The Annals of Statistics
1996, Vol. 24, No. 4, 1493–1510

ROBUST ESTIMATION OF PARAMETERS IN A MIXED UNBALANCED MODEL[1]

BY TADEUSZ BEDNARSKI AND STEFAN ZONTEK

Polish Academy of Sciences

This paper describes a method of robust estimation of shift and scale parameters in a mixed unbalanced interlaboratory model. Estimators presented result from "easily computable" Fréchet differentiable functionals which enjoy some optimal properties in a small neighborhood of the model. A rigorous treatment of their asymptotic behaviour under departures from the model assumptions and a simulation study are given.

1. Introduction. A method of robust analysis of random effects in mixed models is introduced in the paper by Rocke (1991) [see also Iglewicz (1983)] and a general treatment of robust estimators of variance components is presented by Fellner (1986). Rocke proposes estimators of variance components applying robust scale estimators to residuals and laboratory effects. In the first step he uses Huber's method to robustly assess location parameters. A similar point of view is taken by Lischer (1994), who improves breakdown properties of the estimators. He also gives a comprehensive practical motivation for the use of robust methods in interlaboratory experiments. The above-mentioned estimation technique may lead to high-dimensional nonlinear equations and it is "inconvenient" for assessing variability of estimators. A possible improvement may be in the simultaneous estimation of fixed effects and components of variation, where problems of a rigorous asymptotic treatment, smoothness of estimators and their optimality become more feasible. In a preliminary study by Bednarski, Zmyślony and Zontek (1992) it is shown that Fréchet differentiability of statistical functionals easily leads to reasonable robust estimators of variance components and treatment fixed effects in a simple interlaboratory model. The basic idea there is to construct a smooth Fisher consistent functional T for the parameter $\theta = (\mu_1, \ldots, \mu_a, \sigma_\lambda, \sigma_e)^t$ at the model

$$Y_i = \mu_i + \lambda + e_i \quad \text{for } i = 1, \ldots, a,$$

where the laboratory effect λ and errors e_i are independent normal variables with distributions $N(0, \sigma_\lambda^2)$ and $N(0, \sigma_e^2)$, respectively. The estimator has the form $T(\mathcal{F}_n)$, where $\mathcal{F}_n$ is the empirical distribution of the sample $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ with each $\mathcal{Y}_i$ having the distribution of $(Y_1, \ldots, Y_a)^t$. When we let the above model be unbalanced in the sense that the number of treatment replications may depend on both laboratory and treatment, then a similar point of view is

Received July 1993; revised November 1995.
[1]Research supported by KBN Grant 2 1052 9101.
AMS 1991 *subject classifications*. 62J10, 62F35.
Key words and phrases. Mixed models, robust estimation, von Mises method.

possible, though we deal with a number of subpopulations $F^1(\cdot|\theta), \ldots, F^p(\cdot|\theta)$ defined for different designs of treatment replications at various laboratories. The application of von Mises methodology requires then some modifications.

The main objective here is to construct a family of smooth functionals with "reasonable" robust and computational properties. The family of Fréchet differentiable functionals is obtained via "robustification" of the maximum likelihood method for linearly transformed vectors of observations [see Harville (1977) for the maximum likelihood approach to variance components estimation]. The transformations reduce the estimation problem to models with independent normal observations with parameters depending on the design, fixed effects and the components of variation. These make the problem's main ingredient the classical location-scale robust estimation. This reduction to the location-scale problems for independent observations was used already by Huggins (1993) in a robust analysis of variance components models related to random carrier models of Maronna and Yohai (1981). Huggins derived asymptotic properties of his robust M-estimators based on Tukey's bisquare function. We focus our considerations on an optimal choice of the basic function used for location-scale estimation and on smoothness and computability of the implicit functionals. It is shown in particular that the basic function corresponding to Huber's scale proposal minimizes the asymptotic variance of the location estimator under the condition of positive breakdown point for the scale estimator. Motivations for the smoothness analysis can be found in Bednarski, Clarke and Kołkiewicz (1991) and in Bednarski (1993, 1994), where it is shown that important robustness properties are a consequence of Fréchet differentiability. One of the very important statistical implications of the differentiability is a uniform asymptotic consistency of the estimate's variability assessment based on the influence function.

In the following section general information about Fréchet differentiability is presented along with some modifications specific to mixed models. Section 3 gives the construction of a class of Fréchet differentiable functionals and some resulting asymptotic properties for the corresponding estimators. The asymptotic distributions and approximation to the limiting covariance matrix are given under both the model assumptions and small departures from the model. Section 4 discusses the optimal choice for the basic objective function and compares it with Huber's approach and the approach based on the Tukey bisquare function. The final section provides results of a simulation experiment where behavior of the maximum likelihood and the robust estimator are compared. A significant conclusion there is, apart from robustness properties, a remarkable stability of approximated variances of the estimates under various model perturbations. The latter feature is attributed to the strong differentiability condition enjoyed by the class of functionals. Also a comparison with Rocke's estimator is given.

2. General information on Fréchet differentiability. Denote by $\mathscr{G}$ a set of distribution functions defined on R^r and let $\mathscr{D}$ be the convex cone spanned by the differences $F - G$ from $\mathscr{G}$. Here the distance between the

distributions F and G will always be defined by the supremum norm

$$\|F - G\| = \sup_{x \in R^r} |F(x) - G(x)|.$$

A statistical functional T defined from $\mathcal{S}$ to R^k is said to be Fréchet differentiable at $F \in \mathcal{S}$ when there exists a linear functional T^f on $\mathcal{D}$ such that

$$|T(G) - T(F) - T^f(G - F)| = o(\|F - G\|),$$

where $|\cdot|$ is the Euclidean distance. The use of the supremum norm in the context of robust statistics is justified in Bickel (1981). Other choices for the norm are possible depending on what properties of the estimators we search for [Dudley (1992)]. For detailed descriptions of other concepts of differentiability, we refer in particular to Reeds (1976), Fernholz (1983), Gill (1989) and to the classical work of von Mises (1947).

