International Journal of Pure and Applied Mathematics

Volume 21 No. 3 2005, 1-27

ROBUST ESTIMATION AND TESTING IN NONLINEAR REGRESSION MODELS

Ruei-Che Liu[1], Marianthi Markatou[2], Chih-Ling Tsai[3]

[1]PsychoGenics, Inc., Hawthorne, NY 10532
e-mail: rl241@columbia.edu

[2]Department of Biostatistics
722 West 168th Street, Floor 6
Mailman School of Public Health
Columbia University, New York, NY 10032
e-mail: mm168@columbia.edu

[3]Graduate School of Management
One Shields Avenue
University of California
Davis, CA 95616
e-mail: cltsai@ucdavis.edu

Abstract: We study the performance of least squares estimators, robust estimators, and tests in nonlinear regression models under various contamination schemes of the error distribution. We also address the problem of obtaining appropriate initial values for the algorithms that compute the M-estimators. It is shown that a scheme proposed by Smyth in the context of least squares offers good initial values for the computation of M-estimators. The performance of the estimators is measured in terms of their bias and variance, while that of the tests is measured in terms of attaining the nominal level. It is shown that for symmetric contamination with relatively low variance, the nonlinear least squares estimators perform as well as the M-estimators in terms of bias, however their standard deviation is larger than that of the M-estimators. The M-estimators perform well under small levels of both symmetric and asymmetric contamination. In general, the performance of the estimators depends on the nonlinear regression function that is fitted, and the effect of certain types of contamination is more pronounced on the variance of the estimates rather

Received: May 23, 2005

than on their bias.

The tests based on M-estimators have a level close to the nominal level, even for relatively high percentages of symmetric contamination with relatively low variance. Moreover, for the cases studied, with the subhypothesis testing problem $\theta_2 = 0$, θ_1 unspecified, $\theta = (\theta_1^T, \theta_2^T)^T$, it is shown that the chi-squared with $p_2 = \dim(\theta_2)$ degrees of freedom fits the quantiles of the Wald test, while the $F(p_2, N - p)$ distribution fits the drop-in-dispersion test, where $p = \dim(\theta)$, N is the sample size.

AMS Subject Classification: —???—
Key Words: influence function, hypothesis testing, leverage, M-estimator, nonlinear regression, robustness

1. Introduction

The amount of statistical research devoted to robustness has increased in recent years. A lot of this effort is concentrated on robust estimation and, to a lesser extent, on robust testing. Moreover, most of this research addresses location-scale and linear regression models. General references include the books by Huber [8], Hampel, Ronchetti, Rousseeuw and Stahel [4], Staudte and Sheather [21], Morgenthaler and Tukey [13] and Stahel and Weisberg [20].

In this paper, we study the behavior of robust estimation and testing procedures in nonlinear regression, provide simulations that illustrate the performance of estimators and tests under various contamination schemes, and discuss the computational aspects of the robust procedures.

Data inadequacies are often caused by measurement or recording errors. To accommodate violation of the assumptions that such errors cause, robust procedures are often used. These procedures consider the behavior of statistical techniques in the neighborhood of the adopted model; robust techniques consider the model to be only an approximation of the real mechanism that generated the data.

There are two types of robust procedures; bounded influence and high breakdown point methods. Bounded influence procedures provide good stability and efficiency against infinitesimal contamination, but they are not globally stable. Therefore, bounded influence methods pertain to the local stability of estimators and tests. A considerable amount of literature is devoted to the development of robust bounded influence estimation and testing methods, with particular emphasis on the linear regression model. The amount of literature devoted to bounded influence procedures in nonlinear regression models is not

nearly as much. Relatively recent papers include Fraiman [2], Heritier and Ronchetti [7] and Markatou and Manos [10].

High breakdown point procedures are capable of handling multiple outliers in the data; thus, high breakdown point methods refer to the global stability of estimators and tests. In linear regression, high breakdown point estimators have been used to identify outliers and discover problems with masking that other diagnostic procedures miss (Rousseeuw [15]; Rousseeuw and van Zomeren [17]). In nonlinear regression, high breakdown point estimation methods are discussed by Stromberg and Ruppert [25] and Stromberg [22,23].

This paper is organized as follows. Section 2 presents the nonlinear regression model and discusses bounded influence methods. Section 3 discusses the corresponding testing procedures. The behavior of the estimates and tests in nonlinear regression differs from that in linear regression. In particular, it appears that the performance of the nonlinear least squares estimator and the M-estimator depends on the nonlinear regression function as well as the type of contamination. The Wald test statistic based on least squares estimators deteriorates faster than the corresponding statistic based on M-estimators. Simulation results show that the quantiles of the Wald test based on M-estimators are fitted by a χ^2_2 distribution, whereas those of the drop-in-dispersion test based on M-estimators are well approximated by the quantiles of an F distribution. Section 4 presents simulation results and discusses some of the computational issues associated with these procedures. Finally, Section 5 offers concluding remarks and further discussion on unsolved problems.

2. The Nonlinear Regression Model and Robust Estimation Methods

The nonlinear regression model has the form

$$y_i = f(x_i; \theta) + \epsilon_i, \ i = 1, 2, ..., n. \tag{2.1}$$

where the responses y_i are related to the predictors x_i via a known function f which is defined on $X \times \Theta, X \subseteq \mathcal{R}^k, \Theta \subseteq \mathcal{R}^p$ and depends on the unknown parameter vector θ. The errors are usually assumed to be independent, identically distributed random variables with a symmetric distribution $G(\epsilon_i/\sigma)$, $\sigma > 0$ is a scale parameter.

The unknown parameter vector θ is estimated from the data by minimizing a suitable goodness-of-fit expression with respect to θ. The most popular criterion

is the least squares criterion, and the estimates are obtained by

$$\min_{\theta} \Sigma_{i=1}^{n} [y_i - f(x_i; \theta)]^2. \tag{2.2}$$

An M-estimator in a nonlinear regression setting is obtained by minimizing over θ the expression

$$\Sigma_{i=1}^{n} \rho(\frac{y_i - f(x_i;\theta)}{\sigma}), \tag{2.3}$$

where ρ is defined on $\mathcal{R}$ and takes values in $\mathcal{R}^+$. An example of the function ρ is Huber's rho-function defined as

$$\rho_c(\frac{r}{\sigma}) = \left\{ \begin{array}{ll} \frac{1}{2}(\frac{r}{\sigma})^2, & \text{for } |\frac{r}{\sigma}| \leq c \\ c|\frac{r}{\sigma}| - \frac{c^2}{2}, & \text{for } |\frac{r}{\sigma}| > c \end{array} \right\}. \tag{2.4}$$

Here the argument r denotes the residual and c is a positive constant. The value of $c = 1.345$ guarantees a 95% efficiency in the normal model. Taking the derivative, with respect to θ, of (2.3), we can define an M-estimator in a nonlinear model setting by the equation

$$\Sigma_{i=1}^{n} \psi_c(\frac{y_i - f(x_i;T_n)}{\sigma}) \bigtriangledown_\theta f(x_i;T_n) = 0, \tag{2.5}$$

where ψ_c is the derivative of ρ evaluated at T_n. This function has the form $\psi_c(r) = \max(-c, \min(c, r))$ and is called Huber's psi-function. It bounds the residuals as it allows usage of them if they belong in the interval $[-c\hat{\sigma}, c\hat{\sigma}]$; otherwise, it replaces them with $\pm c\hat{\sigma}$. Here $\hat{\sigma}$ is a robust scale estimate of σ. The scale estimate we use in our simulation experiments is defined as $\hat{\sigma} = 1.4826\ med\{|r_i - \underset{j}{med} r_j|\}$, where r_i, $i = 1, 2, ..., n$ are the residuals computed from the least squares fit. We note here that one could also compute the scale estimate simultaneously as an M-estimator of scale (Huber [8], section 6.4). Using an initial robust scale estimate is computationally easier than the simultaneous estimation of the scale.

