The Annals of Statistics
1974, Vol. 2, No. 5, 880–891

AN UNBALANCED JACKKNIFE[1]

BY RUPERT G. MILLER, JR.
Stanford University

It is proved that the jackknife estimate $\tilde{\theta} = n\hat{\theta} - (n-1)(\sum \hat{\theta}_{-i}/n)$ of a function $\theta = f(\boldsymbol{\beta})$ of the regression parameters in a general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is asymptotically normally distributed under conditions that do not require $\mathbf{e}$ to be normally distributed. The jackknife is applied by deleting in succession each row of the $\mathbf{X}$ matrix and $\mathbf{Y}$ vector in order to compute $\hat{\boldsymbol{\beta}}_{-i}$, which is the least squares estimate with the ith row deleted, and $\hat{\theta}_{-i} = f(\hat{\boldsymbol{\beta}}_{-i})$. The standard error of the pseudo-values $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$ is a consistent estimate of the asymptotic standard deviation of $\tilde{\theta}$ under similar conditions. Generalizations and applications are discussed.

1. Introduction. Over the past decade considerable research has been devoted to studying the properties of the jackknife technique which was introduced by Quenouille as a method for bias reduction and which was later proposed as a method for robust interval estimation by Tukey. Conditions under which the jackknife estimator is asymptotically normally distributed with a consistent estimate of its variance have been established. Its performance in the problems of ratio estimation and comparison of variances has been studied at length. The jackknife has since been generalized to handle more general forms of bias, and it has been extended to handle estimation in specialized stochastic processes. For a complete reference on these papers the reader is referred to Gray, Watkins, and Adams (1972).

All of these past papers, with one exception, have assumed the balanced case, i.e., equal sample sizes, equal variances, etc. This is not a criticism; it was necessary in order to achieve the first mathematical grasp of the jackknife's properties. The one exception is a recent technical report by Arvesen and Layard (1971) on unbalanced variance component models. They prove a theorem on jackknifing U-statistics where the expectation of the kernel equals the unknown parameter multiplied by a known constant which varies with the subset of variables. This extends some earlier work of Arvesen (1969) on non-identically distributed random variables. The theorem neatly handles the problem of unbalanced variance components, but it is not clear at the moment which other problems will fit into this framework.

The aim of this paper is to thrust the jackknife into the everyday realm of

Received April 1973; revised September 1973.

[1] This paper was written while the author was a John S. Guggenheim Memorial Fellow visiting at Imperial College of Science and Technology and London School of Hygiene and Tropical Medicine.

AMS 1970 *subject classifications*. 62G05, 15, 35; 62E20.

Key words and phrases. Jackknife, pseudo-value, general linear model, multiple regression, asymptotic normality.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

unbalance, and see how it fares in a fairly general setting. The setting chosen is that of multiple linear regression $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where the error variables are independently, identically distributed but not necessarily with a normal distribution. The parameter to be estimated is $\theta = f(\boldsymbol{\beta})$ where $f(\cdot)$ is a smooth function of the regression parameters $\boldsymbol{\beta}$. The jackknife is applied by deleting successive rows of the $\mathbf{X}$ matrix (and $\mathbf{Y}$ vector), and calculating the pseudo-values for $f(\boldsymbol{\beta})$ in the standard balanced way.

Theorem 1 in Section 4 proves that the jackknife estimator is asymptotically normally distributed provided $f(\cdot)$ has bounded second derivatives in a neighborhood of $\boldsymbol{\beta}$, the random variables e_i have finite fourth moment, and $\mathbf{X}^T\mathbf{X}/n$ converges to a limiting nonsingular matrix. Theorem 2 establishes that the sample variance of the pseudo-values is a consistent estimate for the variance (multiplied by n) of the jackknife estimator under the same conditions $\mathbf{X}^T\mathbf{X}/n \to \boldsymbol{\Sigma}$ and $E(e_i^4) < \infty$ but the weaker condition that $f(\cdot)$ have continuous first derivatives near $\boldsymbol{\beta}$. The structures of the proofs are identical to those used in studying the asymptotic behavior of the jackknifed $f(\bar{Y})$ in Miller (1964) and later extended to functions of U-statistics by Arvesen (1969). The details, however, are considerably messier due to the unbalance, and it is hoped the reader can persevere.

Section 2 lists some matrix lemmas which are needed in the proofs of Theorems 1 and 2. Section 3 contains lemmas which are components of the proofs of the theorems. By separating out some of the components it is hoped the reader will be able to see the forest above the trees. Generalizations of the theorems, bias reduction, and applications are discussed in Section 5.

Throughout this paper the following basic model will be assumed:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\,, \tag{1.1}$$

where $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$, $\mathbf{e} = (e_1, \cdots, e_n)^T$, and

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}. \tag{1.2}$$

For simplicity it will be assumed that $\mathbf{X}$ has rank p. In the theorems the sample size n will tend to infinity, and the reader should be aware that the sizes of the $\mathbf{X}$ matrix and $\mathbf{Y}$ vector are increasing with n. At times a subscript n will be appended to them for clarity, but since it is cumbersome, it will not be retained throughout.

The error variables e_i, $i = 1, \cdots, n$, will be assumed to be independently, identically distributed with zero mean. Their variance will be denoted by σ^2 and their fourth moment by μ_4.

The function $f(\cdot)$ will be assumed to be a real-valued function of p variables defined over a region R of p-dimensional space which contains an open sphere $S(\boldsymbol{\beta}, r)$ with center $\boldsymbol{\beta}$ and radius $r > 0$.

The author gratefully acknowledges some valuable conversations with Bradley Efron about this paper.