Clarke (1983, 1986) gives sets of conditions implying Fréchet differentiability for M-functionals. In the model studied later in this paper we shall consider differentiability of an M-functional, given by some function Ψ, at a product distribution $\mathcal{F}_\theta = F^1(\cdot|\theta) \times \cdots \times F^p(\cdot|\theta)$. If Ψ satisfies Clarke's conditions and $\int \Psi(\cdot|\theta) d\mathcal{F}_\theta(\cdot) = 0$ (this can always be assumed without loss of generality), then

$$T^f(G - \mathcal{F}_\theta) = -M(\theta)^{-1} \int \Psi(x|\theta)\, dG(x),$$

where

$$M(\theta) = \int \frac{\partial}{\partial \theta} \Psi(x|\theta)\, d\mathcal{F}_\theta(x).$$

To obtain the asymptotic distribution of the estimator resulting from the functional T, we need to approximate $\mathcal{F}_\theta$ on the basis of a random sample. This can be done via $\mathcal{F}_n = F^1_{n_1} \times \cdots \times F^p_{n_p}$, where $F^i_{n_i}$ is the empirical distribution function resulting from the subpopulation given by $F^i(\cdot|\theta)$ and $n = \sum n_i$. Another feature to be used is a specific structure of Ψ, namely, we shall have the representation

$$\Psi(Y_1, \ldots, Y_p|\theta) = \Psi_1(Y_1|\theta) + \cdots + \Psi_p(Y_p|\theta),$$

where Y_i is observed in the population with the distribution given by $F^i(\cdot|\theta)$. Then, under the assumption that n increases to infinity while $\min\{n_1/n, \ldots, n_p/n\}$ tends to a positive constant, we obtain

$$\sqrt{n}[T(\mathcal{F}_n) - \theta] = -M(\theta)^{-1} \sum_{i=1}^{p} \sqrt{n} \int \Psi_i(y_i|\theta)\, dF^i_{n_i}(y_i) + o_{G_n}(1)$$

under G_n equal to the product of $G_{n_i}^{\otimes n_i}$, where $\sqrt{n}\|G_{n_i}(\cdot) - F^i(\cdot|\theta)\|$ stays bounded. The asymptotic distribution is then normal with expectation zero (at the model) and the covariance matrix

$$M(\theta)^{-1}\left[\sum_{i=1}^{p} \frac{1}{q_i} \int \Psi_i(y_i|\theta)\Psi_i(y_i|\theta)^t\, dF^i(y_i|\theta)\right]\left[M(\theta)^{-1}\right]^t,$$

where $q_i = \lim_{n\to\infty}(n_i/n)$ for $i = 1, \ldots, p$, at the whole infinitesimal model given by the supremum norm. We shall sometimes call the parameter corresponding to the center of the infinitesimal neighborhood the "true" parameter.

One of the essential implications of the Fréchet differentiability is that the above covariance matrix can be approximated consistently over the infinitesimal neighborhoods by

$$(1)\quad \widehat{M}(\mathscr{F}_n)^{-1}\Big[\sum_{i=1}^{p}\frac{1}{q_i}\int \Psi_i(y_i|T(\mathscr{F}_n))\Psi_i(y_i|T(\mathscr{F}_n))^t\,dF^i_{n_i}(y_i)\Big]\big[\widehat{M}(\mathscr{F}_n)^{-1}\big]^t,$$

where

$$\widehat{M}(\mathscr{F}_n) = \sum_{i=1}^{p}\int\Big[\frac{\partial}{\partial\theta}\Psi_i(y_i|\theta)|_{\theta=T(\mathscr{F}_n)}\Big]\,dF^i_{n_i}(y_i).$$

In the following text we shall give a proposition for Ψ designed for the mixed unbalanced model.

3. The mixed unbalanced model. A single observation y_{ijk} for $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, d_{ij}$ is given in the model by

$$(2)\qquad y_{ijk} = \mu_i + \lambda_j + e_{ijk},$$

where μ_i are treatment fixed effects, λ_j are $N(0, \sigma^2_\lambda)$ independent laboratory effects and e_{ijk} are independent (also of λ_j) random errors with $N(0, \sigma^2_e)$ distribution.

To adopt the already described differentiability approach to this model, we need to single out model distributions for subpopulations for which independent identically distributed observations are available. If $\{d_{ij}\}$ denotes the incidence matrix for the mixed model with i and j corresponding to the treatment and laboratory effects, respectively, then the number p of different subpopulations will correspond to the number of distinct columns of the incidence matrix. Denote these columns further by $N_1, \ldots, N_p$ and let N be the matrix formed by the columns. If b_s is the number of repetitions of the experiment corresponding to $N_s = (n_{1s}, \ldots, n_{as})^t \in R^a$, then we have (after possible permutation of columns in $\{d_{ij}\}$)

$$\{d_{ij}\} = (N_1\mathbf{1}^t_{b_1} : \cdots : N_p\mathbf{1}^t_{b_p}) \quad\text{and}\quad b = \sum_{s=1}^{p} b_s.$$

Referring to our earlier notation we can now write

$$F^s(\cdot|\theta) \sim N(X_s\mu,\ \sigma^2_\lambda\mathbf{1}_{n_{\cdot s}}\mathbf{1}^t_{n_{\cdot s}} + \sigma^2_e I_{n_{\cdot s}}),$$

where $n_{\cdot s} = N^t_s\mathbf{1}_a$ for $i = 1, \ldots, p$, $\theta = (\mu', \sigma'_\lambda, \sigma_e)$, while $\mu = (\mu_1, \ldots, \mu_a)^t$. The matrix $X_s = \operatorname{diag}(\mathbf{1}_{n_{1s}}, \ldots, \mathbf{1}_{n_{as}})$ is a partitioned matrix with the column vectors of ones $\mathbf{1}_{n_{1s}}, \ldots, \mathbf{1}_{n_{as}}$, on the main diagonal and with zeros elsewhere.

To ensure the identifiability of the parameters in the model we further make the following assumptions:

I1. Every row in N has a nonzero element.

I2. There is a column in N for which the sum of its elements exceeds 1.

If a row of the incidence matrix $\{d_{ij}\}$ is zero, then a corresponding fixed effect cannot be identified. If $n_{\cdot s} > 1$, then the matrices $\mathbf{1}_{n_{\cdot s}}\mathbf{1}_{n_{\cdot s}}^t$ and $I_{n_{\cdot s}}$ are linearly independent yielding identifiability of components of variation. A simple condition for estimability of θ is ensured when there are at least two repetitions corresponding to the column in I2.

To validate the asymptotic argument, we shall need the assumption

$$\lim_{b\to\infty}(b_s/b) = q_s > 0 \tag{3}$$

for $s = 1, \ldots, p$.

