Working Paper Series, N. 8, July 2004

Robust prediction limits based on M-estimators

F. Giummolè
Department of Statistics
Università Ca' Foscari di Venezia
Italy

L. Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this paper we discuss a robust solution to the problem of prediction. Following Barndorff-Nielsen and Cox (1996) and Vidoni (1998), we propose improved prediction limits based on M-estimators instead of maximum likelihood estimators. To compute these robust prediction limits, the expressions of the bias and variance of an M-estimator are required. Here a general asymptotic approximation for the bias of an M-estimator is derived. Moreover, by means of comparative studies in the context of affine transformation models, we show that the proposed robust procedure for prediction behaves in a similar manner to the classical one when the model is correctly specified, but it is designed to be stable in a neighborhood of the model.

Keywords: Asymptotic expansion, Bias, Influence function, Prediction, Robustness, Scale and regression model.

Final version (2004-07-02)

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Laura Ventura
tel: +39 049 827 4177
ventura@stat.unipd.it
http://www.stat.unipd.it/~ventura

Robust prediction limits based on M-estimators

F. Giummolè
Department of Statistics
Università Ca' Foscari di Venezia
Italy

L. Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this paper we discuss a robust solution to the problem of prediction. Following Barndorff-Nielsen and Cox (1996) and Vidoni (1998), we propose improved prediction limits based on M-estimators instead of maximum likelihood estimators. To compute these robust prediction limits, the expressions of the bias and variance of an M-estimator are required. Here a general asymptotic approximation for the bias of an M-estimator is derived. Moreover, by means of comparative studies in the context of affine transformation models, we show that the proposed robust procedure for prediction behaves in a similar manner to the classical one when the model is correctly specified, but it is designed to be stable in a neighborhood of the model.

Keywords: Asymptotic expansion, Bias, Influence function, Prediction, Robustness, Scale and regression model.

1 Introduction

The purpose of this paper is to define robust prediction limits for an unobserved absolutely continuous random variable. In particular, we consider the situation where the data $y = (y_1, \ldots, y_n)$ are a realization of a random vector Y, with probability density function $f(y;\theta)$ and cumulative distribution function $F(y;\theta)$, known except for the d-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$. The future random variable Z is assumed independent of Y and with probability density function $f(z;\theta)$ which depends on the same unknown parameter θ.

A prediction statement about Z is often given through prediction limits, i.e. functions of the data, $c(y) \in \mathbb{R}$, such that

$$P_\theta\{Z < c(Y)\} = \alpha \ , \tag{1}$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0,1)$. The above probability is usually called coverage probability and is calculated with respect to the joint density of Z and Y. In practice, prediction limits that satisfy (1) can only be obtained in

very special cases, namely when a pivotal quantity, that is a function of Z and Y which distribution is free of θ, exists. Thus, in practice, the aim becomes that of satisfying (1) to a high order of approximation. An easy way of making predictions about Z is by means of the so-called estimative predictive density $f(z;\hat{\theta})$, where $\hat{\theta}$ is a suitable estimator for θ, usually the maximum likelihood estimator (MLE). However, prediction limits based on the estimative density are usually imprecise, having coverage error of order $O(n^{-1})$. Indeed, Barndorff-Nielsen and Cox (1996) and Vidoni (1998) suggest a way to correct the quantiles of the estimative density, thus obtaining prediction limits with a coverage error of order $o(n^{-1})$. Also Corcuera and Giummolè (2002) propose a solution for the case when Z is a m-dimensional random vector. They find prediction regions with coverage error of asymptotic order $o(n^{-1})$.

The aim here is to study robust prediction limits, that is prediction limits which present a good behaviour both when the model is correctly specified and also in the presence of small deviations from the assumed model. In fact, in many situations of practical interest, there is no certainty that the observed data y come from the specified model $F(y;\theta)$. They may instead come from some neighborhood of the model. It is well known that several standard likelihood procedures are not robust with respect to model misspecifications or presence of outliers, and the need for robust statistical procedures has been stressed by many authors in the statistical literature; see for instance, Huber (1981), Hampel et al. (1986) and Markatou and Ronchetti (1997). However, while robust literature offers many solutions for inference on the parameter θ of the model, the prediction problem has been somehow neglected. Exceptions are given by Fisher and Horn (1994), Basu and Harris (1994) and Ronchetti and Vidoni (1998). Anyway, the above results do not improve on the estimative density in terms of coverage error of related prediction limits.

