2^T\boldsymbol{F}^T \\
&= \widehat{e}_2\boldsymbol{F}_2^T \\
&= \begin{pmatrix} \frac{df_2(\widehat{\boldsymbol{\theta}}(\omega_2))\widehat{e}_2}{d\widehat{\theta}_1(\omega_2)}, & \frac{df_2(\widehat{\boldsymbol{\theta}}(\omega_2))\widehat{e}_2}{d\widehat{\theta}_2(\omega_2)} \end{pmatrix} \\
&= \begin{pmatrix} \frac{x_2\widehat{e}_2}{\widehat{\theta}_2(\omega_2)+x_2}, & \frac{-\widehat{\theta}_1(\omega_2)x_2\widehat{e}_2}{(\widehat{\theta}_2(\omega_2)+x_2)^2} \end{pmatrix}.
\end{aligned}
$$

In the next, consider the remaining two terms in (17). Since $\boldsymbol{F}$ and $\widehat{\mathbf{e}}$ are both functions of $\widehat{\boldsymbol{\theta}}(\omega_2)$, the chain rule will be used to calculate the following derivatives

$$
\begin{aligned}
\frac{d\boldsymbol{F}}{d\omega_2}\left(\boldsymbol{W}^{-1}\widehat{\mathbf{e}}\otimes\boldsymbol{I}_2\right) &= \frac{d\widehat{\boldsymbol{\theta}}(\omega_2)}{d\omega_2}\left(\frac{d\boldsymbol{F}}{d\widehat{\boldsymbol{\theta}}(\omega_2)}\left(\boldsymbol{W}^{-1}\widehat{\mathbf{e}}\otimes\boldsymbol{I}_2\right)\right) \\
\frac{d\widehat{\boldsymbol{e}}}{d\omega_2}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right) &= \frac{d\widehat{\boldsymbol{\theta}}(\omega_2)}{d\omega_2}\left(\frac{d\widehat{\boldsymbol{e}}}{d\widehat{\boldsymbol{\theta}}(\omega_2)}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)\right) \\
&= -\frac{d\widehat{\boldsymbol{\theta}}(\omega_2)}{d\omega_2}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\boldsymbol{\theta}}(\omega_2)}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)\right).
\end{aligned}
$$

Evaluated at $\omega_2 = 1$, let $\widehat{\theta}_1 = \widehat{\theta}_1(\omega_2 = 1)$, $\widehat{\theta}_2 = \widehat{\theta}_2(\omega_2 = 1)$ and $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X},\widehat{\boldsymbol{\theta}}(\omega_k = 1))$ is denoted $\boldsymbol{r}$. Thus, (17) results in the expression for EIC_2^J

$$
EIC_2^J = r_2 F_2^T(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widehat{\boldsymbol{\theta}})\boldsymbol{F}^T(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}(\widehat{\boldsymbol{\theta}})\left(\boldsymbol{r}\otimes\boldsymbol{I}_2\right)\right)^{-1},
$$

where

$$
\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} \\
&= \left(\mathbf{A}_1 \;\middle|\; \mathbf{A}_2 \;\middle|\; \mathbf{A}_3 \right),
\end{aligned}
$$

and where

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

Also, consider the matrix

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

which inverse should be used for calculation of EIC_2^J. First,

$$
\begin{aligned}
\mathbf{F}(\widehat{\boldsymbol{\theta}}) \mathbf{F}^T(\widehat{\boldsymbol{\theta}}) &= \begin{pmatrix} \sum_{i=1}^{3} \left(\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} \right)^2 & \sum_{i=1}^{3} \frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} \frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} \\ \sum_{i=1}^{3} \frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_1} \frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} & \sum_{i=1}^{3} \left(\frac{df_i(\widehat{\boldsymbol{\theta}})}{d\widehat{\theta}_2} \right)^2 \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{3} \left(\frac{x_i}{\widehat{\theta}_2 + x_i} \right)^2 & - \sum_{i=1}^{3} \frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2 + x_i)^3} \\ - \sum_{i=1}^{3} \frac{-\widehat{\theta}_1 x_i^2}{(\widehat{\theta}_2 + x_i)^3} & \sum_{i=1}^{3} \left(\frac{-\widehat{\theta}_1 x_i}{(\widehat{\theta}_2 + x_i)^2} \right)^2 \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{1}^T \boldsymbol{X}_1^* & -\mathbf{1}^T \boldsymbol{X}_2^* \\ -\mathbf{1}^T \boldsymbol{X}_2^* & \mathbf{1}^T \boldsymbol{X}_3^* \end{pmatrix},
\end{aligned}
$$

where

$$\boldsymbol{X}_h^* = (x_{h,i}^*), \qquad x_{h,i}^* = \frac{\widehat{\theta}_1^{h-1} x_i^2}{(\widehat{\theta}_2 + x_i)^{1+h}}.$$

and

$$\begin{aligned}
\mathbf{G}(\widehat{\boldsymbol{\theta}})\,(\mathbf{r}\otimes\mathbf{I}_2) &= \begin{pmatrix} \sum\limits_{i=1}^{3} \frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}}) r_i}{d\widehat{\theta}_1^2} & \sum\limits_{i=1}^{3} \frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}}) r_i}{d\widehat{\theta}_1 d\widehat{\theta}_2} \\ \sum\limits_{i=1}^{3} \frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}}) r_i}{d\widehat{\theta}_2 d\theta_1} & \sum\limits_{i=1}^{3} \frac{d^2 f(x_i,\widehat{\boldsymbol{\theta}}) r_i}{d\widehat{\theta}_2^2} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \sum\limits_{i=1}^{3} \frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} \\ \sum\limits_{i=1}^{3} \frac{-x_i r_i}{(\widehat{\theta}_2+x_i)^2} & \sum\limits_{i=1}^{3} \frac{2\widehat{\theta}_1 x_i r_i}{(\widehat{\theta}_2+x_i)^3} \end{pmatrix} \\
&= \begin{pmatrix} 0 & -\boldsymbol{X}_1^{T*}\mathbf{r}^* \\ -\boldsymbol{X}_1^{T*}\mathbf{r}^* & 2\boldsymbol{X}_1^{T*}\mathbf{r}^* \end{pmatrix},
\end{aligned}$$

where

$$\mathbf{r}^* = (r_i^*), \qquad r_i^* = \frac{r_i}{x_i}.$$

Thus,

$$EIC_2^J = \begin{pmatrix} \frac{x_2 r_2}{\widehat{\theta}_2+x_2} & \frac{-\widehat{\theta}_1 x_2 r_2}{(\widehat{\theta}_2+x_2)^2} \end{pmatrix} \left(\begin{pmatrix} \mathbf{1}^T\boldsymbol{X}_1^* & -\mathbf{1}^T\boldsymbol{X}_2^* \\ -\mathbf{1}^T\boldsymbol{X}_2^* & \mathbf{1}^T\boldsymbol{X}_3^* \end{pmatrix} - \begin{pmatrix} 0 & -\boldsymbol{X}_1^{T*}\mathbf{r}^* \\ -\boldsymbol{X}_1^{T*}\mathbf{r}^* & 2\boldsymbol{X}_1^{T*}\mathbf{r}^* \end{pmatrix} \right)^{-1}.$$

Second, the marginal influence of the 2nd observation on $\widehat{\theta}_2$ will be studied.

