JACKKNIFING IN NON-LINEAR REGRESSION

by

T. Fox, D.V. Hinkley and K. Larntz

Department of Applied Statistics, University of Minnesota

Technical Report No. 322

June 1978

University of Minnesota

St. Paul, Minnesota

*This research was supported in part by National Science Foundation Grant MCS77-00959 and National Institute of Education Grant G76-0094.

ABSTRACT

The standard jackknife and two linear jackknife methods based on a single fit are compared in the context of non-linear regression fitting. Emphasis is on determination of confidence regions for parameters, where we find that the standard jackknife may be inferior.

Key words: Jackknife; Non-linear Regression; Residual; Outlier; Likelihood; Confidence region.

1. INTRODUCTION

In a recent paper, Duncan [4] has discussed the application of the jackknife method to calculation of confidence regions for the parameters of a non-linear regression model. The full jackknife method requires the fitting of $n + 1$ non-linear regressions where n is the number of data vectors. In this paper we describe a "linear jackknife" procedure which requires only one non-linear fit. A corresponding "weighted jackknife," analogous to that defined by Jaeckel [10] and Hinkley [8], is also described.

The general problem we discuss is estimation for the non-linear regression model

$$y_j = f(x_j, \theta) + e_j \qquad (j = 1, \ldots, n), \tag{1.1}$$

where $\theta^T = (\theta_1, \ldots, \theta_t)$ is of primary interest. It is assumed in the estimation that $\text{var}(e_j) \equiv \sigma^2$, an unknown constant. The vector θ is estimated by least-squares, and we wish to set confidence limits on some or all of the components of θ. The jackknife procedure is distribution-free, not requiring a normal error distribution (or even homogeneous errors) for its approximate validity.

Section 2 recaps the definition of the usual "exact" jackknife, and Section 3 describes two linear jackknifes, one of which is weighted according to the independent-variable design. A simple example is given in Section 4, and some simulation results are presented and interpreted in Section 5.

2. THE EXACT JACKKNIFE

Basic theory for the jackknife technique has recently been outlined by Hinkley [7, 8, 9] and Duncan [4], the latter with special reference to non-linear regression.

We define $\hat{\theta}_{-i}$ (Duncan's $\tilde{\theta}_i$) to be the least-squares estimate of θ when the ith data point (x_i, y_i) is deleted from the sample. Then the pseudo-values are

$$P_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i} \quad , (i = 1, \ldots, n) \tag{2.1}$$

with average

$$\overline{P} = \hat{\theta}_J = n^{-1} \sum_{i=1}^{n} P_i \tag{2.2}$$

and variance

$$nV_P = \frac{1}{n-1} \sum_{i=1}^{n} (P_i - \overline{P})(P_i - \overline{P})^T . \tag{2.3}$$

The matrix V_P with elements $v_{P,jk}$ is the jackknife estimate of $\mathrm{Var}(\theta)$ and $\mathrm{Var}(\theta_J)$. If, as is usual, $\hat{\theta}$ has bias of the form

$$E(\hat{\theta}) - \theta = n^{-1}a_1(\theta) + n^{-2}a_2(\theta) + \ldots$$

then $\hat{\theta}_J$ has bias of order n^{-2} ; see Hinkley [8, Section 3] for discussion in the linear regression context.

For moderately large samples one will often assume an approximating normal distribution for $\hat{\theta}$ or $\hat{\theta}_J$, with variance matrix V_P . Then joint $1-\alpha$ confidence regions for θ are ellipses, e.g.

$$\theta : (\theta - \hat{\theta}_J)^T V_P^{-1} (\theta - \hat{\theta}_J) \leq \chi_t^2(1-\alpha) . \tag{2.4}$$

Individual confidence limits for component θ_k are of the form

$$\hat{\theta}_{J,k} \pm \sqrt{\{v_{P,kk}\, \chi_1^2(1-\alpha)\}} . \tag{2.5}$$

As usual, the accuracy of the normal approximation will depend on the choice parametrization.

