The Annals of Statistics
2001, Vol. 29, No. 3, 748–762

WEAK CONVERGENCE OF THE EMPIRICAL PROCESS OF RESIDUALS IN LINEAR MODELS WITH MANY PARAMETERS

BY GEMAI CHEN AND RICHARD A. LOCKHART

University of Manitoba and Simon Fraser University

When fitting, by least squares, a linear model (with an intercept term) with p parameters to n data points, the asymptotic behavior of the residual empirical process is shown to be the same as in the single sample problem provided $p^3 \log^2(p)/n \to 0$ for any error density having finite variance and a bounded first derivative. No further conditions are imposed on the sequence of design matrices. The result is extended to more general estimates with the property that the average error and average squared error in the fitted values are on the same order as for least squares.

1. Introduction. Let Y be a random vector satisfying the linear model

$$Y = X\beta + \sigma\varepsilon,$$

where $X = (x_{ij})$ is a known $n \times q$ matrix of constants, $\beta = (\beta_1, \dots, \beta_q)'$ are unknown regression parameters, σ is an unknown positive constant, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ are iid random variables from a distribution F with mean 0 and variance 1; superscript $'$ denotes transpose. We use x_i to denote the ith row of X. Let p denote the rank of X and put $\mu = X\beta$.

For estimates $\hat{\beta}$ and $\hat{\sigma}$ we define fitted residuals by $\tilde{\varepsilon} = Y - X\hat{\beta}$. The standardized fitted residuals are $\hat{\varepsilon} = \tilde{\varepsilon}/\hat{\sigma}$. The empirical process of fitted residuals is, for $t \in [0, 1]$,

$$\widetilde{Z}(t) = \frac{1}{\sqrt{n}} \sum [1\{F(\tilde{\varepsilon}_i/\sigma) \le t\} - t],$$

while that of the standardized fitted residuals is

$$\widehat{Z}(t) = \frac{1}{\sqrt{n}} \sum [1\{F(\hat{\varepsilon}_i) \le t\} - t].$$

Throughout this paper the quantities X, p, q, β and σ among others depend on n; wherever possible the dependence of quantities on n is suppressed. All limits are taken as $n \to \infty$. Probability calculations are made for true parameter values σ_o and β_o; except to state assumptions we assume for notational simplicity $\sigma_o = 1$, $\beta_o = 0$ and $\mu_o = X\beta_o = 0$.

We analyze $\widetilde{Z}$ using the process

$$Z(t, \beta) = n^{-1/2} \sum [1\{F(Y_i - x_i\beta) \le t\} - t],$$

Received November 1998; revised February 2001.
AMS 2000 *subject classifications*. Primary 62E20; secondary 62J99.
Key words and phrases. Residual, regression, empirical processes, goodness-of-fit.

which we decompose as

$$Z(t,\beta) = Z_1(t) + n^{-1/2}\sum x_i\beta J(t) + R(t,\beta),$$

where

$$Z_1(t) = n^{-1/2}\sum[1\{F(\varepsilon_i)\le t\} - t]$$

and $J(t) = f\{F^{-1}(t)\}$ with f the density of F. For fixed p, under assumptions on F similar to those used below and under some conditions on X, Koul (1969) showed essentially that

$$\sup\{|R(t,\beta)|;\ 0\le t\le 1, \|\beta\|\le C\}\to 0$$

in probability for each fixed $C<\infty$; see also Koul (1984). An important condition imposed by Koul and most later workers is that X has full rank $p=q$ and that

$$\max\{H_{ii};\ 1\le i\le n\}\to 0, \tag{1.1}$$

where $\mathbf{H} = X(X'X)^{-1}X'$ is the usual hat matrix.

Mukantseva (1977) showed that for normal errors, least squares estimates, designs with an intercept, and p fixed we have

$$\sup\{|\widetilde{Z}(t) - Z_1(t) - n^{-1/2}\sum x_i\hat\beta J(t)|; 0\le t\le 1\}\to 0$$

in probability and that

$$Z_1(t) + \Big(n^{-1/2}\sum x_i\hat\beta\Big)J(t)$$

converges weakly in $D[0,1]$ to a mean 0 Gaussian process with covariance function

$$\rho(s,t) = \min(s,t) - st - J(s)J(t). \tag{1.2}$$

The same weak limit arises when the $N(\theta,1)$ model is fitted to an iid sample.

Loynes (1980) fits more general regression models with p fixed. Portnoy (1986) has results for the case where p may grow with n in such a way that $\limsup p^2/n<\infty$. Since this condition is weaker than our condition **N** we discuss Portnoy's results in more detail.

Portnoy shows that when $p^2/n\to 0$, the remainder process $R(t,\hat\beta)$ converges pointwise in t to 0 in probability for many M-estimates, $\hat\beta$, under a variety of conditions on the design matrix, moment conditions, and conditions on the derivatives of f. His conditions, which include (1.1), are strong enough to prove

$$\max\{|x_i\hat\beta|; 1\le i\le n\}\to 0. \tag{1.3}$$

Portnoy gives a weak convergence result which uses (1.3) to prove tightness. However his argument has a gap. The problem arises in Portnoy's (3.18) where he bounds

$$n^{-1/2}\sum\big[1(\varepsilon_i - x_i\hat\beta\le x+\delta) - 1(\varepsilon_i - x_i\hat\beta\le x)\big].$$

It appears that the bound in Portnoy's (3.18) is supposed to be small; in fact it is $O_P(\delta n^{1/2})$. The quantities in question are not increments of the empirical process, as seems to be intended, because they are not centered by their expectations. In Section 3 below we give a sequence of estimates $\hat{\beta}$ and designs [with $p^3 = O(n)$] satisfying (1.3) but for which $R(1/2, \hat{\beta})$ does not go to 0. Though we do not present the somewhat lengthy details, a small generalization of the example can be constructed for which (1.3) holds, $p^3 = O(n)$, $R(t, \hat{\beta}) \to 0$ in probability for each fixed t and yet the conclusion of Theorem 1 below fails. The example is an ANOVA design. The work of Portnoy (1984) makes clear that different rates of convergence may be expected in regression designs where the x_i behave like a random sample from a suitable distribution in R^p.