The influence of regression estimators in linear regression can be decomposed into two parts consisting of the influence due to residuals and the influence due to the position in the factor space. This second influence is generated, for example, from high leverage points. To bound the influence that such points have on the estimates of regression coefficients, the generalized M-estimators (GM-estimators) were introduced. In nonlinear regression GM-estimators were introduced by Fraiman [2]. The GM-estimators are defined by the equation

$$\Sigma_{i=1}^{n} \eta(x_i, \frac{y_i - f(x_i;\theta)}{\sigma}) \bigtriangledown_\theta f(x_i;\theta) = 0, \tag{2.6}$$

where $\eta : \mathcal{R}^k \times \mathcal{R} \to \mathcal{R}^+$ is a given function. Under appropriate conditions (see Fraiman [2] and Markatou and Manos [10]) the estimators obtained by solving the above estimating equation are consistent and asymptotically normal. Their asymptotic variance-covariance matrix is given as $\mathcal{V}=\mathcal{M}^{-1}\mathcal{Q}\mathcal{M}^{-1}$, where

$$\mathcal{M} = E\{\frac{\partial}{\partial\theta}[\eta(x, \frac{y - f(x;\theta)}{\sigma}) \bigtriangledown_\theta f(x;\theta)]\} \text{ and}$$

$$\mathcal{Q} = E\{\eta^2(x, \frac{y - f(x;\theta)}{\sigma})(\bigtriangledown_\theta f(x;\theta))(\bigtriangledown_\theta f(x;\theta))^T\}.$$

Some of the choices for the η-function are Huber's psi-function or functions of the form $\eta(x;r) = v(x)\psi_c(rs(x))$, $v : \mathcal{R}^k \to \mathcal{R}^+$ and $s : \mathcal{R}^k \to \mathcal{R}^+$. This formulation includes Mallows and Schweppe proposals. Recall that the Mallows proposal bounds separately the influence of residuals and the influence due to the position in the factor space, hence it uses $\eta(x;r) = v(x)\psi_c(\frac{r}{\sigma})$, whereas the Schweppe proposal bounds these two influences simultaneously, and it uses $\eta(x;r) = v(x)\psi_c(\frac{r}{\sigma s(x)})$.

A measure of local stability of a statistic is the influence function. The influence function describes the (approximate and standardized) effect of an additional observation in any point x, on a statistic T, given a (large) sample with distribution F. Thus, roughly speaking, the influence function is the first derivative of a statistic T at an underlying distribution F. Formally, it is defined as follows.

Definition. The influence function of a statistic T at a distribution function F and at the point x_0 is given by

$$IF(x_0; T, F) = \lim_{\varepsilon\to 0}\frac{T((1-\varepsilon)F + \varepsilon\Delta_{x_0}) - T(F)}{\varepsilon},$$

in those $x_0 \in \mathcal{X}$ where the limit exists.

In linear regression, the influence function of the least squares estimator of the parameter vector β is a function of the residuals as well as the explanatory variables. Thus, a data point that creates a large residual or an explanatory variable with extreme position in the factor space can unduly influence the value of the least squares estimator. To limit such influences, generalized M-estimation methods are used; these methods provide estimates with totally bounded influence function, therefore local stability is guaranteed.

In the nonlinear regression setting M-estimates are computed using equation (2.5). The influence function of these estimators is given in the following proposition.

Proposition. *The influence function of the estimators obtained by solving (2.5) at* (x_0, y_0) *is given by*

$$IF(x_0; \theta, F) = \psi_c(\frac{y_0 - f(x_0;\theta)}{\sigma}) A^{-1} \bigtriangledown_\theta f(x_0;\theta),$$

where A^{-1} *is the inverse of the matrix*

$$A = \int \psi_c'(\frac{y - f(x;\theta)}{\sigma})(\bigtriangledown_\theta f(x;\theta))(\bigtriangledown_\theta f(x;\theta))^T dF(x,y) \\ - \int \psi_c(\frac{y - f(x;\theta)}{\sigma})(\frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x;\theta)) dF(x,y).$$

Proof. Set $F_\varepsilon(x,y) = (1-\varepsilon)F(x,y) + \varepsilon\Delta_{(x_0,y_0)}$; the functional that induces the M-estimates is then $\int \psi_c(\frac{y-f(x;\theta(F))}{\sigma}) \bigtriangledown_\theta f(x;\theta(F)) dF(x,y)$. Evaluate that functional at F_ε; then taking derivative with respect to ε, and evaluating at $\varepsilon = 0$, produces the above expression for the influence function of the nonlinear M-estimator.

Note that when $f(x;\theta) = \beta^T x$, the influence function given above reduces to the influence function of an M-estimator for the linear model. We also note that the influence function is model dependent; thus, if the nonlinear function $f(x_0;\theta)$ has bounded derivatives with respect to the components of the vector θ, over x, then the M-estimators have a totally bounded influence. However, there are nonlinear models where the derivatives with respect to the various components of the parameter θ are unbounded, hence the M-estimators will no longer have totally bounded influence. Another difference from the linear model is that the influence function can not be written as a product of two terms, the influence due to residuals and the influence due to the position of the $x_i's$ in the factor space. For completeness, we note that high breakdown point estimators in nonlinear regression are studied by Stromberg and Ruppert [25] and Stromberg [23].

3. The Nonlinear Regression Model and Robust Testing

The aim of robust testing is two-fold. The level of the test should be stable under small, arbitrary departures from the null hypothesis, and the test should retain good power under small, arbitrary departures from specified alternatives.

Many authors have proposed robust tests in the linear regression setting, with considerably lesser activity in the nonlinear regression setting.

Heritier and Ronchetti [7] and Markatou and Manos [10] presented robust versions of the Wald, scores and drop-in-dispersion tests for general parametric models and nonlinear regression models specifically. Moreover, Markatou and Manos [10] discuss robust tests for heteroscedastic models.

Assume that $\theta = (\theta_1^T, \theta_2^T)^T$ is a partition of θ, where θ_i are $p_i \times 1$ vectors, $i = 1, 2$, $p_1 + p_2 = p$. We would like to test $H_0 : \theta_2 = 0$, θ_1 unspecified vs. $H_1 : \theta_2 \neq 0$, θ_1 unspecified.

The drop-in-dispersion test is defined as

$$S_n^2 = 2n^{-1}p_2^{-1}\sum_{i=1}^{n}\{\tau(x_i; \frac{y_i - f(x_i;\hat{\theta}_{H_0})}{\hat{\sigma}}) - \tau(x_i; \frac{y_i - f(x_i;\hat{\theta})}{\hat{\sigma}})\}, \tag{3.1}$$

where $\hat{\theta}_{H_0}$ and $\hat{\theta}$ are estimates of θ in the reduced and full model respectively, $\tau : \mathcal{R}^k \times \mathcal{R} \to \mathcal{R}^+$ is a real function, and $\hat{\sigma}$ is a robust estimate of scale.

The Wald-type test is defined as

$$W_n^2 = \hat{\theta}_2^T \hat{\mathcal{V}}_{22}^{-1} \hat{\theta}_2, \tag{3.2}$$

where $\hat{\mathcal{V}}_{22}^{-1}$ is the inverse of an estimated $p_2 \times p_2$ submatrix of the matrix $\mathcal{V}$=$\mathcal{M}^{-1}\mathcal{Q}\mathcal{M}^{-1}$, and $\mathcal{M}$ and $\mathcal{Q}$ are defined as in Section 2.

Finally, a score-type test is defined as

$$T_n^2 = n^{-1} S_2^T(\hat{\theta}_{H_0}) \hat{\mathcal{D}}^{-1} S_2(\hat{\theta}_{H_0}), \tag{3.3}$$

where $\hat{\mathcal{D}}^{-1}$ in the inverse of the estimated matrix $\mathcal{D}$, which is defined as $\mathcal{D} = \mathcal{Q}_{22} - \mathcal{M}_{21}\mathcal{M}_{11}^{-1}\mathcal{Q}_{12} - \mathcal{Q}_{21}\mathcal{M}_{11}^{-1}\mathcal{M}_{12} + \mathcal{M}_{21}\mathcal{M}_{11}^{-1}\mathcal{Q}_{11}\mathcal{M}_{11}^{-1}\mathcal{M}_{12}$, $\mathcal{M}_{ij}$, $\mathcal{Q}_{ij}$, are $p_i \times q_j$ submatrices of $\mathcal{M}$ and $\mathcal{Q}$ respectively, and their indices correspond to the parameter vectors θ_i, θ_j, $i = 1, 2$. Moreover,

$$S_2(\hat{\theta}_{H_0}) = \Sigma_{i=1}^{n} \eta(x_i; \frac{y_i - f(x_i;\hat{\theta}_{H_0})}{\hat{\sigma}}) \bigtriangledown_{\theta_2} f(x_i; \hat{\theta}_{H_0}). \tag{3.4}$$

Heritier and Ronchetti [7] and Markatou and Manos [10] proved that Wald- and score-type tests generally do not have the same asymptotic distribution as the drop-in-dispersion or likelihood -ratio type tests. This is the main difference, with respect to the corresponding classical results, where the three classes of tests are asymptotically equivalent under both the null hypothesis and continuous alternatives. In particular, the asymptotic distribution of the Wald-type and score-type tests under H_0 is a central chi-squared with p_2 degrees of freedom, and under H_1, is a non-central chi-squared with the same degrees of freedom. The asymptotic distribution of the drop-in-dispersion test is more

complicated. Under the null hypothesis, its distribution is a linear combination of differentially weighted chi-squared random variables with one degree of freedom; the weights are the eigenvalues of the matrix $K = Q[M^{-1} - \tilde{M}]$, where $\tilde{M} = \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$.