2. Matrix lemmas. Lemma 2.1 is a standard result in matrix algebra; its proof by direct multiplication is omitted.

LEMMA 2.1. *For a nonsingular matrix* $\mathbf{A}$ *and vectors* $\mathbf{U}$, $\mathbf{V}$,

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{U})(\mathbf{V}^T\mathbf{A}^{-1})}{1 + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}} . \tag{2.1}$$

LEMMA 2.2. *For fixed* $\mathbf{x}$ *the sequence* $\mathbf{x}^T(\mathbf{X}_n{}^T\mathbf{X}_n)^{-1}\mathbf{x}$ *is non-increasing as* n *increases.*

PROOF. From (1.2)

$$\mathbf{X}_{n+1}^T\mathbf{X}_{n+1} = \textstyle\sum_1^{n+1} \mathbf{x}_i{}^T\mathbf{x}_i = \mathbf{X}_n{}^T\mathbf{X}_n + \mathbf{x}_{n+1}^T\mathbf{x}_{n+1} . \tag{2.2}$$

By (2.1) and (2.2)

$$\begin{aligned}\mathbf{x}^T(\mathbf{X}_{n+1}^T\mathbf{X}_{n+1})^{-1}\mathbf{x} &= \mathbf{x}^T(\mathbf{X}_n{}^T\mathbf{X}_n{}^T)^{-1}\mathbf{x} \\ &\quad - \frac{(\mathbf{x}^T(\mathbf{X}_n{}^T\mathbf{X}_n)^{-1}\mathbf{x}_{n+1}^T)(\mathbf{x}_{n+1}(\mathbf{X}_n{}^T\mathbf{X}_n)^{-1}\mathbf{x})}{1 + \mathbf{x}_{n+1}(\mathbf{X}_n{}^T\mathbf{X}_n)^{-1}\mathbf{x}_{n+1}^T} .\end{aligned} \tag{2.3}$$

Since the numerator and denominator of the last term are nonnegative and greater than or equal to one, respectively, the result follows. □

LEMMA 2.3. *If* $\mathbf{X}_n{}^T\mathbf{X}_n/n \to \mathbf{\Sigma}$ *positive-definite, then*

$$\max_{1\leqq i\leqq n} \mathbf{x}_i(\mathbf{X}_n{}^T\mathbf{X}_n)^{-1}\mathbf{x}_i{}^T \to 0 .$$

PROOF. Suppose there is a subsequence m for which $\max \to \varepsilon > 0$. Let

$$\mathbf{x}_{i_m}(\mathbf{X}_m{}^T\mathbf{X}_m)^{-1}\mathbf{x}_{i_m}^T = \max_{1\leqq i\leqq m} \mathbf{x}_i(\mathbf{X}_m{}^T\mathbf{X}_m)^{-1}\mathbf{x}_i{}^T . \tag{2.4}$$

Since $m(\mathbf{X}_m{}^T\mathbf{X}_m)^{-1} \to \mathbf{\Sigma}^{-1}$, the sequence i_m must tend to infinity. By Lemma 2.2

$$\mathbf{x}_{i_m}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}\mathbf{x}_{i_m}^T \geqq \mathbf{x}_{i_m}(\mathbf{X}_m{}^T\mathbf{X}_m)^{-1}\mathbf{x}_{i_m} . \tag{2.5}$$

But

$$\begin{aligned} p &= \operatorname{tr}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}(\mathbf{X}_{i_m-1}^T\mathbf{X}_{i_m-1} + \mathbf{x}_{i_m}^T\mathbf{x}_{i_m}) \\ &= \operatorname{tr}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}(\mathbf{X}_{i_m-1}^T\mathbf{X}_{i_m-1}) + \mathbf{x}_{i_m}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}\mathbf{x}_{i_m}^T .\end{aligned} \tag{2.6}$$

Since $i_m \to \infty$ and $\mathbf{X}_n{}^T\mathbf{X}_n/n \to \mathbf{\Sigma}$ as $n \to \infty$,

$$\operatorname{tr}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}(\mathbf{X}_{i_m-1}^T\mathbf{X}_{i_m-1}) \to p . \tag{2.7}$$

By the identity (2.6) this implies $\mathbf{x}_{i_m}(\mathbf{X}_{i_m}^T\mathbf{X}_{i_m})^{-1}\mathbf{x}_{i_m}^T \to 0$, which is a contradiction. □

Lemma 2.4 is another standard result in matrix algebra; its proof from the definition of eigenvalue is omitted.

LEMMA 2.4. *For a* $p \times p$ *symmetric matrix* $A = (a_{ij})$ *with* $|a_{ij}| \leqq \alpha$ *the eigenvalues of* A *are all less than or equal to* $p\alpha$ *in absolute value.*

3. Component lemmas.

LEMMA 3.1. *If $\sigma^2 < \infty$ and $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$ positive definite, then*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \to_d N(\boldsymbol{\beta}, \sigma^2\mathbf{\Sigma}^{-1}/n)\,.$$

PROOF. Application of the Lindeberg condition (cf. Feller (1966), page 256) to linear combinations $z = \mathbf{l}^T\mathbf{X}^T\mathbf{Y}$ establishes their asymptotic normality so invocation of the Cramér–Wold theorem (cf. Feller (1966), page 495) proves the result. □

The above proof was just barely outlined because a somewhat more general result with similar proof can be found in Gleser (1965, Corollary 3.2).

The jackknife will employ the estimates $\hat{\boldsymbol{\beta}}_{-i}$, $i = 1, \cdots, n$, obtained by successively deleting the ith row $\mathbf{x}_i$ from the $\mathbf{X}$ matrix and recomputing the least squares estimate of $\boldsymbol{\beta}$. To establish the asymptotic results for the jackknife estimator it is necessary to have an expression for $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}$, and this is given in the following lemma. Similar expressions exist in the literature on recursive filtering (cf. Duncan and Horn (1972)).