Mammen (1996) establishes expansions more general than ours which apply uniformly for t in compact subsets of $(0, 1)$ (not enough to get convergence in distribution for standard goodness-of-fit statistics such as Kolmogorov–Smirnov) and under various restrictions on the estimates $\hat{\beta}$. These restrictions permit Mammen to give expansions when $p^2/n = o(n^{1/5})$ which agree with ours provided $p^2/n = o(1)$. Our results extend those of Mammen by lowering moment conditions, dealing with all of $[0, 1]$, and eliminating (1.1) at the cost of requiring a slower rate of growth for p. Mammen imposes a reasonable restriction on the estimators considered which requires that the estimate be nearly independent of those observations within $O(n^{-1/2})$ of a particular point x. Our counterexample in Section 3 shows that such a restriction is necessary. We have not investigated the question of trying to combine our approach with Mammen's to improve our restriction on the rate of growth of p.

In Section 2 we establish under conditions on F and on the average error and squared error in the fitted values that $R(t, \hat{\beta})$ converges to 0 uniformly in t in probability. For least squares estimates the conditions on the average error and squared error are automatic under the growth condition

$$(\mathbf{N}_{\mathrm{LS}}) \qquad \frac{p^3 \log^2(p)}{n} \to 0$$

with no further conditions on the design.

For more general estimators we achieve the same result, that $R(t, \hat{\beta})$ converges to 0 uniformly in t in probability, under a growth condition (see **N** below) which trades possibly slower growth of p (than in condition $\mathbf{N}_{\mathrm{LS}}$) for larger average squared error in the fitted values. In particular, it should be noted that we do not require (1.3); we do not need the condition, (1.1), that the largest leverage tends to 0 and we do not require the estimates $\hat{\beta}$ to be consistent except in the average sense of **E1** and **E2** below. The result is deduced from Lemma 2.1 which asserts that the remainder process $R(t, \beta)$ converges to 0 uniformly in $t \in [0, 1]$ and β in a set to which any reasonable estimate $\hat{\beta}$ is likely to belong. The convergence is almost sure.

Also in Section 2 we give conditions under which $\widetilde{Z}$ has a Gaussian weak limit. For least squares estimates and designs with an intercept term, no further condition on the sequence of designs is necessary. For designs without an intercept term but with normal errors no further condition on the sequence of designs is necessary. For designs without an intercept term and nonnormal errors, a negligibility condition is required to get Gaussian weak limits.

In Section 3 we show that our growth condition is nearly the right rate in the sense that there is a sequence of designs and estimators satisfying our average error and squared error conditions and having $p^3/n \to c < \infty$ but for which $R(1/2, \hat{\beta})$ has a positive limit.

Section 4 records briefly the standard deduction of the behavior of $\widehat{Z}$ from that of $\widetilde{Z}$ via a time transformation argument. We focus on the special case of least squares estimates. We also record a result to the effect that in the case of normal errors and least squares estimation, a weak convergence result for $\widetilde{Z}$ must imply a corresponding result for $\widehat{Z}$, provided $\limsup p/n < 1$.

Section 5 contains proofs.

2. General distributions.

We impose conditions on the error distribution, F, on the estimators $\hat{\beta}$ and on the rate of growth of p. Concerning F we assume the following:

(**F1**) The distribution F is strictly increasing on its support, an interval $[a, b]$ where $a = -\infty$ and $b = \infty$ are permitted. It has density f with derivative f' such that both f and $|f'|$ are bounded on $(-\infty, \infty)$ by some constant M_f.

Note that **F1** implies that the function $J(t) = f\{F^{-1}(t)\}$ has a version which is continuous on $[0, 1]$.

To get Gaussian weak limits for least squares estimates we will also need

(**F2**) The distribution F has mean 0 and variance 1.

Notice that $\widetilde{Z}$ depends on the estimate $\hat{\beta}$ only through the vector $\hat{\mu} = X\hat{\beta}$ of fitted values. We will assume that the average error and average squared error in these fitted values are not too large; letting $\mathbf{1}$ denote a column vector with all entries equal to 1 we need

(**E1**) $n^{-1/2}\sum x_i(\hat{\beta} - \beta_o) = n^{-1/2}\mathbf{1}'X(\hat{\beta} - \beta_o) = n^{-1/2}\mathbf{1}'(\hat{\mu} - \mu_o) = O_P(1)$

and

(**E2**) For a deterministic sequence $d_n \geq 1$,

$$(\hat{\mu} - \mu_o)'(\hat{\mu} - \mu_o) = (\hat{\beta} - \beta_o)'X'X(\hat{\beta} - \beta_o) = O_P(d_n).$$

For least squares estimates with a full rank design matrix X, **E1** is automatic since the quantity in question has mean 0 and variance $n^{-1}\mathbf{1}'\mathbf{H}\mathbf{1} \leq 1$ where $\mathbf{H}$ is the hat matrix $\mathbf{H} = X(X'X)^{-1}X'$. If X does not have full rank then the least squares estimate $\hat{\beta}$ is not uniquely defined. There is, however, a unique $\hat{\mu}$ in the column space of X minimizing $(Y - \mu)'(Y - \mu)$. For any $\hat{\beta}$ for which $\hat{\mu} = X\hat{\beta}$ we see that **E1** holds since $\mathrm{Var}(n^{-1/2}\mathbf{1}'(\hat{\mu} - \mu_o)) = n^{-1}\mathbf{1}'\mathbf{H}\mathbf{1} \leq 1$ where now $\mathbf{H} = X(X'X)^{-}X'$ and $(X'X)^{-}$ is a generalized inverse of $X'X$; see Rao [(1973), page 26]. Condition **E2** holds for least squares estimates with $d_n = p$ since the expectation of the positive quantity involved is simply p.

Finally our condition on the number of parameters is

$$(\mathbf{N})\quad \frac{p^2 d_n \log^2(pd_n)}{n} \to 0 \quad \text{and} \quad \frac{d_n^2}{n} \to 0.$$

THEOREM 1. *Assume conditions* **F1**, **E1**, **E2** *and* **N**. *Then*

$$\sup\{|\widetilde{Z}(t) - Z_1(t) - n^{-1/2}\mathbf{1}'\hat{\mu}J(t)|;\ 0 \le t \le 1\} = o_P(1).$$

Note that conditions **E1** and **E2** replace all moment conditions and that, in particular, **F2** is not needed. To prove the theorem we write

$$R(t, \beta) = R^{(1)}(t, \beta) + R^{(2)}(t, \beta),$$

where

$$R^{(1)}(t, \beta) = n^{-1/2} \sum \big[F\{F^{-1}(t) + x_i\beta\} - t - x_i\beta J(t)\big]$$

and

$$R^{(2)}(t, \beta) = n^{-1/2} \sum R_i(t, \beta)$$

with

$$R_i(t, \beta) = 1[F(\varepsilon_i) \le F\{F^{-1}(t) + x_i\beta\}] - 1\{F(\varepsilon_i) \le t\} - F\{F^{-1}(t) + x_i\beta\} + t.$$

The theorem is an easy consequence of the following lemma.

