Computational Statistics & Data Analysis 13 (1992) 211–220
North-Holland

A note on the computation of robust, bounded influence estimates and test statistics in regression *

Douglas P. Wiens
University of Alberta, Edmonton, Alb., Canada T6G 2G1

Abstract: We give a method of computing bounded influence M-estimates of regression coefficients. The method has the advantage that the accompanying printout will be asymptotically correct, in that the standard errors and p-values have the correct asymptotic values. The method is easily implemented on any package which can perform weighted least squares regression, and Choleski decompositions. The p-values are those which result from substituting robust estimates into the usual F-statistic for testing a general linear hypothesis. The influence function of this test is obtained, and shown to be bounded with respect both to the influence of residuals, and to the influence of the position of the carriers. A numerical example is given, comparing several robust estimators, and the least squares estimator.

Keywords: Robust regression, Bounded influence estimation.

1. Introduction

A barrier to the widespread adoption of robust regression procedures within the statistical community appears to be the perception that such procedures are difficult to compute. Street, Carroll and Ruppert (1988) addressed this problem in the case of Huber-type M-estimation. One of their main points may be summarized as follows.

Consider the linear model

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\theta} + \varepsilon_i, \qquad 1 \le i \le n;$$

with i.i.d. errors ε_i. Suppose that one has obtained a Huber M-estimate $\boldsymbol{\theta}_*$, through solving the defining equations

$$n^{-1}\sum_{i=1}^{n}\psi\left(\frac{y_i - \boldsymbol{x}_i^T\boldsymbol{\theta}_*}{\hat{\sigma}}\right)\boldsymbol{x}_i = \boldsymbol{0}.$$

One can then perform a least squares regression of appropriately defined *pseudovalues* on the independent variables, obtaining an asymptotically equiva-

* Research supported by the Natural Sciences and Engineering Research Council of Canada

lent estimate $\hat{\boldsymbol{\theta}}$. If this is carried out on any of the usual software packages, then the accompanying printout will be asymptotically correct. That is:

(a) The printed estimated covariance matrix of $\hat{\boldsymbol{\theta}}$ will be a consistent estimate of the true covariance matrix.

(b) If $\boldsymbol{\theta}$ is partitioned as

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \tag{1}$$

then the classical test of the hypothesis $H_0: \boldsymbol{\theta}_2 = \mathbf{0}$ consists of rejecting H_0 if $F > F(1-\alpha;\ p-q,\ n-p)$, where

$$F = \{\text{Regression sum of squares due to the last } (p-q) \text{ regressors}\} \div (p-q)S^2. \tag{2}$$

These regression sums of squares are typically printed out, so that F is easily calculated. By virtue of the regression on pseudovalues, the mean residual sum of squares S^2 estimates the "right" quantity. This F-test is then asymptotically of the correct size α, since the distributions of both F and an F_{n-p}^{p-q} random variable tend weakly to the same $\chi_{p-q}^2/(p-q)$ distribution.

In this article we discuss a similar computational approach, valid for *bounded influence estimators*. In contrast to Huber M-estimators, bounded influence estimators are robust against high leverage points. The estimate may be defined as a solution to

$$n^{-1} \sum_{i=1}^{n} \eta\left(\boldsymbol{x}_i, \frac{y_i - \boldsymbol{x}_i^T\boldsymbol{\theta}}{\sigma}\right)\boldsymbol{x}_i = 0. \tag{3}$$

A common class of functions η is that suggested by Schweppe (see Hampel, Ronchetti, Rousseeuw and Stahel (1986)):

$$\eta(\boldsymbol{x}, r) = v(\boldsymbol{x})\psi(r/v(\boldsymbol{x})) \tag{4}$$

for an appropriate positive function $v(\boldsymbol{x})$ and odd function $\psi(r)$. Specific choices of v and ψ are investigated in Sections 3 and 4 below. See Hampel, Ronchetti, et al. (1986), Krasker and Welsh (1982), Huber (1983) for further discussions related to the choice of the function η.