The advertised advantage of the jackknife procedures (2.4) and (2.5) is validity robustness: they are free of error distribution assumptions, and may even be reasonable under heterogeneity of error variance (Hinkley [8]).

In the non-linear regression case the jackknife procedure as described requires $n + 1$ non-linear fits; usually $\hat{\theta}_{-i}$ is computed by iteration from initial value $\hat{\theta}$. The next section describes a simpler jackknife requiring only one non-linear fit.

3. THE LINEAR JACKKNIFE

The exact jackknife procedure of Section 2 is but one of several approximations, as Jaeckel [10], Hinkley [9] and Efron [5] point out. The methods to be described here are not necessarily inferior.

Linear approximations can be defined directly via estimates of the influence function for $\hat{\theta}$. Alternatively we can make a Taylor expansion of the least-squares estimating equation for $\hat{\theta}_{-i}$, assuming this to be a stationary point of

$$\sum_{j \neq i} \{y_j - f(x_j, \theta)\}^2 ,$$

and pick off the linear term. The result is

$$\hat{\theta}_{-i} \doteq \hat{\theta} - (\hat{Z}^T\hat{Z})^{-1} \hat{z}_i\left(\frac{r_i}{1 - \hat{\omega}_i}\right), \qquad (i = 1, \ldots, n) \qquad (3.1)$$

where

$$\hat{z}_i = \nabla f(x_i, \hat{\theta}) = \left(\frac{\partial}{\partial \theta_1} f(x_i, \theta), \ldots, \frac{\partial}{\partial \theta_t} f(x_i, \theta)\right)^T_{\theta=\hat{\theta}}$$

$$\hat{Z}^T = (\hat{z}_1, \ldots, \hat{z}_n) \quad , \quad \hat{\omega}_i = \hat{z}_i^T(\hat{Z}^T\hat{Z})^{-1}\hat{z}_i$$

and

$$r_i = y_i - f(x_i, \hat{\theta}) \quad .$$

The error of the linear approximation in (3.1) is of order $||\hat{\theta} - \hat{\theta}_{-i}||^2$, which typically implies a relative error of order n^{-1}.

Substitution of (3.1) in (2.1), and replacement of $n - 1$ by n, gives the linear pseudo-values

$$LP_i = \hat{\theta} + n(\hat{Z}^T\hat{Z})^{-1}\hat{z}_i\left(\frac{r_i}{1 - \hat{\omega}_i}\right) , \qquad (3.2)$$

whose sample mean $\overline{LP}$ and variance matrix nV_{LP} are equivalent to $\hat{\theta}_J$ and nV_P in (2.2) and (2.3). The calculation required is the original non-linear fit to obtain $\hat{\theta}$, plus calculation of the ancillary quantities $\hat{z}_i$, $\hat{\omega}_i$ and $\hat{r}_i$.

Hinkley [8] suggests probable inaccuracy of the standard jackknife when linear regression design weights w_i are fairly unequal. An alternative weighted jackknife is defined here by replacing pseudo-values P_i or LP_i by

$$LQ_i = \hat{\theta} + n(\hat{Z}^T\hat{Z})^{-1}\hat{z}_i r_i \qquad (i = 1, \ldots, n) . \tag{3.3}$$

There is some evidence in the linear case that the average $\overline{LQ}$ is less biased than $\overline{LP}$, and that the sample variance of the LQ_i leads to a better estimate of $\mathrm{Var}(\hat{\theta})$.