The robust solution proposed in this paper is the analogous of the prediction limits discussed by Barndorff-Nielsen and Cox (1996) and Vidoni (1998), when the MLEs are substituted by suitable M-estimators. Such estimators define a general class, defined through unbiased estimating equations, which includes the MLE as a particular case and plays an important role in the context of robust theory. In order to apply the proposed predictive procedures, the required ingredients are the asymptotic bias and asymptotic variance of the M-estimator used. To this end, in this paper an asymptotic expansion for the bias of an M-estimator is derived, its expression being quite similar to the expansion for the bias of the MLE (see e.g. Pace and Salvan, 1997). Attention is focused on scale and regression models.

The paper is organized as follows. Section 2 presents a brief review on classical prediction limits and their application in the context of scale and regression models. These results are the starting point to obtain the robust prediction limits developed in Section 3. Applications and simulation studies are illustrated in Section 4. In the Appendix the expansion for the bias of an M-estimator is derived.

2 A review of prediction limits

Let us assume that there exists a sufficient reduction of the data of the form $y \leftrightarrow (\hat{\theta}, a)$, where $\hat{\theta}$ is the MLE and a an ancillary statistic. According to the conditionality principle, we keep the value of a fixed when evaluating the goodness of a statistical procedure. Thus, a prediction limit can be written as a function of $\hat{\theta}$, avoiding writing the dependence on a explicitely. The problem becomes that of finding functions $c(\hat{\theta}) \in \mathbb{R}$ such that

$$P_\theta\{Z < c(\hat{\theta})\} = \alpha \ , \quad \forall\theta \ , \tag{2}$$

to a high order of approximation. The above probability is taken with respect to the joint density of Z and $\hat{\theta}$ conditioned on the observed value of a. Notice that a conditional solution to (2) satisfies (1) too.

In this paper we focus on the case of $Y_1, \ldots, Y_n$ and Z being independent random variables with distribution belonging to a parametric scale and regression model of the form

$$f(z;\theta) = \frac{1}{\sigma}\, p_0\left(\frac{z - x^\top \beta}{\sigma}\right) \ ,$$

where $\theta = (\beta, \sigma)$, with $\beta \in \mathbb{R}^p$ a regression coefficient and $\sigma > 0$ a scale parameter, $x^\top$ is a fixed vector of regressors and $p_0(\cdot)$ is a known density on $\mathbb{R}$. Applications of scale and regression models are found in many areas of statistics as, for example, survival analysis or industrial applications.

An easy way to obtain prediction limit satisfying (2), consists of considering the quantiles of the estimative density $f(z;\hat{\theta})$, so that

$$c(\hat{\beta}, \hat{\sigma}) = x^\top \hat{\beta} + \hat{\sigma} q_\alpha \ , \tag{3}$$

where q_α is the α-quantile of $p_0(z)$ and $\hat{\theta} = (\hat{\beta}, \hat{\sigma})$ is the MLE of $\theta = (\beta, \sigma)$. In spite of its intuitiveness, the estimative density $f(z;\hat{\theta})$ may not be entirely adequate for prediction. The estimative prediction limit satisfies (2) up to an error term of order $O(n^{-1})$. Indeed, especially when the dimension of θ is large in comparison with n, it may provide inaccurate results for prediction.