For fixed s, let P_s be an $n_{\cdot s} \times n_{\cdot s}$ matrix with columns $P_{1s}, \ldots, P_{n_{\cdot s}s} \in R^{n_{\cdot s}}$ normed and orthogonal and such that $P_{1s} = (1/\sqrt{n}_{\cdot s})1_{n_{\cdot s}}$. Then the random vector Y_s corresponding to $F^s(\cdot|\theta)$ is transformed into

$$P_s^t Y_s \sim N(P_s^t X_s \mu,\ \operatorname{diag}(n_{\cdot s}\sigma_\lambda^2 + \sigma_e^2, \sigma_e^2, \ldots, \sigma_e^2)).$$

The objective function for the subpopulation s, taken as a simple modification of the loglikelihood function, can be written

$$\Phi_s(y|\theta) = \sum_{i=1}^{n_{\cdot s}}\left[\ln(\delta_{is}) + \phi\left(\frac{1}{c\delta_{is}}P_{is}^t(y - X_s\mu)\right)\right] \tag{4}$$

with a function ϕ properly chosen and δ_{is} the square roots of consecutive elements on the diagonal of the covariance matrix of $P_s^t Y_s$. The function Φ_s becomes proportional to the loglikelihood function when ϕ is taken as a quadratic function and $c = 1$. For a given ϕ, we can frequently choose c to make the functional Fisher consistent.

DEFINITION 3.1 (The functionals). *Define the functional $T^*(G)$ to be the parameter θ for which*

$$\int \Phi(y_1, \ldots, y_p|\theta)\, dG(y_1, \ldots, y_p) \tag{5}$$

attains the minimum value, where

$$\Phi(y_1, \ldots, y_p|\theta) = \sum_{s=1}^{p}\Phi_s(y_s|\theta) \tag{6}$$

is the objective function.

Define $T(G)$ to be the solution of the equation

$$\int \Psi(y_1, \ldots, y_p|\theta)\, dG(y_1, \ldots, y_p) = 0, \tag{7}$$

where Ψ is the vector of partial derivatives of Φ with respect to θ.

Conditions leading to Fréchet differentiability are usually given in terms of M-functionals. Since solutions to M-equations are frequently not unique, one needs an additional criterion to define the functional well [see Clarke (1983)]. This should justify our distinction between T^* and T. Such a definition is sufficient for the future asymptotic considerations and it does not affect the practical aspect of the paper.

Below are stated assumptions concerning ϕ which imply the Fisher consistency of the functional T^* (A1 and A2) and Fréchet differentiability of T (A3 and A4). The differentiability assumptions comply with Clarke's (1983) conditions. It is possible to weaken the conditions as follows from Clarke (1986). However, since one of the objectives here is also a computational feasibility of the method, we shall stay with the smoother case.

A1. The function ϕ defined on the real line is symmetric about 0 and has nonnegative derivative for positive arguments.

A2. The function $x\phi'(x)$ has a nonnegative derivative for $x \geq 0$ and there exist $x_o > 0$ such that $x_o\phi'(x_o) > 1$.

A3. The functions ϕ' and ϕ'' are bounded.

A4. The functions $x\phi'(x)$ and $x^2\phi''(x)$ are bounded.

THEOREM 3.1 (Fisher consistency). *Let $F^i(\cdot|\theta_0)$ be the distribution of Y_i for $i = 1, \ldots, p$, where $\theta_0 = (\mu_1^0, \ldots, \mu_a^0, \sigma_\lambda^0, \sigma_e^0)^t$ and assume that the model is identifiable* (I1 *and* I2).

(i) *If* (A2) *is satisfied, then there is a unique $c > 0$ defining Φ which satisfies*

$$\int \left\{ \frac{x}{c}\phi'\left(\frac{x}{c}\right) - 1 \right\} dF(x) = 0, \tag{8}$$

where F is the distribution function of the standard normal distribution.

(ii) *If ϕ satisfies* (A1) *and* (A2), *then*

$$\int \Phi(y_1, \ldots, y_p|\theta) d(F^1(y_1|\theta_0) \times \cdots \times F^p(y_p|\theta_0)) \tag{9}$$

attains the global minimum if and only if $\theta = \theta_0$.

PROOF. To conclude the uniqueness of c, let us notice that by the first part of (A2),

$$\frac{d}{d\gamma} g(\gamma) = -\frac{1}{\gamma} \int \frac{1}{\gamma}\left[\frac{x}{\gamma}\phi''\left(\frac{x}{\gamma}\right) + \phi'\left(\frac{x}{\gamma}\right)\right] dF(x) < 0,$$

where $g(\gamma) = \int (x/\gamma)\phi'(x/\gamma)\, dF(x)$, $\gamma > 0$. The function g is thus strictly decreasing. Moreover, $g(\gamma) > 1$ for sufficiently small γ by the second part of (A2) and $g(\gamma) < 1$ for sufficiently large γ. This terminates the proof of (i).

For the second part consider first the minimization problem of a single summand given in (4) and corresponding to a subpopulation s. The form of

the objective function and the smoothness assumptions about ϕ imply that we can further reduce the minimization problem to the one dimensional shift and scale case. There, as one can easily see, the global minimum is attained when

$$P_{1s}^t X_s \mu = P_{1s}^t X_s \mu^0 \quad \text{and} \quad n_{\cdot s}\sigma_\lambda^2 + \sigma_e^2 = n_{\cdot s}(\sigma_\lambda^0)^2 + (\sigma_e^0)^2$$

or

$$P_{is}^t X_s \mu = P_{is}^t X_s \mu^0 \quad \text{and} \quad \sigma_e = \sigma_e^0$$

for $i = 2, \ldots, n_{\cdot s}$, where $\mu^0 = (\mu_1^0, \ldots, \mu_a^0)^t$. Thus the minimum of the expected value for the subpopulation s is attained only when $\mu = \mu_0$ and $(\sigma_\lambda, \sigma_e) = (\sigma_\lambda^0, \sigma_e^0)$ under the identifiability assumption for the subpopulation. If, for instance $n_{1s} = 0$, then the first component of μ may vary arbitrarily in the set of shifts realizing the minimum. Similar difficulties incurred by nonidentifiability happen for scales when $n_{\cdot s} = 1$. Under the assumptions I1 and I2 we know, however, that for a given parameter there is always a subpopulation where it is identifiable. Therefore the expectation of the objective function over the whole model distribution attains the global minimum at a single point which is the true parameter value. □

REMARK 3.1. If the parameter in the submodel is identifiable, then (it follows from the proof of Theorem 3.1) the vector of partial derivatives of the expected value of (6) is zero only at the true parameter value. The authors do not know whether this kind of Fisher consistency holds for the whole model. This explains why the consistency in statement (ii) of Theorem 3.1 is given in terms of the objective function. From the computational point of view it is, however, even more desirable to search for the global minimum of the integral taken at the empirical distribution function. This issue will appear in detail later in Theorem 3.3.

Below we give a technical lemma which implies a simple form of the asymptotic covariance matrix.