Notice that when $\hat{\theta}$ is a stationary point of the residual sum of squares, $\overline{LQ} = \hat{\theta}$ because $\Sigma \hat{z}_j r_j = 0$. Hence there is <u>no</u> bias reduction in replacing $\hat{\theta}$ by $\overline{LQ}$. The corresponding variance estimate V_{LQ} is given by

$$V_{LQ} = (\hat{Z}^T\hat{Z})^{-1}\left(\sum_{j=1}^{n} \hat{z}_j \hat{z}_j^T r_j^2\right)(\hat{Z}^T\hat{Z})^{-1} \quad . \tag{3.4}$$

This may be compared to the standard parametric normal-theory estimate

$$V_{EI} = (\text{Fisher information})^{-1} = \frac{1}{n}\Sigma r_j^2 (\hat{Z}^T\hat{Z})^{-1} , \tag{3.5}$$

where EI stands for "Expected Information." These formulae are similar to those in Section 2 of Hinkley [8], but here $\hat{Z}$ is a function of $\hat{\theta}$ and hence random.

In principle the expected values of the several variance estimates V_P , etc. can be obtained by expansion methods, but the results are extremely complicated and preclude simple general interpretations. Although all variance estimates are asymptotically equivalent, noticeable differences may occur in moderately large samples. One obvious cause of difference between V_{ML} and V_P, etc. is lack of homogeneity among residuals, as the comparison between (3.4) and (3.5) will indicate. The example in Section 4 illustrates this.

One general point to note is that all variance estimates V considered so far are unconditional, as are resulting confidence regions

$$\theta : (\theta - \hat{\theta})^T V^{-1} (\theta - \hat{\theta}) \leq \chi_t^2(1 - \alpha) \text{ or } F_{t,n-t}(1 - \alpha) \tag{3.6}$$

These may be inaccurate due to non-normal shape of the likelihood function, and indeed the direct normal-theory likelihood confidence region

$$\theta : \frac{1}{\hat{\sigma}^2} \{\Sigma(y_j - f(x_j, \theta))^2 - \Sigma(y_j - f(x_j, \hat{\theta}))^2\} \leq F_{t,n-t} \tag{3.7}$$

if often recommended (Beale, [1]); the latter method is also preferred on conditional grounds according to Efron and Hinkley [6]. A conditional replacement for V_{EI} in (3.5) is the inverse of observed information

$$V_{OI} = \hat{\sigma}^2 \{\hat{Z}^T \hat{Z} - \sum_{j=1}^{n} \nabla^2 f(x_j, \hat{\theta}) r_j\}^{-1} . \tag{3.8}$$

We shall not consider this further here. If (3.6) "fails" for $V = V_{OI}$ or $V = V_{EI}$, then the jackknife procedures are likely to fail also.

4. AN EXAMPLE

To compare the "exact" and "linear" jackknife methods, we continue the example given by Duncan [4] in his Tables 1 and 2. The model is

$$f(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2}\{\exp(-\theta_1 x) - \exp(-\theta_2 x)\} ,$$

and the observations (x, y) are in columns 2 and 3 of Table 4.1 below. For these data

$$\hat{\theta}_1 = 0.21162 \qquad \text{and} \qquad \hat{\theta}_2 = 0.44614 .$$

The rest of Table 4.1 contains values of r, $\hat{z}^T = (\hat{z}_1, \hat{z}_2)$, $\hat{\omega}$, $P^T = (P_1, P_2)$, $LP^T = (LP_1, LP_2)$ and $LQ^T = (LQ_1, LQ_2)$. Figure 4.1 plots LP_i and LQ_i versus P_i for $i = 1, 2$.

[Table 4.1 and Figure 4.1 about here.]

The jackknifed estimates, which are pseudo-value averages, are

	P	LP	LQ
θ_1 :	0.2103	0.2128	0.2116
θ_2 :	0.4443	0.4655	0.4461

with corresponding estimated variance matrices, computed as in (2.3),

$$V_P = 10^{-4}\begin{pmatrix} 8.43 & 6.40 \\ 6.40 & 26.84 \end{pmatrix}, \quad V_{LP} = 10^{-4}\begin{pmatrix} 8.60 & 3.76 \\ 3.76 & 20.35 \end{pmatrix}, \quad V_{LQ} = 10^{-4}\begin{pmatrix} 7.03 & 4.18 \\ 4.18 & 20.80 \end{pmatrix} .$$

A noteable feature of these estimates is the discrepancy among correlation estimates, which are 0.426, 0.284 and 0.346 for P, LP and LQ respectively. A bad estimate of $\mathrm{corr}(\hat{\theta}_1, \hat{\theta}_2)$ would lead to inaccurate joint confidence regions, even when separate confidence intervals are accurate.