LEMMA 3.2. *For $i = 1, \cdots, n$,*

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i} = \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T(Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})}{1 - \mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T}\,. \tag{3.1}$$

PROOF. By definition

$$\hat{\boldsymbol{\beta}}_{-i} = (\mathbf{X}^T\mathbf{X} - \mathbf{x}_i^T\mathbf{x}_i)^{-1}(\mathbf{X}^T\mathbf{Y} - \mathbf{x}_i^TY_i)\,. \tag{3.2}$$

Direct application of Lemma 2.1 gives (3.1). □

For notational simplicity the ith residual $Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$ will be denoted by r_i, and $\mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T$ will be denoted by Δ_i. The reader should remember that r_i and Δ_i change as n varies, but this dependence on n will be suppressed for notational sanity. With these conventions (3.1) can be rewritten as

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^Tr_i/(1 - \Delta_i)\,.$$

LEMMA 3.3. *If $\sigma^2 < \infty$ and $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$ positive definite,*

$$P\{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})^T(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}) \leqq \varepsilon,\ i = 1, \cdots, n\} \to 1 \tag{3.3}$$

as $n \to \infty$ for any $\varepsilon > 0$.

PROOF.

$$\begin{aligned} &P\{\max\nolimits_{1\leqq i\leqq n} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})^T(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}) > \varepsilon\} \\ &\qquad \leqq \textstyle\sum_1^n P\{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})^T(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}) > \varepsilon\}\,, \\ &\qquad = \textstyle\sum_1^n P\{r_i^2 > \varepsilon(1 - \Delta_i)^2/\mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T\}\,, \\ &\qquad \leqq \textstyle\sum_1^n \dfrac{\mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T}{\varepsilon(1 - \Delta_i)^2} E(r_i^2)\,, \end{aligned} \tag{3.4}$$

where the last inequality follows from the simple inequality $P\{X > \delta\} \leqq E(X)/\delta$ for positive random variables. Since $E(r_i^2) = 1 - \Delta_i$, the sequence (3.4) of equalities/inequalities continues as

$$\begin{aligned}\sum_1^n \frac{\mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T}{\varepsilon(1-\Delta_i)} &\leqq \frac{K}{\varepsilon}\sum_1^n \mathbf{x}_i(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T \\ &= \frac{K}{\varepsilon}\sum_1^n \operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T\mathbf{x}_i \\ &= \frac{K}{\varepsilon}\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1},\end{aligned} \tag{3.5}$$

where the inequality in (3.5) holds for some $K < \infty$ and n sufficiently large by virtue of Lemma 2.3. Since $n(\mathbf{X}^T\mathbf{X})^{-1} \to \mathbf{\Sigma}^{-1}$, the last term in (3.5) tends to zero, which proves the lemma. □

In the proofs of Theorems 1 and 2 the convergence $\sum_1^n r_i^2\mathbf{x}_i^T\mathbf{x}_i/n \to_p \sigma^2\mathbf{\Sigma}$ will be needed. For ordinary least squares estimation the residual variance $\sum_1^n r_i^2/n$ is separate from the normalizing constants $\mathbf{X}^T\mathbf{X} = \sum_1^n \mathbf{x}_i^T\mathbf{x}_i$, but the jackknife estimator weaves them together.

LEMMA 3.4. *If $\mu_4 < \infty$ and $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$ positive definite, then $\sum_1^n r_i^2\mathbf{x}_i^T\mathbf{x}_i/n \to_p \sigma^2\mathbf{\Sigma}$ as $n \to \infty$.*

PROOF. From its definition $r_i = e_i - u_i$ where $e_i = Y_i - \mathbf{x}_i\boldsymbol{\beta}$ and $u_i = \mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Then,

$$\begin{aligned}\frac{1}{n}\sum_1^n r_i^2\mathbf{x}_i^T\mathbf{x}_i &= \frac{1}{n}\sum_1^n (e_i - u_i)^2\mathbf{x}_i^T\mathbf{x}_i, \\ &= \frac{1}{n}\sum_1^n e_i^2\mathbf{x}_i^T\mathbf{x}_i - \frac{2}{n}\sum_1^n e_iu_i\mathbf{x}_i^T\mathbf{x}_i + \frac{1}{n}\sum_1^n u_i^2\mathbf{x}_i^T\mathbf{x}_i.\end{aligned} \tag{3.6}$$

Consider the first sum of squares on the last line.

$$E\left(\frac{1}{n}\sum_1^n e_i^2\mathbf{x}_i^T\mathbf{x}_i\right) = \frac{\sigma^2}{n}\mathbf{X}^T\mathbf{X} \to \sigma^2\mathbf{\Sigma} \tag{3.7}$$

as $n \to \infty$. Since the e_i are independently, identically distributed with finite fourth moment μ_4, the matrix of variances satisfies

$$\operatorname{Var}\left(\frac{1}{n}\sum_1^n e_i^2\mathbf{x}_i^T\mathbf{x}_i\right) = \frac{1}{n^2}(\mu_4 - \sigma^4)\sum_1^n \mathbf{x}_i^T\mathbf{x}_i * \mathbf{x}_i^T\mathbf{x}_i, \tag{3.8}$$

where the Hadamard product $\mathbf{A} * \mathbf{B}$ of two matrices $\mathbf{A} = (a_{kl})$, $\mathbf{B} = (b_{kl})$ is defined to be the matrix of element by element products $(a_{kl}b_{kl})$. Expression (3.8) is bounded above by

$$(\mu_4 - \sigma^4)\frac{\max_{1\leq i\leq n}|\mathbf{x}_i^T\mathbf{x}_i|}{n} * \frac{1}{n}\sum_1^n |\mathbf{x}_i^T\mathbf{x}_i|, \tag{3.9}$$

where $|\mathbf{A}| = (|a_{kl}|)$ and $\max_{1\leq i\leq n}\mathbf{A}_i = (\max_{1\leq i\leq n} a_{ikl})$. Since $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$,