In the linear regression setting, Markatou and Hettmansperger [12] studied the possibility of constructing a practical alternative to the computation of the p-value of the drop-in-dispersion test statistic. Using results from perturbation theory, they studied the quality of the approximation of the asymptotic distribution of the drop-in-dispersion (τ-test) by the distribution of a chi-squared random variable with p_2 degrees of freedom multiplied by the average of the different eigenvalues. The techniques of Markatou and Hettmansperger [12] can be applied to the case of nonlinear regression. This is an area for further investigation, as it is not known under what conditions their results are applicable in the case of general parametric models.

The local stability of a test statistic can be measured via the influence function of the test. The influence function of a test statistic describes the effect of infinitesimal contamination at a point in the sample on the value of the (standardized) test statistic, and therefore on the decision (acceptance or rejection of H_0) based on this value. Analogously, one can define the level influence function and the power influence function that describe the influence of a small amount of contamination at some point on the asymptotic level and power of the test. The level and power influence functions of a test statistic are proportional to the influence function of the test statistic.

A second step in the robust analysis of testing procedures in nonlinear models is the study of high breakdown point tests as in He, Simpson and Portnoy [5], He [6], Markatou and He [11]. This is still an open problem for investigation.

4. Simulation Experiments

In this section we present simulations to illustrate the performance of the methods previously discussed. The aim of the simulation is to understand the behavior of both robust M-estimators and tests that are based on those estimators in the presence of various contamination schemes and nonlinear models. Additionally, we aim to study the role that the nonlinear function has on the performance of the nonlinear least squares estimates in the presence of contamination.

The first model is a separable nonlinear model that is given as

$$y_i = \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_4 e^{\theta_3 x_{3i}} + \varepsilon_i. \tag{4.1}$$

This model is discussed in Gallant [3]. We used the data provided in Gallant with the following alteration. Each row in the matrix $X_{30\times 3} = (X_1, X_2, X_3)$ was repeated three times, so the total sample size is 90. The true values of the parameter vector $(\theta_1, \theta_2, \theta_3, \theta_4)^T$ are $(0, 0, 1, 2)^T$, and the $y_i's$ are computed using (4.1), where ε_i are independent, identically distributed random variables randomly generated from a set of distributions.

The second model is the Michaelis-Menten model defined as

$$y_i = \frac{\beta_0 x_i}{x_i + e^{\beta_1}} + \varepsilon_i, \tag{4.2}$$

where the true value of $(\beta_0, \beta_1) = (10, 0)$ and the $X_i's$ are generated randomly from a $U(0, 10)$ distribution. This parametrization was used by Stromberg [23] and was suggested by Ratkowsky [14].

The third model is a time power model defined as

$$y_i = \alpha x_i^{\beta} + \varepsilon_i, \tag{4.3}$$

where α, β are the model parameters. We generated $x_i's$ from a $U(0, 10)$ distribution. This model does not increase as quickly as the exponential model, and has been considered one of the basic growth models. We generated 100 independent samples from models (4.2) and (4.3), and then computed the ordinary least squares estimator and the M-estimator.

All data were generated using S-plus. To fit the ordinary least squares we used the function 'nlregb' in Splus by supplying the starting values.

To investigate the dependence of the solution on the initial values, we used model (4.1) and generated the error from a $N(0, 1)$ distribution.

Table 1 presents the values of the nonlinear least squares estimators and the M-estimators for various starting points. The value of θ_3 was fixed, and starting values for θ_1, θ_2 and θ_4 in model (4.1) were obtained by regressing y onto X_1, X_2, $e^{\theta_3 X_3}$ (Smyth [19]). We note that when $\theta_3 = 0.1, 0.3, 0.5, 1$ the nonlinear least squares estimators have values close to the true parameter values. When $\theta_3 = 2$ the only parameter that is estimated accurately by least squares is θ_3. In contrast to the least squares estimators (OLSE), the M-estimators appear to be more sensitive to the starting values. As can be seen from table 1 when $\theta_3 = \{0.1, 0.5, 2\}$, the M-estimates do not converge to the true parameter values. Experiments carried out with contaminated data verified the greater sensitivity of the M-estimators to the starting values. Similar results were obtained when models (4.2) and (4.3) were used.

In general, for the separable nonlinear models, we use the following method to obtain starting values. For model (4.1) notice that, for any given value of

Starting Values	OLSE	M-estimator
$\theta_3 = 0.1$ $\theta_1 = -267.86$ $\theta_2 = -21362.28$ $\theta_4 = 15668.56$	$\hat{\theta}_1 = 0.4020$ $\hat{\theta}_2 = -0.1216$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 1.9993$	$\hat{\theta}_1 = -18.8545$ $\hat{\theta}_2 = -4.2561$ $\hat{\theta}_3 = 0.9886$ $\hat{\theta}_4 = 2.2042$
$\theta_3 = 0.3$ $\theta_1 = -463.47$ $\theta_2 = -5691.20$ $\theta_4 = 1700.07$	$\hat{\theta}_1 = 0.4010$ $\hat{\theta}_2 = -0.1217$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 1.9997$	$\hat{\theta}_1 = 0.4688$ $\hat{\theta}_2 = -0.1609$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 1.9997$
$\theta_3 = 0.5$ $\theta_1 = -395.93$ $\theta_2 = -2313.16$ $\theta_4 = 248.57$	$\hat{\theta}_1 = 0.4011$ $\hat{\theta}_2 = -0.1216$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 1.9997$	$\hat{\theta}_1 = -380.0851$ $\hat{\theta}_2 = -2258.158$ $\hat{\theta}_3 = 0.5746$ $\hat{\theta}_4 = 120.0793$
$\theta_3 = 1$ $\theta_1 = 0.39$ $\theta_2 = -0.16$ $\theta_4 = 2.01$	$\hat{\theta}_1 = 0.3902$ $\hat{\theta}_2 = -0.1585$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 2$	$\hat{\theta}_1 = 0.3901$ $\hat{\theta}_2 = -0.1584$ $\hat{\theta}_3 = 1$ $\hat{\theta}_4 = 2$
$\theta_3 = 2$ $\theta_1 = 431.89$ $\theta_2 = 1335.25$ $\theta_4 = 0.0001$	$\hat{\theta}_1 = 60.9511$ $\hat{\theta}_2 = 201.3876$ $\hat{\theta}_3 = 1.0894$ $\hat{\theta}_4 = 0.8349$	$\hat{\theta}_1 = 431.897$ $\hat{\theta}_2 = 1335.251$ $\hat{\theta}_3 = 2$ $\hat{\theta}_4 = 0.0009$

Table 1: OLSE and M-estimators for various starting values in model (4.1)

θ_3, values for θ_1, θ_2 and θ_4 can be obtained from the linear regression of Y on X_1, X_2 and $\exp(\theta_3 X_3)$. These parameters are conditionally linear (Bates and Watts [1]; Smyth [18]). Thus, for any given value of θ_3, the least squares estimators of θ_1, θ_2 and θ_4 are available in a closed-form expression involving θ_3. In this sense, the only nonlinear parameter in model (4.1) is θ_3. Since the true value of θ_3 is 1 we took $\theta_3 \in (0.5, 1.5)$. The grid used was 0.1. Therefore, for a given value of θ_3, in the interval $(0.5, 1.5)$, and a given sample i, we obtain the least squares estimates of θ_1, θ_2 and θ_4. Denote these by $\hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}$ and $\hat{\theta}_4^{(i)}$. Note that the estimators $\hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}$ and $\hat{\theta}_4^{(i)}$ depend on the value of the parameter θ_3. Since $\theta_3 \in (0.5, 1.5)$ and the step of the grid used is 0.1 we denote by $\theta_3(j)$ the value of θ_3 at the j^{th} point of the grid values. We then use as starting values $\hat{\theta}_1^{(i)}, \hat{\theta}_2^{(i)}$, $\theta_3(j)$ and $\hat{\theta}_4^{(i)}$ and compute M-estimators and OLSE for all j. The value of the objective function also depends on θ_3. For the i^{th} sample, denote the value of the objective function by $O_i(\theta_3(j)$. The final estimates of

