Ann. Inst. Statist. Math.
Vol. 46, No. 2, 267–278 (1994)

ON STATISTICAL MODELS FOR REGRESSION DIAGNOSTICS

BO-CHENG WEI AND JIAN-QING SHIH

*Department of Mathematics, Southeast University, Si Pai Lou*2, Nanjing 210018, China*

(Received August 24, 1992; revised March 31, 1993)

Abstract. In regression diagnostics, the case deletion model (CDM) and the mean shift outlier model (MSOM) are commonly used in practice. In this paper we show that the estimates of CDM and MSOM are equal in a wide class of statistical models, which include LSE, MLE, Bayesian estimate and M-estimate in linear and nonlinear regression models; MLE in generalized linear models and exponential family nonlinear models; MLEs of transformation parameters of explanatory variables in a Box-Cox regression models and so on. Furthermore, we study some models, in which, the estimates are not exactly equal but are approximately equal for CDM and MSOM.

Key words and phrases: Case deletion model, exponential family nonlinear models, mean-shift outlier model, nonlinear regression models, regression transformation, proportional hazards models.

1. Introduction

In regression diagnostics, there are three commonly used models: case-deletion model (CDM), mean-shift outlier model (MSOM) and case-weighted model (CWM) (Cook and Weisberg (1982)). Each of those models has its own advantage in practice. In general, CDM is intuitive and easy to understand, MSOM is easy to deal with and CWM may be used to some specific cases (see for example, Pregibon (1981)). However, CDM and MSOM are central frequentist models (Beckman and Cook (1983)). It is well known that in linear models, CDM and MSOM are equivalent in following sense: the least squares estimates (LSE) of the parameters are equal for CDM and MSOM. However, the equivalence of other estimates in linear models are not clear. Furthermore, the equivalence of CDM and MSOM outside of linear models have not been fully studied in the literature. The aim of this article is to show that the estimates of CDM and MSOM are equivalent in a wide class of statistical models, which include LSE, maximum likelihood estimate (MLE), Bayesian estimate and M-estimate in normal linear and nonlinear regression models, MLE in exponential family nonlinear models, MLEs of transformation parameters of explanatory variables in a Box-Cox regression models and so on. Furthermore, we also study some models, in which the estimates are not exactly equal but are approximately equal for CDM and MSOM. The emphasis of our discussion is the equivalence of CDM and MSOM. Some similar results may hold for CWM, but we will skip the details. In Section 2, we discuss the equivalence of CDM and MSOM in normal nonlinear regression models. The exponential family nonlinear models are studied in Section 3. Section 4 investigates two kinds of models, in which the MLEs are approximately equal for CDM and MSOM. Section 5 gives our concluding comments. All proofs of theorems are deferred to the Appendix.

2. CDM and MSOM in nonlinear regression models

Consider a nonlinear regression model as follows:

$$Y = f(\beta) + \epsilon, \tag{2.1}$$

where $Y = (y_1, \ldots, y_n)^T$ is a vector of responses, $f(\beta) = (f(x_1, \beta), \ldots, f(x_n, \beta))^T$ is a vector of model functions, and $x_1, \ldots, x_n$ are explanatory variables. $\beta = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown parameters and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ is a vector of independent random variables having normal distribution $N(0, \sigma^2 I)$ where I is an $n \times n$ identity matrix. We denote that $V(\beta) = \partial f(\beta)/\partial \beta^T$ and $e(\beta) = Y - f(\beta)$.

Now we consider the Bayesian estimates based on the mode of posterior distribution (MBE, see Box and Tiao (1973), p. 308). If (β, σ^2) have noninformative prior, then these estimates is consistent with the ordinary MLE. Here, we assume that σ has inverse gamma distribution, i.e. $p(\sigma) \propto \sigma^{-b} \exp\{-(b-1)a^2/(2\sigma^2)\}$ and the conditional distribution of β given σ is $N(\beta_0, \sigma^2 \Sigma_0)$, where β_0 is a known vector and Σ_0 is a known positive definite matrix. It is easy to show that the posterior density of (β, σ) is

$$p(\beta, \sigma \mid y) \propto \sigma^{-(n+p+b)} \exp\{-M(\beta)/(2\sigma^2)\}, \tag{2.2}$$

where $M(\beta) = (b-1)a^2 + (\beta - \beta_0)^T \Sigma_0^{-1} (\beta - \beta_0) + e^T(\beta) e(\beta)$. Denote the MBEs of β and σ^2 by $\tilde{\beta}$ and $\tilde{\sigma}^2$ which maximize $p(\beta, \sigma \mid y)$. It is easy to show from (2.2) that $\tilde{\beta}$ and $\tilde{\sigma}$ satisfy

$$V^T(\beta) e(\beta) - \Sigma_0^{-1}(\beta - \beta_0) = 0, \tag{2.3}$$

$$\sigma^2 = (n + p + b)^{-1} M(\beta). \tag{2.4}$$

To assess the influence of i-th case (x_i, y_i) on the estimates $\tilde{\beta}$ and $\tilde{\sigma}$, we consider CDM and MSOM of (2.1) as follows:

$$Y_{(i)} = f_{(i)}(\beta) + \epsilon_{(i)}, \tag{2.5}$$

$$Y = f(\beta) + d_i \gamma + \epsilon, \tag{2.6}$$

where $Y_{(i)}$, $f_{(i)}(\beta)$ and $\epsilon_{(i)}$ are obtained from Y, $f(\beta)$ and ϵ respectively with i-th case deleted; d_i is an n-vector having 1 in the i-th position and 0 elsewhere. The prior distributions of σ and β in (2.5) and (2.6) are assumed as before and γ has noninformative prior. The MBEs of β and σ^2 for (2.5) and (2.6) are denoted by $\tilde{\beta}_{(i)}$, $\tilde{\sigma}^2_{(i)}$ and $\tilde{\beta}_{mi}$, $\tilde{\sigma}^2_{mi}$ respectively, which are assumed to be existing and unique.