TABLE 4.1 DATA AND JACKKNIFE PSEUDOVALUES FOR DUNCAN'S EXAMPLE

obs.#	x	y	r	z_1	z_2	w	P_1	P_2	LP_1	LP_2	LQ_1	LQ_2
1	.5	.00530	-.0845	-.4015	.02202	.0305	.0395	.2907	.0445	.3007	.0496	.3052
2	"	.04356	-.0463	"	"	"	.1187	.3620	.1202	.3666	.1230	.3690
3	"	.00603	-.0838	"	"	"	.0411	.2921	.0460	.3020	.0510	.3063
4	"	.05198	-.0378	"	"	"	.1359	.3775	.1368	.3810	.1391	.3830
5	1.0	.15303	.0004	-.6421	.07335	.0709	.2126	.4466	.2128	.4469	.2127	.4468
6	"	.17526	.0226	"	"	"	.2803	.4936	.2812	.4910	.2763	.4878
7	"	.15337	.0007	"	"	"	.2134	.4471	.2138	.4476	.2137	.4475
8	"	.20580	.0531	"	"	"	.3712	.5571	.3753	.5517	.3637	.5442
9	2.0	.36962	.1486	-.8071	.2040	.1017	.6897	.4816	.7080	.4310	.6576	.4326
10	"	.18513	-.0361	"	"	"	.0915	.4448	.0910	.4498	.1032	.4494
11	"	.25143	.0302	"	"	"	.3108	.4492	.3124	.4431	.3021	.4434
12	"	.25610	.0348	"	"	"	.3261	.4500	.3280	.4426	.3161	.4430
13	4.0	.18093	-.0546	-.5694	.3985	.1102	.1661	.7626	.1651	.7870	.1702	.7494
14	"	.19627	-.0393	"	"	"	.1793	.6762	.1781	.6913	.1818	.6643
15	"	.26221	.0267	"	"	"	.2320	.2821	.2343	.2799	.2318	.2982
16	"	.15962	-.0759	"	"	"	.1470	.8789	.1469	.9199	.1540	.8677
17	8.0	.11619	-.0244	.0643	.3960	.1530	.2823	.6977	.2819	.7018	.2711	.6627
18	"	.20856	.0680	"	"	"	.0026	-.2757	.0159	-.2665	.0458	-.1575
19	"	.18540	.0448	"	"	"	.0756	-.0270	.0826	-.0237	.1023	.0482
20	"	.09583	-.0448	"	"	"	.3392	.9037	.3405	.9153	.3208	.8435
21	16.0	.05278	.0230	.2206	.1157	.0336	.1713	.3614	.1704	.3589	.1718	.3619
22	"	.01473	-.0151	"	"	"	.2385	.5029	.2387	.5035	.2378	.5016
23	"	.05738	.0276	"	"	"	.1629	.3440	.1622	.3415	.1638	.3450
24	"	.02519	-.0046	"	"	"	.2198	.4635	.2199	.4638	.2197	.4632

FIGURE 4.1 Scatter plot of linear pseudo-values LP(∘) and LQ(•) versus standard pseudo-values P for example in Table 4.1