LEMMA 1. *Assume condition* **F1**. *Then for any sequence* $\eta_n \to 0$,

$$\sup\{|R^{(1)}(t, \beta)|;\ 0 \le t \le 1,\ \beta'(X'X)\beta \le \eta_n n^{1/2}\} \to 0$$

almost surely. *If*, *in addition*, **N** *holds then*

$$\sup\{|R^{(2)}(t, \beta)|;\ 0 \le t \le 1,\ \beta \in D_L\} \to 0 \tag{2.1}$$

and

$$\sup\{|Z(t, \beta) - Z_1(t) - n^{-1/2}\mathbf{1}'X\beta J(t)|;\ 0 \le t \le 1,\ \beta \in D_L\} \to 0 \tag{2.2}$$

almost surely for each L *where*

$$D_L = \{\beta\colon \beta'(X'X)\beta \le Ld_n, |n^{-1/2}\mathbf{1}'X\beta| \le L\}.$$

To get a weak convergence result for a given sequence of estimators we need to verify **E1** and **E2** and then check convergence of finite-dimensional distributions for $(Z_1, n^{-1/2}\mathbf{1}'\hat{\mu})$. In the next subsection we do this for least squares estimates identifying all possible limit laws and characterizing those situations where the weak limit is Gaussian. For more general estimators we note, as an example, that Portnoy (1985) gives conditions under which an M-estimate $\hat{\beta}$, defined as a root of $\sum x_i\psi(Y_i - x_i\beta)$ for suitable ψ, satisfies both **E1** (see Portnoy's Theorem 3.1) and **E2** with $d_n = p$ (see Portnoy's Theorem 3.2 and his discussion of his condition X1). Put $\mathbf{a} = n^{-1/2}\mathbf{H1}$ and $\nu = \mathrm{E}\{\psi'(\varepsilon)\}$.

Under Portnoy's conditions,

$$n^{-1/2}\mathbf{1}'\hat{\mu} = \sum a_i \psi(\varepsilon_i)/\nu + o_P(1) \tag{2.3}$$

and

$$\max\{|a_i|;\ 1 \le i \le n\} \to 0. \tag{2.4}$$

[An application of the Cauchy–Schwarz inequality shows that (1.1) implies (2.4); the converse is false so (2.4) is weaker.] Portnoy's assumptions give $\mathrm{E}\{\psi(\varepsilon)\} = 0$ and $\mathrm{Var}\{\psi(\varepsilon)\} = \tau^2 < \infty$; the Lindeberg central limit theorem then shows that $\widetilde{Z}$ is approximately Gaussian. To state the result let Δ be a metric for which the set of distributions on $D[0, 1]$ is a complete separable metric space. Let $\mathscr{L}(Z)$ denote the law of a process Z in $D[0, 1]$. Define $c = c_n = \mathbf{a}'\mathbf{a}$. Put $m(s) = \mathrm{E}[\psi(\varepsilon)1\{F(\varepsilon) \le s\}]$. For $0 \le \gamma \le 1$ let $\widetilde{W}_{\gamma,\nu,\tau}$ be a mean 0 Gaussian process with covariance

$$\tilde{\rho}_{\gamma,\nu,\tau}(s, t) = \min(s, t) - st + \gamma\{m(s)J(t)/\nu + m(t)J(s)/\nu + \tau^2 J(s)J(t)/\nu^2\}.$$

COROLLARY 1. *Assume* **F1** *and* **N**. *Assume that* $\hat{\beta}$ *is an* M*-estimate with* $\mathrm{E}\{\psi(\varepsilon)\} = 0$ *and* $\mathrm{Var}\{\psi(\varepsilon)\} = \tau^2 < \infty$. *Let* $\nu = \mathrm{E}\{\psi'(\varepsilon)\}$ *and assume* **E2** *and that* (2.3) *and* (2.4) *hold. Then* $\Delta\{\mathscr{L}(\widetilde{Z}), \mathscr{L}(\widetilde{W}_{c,\nu,\tau})\} \to 0$. *If* **1** *is in the column space of* X *then* $\widetilde{Z} \Rightarrow \widetilde{W}_{1,\nu,\tau}$ *in* $D[0, 1]$.

When ψ is the score function f'/f we have $m(s) = J(s)$ and, under standard regularity conditions, $\nu = -\tau^2$. In this case $\tilde{\rho}$ simplifies to $\min(s, t) - st - \gamma J(s)J(t)/\tau^2$.

The conditions in Portnoy (1985) are rather stronger than those labelled **F** and impose a number of conditions on the sequence of design matrices. In our context we need far less than the conclusions of Portnoy's Theorems 3.2 and 3.3; it seems likely to us that Portnoy's work can be followed to establish **E1**, **E2**, and representation (2.3) under weaker conditions. However, for least squares estimates we can proceed directly and obtain the conclusions of Corollary 1.

2.1. *Least squares.* For least squares estimates the conclusion is that of the previous result with $m(s) = \int 1\{F(y) \le s\} y f(y)\,dy$, $\tau = \sigma = 1$ and $\nu = 1$. Let $\widetilde{W}_{\gamma,\mathrm{LS}}$ be a mean 0 Gaussian process with covariance

$$\tilde{\rho}_{\gamma,\mathrm{LS}}(s, t) = \min(s, t) - st + \gamma\{m(s)J(t) + m(t)J(s) + J(s)J(t)\}.$$

COROLLARY 2. *Assume that* $\hat{\mu}$ *is the least squares estimate. Assume* **F1**, **F2**, *and* $\mathbf{N}_{\mathrm{LS}}$. *Then the family* $\mathscr{L}(\widetilde{Z})$ *is tight. If* F *is standard normal or* (2.4) *holds then* $\Delta\{\mathscr{L}(\widetilde{Z}), \mathscr{L}(\widetilde{W}_{c,\mathrm{LS}})\} \to 0$. *If* **1** *is in the column space of* X *then* $\widetilde{Z} \Rightarrow \widetilde{W}_{1,\mathrm{LS}}$.