Let $\boldsymbol{F} = \boldsymbol{F}(\widehat{\theta}_2(\omega_2))$, $\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}(\omega_2)$, $\widehat{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\theta}_1, \widehat{\theta}_j(\omega_2))$. The derivative in (15) equals

$$\frac{d}{d\omega_2}\left(\boldsymbol{F}\boldsymbol{W}^{-1}\widehat{\mathbf{e}}\right) = \frac{d\boldsymbol{F}}{d\omega_2}\boldsymbol{W}^{-1}\widehat{\mathbf{e}} + \frac{d\boldsymbol{W}^{-1}}{d\omega_2}\left(\widehat{\mathbf{e}}\otimes\boldsymbol{F}^T\right) + \frac{d\widehat{\mathbf{e}}}{d\omega_2}\boldsymbol{W}^{-1}\boldsymbol{F}^T, \quad (18)$$

where

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

To calculate the derivative in (18), each term will be considered separately.

$$\begin{aligned}
\frac{d\boldsymbol{W}^{-1}}{d\omega_2}\left(\widehat{\mathbf{e}}\otimes\boldsymbol{F}^T\right) &= \boldsymbol{d}_2^T\widehat{\boldsymbol{e}}\otimes\boldsymbol{d}_2^T\boldsymbol{F}^T \\
&= \widehat{e}_2\boldsymbol{F}_2^T \\
&= \frac{df_2(\widehat{\theta}_2(\omega_2))\widehat{e}_2}{\widehat{\theta}_2(\omega_2)} \\
&= \frac{-\widehat{\theta}_1 x_2\widehat{e}_2}{(\widehat{\theta}_2(\omega_2)+x_2)^2}.
\end{aligned}$$

Finally, consider the remaining two terms in (18). Since $\boldsymbol{F}$ and $\widehat{\mathbf{e}}$ are both functions of $\widehat{\theta}_2(\omega_2)$, the chain rule will be used to calculate the following derivatives

$$\begin{aligned}
\frac{d\boldsymbol{F}}{d\omega_2}\boldsymbol{W}^{-1}\widehat{\mathbf{e}} &= \frac{d\widehat{\theta}_2(\omega_2)}{d\omega_2}\left(\frac{d\boldsymbol{F}}{d\widehat{\theta}_2(\omega_2)}\boldsymbol{W}^{-1}\widehat{\mathbf{e}}\right) \\
\frac{d\widehat{\boldsymbol{e}}}{d\omega_2}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right) &= \frac{d\widehat{\theta}_2(\omega_2)}{d\omega_2}\left(\frac{d\widehat{\boldsymbol{e}}}{d\widehat{\theta}_2(\omega_2)}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)\right) \\
&= -\frac{d\widehat{\theta}_2(\omega_2)}{d\omega_2}\left(\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widehat{\theta}_2(\omega_2)}\left(\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)\right).
\end{aligned}$$

Evaluated at $\omega_2 = 1$, $\widehat{\theta}_2 = \widehat{\theta}_2(\omega_2 = 1)$, $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widehat{\theta}_1, \widehat{\theta}_2(\omega_2 = 1))$ is denoted $\boldsymbol{r}$ and (18) results in the expression for $EIC^M_{2,2}$

$$EIC^M_{2,2} = r_2F_2(\widehat{\theta}_2)\left(\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2) - \boldsymbol{G}(\widehat{\theta}_2)\boldsymbol{r}\right)^{-1},$$

where

$$\begin{aligned}
\mathbf{G}(\widehat{\theta}_2) &= \frac{d\boldsymbol{F}(\widehat{\theta}_2)}{d\widehat{\theta}_2} \\
&= \begin{pmatrix} \frac{d^2f_1(\widehat{\theta}_2)}{d\widehat{\theta}_2^2} & \frac{d^2f_2(\widehat{\theta}_2)}{d\widehat{\theta}_2^2} & \frac{d^2f_3(\widehat{\theta}_2)}{d\widehat{\theta}_2^2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{2\widehat{\theta}_1x_1}{(\widehat{\theta}_2+x_1)^3} & \frac{2\widehat{\theta}_1x_2}{(\widehat{\theta}_2+x_2)^3} & \frac{2\widehat{\theta}_1x_3}{(\widehat{\theta}_2+x_3)^3} \end{pmatrix}.
\end{aligned}$$

Also consider $\left(\boldsymbol{F}(\widehat{\theta}_2)\boldsymbol{F}^T(\widehat{\theta}_2) - \boldsymbol{G}(\widehat{\theta}_2)\boldsymbol{r}\right)$ which inverse should be used for calculation of $EIC_{2,2}^M$. First,

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

and

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

Thus,

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

4. The EIC of the score test statistic in nonlinear regression

Numerous influence diagnostics are available for use in linear regression, for example Cook's Distance, Welch-Kuh's Distance, DFBETAS and the Andrews-Pregibon Statistic presented in e.g. Belsley *et al.* (1980) and Chatterjee and Hadi (1988). By contrast, very little has been done for nonlinear models. The influence measures for nonlinear models have been primary concerned with measuring the impact of observations on the parameter estimates. However, it is of interest to assess the observations' influence on other aspects of the statistical inference as well, such as a hypothesis test about regression parameters.

Some observations have more impact on the outcome of a hypothesis test than others. In fact, the hypothesis testing procedure can result in a significant (non significant) result due to the influence of a single observation. If the data contains such influential observation it is beneficial for the analyst to know about it, since this observation may carry a lot of information. In this article an important aspect of the influence analysis is highlighted, that is; how do individual observations contribute to the conclusions regarding a null hypothesis? An explicit expression of an influence measure of the score test statistic is derived to address this question. 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.

The score test is, under certain regularity conditions (see e.g. Lehmann (1999)) asymptotically equivalent to the Wald test and the likelihood ratio test - LRT. It can be used for testing hypotheses about parameters in nonlinear regression models. The score test has the advantage that it only requires evaluation under the null hypothesis, whereas the Wald test and the LRT require maximum likelihood estimates under both the null hypothesis and its alternative. Moreover, the score test is a locally most powerful test. The Added Parameter Plot, see Stål (2011), illustrates the effect of the observations on the score test statistic. To read more about the score test, see e.g. Rao (1973) and Lehmann and Romano (2005).