Define, for $\sigma = (\sigma_\lambda, \sigma_e)^t$ and $s = 1, \ldots, p$, the matrices

$$U_s^{(1)}(\sigma) = \sigma_e^{-2}(\operatorname{diag} N_s - \sigma_\lambda^2 \delta_{1s}^{-2} N_s N_s^t)$$

and

$$U_s^{(2)}(\sigma) = 2[\delta_{1s}^{-2}(n_{\cdot s}\sigma_\lambda, \sigma_e)^t(n_{\cdot s}\sigma_\lambda, \sigma_e) + \sigma_e^{-2}\operatorname{diag}(0, n_{\cdot s} - 1)].$$

Let $\Psi_s^{(1)}(\cdot|\theta)$ and $\Psi_s^{(2)}(\cdot|\theta)$, $s = 1, \ldots, p$, be partitions of the vector function $\Psi_s(\cdot|\theta)$ composed of partial derivatives of Φ with respect to shift and scale parameters, respectively. Moreover, let

$$\Delta_s(\cdot|\theta) = \begin{bmatrix} \Delta_s^{11}(\cdot|\theta) & \Delta_s^{12}(\cdot|\theta) \\ \Delta_s^{12}(\cdot|\theta)^t & \Delta_s^{22}(\cdot|\theta) \end{bmatrix}$$

be the corresponding partition of the matrix $(\partial/\partial\theta)\Psi_s(\cdot|\theta)$.

LEMMA 3.1. *Assume* A1 *and* A2 *are satisfied and let the constant* c *be given by* (8). *Let* X *be a standard normal random variable. Then for* $s = 1, \ldots, p$ *we have*:

$$\text{(i)} \int \Psi_s^{(1)}(y_s|\theta)[\Psi_s^{(1)}(y_s|\theta)]^t \, dF^s(y_s|\theta) = c^{-2} E[\phi'(X/c)^2] U_s^{(1)}(\sigma),$$

$$\int \Delta_s^{11}(y_s|\theta) \, dF^s(y_s|\theta) = c^{-2} E[\phi''(X/c)] U_s^{(1)}(\sigma),$$

$$\text{(ii)} \int \Psi_s^{(2)}(y_s|\theta)[\Psi_s^{(2)}(y_s|\theta)]^t \, dF^s(y_s|\theta) = (1/2) E\{[(X/c)\phi'(X/c) - 1]^2\} U_s^{(2)}(\sigma),$$

$$\int \Delta_s^{22}(y_s|\theta) \, dF^s(y_s|\theta) = (1/2) E[(X/c)^2 \phi''(X/c) + 1] U_s^{(2)}(\sigma),$$

$$\text{(iii)} \int \Psi_s^{(1)}(y_s|\theta)[\Psi_s^{(2)}(y_s|\theta)]^t \, dF^s(y_s|\theta) = \int \Delta_s^{12}(y_s|\theta) \, dF^s(y_s|\theta) = 0.$$

PROOF. Let us fix the index s and let $\theta = (\mu^t, \sigma^t)^t$ be an arbitrary parameter. Define a random vector

$$Z = (Z_1, \ldots, Z_{n_{\cdot s}})^t = \operatorname{diag}^{-1}(\delta_{1s}, \ldots, \delta_{n_{\cdot s} s}) \; P_s^t (Y - X_s \mu),$$

where Y has the distribution $F^s(\cdot|\theta)$. Then Z has multivariate standard normal distribution.

(i) Note, that the random vector $\Psi_s^{(1)}(Y|\theta)$ can be represented as

$$\Psi_s^{(1)}(Y|\theta) = -\sum_{i=1}^{n_{\cdot s}} \frac{\phi'(Z_i/c)}{c\delta_{is}} \; T_{is},$$

where T_{is} is the ith column of $(P_s^t X_s)^t$, $i = 1, \ldots, n_{\cdot s}$. Since

$$E[\phi'(Z_i/c)\phi'(Z_j/c)] = \begin{cases} E[\phi'(X/c)^2], & i = j, \\ 0, & i \neq j, \end{cases}$$

we have

$$E\{\Psi_s^{(1)}(Y|\theta)[\Psi_s^{(1)}(Y|\theta)]^t\} = \frac{E[\phi'(X/c)^2]}{c^2} \sum_{i=1}^{n_{\cdot s}} \frac{1}{\delta_{is}^2} T_{is} T_{is}^t.$$

The first equation in part (i) of the lemma follows now by noting that $T_{1s} T_{1s}^t = (1/n_{\cdot s}) \mathbf{1}_{n_{\cdot s}} \mathbf{1}_{n_{\cdot s}}^t$ and $\sum_{i=1}^{n_{\cdot s}} T_{is} T_{is}^t = X_s^t P_s P_s^t X_s = \operatorname{diag} N_s$. The second equation can be shown to hold in a similar way, by using the relation

$$\Delta_s^{11}(Y|\theta) = \sum_{i=1}^{n_{\cdot s}} \frac{\phi''(Z_i/c)}{c^2 \delta_{is}^2} T_{is} T_{is}^t.$$

(ii) The first and the second partial derivatives of $\Psi_s(Y|\theta)$ with respect to σ_λ and σ_e are given by

$$\frac{\partial}{\partial\sigma_\lambda}\Phi_s(Y|\theta) = n_{\cdot s}\sigma_\lambda\delta_{1s}^{-2}\left[1-\left(\frac{Z_1}{c}\right)\phi'\left(\frac{Z_1}{c}\right)\right],$$

$$\frac{\partial}{\partial\sigma_e}\Phi_s(Y|\theta) = \sigma_e\sum_{i=1}^{n_{\cdot s}}\delta_{is}^{-2}\left[1-\left(\frac{Z_i}{c}\right)\phi'\left(\frac{Z_i}{c}\right)\right],$$

$$\frac{\partial^2}{\partial\sigma_\lambda^2}\Phi_s(Y|\theta) = n_{\cdot s}^2\sigma_\lambda^2\delta_{1s}^{-4}\left\{\left(\frac{Z_1}{c}\right)^2\phi''\left(\frac{Z_1}{c}\right)+1 + \frac{\sigma_e^2-2n_{\cdot s}\sigma_\lambda^2}{n_{\cdot s}\sigma_\lambda^2}\left[1-\left(\frac{Z_1}{c}\right)\phi'\left(\frac{Z_1}{c}\right)\right]\right\},$$

$$\frac{\partial^2}{\partial\sigma_\lambda\partial\sigma_e}\Phi_s(Y|\theta) = n_{\cdot s}\sigma_\lambda\sigma_e\delta_{1s}^{-4}\left\{\left(\frac{Z_1}{c}\right)^2\phi''\left(\frac{Z_1}{c}\right)+1 - 3\left[1-\left(\frac{Z_1}{c}\right)\phi'\left(\frac{Z_1}{c}\right)\right]\right\}$$

and

$$\frac{\partial^2}{\partial\sigma_e^2}\Phi_s(Y|\theta) = \sum_{i=1}^{n_{\cdot s}}\sigma_e^2\delta_{is}^{-4}\left\{\left(\frac{Z_i}{c}\right)^2\phi''\left(\frac{Z_i}{c}\right)+1 + (\delta_{is}^2-3\sigma_e^2)\left[1-\left(\frac{Z_i}{c}\right)\phi'\left(\frac{Z_i}{c}\right)\right]\right\}.$$

Applying now the equation (8) defining the constant c, one can easily derive the formulae (ii).