When (2.4) fails we can describe the set of possible weak limit points of $\mathscr{L}(\widetilde{Z})$. Let $\widetilde{W}^*_{\gamma,\zeta}$ be a mean 0 Gaussian process with covariance $\tilde{\rho}^*(s,t) = \min(s,t) - st + (\gamma-\zeta)J(s)J(t) + \gamma\{m(s)J(t)+m(t)J(s)\}$ with $\gamma \in [0,1]$. Let V independent of W^*_γ be distributed as $\sum \alpha_i \varepsilon_i$ for a set of constants α_i with $\sum \alpha_i^2 = \zeta \le \gamma$. Then any weak limit of $\widetilde{Z}$ has the distribution of $\widetilde{W}^*_{\gamma,\zeta}(t) + VJ(t)$ for some (γ, ζ). Note that the assertion of the corollary for normal errors is a special case of this assertion since V would then be $N(0,\zeta)$.

On the other hand, suppose F is not normal and (2.4) is not satisfied. Let $|a_{(1)}| \ge |a_{(2)}| > \cdots$ denote the entries in a_i sorted in order of decreasing absolute value. By a diagonalization argument we can pick a subsequence along which $a_{(1)} \to \alpha_1$, $a_{(2)} \to \alpha_2$ and so on with $\alpha_1 \ne 0$. Along this subsequence $\mathscr{L}(\widetilde{Z})$ converges weakly to $W^*_\gamma(t) + VJ(t)$. If F is not normal then V is not Gaussian. To get a precise example, take $p=1$ and put $x_1 = n^{1/2}$ and $x_2 = \cdots = x_n = 1$. Then it is easily checked that $a_{(1)} = a_1 \to \alpha_1 = 1/2$ and $\max\{|a_i|;\ 2 \le i \le n\} = n^{1/2}/(2n-1) \to 0$. We find $c_n \to 1/2$ and $\widetilde{Z} \Rightarrow W^*_{1/2} + \varepsilon J/2$ which is not Gaussian unless ε is normal.

When F is normal we find $m = -J$; the covariance of $\widetilde{W}_{\gamma,\mathrm{LS}}$ simplifies to $\min(s,t) - st - \gamma J(s)J(t)$ matching (1.2) when the design matrix has an intercept.

3. Counterexamples. The condition **N** cannot be much weakened in either Theorem 1 or Lemma 1 without strengthening the other conditions. Here we present an example satisfying **F1** and **F2** with $p^3/n \to c \in (0,\infty)$ for which there is an estimator $\tilde{\beta}$ satisfying **E1** and **E2** with $d_n = p$ such that $R(1/2,\tilde{\beta})$ does not converge to 0 in probability. On the event $(\tilde{\beta}-\beta_o)'X'X \times (\tilde{\beta}-\beta_o) \le L$ (which has high probability for large L) we have

$$\sup\{|R^{(2)}(t,\beta)|;\ 0 \le t \le 1,\ \beta \in D_L\} \ge |R(1/2,\tilde{\beta})|$$

so that the last two conclusions [(2.1) and (2.2)] of Lemma 1 also fail for this example. It will be seen that the estimator $\tilde{\beta}$ is equivariant. We emphasize the point since it is conceptually possible that (2.1) could fail while at the same time,

$$\sup\{|R^{(2)}(t,\hat{\beta})|;\ 0 \le t \le 1\} \to 0$$

in probability for every sequence of equivariant estimators $\hat{\beta}$.

Our example has p samples of size $m = p^2$. The ith sample is $\{Y_{ij},\ j = 1,\dots,m\}$, with $Y_{ij} = \mu_i + \varepsilon_{ij}$ for iid $N(0,1)$ errors ε_{ij}. With $\beta' = (\mu_1,\dots,\mu_p)$ and standard notation for sample means, the remainder process, $R^{(2)}$, then has the form (remember that $n^{1/2} = p^{1/2}m^{1/2} = pm^{1/4}$)

$$R^{(2)}(t,\beta) = \frac{1}{p}\sum_{i=1}^{p} R_i(t,\mu_i),$$

where the R_i are the iid processes

$$R_i(t, \mu) = \frac{1}{m^{1/4}} \sum_{j=1}^{m} R_{ij}(t, \mu)$$

and

$$R_{ij}(t, \mu) = 1\{Y_{ij} - \mu \le \Phi^{-1}(t)\} - 1\{Y_{ij} \le \Phi^{-1}(t)\} - \Phi\{\Phi^{-1}(t) + \mu\} + t.$$

We define $\tilde{\mu}_i = \overline{Y}_i + \tilde{\delta}_i / m^{1/2}$ where $\tilde{\delta}_i$ is chosen to maximize

$$W_i(\delta) \equiv \frac{1}{m^{1/4}} \sum_{j=1}^{m} \big\{1(Y_{ij} - \overline{Y}_i \le \delta/m^{1/2}) - 1(Y_{ij} - \overline{Y}_i \le 0) - \Phi(\delta/m^{1/2}) + 1/2\big\}.$$

over $|\delta| \le 1$. We see the $\tilde{\mu}_i$ are iid and that each $\tilde{\delta}_i$ has a symmetric distribution. Since $Y_{ij} - \overline{Y}_i = \varepsilon_{ij} - \bar{\varepsilon}_i$ is location invariant, $\tilde{\delta}_i$ is location invariant and $\tilde{\mu}_i$ is location equivariant. Then

$$\begin{aligned} \mathbf{1}' X(\tilde{\beta} - \beta_o) &= m \sum (\tilde{\mu}_i - \mu_{i,o}) \\ &= m \sum (\overline{Y}_i - \mu_{i,o}) + m^{1/2} \sum \tilde{\delta}_i \\ &= O_P(n^{1/2}) \end{aligned}$$

so that $\tilde{\beta}$ satisfies **E1**. Moreover, $\tilde{\beta}$ satisfies **E2** since

$$\begin{aligned} (\tilde{\beta} - \beta_o)' X' X (\tilde{\beta} - \beta_o) &= m \sum_{1}^{p} (\tilde{\mu}_i - \mu_{i,o})^2 \\ &\le 2\big\{m \sum (\overline{Y}_i - \mu_{i,o})^2 + \sum \tilde{\delta}_i^2\big\} \\ &= O_P(p). \end{aligned}$$