$\max_{1 \leqq i \leqq n} |\mathbf{x}_i{}^T\mathbf{x}_i|/n \to \mathbf{0}$, a matrix of zeroes. Furthermore,

$$\frac{1}{n} \textstyle\sum_1^n |\mathbf{x}_i{}^T\mathbf{x}_i| \leqq \mathbf{C}_n , \tag{3.10}$$

where $\mathbf{A} \leqq \mathbf{B}$ means $a_{kl} \leqq b_{kl}$ for all k, l, and $\mathbf{C}_n$ is a matrix whose (k, l) element is

$$c_{nkl} = \left(\frac{1}{n} \textstyle\sum_{i=1}^n x_{ik}^2\right)^{\frac{1}{2}} \left(\frac{1}{n} \textstyle\sum_{i=1}^n x_{il}^2\right)^{\frac{1}{2}} . \tag{3.11}$$

Since $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma} = (\sigma_{kl})$, $c_{nkl} \to \sigma_{kk}^{\frac{1}{2}}\sigma_{ll}^{\frac{1}{2}}$. Thus, the last factor in (3.9) remains bounded so the whole expression converges to $\mathbf{0}$. This, in turn, implies that (3.8) converges to a zero matrix so

$$\frac{1}{n} \textstyle\sum_1^n e_i{}^2\mathbf{x}_i{}^T\mathbf{x}_i \to_p \sigma^2\mathbf{\Sigma} . \tag{3.12}$$

Next consider the last sum of squares in (3.6).

$$\left|\frac{1}{n} \textstyle\sum_1^n u_i{}^2\mathbf{x}_i{}^T\mathbf{x}_i\right| \leqq [\max_{1 \leqq i \leqq n} u_i{}^2] \frac{1}{n} \textstyle\sum_1^n |\mathbf{x}_i{}^T\mathbf{x}_i| . \tag{3.13}$$

It has been established in the previous paragraph that $\sum_1^n |\mathbf{x}_i{}^T\mathbf{x}_i|/n$ remains bounded. Since $|\mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|^2 \leqq \mathbf{x}_i\mathbf{x}_i{}^T \cdot (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$,

$$\begin{aligned} P\{\max_{1 \leqq i \leqq n} u_i{}^2 > \varepsilon\} &\leqq P\{\max_{1 \leqq i \leqq n} \mathbf{x}_i\mathbf{x}_i{}^T \cdot (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) > \varepsilon\} , \\ &\leqq \frac{1}{\varepsilon} \max_{1 \leqq i \leqq n} \mathbf{x}_i\mathbf{x}_i{}^T E((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) , \\ &= \frac{1}{\varepsilon} \max_{1 \leqq i \leqq n} \mathbf{x}_i\mathbf{x}_i{}^T E(\operatorname{tr}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T) , \\ &= \frac{\sigma^2}{\varepsilon} \frac{\max_{1 \leqq i \leqq n} \mathbf{x}_i\mathbf{x}_i{}^T}{n} n \operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} , \end{aligned} \tag{3.14}$$

where the second inequality holds because $P\{X > \delta\} \leqq E(X)/\delta$ for positive random variables. Since $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$ positive definite, $\max_{1 \leqq i \leqq n} \mathbf{x}_i\mathbf{x}_i{}^T/n \to 0$ and $n \operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} \to \operatorname{tr} \mathbf{\Sigma}^{-1}$. Thus, (3.14) converges to zero so

$$\frac{1}{n} \textstyle\sum_1^n u_i{}^2\mathbf{x}_i{}^T\mathbf{x}_i \to_p \mathbf{0} . \tag{3.15}$$

Because of (3.12) and (3.15) the cross-product term $\sum_1^n e_i u_i \mathbf{x}_i{}^T\mathbf{x}_i/n$ also converges in probability to zero by the Cauchy–Schwarz inequality. □

4. Main theorems. The jackknife estimator $\tilde{\theta}$ is defined in the usual manner. Let $\theta = f(\boldsymbol{\beta})$ and $\hat{\theta} = f(\hat{\boldsymbol{\beta}})$. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_{-i}$ the least squares estimate when the ith row is deleted from the $\mathbf{X}$ matrix and $\mathbf{Y}$ vector. Then the pseudo-values are defined by

$$\tilde{\theta}_i = nf(\hat{\boldsymbol{\beta}}) - (n - 1)f(\hat{\boldsymbol{\beta}}_{-i}) , \tag{4.1}$$

and the jackknife estimator is the average of the pseudo-values, i.e.,

$$\tilde{\theta} = \frac{1}{n} \sum_1^n \tilde{\theta}_i \,. \tag{4.2}$$

THEOREM 1. *If* $\mu_4 < \infty$ *and* $\mathbf{X}^T\mathbf{X}/n \to \mathbf{\Sigma}$ *positive definite,* $n^{\frac{1}{2}}(\tilde{\theta} - \theta) \to_d N(0, \sigma^2 \mathbf{f}'(\boldsymbol{\beta})^T \mathbf{\Sigma}^{-1} \mathbf{f}'(\boldsymbol{\beta}))$ *as* $n \to \infty$ *for any function* $f(\cdot)$ *with bounded second derivatives in an open sphere about* $\boldsymbol{\beta}$, *where* $\mathbf{f}'(\boldsymbol{\beta}) = (\partial f(\boldsymbol{\beta})/\partial\beta_1, \cdots, \partial f(\boldsymbol{\beta})/\partial\beta_p)^T$.

PROOF. Let $f(\cdot)$ be defined and have bounded second derivatives in the sphere $S(\boldsymbol{\beta}, r)$ with radius r about $\boldsymbol{\beta}$. From Lemma 3.1 $P\{\hat{\boldsymbol{\beta}} \in S(\boldsymbol{\beta}, r/2)\} \to 1$, and by Lemma 3.3 $P\{\hat{\boldsymbol{\beta}}_{-i} \in S(\hat{\boldsymbol{\beta}}, r/2), i = 1, \cdots, n\} \to 1$. Thus, $P\{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{-1}, \cdots, \hat{\boldsymbol{\beta}}_{-n} \in S(\boldsymbol{\beta}, r)\} \to 1$ as $n \to \infty$.