For model (1) let $\boldsymbol{\Psi} = \left(\boldsymbol{\theta}, \sigma^2\right)$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$. The score test statistic is based on the score vector

$$\mathbf{U}(\boldsymbol{\Psi}) = \left(\frac{d \log L(\boldsymbol{\Psi}, \mathbf{y})}{d\theta_1}, \ldots, \frac{d \log L(\boldsymbol{\Psi}, \mathbf{y})}{d\theta_q}, \frac{d \log L(\boldsymbol{\Psi}, \mathbf{y})}{d\sigma^2}\right)^T,$$

for the $q+1-$vector $\boldsymbol{\Psi}$, where $L(\boldsymbol{\Psi}, \mathbf{y})$ is the likelihood function. Without loss of generality, consider a hypothesis test of a single parameter

$$H_0 : \boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}, \tag{19}$$

where $\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\theta}_1, \ldots, \widetilde{\theta}_{q-1}, 0\right)^T$. Let $\widetilde{\boldsymbol{\Psi}}$ be the maximum likelihood estimate of $\boldsymbol{\Psi}$ under the null hypothesis (19). The score test statistic equals

$$\begin{aligned} S(\widetilde{\boldsymbol{\theta}}) &= \mathbf{U}^T(\widetilde{\boldsymbol{\Psi}})\mathbf{I}^{-1}(\widetilde{\boldsymbol{\Psi}})\mathbf{U}(\widetilde{\boldsymbol{\Psi}}) \\ &= \mathbf{U}^T(\widetilde{\boldsymbol{\theta}})\mathbf{I}^{-1}(\widetilde{\boldsymbol{\theta}})\mathbf{U}(\widetilde{\boldsymbol{\theta}}), \end{aligned} \tag{20}$$

since $\boldsymbol{I}(\widetilde{\boldsymbol{\Psi}})$ is block-diagonal and

$$\left.\frac{d \log L(\boldsymbol{\Psi}, \mathbf{y})}{d\sigma^2}\right|_{\sigma^2=\widetilde{\sigma}^2} = 0.$$

Furthermore, under H_0 the score test statistic in (20) has, asymptotically, a χ^2 distribution with degrees of freedom equal to the number of restrictions on the parameters under the null hypothesis, see e.g. Rao (2005).

The score vector and information matrix in (20) are the following

$$\mathbf{U}(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\widetilde{\sigma}^2}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}},$$

and

$$\begin{aligned}\mathbf{I}(\widetilde{\boldsymbol{\theta}}) &= E\left(\mathbf{U}(\widetilde{\boldsymbol{\theta}})\mathbf{U}^T(\widetilde{\boldsymbol{\theta}})\right) \\ &= \frac{1}{\widetilde{\sigma}^2}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}}).\end{aligned}$$

Hence, the score test statistic can be written as

$$S(\widetilde{\boldsymbol{\theta}}) = \frac{1}{\widetilde{\sigma}^2}\widetilde{\mathbf{r}}^T\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\left(\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)^{-1}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}}, \tag{21}$$

where

$$\widetilde{\mathbf{r}} = (\widetilde{r}_k) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}) \tag{22}$$

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

$$\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) = \left(\boldsymbol{F}_1(\widetilde{\boldsymbol{\theta}}), \ldots, \boldsymbol{F}_n(\widetilde{\boldsymbol{\theta}})\right) = \left.\frac{d}{d\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}. \tag{23}$$

Using the same methodology as in the previous section, the influence measure of the score test statistic, $EIC_{score,k}$, is defined as follows

Definition 4.1. *Consider the nonlinear model presented in Definition 3.1. The $EIC_{score,k}$ is defined as*

$$EIC_{score,k} = \left.\frac{dS(\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k}\right|_{\omega_k=1},$$

where

$$S(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \mathbf{U}^T(\widetilde{\boldsymbol{\theta}}(\omega_k))\mathbf{I}^{-1}(\widetilde{\boldsymbol{\theta}}(\omega_k))\mathbf{U}(\widetilde{\boldsymbol{\theta}}(\omega_k)), \tag{24}$$

$$\mathbf{U}(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \frac{1}{\widetilde{\sigma}^2(\omega_k)}\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))\boldsymbol{W}^{-1}(\omega_k)\left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\omega_k))\right),$$

$$\boldsymbol{I}(\widetilde{\boldsymbol{\theta}}(\omega_k)) = \frac{1}{\widetilde{\sigma}^2(\omega_k)}\left(\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))\boldsymbol{W}^{-1}(\omega_k)\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}}(\omega_k))\right)$$

and $\widetilde{\boldsymbol{\theta}}(\omega_k)$ *is the maximum likelihood estimate of* $\boldsymbol{\theta}$ *under the null hypothesis (19) in the perturbed model, i.e.*

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{W}(\omega_k)),$$

when

$$\boldsymbol{W}(\omega_k) = diag(1, \ldots, \omega_k^{-1}, \ldots, 1).$$

The next theorem provides an explicit expression of the $EIC_{score,k}$.

Theorem 4.1. *The* $EIC_{score,k}$ *given in Definition 4.1 equals*

$$\begin{aligned} EIC_{score,k} = \frac{1}{\widetilde{\sigma}^2}\Big[2\left(\widetilde{r}_k\boldsymbol{F}_k^T + EIC^J_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}\left(\widetilde{\mathbf{r}} \otimes \boldsymbol{I}_q\right)\right)\mathbf{g} \\ - \left(EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T \otimes \boldsymbol{I}_q\right) + \mathbf{G}^*\left(\boldsymbol{I}_q \otimes \boldsymbol{F}^T\right)\right) + vec^T\left(\boldsymbol{F}_k\boldsymbol{F}_k^T\right)\right) \\ \times (\mathbf{g} \otimes \mathbf{g}) - S(\widetilde{\boldsymbol{\theta}})\frac{\widetilde{r}_k^2}{n}\Big]. \end{aligned}$$

where $\widetilde{\mathbf{r}}$ *and* $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$ *are defined in (22) and (23), respectively,* $\mathbf{G} = \mathbf{G}(\widetilde{\boldsymbol{\theta}})$ *is a* $q \times nq$ *matrix derivative*

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \frac{d\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}},$$

$\mathbf{G}^* = \mathbf{G}^*(\widetilde{\boldsymbol{\theta}})$ *is a* $q \times nq$ *matrix derivative*

$$\mathbf{G}^*(\widetilde{\boldsymbol{\theta}}) = \frac{d\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}},$$

$\mathbf{g}$ *is a* $q-$*vector such that*

$$\mathbf{g} = \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}}.$$

and $\left(EIC^J_{\widetilde{\boldsymbol{\theta}},k}\right)^T$ *is the* $q-$*vector of joint influence of the parameter estimates under the null hypothesis, and hence, the* $q-th$ *value of the vector equals zero.*

Proof. For simplicity, let $\boldsymbol{F} = \boldsymbol{F}(\widetilde{\boldsymbol{\theta}}(\omega_k))$, $\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}(\omega_k)$ and $\widetilde{\mathbf{e}} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{X}, \widetilde{\boldsymbol{\theta}}(\omega_k))$.