Error Distribution	OLSE: starting values are the true values			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
N(0,1)	0.0156 (0.2392)	0.0148 (0.2071)	1 ($2.49e-5$)	1.9999 (0.0005)
0.95N(0,1)+0.05N(0,9)	−0.0128 (0.2435)	0.0056 (0.2000)	1 ($2.77e-5$)	1.9999 (0.0005)
0.90N(0,1)+0.10N(0,9)	0.0033 (0.2940)	−0.0061 (0.2358)	1 ($3.30e-5$)	1.9999 (0.0007)
0.85N(0,1)+0.15N(0,9)	−0.0472 (0.3299)	0.0287 (0.2539)	1 ($3.51e-5$)	1.9999 (0.0007)
0.80N(0,1)+0.20N(0,9)	0.0246 (0.3204)	−0.0053 (0.2426)	1 ($3.90e-5$)	1.9999 (0.0008)
0.75N(0,1)+0.25N(0,9)	−0.0244 (0.3664)	0.0065 (0.2652)	1 ($4.59e-5$)	1.9999 (0.0009)
0.70N(0,1)+0.30N(0,9)	−0.0355 (0.3932)	0.0248 (0.2666)	1 ($4.47e-5$)	1.9998 (0.0009)
0.60N(0,1)+0.40N(0,9)	−0.1019 (0.3912)	0.0483 (0.3285)	1 ($5.06e-5$)	1.9999 (0.0010)
0.50N(0,1)+0.50N(0,9)	−0.0100 (0.4520)	−0.0065 (0.3971)	0.99996 ($5.72e-5$)	2 (0.0011)

Table 2: Least squares estimator for model (4.1). Starting values are the true parameter values, $\theta_1 = \theta_2 = 0$, $\theta_3 = 1$, $\theta_4 = 2$. The numbers in parenthesis are the standard deviations of the estimates

sample i are those for which we have $\min_j O_i(\theta_3(j))$. This procedure is repeated for $i = 1, 2, ..., 100$, and the average estimates are calculated together with their standard deviation. This scheme for selecting starting values was proposed by Smyth [18] in the context of computing least squares estimators. We extend its use in the context of M-estimation by using it to obtain good initial least squares estimators as starting values in an M-estimation algorithm. The true values of the parameters were used as additional starting values in the calculation of both OLSE and M-estimators; in addition, for M-estimators, we initialize the algorithm using the nonlinear least squares estimators. We present detailed results for model (4.1); the results obtained for models (4.2) and (4.3) are similar to those presented here, and are discussed briefly.

Table 2 presents the least squares estimators when the starting values are the true parameter values for the parameters in model (4.1) under various error distributions. We notice that, for symmetric contamination of the error distribution, the least squares estimators have little bias under contamination but

Error Distribution	OLSE: starting values are the true values			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
0.95N(0,1)+0.05N(3,1)	−0.0163 (0.2480)	0.1515 (0.1979)	0.9999 ($2.95e-5$)	2 (0.0006)
0.90N(0,1)+0.10N(3,1)	−0.0178 (0.2893)	0.3074 (0.2365)	0.9999 ($3.51e-5$)	2 (0.0007)
0.85N(0,1)+0.15N(3,1)	−0.0396 (0.3188)	0.4553 (0.2608)	0.9999 ($3.58e-5$)	2 (0.0007)
0.80N(0,1)+0.20N(3,1)	−0.0040 (0.3045)	0.5982 (0.2296)	0.9999 ($3.75e-5$)	2 (0.0007)
0.75N(0,1)+0.25N(3,1)	−0.0093 (0.3634)	0.7613 (0.2424)	1 ($4.16e-5$)	2 (0.0008)
0.70N(0,1)+0.30N(3,1)	−0.0008 (0.3840)	0.9249 (0.2665)	1 ($4.21e-5$)	1.9999 (0.0008)
0.60N(0,1)+0.40N(3,1)	−0.0487 (0.4080)	1.2522 (0.2978)	1 ($4.53e-5$)	1.9998 (0.0009)
0.50N(0,1)+0.50N(3,1)	−0.0015 (0.3900)	1.4862 (0.3213)	1 ($4.5e-5$)	2 (0.0009)

Table 3: Least squares estimators for the parameters in model (4.1). Starting values are the true parameter values. The numbers in parenthesis are the standard deviations of the estimates

their standard deviation approximately doubles as the percentage of contamination increases. In particular, the nonlinear parameter θ_3 and the parameter θ_4 are estimated almost without bias and accurately; their standard deviations are rather small across all error distributions. The linear parameters θ_1, θ_2 are estimated with low bias, but their variance increases as the amount of contamination of the error distribution increases.

We observe similar behavior, under symmetric contamination, of the ordinary least squares estimators of the parameters of the Michaelis-Menten model. Both parameter estimates exhibit low bias, however the standard deviation becomes large as the amount of contamination increases. For contamination with higher variance than is presented here, the least squares estimates become unreliable in terms of both bias and variance for relatively small amounts of contamination. The same behavior was observed for model (4.3); when the standard deviation of the symmetric contamination was increased, the least squares estimate exhibits high bias and standard deviation. For that model, reliable results were obtained for contamination up to 5%, and for normal contaminating distribution with standard deviation of 5 and 6.

Table 3 presents the least squares estimators for the parameters in model

Error Distribution	M-estimators			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
N(0,1)	0 $(9.15e-13)$	0 $(1.16e-11)$	1 $(1.24e-6)$	1.9999 $(6.46e-8)$
0.95N(0,1)+0.05N(0,9)	-0 $(7.05e-12)$	0 $(1.05e-11)$	0.9999 $(1.29e-6)$	2 $(6.67e-8)$
0.90N(0,1)+0.10N(0,9)	0 $(8.16e-12)$	0 $(1.33e-11)$	1 $(1.4e-6)$	2 $(7.25e-8)$
0.85N(0,1)+0.15N(0,9)	0 $(9.16e-12)$	0 $(1.27e-11)$	1 $(1.53e-6)$	2 $(7.89e-8)$
0.80N(0,1)+0.20N(0,9)	0 $(8.81e-12)$	0 $(1.27e-11)$	0.9999 $(1.43e-6)$	1.9999 $(7.4e-8)$
0.75N(0,1)+0.25N(0,9)	0 $(8.74e-12)$	0 $(1.27e-11)$	0.9999 $(1.54e-6)$	1.9999 $(8.05e-8)$
0.70N(0,1)+0.30N(0,9)	0 $(1.03e-11)$	0 $(1.3e-11)$	0.9999 $(1.64e-6)$	1.9998 $(8.55e-8)$
0.60N(0,1)+0.40N(0,9)	0 $(1.02e-11)$	0 $(1.49e-11)$	0.9999 $(1.69e-6)$	1.9999 $(8.69e-8)$
0.50N(0,1)+0.50N(0,9)	0 $(1.47e-11)$	0 $(1.86e-11)$	0.9999 $(2.14e-6)$	1.9999 $(1.12e-7)$

Table 4: M-estimators for the parameters in model (4.1). Starting values are the true parameter values, $\theta_1 = \theta_2 = 0$, $\theta_3 = 1$, $\theta_4 = 2$. The numbers in parenthesis are the standard deviations of the estimates

(4.1) when the errors are asymmetrically contaminated and the starting values are the true parameter values. In this case, the nonlinear parameter θ_3 and the parameter θ_4 are estimated accurately and with low bias. The other two linear parameters have standard deviations that increase as the amount of contamination increases. In particular, for moderate and large amounts of contamination of the error distribution, the estimate for θ_2 is unreliable.

The effect of asymmetric contamination of the error distribution on the parameter estimates of the Michaelis-Menten model is increased bias for both least squares parameter estimates, as well as increased variance. For example, when the amount of contamination is 10%, $\hat{\theta}_1 = 10.1879$ (with standard deviation 0.3541) and $\hat{\theta}_2 = -0.1173$ (with standard deviation 0.1904). Similar behavior was observed for model (4.3), where both parameter estimates were biased.