For a double sequence of events E_n and C_n in which $P\{C_n\} \to 1$

$$\lim P\{E_n\} = \lim [P\{E_n C_n\} + P\{E_n C_n^c\}] = \lim P\{E_n C_n\} \tag{4.3}$$

so convenient imposition or removal of the condition $C_n = \{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{-1}, \cdots, \hat{\boldsymbol{\beta}}_{-n} \in S(\boldsymbol{\beta}, r)\}$ has no effect on any limiting probabilities.

For $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{-1}, \cdots, \hat{\boldsymbol{\beta}}_{-n} \in S(\boldsymbol{\beta}, r)$

$$f(\hat{\boldsymbol{\beta}}_{-i}) = f(\hat{\boldsymbol{\beta}}) + \mathbf{f}'(\hat{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) + \tfrac{1}{2}(\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}})^T \mathbf{f}''(\boldsymbol{\zeta}_i)(\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) \,, \tag{4.4}$$

where $\mathbf{f}''(\boldsymbol{\beta}) = (\partial^2 f(\boldsymbol{\beta})/\partial\beta_k\partial\beta_l)$ and $\boldsymbol{\zeta}_i$ is a point on the line segment between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{-i}$. From (4.1), (4.2), and (4.4)

$$\begin{aligned} \tilde{\theta} &= f(\hat{\boldsymbol{\beta}}) - [(n-1)/n]\mathbf{f}'(\hat{\boldsymbol{\beta}})^T \textstyle\sum_1^n (\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) \\ &\quad - [(n-1)/2n] \textstyle\sum_1^n (\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}})^T \mathbf{f}''(\boldsymbol{\zeta}_i)(\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) \,. \end{aligned} \tag{4.5}$$

By Lemma 3.1 and Slutsky's theorem $n^{\frac{1}{2}}(f(\hat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta})) \to_d N(0, \sigma^2\mathbf{f}'(\boldsymbol{\beta})^T\mathbf{\Sigma}^{-1}\mathbf{f}'(\boldsymbol{\beta}))$ as $n \to \infty$, so the theorem is proved if it can be shown that the second and third terms on the right in (4.5) when multiplied by $n^{\frac{1}{2}}$ tend to zero in probability.

By Lemma 3.2

$$\begin{aligned} n^{\frac{1}{2}} \textstyle\sum_1^n (\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) &= -n^{\frac{1}{2}}(\mathbf{X}^T\mathbf{X})^{-1} \textstyle\sum_1^n \dfrac{\mathbf{x}_i^T r_i}{1 - \Delta_i} \,, \\ &= -n^{\frac{1}{2}}(\mathbf{X}^T\mathbf{X})^{-1} \textstyle\sum_1^n \dfrac{\Delta_i \mathbf{x}_i^T r_i}{1 - \Delta_i} \,, \end{aligned} \tag{4.6}$$

where the second equality holds because $\sum_1^n \mathbf{x}_i^T r_i = \mathbf{0}$, a vector of zeroes. Since $n(\mathbf{X}^T\mathbf{X})^{-1} \to \mathbf{\Sigma}^{-1}$, it is sufficient to show that $\sum_1^n \Delta_i \mathbf{x}_i^T r_i / n^{\frac{1}{2}}(1 - \Delta_i) \to_p \mathbf{0}$. The expectation of each coordinate is zero, and the variance of the kth coordinate is

$$\begin{aligned} \operatorname{Var}\left(\frac{1}{n^{\frac{1}{2}}} \sum_1^n \frac{\Delta_i x_{ik} r_i}{1 - \Delta_i}\right) &= \frac{1}{n}\left(\frac{\Delta_1 x_{1k}}{1 - \Delta_1}, \cdots, \frac{\Delta_n x_{nk}}{1 - \Delta_n}\right)(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\ &\quad \times \left(\frac{\Delta_1 x_{1k}}{1 - \Delta_1}, \cdots, \frac{\Delta_n x_{nk}}{1 - \Delta_n}\right)^T, \\ &\leqq \frac{1}{n} \sum_1^n \left(\frac{\Delta_i}{1 - \Delta_i}\right)^2 x_{ik}^2 \,, \\ &\leqq \max_{1 \leqq i \leqq n} \frac{\Delta_i^2}{(1 - \Delta_i)^2} \cdot \frac{1}{n} \sum_1^n x_{ik}^2 \,, \end{aligned} \tag{4.7}$$

where the first inequality holds because $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is positive definite. By Lemma 2.3 $\max_{1 \le i \le n} \Delta_i^2/(1-\Delta_i)^2 \to 0$, and $\sum_1^n x_{ik}^2/n \to \sigma_{kk}$. Thus, the variance converges to zero, and, since $\mathbf{f}'(\hat{\boldsymbol\beta}) \to_p \mathbf{f}'(\boldsymbol\beta)$, the second term (multiplied by $n^{\frac{1}{2}}$) in (4.5) vanishes asymptotically.

By Lemma 2.4 and the assumption that the second derivatives of $f(\cdot)$ are bounded in $S(\boldsymbol\beta, r)$, there exists a constant $M < \infty$ such that for $\hat{\boldsymbol\beta}, \hat{\boldsymbol\beta}_{-1}, \cdots, \hat{\boldsymbol\beta}_{-n} \in S(\boldsymbol\beta, r)$

$$\begin{aligned} n^{\frac{1}{2}} \textstyle\sum_1^n (\hat{\boldsymbol\beta}_{-i} - \hat{\boldsymbol\beta})^T \mathbf{f}''(\boldsymbol\zeta_i)(\hat{\boldsymbol\beta}_{-i} - \hat{\boldsymbol\beta}) & \\ \le M n^{\frac{1}{2}} \textstyle\sum_1^n (\hat{\boldsymbol\beta}_{-i} - \hat{\boldsymbol\beta})^T (\hat{\boldsymbol\beta}_{-i} - \hat{\boldsymbol\beta})\,, & \\ = M n^{\frac{1}{2}} \textstyle\sum_1^n \dfrac{1}{(1-\Delta_i)^2}\, r_i \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T r_i\,, & \\ = M n^{\frac{1}{2}} \textstyle\sum_1^n \dfrac{r_i^2}{(1-\Delta_i)^2}\, \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T\,. & \end{aligned} \tag{4.8}$$