In (24), let

$$a(\omega_k) = \widetilde{\mathbf{e}}^T\boldsymbol{W}^{-1}\boldsymbol{F}^T\left(\boldsymbol{F}\mathbf{W}^{-1}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\mathbf{W}^{-1}\widetilde{\mathbf{e}}$$

and

$$b(\omega_k) = \widetilde{\sigma}^2(\omega_k),$$

i.e, when differentiating $S(\widetilde{\boldsymbol{\theta}}(\omega_k))$, the quotient rule is used,

$$\frac{dS(\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k} = \frac{a'(\omega_k)b(\omega_k) - a(\omega_k)b'(\omega_k)}{b^2(\omega_k)},$$

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 $\mathbf{C} = \boldsymbol{F}\mathbf{W}^{-1}\widetilde{\mathbf{e}}$ and $\mathbf{D} = \boldsymbol{F}\mathbf{W}^{-1}\boldsymbol{F}^T$, 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

$$\begin{aligned}\frac{d\mathbf{C}}{d\omega_k} &= \frac{d}{d\omega_k}\boldsymbol{F}\mathbf{W}^{-1}\widetilde{\mathbf{e}} \\ &= \frac{d\boldsymbol{F}}{d\omega_k}\left(\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\otimes\mathbf{I}_q\right) + \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\mathbf{F}^T\right) + \frac{d\widetilde{\mathbf{e}}}{d\omega_k}\boldsymbol{W}^{-1}\boldsymbol{F}^T. \quad (25)\end{aligned}$$

Applying the chain rule to (25) gives

$$\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\otimes \boldsymbol{I}_q\right)+\boldsymbol{d}_k^T\widetilde{\boldsymbol{e}}\ \otimes\ \boldsymbol{d}_k^T\boldsymbol{F}^T \qquad -\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\frac{d\boldsymbol{f}(\boldsymbol{X})}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\boldsymbol{W}^{-1}\boldsymbol{F}^T. \tag{26}$$

Next the derivative of $\mathbf{D}$ with respect to ω_k is considered. Using the rule for differentiation of a matrix inverse, 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,

$$\begin{aligned}\frac{d\mathbf{D}}{d\omega_k} &= \frac{d}{d\omega_k}\left(\mathbf{F}\mathbf{W}^{-1}\boldsymbol{F}^T\right)\\ &= \frac{d\boldsymbol{F}}{d\omega_k}(\boldsymbol{W}^{-1}\boldsymbol{F}^T\otimes\boldsymbol{I}_q)+\frac{d\boldsymbol{W}^{-1}}{d\omega_k}(\boldsymbol{F}^T\otimes\boldsymbol{F}^T)+\frac{d\boldsymbol{F}^T}{d\omega_k}(\mathbf{I}_q\otimes\boldsymbol{W}^{-1}\boldsymbol{F}^T)\\ &= \frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{W}^{-1}\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\mathbf{I}_q\otimes\mathbf{W}^{-1}\boldsymbol{F}^T\right)\right)\\ &+ \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\boldsymbol{F}^T\otimes\mathbf{F}^T\right),\end{aligned}$$

and

$$\frac{d\mathbf{D}^{-1}}{d\omega_k}=-\left[\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{F}}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}(\boldsymbol{W}^{-1}\boldsymbol{F}^T\otimes\boldsymbol{I}_q)+\frac{d\boldsymbol{F}^T}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}(\boldsymbol{I}_q\otimes\boldsymbol{W}^{-1}\boldsymbol{F}^T)\right) + \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\boldsymbol{F}^T\otimes\boldsymbol{F}^T\right)\left(\left(\boldsymbol{F}\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{W}^{-1}\boldsymbol{F}^T\right)^{-1}\right)\right]. \tag{27}$$

Evaluated at $\omega_k=1$, $\widetilde{\boldsymbol{\theta}}=\widetilde{\boldsymbol{\theta}}(\omega_k=1)$ and $\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k=1))$ is denoted $\widetilde{\boldsymbol{r}}$. Let $\boldsymbol{F}=\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})$, $\mathbf{G}=\mathbf{G}(\widetilde{\boldsymbol{\theta}})$ and $\mathbf{G}^*=\mathbf{G}^*(\widetilde{\boldsymbol{\theta}})$. The derivatives in (26) and (27) become

$$\left.\frac{d}{d\omega_k}\boldsymbol{F}\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right|_{\omega_k=1}=\widetilde{r}_k\boldsymbol{F}_k^T-EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\boldsymbol{F}\boldsymbol{F}^T-\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\boldsymbol{I}_q\right)\right),$$

$$\begin{aligned}\left.\frac{d(\boldsymbol{F}\boldsymbol{W}^{-1}\boldsymbol{F}^T)^{-1}}{d\omega_k}\right|_{\omega_k=1} &= -\left[EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\mathbf{I}_q\right)+\mathbf{G}^*\left(\mathbf{I}_q\otimes\mathbf{F}^T\right)\right)\right.\\ &+ \left. vec^T\left(\boldsymbol{F}_k\boldsymbol{F}_k^T\right)\right]\times\left(\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\right).\end{aligned}$$

Let

$$\mathbf{g} = \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}},$$

then

$$\begin{aligned}
\left.\frac{da(\omega_k)}{d\omega_k}\right|_{\omega_k=1} &= 2\left(\widetilde{r}_k\boldsymbol{F}_k^T - EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\boldsymbol{F}\boldsymbol{F}^T - \mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\boldsymbol{I}_q\right)\right)\right)\mathbf{g} \\
&- \Big[EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\boldsymbol{I}_q\right) + \mathbf{G}^*\left(\boldsymbol{I}_q\otimes\boldsymbol{F}^T\right)\right) \\
&+ vec^T\left(\boldsymbol{F}_k\boldsymbol{F}_k^T\right)\Big]\times\left(\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\otimes\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\right)\left(\boldsymbol{F}\widetilde{\mathbf{r}}\otimes\boldsymbol{F}\widetilde{\mathbf{r}}\right) \\
&= 2\left(\widetilde{r}_k\boldsymbol{F}_k^T + EIC^J_{\widetilde{\boldsymbol{\theta}},k}\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\boldsymbol{I}_q\right)\right)\mathbf{g} \\
&- \Big[EIC^J_{\widetilde{\boldsymbol{\theta}},k}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes I_q\right) + \mathbf{G}^*\left(\mathbf{I}_q\otimes\boldsymbol{F}^T\right)\right) \\
&+ vec^T\left(\boldsymbol{F}_k\boldsymbol{F}_k^T\right)\Big]\times\left(\mathbf{g}\otimes\mathbf{g}\right).
\end{aligned}$$

In the expression above, $EIC^J_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\boldsymbol{F}^T\mathbf{g} = 0$. This is due to the fact that the normal equations in (3) for estimating $\boldsymbol{\theta}$ is set to zero and $EIC^J_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}\boldsymbol{F}\widetilde{\mathbf{r}} = EIC^J_{\widetilde{\boldsymbol{\theta}},k}\boldsymbol{F}\widetilde{\mathbf{r}} = 0$.