A common, easily programmed algorithm for solving (3) is as follows:

(i) From $\boldsymbol{\theta}_{(k)}$ $(k = 0, 1, \ldots)$ compute a scale estimate $\sigma_{(k)}$. Put $e_{i,k} = y_i - \boldsymbol{x}_i^T\boldsymbol{\theta}_{(k)}$, $r_{i,k} = e_{i,k}/\sigma_{(k)}$.

(ii) Form weights $w_{i,k} = \eta(\boldsymbol{x}_i, r_{i,k})/r_{i,k}$. With $\boldsymbol{\theta} = \boldsymbol{\theta}_{(k)}$, (3) now becomes the weighted least squares problem

$$\sum_{i=1}^{n} w_{i,k} y_i \boldsymbol{x}_i = \left(\sum_{i=1}^{n} w_{i,k} \boldsymbol{x}_i \boldsymbol{x}_i^T\right)\boldsymbol{\theta} \tag{5}$$

with weights dependent upon $\boldsymbol{\theta}_{(k)}$.

(iii) "Solve" (5), thus obtaining $\boldsymbol{\theta}_{(k+1)}$, by performing a weighted least squares regression of the y's on the $\boldsymbol{x}$'s, using weights $w_{i,k}$.
(iv) Iterate to convergence.

A frequent choice of $\sigma_{(k)}$, suggested by Hill and Holland (1977), is

$$\sigma_{(k)} = 1.48 \times \{\text{Median of the largest } n-p+1 \text{ of the } |e_{i,k}|\}. \tag{6}$$

The factor $1.48 (= 1/\Phi^{-1}(0.75))$ is for consistency at the normal distribution.

The final values $(\boldsymbol{\theta}_*, \sigma_*)$ of the iterates are, under mild conditions (see Maronna and Yohai (1981)), consistent and asymptotically normally distributed, and

$$\sqrt{n}(\boldsymbol{\theta}_* - \boldsymbol{\theta}) \xrightarrow{w} N(\mathbf{0}, \sigma^2 C^{-1}). \tag{7}$$

Here, $\sigma = \text{plim } \hat{\sigma}$ and $n \to \infty$

$$C = MQ^{-1}M,$$

$$M = E\left[\eta'\left(x, \frac{\varepsilon}{\sigma}\right)\boldsymbol{x}\boldsymbol{x}^T\right], \left(\eta'(x, \varepsilon) = \frac{\partial}{\partial \varepsilon}\eta(x, \varepsilon)\right)$$

$$Q = E\left[\eta^2\left(x, \frac{\varepsilon}{\sigma}\right)\boldsymbol{x}\boldsymbol{x}^T\right].$$

The regressors x may be random, in which case they are assumed to be distributed independently of ε. Note that if (4) is used, then $\eta'(\boldsymbol{x}, \varepsilon) = \psi'(\varepsilon/v(\boldsymbol{x}))$.

A further problem is the estimation of the matrix C, and the computation of test procedures for which estimates of C are required. Street, Carroll and Ruppert (1988) pointed out that the estimated standard errors normally appearing on the weighted least squares printout are inconsistent, in the case of Huber M-estimation. For bounded influence estimation such estimates are not merely inconsistent but meaningless, due to the relatively complex structure of C.

In Section 2 below we give an easily implemented method, involving a further least squares regression on pseudovalues, which has both properties (a) and (b) above. We give as well the non-centrality parameter in the limiting χ^2-distribution of the test statistic, when H_0 is false. In Section 3 the influence function of the test is presented. It is shown that, provided $\psi(r)$ and $\|\boldsymbol{x}\| v(\boldsymbol{x})$ are bounded, the influence function is bounded with respect to both the influence of residuals, and to the influence of the position of the carriers. Two appropriate choices of $v(\boldsymbol{x})$ are given. Some numerical comparisons are made in the example of Section 4.