Part (iii) follows similarly. □

It results from Clarke (1983) that under assumptions A3 and A4 the M-functional T is Fréchet differentiable. We phrase it in the following theorem.

THEOREM 3.2 (Fréchet differentiability). *Under* A3 *and* A4 *we have*

$$[T(G)-T(\mathscr{F}_\theta)] = \int IF(y_1,\ldots,y_p|\theta)d(G-\mathscr{F}_\theta)(y_1,\ldots,y_p)+o(\|G-\mathscr{F}_\theta\|),$$

where

$$IF(y_1,\ldots,y_s|\theta) = \left[\sum_{s=1}^{p}\int\Delta_s(y_s|\theta)\,dF^s(y_s|\theta)\right]^{-1}\Psi(y_1,\ldots,y_p|\theta)$$

and $T(G)$ *is the functional given in Definition* 3.1.

The differentiability itself does not settle applicability of the estimation procedure. It is necessary to show that in a neighborhood of the true model distribution we can effectively find the functional's value. In fact, if we add to

A1–A4 the assumption that the second derivative of ϕ is uniformly continuous, then such a functional becomes effectively computable, at least in a small neighborhood of the model.

Take Θ_0 to be an open bounded subset of the parameter space such that $\mathrm{cl}(\Theta_0) \subset \Theta$. Let G_b be a sequence of distributions so that $\|G_b - \mathscr{F}_{\theta_0}\| < C_o/\sqrt{b}$ for a constant C_o and assume the integral of the objective function with respect to G_b is finite.

THEOREM 3.3. *Suppose* A1–A4 *hold and* ϕ'' *is uniformly continuous. Then there is a neighborhood of* $\theta_0 \in \Theta_0$, $U(\theta_0)$ *and* b_0 *such that, for all* $b > b_0$, *the integral* $\int \Phi(\cdot|\theta)\, dG_b$ *is strictly convex on* $U(\theta_0)$ *and it attains the minimum in a single point there* [*the functional* $T(G_b)$ *is uniquely defined in* $U(\theta_0)$].

PROOF. By Theorem 6.1 from Clarke (1983), it follows that

$$\sup_{\theta\in\Theta_0}\left[\left|\int \Psi(\cdot|\theta)d(\mathscr{F}_{\theta_0} - G)\right| + \left|\int \sum_{s=1}^{p} \Delta_s(\cdot|\theta)d(\mathscr{F}_{\theta_0} - G)\right|\right] = O(\|\mathscr{F}_{\theta_0} - G\|).$$

The uniform continuity of ϕ'' implies that there is a neighborhood $U^*(\theta_0)$ of θ_0 such that $\int \sum \Delta_s(\cdot|\theta)\, d\mathscr{F}_{\theta_0}$ is positive definite for all $\theta \in U^*(\theta_0) \subset \Theta$. Thus if $\|G_b - \mathscr{F}_{\theta_0}\| < \epsilon$ (for some $\epsilon > 0$), then $\int \Phi\, dG_b$ is also strictly convex in a neighborhood $U(\theta_0) \subset U^*(\theta_0)$ of θ_0 (if the integral exists). From the general properties of Fréchet differentiability [see Bednarski, Clarke and Kołkiewicz (1991)] it follows in turn that in neighborhoods of θ_0 of size $\|G_b - \mathscr{F}_{\theta_0}\|$ there are solutions to the M-equation. By strict convexity, which follows from Lemma 3.1, only one such solution exists in $U(\theta_0)$ if b is large. □

REMARK 3.2. When $\mathscr{F}_b$ are empirical distribution functions corresponding to G_b, then the integrals of the objective function with respect to $\mathscr{F}_b$ exist and since $\sqrt{b}\|\mathscr{F}_b - \mathscr{F}_{\theta_0}\|$ is bounded in probability G_b we can conclude that for large b there is a unique solution $T(\mathscr{F}_b)$ in $U(\theta_0)$ with probability approaching 1. From the practical point of view it might be important to give an initial approximation to the interval $U(\theta_0)$. This problem can be solved by a preliminary robust estimator, which is asymptotically consistent under the infinitesimal model. It is not difficult to show that such an estimator exists and it can be constructed. We can define it as an argument for which the global minimum of

$$\int_{K_b} \Phi(\cdot|\theta)\, d\mathscr{F}_b$$

in Θ is attained, where K_b is a sequence of compacts defined by $K_b = [-k\log b, k\log b]^{n_{\cdot\cdot}}$, where k is a positive constant while $n_{\cdot\cdot} = \sum_{s=1}^{p} n_{\cdot s}$. The consistency follows because $\arg\min(\int_{K_b} \Phi(\cdot|\theta)\, d\mathscr{F})$ converges uniformly to θ_0 while

$$\sup_{\theta\in\Theta}\left|\int_{K_b} \Phi(\cdot|\theta)\, d\mathscr{F}_{\theta_o} - \int_{K_b} \Phi(\cdot|\theta)\, d\mathscr{F}_b\right|$$

tends to zero in G_b when $b \to \infty$. The proximity of integral functions follows whenever any fixed power of the side length of K_b is of order $o(\sqrt{b})$.

REMARK 3.3. The two theorems describe both computational aspects and the asymptotic behavior of the functional (estimator) under general assumptions. Because of Fréchet differentiability we know that the linear expansion of the estimator holds in a small vicinity of the model (formally in the infinitesimal neighborhoods given by the supremum norm for distribution functions). The inequality of Kiefer (1961) and the central limit theorem yield the asymptotic normality of the estimator in the infinitesimal supermodel. The influence function obtained here *can really* be used to assess the variance of the obtained estimates.