Since the estimator $\tilde{\mu}_i$ is equivariant we may take $\mu_{i,o} = 0$ for all i to make probability calculations. Then

$$\begin{aligned} R_i(1/2, \tilde{\mu}_i) = {} & R_i(1/2, \bar{\varepsilon}_i) + W_i(\tilde{\delta}_i) - W_i(0) \\ & + m^{3/4}\big\{\Phi(\tilde{\delta}_i/m^{1/2}) + \Phi(\bar{\varepsilon}_i) - \Phi(\bar{\varepsilon}_i + \tilde{\delta}_i/m^{1/2}) - \Phi(0)\big\}. \end{aligned}$$

Routine moment calculations may be used to check that

$$\frac{1}{p} \sum R_i(1/2, \bar{\varepsilon}_i) \to 0$$

in probability. A two-term Taylor expansion shows that

$$\frac{m^{3/4}}{p} \sum_i \big\{\Phi(\tilde{\delta}_i/m^{1/2}) + \Phi(\bar{\varepsilon}_i) - \Phi(\bar{\varepsilon}_i + \tilde{\delta}_i/m^{1/2}) - \Phi(0)\big\}$$

is at least

$$-2p^{1/2} \|\Phi''\|_\infty \Big\{\sum \bar{\varepsilon}_i^2 + \sum \tilde{\delta}_i^2/m\Big\},$$

which evidently converges to 0 in probability.

Consider the process $\{W_1(\delta);\ |\delta| \le 1\}$. Define, for real t,

$$V_m(t) \equiv \frac{1}{m^{1/4}} \sum_{j=1}^{m} \Big\{1(\varepsilon_{1j} \le t/m^{1/2}) - 1(\varepsilon_{1j} \le 0) - \Phi(t/m^{1/2}) + 1/2\Big\}.$$

Then $(V_m, m^{1/2}\bar{\varepsilon}_1)$ converges weakly as $m \to \infty$. The weak limit has independent components. The weak limit of V_m is a Gaussian process B on the real line where $\{B(t);\ t \ge 0\}$ and $\{B(-t);\ t \ge 0\}$ are independent Brownian motions. If $\widetilde{V}_m(\cdot) = V_m(\cdot + m^{1/2}\bar{\varepsilon}_1) - V_m(m^{1/2}\bar{\varepsilon}_1)$ then the asymptotic independence of V_m and $m^{1/2}\bar{\varepsilon}_1$ shows that $\widetilde{V}_m$ also converges weakly to B. Finally $\sup\{|\widetilde{V}_m(t) - W_1(t)|;\ |t| \le T\}$ converges to 0 in probability for each fixed T so that W_1 also converges weakly to B. It follows that $W_1(\tilde{\delta}_1)$ converges in distribution to $B(\tilde{\delta})$ where $\tilde{\delta}$ maximizes B over $[-1, 1]$. Note that $B(\tilde{\delta}) > 0$ almost surely and that $(W_1(\tilde{\delta}_1), \dots, W_p(\tilde{\delta}_p))$ are iid. It follows that

$$\liminf \frac{1}{p} \sum_i W_i(\tilde{\delta}_i) \ge \mathrm{E}\{B(\tilde{\delta})\} > 0$$

in probability and, since $W_i(0) = 0$,

$$\lim \mathrm{P}\big[R(1/2, \tilde{\beta}) > \mathrm{E}\{B(\tilde{\delta})\}/2\big] = 1.$$

4. Unknown *σ*. Write $\widehat{Z}(t) = \widetilde{Z}\{T(t)\} + n^{1/2}\{T(t) - t\}$ where $T(t) = F\{\hat{\sigma} F^{-1}(t)\}$. Expanding the second term we have $n^{1/2}\{T(t) - t\} = n^{1/2}(\hat{\sigma}^2 - 1)F^{-1}(t) f\{\eta(t)\}/(1 + \hat{\sigma})$ where $\eta(t)$ lies between $F^{-1}(t)$ and $\hat{\sigma} F^{-1}(t)$. For consistent $\hat{\sigma}$, and assuming **F1** and **F2**, it is elementary to check that the process $F^{-1}(t) f\{\eta(t)\}$ converges uniformly to $J_2(t) \equiv F^{-1}(t)J(t)$ and that J_2 is in $C[0, 1]$. A weak convergence result for $\widehat{Z}$ then follows (via standard time transform arguments) from our results for $\widetilde{Z}$ provided we verify that the finite-dimensional distributions of $(Z_1, n^{-1/2}\mathbf{1}' X \hat{\beta}, n^{1/2}(\hat{\sigma}^2 - 1))$ converge.

We do not have general results along these lines; Portnoy's work deals with known σ only. For least squares we have two results. First, for general distributions we add only the following condition.

F3 F has a finite fourth moment, $\mu_4 = \mathrm{E}(\varepsilon^4)$.

If we then estimate σ^2 by the usual least squares estimate,

$$\hat{\sigma}^2_{\mathrm{LS}} = \|Y - \hat{\mu}_{\mathrm{LS}}\|^2/(n - p),$$

we have an obvious extension of our results for $\widetilde{Z}$. Define $m_2(s) = \int 1\{F(y) \le s\}(y^2 - 1) f(y)\, dy$. Let $\mu_3 = \mathrm{E}(\varepsilon^3)$. Let $\widehat{W}_c$ be a continuous Gaussian process with mean 0 and covariance function

$$\begin{aligned}\hat{\rho}_c(s, t) = \min(s, t) - st &\\ &+ c\big[J(s)J(t) + m(s)J(t) + m(t)J(s) + \mu_3\{J(s)J_2(t) + J(t)J_2(s)\}\big] \\ &+ (\mu_4 - 1)J_2(s)J_2(t)/4 + \{m_2(s)J_2(t) + m_2(t)J_2(s)\}/2\,.\end{aligned}$$

THEOREM 2. *Assume* **F1**, **F2**, **F3** *and* $\mathbf{N}_{\text{LS}}$. *Assume also* (2.4). *Then for least squares estimates* $\Delta\{\mathscr{L}(\widehat{Z}_n), \mathscr{L}(\widehat{W}_c)\} \to 0$. *If the column space of* X *contains* **1** *then* (2.4) *is automatic and* $\widehat{Z} \Rightarrow \widehat{W}_1$.