By Lemma 2.3 there exists a constant $M' < \infty$ such that for n sufficiently large the last expression in (4.8) is bounded above by $M' n^{\frac{1}{2}} \sum_1^n r_i^2 \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-2}\mathbf{x}_i^T$, which in turn is bounded above by

$$\frac{1}{n^{\frac{1}{2}}} M' n^2 \lambda_{np}^{-2} \frac{1}{n} \textstyle\sum_1^n r_i^2 \mathbf{x}_i \mathbf{x}_i^T\,, \tag{4.9}$$

where λ_{np}^{-2} is the maximum eigenvalue of $(\mathbf{X}^T\mathbf{X})^{-2}$. Since $n^2(\mathbf{X}^T\mathbf{X})^{-2} \to \boldsymbol\Sigma^{-2}$, $n^2\lambda_{np}^{-2} \to \lambda_p^{-2}$, the maximum eigenvalue of $\boldsymbol\Sigma^{-2}$. From Lemma 3.4 $\sum_1^n r_i^2 \mathbf{x}_i^T \mathbf{x}_i/n \to_p \sigma^2 \boldsymbol\Sigma$ so

$$\begin{aligned} \frac{1}{n} \textstyle\sum_1^n r_i^2 \mathbf{x}_i \mathbf{x}_i^T &= \frac{1}{n} \textstyle\sum_1^n r_i^2 \operatorname{tr}(\mathbf{x}_i \mathbf{x}_i^T)\,, \\ &= \operatorname{tr}\left(\frac{1}{n} \textstyle\sum_1^n r_i^2 \mathbf{x}_i^T \mathbf{x}_i\right), \\ &\to_p \sigma^2 \operatorname{tr} \boldsymbol\Sigma\,. \end{aligned} \tag{4.10}$$

Since $n^2\lambda_{np}^{-2} \to \lambda_p^{-2}$ and $\sum_1^n r_i^2 \mathbf{x}_i \mathbf{x}_i^T/n \to \sigma^2 \operatorname{tr} \boldsymbol\Sigma$, the extra $n^{\frac{1}{2}}$ in the denominator of (4.9) makes (4.9) converge to zero in probability. Thus, the third term (multiplied by $n^{\frac{1}{2}}$) in (4.5) converges to zero in probability, and the theorem is proved. □

Tukey proposed that the sample variance of the pseudo-values

$$\tilde{s}^2 = \frac{1}{n-1} \textstyle\sum_1^n (\tilde\theta_i - \tilde\theta)^2 \tag{4.11}$$

divided by n would estimate the variance of $\tilde\theta$.

THEOREM 2. *If $\mu_4 < \infty$ and $\mathbf{X}^T\mathbf{X}/n \to \boldsymbol\Sigma$ positive definite, $\tilde{s}^2 \to_p \sigma^2 \mathbf{f}'(\boldsymbol\beta)^T \boldsymbol\Sigma^{-1} \mathbf{f}'(\boldsymbol\beta)$ as $n \to \infty$ for any function $f(\cdot)$ with continuous first derivatives in an open sphere about $\boldsymbol\beta$.*

PROOF. By the argument in Theorem 1 $P\{\hat{\boldsymbol\beta}, \hat{\boldsymbol\beta}_{-1}, \cdots, \hat{\boldsymbol\beta}_{-n} \in S(\boldsymbol\beta, r)\} \to 1$ as

$n \to \infty$ where $S(\boldsymbol{\beta}, r)$ is a sphere of radius r about $\boldsymbol{\beta}$ in which $f(\cdot)$ has continuous first derivatives. Also, because of (4.3), convenient imposition or removal of the condition $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{-1}, \cdots, \hat{\boldsymbol{\beta}}_{-n} \in S(\boldsymbol{\beta}, r)$ has no effect on any limiting probabilities.

For $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{-1}, \cdots, \hat{\boldsymbol{\beta}}_{-n} \in S(\boldsymbol{\beta}, r)$

$$f(\hat{\boldsymbol{\beta}}_{-i}) = f(\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}})^T \mathbf{f}'(\boldsymbol{\zeta}_i)\,, \tag{4.12}$$

where $\boldsymbol{\zeta}_i$ is a point on the line segment between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{-i}$. The expansion (4.12) and Lemma 3.2 yield

$$\begin{aligned}\textstyle\sum_1^n (\tilde{\theta}_i - \tilde{\theta})^2 &= (n-1)^2 \textstyle\sum_1^n \Big((\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}})^T \mathbf{f}'(\boldsymbol{\zeta}_i) \\ &\qquad - \frac{1}{n} \textstyle\sum_1^n (\hat{\boldsymbol{\beta}}_{-j} - \hat{\boldsymbol{\beta}})^T \mathbf{f}'(\boldsymbol{\zeta}_j)\Big)^2\,, \\ &= (n-1)^2 \textstyle\sum_1^n \Big(\frac{r_i}{1-\Delta_i} \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\boldsymbol{\zeta}_i) \\ &\qquad - \frac{1}{n} \textstyle\sum_1^n \frac{r_j}{1-\Delta_j} \mathbf{x}_j (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\boldsymbol{\zeta}_j)\Big)^2\,.\end{aligned} \tag{4.13}$$

For $i = 1, \cdots, n$, let

$$\begin{aligned} U_i &= r_i \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1} (\mathbf{f}'(\boldsymbol{\zeta}_i) - \mathbf{f}'(\hat{\boldsymbol{\beta}}))\,, \\ V_i &= \frac{\Delta_i}{1-\Delta_i} r_i \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\boldsymbol{\zeta}_i)\,. \end{aligned} \tag{4.14}$$