The test statistic determined by (2), following the regression on pseudovalues, is in fact identical to $R_n^2/(p-q)$, in the notation of Hampel et al. (1986, p. 364). Hampel et al. discuss $R_n^2/(p-q)$ as well as another statistic for testing H_0 – their $W_n^2/\hat{\sigma}^2$ – and several asymptotically equivalent versions. In two of the cases in which they are able to explicitly evaluate the statistics and their asymptotic distributions – Huber M-estimation, and the case $p-q=1$ – it turns out that the F of (2), $R_n^2/(p-q)$, and $W_n^2/\hat{\sigma}^2$ are in fact all identical. In

general, however, the F of (2) does not agree with $W_n^2/\hat{\sigma}^2$. In such cases, the F-based test enjoys the advantage that, as described in Section 2, both the test statistic and its asymptotic distribution are easier to compute than are those for $W_n^2/\hat{\sigma}^2$.

We note that there are several sophisticated, main-frame based computer packages available for the computation of bounded influence estimates. See Marazzi (1987) for one such package. The procedures outlined here have the advantage of being easily implemented by the casual user, or by students. Indeed, a program which runs on MINITAB has been used successfully in classes, and is available from the author.

2. Computation of the estimates

Suppose that (θ_*, σ_*) satisfy (3), with σ_* determined from θ_* as at (6), or in any other manner which ensures its consistency for σ. Let M_n, Q_n be consistent estimates of M and Q, e.g.

$$M_n = n^{-1} \sum_{i=1}^{n} \eta'(x_i, r_i) x_i x_i^T,$$

$$Q_n = n^{-1} \sum_{i=1}^{n} \eta^2(x_i, r_i) x_i x_i^T,$$

where $r_i = (y_i - x_i^T \theta_*)/\sigma_*$. Assume that M_n, Q_n and the design matrix X are of full rank p. Let

$$C_n = M_n Q_n^{-1} M_n,$$

and let A_n be an upper triangular matrix satisfying

$$A_n^T A_n = n C_n. \tag{8}$$

Decompose the design matrix as

$$X = \Gamma U, \tag{9}$$

where $\Gamma : n \times p$ has orthonormal columns and U is upper triangular. Note that (8), (9) involve only Choleski decompositions. Define

$$V_n = \Gamma A_n = X U^{-1} A_n, \quad \eta(x, r) = (\eta(x_1, r_1), \ldots, \eta(x_n, r_n))^T. \tag{10}$$

Then by (3) and (10),

$$X^T \eta(x, r) = V_n^T \eta(x, r) = 0. \tag{11}$$

Compute a vector of pseudovalues

$$y_* = V_n \theta_* + k_n \eta(x, r),$$

where

$$k_n = \sqrt{n - p}\, \sigma_* / \| \eta(x, r) \|.$$

Regress y_* on the columns of V_n-remembering to fit a no-intercept model since V_n does not have a column of ones – to obtain a final estimate $\hat{\boldsymbol{\theta}}$. By virtue of (11),

$$\hat{\boldsymbol{\theta}} = (V_n^T V_n)^{-1} V_n^T \mathbf{y}_* = \boldsymbol{\theta}_*.$$

Define $\hat{\sigma}$ to be σ_*.

On typical regression packages, the printout for this final regression will include standard errors of estimates, p-values, etc. determined from the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$. This matrix is calculated as

$$\text{"}est.cov.(\hat{\boldsymbol{\theta}})\text{"} = s^2(V_n^T V_n)^{-1} = \frac{s^2}{n} C_n^{-1},$$

where

$$\begin{aligned} s^2 &= \| \left(I - V_n(V_n^T V_n)^{-1} V_n^T\right) \mathbf{y}_* \|^2/(n-p) \\ &= k_n^2 \| \boldsymbol{\eta}(\boldsymbol{x}, r) \|^2/(n-p) \\ &= \hat{\sigma}^2. \end{aligned}$$