In the special case of normal errors we have $m = -J$, $m_2 = -J_2$, $\mu_3 = 0$ and $\mu_4 = 3$ so that the covariance simplifies to

$$\min(s,t) - st - cJ(s)J(t) - J_2(s)J_2(t)/2.$$

Our other result for least squares estimates applies only in the important special case of normal errors. Consider any estimate $\hat{\beta}$ with the property that the standardized residual vector $\hat{\varepsilon}$ has a distribution free of β and σ. Then $\hat{\sigma}_{\text{LS}}$ is independent of $\hat{\varepsilon}$. We use this to show that virtually any weak convergence result for σ known can be extended to one for unknown σ under the weak condition $\limsup p/n < 1$. We begin by motivating the consideration of a slightly more general empirical process.

When using least squares estimates, the fitted residuals $\tilde{\varepsilon}_i = \varepsilon_i - x_i\hat{\beta}$ have variance covariance matrix $\mathbf{I}-\mathbf{H}$. Thus the exact probability integral transform of $\tilde{\varepsilon}_i$ is $\Phi(w_i\tilde{\varepsilon}_i)$ with $w_i^{-2} = 1 - \mathbf{H}_{ii}$. Portnoy (1986) and Mammen (1996) both show that when $\limsup p^2/n > 0$, the process $\widetilde{Z}$ has, for general M-estimates, a nonnegligible asymptotic mean. Though this correction term vanishes for least squares and normal errors, there are examples for $p/n^{1/2} \to c \in (0,\infty)$ with least squares estimates and normal errors for which the process $\widetilde{Z}$ does indeed have a nonnegligible asymptotic mean.

This problem can sometimes be corrected by consideration of the process

$$n^{-1/2}\sum[1\{\Phi(w_i\tilde{\varepsilon}_i) \le t\} - t],$$

using the exact probability integral transform; Meester and Lockhart (1988) for example, give weak convergence results for this process for highly structured designs with $p/n \to c \in (0,1)$.

This prompts us to consider a slightly more general empirical process,

$$Q_n(t) = n^{-1/2}\sum[1\{F(r_i\tilde{\varepsilon}_i) \le t\} - t]$$

for an arbitrary set of constants r_i and the corresponding σ unknown process

$$Y_n(t) = n^{-1/2}\sum[1\{F(r_i\tilde{\varepsilon}_i/\hat{\sigma}_{\text{LS}}) \le t\} - t].$$

[Meester (1984) has studied the use of the exact distribution of $w_i\tilde{\varepsilon}_i/\hat{\sigma}_{\text{LS}}$ for the probability integral transform.] Note that Y_n is independent of $\hat{\sigma}_{\text{LS}}$.

THEOREM 3. *Consider any sequence of models and constants* r_i *for which* Q_n *converges in* $D[0,1]$ *to some continuous Gaussian process* Q *with mean function* μ *and covariance function* $\rho(s,t)$. *If the errors are normal and* $\limsup p/n < 1$ *then* $\Delta\{\mathscr{L}(Y_n), \mathscr{L}(Y_n^*)\} \to 0$ *where* Y_n^* *is a continuous Gaussian process with mean function* μ *and covariance function* $\rho_n^*(s,t) = \rho(s,t) - (n-p)J_2(s)J_2(t)/(2n)$.

5. Proofs.

PROOF OF LEMMA 1. The assertion for $R^{(1)}$ is an elementary Taylor expansion. We will prove by a chaining argument that $R^{-} = \sup\{R^{(2)}(t, \beta);\ 0 \le t \le 1,\ \beta \in D_L\} \to 0$ almost surely. (The parallel argument for the infimum is omitted.)

For a given $\delta > 0$ let $M(\delta)$ be the smallest integer m for which there exist $\beta_1^*, \dots, \beta_m^*$ such that $\beta'(X'X)\beta \le Ld_n$ implies that there is a k for which

$$(\beta - \beta_k^*)'(X'X)(\beta - \beta_k^*) \le \delta. \tag{5.1}$$

(M is called a covering number.) Let C_k denote the set of β in D_L satisfying (5.1). If β_k^* is not in C_k let β_k be any point in C_k minimizing $(\beta - \beta_k^*)'(X'X)(\beta - \beta_k^*)$; otherwise let $\beta_k = \beta_k^*$. Let N be a large integer. Put

$$R_{j,k} = \sup\left\{R(t, \beta); \frac{j}{N} \le t \le \frac{j+1}{N},\ \beta \in C_k\right\}.$$

Note that

$$P(R^{-} > \eta) < \sum_{j,k} P(R_{j,k} > \eta),$$

where the sum extends over $j = 0, \dots, N-1$, and all indices k for which C_k is not empty. The number of terms in this sum is no more than $NM(\delta)$.

Put $R_i(t, \beta) = 1[F(\varepsilon_i) \le F\{F^{-1}(t) + x_i\beta\}] - 1\{F(\varepsilon_i) \le t\} - F\{F^{-1}(t) + x_i\beta\} + t$. Fix j, k. Put $t_j = j/N$ and $c_j = F^{-1}(t_j)$. For any $\beta \in C_k$, and $t_j \le t \le t_{j+1}$ we have $R_i(t, \beta)$ is less than or equal to

$$1\{F(\varepsilon_i) \le F(c_{j+1} + a_{i,k})\} - 1\{F(\varepsilon_i) \le t_j\} - F(c_j + b_{i,k}) + t_{j+1}, \tag{5.2}$$

where $a_{i,k}$ is the supremum of $x_i\beta$ over C_k and $b_{i,k}$ the corresponding infimum. This eliminates t and β from the bound. As in Loynes (1980) write (5.2) as

$$\omega_{i,j,k}(X_{i,j,k} - p_{i,j,k}) + B_{i,j,k},$$

where $X_{i,j,k} = |1\{F(\varepsilon_i) \le F(c_{j+1} + a_{i,k})\} - 1\{F(\varepsilon_i) \le t_j\}|$ is a Bernoulli variable, $p_{i,j,k} = \mathrm{E}(X_{i,j,k})$, $\omega_{i,j,k}$ is the sign of $F\{F^{-1}(t_{j+1}) + a_{i,k}\} - j/N$ and

$$B_{i,j,k} = F(c_{j+1} + a_{i,k}) - F(c_j + b_{i,k}) + 1/N.$$

Consider now

$$\begin{aligned}\sum_i B_{i,j,k} = {} & \sum_i \{F(c_{j+1} + a_{i,k}) - F(c_{j+1} + x_i\beta_k)\} \\ & + \sum_i \{F(c_{j+1} + x_i\beta_k) - F(c_j + x_i\beta_k)\} \\ & + \sum_i \{F(c_j + x_i\beta_k) - F(c_j + b_{i,k})\} + n/N.\end{aligned} \tag{5.3}$$