THEOREM 2.1. *Under above assumptions, we have*

$$\tilde{\beta}_{mi} = \tilde{\beta}_{(i)}, \quad \tilde{\sigma}^2_{mi} = (n+p+b)^{-1}(n+p+b-1)\tilde{\sigma}^2_{(i)} \approx \tilde{\sigma}^2_{(i)}. \tag{2.7}$$

This theorem shows that diagnostic models (2.5) and (2.6) of nonlinear model (2.1) play the same role for investigating the influence of i-th case on $\tilde{\beta}$ and $\tilde{\sigma}^2$. In other words, CDM is equivalent to MSOM. For the diagnostics purpose, we can define the Cook distance and W-K statistic from $\tilde{\beta}$ and $\tilde{\beta}_{(i)}$ based on models (2.5) or (2.6), and we also can make a testing of hypothesis $H_0 : \gamma = 0$ based on model (2.6).

As the special case of Theorem 2.1, we can get several important corollaries as follows which have not been seen in the literature.

(1) If we set $b = 1$, $\Sigma_0 \to \infty$ in equation (2.2) which corresponds to noninformative prior, we have $p(\beta, \sigma \mid y) \propto \sigma^{-1}\exp\{-e^T(\beta)e(\beta)/(2\sigma^2)\}$. Then $\tilde{\beta}$ and $\tilde{\sigma}^2$ is consistent with the ordinary MLE or LSE. Therefore, CDM is equivalent to MSOM for MLE or LSE in normal nonlinear regression models. Ross (1987) and Dzieciolowski and Ross (1990) have used this fact implicitly.

(2) If $f(\beta) = X\beta$ in (2.1), then it reduces to a linear regression model and $V(\beta) = X$, $V_{(i)}(\beta) = X_{(i)}$ which is obtained from X by deleting i-th row. So equations (2.3) and (A.1) become

$$X^T(Y - X\beta) = \Sigma_0^{-1}(\beta - \beta_0),$$
$$X^T_{(i)}(Y_{(i)} - X_{(i)}\beta) = \Sigma_0^{-1}(\beta - \beta_0).$$

Therefore, the MBE of β in model (2.1) and (2.5) (or (2.6)) can be expressed as

$$\begin{aligned} \tilde{\beta} &= (X^TX + \Sigma_0^{-1})^{-1}(X^TY + \Sigma_0^{-1}\beta_0), \\ \tilde{\beta}_{(i)} &= \tilde{\beta} - (1 - \tilde{p}_{ii})^{-1}(X^TX + \Sigma_0^{-1})^{-1}x_i\tilde{e}_i, \end{aligned} \tag{2.8}$$

where $\tilde{p}_{ii} = x_i^T(X^TX + \Sigma_0^{-1})^{-1}x_i$, $\tilde{e} = Y - X\tilde{\beta}$, $\tilde{e}_i = y_i - x_i^T\tilde{\beta}$ and x_i^T is i-th row of X. It is well known that (2.8) is also the Bayesian estimate of β based on minimizing the posterior risk under squared error loss. It is also known that, if $\Sigma_0 = k^{-1}I$ in equation (2.8), then $\tilde{\beta}$ is the ridge estimate; if $\Sigma_0 = (QKQ^T)^{-1}$ where Q is composed of the orthonormal eigenvectors of X^TX, then $\tilde{\beta}$ is the generalized ridge estimate; if $\Sigma_0 = \sigma^{-2}\sigma^2_\beta(X^TX)^{-1}$, then $\tilde{\beta}$ is the James-Stein estimate (Lindley and Smith (1972), Weisberg (1985)). Theorem 2.1 implies that several important biased estimates in linear regression models, such as ridge estimate, generalized ridge estimate and James-Stein estimate, are equal for CDM and MSOM.

(3) The results of Theorem 2.1 also can be extended to the diagnostics of k cases (x_i, y_i), $i \in J$, $J = \{i_1, \ldots, i_k\}$, i.e. CDM and MSOM are equivalent for investigating the influence of k cases on the estimates $\tilde{\beta}$ and $\tilde{\sigma}^2$.