Then, since $\sum_1^n \mathbf{x}_i^T r_i = \mathbf{0}$, the relation (4.13) can be rewritten as

$$\textstyle\sum_1^n (\tilde{\theta}_i - \tilde{\theta})^2 = (n-1)^2 \sum_1^n (r_i \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\hat{\boldsymbol{\beta}}) + U_i - \bar{U} + V_i - \bar{V})^2\,. \tag{4.15}$$

The sum of squares on the right in (4.15) can be expanded into three sums of squares and three sums of cross-products. Consider the first sum of squares:

$$\begin{aligned} &(n-1)^2 \textstyle\sum_1^n (r_i \mathbf{x}_i (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\hat{\boldsymbol{\beta}}))^2 \\ &\qquad = \mathbf{f}'(\hat{\boldsymbol{\beta}})^T (n-1)(\mathbf{X}^T\mathbf{X})^{-1} [\textstyle\sum_1^n r_i^2 \mathbf{x}_i^T \mathbf{x}_i](n-1)(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{f}'(\hat{\boldsymbol{\beta}})\,. \end{aligned} \tag{4.16}$$

Since $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$, $\mathbf{f}'(\hat{\boldsymbol{\beta}}) \to_p \mathbf{f}'(\boldsymbol{\beta})$. Also, $(n-1)(\mathbf{X}^T\mathbf{X})^{-1} \to \boldsymbol{\Sigma}^{-1}$. Thus, since $\sum_1^n r_i^2 \mathbf{x}_i^T \mathbf{x}_i / n \to_p \sigma^2 \boldsymbol{\Sigma}$ by Lemma 3.4, the sum of squares (4.16) divided by $n-1$ converges in probability to

$$\sigma^2 \mathbf{f}'(\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} \mathbf{f}'(\boldsymbol{\beta}) \tag{4.17}$$

as $n \to \infty$. Since from Lemma 3.3 $\mathbf{f}'(\boldsymbol{\zeta}_i) - \mathbf{f}'(\hat{\boldsymbol{\beta}}) \to_p \mathbf{0}$ uniformly in $i = 1, \cdots, n$, the convergence of (4.16) divided by $n-1$ to (4.17) implies that $\sum_1^n U_i^2/(n-1)$, and hence $\sum_1^n (U_i - \bar{U})^2/(n-1)$, converges in probability to zero. Similarly, since $\Delta_i \to 0$ uniformly in $i = 1, \cdots, n$ by Lemma 2.3, $\sum_1^n (V_i - \bar{V})^2/(n-1) \to_p 0$ as $n \to \infty$. Since these last two sums of squares (divided by $n-1$) converge to zero in probability, the three cross-product sums (divided by $n-1$) will also converge to zero in probability by the Cauchy–Schwarz inequality. Thus $\sum_1^n (\tilde{\theta}_i - \tilde{\theta})^2/(n-1)$ converges in probability to (4.17). □

5. Discussion. The unbalanced jackknife technique can be extended to the case where the random variables $Y_i - \mathbf{x}_i\boldsymbol{\beta}$ have the same distribution except for known scalar constants $c_i(\mathrm{Var}(Y_i) = \sigma^2 c_i)$. The model can be transformed to the new variables $Y_i^* = c_i^{-\frac{1}{2}} Y_i$ and $\mathbf{x}_i^* = c_i^{-\frac{1}{2}}\mathbf{x}_i$, and the theorems then apply to these new variables provided the c_i and $\mathbf{x}_i$ are sufficiently well-behaved together that the condition $\mathbf{X}^{*T}\mathbf{X}^*/n \to \boldsymbol{\Sigma}^*$ positive definite applies.

The theorems do not immediately extend to the more general case where the random vector $\mathbf{Y}$ has a general covariance matrix $\sigma^2\mathbf{C}$ where $\mathbf{C}$ is known. The transformation $\mathbf{Y}^* = \mathbf{C}^{-\frac{1}{2}}\mathbf{Y}$ will not in general lead to independent random variables Y_i^*, and independence (not just uncorrelatedness) is specifically used in Lemmas 3.1 and 3.4. One could achieve this independence by assuming the Y_i are normally distributed, or assuming specifically that the transformation produces independent Y_i^*. Unfortunately, neither assumption may be very palatable.

Although the proofs have not been worked out, the unbalanced jackknife should extend to the case of non-linear least squares, i.e., $\sum_1^n (Y_i - g(\mathbf{x}_i, \boldsymbol{\beta}))^2$. For sufficiently smooth, though non-linear $g(\cdot, \cdot)$, the estimate $\hat{\boldsymbol{\beta}}$ should be behaving like

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{G}'(\boldsymbol{\beta})^T\mathbf{G}'(\boldsymbol{\beta}))^{-1}\mathbf{G}'(\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{g}(\boldsymbol{\beta})) \tag{5.1}$$

in large samples where

$$\begin{aligned} \mathbf{g}(\boldsymbol{\beta}) &= (g(\mathbf{x}_1, \boldsymbol{\beta}), \cdots, g(\mathbf{x}_n, \boldsymbol{\beta}))^T, \\ \mathbf{G}'(\boldsymbol{\beta}) &= (\partial g(\mathbf{x}_i, \boldsymbol{\beta})/\partial \beta_j). \end{aligned} \tag{5.2}$$

Theorems like 1 and 2 should then cover the asymptotic behavior of a jackknifed $f(\hat{\boldsymbol{\beta}})$. Weighted non-linear least squares could be included as well.

The theorems in this paper should also be extendable to prove that jackknifing the usual estimates of σ^2 and ρ^2, the multiple correlation coefficient, will produce robust confidence intervals.