Since C_n is consistent for C and $\hat{\sigma}$ for σ, objective (a) of Section 1 is met. To see that objective (b) is met as well, first decompose $\hat{\boldsymbol{\theta}}$ as $(\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T)^T$, compatibly with (1). Standard algebraic manipulations show that the F statistic, calculated from a printout as at (2), is in fact given by

$$F = \left(n\hat{\boldsymbol{\theta}}_2^T \frac{C_{n;22.1}}{\hat{\sigma}^2} \hat{\boldsymbol{\theta}}_2 \right)/(p-q), \tag{12}$$

where

$$\hat{\sigma}^2 C_{n;22.1}^{-1} = \hat{\sigma}^2 (C_n^{-1})_{22}$$

is the estimated covariance matrix of $\sqrt{n}\,\hat{\boldsymbol{\theta}}_2$. Assuming that (7) holds, the limiting distribution of F, under alternatives

$$H_a^{(n)}: \boldsymbol{\theta}_2 = \boldsymbol{\Delta}/\sqrt{n},$$

is that of a $\chi^2_{p-q}(\delta^2)/(p-q)$ random variable. The non-centrality parameter is given by

$$\delta^2 = \boldsymbol{\Delta}^T C_{22.1} \boldsymbol{\Delta}/\sigma^2,$$

with $C_{22.1} = ((C^{-1})_{22})^{-1}$. Thus, objective (b) is met.

We note that it is not necessary that an exact zero be attained, at (3), by $(\boldsymbol{\theta}_*, \sigma_*)$. It suffices if

$$n^{-1/2} V_n^T \boldsymbol{\eta}(\boldsymbol{x}, r) \xrightarrow{pr} \mathbf{0}.$$

3. Influence function

Recall (12), and put

$$T^2 = \hat{\theta}_2^T C_{n;22.1}\theta_2/(p-q).$$

The influence function of T is defined by

$$IF(z_0; T, H_\theta) = \lim_{t\to 0} \frac{T\{(1-t)H_\theta + t\Delta_{z_0}\} - T(H_\theta)}{t}.$$

Here, Δ_{z_0} is the distribution function which places all mass at $(x_0^T, y_0)^T$, and H_θ is the true distribution function, under the model.

The influence function represents the limiting influence of an observation at z_0 on the test statistic, normalized by the amount of mass at z_0. For robustness against outlying y-values, and for bounded influence, we require that the gross error sensitivity, defined by

$$GES(T, H_\theta) = \sup_{z_0} |IF(z_0; T, H_\theta)|$$

be finite.

For the calculations, we follow the procedures in chapter 6 of Hampel et al. (1986). We take $\sigma^2 = 1$, and assume that the null hypothesis $H_0: \theta_2 = 0$ is true. Partition M^{-1} as

$$M^{-1} = \begin{pmatrix} M_{(1)} \\ M_{(2)} \end{pmatrix} \begin{matrix} q\times p \\ (p-q)\times p \end{matrix}.$$

Define

$$\tilde{\theta} = \begin{pmatrix} \theta_1 \\ 0 \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \quad A_{p\times p} = M_{(2)}^T C_{22.1} M_{(2)}.$$

Under the regularity conditions of Maronna and Yohai (1981), we then find that

$$IF(z_0; T, H_{\tilde{\theta}}) = \left|\eta\left(x_0, y_0 - x_0^T\tilde{\theta}\right)\right| \{x_0^T A x_0/(p-q)\}^{1/2}. \tag{13}$$

For $\eta(\cdot)$ as at (4), (13) becomes

$$IF(z_0; T, H_{\tilde{\theta}}) = \|x\|\nu(x_0)\cdot\left|\psi\left(\frac{y_0 - x_0^T\tilde{\theta}}{\nu(x_0)}\right)\right|\cdot\left\{\frac{x_0^T A x_0}{(p-q)x_0^T x_0}\right\}^{1/2}. \tag{14}$$

Since A does not depend upon z_0, (14) then gives

$$GES(T, H_{\tilde{\theta}}) \le \sup_{x_0}\|x_0\|\nu(x_0)\cdot\sup_r|\psi(r)|\cdot\left\{\frac{ch_{\max}(A)}{p-q}\right\}^{1/2}, \tag{15}$$

where $ch_{\max}$ denotes the largest characteristic root. We thus require that $\psi(r)$ and $\|x_0\|\nu(x_0)$ be bounded.