In order to improve the estimative solution Barndorff-Nielsen and Cox (1996) and Vidoni (1998) discuss corrections to the estimative quantiles. In particular, denote by $b(\theta) = E_\theta\{\hat{\theta} - \theta\}$ and $i(\theta)^{-1} = E_\theta\{(\hat{\theta} - \theta)^\top(\hat{\theta} - \theta)\}$ the bias and variance of $\hat{\theta}$, respectively. To present the corrected prediction limit it is convenient to use index notation and Einstein summation convention, so that summation is intended over indices that appear twice in an expression. The components of β are denoted by β^j, the corresponding components of $b(\beta)$ are b^j, while the bias of $\hat{\sigma}$ is b^σ. The blocks of $i(\theta)^{-1}$ corresponding to (β, β), (β, σ) and (σ, σ) are denoted, respectively, by i^{ij}, $i^{\sigma j}$ and $i^{\sigma\sigma}$. A corrected prediction limit is then

$$\begin{aligned} c_C(\hat{\theta}) &= c_C(\hat{\beta}, \hat{\sigma}) = (\hat{\beta}^j - \hat{b}^j) x_j^{n+1} + (\hat{\sigma} - \hat{b}^\sigma) q_\alpha \\ &+ \frac{g_0(q_\alpha)}{2\hat{\sigma}} \left(\hat{i}^{\sigma\sigma} q_\alpha^2 + 2\hat{i}^{\sigma j} q_\alpha x_j^{n+1} + \hat{i}^{ij} x_i^{n+1} x_j^{n+1} \right) \ , \end{aligned} \tag{4}$$

where a hat over a quantity means evaluation at $\theta = \hat{\theta}$, x^{n+1} is the future value of the regressors and $g_0(z) = -d \log p_0(z)/dz$. Prediction limits (4) give coverage error of asymptotic order $o(n^{-1})$, thus improving the estimative solution. Moreover it is important to notice that prediction limits (4) are equivariant with respect to regression and scale transformations in the observations. General expressions for the computation of asymptotic bias and variance of the MLE can be found in Pace and Salvan (1997).

3 Prediction limits based on M-estimators

The approach to robustness followed here is the one based on the influence function (IF) (see Hampel *et al.*, 1986).

It is easy to verify that prediction limits of the form (4) are sensitive with respect to model misspecifications or presence of outliers, since they depend on the MLE. It is well-known that the IF of the MLE is proportional to its score function $(\partial/\partial\theta) \log f(y;\theta)$. Thus, if the score function is unbounded, also is the IF. This is tipically the case for scale and regression models. Prediction limits of the form (4) are not B-robust (bias-robust) since they depend on the MLE of θ, which is not in general B-robust.

The aim of this section is to introduce a robust version of prediction limits (4). More precisely, we define robust prediction limits based on suitable M-estimators for θ. An M-estimator for θ is a generalization of the MLE and in general is defined as the solution $\tilde{\theta}$ of the system of unbiased estimating equations

$$\Psi_r = \sum_{i=1}^{n} \psi_r(y_i;\theta) = 0 \ , \quad r = 1, \ldots, d \ , \tag{5}$$

where $\psi_r(\cdot)$ is the r-th component of a suitable function $\psi(\cdot) : \mathcal{Y} \times \mathbb{R}^d \to \mathbb{R}^d$. By setting Ψ_r equal to the components of the score function, $\tilde{\theta}$ coincides with the MLE for θ. Under broad conditions which we will assume throughout this paper (cfr. Hampel et al., 1986), it can be shown that $\tilde{\theta}$ is consistent and asymptotically normal, with mean θ and asymptotic variance

$$V(\theta) = M(\theta)^{-1} \Omega(\theta) (M(\theta)^{-1})^{\top} \ , \tag{6}$$

where $M(\theta) = -\int \partial\psi(y;\theta)/\partial\theta^{\top} \, dF_\theta$ and $\Omega(\theta) = \int \psi(y;\theta)\psi(y;\theta)^{\top} \, dF_\theta$. Since the influence function of $\tilde{\theta}$ at a point x is given by

$$\mathrm{IF}_{\tilde{\theta}}(x) = M(\theta)^{-1} \psi(x;\theta) \ ,$$

the M-estimator is B-robust at the assumed model F_θ if and only if $\psi(x;\theta)$ is bounded. By bounding the IF, we are able to ensure that small deviations from the model distribution do not cause large changes in the estimates.