(a) First component

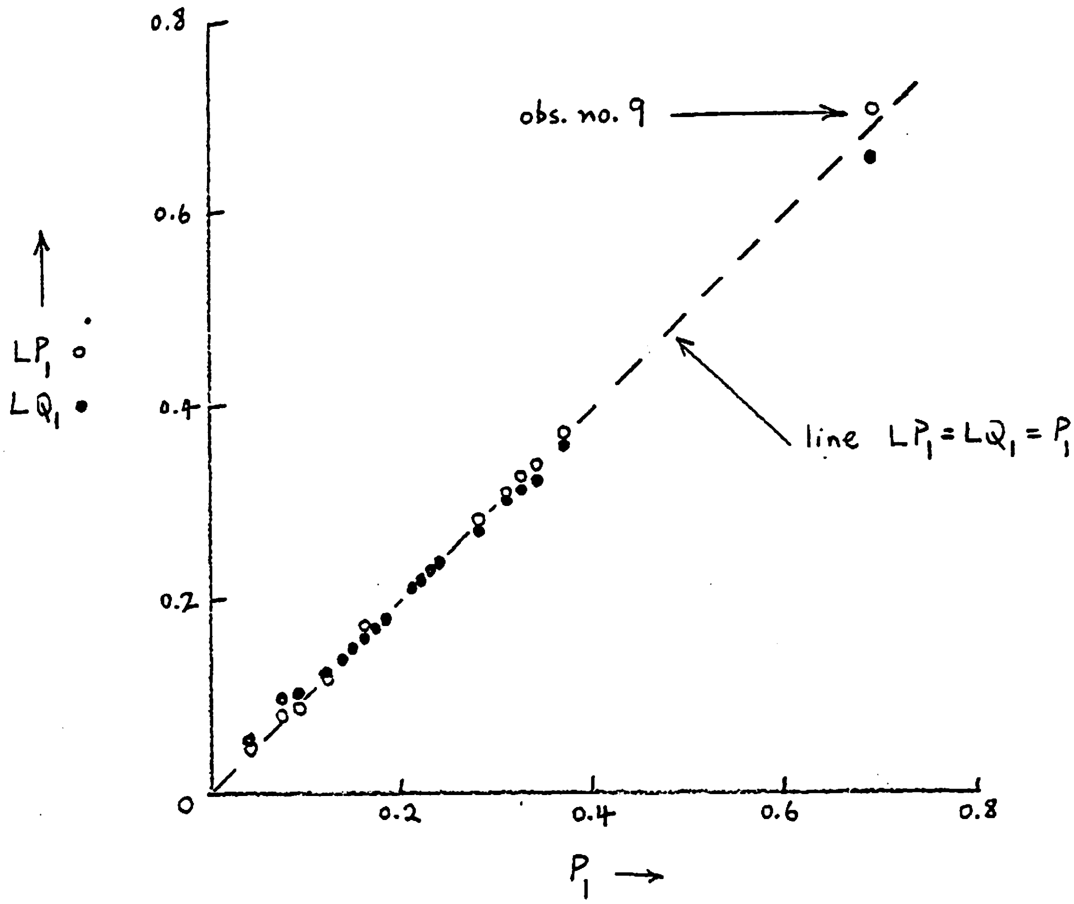