Condition (3) leads to the kind of asymptotics where we are to imagine adding successive laboratories, where the design pattern for each laboratory must be chosen from a finite set. This, from the practical point of view, is an obvious inconvenience. It can be understood, however, as an indication for the design ensuring high stability of estimates of interest. Asymptotic approximations (not uniform) hold under much weaker conditions as proved by Huggins (1993) for the biweight function. His results can easily be transferred to cover the family of objective functions considered here. The simulation results in Section 5 show that the asymptotic approximation works very well even for amazingly small number of repetitions per subpopulation. We suppose that the differentiability approach combined with the Kiefer type inequality proved by Le Cam (1982) should lead to uniform expansions under a much wider range of practically important situations.

4. The choice of the objective function. Conditions A1 and A2 are sufficient for the Fisher consistency of location and scale, while requirements posed in A3 and A4 ensure the Fréchet differentiability. Since the function $x\phi'(x)$ must have a nonnegative derivative for $x \geq 0$ and it must be bounded [A2 and A4], the conditions specify indeed a narrow range of ϕ functions. In particular, they must descend in the tails at exactly x^{-1} rate. Below we argue that the class in fact contains the "good" robust estimators for a much larger family of "false likelihoods" given by (4) and (6).

Let us notice first that by (4) and (6) the estimation problem is reduced to a number of simultaneous one-dimensional robust estimation problems of location and scale under the normal model. In the one-dimensional case one chooses μ and σ to minimize $n \ln(\sigma) + \sum \phi((x_i - \mu)/\sigma)$. The defining equation for scale is then

$$n/\sigma - \sum (x_i - \mu)\phi'((x_i - \mu)/\sigma)/\sigma = 0 \, .$$

Now if $x\phi'(x)$ is not bounded for $|x|$ large, then the breakdown for σ must be zero. This leads naturally to the following question: under $|x\phi'(x)| \leq M$ for some constant M, when is the minimum variance for the location estimator achieved?

Using results from Luenberger (1969) or those in Appendix B of Rieder (1994) for the Lagrange multipliers and then arguing as in Huber [(1981), page 288], we easily arrive at a family of functions ϕ' of the form

$$\phi'(x) = \begin{cases} x, & |x| \le k, \\ k^2/x, & |x| < k, \end{cases}$$

which satisfies conditions complying with Clarke (1986).

Moreover, this same family of functions minimizes the variance of scale estimators given a bound on their bias span under the infinitesimal contamination of the normal law (bounding gross error sensitivity). In fact the same center and tail behavior is obtained when we minimize the variance for the location under a bound on gross error sensitivity for the scale. The false likelihood method cannot lead here to optimal robust location estimation if we want the scale estimator to have its breakdown greater than zero (the Fréchet differentiability "automatically" enforces nonzero breakdown for the scale). One should realize at this point that all this does not preclude construction of robust functionals based on other principles. For example, Huber's [(1981), pages 176 and 177] location-scale proposition resulting from a minimization problem leads also to Fréchet differentiable and Fisher consistent functionals at the mixed model. The important merit of the method presented here is that it also leads to an "easy" computational method in the case of interlaboratory designs. It is intuitively clear that the method is a compromise between the simultaneous location-scale solution given by Huber's solution [Huber (1981), page 137, Example 4.1] and the Huggins solution based on the bisquare function. It is less sensitive to large outliers than Huber's function and at the same time it does not lose as much efficiency as the biweight function with nearly as good resistance to outliers.

The function ϕ (defined below by its derivative),

$$\phi'(x) = \begin{cases} x, & |x| \le t, \\ -x - 4t - \dfrac{2t^2}{x}, & -2t < x < -t, \\ -x + 4t - \dfrac{2t^2}{x}, & t < x < 2t, \\ \dfrac{2t^2}{x}, & |x| \ge 2t, \end{cases} \tag{10}$$

satisfies the assumptions A3 and A4. It is in fact a smoothed version of Huber's proposition for scale estimation.

It follows from Lemma 3.1 that the covariance matrix for the Fréchet differentiable estimator generated by (6) is of the form

$$\begin{bmatrix} w_1 V_1(\theta) & 0 \\ 0 & w_2 V_2(\theta) \end{bmatrix},$$

where $V_1(\theta)$ and $V_2(\theta)$ are respective asymptotic covariance matrices for the shift and scale of the maximum likelihood estimator given by

$$V_i(\theta) = \left[\sum_{s=1}^{p} U_s^{(i)}(\sigma)\right]^{-1}\left[\sum_{s=1}^{p} \frac{1}{q_s} U_s^{(i)}(\sigma)\right]\left[\sum_{s=1}^{p} U_s^{(i)}(\sigma)\right]^{-1}.$$

The positive constants w_1 and w_2 are given by the formulas

$$w_1 = \frac{c^2 E[\phi'(X/c)^2]}{[E\phi''(X/c)]^2},$$

$$w_2 = \frac{2E\{[(X/c)\phi'(X/c) - 1]^2\}}{\{E[(X/c)^2\phi''(X/c) + 1]\}^2}.$$

For each t in (10) we choose c satisfying (8) so as to make the functional Fisher consistent.

5. Simulation results. The robust estimator applied in the simulation was given by (9) with constant $t = 1.253$ selected in such a way that the loss of efficiency for the shift estimator was 10%. More exactly, it then gave $w_1 = 1.1$ and $w_2 = 1.29$ ($c = 0.873$). We considered the following three unbalanced mixed models in our simulation study:

M1. The shift parameter is two dimensional and equal to $(-4, 4)^t$. The scales are $\sigma_\lambda = 2$ and $\sigma_e = 1$ and subpopulation sample sizes are $b_1 = 5$ and $b_2 = 5$, while

$$N = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

M2. Except for $b_1 = 15$ and $b_2 = 15$, other quantities defining the model remain the same as in M1.

M3. The shift parameter is a three-dimensional zero vector, $\sigma_\lambda = 2$, $\sigma_e = 1$, $b_1 = 7$, $b_2 = 5$ and $b_3 = 3$, while

$$N = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 0 & 2 \\ 1 & 4 & 3 \end{bmatrix}.$$

Models M1 and M2 differ in sample sizes only and they were considered to see how the number of laboratories affects the precision of estimation of the covariance matrix via the influence function both under the pure and contaminated model for the maximum likelihood estimator and robust estimator. The third model has a "more complex" structure, with empty cells and relatively small and different number of observations per subpopulation. It is meant to give an overall impression of the value of the robust estimator. Contaminations of the data simulated from the model were taken to imitate (1) exchange of samples in a laboratory, (2) wrong scaling of an instrument in

a laboratory, (3) higher occasional variability in laboratory effects, (4) higher occasional effects in error effects and (5) occasional interaction.