Since f is bounded the first term on the right-hand side is bounded by

$$M_f \sum_i |a_{i,k} - x_i\beta_k|.$$

Write $a_{i,k}$ as $x_i\beta_{i,k}$ for some $\beta_{i,k} \in C_k$; use the Cauchy–Schwarz inequality to get

$$|a_{i,k} - x_i\beta_k|^2 \le (\beta_{i,k} - \beta_k)'(X'X)(\beta_{i,k} - \beta_k)x_i(X'X)^- x_i' \le \delta x_i(X'X)^- x_i'.$$

Use the inequality $(\sum |u_i|)^2 \le n\sum u_i^2$ for any real $u_1, \dots, u_n$ and recall that $\sum x_i(X'X)^- x_i' = p$ to see that the first and third terms in (5.3) are each bounded by $M_f\sqrt{np\delta}$. For the second term in (5.3), two-term Taylor expansions give

$$\begin{aligned}\sum_i &\big[F\{F^{-1}(t_{j+1}) + x_i\beta_k\} - F\{F^{-1}(t_j) + x_i\beta_k\}\big]\\ &= \sum_i \big[F\big\{F^{-1}(t_{j+1})\big\} + x_i\beta_k J(t_{j+1}) + \tfrac{1}{2}(x_i\beta_k)^2 f'(\alpha_i)\big]\\ &\quad - \sum_i \big[F\big\{F^{-1}(t_j)\big\} + x_i\beta_k J(t_j) + \tfrac{1}{2}(x_i\beta_k)^2 f'(\alpha_i^*)\big]\end{aligned}$$

for suitable α_i and α_i^*. The terms other than the remainders sum to

$$n/N + \mathbf{1}'X\beta_k\{J(t_{j+1}) - J(t_j)\}.$$

Let $\omega_J(r) = \sup\{|J(x) - J(y)|; |x - y| \le r\}$ be the modulus of continuity of J. Then the second term in (5.3) is no more than

$$n/N + \sqrt{n}L\omega_J(1/N) + M_f L d_n,$$

where we have used the fact that $\sum(x_i\beta_k)^2 = \beta_k'(X'X)\beta_k$. Now choose $N = \sqrt{n}/\gamma_n$ where $\gamma_n = 1/(pd_n)$ converges to 0. For any sequence of parameters p with $p/\sqrt{n} \to 0$ we then have

$$\begin{aligned}n^{-1/2}\sum_i R_i(t, \beta) &\le n^{-1/2}\sum \omega_{i,j,k}(X_{i,j,k} - p_{i,j,k})\\ &\qquad + 2\gamma_n + L\omega_J(1/N) + M_f\frac{Ld_n}{\sqrt{n}} + 2M_f\sqrt{p\delta}.\end{aligned}$$

We will choose $\delta = \gamma_n/p = 1/(p^2 d_n)$. For each fixed $\rho > 0$ there is then an n_0 (not depending on j or k) such that

$$\text{(5.4)} \qquad P(R_{j,k} > \rho) \le P\Big(n^{-1/2}\sum_i \omega_{i,j,k}(X_{i,j,k} - p_{i,j,k}) > \rho/2\Big)$$

for all $n \ge n_0$. Put $\tau_{j,k} = \sum_i p_{i,j,k}(1 - p_{i,j,k})$. According to Bernstein's inequality [see Pollard (1984), page 193], the right-hand side of (5.4) is no more than

$$2\exp\{-n\rho^2/\{8(\tau_{j,k} + n^{1/2}\rho/6)\}].$$

The arguments surrounding (5.3) can be followed to show

$$\begin{aligned}\tau_{j,k} \le \sum_i p_{i,j,k} &= \sum_i |F(c_{j+1} + a_{i,k}) - j/N| \\ &\le M_f\big(\sqrt{np\delta} + \sqrt{nLd_n}\big) + \sqrt{n}\gamma_n \\ &\le 2M_f\sqrt{nLd_n}\end{aligned}$$

for all sufficiently large n. For such n,

$$P(R^- > \rho) < 2NM(\gamma_n/p)\exp[-n\rho^2/\{8(2\sqrt{nLd_n} + \sqrt{n}\rho/6)\}].$$

We use a bound on the covering number $M(\delta)$, which we learned from David Pollard.

LEMMA 2. *The covering number is bounded by* $M(\delta) \le (1 + 2\sqrt{Ld_n/\delta})^p$.

For L^* slightly larger than $2L^{1/2}$ we have $M(\gamma_n/p) \le (L^* dp)^p$ for all large n. Combining these we obtain

$$\begin{aligned}\log\{P(R^- > \rho)\} \le{}& \log(2n^{1/2}) - \log(\gamma_n) + p\log(L^*) - p\log(pd_n) \\ & - \frac{\rho^2}{16}\frac{\sqrt{n}}{\sqrt{Ld_n} + \rho/12}.\end{aligned}$$

Factor out $(n/d_n)^{1/2}$ to see the bound goes to $-\infty$ under condition **N**. The argument shows $\sum_n P(R^- > \rho) < \infty$ so Borel–Cantelli gives almost sure convergence.

PROOF OF LEMMA 2. Consider first X of full rank. Define $M^*(\delta)$ to be the largest integer m for which there exist $\beta_1^*, \dots, \beta_m^*$ with each $\beta^* \in D_L$ and such that $(\beta_i^* - \beta_j^*)'(X'X)(\beta_i^* - \beta_j^*) > \delta$ for all $i \ne j$. Evidently $M(\delta) \le M^*(\delta)$. Define

$$B_k(r) = \{\beta; (\beta - \beta_k^*)'(X'X)(\beta - \beta_k^*) < r^2\}.$$

The triangle inequality [for the metric d defined by $d^2(x, y) = (x-y)'(X'X) \times (x-y)$] shows that the sets $B_1(\sqrt{\delta}/2), \dots, B_{M^*}(\sqrt{\delta}/2)$ are disjoint. Let $V(r)$ denote the volume of $\{\beta; \beta'(X'X)\beta \le r^2\}$. All the $B_k(\sqrt{\delta}/2)$ lie within the ellipsoid $\{\beta; \beta'(X'X)\beta \le (\sqrt{Lp} + \sqrt{\delta}/2)^2\}$ and so

$$M^*(\delta)V(\sqrt{\delta}/2) \le V(\sqrt{Lp} + \sqrt{\delta}/2).$$

Since $V(r) = r^p V(1)$ the lemma follows. For X of less than full rank we take $V(r)$ to be the p-dimensional volume in the orthogonal complement of the kernel of X of $\{\beta \in \ker(X)^\perp; \beta'(X'X)\beta \le r^2\}$; otherwise the argument remains the same. □

PROOF OF COROLLARIES 1 AND 2. Condition (2.4) is used in both corollaries to apply the Lindeberg theorem to $n^{-1/2}\mathbf{1}'\hat{\mu}$. It remains only to show convergence of finite-dimensional distributions in Corollary 2 without using (2.4) for normal F.