(4) The results of Theorem 2.1 can be extended to regression models in which random error vector ϵ has elliptically contoured distribution, i.e. $\epsilon \sim EC_n(0, \sigma^2 I, g)$ (see Fang *et al.* (1990)) where the density function of ϵ is $\sigma^{-1}g(-\epsilon^T\epsilon/(2\sigma^2))$ and $g(\cdot)$ is twice continuously differentiable. For the sake of simplicity, we assume

that σ is known and β has elliptically contoured prior whose density function is $\sigma^{-1}|\Sigma_0|^{-1/2}h(T(\beta))$, $T(\beta) = -(2\sigma^2)^{-1}(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0)$. Now we consider the MBE of β based on the mode of posterior distribution. By an analogous analysis to Theorem 2.1, we can show that $\tilde{\beta}_{(i)} = \tilde{\beta}_{mi}$ under the conditions in which both $\tilde{\beta}_{(i)}$ and $\tilde{\beta}_{mi}$ exist and are unique. As a corollary of above statement, for nonlinear model (2.1) with elliptically contoured random error $\epsilon \sim EC_n(0, \sigma^2 I, g)$, if σ^2 is known, then the ordinary MLEs of β are equal for CDM (2.5) and MSOM (2.6).

3. CDM and MSOM in exponential family nonlinear models

Suppose that the data are given by (x_i, y_i) with $\mu_i = E(y_i)$ and the monotonic transformation $\eta_i = g(\mu_i)$ is a nonlinear function of β, i.e.

$$E(y_i) = \mu_i, \quad \eta_i = g(\mu_i) = f(x_i; \beta) \equiv f_i(\beta), \qquad i = 1, \ldots, n, \tag{3.1}$$

which may be called generalized nonlinear model. If $f_i(\beta) = x_i^T\beta$, then (3.1) is the well known generalized linear model and if $g(\mu_i) = \mu_i$, then (3.1) is just the ordinary nonlinear regression model. We usually assume that $Y = (y_1, \ldots, y_n)$ is an n-vector of independent variables having exponential family distribution as

$$p(y_i, \theta_i) = \exp[\sigma^{-2}\{y_i\theta_i - b(\theta_i) + c(y_i, \sigma)\}], \qquad i = 1, \ldots, n,$$

where $b(\cdot)$ and $c(\cdot, \cdot)$ are known differentiable functions, and θ_i is the nature parameter. If $b(\theta_i) = \theta_i^2/2$, then y_i is normally distributed. Since $\mu_i = \dot{b}(\theta_i)$, we have $\theta_i = k(f_i(\beta))$ where $k(\cdot) = \dot{b}^{-1}g^{-1}(\cdot)$. The MLEs of β and σ are respectively denoted by $\hat{\beta}$ and $\hat{\sigma}$ that minimize following objective function

$$Q(\beta, \sigma) = -\sigma^{-2}\sum_{i=1}^{n}\{y_i k(f_i(\beta)) - b(k(f_i(\beta))) + c(y_i, \sigma)\}, \tag{3.2}$$

which is called exponential family nonlinear models (Cook and Tsai (1990)). Furthermore, we can consider more general models in which $\hat{\beta}$ and $\hat{\sigma}$ minimize following objective function:

$$Q(\beta, \sigma) = \sum_{i=1}^{n} \rho_i(f_i(\beta), \sigma), \tag{3.3}$$

where $f_i(\cdot)$ and $\rho_i(\cdot, \cdot)$ are known differentiable functions and may depend on both x_i and y_i, $i = 1, \ldots, n$. This model is proposed by Gay and Welsch (1988) and includes a wide class of models as its special cases. Clearly, model (3.2) is an important example of (3.3).

Now we consider the diagnostic problem for model (3.3). To assess the influence of i-th case on the estimates $\hat{\beta}$ and $\hat{\sigma}$, we consider CDM and MSOM of model (3.3) as follows:

$$Q_{(i)}(\beta, \sigma) = \sum_{j \neq i} \rho_j(f_j(\beta), \sigma), \tag{3.4}$$

$$Q_{mi}(\beta, \gamma, \sigma) = \left\{\sum_{j \neq i} \rho_j(f_j(\beta), \sigma)\right\} + \rho_i(f_i(\beta) + \gamma, \sigma). \tag{3.5}$$

Now let estimates $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ and $\hat{\beta}_{mi}$, $\hat{\gamma}$, $\hat{\sigma}_{mi}$ minimize (3.4) and (3.5) respectively. We suppose that all above estimates exist and are unique.

THEOREM 3.1. *Under above assumptions, if σ is known, then $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$; if σ is unknown, we have*

(a) *If $\rho_i(f_i(\beta), \sigma) = \rho_i(a_i)$, $a_i = g(\sigma) f_i(\beta) + h_i(\sigma)$, $i = 1, \ldots, n$, where $g(\cdot)$ and $h_i(\cdot)$ are differentiable functions and $g(\sigma) \neq 0$, then $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$, $\hat{\sigma}_{mi} = \hat{\sigma}_{(i)}$.*

(b) *If $\rho_i(f_i(\beta), \sigma) = g(\sigma)\xi_i(f_i(\beta)) + h_i(\sigma)$, where $\xi_i(\cdot)$ is a differentiable function, then $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$ (but not necessary to have $\hat{\sigma}_{mi} = \hat{\sigma}_{(i)}$).*

This theorem states that, for the general model (3.3) CDM is equivalent to MSOM under some mild conditions when we investigate the influence of i-th case on estimates $\hat{\beta}$ and $\hat{\sigma}^2$. From Theorem 3.1, we have several important corollaries and comments as follows.