Table 4 and 5 present the M-estimators of the parameters of model (4.1) under symmetric and asymmetric error contamination. The starting values are the true parameter values. Comparing tables 4 and 5 with table 2 and 3 the advantage of M-estimators in terms of both, bias and efficiency is obvious. The M-estimators overall out perform the OLSE on both accounts and for both

Error Distribution	M-estimators			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
0.95N(0,1)+0.05N(3,1)	0 $(8.23e-12)$	0 $(1.15e-11)$	1 $(1.39e-6)$	2 $(7.21e-8)$
0.90N(0,1)+0.10N(3,1)	0 $(9.95e-12)$	0 $(1.37e-11)$	1 $(1.60e-6)$	2 $(8.32e-8)$
0.85N(0,1)+0.15N(3,1)	0 $(1.14e-11)$	0 $(1.78e-11$	1 $(1.95e-6)$	2 $(1.01e-7)$
0.80N(0,1)+0.20N(3,1)	0 $(1.37e-11)$	0 $(2.16e-11)$	1 $(2.22e-6)$	2 $(1.16e-7)$
0.75N(0,1)+0.25N(3,1)	0 $(1.76e-11)$	0 $(2.75e-11)$	1 $(2.70e-6)$	2 $(1.42e-7)$
0.70N(0,1)+0.30N(3,1)	0 $(1.97e-11)$	0 $(3.22e-11)$	1 $(2.82e-6)$	2 $(1.48e-7)$
0.60N(0,1)+0.40N(3,1)	0 $(2.37e-11)$	0 $(4.11e-11)$	1 $(3.17e-6)$	2 $(1.68e-7)$
0.50N(0,1)+0.50N(3,1)	0 $(2.04e-11)$	0 $(3.95e-11)$	1 $(2.90e-6)$	2 $(1.55e-7)$

Table 5: M-estimators for the parameters in model (4.1). Starting values are the true parameter values. The numbers in parenthesis are the standard deviations of the estimates

contamination schemes. However, when the starting values are the OLSE the advantage of M-estimators deteriorates, as they perform essentially as the least squares estimators. Therefore, it is imperative that a careful selection of the starting values is needed in order to gain advantages by using M-estimators. To select starting values for use in the M-estimation algorithm we recommend to use the scheme suggested by Smyth.

Tables 6 to 9 present the least squares estimators and the M-estimators for the parameters of model (4.1), when starting values are obtained using the Smyth procedure. When the contamination of the error distribution is symmetric, both the least squares estimators and the M-estimators have low bias; however, the standard deviation of the least squares estimators is higher than that of the M-estimators. When the contamination of the error distribution is asymmetric, the bias of the M-estimators of the parameter θ_2 is considerably lower than that of the corresponding least squares estimator, so that for the error distribution 0.95N(0,1)+0.05N(3,1) the M-estimators have very low bias and low standard deviation. The bias of the M-estimators of the parameter θ_1 is comparable to the least squares estimate for percentage of contamination up to 15%; for higher percentages of contamination, the bias of the M-estimator of θ_1 is low, however, in certain cases it is higher than that of the corresponding

Error Distribution	OLSE			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
N(0,1)	0.1559 (0.2392)	0.0148 (0.2071)	1 $(2.49e-5)$	1.9999 (0.0005)
0.95N(0,1)+0.05N(0,9)	−0.0129 (0.2435)	0.0056 (0.2000)	1 $(2.76e-5)$	1.9999 (0.0005)
0.90N(0,1)+0.10N(0,9)	0.0034 (0.2940)	−0.0061 (0.2358)	1 $(3.30e-5)$	1.9997 (0.0006)
0.85N(0,1)+0.15N(0,9)	−0.0473 (0.3299)	−0.0053 (0.2426)	1 $(3.51e-5)$	1.9999 (0.0007)
0.80N(0,1)+0.20N(0,9)	0.0246 (0.3203)	0.0065 (0.2652)	1 $(3.90e-5)$	1.9999 (0.0008)
0.75N(0,1)+0.25N(0,9)	−0.0244 (0.3664)	0.0065 (0.2652)	1 $(4.59e-5)$	1.9997 (0.0009)
0.70N(0,1)+0.30N(0,9)	−0.0355 (0.3932)	0.0248 (0.2665)	1 $(4.47e-5)$	1.9998 (0.0009)
0.60N(0,1)+0.40N(0,9)	−0.1020 (0.3912)	0.0483 (0.3285)	1 $(5.06e-5)$	1.9995 (0.0010)
0.50N(0,1)+0.50N(0,9)	−0.0100 (0.4520)	−0.0066 (0.3971)	0.9999 $(5.72e-5)$	2 (0.0011)

Table 6: Least squares estimators for the parameters in model (4.1). Starting values are obtained according to the scheme by Smyth. The numbers in parenthesis are the standard deviations of the estimates

least squares estimator. The bias of the M-estimator of θ_2 is consistently lower than the bias of the least squares estimate. When the amount of contamination increases, both estimators have unsatisfactory performance.

On the other hand, the performance of the M-estimates in the Michaelis-Menten model is different when OLSE is used as a starting value, with greater advantages in the case of asymmetric contamination. When the error distribution is 0.9N(0,1)+0.1N(3,1), the least squares estimates and their standard deviations are $\hat{\theta}_1 = 10.1879$ (0.354) and $\hat{\theta}_2 = -0.118$ (0.19). On the other hand the M-estimates are $\hat{\theta}_{1,M} = 10.11$ (0.327) and $\hat{\theta}_{2,M} = -0.08$ (0.17).The M-estimates improve upon the least squares estimates, and exhibit smaller bias and considerably smaller standard derivations.

In summary, it appears that, for the models studied and when the contamination of the error distribution is symmetric with relatively low variance, the models are flexible enough so that the least squares estimates obtained have low bias. When the standard deviation of the contaminating distribution increases, the bias of the least squares estimators increases as the percentage of

Error Distribution	M-estimators			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
N(0,1)	0.0183 (0.2496)	0.0093 (0.2105)	1 ($2.45e-5$)	2 (0.0005)
0.95N(0,1)+0.05N(0,9)	−0.0078 (0.2248)	0.0086 (0.1798)	1 ($2.37e-5$)	2 (0.0005)
0.90N(0,1)+0.10N(0,9)	−0.0063 (0.2442)	0.0058 (0.1968)	1 ($2.66e-5$)	2 (0.0005)
0.85N(0,1)+0.15N(0,9)	−0.0356 (0.2619)	0.0203 (0.1957)	1 ($2.94e-5$)	2 (0.0006)
0.80N(0,1)+0.20N(0,9)	0.0104 (0.2408)	0.0003 (0.1976)	1 ($3.10e-5$)	2 (0.0006)
0.75N(0,1)+0.25N(0,9)	−0.0205 (0.2747)	0.0132 (0.2115)	1 ($3.51e-5$)	2 (0.0007)
0.70N(0,1)+0.30N(0,9)	−0.0223 (0.3079)	0.0197 (0.2081)	1 ($3.23e-5$)	2 (0.0006)
0.60N(0,1)+0.40N(0,9)	−0.0815 (0.3141)	0.0337 (0.2643)	1 ($4.20e-5$)	2 (0.0008)
0.50N(0,1)+0.50N(0,9)	−0.0247 (0.3826)	0.0087 (0.3417)	0.9999 ($4.97e-5$)	2 (0.0010)

Table 7: M-estimators for the parameters in model (4.1). Starting values are obtained according to the scheme by Smyth. The numbers in parenthesis are the standard deviations of the estimates

contamination increases. This bias also increases for small amounts of contamination if the variance of the contaminating distribution is large. When the contamination of the error distribution is asymmetric, the least squares estimates exhibit higher bias than the corresponding M-estimators. In comparing the performance of the estimators for models (4.1), (4.2) and (4.3) we notice that there is a difference in M-estimators when starting values are the least squares estimates. The M-estimates for the Michaelis-Menten parameters are not similar to least squares, as in the case for model (4.1). This may be due to the nonlinear function used.