For the numerical example in Section 4 below we take ψ to be Huber's ψ_c, defined by

$$\psi_c(r) = \begin{cases} r, & |r| \le c \\ c \cdot \operatorname{sign}(r), & |r| > c \end{cases}. \tag{16}$$

Two attractive choices of $v(x_0)$ are those discussed in Markatou and Hettmansperger (1990). Let $h_0 = x_0^T(X^TX)^{-1}x_0$, the leverage value of x_0. Then for both

$$v(x_0) = (1 - h_0)^{1/2} \quad \text{and} \tag{17}$$

$$v(x_0) = (1 - h_0)/\sqrt{h_0}, \tag{18}$$

it is easy to show that

$$\sup_{x_0} \| x_0 \| v(x_0) = \{ch_{\max}(X_1^TX_1)\}^{1/2},$$

where X_1 is the design matrix X, without the row x_0^T. For (17), $\| x_0 \| v(x_0)$ is maximized at $\| x_0 \| = \infty$, while for (18) the maximum is attained at $\| x_0 \| = 0$.

4. An example

We have computed estimates and standard errors for the 'stackloss' data set, described, and analyzed exhaustively and ingeniously, in Daniel and Wood (1980). There 21 points in the data set, obtained over 21 successive days of operation of a plant oxidizing ammonia to nitric acid. The model is

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \varepsilon,$$

where

y = 10 × (% ingoing ammonia lost as unabsorbed nitric oxides),
x_1 = air flow to the plant,
x_2 = temperature of cooling water in the absorption tower,
x_3 = concentration of nitric acid in the absorbing liquid.

Daniel and Wood determined that four of the data points represented 'transitional states', and should be removed. Subsequently, variable x_3 was dropped, and x_1^2 added.

We ran six regressions on these data. The methods used were:

A. Least squares, on all 21 points;
B. Least squares, after removing the 4 aforementioned points;
C. Huber-type M-estimation, with $\psi(r)$ given by (16) and $c = 1.5$;
D. Huber-type M-estimation, with $\psi(r)$ given by (16) and $c = 2\sqrt{p/n} = .873$;
E. Bounded influence estimation, using (16) with $c = 2\sqrt{p/n}$ and $v(x) = (1 - h_0)^{1/2}$, as at (17);
F. Bounded influence estimation, using (16) with $c = 2\sqrt{p/n}$ and $v(x) = (1 - h_0)/\sqrt{h_0}$, as at (18).

Table 1
Estimates, standard errors and p-values for the Example of Section 4

Method	Parameter estimates (standard errors in parentheses)					p-value for $H_0: \theta_2 = \theta_3 = 0$
	θ_0	θ_1	θ_2	θ_3	σ	
A	−39.92	0.7156	1.2953	0.1521	3.243	0.0073
	(11.90)	(0.1349)	(0.3680)	(0.1563)		
B	−37.65	0.7977	0.5773	−0.0671	1.253	0.0112
	(4.732)	(0.0674)	(0.1660)	(0.0616)		
C	−41.07	0.7962	1.0562	−0.1355	2.942	0.0147
	(10.79)	(0.1223)	(0.3338)	(0.1418)		
D	−39.33	0.8288	0.7590	−0.1087	2.303	0.0237
	(8.447)	(0.0958)	(0.2613)	(0.1110)		
E	−38.82	0.8326	0.7174	−0.1075	2.118	0.0074
	(3.883)	(0.1106)	(0.2258)	(0.0614)		
F	−41.749	0.7995	1.0639	−0.1310	3.194	0.0236
	(5.426)	(0.1442)	(0.3945)	(0.0734)		

The choice $c = 2\sqrt{p/n}$ is recommended by Belsley, Kuh and Welsch (1980). In each of methods $C - F$, all 21 cases are used and $\hat{\sigma}$ is estimated as at (6). Convergence, defined as

$$|\theta_{(k+1),j} - \theta_{(k),j}| < 0.01, \qquad j = 0, 1, 2, 3,$$

was obtained, in each case, in at most 7 iterations of the iteratively reweighted least squares algorithm of Section 1. This was followed by a regression on pseudovalues, as in Street, Carroll and Ruppert (1988) for methods C and D, and as in Section 2 for methods E and F.