(1) The exponential family nonlinear models (3.2) is the special case of (b) of Theorem 3.1 with $g(\sigma) = \sigma^{-2}$, $h_i(\sigma) = \sigma^{-2} c(y_i, \sigma)$ and $\xi_i(f_i(\beta)) = y_i k(f_i(\beta)) - b(k(f_i(\beta)))$ respectively. Therefore, we always have $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$. Note that we may have not $\hat{\sigma}_{mi} = \hat{\sigma}_{(i)}$. As a counter example, the MLE of σ in normal linear regression is a trivial one, in which $\hat{\sigma}^2_{mi} = n^{-1}(n-1)\hat{\sigma}^2_{(i)} \neq \hat{\sigma}^2_{(i)}$, but $\hat{\sigma}^2_{mi} \approx \hat{\sigma}^2_{(i)}$.

(2) The M-estimates of (2.1) can be obtained from (3.3) with $\rho_i(f_i(\beta), \sigma) = \rho_i(a_i)$, $a_i = \sigma^{-1}(y_i - f_i(\beta))$ and $\rho(\cdot)$ is a convex positive function. This is the special case of (a) of Theorem 3.1 with $g(\sigma) = -\sigma^{-1}$ and $h_i(\sigma) = \sigma^{-1} y_i$. Therefore the M-estimates of nonlinear regression models are equal for CDM and MSOM.

(3) If $\rho_i(f_i(\beta), \sigma) = -\log\{p(y_i; f_i(\beta), \sigma)\}$ where $p(y_i; f_i(\beta), \sigma)$ is the density function of y_i (Davison and Tsai (1992)), then $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$ when σ is known. As a special case, for nonlinear regression models whose random error $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distribution but not necessarily normal, we have $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$ when σ is known.

(4) The accelerated life models described by Cox and Oakes (1984) can be represented as

$$Z_i = \log T_i = x_i^T \beta + \sigma \epsilon_i, \qquad i = 1, \ldots, n. \tag{3.6}$$

We consider one of the typical case in which each T_i has Weibull distribution and hence each ϵ_i has standard extreme-value distribution function $F(y) = 1 - \exp(-e^y)$. Therefore, the log likelihood function of $Z = (Z_1, \ldots, Z_n)^T$ can be expressed as

$$L(\beta, \sigma) = \sum_{i=1}^{n} \{\delta_i a_i - \exp(a_i) - \delta_i \log \sigma\}, \tag{3.7}$$

where $a_i = \sigma^{-1}(Z_i - x_i^T \beta)$, $\delta_i = 1$ for an uncensored observation and $\delta_i = 0$ for a censored one. Theorem 3.1 shows that if σ is known, then the MLEs of β of model (3.6) are equal for CDM and MSOM.

Table 1. The estimators corresponding to CDM and MSOM.

	CDM			MSOM			
	$\hat{\alpha}_{(i)}$	$\hat{\beta}_{(i)}$	$\hat{\sigma}_{(i)}$	$\hat{\alpha}_{mi}$	$\hat{\beta}_{mi}$	$\hat{\gamma}_{mi}$	$\hat{\sigma}_{mi}$
1	5.157	−.505	.064	5.161	−.506	.424	.059
2	5.663	−.629	.129	5.669	−.629	−.161	.120
3	5.663	−.627	.123	5.669	−.628	−.232	.114
4	5.656	−.625	.123	5.662	−.626	−.233	.114
5	5.646	−.622	.123	5.653	−.623	−.239	.114
6	5.631	−.620	.132	5.638	−.621	−.130	.123
7	5.622	−.619	.138	5.629	−.620	−.046	.129
8	5.631	−.626	.150	5.637	−.626	.051	.131
9	5.704	−.655	.119	5.710	−.656	.200	.111
Modified data							
1*	2.817	.299	1.015	2.817	.299	29.622	1.015
2	5.786	−.475	1.195	5.683	−.439	−.612	1.109
3	5.777	−.432	.988	5.667	−.394	−3.166	.921
4	5.610	−.421	1.172	5.520	−.387	−.805	1.086
5	5.477	−.380	1.157	5.388	−.347	−.914	1.071
6	5.378	−.352	1.154	5.290	−.319	−.896	1.067
7	5.756	−.668	.427	5.736	−.661	2.319	.379
8	5.132	−.278	1.134	5.042	−.243	−1.032	1.046
9	5.053	−.255	1.130	4.962	−.219	−1.063	1.042

If σ is unknown, then (3.7) is neither case (a) nor case (b) of Theorem 3.1. In this case, CDM may be not equivalent to MSOM. In fact, it is easily shown that $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ and $\hat{\beta}_{mi}$, $\hat{\sigma}_{mi}$ respectively satisfy following equations:

$$\sum_{j\neq i} b_j x_j = 0 \quad \text{and} \quad (m - \delta_i) + \sum_{j\neq i} b_j a_j = 0; \tag{3.8}$$

$$\sum_{j\neq i} b_j x_j = 0 \quad \text{and} \quad m + \sum_{j\neq i} b_j a_j = 0, \tag{3.9}$$

where $b_j = \delta_j - \exp(a_j)$ and m is the number of uncensored observations. Comparing (3.8) and (3.9), it is easily seen that if i-th case is a censored observation, then MSOM is equivalent to CDM; if i-th case is an uncensored one then the MLEs of (β, σ) are not exactly equal for CDM and MSOM, but they may be approximately equal if m is relatively large. Similar result also holds for (3.6) if T_i have log-normal distributions.