Now the derivative of the variance term needs to be calculated. The MLE of $\sigma^2(\omega_k)$ under the null hypothesis of (19) is

$$\widetilde{\sigma}^2(\omega_k) = \frac{1}{n}\left(\widetilde{\mathbf{e}}^T\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right).$$

Using the product rule and the chain rule, the derivative of $\widetilde{\sigma}^2(\omega_k)$ with respect to ω_k is the following

$$
\begin{aligned}
\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k} &= \frac{1}{n}\frac{d}{d\omega_k}\left(\widetilde{\mathbf{e}}^T \boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right) \\
&= \frac{1}{n}\left(2\frac{d\widetilde{\mathbf{e}}}{d\omega_k}\boldsymbol{W}^{-1}\widetilde{\mathbf{e}} + \frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&= \frac{1}{n}\left(\frac{d\boldsymbol{W}^{-1}}{d\omega_k}\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right) - 2\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\omega_k}\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right) \\
&= \frac{1}{n}\left(\left(\mathbf{d}_k^T\otimes\mathbf{d}_k^T\right)\left(\widetilde{\mathbf{e}}\otimes\widetilde{\mathbf{e}}\right)\right) \\
&- \frac{1}{n}\left(2\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right)\right)\right) \\
&= \frac{1}{n}\left(\widetilde{e}_k^2 - 2\frac{d\widetilde{\boldsymbol{\theta}}(\omega_k)}{d\omega_k}\left(\frac{d\boldsymbol{f}(\boldsymbol{X},\widetilde{\boldsymbol{\theta}}(\omega_k))}{d\widetilde{\boldsymbol{\theta}}(\omega_k)}\left(\boldsymbol{W}^{-1}\widetilde{\mathbf{e}}\right)\right)\right).
\end{aligned}
$$

Evaluated at $\omega_k = 1$, $\widetilde{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}(\omega_k = 1)$, $\widetilde{\mathbf{e}}$ is denoted $\widetilde{\mathbf{r}}$ and hence

$$
\begin{aligned}
\left.\frac{d\widetilde{\sigma}^2(\omega_k)}{d\omega_k}\right|_{\omega_k=1} &= \frac{1}{n}\left(\widetilde{r}_k^2 - 2EIC_{\widetilde{\theta},k}^J\mathbf{F}\widetilde{\mathbf{r}}\right) \\
&= \frac{\widetilde{r}_k^2}{n},
\end{aligned}
$$

since $EIC_{\widetilde{\theta},k}\mathbf{F}\widetilde{\mathbf{r}} = 0$.

Finally, the expression for $EIC_{score,k}$ is given by

$$
\begin{aligned}
EIC_{score,k} &= \frac{1}{\widetilde{\sigma}^2}\left[2\left(\widetilde{r}_k\boldsymbol{F}_k^T + EIC_{\widetilde{\boldsymbol{\theta}},k}^J\mathbf{G}\left(\widetilde{\mathbf{r}}\otimes\boldsymbol{I}_q\right)\right)\mathbf{g}\right. \\
&- \left(EIC_{\widetilde{\boldsymbol{\theta}},k}^J\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\boldsymbol{I}_q\right) + \mathbf{G}^*\left(\boldsymbol{I}_q\otimes\boldsymbol{F}^T\right)\right) + vec^T\left(\boldsymbol{F}_k\boldsymbol{F}_k^T\right)\right) \\
&\left.\times\left(\mathbf{g}\otimes\mathbf{g}\right) - S(\widetilde{\boldsymbol{\theta}})\frac{\widetilde{r}_k^2}{n}\right].
\end{aligned}
$$

□

In the next example the results from Theorem 4.1 will be illustrated.

Example 4.1
Consider the same model as in Example 3.1, where

$$\boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) = \begin{pmatrix} \frac{\theta_1 x_1}{\theta_2+x_1} & \frac{\theta_1 x_2}{\theta_2+x_2} & \frac{\theta_1 x_3}{\theta_2+x_3} \end{pmatrix}^T.$$

Let the hypothesis of interest be

$$H_0 : \theta_2 = 0.$$

The Score test statistic is given in (21), and $\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\theta}_1, 0\right)^T$. For this particular test, the residuals under the null hypothesis are given by

$$\widetilde{\mathbf{r}} = \boldsymbol{y} - \widetilde{\theta}_1 \mathbf{1}_3$$

and

$$\boldsymbol{F}(\widetilde{\boldsymbol{\theta}}) = \frac{d}{d\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$$

is a 2×3 matrix such that its first row is

$$\frac{d}{d\theta_1} \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \mathbf{1}_3^T,$$

and its second row is

$$\frac{d}{d\theta_2} \boldsymbol{f}(\boldsymbol{X},\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \begin{pmatrix} -\frac{\widetilde{\theta}_1}{x_1} & -\frac{\widetilde{\theta}_1}{x_2} & -\frac{\widetilde{\theta}_1}{x_3} \end{pmatrix}.$$

For illustration, each term in the expression of $EIC_{score,2}$ will be evaluated, starting with

$$\begin{aligned} \widetilde{r}_2 \boldsymbol{F}_2^T(\widetilde{\boldsymbol{\theta}}) &= \begin{pmatrix} \frac{df_2(\widetilde{\boldsymbol{\theta}})\widetilde{r}_2}{d\widetilde{\theta}_1}, & \frac{df_2(\widetilde{\boldsymbol{\theta}})\widetilde{r}_2}{d\widetilde{\theta}_2} \end{pmatrix} \\ &= \begin{pmatrix} \widetilde{r}_2, & -\frac{\widetilde{\theta}_1}{x_2}\widetilde{r}_2 \end{pmatrix}. \end{aligned}$$

The next term is $EIC^J_{\widetilde{\boldsymbol{\theta}},2} \boldsymbol{G}(\widetilde{\boldsymbol{\theta}}) \left(\widetilde{\mathbf{r}} \otimes \boldsymbol{I}_2\right) \mathbf{g}$, where

$$EIC^J_{\widetilde{\boldsymbol{\theta}},2} = \begin{pmatrix} EIC^J_{\widetilde{\theta}_1,2}, & EIC^J_{\widetilde{\theta}_2,2} \end{pmatrix} = \begin{pmatrix} EIC^J_{\widetilde{\theta}_1,2}, & 0 \end{pmatrix},$$

and

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} \\ \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} \end{pmatrix}.$$