Let $\delta_n > 0$ be some sequence tending to 0 but with $\sqrt{n}\delta_n^2 \to \infty$. Define $I = \{i: |a_i| > \delta_n\}$. As in Theorem 2.1 the observation $\max\{|a_i|;\ i \notin I\} \le \delta_n \to 0$ shows that the process

$$n^{-1/2} \sum_{i \notin I} [1\{\Phi(\varepsilon_i) \le t\} - t + n^{1/2} a_i \varepsilon_i J(t)]$$

has asymptotically Gaussian finite-dimensional distributions.

Now $\mathrm{card}(I) \le \sum_{i\in I} a_i^2/\delta_n^2 \le \sum_i a_i^2/\delta_n^2 \le \mathbf{1}'\mathbf{H}\mathbf{1}/(n\delta_n^2) \le 1/\delta_n^2$. Hence the process $n^{-1/2}\sum_{i\in I}[1\{\Phi(\varepsilon_i) \le t\} - t]$ converges uniformly in t to 0. Finally the process $\sum_{i\in I} a_i \varepsilon_i J(t)$ has exactly normal finite-dimensional distributions. Thus $Z_1 + n^{-1/2}\mathbf{1}'\hat{\mu}J(t)$ is the sum of two independent Gaussian processes and a negligible remainder. The corollary follows. □

PROOF OF THEOREM 2. Apply a compactness argument, our previous results and the fact that $n^{1/2}(\hat{\sigma}^2 - 1) = n^{-1/2}\sum(\varepsilon_i^2 - 1) + o_P(1)$. □

PROOF OF THEOREM 3. To prove the theorem pick any subsequence along which $p/n \to \lambda$ for some $\lambda < 1$. We need only prove that Y_n then converges to a Gaussian process with mean μ and covariance $\rho^*(s,t) = \rho(s,t) - (1-\lambda)J_2(s)J_2(t)/2$.

The condition $n - p \to \infty$ guarantees [because $(n-p)\hat{\sigma}^2$ has a chi-squared distribution on $n - p$ degrees of freedom] that

$$\sqrt{n-p}(\hat{\sigma}^2 - 1) \overset{\mathscr{D}}{\Rightarrow} N(0, 2). \tag{5.5}$$

Write $Y_n(t) = Q_n\{T_n(t)\} + n^{1/2}[\Phi\{\hat{\sigma}\Phi^{-1}(t)\} - t] = Q_n^*(t) + V_n(t)$, say, where $T_n(t) = \Phi\{\hat{\sigma}\Phi^{-1}(t)\}$. Since $V_n(t) = n^{1/2}(\hat{\sigma} - 1)\Phi^{-1}(t)\phi\{\theta(t)\Phi^{-1}(t)\} = n^{1/2}(\hat{\sigma}^2 - 1)\Phi^{-1}(t)\phi\{\theta(t)\Phi^{-1}(t)\}/(1 + \hat{\sigma})$ with $\theta(t)$ between 1 and $\hat{\sigma}$ it is easy to check that, uniformly in t in probability,

$$V_n(t) - n^{1/2}(\hat{\sigma}^2 - 1)\Phi^{-1}(t)\phi\{\Phi^{-1}(t)\}/2 \to 0.$$

In view of (5.5) the process V_n converges weakly to a mean 0 Gaussian process V with variance covariance function $(1-\lambda)J_2(s)J_2(t)/2$. Moreover, T_n converges uniformly to the identity map on $[0, 1]$ in probability so that Q_n^* converges weakly to Q; see Billingsley [(1968), pages 144 and 145]. This shows that the sequence Y_n is tight (notice that this conclusion requires that Q and J_2 be in $C[0, 1]$). Thus the pair Y_n, V_n is tight. Since Y_n is independent of V_n any weak limit (Y, V) for the pair has Y independent of V. Thus $Q \overset{\mathscr{D}}{=} Y + V$ and the finite-dimensional distributions of Y are determined by the fact that Y and V are independent and their sum has a Gaussian law. Thus the mean function of Y is that of Q minus that of V, or μ, and the covariance function of Y is that of Q minus that of V. □

Acknowledgments. The authors are grateful to David Pollard for pointing out the bound on the covering numbers in the Lemma 2.2. We also thank two referees and an Associate Editor who suggested we extend our results to estimates other than least squares and whose comments resulted in a far better paper.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

KOUL, H. L. (1969). Asymptotic behaviour of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

KOUL, H. L. (1984). Tests of goodness-of-fit in linear regression. *Colloq. Math. Soc. János Bolyai* **45** 279–315.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

MAMMEN, E. (1996). Empirical process of residuals for high dimension linear models. *Ann. Statist.* **24** 307–335.

MEESTER, S. G. (1984). Testing for normally distributed errors in block design experiments. M.Sc. thesis, Dept. Mathematics and Statistics, Simon Fraser Univ.

MEESTER, S. G. and LOCKHART, R. A. (1988). Testing for normal errors in designs with many blocks. *Biometrika* **75** 569–575.

MUKANTSEVA, L. A. (1977). Testing normality in one-dimensional and multi-dimensional linear regression. *Theory Probab. Appl.* **22** 591–602.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large I. Consistency. *Ann. Statist.* **12** 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

PORTNOY, S. (1986). Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.* **14** 1152–1170.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MANITOBA
WINNIPEG, MANITOBA
R3T 2N2 CANADA
E-MAIL: gchen@cc.umanitoba.ca

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
SIMON FRASER UNIVERSITY
BURNABY, BC
V5A 1S6 CANADA
E-MAIL: lockhart@sfu.ca