Quenouille originally proposed the jackknife to eliminate a bias term of order $1/n$. This it does exactly in balanced problems (cf. Gray *et al.* (1972)). However, in the model of this paper it is not at all clear what jackknifing does to the bias of $\hat{\theta} = f(\hat{\boldsymbol{\beta}})$. From the expansion

$$f(\hat{\boldsymbol{\beta}}) = f(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbf{f}'(\boldsymbol{\beta}) + \tfrac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbf{f}''(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \cdots, \tag{5.3}$$

one obtains

$$E(f(\hat{\boldsymbol{\beta}})) = f(\boldsymbol{\beta}) + \frac{\sigma^2}{2}\operatorname{tr}\mathbf{f}''(\boldsymbol{\beta})(\mathbf{X}_n^T\mathbf{X}_n)^{-1} + \cdots. \tag{5.4}$$

After jackknifing, the first order bias term becomes

$$\frac{\sigma^2}{2}\left\{\operatorname{tr}\mathbf{f}''(\boldsymbol{\beta})(\mathbf{X}_n^T\mathbf{X}_n)^{-1} - \frac{(n-1)}{n}\sum_1^n \frac{\mathbf{x}_i(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{f}''(\boldsymbol{\beta})(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{x}_i^T}{1 - \mathbf{x}_i(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{x}_i^T}\right\}. \tag{5.5}$$

What the second term does to the absolute size of the bias is still a mystery. Perhaps results on this will be forthcoming.

Many estimation problems can be fit into the framework of the model in this paper. A classical example is inverse linear calibration. A sample $(x_1, Y_1), \cdots, (x_n, Y_n)$ is used to estimate the calibration line $y = \alpha + \beta x$. The problem is to estimate the value of x which gives a specified value of y. The classic approach is to estimate x by $\hat{x} = (y - a)/b$ and apply Fieller's theorem for the construction of confidence intervals or tests. The endpoints of the Fieller interval are

$$\frac{\hat{x} - g\bar{x} \pm g^{\frac{1}{2}}[(\hat{x} - \bar{x})^2 + (1 - g) \sum_1^n (x_i - \bar{x})^2/n]^{\frac{1}{2}}}{1 - g} \tag{5.6}$$

provided $g < 1$, where

$$g = (t_{n-2}^{\alpha/2})^2 s^2/b^2 \sum_1^n (x_i - \bar{x})^2 \tag{5.7}$$

and $t_{n-2}^{\alpha/2}$ is the $1 - \alpha/2$ percentile point of the t distribution with $n - 2$ degrees of freedom.

The jackknife was tried on two numerical examples with $n = 10$ and $\alpha = \beta = \sigma = 1$ to learn if the asymptotics in this paper are at all valid when the sample size is not large. In the first example the x's were chosen equally-spaced: $x = 1(1)10$. In the second the x's were bunched toward one end with repeated values: $x = 1, 3, 5, 6, 6, 7, 8, 8.5, 9, 10$. In the first example the least squares estimates were $a = 1.20$, $b = .99$, and $s = .85$; the jackknifed estimates were $\tilde{a} = 1.04$ and $\tilde{b} = 1.01$. The estimated standard deviations for a and b were .82 and .13 and for $\tilde{a}$ and $\tilde{b}$ they were also .82 and .13. In the second example $a = .85$, $b = 1.05$, $s = .84$ and $\tilde{a} = .42$, $\tilde{b} = 1.10$. The estimated standard deviations of a and b were .69 and .10 and for $\tilde{a}$ and $\tilde{b}$ they were 1.16 and .16.

Values of y which would give x values near the middle and ends of the x range were selected for comparison purposes. They were $y = 3, 6.5, 10$ which correspond to $x = 2, 5.5, 9$. The classic and jackknife estimates and intervals for $\alpha = .05$ are presented in Table 1. The critical constant $t_{n-1}^{\alpha/2}$ was used in the jackknife intervals.

The jackknife performed very well in the first example in terms of reproducing the Fieller intervals and improving the point estimates. Its performance was similar in the second example except towards the low end of the x range. That

TABLE 1
Estimates and 95% confidence intervals for x

x	Type	Lower		Estimate		Upper	
		Ex. 1	Ex. 2	Ex. 1	Ex. 2	Ex. 1	Ex. 2
2	Fieller	.58	.67	1.82	2.05	2.69	2.99
	Jackknife	.49	.39	2.02	2.50	3.54	4.60
5.5	Fieller	4.70	4.71	5.35	5.39	5.98	5.98
	Jackknife	4.70	4.64	5.42	5.58	6.14	6.51
9	Fieller	8.04	8.04	8.88	8.73	10.04	9.67
	Jackknife	7.79	7.78	8.82	8.65	9.85	9.53

is the region where x values are sparse, so one might expect the jackknife not to do well there.

Another form of the inverse calibration problem is to estimate the value of x which gives the mean value of Y^0 which is an independent future observation. This problem is not covered by the results in this paper. If the estimate $(Y^0 - a)/b$ is jackknifed on $(x_1, Y_1), \cdots, (x_n, Y_n)$, the variability due to Y^0 is not incorporated into the confidence interval. The jackknife would give the same interval as if Y^0 were a fixed value $y = Y^0$ and not a random observation on $E(Y^0)$. If, instead, there are repeated observations $Y_1{}^0, \cdots, Y_m{}^0$ at the same x value, then the theorems do apply by deleting each of $Y_1{}^0, \cdots, Y_m{}^0, (x_1, Y_1), \cdots, (x_n, Y_n)$ successively provided $m/n \rightarrow c$, $0 < c < \infty$.

REFERENCES

[1] ARVESEN, J. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40** 2076–2100.

[2] ARVESEN, J. N. and LAYARD, M. W. J. (1971). Asymptotically robust tests in unbalanced variance component models. Technical Report No. 263, Department of Statistics, Purdue Univ.

[3] DUNCAN, D. B. and HORN, S. D. (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Amer. Statist. Assoc.* **67** 815–821.

[4] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*, **2**. Wiley, New York.

[5] GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467.

[6] GRAY, H. L., WATKINS, T. A. and Adams, J. E. (1972). On the jackknife statistic, its extensions, and its relation to e_n-transformations. *Ann. Math. Statist.* **43** 1–30.

[7] MILLER, R. G. JR. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305