Now let us consider a numerical example of clotting times of blood data discussed by McCullagh and Nelder ((1989), p. 302). We use the accelerated life models with Weibull distribution and set $Z_i = \log(y_i) = \alpha + \beta x_i + \sigma\epsilon_i$, $i = 1, \ldots, 9$, where $x_i = \log(u_i)$ and $y_i = (\text{Lot})_i$. The parameter estimates are $\hat{\alpha} = 5.63$, $\hat{\beta} = -.622$ and $\hat{\sigma} = .130$. Parameter estimates corresponding to CDM and MSOM are listed in Table 1 which shows that the differences of estimates between CDM and MSOM are very small. Now we change the values of y_3 and y_7 for 4.2 and 210 respectively and regard y_1 as a right censored observation, then the difference of estimates between CDM and MSOM seems to be large. The last nine rows of Table 1 give parameter estimates corresponding to CDM and MSOM for modified data. These results show that some of relative differences of estimates of β between CDM and MSOM are quite large (for example, $(\hat{\beta}_{(9)} - \hat{\beta}_{m9})/\hat{\beta}_{(9)} = 14.1\%$ and $(\hat{\beta}_{(8)} - \hat{\beta}_{m8})/\hat{\beta}_{(8)} = 12.6\%$). Note that $\hat{\alpha}_{(1)} = \hat{\alpha}_{m1}$ and $\hat{\beta}_{(1)} = \hat{\beta}_{m1}$ since case 1 is assumed to be a censored observation.

(5) For general case of (3.3) with unknown σ, equations (A.9) and (A.11) of Appendix show that $\hat{\sigma}_{mi}$ is usually not equal to $\hat{\sigma}_{(i)}$ except $\partial\rho_i(f_i(\beta)+\gamma,\sigma)/\partial\sigma = 0$ in (A.12). The case (a) of Theorem 3.1 is a typical example of this. If $\hat{\sigma}_{mi} \neq \hat{\sigma}_{(i)}$, (A.8) and (A.10) show that $\hat{\beta}_{mi}$ is usually not equal to $\hat{\beta}_{(i)}$ except (A.8) or (A.10) is independent of σ. The case (b) of Theorem 3.1 is a typical example of this. For other cases of (3.3) with unknown σ, the equivalent of CDM and MSOM is still not known. But fortunately for the most of commonly encountered practical models, such as (2.1) and (3.2), this equivalence is clarified as stated in items (1)–(4).

4. CDM and MSOM in other models

Since MSOM is easy to deal with in most situations, one usually use MSOM instead of CDM to derive diagnostic statistics. But MSOM may be not equivalent to CDM for some models. In this section, we study regression transformation models and proportional hazards regression models, in which the MLEs of MSOM and CDM are not exactly equal, but may be approximately equal if the number of observations is relatively large.

The regression transformation model of (2.1) with $f(\beta) = X\beta$ is

$$Y(\lambda) = X\beta + \epsilon, \tag{4.1}$$

where $Y(\lambda) = (y_1(\lambda), \ldots, y_n(\lambda))^T$. $y_i(\lambda) = h(y_i, \lambda)$ is a monotonic transformation of y_i indexed by the unknown parameter λ, $h(\cdot,\cdot)$ is a known differentiable function, $i = 1, \ldots, n$, and Box-Cox power transformations are commonly used in practice. The CDM and MSOM for model (4.1) can be expressed as $Y_{(i)}(\lambda) = X_{(i)}\beta + \epsilon_{(i)}$ and $Y(\lambda) = X\beta + \gamma d_i + \epsilon$ respectively, where $Y_{(i)}(\lambda)$ is obtained from $Y(\lambda)$ with i-th element deleted. Tsai and Wu (1990) reported by numerical examples that the MLE $\hat{\lambda}_{mi}$ based on MSOM is not necessarily equal to the MLE $\hat{\lambda}_{(i)}$ based on CDM. Now we give a theoretical analysis for this statement.

By the equivalence of CDM and MSOM in linear regression, for every fixed λ, the MLEs of β and the residual sum of squares in CDM and MSOM are equal, which are denoted by $\hat{\beta}_{(i)}(\lambda)$ and

$$\mathrm{RSS}_{(i)}(\lambda) = \sum_{j \neq i} \{y_j - x_j^T \hat{\beta}_{(i)}(\lambda)\}^2.$$

According to the results given by Cook and Weisberg (1982), the log likelihood of λ associated with CDM is

$$L_{(i)}(\lambda) = -\frac{(n-1)}{2}\log\{J_{(i)}(\lambda)^{-2/(n-1)}\mathrm{RSS}_{(i)}(\lambda)\},$$
$$J_{(i)}(\lambda) = \prod_{j\neq i}\partial h(y_j,\lambda)/\partial y_j,$$

the log likelihood of λ associated with MSOM is

$$L_{mi}(\lambda) = -\frac{n}{2}\log\{J(\lambda)^{-2/n}\mathrm{RSS}_{(i)}(\lambda)\}, \qquad J(\lambda) = \prod_{j=1}^{n}\partial h(y_j,\lambda)/\partial y_j.$$