Only the first type of contamination may be identified with gross errors. For the other types of discrepancies the "erroneous" data were mildly contaminated: one laboratory was randomly chosen to be "contaminated" for the model M1 and exactly two laboratories were chosen in the case of M2. The following modifications of the original model distribution were applied to the randomly selected laboratories:

1. The mean was taken as $(4, -4)$ instead of $(-4, 4)$.
2. The mean was taken as $(-3, 5)$.
3. σ_λ was taken equal to 4.
4. σ_e was taken equal to 2.
5. An interaction random effect between a laboratory and a sample was taken to be $N(0, 2.25)$.

In the tables herein typical results of simulations are given. In each case the estimation was repeated 500 times. The initial values for the estimation were randomly chosen (according to the uniform distribution) from the intervals of length 1 and centered at the true parameter value. Table 1 lets us judge how

TABLE 1

	Model M1 estimates for parameters				**Model M2 estimates for parameters**			
	μ_1	μ_2	σ_λ	σ_e	μ_1	μ_2	σ_λ	σ_e
m.l.e.	−4.01	4.01	1.83	0.96	−3.99	4.00	1.93	0.99
	(0.68)	(0.69)	(0.50)	(0.16)	(0.38)	(0.40)	(0.27)	(0.09)
	(0.65)	(0.65)	(0.39)	(0.14)	(0.39)	(0.39)	(0.26)	(0.09)
	(0.69)	(0.69)	(0.49)	(0.16)	(0.40)	(0.40)	(0.28)	(0.09)
r.e.	−4.02	4.00	1.85	0.97	−3.99	4.00	1.93	0.99
	(0.70)	(0.72)	(0.57)	(0.17)	(0.40)	(0.42)	(0.31)	(0.11)
	(0.77)	(0.78)	(0.68)	(0.19)	(0.42)	(0.42)	(0.33)	(0.10)
	(0.72)	(0.72)	(0.55)	(0.18)	(0.42)	(0.42)	(0.32)	(0.10)
m.l.e.	−3.19	3.19	1.04	2.90	−3.47	3.48	1.55	2.51
	(0.69)	(0.71)	(0.79)	(0.21)	(0.40)	(0.40)	(0.40)	(0.13)
Case 1	(1.03)	(1.02)	(1.28)	(1.04)	(0.55)	(0.55)	(0.53)	(0.68)
	(0.69)	(0.69)	(0.49)	(0.16)	(0.40)	(0.40)	(0.28)	(0.09)
r.e.	−3.97	3.97	1.94	1.06	−3.98	4.01	2.05	1.05
	(0.73)	(0.74)	(0.60)	(0.20)	(0.42)	(0.42)	(0.33)	(0.12)
Case 1	(0.81)	(0.82)	(0.74)	(0.23)	(0.44)	(0.44)	(0.35)	(0.12)
	(0.72)	(0.72)	(0.55)	(0.18)	(0.42)	(0.42)	(0.32)	(0.11)
m.l.e.	−3.98	4.02	2.10	0.95	−3.98	4.02	2.13	0.99
	(0.70)	(0.71)	(0.63)	(0.15)	(0.41)	(0.41)	(0.34)	(0.09)
Case 3	(0.74)	(0.74)	(0.50)	(0.14)	(0.42)	(0.42)	(0.33)	(0.09)
	(0.69)	(0.69)	(0.49)	(0.16)	(0.40)	(0.40)	(0.28)	(0.09)
r.e.	−4.00	4.01	2.00	0.96	−3.98	4.02	2.05	0.99
	(0.72)	(0.72)	(0.63)	(0.17)	(0.41)	(0.41)	(0.33)	(0.10)
Case 3	(0.82)	(0.82)	(0.75)	(0.18)	(0.44)	(0.44)	(0.35)	(0.10)
	(0.72)	(0.72)	(0.55)	(0.18)	(0.42)	(0.42)	(0.32)	(0.10)

rapidly the number of "observations" improves the credibility of the variability estimator obtained via the influence function (models M1 and M2). The tables are set up in two-row blocks named by the model and, if necessary, also by contamination type. The first row in each block is always for the maximum likelihood estimator, while the second row is for the robust estimator. Each model row consists of four subrows: the first one is for the average of the 500 estimates, the second one gives sample standard deviation for the estimates, the third one shows the standard deviations for the estimates obtained via the influence functions and the last subrow gives the "true" asymptotic value for the standard deviation. Even though it is better to evaluate the variability of the maximum likelihood estimate via the asymptotic covariance matrix (under the model), we decided to use the "influence function approach" to make the two methods comparable under departures from the hypothetical normal model. We used

$$\widehat{M}(\mathcal{F}_b) = \sum_{s=1}^{p} \int \Delta_s(\cdot|T(\mathcal{F}_b))\, dF^s_{b_s}.$$

Still another possibility of assessing the covariance matrix of the estimators would be by the approximation of $M(\theta)$ via $M[T(\mathcal{F}_b)]$. It turned out, however, to be less accurate when sample sizes were relatively small.

Under the model distribution the robust estimator has about 5% higher standard deviation for the shift and about 10% for the scale estimators. The robust estimator shows, however, much higher performance under small contaminations in both the accuracy of estimation in terms of unbiasedness and in terms of the assessment of estimate's variability via the influence function. In the case of modification 1 applied to M1 and M2 these effects are especially visible. Apart from a clear bias superiority of the robust estimator under the model violations, let us also notice higher accuracy of the variability assessment based on influence functions. This is ensured by the Fréchet differentiability of the functional generating the robust estimator.

Simulation results for models M1 and M2 in the cases of modifications 2, 4, and 5 are similar to those of modification 3 and are not presented. Mixtures of different contamination patterns do not much change the picture. In fact comparable simulation effects were obtained for M3 in the sense that only "gross error" of type 1 led to the breakdown of the maximum likelihood estimator. To validate our simulations with a richer spectrum of contamination patterns, we include results of Cauchy type contaminations for M3 (Table 2). For example, the meaning of M3 1% is that each laboratory normal random effect was substituted with probability 0.01 by Cauchy random effect (with scales of the model random effect) and independently with the same probability the random error effect was replaced by the Cauchy random error. The outcome again strongly favors the robust estimator under departures from normality.