The results are presented in Table 1. Of the four robust methods, E appears to have been the most successful at automatically down-weighting the four erroneous points, and then efficiently estimating $\boldsymbol{\theta}$. The estimates of the covariance matrix of $\hat{\boldsymbol{\theta}}$, and the resulting p-values for the trial hypothesis $H_0: \theta_2 = \theta_3 = 0$, appear to be quite sensitive to the choice of method. The complete output is given in the technical report of Wiens (1990).

Acknowledgement

The author wishes to thank the Associate Editor and a referee for their helpful suggestions.

Appendix I: Verification of equation (12)

Partition X, Γ, U as

$$X = \underset{q}{(X_1} \;\vdots\; \underset{(p-q)}{X_2)} = \underset{q}{(\Gamma_1} \;\vdots\; \underset{(p-q)}{\Gamma_2)} \begin{pmatrix} U_{11} & U_{12} \\ 0 & u_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix} = \Gamma U.$$

Since $s^2 = \hat{\sigma}^2$, the statistic F at (2) has

$$(p-q)\hat{\sigma}^2 F = \text{"SSE in reduced model"} - \text{"SSE in full model"}$$
$$= \|(I - H_{X_1})\boldsymbol{y}_*\|^2 - \|(I - H_X)\boldsymbol{y}_*\|^2,$$

where

$$H_{X_1} = X_1(X_1^T X_1)^{-1} X_1^T = \Gamma_1 \Gamma_1^T,$$
$$H_X = X(X^T X)^{-1} X^T = \Gamma \Gamma^T.$$

Thus

$$(p-q)\hat{\sigma}^2 F = \boldsymbol{y}_* \Gamma_2 \Gamma_2^T \boldsymbol{y}_* = \|\Gamma_2^T(V_n\boldsymbol{\theta}_* + k_n\boldsymbol{\eta}(\boldsymbol{x}, r))\|^2$$
$$= \|\Gamma_2^T V_n \hat{\boldsymbol{\theta}}\|^2 \ (by\ (7)) = \|(0 \vdots I_{p-q}) A_n \hat{\boldsymbol{\theta}}\|^2 \ (by\ (10))$$
$$= \|A_{n:22}\boldsymbol{\theta}_2\|^2, \tag{A.1}$$

where

$$A_n = \begin{pmatrix} A_{n:11} & A_{n:12} \\ 0 & A_{n:22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}.$$

Now

$$\left((nC_n)^{-1}\right)_{22} = \left(A_n^{-1} A_n^{-T}\right)_{22} = \left(A_{n:22}^T A_{n:22}\right)^{-1},$$

so that

$$A_{n:22}^T A_{n:22} = (nC_n)_{22.1}. \tag{A.2}$$

Now (12) follows from (A.1) and (A.2).

References

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

Daniel, C. and F.S. Wood, *Fitting equations to data* (Wiley, New York, 1980).

Hampel, F.R., R. Ronchetti, P.J. Rousseeuw and W. Stahel, *Robust statistics: The approach based on influence functions* (Wiley, New York, 1986)

Hill, R.W. and P.W. Holland, Two robust alternatives to robust regression, *Journal of the American Statistical Association*, **72** (1977) 828–833.

Huber, P.J., Minimax Aspects of Bounded-Influence Regression (with discussion), *Journal of the American Statistical Association*, **78** (1983) 66–80.

Krasker, W.S. and R.E. Welsch, Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, **77** (1982) 595–604.

Marazzi, A., Testing in linear models and model selection in Robeth, Robeth-85 document no. 4 (Institut Universitaire de Médicine Sociale et Préventive, Lausanne, (1987).

Markatou, M. and T.P. Hettmansperger, Robust bounded influence tests in linear models, *Journal of the American Statistical Association*, **85** (1990) 187–190.

Maronna, R.A. and V.J. Yohai, Asymptotic behaviour of general M-estimates for regression and scale with random carriers, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **58** (1981) 7–20.

Street, J.O., R.J. Carroll and D. Ruppert, A note on computing robust regression estimates via iteratively reweighted least squares, *The American Statistician*, **42** (1988) 152–154.

Wiens, D.P., A note on the computation of robust, bounded influence estimates and test statistics in regression, Technical report 90.02 (University of Alberta, Dept. of Statistics and Applied Probability, Edmonton, 1990).