$\hat{\lambda}_{(i)}$ and $\hat{\lambda}_{mi}$ minimize $S_{(i)}(\lambda) = J_{(i)}(\lambda)^{-2/(n-1)}\mathrm{RSS}_{(i)}(\lambda)$ and $S_{mi}(\lambda) = J(\lambda)^{-2/n}\cdot \mathrm{RSS}_{(i)}(\lambda)$ respectively, therefore, $\hat{\lambda}_{mi}$ is not necessarily equal to $\hat{\lambda}_{(i)}$. But when n is sufficiently large, we have $J_{(i)}(\lambda)^{1/(n-1)} \approx J(\lambda)^{1/n}$ and therefore $\hat{\lambda}_{mi} \approx \hat{\lambda}_{(i)}$. In the example given by Tsai and Wu (1990), because $n = 11$ is relatively small, the difference of $\hat{\lambda}_{mi}$ and $\hat{\lambda}_{(i)}$ is relatively large.

It is very interesting that we have a different result for regression transformation model of explanatory variables

$$Y = X(\lambda)\beta + \epsilon, \tag{4.2}$$

where some columns of X are transformed and denoted by $X(\lambda)$. The log-likelihood of λ for model (4.2) is $L(\lambda) = -(n/2)\log[\mathrm{RSS}(\lambda)]$, where $\mathrm{RSS}(\lambda)$ is the residual sum of squares given by

$$\mathrm{RSS}(\lambda) = \sum_{i=1}^{n}\{y_i - x_i^T\hat{\beta}(\lambda)\}^2,$$

$\hat{\beta}(\lambda)$ is the MLE of β for fixed λ in (4.2). So the MLE of λ minimizes $\mathrm{RSS}(\lambda)$.

For model (4.2), its associated CDM and MSOM are respectively expressed as $Y_{(i)} = X_{(i)}(\lambda) + \epsilon_{(i)}$, and $Y = X(\lambda) + \gamma d_i + \epsilon$, where $X_{(i)}(\lambda)$ is obtained from $X(\lambda)$ with i-th row deleted. It is easily seen that for any fixed λ, the residual sum of squares and the MLEs of β are exactly equal for above two models, which are denoted by $\hat{\beta}_{(i)}(\lambda)$ and $\mathrm{RSS}_{(i)}(\lambda)$ respectively. Because the MLEs of λ for both CDM and MSOM minimize the same residual sum of squares $\mathrm{RSS}_{(i)}(\lambda)$, these two MLEs of λ are exactly equal.

Now we investigate the proportional hazards model described by McCullagh and Nelder ((1989), pp. 421–422), the hazards function is represented as

$$h(t_i, x_i) = \lambda(t_i)\exp(\eta_i), \quad \eta_i = G(x_i, \beta), \quad i = 1, \ldots, n, \tag{4.3}$$

where $\lambda(t)$ is a positive continuous function indexed by an unknown parameter α, and β is a p-vector of unknown parameters. The log-likelihood for (β, α) can be expressed as

$$L(\beta, \alpha) = \sum_{j=1}^{n}(\delta_j\log\mu_j - \mu_j) + \sum_{j=1}^{n}\delta_j\log[\lambda(t_j)/\Lambda(t_j)], \tag{4.4}$$

where $\Lambda(t) = \int_{-\infty}^{t} \lambda(u)du$, $\mu_j = \Lambda(t_j)\exp(\eta_j)$. The MLEs of parameters (β, α) must satisfy following equations:

$$\sum_{j=1}^{n} e_j \partial\eta_j/\partial\beta = 0, \quad \sum_{j=1}^{n} c_j = 0, \tag{4.5}$$

where $c_j = e_j f_{1j} + \delta_j f_{2j}$, $e_j = \delta_j - \mu_j$, $f_{1j} = \partial \log[\Lambda(t_j)]/\partial\alpha$, and $f_{2j} = \partial \log[\lambda(t_j)/\Lambda(t_j)]/\partial\alpha$. Similarly, the MLEs of (β, α) for CDM of (4.4) satisfy

$$\left\{\sum_{j=1}^{n} e_j \partial\eta_j/\partial\beta\right\} - e_i\partial\eta_i/\partial\beta = 0, \quad \sum_{j\neq i} c_j = 0. \tag{4.6}$$

The log-likelihood of MSOM associated (4.4) is given by

$$L_{mi}(\beta,\alpha,\gamma) = \sum_{j\neq i}(\delta_j \log\mu_j - \mu_j) + \sum_{j=1}^{n} \delta_j \log\{\lambda(t_j)/\Lambda(t_j)\} + \delta_i \log\mu_i' - \mu_i',$$

where $\mu_i' = \Lambda(t_i)\exp(\eta_i + \gamma)$. It is easy to show that $L_{mi}(\beta,\alpha,\gamma) = L(\beta,\alpha) + \delta_i\gamma + (1 - e^{\gamma})\mu_i$. So the MLEs of (β, α) for MSOM of (4.4) satisfy following equations

$$\left\{\sum_{j=1}^{n} e_j \partial\eta_j/\partial\beta\right\} - e_i\partial\eta_i/\partial\beta = 0, \quad \left\{\sum_{j=1}^{n} c_j\right\} - e_i f_{1i} = 0 \tag{4.7}$$