Table 3 compares our results of simulations with two methods for estimation of the laboratory variance component presented in Rocke (1991) in the case of 10 laboratories and 10 samples. We use Rocke's notation: standard normal (N),

TABLE 2

	Model M3 estimates for parameters				
	μ_1	μ_2	μ_3	σ_λ	σ_e
m.l.e.	−0.06	−0.02	−0.01	1.88	0.98
	(0.66)	(0.59)	(0.58)	(0.41)	(0.11)
	(0.64)	(0.58)	(0.56)	(0.34)	(0.10)
	(0.67)	(0.61)	(0.58)	(0.41)	(0.11)
r.e.	−0.06	−0.02	−0.00	1.89	0.98
	(0.70)	(0.64)	(0.61)	(0.48)	(0.13)
	(0.75)	(0.67)	(0.65)	(0.51)	(0.12)
	(0.71)	(0.64)	(0.61)	(0.47)	(0.12)
m.l.e.	−0.05	−0.09	−0.01	1.98	1.28
	(0.88)	(0.92)	(1.10)	(1.52)	(3.08)
1%	(0.81)	(0.85)	(1.06)	(0.77)	(1.40)
	(0.67)	(0.61)	(0.58)	(0.41)	(0.11)
r.e.	−0.00	−0.03	−0.01	1.88	0.98
	(0.74)	(0.68)	(0.62)	(0.50)	(0.13)
1%	(0.76)	(0.66)	(0.64)	(0.53)	(0.13)
	(0.71)	(0.64)	(0.61)	(0.47)	(0.12)
m.l.e.	0.04	−0.09	0.01	2.70	2.22
	(4.07)	(2.25)	(2.14)	(6.43)	(5.79)
5%	(3.82)	(2.11)	(2.06)	(3.18)	(2.75)
	(0.67)	(0.61)	(0.58)	(0.41)	(0.11)
r.e.	0.00	−0.02	−0.02	1.98	1.01
	(0.73)	(0.68)	(0.63)	(0.53)	(0.15)
5%	(0.76)	(0.71)	(0.66)	(0.62)	(0.15)
	(0.71)	(0.64)	(0.61)	(0.47)	(0.12)
m.l.e.	−0.35	−0.41	−0.91	4.02	4.86
	(3.61)	(3.41)	(13.83)	(12.08)	(48.37)
10%	(4.38)	(4.28)	(13.05)	(5.70)	(20.89)
	(0.67)	(0.61)	(0.58)	(0.41)	(0.11)
r.e.	−0.01	-0.04	−0.05	2.10	1.10
	(0.76)	(0.70)	(0.66)	(0.60)	(0.19)
10%	(0.82)	(0.75)	(0.72)	(0.70)	(0.20)
	(0.71)	(0.64)	(0.61)	(0.47)	(0.12)

a mixture of 90% standard normal and 10% normal with a standard deviation of 3 (LT) or a mixture in which 5% of observations have 10 times the standard deviation (VLT). The comparison is given in terms of the averages (first row) and standard deviations (second row) across the 500 replications. The indexes H and B correspond to the application of Huber's and biweight functions, respectively. The simulations show a fairly uniform advantage of the method presented here in the case of contaminated data.

Acknowledgment. The authors would like to thank the referees for valuable comments and suggestions.

TABLE 3

Labs	Error	$\hat{V}_1^H$	$\hat{V}_1^B$	$\hat{V}_2^H$	$\hat{V}_2^B$	r.e.
N	N	1.01	0.75	1.01	1.02	0.92
		(0.55)	(0.48)	(0.52)	(0.53)	(0.53)
N	VLT	1.01	0.76	1.22	1.18	1.12
		(0.56)	(0.49)	(0.71)	(0.66)	(0.63)
LT	N	1.62	1.10	1.72	1.66	1.17
		(1.38)	(0.99)	(1.41)	(1.31)	(0.79)
LT	LT	1.64	1.11	1.90	1.88	1.37
		(1.43)	(1.01)	(1.56)	(1.47)	(1.18)
VLT	N	2.16	1.21	2.32	1.64	1.21
		(3.87)	(2.69)	(3.94)	(2.79)	(1.10)
VLT	VLT	2.17	1.23	2.80	1.95	1.38
		(3.91)	(2.72)	(4.65)	(3.14)	(1.28)

REFERENCES

BEDNARSKI, T. (1993). Robust estimation in the Cox regression model. *Scand. J. Statist.* **20** 213–225.

BEDNARSKI, T. (1994). Fréchet differentiability and robust estimation. In *Asymptotic Statistics. Proceedings of the Fifth Prague Symposium* 49–58. Springer, Berlin.

BEDNARSKI, T., CLARKE, B. R. and KOLKIEWICZ, W. (1991). Statistical expansions and locally uniform Fréchet differentiability. *J. Austral. Math. Soc. Ser. A* **50** 88–97.

BEDNARSKI, T., ZMYŚLONY, R. and ZONTEK, S. (1992). On robust estimation of variance components via von Mises functionals. Preprint 504, Institute of Mathematics, Polish Academy of Sciences.

BICKEL, P. J. (1981). *Quelques aspects de la Statistique Robuste. Lecture Notes in Math.* **876** 1–72. Springer, Berlin.

CLARKE, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11** 1196–1206.

CLARKE, B. R. (1986). Nonsmooth analysis and Fréchet differentiability of M-functionals. *Probab. Theory Related Fields* **73** 197–209.

DUDLEY, R. M. (1992). Fréchet differentiability, p-variation and uniform Donsker classes. *Ann. Probab.* **20** 1968–1982.

FELLNER, W. H. (1986). Robust estimation of variance components. *Technometrics* **28** 51–60.

FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals. Lecture Notes in Statistics* **19**. Springer, New York.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. Part 1. *Scand. J. Statist.* **16** 97–128.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–338.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

HUGGINS, R. M. (1993). On the robust analysis of variance components models for pedigree data. *Austral. J. Statist.* **35** 43–57.

IGLEWICZ, B. (1983). Robust scale estimators and confidence intervals for location In *Understanding Robust and Exploratory Data Analysis* (D. C. Hoagling, F. Mosteller and J. W. Tukey, eds.) 404–431. Wiley, New York.

KIEFER, J. (1961). On large deviations of the empiric D.F. of vector chance variables and a law of iterated logarithm. *Pacific J. Math.* **11** 649–660.

LE CAM, L. (1982). Limit theorems for empirical measures and Poissonization. In *Statistics and Probability: Essays in Honor of C. R. Rao* 455–463. North-Holland, Amsterdam.

LISCHER, P. (1994). Robust statistical methods in interlaboratory analytical studies. Unpublished manuscript.

LUENBERGER, D. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behaviour of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. Verw. Gebiete* **58** 7–20.

REEDS, J. A. (1976). On the definition of von Mises functions. Ph.D. dissertation, Dept. Statistics, Harvard Univ.

RIEDER, H. (1994). *Robust Asymptotic Statistics*. Springer, New York.

ROCKE, D. M. (1991). Robustness and balance in the mixed model. *Biometrics* **47** 303–309.

VON MISES, R. (1947). On the asymptotic distributions of differentiable statistical functionals. *Ann. Math. Statist.* **18** 309–348.

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
UL. KOPERNIKA 18
51-617 WROCŁAW
POLAND