Next, we study the performance of the Wald and drop-in-dispersion tests in terms of their level under symmetric and asymmetric contamination. The nominal level was set at 0.05; the level was obtained by comparing the value of the test statistic with the 95^{th}percentile of a chi-squared random variable with 2 degrees of freedom. The results are presented for model (4.1) with two different sample sizes, n=90 and n=900. The original design matrix was of dimension 30×3 and the rows were repeated 3 and 30 times respectively to create the

Error Distribution	OLSE			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
0.95N(0,1)+0.05N(3,1)	−0.0163 (0.2480)	0.1515 (0.1979)	0.9999 ($2.95e-5$)	2 (0.0006)
0.90N(0,1)+0.10N(3,1)	−0.0178 (0.2893)	0.3074 (0.2365)	0.9999 ($3.52e-5$)	2 (0.0007)
0.85N(0,1)+0.15N(3,1)	−0.03958 (0.3188)	0.4539 (0.2608)	0.9999 ($3.59e-5$)	2 (0.0007)
0.80N(0,1)+0.20N(3,1)	−0.0040 (0.3045)	0.5983 (0.2295)	0.9999 ($3.75e-5$)	2 (0.0007)
0.75N(0,1)+0.25N(3,1)	−0.0093 (0.3634)	0.7613 (0.2423)	1 ($4.15e-5$)	1.9999 (0.0008)
0.70N(0,1)+0.30N(3,1)	−0.0008 (0.3840)	0.9248 (0.2665)	1 ($4.21e-5$)	1.9998 (0.0008)
0.60N(0,1)+0.40N(3,1)	−0.04878 (0.4080)	1.2522 (0.2980)	1 ($4.53e-5$)	1.9998 (0.0009)
0.50N(0,1)+0.50N(3,1)	0.0015 (0.3901)	1.4862 (0.3214)	0.9999 ($4.50e-5$)	2 (0.0009)

Table 8: Least squares estimators for the parameters in model (4.1). Starting values are obtained according to the scheme by Smyth. The numbers in parenthesis are the standard deviations of the estimates

two different samples. The errors were generated from the listed contaminating distribution and the true values of the vector θ are as before, (0,0,1,2). The hypothesis tested is $H_0 : \theta_1 = \theta_2 = 0$, θ_3, θ_4, unspecified.

Figures 1 and 2 plot the quantiles of a chi-squared distribution with 2 degrees of freedom versus the empirical quantiles of the Wald test when based on the OLSE (in figure 1) or on the M-estimator (in figure 2). As it can be seen from these figures, the chi-squared approximation for the M-estimator-based Wald test works relatively well. Further simulations that used the M-estimator-based Wald test for testing the hypothesis $H_0 : \theta_1 = \theta_2 = 0$, θ_3, θ_4, unspecified in model (4.1), showed that the chi-squared approximation held for higher amounts of symmetric contamination of the error distribution, approximately until about 20%. When the variance of the contaminating normal distribution is 16, the chi-squared approximation for the M-estimator-based Wald test held up to 10% of contamination; generally, the higher the variance of the contaminating distribution, the smaller the amount of contamination that could be tolerated.

Table 10 presents the level of the Wald test for a variety of symmetric and asymmetric error contaminating distributions. When the Wald test is based

Error Distribution	M-estimators			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
0.95N(0,1)+0.05N(3,1)	−0.0147 (0.2322)	0.0946 (0.1784)	0.9999 ($4.97e-5$)	2 (0.0005)
0.90N(0,1)+0.10N(3,1)	−0.0218 (0.2701)	0.2120 (0.2122)	0.9999 ($2.61e-5$)	2 (0.0006)
0.85N(0,1)+0.15N(3,1)	−0.0404 (0.3170)	0.3354 (0.2485)	0.9999 ($3.21e-5$)	2 (0.0006)
0.80N(0,1)+0.20N(3,1)	−0.0191 (0.3115)	0.4787 (0.2430)	0.9999 ($3.43e-5$)	2 (0.0007)
0.75N(0,1)+0.25N(3,1)	−0.0046 (0.3840)	0.6502 (0.2596)	0.9999 ($3.73e-5$)	1.9999 (0.0008)
0.70N(0,1)+0.30N(3,1)	−0.0024 (0.4038)	0.8451 (0.2913)	1 ($4.24e-5$)	1.9998 (0.0008)
0.60N(0,1)+0.40N(3,1)	−0.0502 (0.4317)	1.2225 (0.3197)	1 ($4.42e-5$)	1.9999 (0.0009)
0.50N(0,1)+0.50N(3,1)	−0.0012 (0.4108)	1.4843 (0.3388)	1 ($4.74e-5$)	2 (0.00113)

Table 9: M-estimators for the parameters in model (4.1). Starting values are obtained according to the scheme by Smyth. The numbers in parenthesis are the standard deviations of the estimates

on the nonlinear least squares estimates, and for symmetric contamination, the level of the test inflates as the percentage of contamination increases. In contrast, the level of the test that uses the M-estimators holds quite well. When the contamination is asymmetric, the level of the Wald test based on the nonlinear least squares estimators is inflated considerably; on the other hand, the Wald test based on M-estimators holds its level for small amounts of asymmetric contamination with mean less than or equal to three standard deviations away from the mean of the central distribution. In any other case the level is inflated, but this inflation is considerably smaller than the least squares-based level of the test.

The behavior of this test can be explained by looking at the estimation results. The least squares estimator of the parameter θ_2 breaks down quickly under asymmetric contamination; the corresponding M-estimator also breaks down, but at a much slower rate. Moreover, the covariance matrix that enters in the computation of the Wald test breaks down, and thus causes the inflation of the level of the test. We examined the performance of the Wald test in terms of level as a function of the sample size. When the sample size is 900 then the Wald test achieves a 4% level for the error distribution $N(0,1)$,

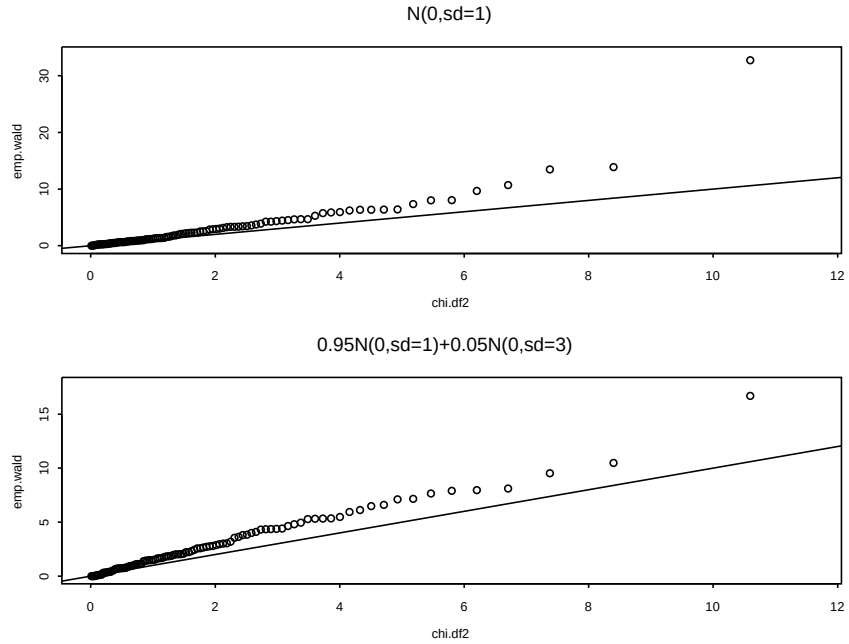

Figure 1: Quantile plots of the Wald test based on nonlinear least squares estimators. Data are from N(0,1) and 0.95N(0,1)+0.05N(0,9).

$0.95N(0,1) + 0.05N(0,9)$, $0.90N(0,1) + 0.10N(0,9)$. When the error distribution is $0.85N(0,1) + 0.15N(0,9)$ then the level of the Wald test becomes 0.08.

To evaluate the performance of a test across a range of nominal sizes Lloyd [9] recommended plotting the estimated levels versus a range of practical levels of significance. We therefore plotted the nominal levels of significance up to 0.2 level versus the estimated levels of significance of the Wald test that is based on M-estimators for the error distributions $0.95N(0,1)+0.05N(0,9)$ and $0.90N(0,1) + 0.10N(0,9)$, and for sample sizes n=90 and 900.