where we use the fact that $\partial L_{mi}/\partial\gamma = \delta_i - \mu_i\exp(\gamma) = 0$. Comparing (4.6) and (4.7), it is easily seen that if i-th case is a censored observation (i.e. $\delta_i = 0$) or $\lambda(t) = \lambda$ which corresponds to exponential distribution, then MSOM is equivalent to CDM; if the model has Weibull distribution and the i-th case is an uncensored observation (i.e. $\delta_i = 1$), then the second equations of (4.6) and (4.7) are respectively expressed as

$$\left\{\sum_{j=1}^{n} e_j f_{1j}\right\} - e_i f_{1i} + (m-1)/\alpha = 0, \quad \left\{\sum_{j=1}^{n} e_j f_{1j}\right\} - e_i f_{1i} + m/\alpha = 0.$$

Clearly, if the number of uncensored observations m is sufficiently large, above two equations are approximately equivalent, therefore, the MLEs of (β, α) are approximately equal for CDM and MSOM. It is easy to show that similar result holds for extreme-value distribution.

It is easily seen that for proportional hazards model (4.3), if α is a known parameter or $\lambda(t)$ is a known positive continuous function, then (4.4) is the special case of (3.3). According to Theorem 3.1, the MLEs of β are equal for CDM and MSOM.

5. Discussion

To assess the influence of individual cases on the estimates in regression analysis, CDM and MSOM are commonly used. In most of situations, MSOM is usually easier to deal with than CDM, so we often use MSOM instead of CDM to get diagnostic statistics in practice. Many authors have mentioned and used this fact in the literature, such as Cook and Wang (1983), Storer and Crowley (1985), Williams (1987), Ross (1987), Tsai and Wu (1990), Dzieciolowski and Ross (1990) and so on. But there have been almost no proofs outside of linear regression models about that. In this paper, we have clarified this fact and shown that for several commonly encountered practical models, the estimates of CDM are equal to the estimates of MSOM, which may help us to derive diagnostic measures.

On the other hand, as shown by numerical example in Section 3 and as discussed in Sections 3 and 4, MSOM is not always equivalent to CDM. In some situations, it may cause problems if one use MSOM instead of CDM as pointed out by Tsai and Wu (1990).

Acknowledgements

This work was supported in part by National Science Foundation of China under grants 19071017. The authors thank the referee and the Associate Editor for providing many valuable comments on an earlier version of this paper.

Appendix

PROOF OF THEOREM 2.1. It is similar to (2.3) and (2.4) that $\tilde{\beta}_{(i)}$ and $\tilde{\sigma}^2_{(i)}$ satisfy following equations:

$$V_{(i)}^T(\beta)e_{(i)}(\beta) - \Sigma_0^{-1}(\beta - \beta_0) = 0, \tag{A.1}$$

$$\sigma^2 = (n+p+b-1)^{-1}M_{(i)}(\beta), \tag{A.2}$$

where

$$M_{(i)}(\beta) = (b-1)a^2 + (\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) + e_{(i)}^T(\beta)e_{(i)}(\beta), \tag{A.3}$$

$V_{(i)}(\beta)$ and $e_{(i)}(\beta)$ are obtained from $V(\beta)$ and $e(\beta)$ with i-th row and i-th element deleted respectively.

For model (2.6), it is similar to (2.2) that

$$\begin{aligned} p(\sigma,\beta,\gamma \mid y) &\propto \sigma^{-(n+p+b)}\exp\{-M(\beta,\gamma)/(2\sigma^2)\}, \\ M(\beta,\gamma) &= (b-1)a^2 + (\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) \\ &\quad + \{e(\beta)-\gamma d_i\}^T\{e(\beta)-\gamma d_i\}. \end{aligned} \tag{A.4}$$

Differentiating $L(\sigma,\beta,\gamma \mid y) = \log\{p(\sigma,\beta,\gamma \mid y)\}$ with respect to β, γ and σ, we have

$$\begin{aligned} \partial L/\partial\beta &= -\sigma^{-2}\{\Sigma_0^{-1}(\beta-\beta_0) - V^T(\beta)(e(\beta)-\gamma d_i)\}, \\ \partial L/\partial\gamma &= -\sigma^{-2}d_i^T\{e(\beta)-\gamma d_i\}, \end{aligned} \tag{A.5}$$

$$\partial L/\partial\sigma = -(n+p+b)\sigma^{-1} + \sigma^{-3}M(\beta,\gamma). \tag{A.6}$$

$\partial L/\partial\gamma = 0$ gives $\gamma = e_i(\beta)$ which is the i-th element of $e(\beta)$. Substituting $\gamma = e_i(\beta)$ into (A.5) shows that $\tilde{\beta}_{mi}$ satisfies

$$\text{(A.7)} \qquad \partial L/\partial\beta = -\sigma^2\{\Sigma_0^{-1}(\beta-\beta_0) - V_{(i)}^T(\beta)e_{(i)}(\beta)\} = 0.$$

Since we assume that $\tilde{\beta}_{mi}$ and $\tilde{\beta}_{(i)}$ exist and are unique, we get $\tilde{\beta}_{mi} = \tilde{\beta}_{(i)}$ by comparing (A.1) and (A.7). Substituting $\tilde{\gamma} = e_i(\tilde{\beta}_{mi})$ into equation (A.4) and comparing with equation (A.3) we get $M(\tilde{\beta}_{mi}, \tilde{\gamma}) = M_{(i)}(\tilde{\beta}_{(i)})$. Combining (A.2) and (A.6), the second equation of (2.7) is proved.