In the matrix $\mathbf{G}(\widetilde{\boldsymbol{\theta}})$

$$\begin{aligned} \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} &= 0, \\ \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} &= \frac{2\widetilde{\theta}_1}{x_i^2} \end{aligned}$$

and

$$\begin{aligned} \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} &= \frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} \\ = -\frac{x_i}{(\widetilde{\theta}_2 + x_i)^2} &= -\frac{1}{x_i}, \end{aligned}$$

so that

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}}) = \begin{pmatrix} 0 & -\frac{1}{x_1} & 0 & -\frac{1}{x_2} & 0 & -\frac{1}{x_3} \\ -\frac{1}{x_1} & \frac{2\widetilde{\theta}_1}{x_1^2} & -\frac{1}{x_2} & \frac{2\widetilde{\theta}_1}{x_2^2} & -\frac{1}{x_3} & \frac{2\widetilde{\theta}_1}{x_3^2} \end{pmatrix}.$$

Also

$$\begin{aligned} (\widetilde{\mathbf{r}} \otimes \mathbf{I}_2) &= \begin{pmatrix} \widetilde{r}_1 \\ \widetilde{r}_2 \\ \widetilde{r}_3 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \widetilde{r}_1 \mathbf{I}_2 & \widetilde{r}_2 \mathbf{I}_2 & \widetilde{r}_3 \mathbf{I}_2 \end{pmatrix}^T, \end{aligned}$$

and

$$\begin{aligned} \mathbf{g} &= \left(\boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)^{-1} \boldsymbol{F}(\widetilde{\boldsymbol{\theta}})\widetilde{\mathbf{r}} \\ &= \begin{pmatrix} \sum_{i=1}^{3} 1 & -\sum_{i=1}^{3} \frac{\widetilde{\theta}_1}{x_i} \\ -\sum_{i=1}^{3} \frac{\widetilde{\theta}_1}{x_i} & \sum_{i=1}^{3} (\frac{\widetilde{\theta}_1}{x_i})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{3} \widetilde{r}_i \\ -\sum_{i=1}^{3} \widetilde{\theta}_1 \widetilde{r}_i^* \end{pmatrix}, \end{aligned}$$

where

$$\widetilde{\mathbf{r}}^* = (\widetilde{r}_i^*), \quad \widetilde{r}_i^* = \frac{\widetilde{r}_i}{x_i}.$$

Using the rule for inversion of a 2×2-matrix, one gets $\mathbf{g} = c\mathbf{B}$, where c is a constant such that

$$c = \frac{1}{3\sum_{i=1}^{3}\left(\frac{\widetilde{\theta}_1}{x_i}\right)^2 - \left(\sum_{i=1}^{3}\frac{\widetilde{\theta}_1}{x_i}\right)^2},$$

and $\mathbf{B}$ is a $2-$vector such that

$$\mathbf{B} = \begin{pmatrix} \sum_{i=1}^{3}\left(\frac{\widetilde{\theta}_1}{x_i}\right)^2 \sum_{j=1}^{3}\widetilde{r}_j - \sum_{i=1}^{3}\left(\frac{\widetilde{\theta}_1}{x_i}\right)\sum_{j=1}^{3}\widetilde{\theta}_1\widetilde{r}_j^* \\ \sum_{i=1}^{3}\left(\frac{\widetilde{\theta}_1}{x_i}\right)\sum_{j=1}^{3}\widetilde{r}_j - 3\sum_{j=1}^{3}\widetilde{\theta}_1\widetilde{r}_j^* \end{pmatrix}.$$

Thus

$$\begin{aligned}\left(EIC^J_{\widetilde{\boldsymbol{\theta}},2}\mathbf{G}(\widetilde{\boldsymbol{\theta}})\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_2\right)\right)^T &= \begin{pmatrix} EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2}\widetilde{r}_i \\ EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2}\widetilde{r}_i \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\widetilde{r}_i^* \end{pmatrix},\end{aligned}$$

and

$$2\left(\widetilde{r}_2\boldsymbol{F}_2^T + EIC^J_{\widetilde{\boldsymbol{\theta}},2}\mathbf{G}(\widetilde{\boldsymbol{\theta}})\left(\widetilde{\mathbf{r}}\otimes\mathbf{I}_2\right)\right)\mathbf{g}$$

$$= 2c\begin{pmatrix} \widetilde{r}_2 & -\widetilde{\theta}_1\widetilde{r}_2^* - EIC^J_{\widetilde{\theta}_1,2}\sum\widetilde{r}_i^* \end{pmatrix}\mathbf{B}.$$

Moving on to

$$EIC^J_{\widetilde{\boldsymbol{\theta}},2}\Big(\mathbf{G}(\widetilde{\boldsymbol{\theta}})\Big(\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\otimes\mathbf{I}_2\Big)+\mathbf{G}^*(\widetilde{\boldsymbol{\theta}})\left(\mathbf{I}_2\otimes\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)\Big)+vec^T\left(\boldsymbol{F}_2(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}_2^T(\widetilde{\boldsymbol{\theta}})\right),$$

where $\mathbf{G}(\widetilde{\boldsymbol{\theta}})$ is given above and

$$\begin{aligned}
\mathbf{G}^*(\widetilde{\boldsymbol{\theta}}) &= \frac{d\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\boldsymbol{\theta}}} \\
&= \begin{pmatrix} \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} \\ \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \frac{d^2 f_1(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} & \frac{d^2 f_3(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & 0 & -\frac{1}{x_1} & -\frac{1}{x_2} & -\frac{1}{x_3} \\ -\frac{1}{x_1} & -\frac{1}{x_2} & -\frac{1}{x_3} & \frac{2\widetilde{\theta}_1}{{x_1}^2} & \frac{2\widetilde{\theta}_1}{{x_2}^2} & \frac{2\widetilde{\theta}_1}{{x_3}^2} \end{pmatrix}.
\end{aligned}$$