When $\tilde{\theta}$ is an M-estimator for θ, approximate expressions for the calculation of the bias and variance can be made available. These approximations hold up to terms of order n^{-1} and do not change the order of the coverage error associated to limits (4). To give the expression of these approximations it is again convenient to use index

notation. The components of θ are denoted by θ^r and the derivatives of Ψ_r with respect to the components of θ are denoted by $\Psi_{rs} = (\partial/\partial\theta^s)\Psi_r$, $\Psi_{rst} = (\partial/\partial\theta^s\theta^t)\Psi_r$, etc, where the indices $r, s, t, \ldots$ range over $1, \ldots, d$. For the expected values of these derivatives, we use the notation $\nu_{rs} = E(\Psi_{rs})$, $\nu_{rst} = E(\Psi_{rst})$, etc, and we assume that these quantities are of order $O(n)$. Further, the zero–mean variables Ψ_r and $\Psi_{rs} - \nu_{rs}$ are assumed to be of order $O_p(n^{1/2})$. These assumptions are satisfied in practice, since Ψ_θ asymptotically behaves like the sum of n independent random variables. In addition, κ^{rs} denotes the inverse matrix of $-\nu_{rs}$.

By using the expansions in the Appendix, we find that

$$b_M^r = E_\theta(\tilde{\theta} - \theta)^r = \frac{1}{2}\kappa^{ij}\kappa^{hk}\kappa^{lm}E(\Psi_k\Psi_m) + \kappa^{il}\kappa^{hj}E(\Psi_{hl}\Psi_j) + O(n^{-2}) \ . \qquad (7)$$

The expression for the asymptotic variance is simply given by $\hat{i}_M^{rs} = \hat{v}^{rs} + o(n^{-1})$, where v^{rs} is the (r, s) element of the matrix (6).

Now all we need is to calculate the above quantities for scale and regression models. We find that

$$\begin{aligned} c_M(\tilde{\theta}) &= c_M(\tilde{\beta}, \tilde{\sigma}) = (\tilde{\beta}^j - \tilde{b}_M^j)x_j^{n+1} + (\tilde{\sigma} - \tilde{b}_M^\sigma)q_\alpha \\ &+ \frac{g_0(q_\alpha)}{2\tilde{\sigma}}\left(\tilde{v}^{\sigma\sigma}q_\alpha^2 + 2\tilde{v}^{\sigma j}q_\alpha x_j^{n+1} + \tilde{v}^{ij}x_i^{n+1}x_j^{n+1}\right) \ , \qquad (8) \end{aligned}$$

where a tilde over a quantity indicates evaluation at $\tilde{\theta} = (\tilde{\beta}, \tilde{\sigma})$. The robust prediction limit (8) is a modification of the ordinary prediction limit based on the MLE when robustness is required.

Observe also that (8) is valid in general for any estimator $\tilde{\theta}$ which is defined as the solution of an unbiased estimating equation that behaves as the sum of n independent random variables.

4 Examples and Monte Carlo studies

In this section we consider two examples in order to compare the finite-sample behaviour of the predictive limits based on robust M-estimators with those based on classical MLEs. Interest is focused on scale and regression models, which include location and scale models. Referring to the construction of prediction limits in the robust setting, we perform Monte Carlo experiments whose objective is both to evaluate the accuracy of the prediction limits when the model is correctly specified and to assess the stability of the coverage levels under small, arbitrary departures from the assumed model. We consider a contamination model of the form $F_\varepsilon = (1 - \varepsilon)F_\theta + \varepsilon G$, where $G(\cdot)$ denotes the contaminating distribution and the contamination percentage ε is set at 5%.

Example 1: Exponential model. Let $\theta > 0$ be a scale parameter and let $(y_1, \ldots, y_n)$ be a random sample from a scale model $F(y; \theta)$. For a scale model with standard distribution $F_0(\cdot)$, we have $F(y; \theta) = F_0(y\theta)$ and $\psi(y; \theta) = \psi(y\theta)$. Moreover, $\Omega(\theta) = n\int \psi(x)^2\, dF_0(x) = \Omega$, $B(\theta) = -(n/\theta)\int x\dot{\psi}(x)\, dF_0(x) = -B/\theta$, where $\dot{\psi}(x) = \partial\psi(x)/\partial x$. In this example we focus on the optimal Hampel estimator

(see Hampel *et al.* (1986), Chap. 2) for the parameter of the exponential model $F(y;\theta) = 1 - \exp(-y\theta)$, with $\psi(x) = \max\{-b, \min\{b, a - x\}\}$, for appropriate constants a and b. Quantities involved in (4) and (8) are easy to obtain.