Figure 4 shows that when the sample size is 90 and the error distribution is $0.95N(0,1) + 0.05N(0,9)$ the Wald test based on M-estimators slightly overestimated the nominal level of significance with the largest discrepancy observed for nominal sizes in the interval $[0.05, 0.12]$. For the same error distribution and when the sample size is 900 the agreement between the nominal and the estimated levels of significance is almost perfect for all practical nominal sizes used. When the error distribution is $0.90N(0,1) + 0.10N(0,9)$ for sample size n=90

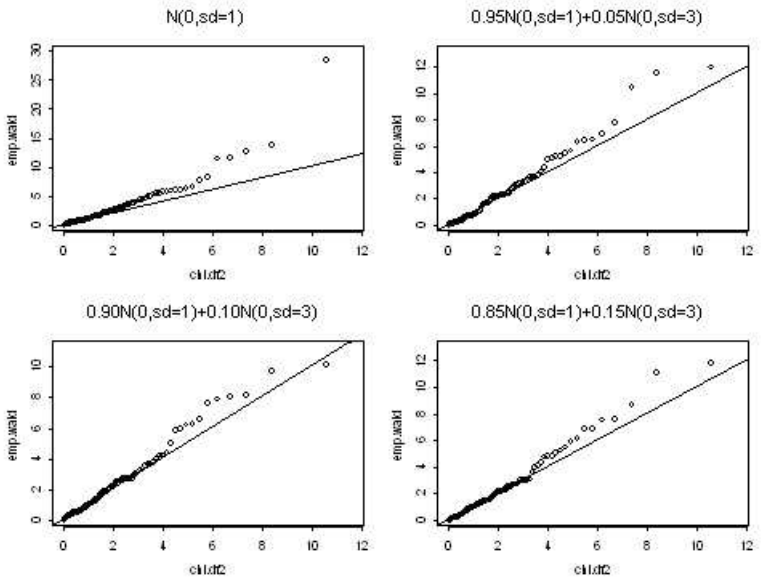

Figure 2: Quantile plots of Wald test based on M-estimators. Data are generated from N(0,1), 0.95N(0,1)+0.05N(0,9), 0.90N(0,1)+0.10N(0,9) and 0.85N(0,1)+0.15N(0,9)

the largest discrepancy is observed for nominal sizes in the interval $[0.05, 0.10]$, while when the sample size is 900, for the same error distribution, the largest discrepancy is observed in the interval $[0.15, 0.20]$. Therefore, the Wald-test based on M-estimators performs rather well in terms of level across a number of nominal sample sizes, exhibiting almost perfect performance across all practical nominal sizes for rather large samples and for relatively low (less than or equal to 10%) contamination levels. If one is willing to restrict the range of levels of significance, then the Wald-test based on M-estimators performs well for up to slightly over 10% contamination level.

We also computed the levels of the drop-in-dispersion test for symmetric and asymmetric contamination. Table 11 presents the level of the likelihood ratio test based on least squares, and the drop-in-dispersion test based on M-estimates for testing $H_0 : \theta_1 = \theta_2 = 0$, θ_3, θ_4, unspecified for model (4.1). To obtain the level of the likelihood ratio test, the cutoff of a chi-squared distribution with 2 degrees of freedom was compared with the value of the likelihood statistic. Note that when the variance of the symmetric contaminating distribution is relatively low and the percentage of contamination is relatively small

Error Distribution	Level of Wald test (based on OLSE)	Level of Wald test (based on M-estimators)
N(0,1)	0.08	0.11
0.95N(0,1)+0.05N(0,9)	0.09	0.08
0.90N(0,1)+0.10N(0,9)	0.14	0.09
0.85N(0,1)+0.15N(0,9)	0.14	0.08
0.80N(0,1)+0.20N(0,9)	0.14	0.06
0.95N(0,1)+0.05N(0,16)	0.16	0.08
0.90N(0,1)+0.10N(0,16)	0.23	0.09
0.85N(0,1)+0.15N(0,16)	0.24	0.10
0.95N(0,1)+0.05N(0,25)	0.20	0.07
0.90N(0,1)+0.10N(0,25)	0.29	0.09
0.85N(0,1)+0.15N(0,25)	0.33	0.10
0.98N(0,1)+0.02N(3,1)	0.10	0.08
0.95N(0,1)+0.05N(3,1)	0.11	0.09
0.90N(0,1)+0.10N(3,1)	0.38	0.16

5. Levels of the Wald test based on nonlinear least squares estimates and M-estimates. Sample size n=90

(up to about 8%), then the likelihood ratio test holds its level. If the amount of contamination increases, and if the variance of the symmetric contaminating distribution increases, the level of the likelihood test based on the least squares estimates inflates. The level also inflates when the error contamination is asymmetric. On the other hand, table 11 shows that the drop-in-dispersion test that is based on M-estimator underestimates the level as the amount of symmetric contamination increases. In contrast to the likelihood test, the level of the M-estimator based test holds well for small amounts of asymmetric contamination. Our experiments indicate that the level holds for up to about 8% asymmetric contamination with mean less than or equal to 3 standard deviations away from 0.

Figure 5 plots the quantiles of the drop-in-dispersion test versus those of a chi-squared distribution with 2 degrees of freedom (left panel), and those of an $F_{p_2,N-p}$ distribution (right panel). In our case $N = 90$, $p = 4$ and $p_2 = 2$, thus we employ the $F_{2,86}$ distribution. We see that the theoretical chi-squared approximation underestimates the level of the test as the percentage of symmetric contamination increases. This can also be seen from table 11. On the other hand, the $F_{p_2,N-p}$ distribution appears to provide a better fit to

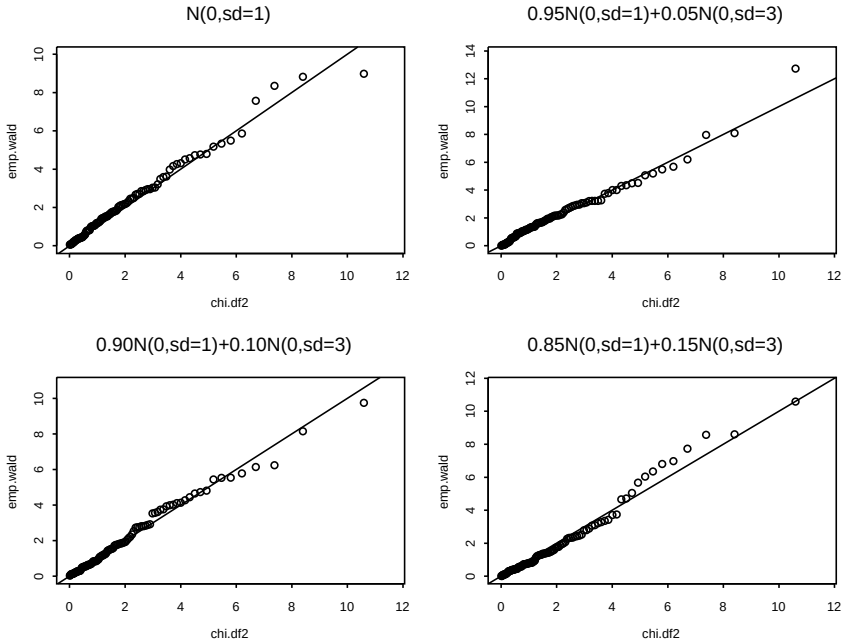

Figure 3: Quantile plots of Wald test based on M-estimators against the χ^2_2 quantiles. Sample size n=900

the quantiles of the drop-in-dispersion test for percentages of contamination as high as 15%. Similar results were obtained for the Michaelis-Menten model.

6. Summary

We studied the behavior of OLSE and Huber's M-estimators in nonlinear regression models and presented simulations that demonstrated their behavior under various contamination schemes for certain partially linear models and the Michaelis-Menten model. The results here indicate that the Smyth procedure provides good starting values for the computation of M-estimators. The results also indicate that M-estimators are an useful alternative to OLSE under both, symmetric and asymmetric contamination, as they are much more efficient than the corresponding OLSE and have lower bias.

We also studied the behavior of the Wald and drop-in-dispersion type tests in those models. It was found that a chi-squared distribution with two degrees of freedom is a reasonable fit to the Wald test quantiles for the models tested

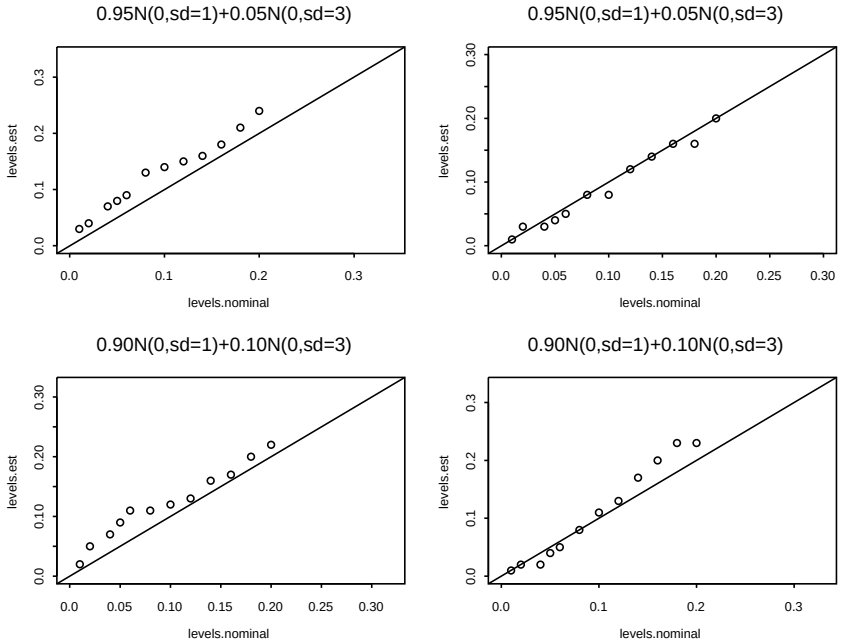

Figure 4: Plots of nominal size versus the estimated true size of the Wald test based on M-estimators. Left panel uses n=90, right panel uses n=900

here. This contrasts with the result in Markatou and Manos [10], where a different form of a separable model was used, indicating that the empirical performance depends on the regression function. The models used seem to be flexible enough so that even with high percentages of symmetric contamination, as long as the variance of the contaminating distribution is relatively low, the bias of the M-estimator and of the nonlinear least squares estimators is low.