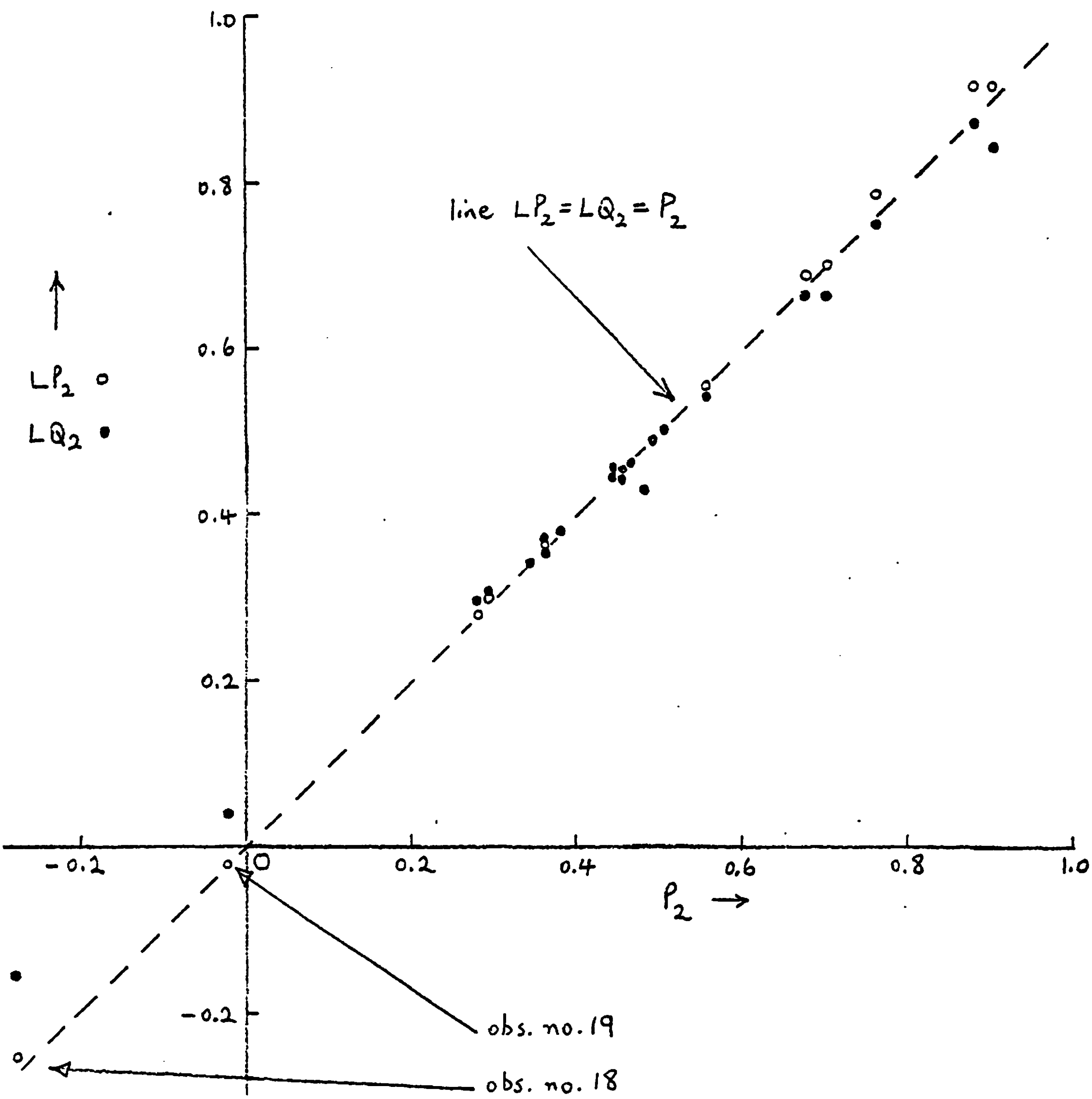