Table 1 shows the results of a Monte Carlo experiment (based on 10000 trials) that compares prediction limits, when data are generated from the exponential with mean 1 ($Exp(1)$) and the $Exp(1)$ contaminated by the Weibull $G(x) = 1 - \exp\{-(10x)^{1.5}\}$ (contamination on the left). The Hampel estimator is used with $a = 0.8636$ and $b = 1.129$. From Table 1 we can see that prediction limits based on this robust M-estimator are accurate for every sample size and under contamination are always preferable to those based on MLE.

(Table 1 about here)

Example 2: Linear model. A wide class of M-estimators for scale and regression parameters is defined by estimating functions of the form

$$\Psi_{\beta,\sigma} = \sum_{i=1}^{n} \psi(y_i;\beta,\sigma) = \begin{pmatrix} \sum s(x_i)\psi_\beta\{r_i v(x_i)\}x_i \\ \sum \psi_\sigma(r_i) \end{pmatrix}, \tag{9}$$

where $r_i = (y_i - x_i^\top \beta)/\sigma$ and $s(\cdot)$, $v(\cdot)$, $\psi_\beta(\cdot)$, $\psi_\sigma(\cdot)$ are appropriate functions (see Hampel et al., 1986, chap. 6). When $s(x) = v(x) = 1$ and $\psi_\beta(\cdot) = \psi_{HF}(\cdot;k_1) = \max\{-k_1, \min\{k_1, k_1 - x\}\}$ we obtain the Huber estimator. Alternatively, the choice $s(x) = 1/v(x)$, $v(x) = ||x||$ and $\psi_\beta(\cdot) = \psi_{HF}(\cdot;k_1)$ defines the Hampel-Krasker estimator. A popular choice for ψ_σ is $\psi_\sigma(\cdot) = \psi^2_{HF}(\cdot;k_2) - \gamma(k_2)$, for appropriate k_2 and $\gamma(k_2)$. However, for a general M-estimator defined by (9) with ψ_β and ψ_σ odd and even functions respectively, we have $Var\{\Psi_{\beta,\sigma}\} = \Omega = \mathrm{diag}(\Omega_{\beta,\beta}, \Omega_{\sigma,\sigma})$ and $-E\{\partial\Psi_{\beta,\sigma}/\partial(\beta,\sigma)^\top\} = (1/\sigma)B$, with $B = \mathrm{diag}(B_{\beta,\beta}, B_{\sigma,\sigma})$, where $\Omega_{\beta,\beta} = \sum s^2(x_i)g_1(x_i)x_i x_i^\top$, $\Omega_{\sigma,\sigma} = n\int \psi^2_\sigma(r)rdF_0(r)$, $B_{\beta,\beta} = \sum s(x_i)v(x_i)g_2(x_i)x_i x_i^\top$ and $B_{\sigma,\sigma} = n\int \dot{\psi}_\sigma(r)rdF_0(r)$, with $g_1(x) = \int \psi^2_\beta\{rv(x)\}dF_0(r)$ and $g_2(x) = \int \dot{\psi}_\beta\{rv(x)\}dF_0(r)$.

Consider a normal regression model with $p = 2$. The Huber estimator is used with $k_1 = 1.345$. Table 2 gives the results of a Monte Carlo experiment (based on 10000 trials) performed to assess the coverage error of the prediction limits based on (4) and (8). The central model is the $N(0,1)$, while the contaminated scenario considers the $N(0,1)$ contaminated by a $N(0,25)$. From Table 2 we can see that, in the case of a moderate sample size, both (4) and (8) can be used succesfully to construct predictive regions when the model is correct, while (8) seems slightly preferable under a small departure from the central model.

(Table 2 about here)

ACKNOWLEDGMENTS. This work was partially supported by grants from Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Italy.