One reason that robust methods are often difficult to work with is the lack of affordable, easy to use, software. Here, we worked with M-estimators as we were able to easily write programs in S-plus whenever we could not use the S-plus function directly. Stromberg [24] addresses this issue by providing instructions, examples, and code for using Fortran and LISPSTAT to compute MM-estimates and other robust regression estimates. Stromberg [24] describes how to compute several high breakdown estimators in linear regression, as well as the LMS and MM-estimators in nonlinear regression.

An aspect of the robust analysis that has not been dealt with in this paper

Error Distribution	Level of likelihood ratio test (based on OLSE)	Level of drop-in-disperson test (based on M-estimators)
N(0,1)	0.06	0.05
0.95N(0,1)+0.05N(0,4)	0.05	0.04
0.90N(0,1)+0.10N(0,4)	0.08	0.03
0.85N(0,1)+0.15N(0,4)	0.08	0.02
0.95N(0,1)+0.05N(0,9)	0.05	0.04
0.90N(0,1)+0.10N(0,9)	0.07	0.01
0.98N(0,1)+0.02N(3,1)	0.08	0.04
0.95N(0,1)+0.05N(3,1)	0.10	0.04
0.90N(0,1)+0.10N(3,1)	0.32	0.03

Table 10: Levels of the likelihood ratio test based on nonlinear least squares estimates and the drop-in-disperson test based on M-estimates. Sample size n=90

is the performance of estimators and tests in the presence of leverage points. St. Laurent and Cook [26,27] discuss several measures of leverage constructed from first- and second-order approximation to the nonlinear response function. This aspect of the research will open an avenue for future research.

Acknowledgement

The work of the second author was partially supported by NSF grant DMS-00-72310.

References

[1] D.M. Bates and D.G. Watts, *Nonlinear Regression Analysis and Its Application*, Wiley, New York (1988).

[2] R. Fraiman, General M-estimators and applications to bounded influence estimation for nonlinear regression, *Communications in Statistics, Theory and Methods*, **12** (1983), 2617-2631.

[3] A.R. Gallant, Nonlinear regression, *The American Statistician*, **29** (1975), 73-81.

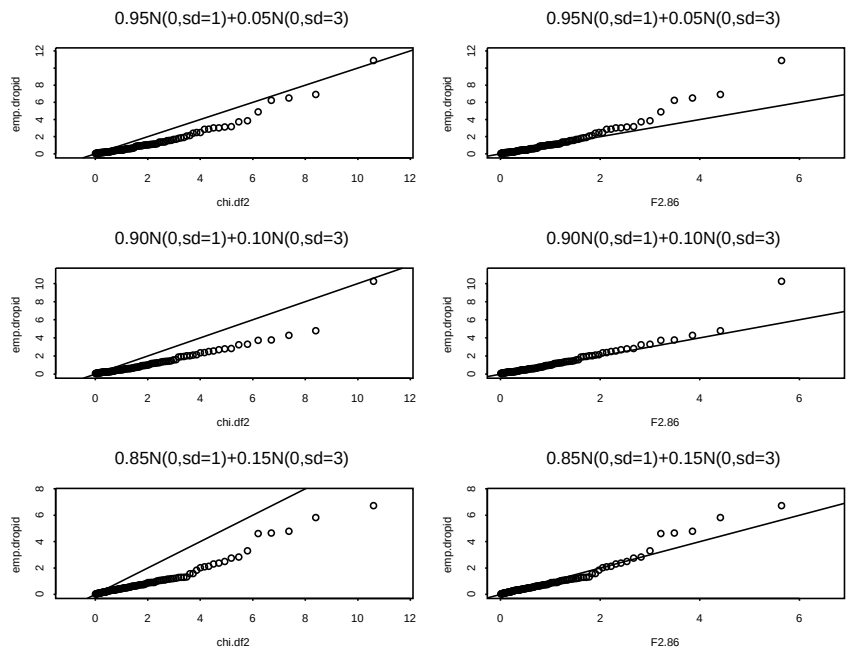

Figure 5: Quantile plots for the drop-in-dispersion test based on M-estimators against a χ^2_2 distribution (left) and an $F_{2,86}$ distribution (right)

[4] F.R. Hampel, E.M. Ronchetti, P.I. Rousseeuw, and W.A. Stahel, *Robust Statistics: The Approach based on Influence Function*, John Wiley, New York (1986)..

[5] X. He, D.G. Simpson, and S.L. Portnoy, Breakdown robustness of tests, *Journal of the American Statistical Association,* **85** (1990), 446-452.

[6] X. He, A local breakdown property of robust tests in linear regression, *Journal of Multivariate Analysis*, **38** (1991), 294-305.

[7] S. Heritier and E. Ronchetti, Robust bounded-influence tests in general parametric models, *Journal of the American Statistical Association*, **89** (1994), 897-904.

[8] P.J. Huber, *Robust Statistics*, John Wiley, New York (1981).

[9] C.J. Lloyd, On comparing the accuracy of competing tests of the same

hypotheses from simulation data, *Journal of Statistical Planning and Inference*, **128** (2005), 497-508.

[10] M. Markatou, and G. Manos, Robust tests in nonlinear models, *Journal of Statistical Planning and Inference*, **55** (1996), 205-217.

[11] M. Markatou and X. He, Bounded influence and high breakdown point testing procedures in linear models, *Journal of the American Statistical Association*, **39** (1994), 543-549.

[12] M. Markatou and T.P. Hettmansperger, Applications of the asymmetric eigenvalue problem techniques to robust testing, *Journal of Statistical Planning and Inference*, **31** (1992), 51-65.

[13] S. Morgenthaler and J.W. Tukey, *Configural Polysampling*, John Wiley, New York (1991).

[14] D.A. Ratkowsky, *Nonlinear Regression Modeling*, Marcel Dekker, New York (1983).

[15] P.J. Rousseeuw, Least median of squares regression, *Journal of the American Statistical Association*, **79** (1984), 871-880.

[16] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, New York (1987).

[17] P.J. Rousseeuw and B.C. van Zomeren, Unmasking multivariate outliers and leverage points (with discussion). *Journal of the American Statistical Association*, **85** (1990), 633-651.

[18] G.K. Smyth, Partitioned algorithms for maximum likelihood and other nonlinear estimation, *Statistics and Computing*, **6** (1996), 201-216.

[19] G.K. Smyth, Nonlinear regression. In *Encyclopedia of Environmetrics,* El-Shaarawi & W.W. Piegorsch (eds), John Wiley, New York (2002).

[20] W.A. Stahel and S. Weisberg, S. (eds), *Directions in Robust Statistics and Diagnostics*, Springer-Verlag, New York (1991).

[21] R.G. Staudte and S.J. Sheather, *Robust Estimation and Testing*, John Wiley, New York (1990).

[22] A.J. Stromberg, High breakdown point estimators in nonlinear regression, In L_1-Statistical Analysis and Related Methods, ed. Y. Dodge (1992), Amsterdam: North-Holland, p.103-112.

[23] A.J. Stromberg, Computation of high breakdown nonlinear regression parameters, *Journal of the American Statistical Association*, **88** (1993), 237-244.

[24] A.J. Stromberg, Some software for computing robust linear or nonlinear regression estimators, *Communications in Statistics - Simulation and Computation*, **26** (1997), 947-959.

[25] A.J. Stromberg and D. Ruppert, Breakdown in nonlinear regression, *Journal of the American Statistical Association*, **87** (1992), 991-997.

[26] R.T. St. Laurent and D.R. Cook, Leverage and superleverage in nonlinear regression, *Journal of the American Statistical Association*, **87** (1992), 985-990.

[27] R.T. St. Laurent and D.R. Cook, Leverage local influence and curvature in nonlinear regression, *Biometrika*, **80** (1993), 99-106.