Now, $\mathbf{G}(\widetilde{\boldsymbol{\theta}})\left(\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\otimes\mathbf{I}_2\right)$ and $\mathbf{G}^*(\widetilde{\boldsymbol{\theta}})\left(\mathbf{I}_2\otimes\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)$ are matrices of sums of first order derivatives multiplied by second order derivatives such that

$$\mathbf{G}(\widetilde{\boldsymbol{\theta}})\left(\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\otimes\mathbf{I}_2\right)$$

$$= \begin{pmatrix} \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_1^2} & \sum_{i=1}^{3}\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} & \cdots & \sum_{i=1}^{3}\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} \\ \sum_{i=1}^{3}\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \sum_{i=1}^{3}\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_2^2} & \cdots & \sum_{i=1}^{3}\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\theta})}{d\widetilde{\theta}_2^2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -\sum_{i=1}^{3}\frac{1}{x_i} & 0 & \sum_{i=1}^{3}\frac{\widetilde{\theta}_1}{x_i^2} \\ -\sum_{i=1}^{3}\frac{1}{x_i} & -\sum_{i=1}^{3}\frac{2\widetilde{\theta}_1}{x_i} & \sum_{i=1}^{3}\frac{\widetilde{\theta}_1}{x_i^2} & -\sum_{i=1}^{3}\frac{2(\widetilde{\theta}_1)^2}{(x_i)^2} \end{pmatrix}$$

and

$$\mathbf{G}^*(\widetilde{\boldsymbol{\theta}})\left(\mathbf{I}_2\otimes\boldsymbol{F}^T(\widetilde{\boldsymbol{\theta}})\right)$$

$$= \begin{pmatrix} \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2} & \cdots & \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2} \\ \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2 d\widetilde{\theta}_1} & \cdots & \sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2 f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2^2} \end{pmatrix}$$

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

Moreover,

$$vec\left(\boldsymbol{F}_2(\widetilde{\boldsymbol{\theta}})\boldsymbol{F}_2^T(\widetilde{\boldsymbol{\theta}})\right) \quad = \quad \begin{pmatrix} \left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\right)^2 \\ \left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\right)\left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\right) \\ \left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_2}\right)\left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_1}\right) \\ \left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_2}\right)^2 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{-\widetilde{\theta}_1}{x_2} \\ \frac{-\widetilde{\theta}_1}{x_2} \\ (\frac{\widetilde{\theta}_1}{x_2})^2 \end{pmatrix}$$

and this results in the following expression

$$\left(EIC^J_{\widetilde{\boldsymbol{\theta}},2}\left(\mathbf{G}\left(\boldsymbol{F}^T\otimes\boldsymbol{I}_2\right)+\mathbf{G}^T\left(\mathbf{I}_2\otimes\boldsymbol{F}^T\right)\right)+vec^T\left(\boldsymbol{F}_2\boldsymbol{F}_2^T\right)\right)^T$$

$$= \begin{pmatrix} 2EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2}+\left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\right)^2 \\ EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\left(\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2}+\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1^2}\right)+\left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\right)\left(\frac{df_2(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\right) \\ EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\left(\frac{df_i(\widetilde{\theta})}{d\widetilde{\theta}_2}\frac{d^2f_i(\widetilde{\theta})}{d\widetilde{\theta}_1^2}+\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1}\frac{d^2f_i(\widetilde{\theta})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2}\right)+\left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_2}\right)\left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_1}\right) \\ 2EIC^J_{\widetilde{\theta}_1,2}\left(\sum_{i=1}^{3}\frac{df_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_2}\frac{d^2f_i(\widetilde{\boldsymbol{\theta}})}{d\widetilde{\theta}_1 d\widetilde{\theta}_2}\right)+\left(\frac{df_2(\widetilde{\theta})}{d\widetilde{\theta}_2}\right)^2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{1}{x_i}-\frac{\widetilde{\theta}_1}{x_2} \\ -EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{1}{x_i}-\frac{\widetilde{\theta}_1}{x_2} \\ 2EIC^J_{\widetilde{\theta}_1,2}\sum_{i=1}^{3}\frac{\widehat{\theta}_1}{x_i^2}+\left(\frac{\widetilde{\theta}_1}{x_2}\right)^2 \end{pmatrix}.$$

The final expression for $EIC_{score,2}$ is the following

$$EIC_{score,2} = \frac{2c}{\widetilde{\sigma}^2}\left(\; \widetilde{r}_2 \quad -\widetilde{\theta}_1 \widetilde{r}_2^* - EIC^J_{\widetilde{\theta}_1,2} \sum_{i=1}^{3} \widetilde{r}_i^* \; \right) \mathbf{B}$$

$$- \frac{c^2}{\widetilde{\sigma}^2} \begin{pmatrix} 1 \\ -EIC^J_{\widetilde{\theta}_1,2} \sum_{i=1}^{3} \frac{1}{x_i} - \frac{\widetilde{\theta}_1}{x_2} \\ -EIC^J_{\widetilde{\theta}_1,2} \sum_{i=1}^{3} \frac{1}{x_i} - \frac{\widetilde{\theta}_1}{x_2} \\ 2EIC^J_{\widetilde{\theta}_1,2} \sum_{i=1}^{3} \frac{\widehat{\theta}_1}{x_i^2} + \left(\frac{\widetilde{\theta}_1}{x_2}\right)^2 \end{pmatrix}^T \times \mathbf{B} \otimes \mathbf{B} - S(\widetilde{\boldsymbol{\theta}}) \frac{\widetilde{r}_2^2}{3\widetilde{\sigma}^2}.$$

5. Concluding remarks

This article contributes to the influence analysis as it provides influence diagnostics for use in nonlinear regression analysis. Explicit expressions of the EIC for both the parameter estimates and the score test statistic are derived. The EIC for the parameter estimates are constructed in two different ways to assess the marginal and the joint influence of the observations on the parameter estimates.

The EIC-measures derived in this article are useful since they make it possible to identify influential observations that might compromise the reliability of the inference and results in the regression analysis. If the primary interest is to investigate the influence of observations on the parameter estimates, the EIC_k^J or the $EIC_{j,k}^M$ given in Theorem 3.1 and 3.2 respectively, should be used.

The EIC_k^J is a joint influence diagnostic and measures the influence of the $k-$th observation on all parameter estimates in the model simultaneously. The $EIC_{j,k}^M$ is a marginal influence diagnostic that measures the influence of the $k-$th observation on the $j-$th parameter estimate, when the other parameters in the model are regarded as known.

On the other hand, if the aim is to study how the individual observations contribute to the conclusion regarding the null hypothesis of the score test, the $EIC_{score,k}$ given in Theorem 4.1 should be used. This diagnostic can be regarded as a joint influence measure since all parameters in the model are estimated from the perturbed model. The $EIC_{score,k}$

could be constructed as a marginal influence measure if the parameters in the model under the null hypothesis would be estimated from the unperturbed model.

In practice, if an observation is connected to a notably high absolute value of the EIC it is suspected that this observation is influential. If the $EIC-$value is positive it means that the presence of that observation increases the value of the parameter estimate(s) or the score test statistic. On the other hand, if the $EIC-$value is negative it means that the presence of that observation decreases the value of the parameter estimate(s) or the score test statistic. Moreover, if the absolute value of the EIC is substantially large, a removal of that observation from the analysis can change a significant result to a non-significant result, or vice versa.

Once the influential observations are identified, a search for the reason to their deviant influence value should be conducted. There are numerous reasons for existence of the influential observations; for example, invalid data entry, measurement error, invalid model assumptions, etc. How to deal with the influential observations depends on the cause of their presence. The emphasis in this article is, however, on identifying influential observations rather than how to deal with them once they are found.