Appendix

The first step in deriving expansions for estimating equations is, most commonly, to make a Taylor expansion in $\tilde{\theta}$ around θ. Thus, we begin by expanding the estimating

function $\tilde{\Psi}_\theta$ about the true parameter value θ to give

$$\tilde{\Psi}_r = \Psi_r + (\tilde{\theta}-\theta)^s\Psi_{rs} + \frac{1}{2}(\tilde{\theta}-\theta)^{st}\Psi_{rst} + \frac{1}{6}(\tilde{\theta}-\theta)^{stu}\Psi_{rstu} + O_p(n^{-1}) \ , \qquad (10)$$

where $(\tilde{\theta}-\theta)^r = \tilde{\theta}^r - \theta^r$, $(\tilde{\theta}-\theta)^{rs} = (\tilde{\theta}-\theta)^r(\tilde{\theta}-\theta)^s$, etc. Under the usual regularity conditions, which assure that the global estimator $\tilde{\theta}$ is consistent and asymptotically normal, the summands on the right-hand side of (10) are $O_p(n^{1/2})$, $O_p(n^{1/2})$, $O_p(1)$ and $O_p(n^{-1/2})$ respectively. The sample size does not appear explicitly in (10) but is incorporated into the random variables.

Let us introduce the general notation $H_{Rm} = \Psi_{Rm} - \nu_{Rm}$, for any set $Rm = r_1 \dots r_m$ of coordinate indices, so that ν_{Rm} is of order $O(n)$ and H_{Rm} of order $O_p(n^{1/2})$. Now, inserting H_{Rm} in (10) and collecting terms of the same asymptotic order, we obtain

$$\begin{aligned}(\tilde{\theta}-\theta)^s(-\nu_{rs}) &= \Psi_r + (\tilde{\theta}-\theta)^s H_{rs} + \frac{1}{2}(\tilde{\theta}-\theta)^{st}\nu_{rst} + \frac{1}{2}(\tilde{\theta}-\theta)^{st}H_{rst} \\ &+ \frac{1}{6}(\tilde{\theta}-\theta)^{stu}\nu_{rstu} + O_p(n^{-1}) \ .\end{aligned}$$

Multiplying both sides by $(-\nu^{rr'})$, we obtain the implicit version of the expansion for $(\tilde{\theta}-\theta)^r$, that is

$$\begin{aligned}(\tilde{\theta}-\theta)^r &= \Psi^r + (\tilde{\theta}-\theta)^s\kappa^{rt}H^r_s + \frac{1}{2}(\tilde{\theta}-\theta)^{st}\nu^r_{st} + \frac{1}{2}(\tilde{\theta}-\theta)^{st}H^r_{st} \\ &+ \frac{1}{6}(\tilde{\theta}-\theta)^{stu}\nu^r_{stu} + O_p(n^{-2}) \ ,\end{aligned}$$

where $\Psi^r = \kappa^{rs}\Psi_s$, $H^r_s = \kappa^{rt}H_{st}$, $\nu^r_{st} = \kappa^{ru}\nu_{stu}$, $H^r_{st} = \kappa^{ru}H_{stu}$ and $\nu^r_{stu} = \kappa^{rv}\nu_{stuv}$. Applying the method of recursive substitutions we obtain

$$\begin{aligned}(\tilde{\theta}-\theta)^r &= \Psi^r + \frac{1}{2}\nu^r_{st}\Psi^s\Psi^t + H^r_s\Psi^s + \frac{1}{6}\nu^r_{stu}\Psi^s\Psi^t\Psi^u + \frac{1}{2}\nu^r_{st}\nu^t_{wu}\Psi^s\Psi^w\Psi^u \\ &+ \frac{1}{2}H^r_{st}\Psi^s\Psi^t + \nu^r_{st}H^t_w\Psi^s\Psi^w + \frac{1}{2}H^r_s\nu^s_{ut}\Psi^u\Psi^t + H^r_sH^s_t\Psi^t + O_p(n^{-2}) \qquad (11)\end{aligned}$$

There is a formal similarity between equation (11) and the same expression for the MLE given e.g. in Pace and Salvan (1997, chap. 9). The first term of (11) gives

$$(\tilde{\theta}-\theta)^r = \kappa^{rs}\Psi_s + O_p(n^{-1}) \ ,$$

that coincides with the usual first order expansion for an M-estimator.