FIGURE 4.1 (continued)

(b) Second component



The plots of LP and LQ show that (i) P and LP are in close agreement, and (ii) that LQ matches P and LP except at the extremes. One might conclude that the linear pseudo-values LP are about as useful as the standard pseudo-values, but the potential advantage of LQ is not evidenced.

An important aspect of pseudo-values is illustrated by the plots, namely their analogy to residuals (Devlin et al [3], Hinkley [9]). It is quite clear that observation 9 is an outlier for estimating θ_1 , also that observations 18 and 19 may be outliers for estimating θ_2 . Cook's [2] distance measure would give essentially the same indications. Usually one would have performed a normal plot of the pseudo-values, and the above-mentioned points would show up clearly.

If the data are re-analyzed with observations 9, 18, and 19 removed, then the estimates of θ_1 , θ_2 and their estimated variance matrices become as in Table 4.2. The estimates are now close to the supposed true values

TABLE 4.2 ESTIMATES AND VARIANCES FOR EXAMPLE WITH DATA POINTS 9, 18, 19 OMITTED

			Jackknife		
Method:		LS*	P	LP	LQ
estimates	θ_1	0.208	0.208	0.211	0.208
	θ_2	0.512	0.514	0.519	0.512
estimated variances	v_{11}	5.08×10^{-4}	5.46×10^{-4}	5.67×10^{-4}	5.08×10^{-4}
	v_{12}	7.86×10^{-4}	5.65×10^{-4}	5.94×10^{-4}	5.34×10^{-4}
	v_{22}	33.49×10^{-4}	24.23×10^{-4}	27.56×10^{-4}	22.08×10^{-4}

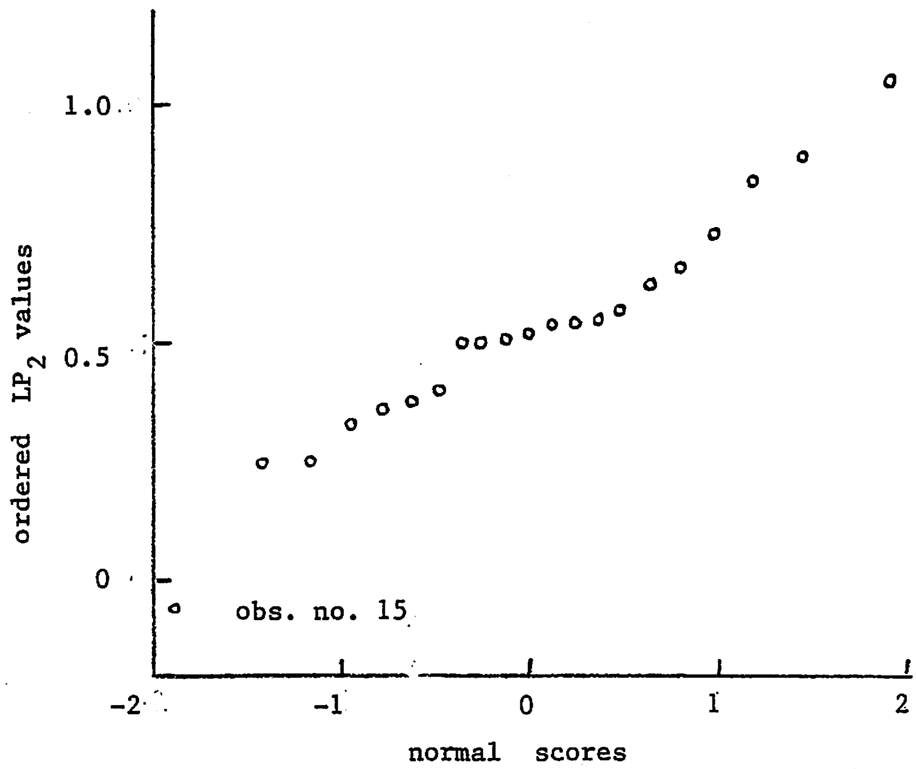
* estimated variance matrix is V_{EI}

($\theta_1 = 0.2$, $\theta_2 = 0.5$) used to generate the data. Also the estimates of $\mathrm{corr}(\hat{\theta}_1, \hat{\theta}_2)$ are much higher. The jackknife estimates of $\mathrm{var}(\hat{\theta}_2)$ have not decreased because observation 15 has somewhat discrepant values of P_1, LP_1, LQ_1. Figure 4.2 shows a normal plot of the values of LP_2.

We have not pursued the analysis of the example to a final conclusion, but one reasonable interpretation of the pseudo-value analysis up to this point is that errors are non-homogeneous, as Duncan's Figure 1 suggests.

[Figure 4.2 about here.]