Influence analysis is a complicated task. The EIC_k^J, $EIC_{j,k}^M$ and the $EIC_{score,k}$ only measure the influence of a single observation but the number of influential observations can be more than one. Some observations may influence the parameter estimates and/or the score test statistic mutually. If an influential observation is deleted another observation may emerge as extremely influential. This situation is referred to as masking. However, this work is not extended to discuss this possibility specifically. Further studies can be done to examine if the EIC-measures derived in this article are affected by masking. If it is the case, that the EIC is sensitive to masking, important observations may go unnoticed. Development of the measures can then be a future task and the question to be answered is how to make these measures robust against the masking effect.

EIC-measures can be extended to study the influence of several observations simultaneously. For example, in different applications this could be done in order to examine if experiments conducted on a particular day have a substantial influence on the inference.

6. Appendix - Matrix calculus

In this section rules for matrix differentiation from Kollo and von Rosen (2010) are presented.

Definition 6.1. *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* $\frac{\partial y_{kl}}{\partial x_{ij}}$ *exist, are continues in* A *and*

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} vec^T \mathbf{Y}, \qquad (28)$$

where

$$\frac{\partial}{\partial \mathbf{X}} = \left(\frac{\partial}{\partial x_{11}}, \ldots, \frac{\partial}{\partial x_{p1}}, \frac{\partial}{\partial x_{12}}, \ldots, \frac{\partial}{\partial x_{p2}}, \ldots, \frac{\partial}{\partial x_{1q}}, \ldots, \frac{\partial}{\partial x_{pq}}\right)^T.$$

Properties of the matrix derivative given in (28) is presented in the following table.

Table 1: Matrix differentiation

Differentiated function	Derivative
$\mathbf{Z} = \mathbf{Z}(\mathbf{Y}),\ \mathbf{Y} = \mathbf{Y}(\mathbf{X})$	$\frac{d\mathbf{Z}}{\mathbf{X}} = \frac{d\mathbf{Y}}{d\mathbf{X}}\frac{d\mathbf{Z}}{d\mathbf{Y}}$
$\mathbf{Y} = \mathbf{A}\boldsymbol{X}\mathbf{B}$	$\frac{d\mathbf{Y}}{d\boldsymbol{X}} = \mathbf{A} \otimes \mathbf{B}$
$\mathbf{Z} = \mathbf{AYB}$	$\frac{d\mathbf{Z}}{d\boldsymbol{X}} = \frac{d\mathbf{Y}}{d\boldsymbol{X}}(\mathbf{A} \otimes \mathbf{B})$
$\boldsymbol{W} = \mathbf{YZ},\ \mathbf{Z} \in \mathbb{R}^{s\times t}$	$\frac{d\mathbf{W}}{d\boldsymbol{X}} = \frac{d\mathbf{Y}}{d\boldsymbol{X}}(\mathbf{Z} \otimes \mathbf{I}_r) + \frac{d\mathbf{Z}}{d\boldsymbol{X}}\left(\mathbf{I}_t \otimes \mathbf{Y}^T\right)$
$\mathbf{W} = \mathbf{RYZ},\ \mathbf{R} \in \mathbb{R}^{p\times r}$	$\frac{d\mathbf{W}}{d\mathbf{X}} = \frac{d\mathbf{R}}{d\mathbf{X}}(\mathbf{YZ} \otimes \mathbf{I}_p) + \frac{d\mathbf{Y}}{d\mathbf{X}}\left(\mathbf{Z} \otimes \mathbf{R}^T\right) + \frac{d\mathbf{Z}}{d\mathbf{X}}\left(\mathbf{I}_t \otimes (\mathbf{RY})^T\right)$
$\mathbf{Y}^{-1}$	$\frac{d\mathbf{Y}^{-1}}{d\mathbf{X}} = -\frac{d\mathbf{Y}}{d\mathbf{X}}\left(\mathbf{Y}^{-1} \otimes \mathbf{Y}^{-1}\right)$

References

Bates, D. M. and Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* Wiley, New york.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* Wiley, New york.

Briggs, G. E. and Haldane, J. B. S. (1925). A note on the kinetics of enzyme action. *Biochemical journal*, **19**, 338.

Castillo, E., Castillo, C., Hadi, A. S. and Sarabia, J. M. (2008). Local sensitivity analysis in estimation problems. *Journal of Computational and Graphical Statistics*, **17**, 703–725.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity analysis in linear regression.* Wiley, New York.

Cleland, W. *et al.* (1967). The statistical analysis of enzyme kinetic data. *Advances in enzymology and related areas of molecular biology*, **29**, 1–32.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, **48**, 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression.* Chapman and Hall, New York.

Gallant, A. R. (1987). *Nonlinear statistical models.* Wiley, New York.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383–393.

Kollo, T. and von Rosen, D. (2010). *Advanced multivariate statistics with matrices.* Springer, Dordrecht.

Laurent, R. T. S. and Cook, R. D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika*, **80**, 99–106.

Lee, A. H., Xiang, L. and Fung, W. K. (2004). Sensitivity of score tests for zero-inflation in count data. *Statistics in medicine*, **23**, 2757–2769.

Lehmann, E. L. (1999). *Elements of large-sample theory.* Springer, New York.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses, 3rd edition.* Springer, New York.

Lustbader, E. D. and Moolgavkar, S. H. (1985). A diagnostic statistic for the score test. *Journal of the American Statistical Association*, **80**, 375–379.

Nocedal, J. and Wright, S. J. (2006). *Numerical optimization.* Springer, New York.

Pázman, A. (1993). *Nonlinear statistical models.* Kluwer, Dordrecht.

Rao, C. R. (1973). *Linear statistical inference and its applications.* Wiley, New York.

Rao, C. R. (2005). Score test: Historical review and recent developments. In *Advances in ranking and selection, multiple comparisons, and reliability.* 3–20, Birkhäuser, Boston.

Ritchie, R. J. and Prvan, T. (1996). Current statistical methods for estimating the k_m and v_{max} of michaelis-menten kinetics. *Biochemical Education*, **24**, 196–206.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global sensitivity analysis: the primer.* Wiley, Chichester.

Seber, G. and Wild, C. (1989). *Nonlinear regression.* Wiley, New York.

Stål, K. (2011). A graphical representation of the score test in regression. *JP Journal of Fundamental and Applied Statistics*, **1**, 1–14.

Tang, N. S., Wei, B. C. and Wang, X. R. (2002). Detection of influential observations in nonlinear reproductive dispersion models. *Australian & New Zealand Journal of Statistics*, **44**, 185–192.

Wei, B. (1995). Score statistic and its local influence analysis in weighted nonlinear regression. *Chinese Journal of Applied Probability and Statistics*, **11**, 147–156.