PROOF OF THEOREM 3.1. Differentiating (3.4) with respect to β and σ shows that $\hat{\beta}_{(i)}$ and $\hat{\sigma}_{(i)}$ satisfy following equations:

$$\text{(A.8)} \qquad \sum_{j\neq i}\left[\frac{\partial\rho_j(f_j(\beta),\sigma)}{\partial f_j}\frac{\partial f_j(\beta)}{\partial\beta_a}\right] = 0, \qquad a = 1,\ldots,p,$$

$$\text{(A.9)} \qquad \sum_{j\neq i}\left[\frac{\partial\rho_j(f_j(\beta),\sigma)}{\partial\sigma}\right] = 0.$$

Similarly, $\hat{\beta}_{mi}$, $\hat{\gamma}$ and $\hat{\sigma}_{mi}$ for (3.5) satisfy following equations:

$$\text{(A.10)} \qquad \sum_{j\neq i}\left[\frac{\partial\rho_j(f_j(\beta),\sigma)}{\partial f_j}\frac{\partial f_j(\beta)}{\partial\beta_a}\right] = 0, \qquad a = 1,\ldots,p,$$

$$\text{(A.11)} \qquad \frac{\partial\rho_i(f_i(\beta)+\gamma,\sigma)}{\partial f_i} = 0,$$

$$\text{(A.12)} \qquad \sum_{j\neq i}\left[\frac{\partial\rho_j(f_j(\beta),\sigma)}{\partial\sigma}\right] + \frac{\partial\rho_i(f_i(\beta)+\gamma,\sigma)}{\partial\sigma} = 0,$$

where (A.10) is obtained from $\{\partial Q_{mi}(\beta,\gamma,\sigma)/\partial\beta_a\} = 0$ and using $\{\partial Q_{mi}(\beta,\gamma,\sigma)/\partial\gamma\} = 0$.

If σ is known, (A.8) and (A.10) show that $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$.

Now we prove case (a) of Theorem 3.1. (A.11) shows that $\dot{\rho}_i(a_i + g(\sigma)\gamma) = 0$ at $\hat{\beta}_{mi}$, $\hat{\gamma}$ and $\hat{\sigma}_{mi}$, where $\rho_i(f_i(\beta)+\gamma,\sigma) = \rho_i(a_i+g(\sigma)\gamma,\sigma)$ and $\dot{\rho}_i(a_i) = \partial\rho_i(a_i)/\partial a_i$. Then we have $\partial\rho_i(f_i(\beta)+\gamma,\sigma)/\partial\sigma = 0$ in (A.12). Thus (A.9) and (A.12) are identical for case (a), which results in $\hat{\sigma}_{mi} = \hat{\sigma}_{(i)}$. It follows from (A.8) and (A.10) that $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$.

For case (b) of Theorem 3.1, since $\partial\rho_j(f_j(\beta),\sigma)/\partial f_j = g(\sigma)\partial\xi_j(f_j(\beta))/\partial f_j$, equations (A.8) and (A.10) are independent of σ, which results in $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$. Since (A.9) and (A.12) are not necessary identical in this case, we may have not $\hat{\sigma}_{mi} = \hat{\sigma}_{(i)}$.

REFERENCES

Beckman, R. J. and Cook, R. D. (1983). Outlier ... s, *Technometrics*, **25**, 119–149.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Weley, Reading, Massachusetts.

Cook, R. D. and Tsai, C. L. (1990). Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models, *J. Amer. Statist. Assoc.*, **85**, 770–777.

Cook, R. D. and Wang, P. C. (1983). Transformation and influential cases in regression, *Technometrics*, **25**, 337–343.

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall, London.

Davison, A. C. and Tsai, C. L. (1992). Regression model diagnostics, *Internat. Statist. Rev.*, **60**, 337–353.

Dzieciolowski, K. and Ross, W. H. (1990). Assessing case influence on confidence intervals in nonlinear regression, *Canad. J. Statist.*, **18**, 127–139.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

Gay, D. M. and Welsch, R. E. (1988). Maximum likelihood and quasi-likelihood for nonlinear exponential family regression models, *J. Amer. Statist. Assoc.*, **83**, 990–998.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model, *J. Roy. Statist. Soc. Ser. B*, **34**, 1–42.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman and Hall, London.

Pregibon, D. (1981). Logistic regression diagnostics, *Ann. Statist.*, **9**, 705–724.

Ross, W. H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression, *Canad. J. Statist.*, **15**, 91–103.

Storer, B. E. and Crowley, J. (1985). A diagnostic for Cox regression and general conditional likelihoods, *J. Amer. Statist. Assoc.*, **80**, 139–147.

Tsai, C. L. and Wu, X. (1990). Diagnostics in transformation and weighted regression, *Technometrics*, **32**, 315–322.

Weisberg, S. (1985). *Applied Linear Regression*, 2nd ed., John Wiley, New York.

Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions, *Appl. Statist.*, **36**, 181–191.