FIGURE 4.2 Normal plot of LP_2 values for analysis of data in Table 4.1 with observations 9,18 and 19 removed



5. SOME SIMULATION RESULTS

According to Duncan [4], the standard jackknife works quite well in obtaining separate confidence limits for components of θ, but misbehaves in joint confidence region procedures, as in (2.4). In the small-scale simulations to be described here we have concentrated on comparisons among the jackknifes, but some attention has been given to Duncan's disturbing findings.

The main simulations have been carried out for Duncan's Model II with additive error, that is (1.1) with

$$f(x, \theta) = 1 - \frac{1}{\theta_1 - \theta_2}(\theta_1 e^{-\theta_2 x} - \theta_2 e^{-\theta_1 x}) ; \quad \theta_1, \theta_2 \geq 0 . \quad (5.1)$$

Note that θ_1 and θ_2 are interchangeable, so that estimation should strictly be confined to $\theta_1 \geq \theta_2$. Likelihood contours have a tendency toward boomerang shape, rather than elliptical, when plotted in the full space. This is associated with a breakdown of the elliptical confidence regions (2.4), as we shall see.

The main set of results on model (5.1) was obtained with $\theta_1 = 0.2$, $\theta_2 = 0.5$, standard normal errors and the same x-design as in Table 4.1. Coverage frequencies of nominal 95% confidence regions in 100 simulated samples are given in Table 5.1. In the overlap with Duncan's results [4, Table 3] there is general agreement, although both studies are small. The general tentative conclusions that we would draw are:

(i) the linear jackknife using LP is better than the standard jackknife and corresponds to the normal-theory maximum likelihood method, i.e. (3.6) with $V = V_{EI}$;

(ii) all methods based on normal approximation (3.6) are poor for joint confidence regions, whereas the direct likelihood method (3.7) is reasonably good in this model.

TABLE 5.1. COVERAGE OF NOMINAL 95% CONFIDENCE REGIONS *
FOR MODEL (5.1) WITH $\theta_1 = 0.2$, $\theta_2 = 0.5$,
AND DESIGN OF TABLE 4.1.

Method / Parameter(s)	Direct Likelihood (3.7)	Normal approximation			
		$V=V_{EI}$	$V=V_P$	$V=V_{LP}$	$V=V_{LQ}$
θ_1	94%	92%	84%	91%	88%
θ_2	91%	99%	89%	96%	96%
(θ_1, θ_2)	91%	79%	57%	73%	72%

* In these calculations Student-t percentage points were used in confidence limit formula (2.5) for all V.

A second set of results duplicated the first except that errors were simulated from a Student t_2 distribution. The patterns were almost identical to those in Table 5.1, although the coverage rates were all higher.

REFERENCES

[1] Beale, M.L. (1960) Confidence regions in non-linear estimation (with discussion). J. Roy. Statist. Soc. B, 22, 41-88.

[2] Cook, R.D. (1977) Detection of influential observations in linear regression. Technometrics 19, 15-18.

[3] Devlin, S. and Gnanadesikan, R. (1975) Robust estimation and outlier detection with correlation coefficients. Biometrika 62, 531-545.

[4] Duncan, G.T. (1978) An empirical study of jackknife-constructed confidence regions in nonlinear regression. Technometrics 20, 123-129.

[5] Efron, B. (1978) Bootstrap methods: another look at the jackknife. Ann. Statist. 6 (to appear).

[6] Efron, B. and Hinkley, D.V. (1978) Assessing the accuracy of the maximum likelihood estimate: observed versus expected information. Biometrika 65 (to appear).

[7] Hinkley, D.V. (1977) Jackknife confidence limits using Student t approximations. Biometrika 64, 21-8.

[8] Hinkley, D.V. (1977) Jackknifing in unbalanced situations. Technometrics 19, 285-292.

[9] Hinkley, D.V. (1978) Improving the jackknife with special reference to correlation estimation. Biometrika 65, 13-21.

[10] Jaeckel, L.A. (1972) The infinitesmal jackknife. Unpublished Bell Labs. Memorandum.