An expansion for the bias of $\tilde{\theta}$ is readily obtained by taking termwise expectations in (11). We find

$$E(\tilde{\theta}-\theta)^r = \frac{1}{2}\kappa^{ij}\kappa^{hk}\kappa^{lm}E(\Psi_k\Psi_m) + \kappa^{il}\kappa^{hj}E(\Psi_{hl}\Psi_j) + O(n^{-2}) \ . \qquad (12)$$

Also in this case there is a formal similarity between equation (12) and the expression of the bias of the ordinary MLE given e.g. in Pace and Salvan (1997, chap. 9).

References

Barndorff-Nielsen, O.E., Cox, D.R. (1996). Prediction and asymptotics. *Bernoulli*, **2**, 319–340.

Basu, A., Harris, I.R. (1994). Robust predictive distributions for exponential families. *Biometrika*, **81**, 790–794.

Corcuera, J.M., Giummolè, F. (2002). A note about simultaneous prediction. *Mathematics Preprint Series*, n. 229.2002, Institut de Matemàtica, University of Barcelona.

Fisher, A., Horn, P.S. (1994). Robust prediction intervals in a regression setting. *Computational Statistics and Data Analysis*, **17**, 129-140.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986). *Robust Statistics. The approach based on influence functions.* Wiley, New York.

Huber, P.J. (1981). *Robust statistics.* Wiley, New York.

Markatou, M., Ronchetti, E. (1997). Robust Inference: The approach based on influence functions. *Handbook of Statistics*, **15**, 49–75.

Pace, L., Salvan, A. (1997). *Principles of Statistical Inference.* World Scientific, Singapore.

Ronchetti, E., Vidoni, P. (1998). A robust predictive density based on the saddlepoint approximation for M-estimators. *Working Paper N. 8.1998*, Department of Econometrics, University of Geneva.

Vidoni, P. (1998). A note on modified estimative prediction limits and distributions. *Biometrika*, **85**, 949–953.

		true model		cont. model	
α	n	MLE	HAMPEL	MLE	HAMPEL
0.1	10	0.089	0.097	0.087	0.094
		(0.027)	(0.030)	(0.027)	(0.030)
	20	0.095	0.099	0.092	0.096
		(0.021)	(0.022)	(0.020)	(0.021)
	50	0.098	0.101	0.093	0.096
		(0.013)	(0.013)	(0.013)	(0.013)
	100	0.099	0.099	0.094	0.096
		(0.009)	(0.009)	(0.010)	(0.009)
0.2	10	0.178	0.192	0.176	0.188
		(0.051)	(0.055)	(0.052)	(0.056)
	20	0.192	0.199	0.185	0.192
		(0.039)	(0.040)	(0.038)	(0.039)
	50	0.197	0.199	0.188	0.192
		(0.024)	(0.025)	(0.025)	(0.025)
	100	0.198	0.199	0.189	0.192
		(0.017)	(0.017)	(0.018)	(0.018)
0.3	10	0.275	0.282	0.266	0.281
		(0.072)	(0.076)	(0.074)	(0.078)
	20	0.287	0.295	0.279	0.287
		(0.052)	(0.053)	(0.053)	(0.056)
	50	0.295	0.298	0.283	0.288
		(0.034)	(0.035)	(0.035)	(0.035)
	100	0.298	0.300	0.285	0.290
		(0.025)	(0.025)	(0.026)	(0.026)
0.4	10	0.367	0.385	0.359	0.378
		(0.087)	(0.091)	(0.092)	(0.0970)
	20	0.384	0.393	0.375	0.387
		(0.065)	(0.067)	(0.066)	(0.069)
	50	0.396	0.399	0.379	0.388
		(0.042)	(0.043)	(0.043)	(0.044)
	100	0.397	0.399	0.381	0.390
		(0.029)	(0.029)	(0.032)	(0.032)
0.5	10	0.463	0.480	0.453	0.469
		(0.100)	(0.104)	(0.107)	(0.110)
	20	0.482	0.491	0.471	0.484
		(0.074)	(0.076)	(0.076)	(0.078)
	50	0.495	0.499	0.476	0.488
		(0.048)	(0.049)	(0.049)	(0.049)
	100	0.496	0.499	0.479	0.492
		(0.033)	(0.033)	(0.037)	(0.037)
		true model		cont. model	
α	n	MLE	HAMPEL	MLE	HAMPEL
0.6	10	0.561	0.575	0.550	0.563
		(0.108)	(0.110)	(0.115)	(0.118)
	20	0.581	0.589	0.569	0.579
		(0.079)	(0.081)	(0.082)	(0.084)
	50	0.595	0.598	0.574	0.583
		(0.051)	(0.051)	(0.052)	(0.053)
	100	0.596	0.598	0.578	0.589
		(0.035)	(0.035)	(0.039)	(0.039)
0.7	10	0.662	0.671	0.650	0.659
		(0.110)	(0.111)	(0.118)	(0.120)
	20	0.681	0.686	0.670	0.679
		(0.080)	(0.081)	(0.082)	(0.084)
	50	0.695	0.697	0.674	0.677
		(0.050)	(0.051)	(0.053)	(0.053)
	100	0.696	0.697	0.678	0.684
		(0.035)	(0.035)	(0.039)	(0.040)
0.8	10	0.766	0.768	0.755	0.764
		(0.102)	(0.103)	(0.111)	(0.113)
	20	0.783	0.785	0.773	0.783
		(0.073)	(0.074)	(0.075)	(0.078)
	50	0.795	0.796	0.777	0.789
		(0.045)	(0.045)	(0.048)	(0.048)
	100	0.797	0.797	0.780	0.790
		(0.031)	(0.031)	(0.036)	(0.036)
0.9	10	0.876	0.869	0.866	0.871
		(0.080)	(0.083)	(0.087)	(0.093)
	20	0.888	0.885	0.880	0.885
		(0.054)	(0.056)	(0.057)	(0.060)
	50	0.897	0.896	0.883	0.889
		(0.032)	(0.033)	(0.036)	(0.036)
	100	0.898	0.897	0.884	0.890
		(0.022)	(0.023)	(0.027)	(0.027)

Table 1: Simulation results for the scale model.

α	n	true model MLE	true model HUBER	cont. model MLE	cont. model HUBER
0.5	20	0.500	0.500	0.506	0.500
		(0.094)	(0.096)	(0.200)	(0.107)
	50	0.495	0.494	0.502	0.500
		(0.084)	(0.086)	(0.138)	(0.071)
	100	0.499	0.498	0.497	0.499
		(0.053)	(0.054)	(0.084)	(0.058)
0.6	20	0.592	0.624	0.694	0.631
		(0.116)	(0.118)	(0.139)	(0.105)
	50	0.592	0.608	0.689	0.613
		(0.083)	(0.085)	(0.122)	(0.068)
	100	0.597	0.603	0.664	0.604
		(0.053)	(0.053)	(0.083)	(0.057)
0.7	20	0.686	0.741	0.831	0.754
		(0.109)	(0.108)	(0.082)	(0.097)
	50	0.689	0.714	0.836	0.724
		(0.078)	(0.079)	(0.097)	(0.067)
	100	0.696	0.707	0.774	0.710
		(0.049)	(0.049)	(0.080)	(0.052)
0.8	20	0.782	0.848	0.880	0.861
		(0.095)	(0.084)	(0.035)	(0.079)
	50	0.788	0.817	0.900	0.829
		(0.067)	(0.065)	(0.058)	(0.052)
	100	0.795	0.810	0.874	0.813
		(0.042)	(0.042)	(0.068)	(0.043)
0.9	20	0.882	0.936	0.997	0.945
		(0.071)	(0.047)	(0.099)	(0.050)
	50	0.889	0.915	0.981	0.925
		(0.047)	(0.042)	(0.021)	(0.034)
	100	0.895	0.909	0.956	0.912
		(0.029)	(0.028)	(0.043)	(0.029)

Table 2: Simulation results for the regression model.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy