The Annals of Statistics
1987, Vol. 15, No. 1, 325–345

AN EXTENSION OF PARTIAL LIKELIHOOD METHODS FOR PROPORTIONAL HAZARD MODELS TO GENERAL TRANSFORMATION MODELS[1]

BY KJELL A. DOKSUM

University of California, Berkeley

Estimates of the linear model parameters in a linear transformation model with unknown increasing transformation are obtained by maximizing a partial likelihood. A resampling scheme (likelihood sampler) is used to compute the maximum partial likelihood estimates. It is shown that for a certain "local" parameter set where the "signal to noise ratio" is small, it is asymptotically possible to estimate the linear model parameters using the partial likelihood as well as if the transformation were known. In the case of the power transformation model with symmetric error distribution, this result is shown to also hold when the distribution of the error in the transformed linear model is unknown and is estimated. Monte Carlo results are used to show that for moderate sample size and small to moderate signal to noise ratio, the asymptotic results are approximately in effect and thus the partial likelihood estimates perform very well. Estimates of the transformation are introduced and it is shown that the estimates, when centered at the transformation and multiplied by $\sqrt{n}$, converge weakly to Gaussian processes.

1. Introduction. Consider the transformation model where an unknown increasing transformation $h(Y)$ of the response variable Y follows a linear model with p covariates $x_1, \ldots, x_p$. Since the Cox (1972, 1975) proportional hazard model with time independent covariates is a special case of such a transformation model, and since partial (marginal) likelihood methods have proven so useful in the proportional hazard model, we investigate properties of partial likelihood methods in the transformation model.

In the preceding model the partial likelihood is proportional to the projection of the standardized likelihood onto the space of rank statistics and it is proportional to the distribution of the ranks. Thus it could be called the rank likelihood or marginal likelihood. Starting with Hoeffding (1951), this rank likelihood has been used very successfully to generate test statistics, e.g., Terry (1952), Lehmann (1953, 1959), Savage (1956, 1957), Hájek and Šidák (1967) and Kalbfleisch and Prentice (1973).

In estimation, it has been used in the proportional hazard case by Cox (1972, 1975). A local approximation to the partial likelihood has been used by Pettitt (1982, 1983, 1984), but otherwise its use in estimation has been limited by computational difficulties. In this paper, a resampling scheme called the

Received March 1985; revised December 1985.

[1]This work was supported in part by National Science Foundation Grant MCS83-01716, the Royal Norwegian Council for Scientific and Industrial Research, and the Department of Statistics, University of Trondheim, NTH.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62J02.

Key words and phrases. Partial likelihood, marginal likelihood, rank likelihood, semiparametric transformation models, proportional hazard model.

likelihood sampler is introduced to compute the partial (marginal, rank) likelihood and the maximum partial likelihood estimates (MPLE's). Monte Carlo techniques are used to show that the estimates perform very well for moderate sample sizes and a certain range of parameter values.

When the results of this paper are applied to parameters that have interpretations unrelated to h and to prediction, there is no controversy as to the relevance and interpretation of the results. However, for parameters defined in terms of h, the interpretation and properties of the linear model parameters and their estimates is a sticky problem full of controversies, interesting questions, and different approaches (see Hinkley and Runger and discussants (1984)). We consider a parameter space for which the various different approaches asymptotically coincide and the parameters and their estimates have simple slope interpretations. This parameter set can be regarded as a *domain of adaptability*, i.e., a space on which the linear model parameters can be estimated as well as if h were known. It is a local parameter set that can be described as the set of parameter values for which the probability distribution of the data is contiguous to a power probability measure (Le Cam (1960), Hájek (1962), Hájek and Šidák (1967)).

This local parameter set is the one that has been used successfully in testing problems to obtain approximations to the power of tests. Here it is being used to obtain useful approximations to the biases and mean squared errors of estimates of the linear model parameters in the transformed linear regression model. Monte Carlo results show that the approximations are surprisingly accurate for moderate sample sizes and a range of parameter values with small to moderate signal to noise ratio.

The finding that for the local parameter set, parameters can be estimated as well as if h is known is consistent with the findings of Doksum and Wong (1983) and Carroll (1982) for testing problems. It is different from the results of Bickel and Doksum (1981), who considered nonlocal alternatives and found a severe increase in variability of the estimates of the linear model parameters *relative* to the h-known case when $\mathrm{Var}(h(Y))$ is small. Carroll and Ruppert (1981) and Taylor (1986) also considered nonlocal alternatives. They found a moderate increase in variability for prediction problems relative to the h-known case.

In Section 5 we combine results of Hájek (1962) and Hinkley (1975) to construct estimates that, in the case of power transformations and the local parameter set, are adaptive when both the transformation and the distribution of the error in the transformed linear model are unknown.

The problem of estimating a nonparametric transformation h has been considered by Fisher (1946), Kruskal (1965) and Breiman and Friedman (1985), among others. In Section 6 we introduce estimates $\hat{h}$ with the property that $\sqrt{n}(\hat{h} - h)$ converges weakly to Gaussian processes.

2. Preliminaries.

2.1. *Transformation models.* The independent random variables $Y_1, \ldots, Y_n$ are said to follow a linear transformation model if for some increasing transfor-

mation h,

$$h(Y_i) = \alpha + \beta \mathbf{x}_i + \sigma\varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\mathbf{x}_i' = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$, are known constants, $\beta = (\beta_1, \ldots, \beta_p)$ is a vector of regression parameters, α is an intercept parameter and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. with distribution F. The model (2.1) can be regarded as a special case of the model of Breiman and Friedman (1985) where the covariates are also transformed.

Some examples of h are $h(y) = (y + c)^\lambda$, $h(y) = \text{sign}(y)|y|^\lambda$ and $h(y) = (y^\lambda - 1)/\lambda$, $\lambda \neq 0$, $h(y) = \log y$, $\lambda = 0$. See for instance Anscombe and Tukey (1954), Tukey (1957), Box and Cox (1964) and Bickel and Doksum (1981). For these examples, which we refer to as power transformations, F is often taken to be the standard normal distribution function Φ.

Another parametric family of increasing transformations is given by the Beall (1942) transform $h(y) = \sinh^{-1}(\sqrt{\lambda}y)/\sqrt{\lambda} = \log(\sqrt{\lambda}y + \sqrt{1 + \lambda y})/\sqrt{\lambda}$, $\lambda > 0$, $h(y) = \sqrt{y}$, $\lambda = 0$. Sometimes it is not clear which parametric family is appropriate for a given experiment. We will primarily consider the nonparametric case where h is continuous and increasing, but otherwise arbitrary; however, in Sections 5 and 7, power transformations are considered.

2.2. *The proportional hazard model as a transformation model.* Suppose Y_i is a survival time with distribution F_i and hazard rate $r_i = f_i/[1 - F_i]$, $i = 1, \ldots, n$. The proportional hazard model (Cox (1972, 1975)) is

$$r_i(t) = \Delta_i r(t), \quad \text{some } r(t), \text{ where } \Delta_i = \exp(\beta \mathbf{x}_i).$$

An equivalent form of this model is the Lehmann (1953) form,

$$F_i(t) = 1 - [1 - F_0(t)]^{\Delta_i}, \quad \text{where } F_0(t) = 1 - \exp\left[-\int_0^t r(x)\,dx\right].$$

It follows that we can write

$$\log\{-\log[1 - F_0(Y_i)]\} = -\beta \mathbf{x}_i + \varepsilon_i,$$

where the $\{\varepsilon_i\}$ are i.i.d. with the extreme value distribution $1 - \exp(-e^t)$. In other words, the proportional hazard model is a transformation model of the form (2.1) with $h(y) = \log\{-\log[1 - F(y)]\}$.

2.3. *Interpretation and properties of the parameters and estimates.* The model for one response Y is

$$h(Y) = \alpha + \sum_{j=1}^{p} x_j\beta_j + \sigma\varepsilon. \tag{2.2}$$

Box and Cox (1982), Hinkley and Runger (1984), Rubin (1984), Bickel (1984), Carroll and Ruppert (1984) and Doksum (1984) discuss the relevance, interpretation and properties of the parameters $\beta_1, \ldots, \beta_p$ and their estimates. Here we

consider two cases:

(i) *β_j has an interpretation unrelated to the transformation h.* In the proportional hazard model, β_j is the decrease in log hazard as x_j is increased one unit while the x_k, $k \neq j$, are held fixed. Thus β_j has an interpretation independent of the transformation and h (or the hazard rate r) is treated as an unknown nuisance parameter. (Note, however, that the relative decrease in log hazard depends on h. Often relative decrease is important; for instance, a smoker will pay more attention if told that quitting will result in a 10% reduction of the log hazard than if told that the log hazard will drop by 0.04.)

(ii) *β_j has an interpretation related to h.* Assume that the expected value of ε_i exists. Then, in the power transformation model, and in the general model (2.2), β_j is the increase in the mean of $h(Y)$ as x_j is increased one unit with x_k, $k \neq j$, held fixed. Thus we call β_j the *slope parameter* for the jth covariate on the scale h. Since the interpretation of β_j depends on h, it is necessary to report an estimate $\hat{h}$ of h as well as an estimate $\hat{\beta}_j$ of β_j. Hinkley and Runger (1984) argue that the analysis should proceed on the scale $\hat{h}$ with $\hat{h}$ regarded as fixed. In this approach, estimates are reported on the estimated scale $\hat{h}(Y)$ and no allowance is made for the randomness of $\hat{h}$ since $\hat{h}$ is rendered fixed and nonrandom by a conditioning argument. This approach makes sense when there is a unique, natural transformation h that will produce a linear relationship, and when this transformation becomes apparent after contemplating $\hat{h}$ and the mechanisms producing the data. One good example is x = weight of an automobile, y = miles per gallon. Here $h(y) = -1/y$ emerges as the natural transformation (e.g., Hocking (1976)).

Next consider the case where no such natural h emerges, and $\hat{h}$ is an estimate of the transformation h that produces a linear relationship. An example occurs in meteorology where the third root and fourth root of precipitation appear to be about equally popular transformations (see Woodley et al. (1977), fourth root, and Miller et al. (1979), third root). Thus it is natural to ask, if the correct scale is $h(y) = y^{1/4}$, but we report the estimate on the wrong scale $h(y) = y^{1/3}$, what errors are committed?

Returning to the general case, note that if we use the conditional approach where $\hat{h}$ is fixed and we regard $\{\hat{h}(y)\}$ as our "data," $\hat{h}(y)$ satisfies

$$\hat{h}(Y) = \hat{h}h^{-1}\left(\alpha + \sum_{j=1}^{p} x_j\beta_j + \sigma\varepsilon\right). \tag{2.3}$$

Thus, β_j is not a slope parameter on the scale $\hat{h}$ and on this scale, β_j does not have a simple interpretation. Moreover, in this conditional approach, the function $\hat{h}h^{-1}$ relating β_j to $\hat{h}(y)$ is unknown. See also Rubin (1984).

If we want to keep the slope interpretation, we would say that β_j is a slope parameter on the linear model scale h, and since h is unknown, we would, in addition to giving an estimate $\hat{\beta}_j$ of β_j, also give an estimate $\hat{h}$ of h. Now an allowance needs to be made for $\hat{h}$ being estimated and random. One of the problems with this approach is that β_j is a slope parameter on the unknown

scale h, but it does have the advantage that it is compatible with testing, confidence intervals derived from tests and prediction (see Doksum (1984)).

However, it would be preferable to have it both ways; that is, to report $\hat{\beta}_j$ on the estimated scale $\hat{h}$ *and* keep the slope parameter interpretation. We heuristically derive conditions under which this is possible in an approximate sense. These conditions limit the range of the parameters as explained in Sections 4 and 7.

Suppose we operate with the wrong reported scale $h_1(y) = y^{\lambda_1}$ when the true scale is $h(y) = y^{\lambda}$. Let $\delta = (\lambda_1/\lambda) - 1$; then the slope for the jth covariate on the reported scale is

$$\beta_j(\alpha, \beta, \sigma) = \frac{d}{dx_j} E h_1(y) = (\delta + 1) E(\alpha + \Sigma x_k \beta_k + \sigma\varepsilon)^{\delta} \beta_j.$$

Thus if $\hat{\beta}_j$ is reported as the estimate of the slope β_j in $\hat{h}$ units, the $\sqrt{n}$ scaled error is $D_n = \sqrt{n}[\hat{\beta}_j - \beta_j(\alpha, \beta, \sigma)]$. Setting $\delta = a/\sqrt{n}$, a Taylor expansion gives

$$D_n \simeq \sqrt{n}\left(\hat{\beta}_j - \beta_j\right) + a\left[1 + E \log\left(\alpha + \sum_{k=1}^{p} x_k \beta_k + \sigma\varepsilon\right)\right]\beta_j + R_n,$$

where R_n tends in probability to zero as $n \to \infty$. See also Bickel (1984).

From this expression, we see that plausible conditions that ensure $D_n \simeq \sqrt{n}(\hat{\beta}_j - \beta_j)$ are $\beta_j = o(1)$, $\Sigma_{k=1}^{n} x_k \beta_k = O(1)$ and $\delta = O(n^{-1/2})$. In later sections we will see that $\sqrt{n}(\hat{\beta}_j - \beta_j)$ has a nondegenerate limiting normal distribution and that $\hat{h}(y) = h(y) + O_p(n^{-1/2})$. Thus, heuristically, $D_n \simeq \sqrt{n}(\hat{\beta}_j - \beta_j)$. These conditions are compatible with the conditions of Section 4.

3. The maximum partial likelihood estimates (MPLE) and the likelihood sampler. Cox's partial likelihood idea can be applied to the transformation model (2.1) with h increasing but otherwise unknown, and F continuous. See Kalbfleisch (1978) and Pettitt (1982, 1983). The partial likelihood for β is equivalent to the likelihood of the rank vector $\mathbf{R} = (R_1, \ldots, R_n)$, where $R_i = \text{Rank}(Y_i) = \text{Rank}(h(Y_i))$. Let $\mathbf{r} = (r_1, \ldots, r_n)$ be the vector of ranks obtained in an experiment. The MPLE is the vector $\hat{\beta}$ that maximizes

$$\tilde{L}_F(\beta) = P(\mathbf{R} = \mathbf{r}).$$

The case where F is known will be considered first. Later, the case F unknown will be considered, as well as the problem of estimating h.

Note that in addition to the unknown h case, $\tilde{L}_F(\beta)$ is a natural basis for statistical inference in regression or ANOVA experiments where only the ranks of the responses are available.

Next we consider identifiability conditions. Since the ranks are invariant under shift, $\tilde{L}_F(\beta)$ cannot be used to estimate an intercept parameter and we set $\alpha = 0$. Moreover, because of this shift invariance, we can and will reparametrize

to have $\sum_{i=1}^{n} x_{ij} = 0$, $j = 1, \ldots, p$. Similarly, if we divide by σ in each term of the model (2.1), we find that the ranks remain unchanged while the parameters change from $(\beta_1, \ldots, \beta_p, \sigma)$ to $(\beta_1/\sigma, \ldots, \beta_p/\sigma, 1)$. Thus we reparametrize to have $\sigma = 1$. For simplicity of notation, we assume this reparametrization has already been incorporated into the model (2.1).

Next note that if $p = 1$, $x_{i1} = i$, and if the support of F is finite, it is possible to choose the β's so that the ranks are equal to $(1, 2, \ldots, n)$, and, for a range of β's, we do not have identifiability. We avoid such problems by using a condition that implies that the support of F is the whole real line. Finally, let $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$ be the design matrix; we will make the familiar assumption that $\mathbf{X}$ has rank p. Here, then, is a summary of assumptions and notation:

$$h(Y_i) = \beta \mathbf{x}_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$

$$\sum_{i=1}^{n} x_{ij} = 0, \qquad \mathbf{X} \text{ has rank } p, \tag{3.1}$$

where h is increasing on the real line and

$$F \text{ is absolutely continuous with a density } f \text{ that satisfies } f(x) > 0,\ x \in R. \tag{3.2}$$

For convenience, we introduce the standardized partial likelihood

$$L_F(\beta) = \tilde{L}_F(\beta)/\tilde{L}_F(\mathbf{0}) = n!\tilde{L}(\beta).$$

PROPOSITION 3.1 (Hoeffding, 1951). *If* (3.1) *and* (3.2) *hold, then*

$$L_F(\beta) = E\left\{ \prod_{i=1}^{n} \frac{f\left(V^{(r_i)} - \beta \mathbf{x}_i\right)}{f\left(V^{(r_i)}\right)} \right\}, \tag{3.3}$$

where $V^{(1)} < \cdots < V^{(n)}$ *are the order statistics in a sample of size n from F.*

Typically, there is no explicit solution to the integral in (3.3). In this case, we can use a resampling scheme to approximate $L_F(\beta)$: On the computer, generate M independent ordered samples $V_k^{(1)} < \cdots < V_k^{(n)}$, $k = 1, \ldots, M$, where each ordered sample is the order statistics of a sample of size n from F. Now we approximate $L_F(\beta)$ by $\hat{L}_{F,M}(\beta)$, where

$$\hat{L}_{F,M}(\beta) = \frac{1}{M} \sum_{k=1}^{M} f_k(\beta) \quad \text{and} \quad f_k(\beta) = \prod_{i=1}^{n} \frac{f\left(V_k^{(r_i)} - \beta \mathbf{x}_i\right)}{f\left(V_k^{(r_i)}\right)}. \tag{3.4}$$

PROPOSITION 3.2. *If* (3.1) *and* (3.2) *hold, then* $\hat{L}_{F,M}(\beta)$ *converges almost surely to* $L_F(\beta)$ *as* $M \to \infty$.

PROOF. By the strong law of large numbers, it is enough to show that $f_k(\beta)$ has finite expected value. This follows since $E[f_k(\beta)] = L_F(\beta)$. □

Now the procedure is to maximize $\hat{L}_{F,M}(\beta)$ for $M = 100, 200, \ldots$ and to stop when the change in the resulting estimates $\hat{\beta}_M$, $M = 100, 200, \ldots$ from one M to the next is within prescribed precision.

Note that at each sampling stage k the preceding resampling scheme introduces variables $V_k^{(r_1)}, \ldots, V_k^{(r_n)}$ that have the same order as the original data $Y_1, \ldots, Y_n$. Thus it preserves order, but introduces new random order statistics at each stage k. This is in contrast to Efron's (1979) resampling scheme, the bootstrap, which at each sampling stage uses the original order statistics but in a different (random) order. We call the scheme based on (3.4) the likelihood sampler. A different approximation to $L_F(\beta)$ has been considered by Pettitt (1982, 1983).

EXAMPLE 3.1. Suppose, as in Box and Cox (1964), that $F = \Phi$, the standard normal distribution. Then the partial likelihood is

$$L_\Phi(\beta) = \frac{1}{n!} E\left\{\exp\left[\sum_{i=1}^{n} \mu_i Z^{(r_i)} - \frac{1}{2}\sum_{i=1}^{n} \mu_i^2\right]\right\},$$

where $Z^{(1)} < \cdots < Z^{(n)}$ are the order statistics of a sample of size n from Φ and $\mu_i = \beta \mathbf{x}_i = \sum_{j=1}^{p} \beta_j x_{ij}$. We illustrate the likelihood sampler on the steam data of Draper and Smith ((1981), page 9) in Table 1.

Table 2 shows how, for fixed M, $\hat{\beta}_M$ and $\hat{L}_{\Phi,M}(\hat{\beta}_M)$ change with the normal random deviates used in the likelihood sampler and it shows how they change with M. The standard error of $\hat{\beta}$ is approximately $1/\sqrt{\Sigma x_i^2} = 0.012$. For comparison, note that the corresponding normal model estimate is $\hat{\beta}_{LS}/\hat{\sigma} = -0.090$.

TABLE 1

Steam data. t is the atmospheric temperature in degrees Fahrenheit, $x_i = t_i - t$, and y is pounds of steam used per month.

t_i	35.3	29.7	30.8	58.8	61.4	71.3	74.4	76.7	70.7	57.5	46.4
y_i	10.98	11.13	12.51	8.40	9.27	8.73	6.36	8.50	7.82	9.14	8.24
r_i	20	22	25	7	14	10	1	9	4	13	6

t_i	28.9	28.1	39.1	46.8	48.5	59.3	70.0	70.0	74.5	72.1	58.1	44.6	33.4	28.6
y_i	12.19	11.88	9.57	10.94	9.58	10.09	8.11	6.83	8.88	7.68	8.47	8.86	10.36	11.08
r_i	24	23	15	19	16	17	5	2	12	3	8	11	18	21

TABLE 2

The likelihood sampler for the steam data.

M	100	100	100	200	200	200	400	400	400
$\hat{\beta}_M$	−0.060	−0.061	−0.059	0.064	−0.062	−0.061	−0.064	−0.063	−0.064
$10^{-4}L(\hat{\beta}_M)$	3.82	6.46	3.10	0.04	7.69	5.79	8.41	6.80	8.48

We return to the general case and consider the problem of obtaining relationships between the likelihood and partial likelihood. The likelihood for β in model (3.1) assuming that h is known, expressed in terms of $Z_i = h(Y_i)$, is

$$\tilde{\mathscr{L}}_{F,h}(\beta) = \prod_{i=1}^{n} f(Z_i - \beta \mathbf{x}_i).$$

For convenience, we introduce the standardized likelihood

$$\mathscr{L}_{F,h}(\beta) = \tilde{\mathscr{L}}_{F,h}(\beta)/\tilde{\mathscr{L}}_{F,h}(\mathbf{0}).$$

PROPOSITION 3.3. *For the model* (3.1), *if* (3.2) *holds, then*

$$E_{P_0}(\mathscr{L}_{F,h}(\beta)|\mathbf{R} = \mathbf{r}) = L_F(\beta),$$

where P_0 *represents the distribution when* $\beta = 0$.

PROOF. This follows from Proposition 3.1 since on $[\mathbf{R} = \mathbf{r}]$, $V_i = V^{(r_i)}$, and since the ranks and order statistics of $h(Y_1), \ldots, h(Y_n)$ are independent under P_0. □

In other words, the (standardized) partial likelihood $L_F(\beta)$ is the projection of the (standardized) likelihood $\mathscr{L}_{F,h}(\beta)$ onto the space of rank statistics. We readily obtain

COROLLARY 3.1. *Under* (3.1) *and* (3.2), $E_{P_0}(\mathscr{L}_{F,h}(\beta)) = E_{P_0}(L_F(\beta))$.

In the two-sample case, it is possible to obtain limits as well as asymptotic normality results for $L_F(\beta)$ as we now show.

REMARK 3.1. Consider the transformed two-sample shift model where $h(Y_i) \sim F(\cdot - \theta)$, $i = 1, \ldots, n_1$; $h(Y_i) \sim F(\cdot)$, $i = n_1 + 1, \ldots, n$. By shifting the $h(Y_i)$ by a constant amount, model (3.1) will be satisfied with $\beta = \theta$. Let $Q_F(\beta) = (1/n)\log L_F(\beta)$; then, using Hájek (1974), we can conclude that if F is absolutely continuous and $\log[\bar{f}_\theta(u)/\bar{g}_\theta(u)]$ is integrable and has bounded variation on every closed subinterval of $(0, 1)$, then, with probability one,

$$\lim_{n\to\infty} Q_F(\beta) = \lambda \int_0^1 \bar{f}_\theta(u) \log \bar{f}_\theta(u)\, du + (1 - \lambda) \int_0^1 \bar{g}_\theta(u) \log \bar{g}_\theta(u)\, du,$$

where

$$\bar{f}_\theta(u) = \frac{d}{du} F\big(H_\theta^{-1}(u) - \theta\big), \qquad \bar{g}_\theta(u) = \frac{d}{du} F\big(H_\theta^{-1}(u)\big),$$

$$H_\theta(t) = \lambda F(t - \theta) + (1 - \lambda) F(t) \quad \text{and} \quad \lambda = \frac{n_1}{n}.$$

See also Berk and Savage (1968) for a similar result.

REMARK 3.2. Consider the proportional hazard two-sample model where $Y_i \sim G_0^\theta$, $i = 1, \ldots, n_1$; $Y_i \sim G_0$, $i = n_1 + 1, \ldots, n$. This model is equivalent to

the transformation model $h(Y_i) \sim \beta + \varepsilon_i$, $i = 1, \ldots, n_1$; $h(Y_i) \sim \varepsilon_i$, $i = n_1 + 1, \ldots, n$, where $h(y) = \log(-\log G_0(y))$, $\beta = e^{\theta}$ and ε_i has distribution $F(t) = 1 - \exp(-e^t)$. Let $F_\theta(t) = G_0^\theta(t)$,

$$\mu(F_\theta, G_0, \theta) = \log(4\theta) - 2 - \int \log(F_\theta(x) + \theta G_0(x))(dF_\theta(x) + dG_0(x))$$

and

$$\sigma^2(F_\theta, G_0, \theta)$$

$$= 2(\theta - 1)^2 \Big\{ \iint_{x<y} [G_0(x)(1 - G_0(y))/W(x)W(y)]\, dF_\theta(x)\, dF_\theta(y)$$

$$+ \iint_{x<y} [F_\theta(x)(1 - F_\theta(y))/W(x)W(y)]\, dG_0(x)\, dG_0(y) \Big\},$$

$$W = F_\theta + \theta G_0.$$

From Sethuraman (1970) and Lai (1975), it follows that, with Q_F as in Remark 3.1. $\sqrt{n}(Q_F(\beta) - \mu(F_\theta, G_0, \theta)) \to_d N(0, \sigma^2(F_\theta, G_0, \theta))$ where $\to_d$ denotes convergence in distribution. This result can be used to establish the asymptotic normality of $\hat{\theta}$ and $\hat{\beta}$. See also Begun (1981).

4. A local approximation to the partial likelihood.

4.1. *A local parameter set.* If $E(\varepsilon_i)$ exists, we can without loss of generality think of $\mu_i = \sum_{j=1}^{p} x_{ij}\beta_j$ as the mean of $h(Y_i)$. Moreover, in our parametrization, $\bar{\mu} = n^{-1}\sum \mu_i = 0$. We will assume that $\boldsymbol{\beta} \in \Omega_n$, where

$$\Omega_n = \left\{ \boldsymbol{\beta};\ \sum_{i=1}^{n} \mu_i^2 \le K^2,\ \max_{1 \le i \le n} |\mu_i| \to 0 \right\}. \tag{4.1}$$

In (4.1), K^2 is a constant not dependent on n, while $\boldsymbol{\beta}$ and μ_i may depend on n although this is suppressed in the notation. If we let $P_{\boldsymbol{\beta}}$ denote the probability distribution of $h(Y_1), \ldots, h(Y_n)$, it follows from Hájek and Šidák that for $\boldsymbol{\beta} \in \Omega_n$, $P_{\boldsymbol{\beta}}$ is contiguous to $P_{\mathbf{0}}$ provided f has finite and positive Fisher information.

We can think of the restrictions on the μ_i's in (4.1) as imposing conditions on the design matrix $\mathbf{X}$, on the parameter $\boldsymbol{\beta}$ or on both. Moreover, remembering that we arrived at model (3.1) by dividing through by σ in model (2.1), we see that in the context of model (2.1), (4.1) becomes $\sigma^{-2}\Sigma(\boldsymbol{\beta}\mathbf{x}_i) \le K^2$ and $\sigma^{-1}\max|\boldsymbol{\beta}\mathbf{x}_i| \to 0$. Thus, (4.1) could be interpreted as $\sigma \to \infty$ at the rate determined by (4.1). This is the opposite of the $\sigma \to 0$ case studied by Bickel and Doksum (1981). The $\sigma \to \infty$ case is also the case where transformation models are most useful since transformations can be reliably estimated. See Bickel and Doksum (1981) and Box and Cox (1982). For the purpose of obtaining approximations to bias and MSE and using Monte Carlo procedures to check for which parameter values these approximations are accurate for finite n, it is better to think of (4.1) as imposing restrictions on $\boldsymbol{\beta}$.

Since, for $\beta \in \Omega_n$, P_β is contiguous to P_0, we think of Ω_n as a *local* (near **0**) parameter set. It has been used successfully to obtain theoretical results and useful approximations for power functions. The results of this section in conjunction with the Monte Carlo results of Section 7 show that asymptotic results obtained for $\beta \in \Omega_n$ lead to useful approximations to bias and MSE for moderate n and certain ranges of β's. Borrowing from the testing literature and considering the Monte Carlo results of this paper, we find that an approximate ball park rule is that the Ω_n approximations are good for parameter values where the power of the level 0.05 likelihood ratio test of H_0: $\beta_1 = \cdots = \beta_p$ based on $h(Y_i)$ (assuming h known) has asymptotic power at most 0.95. In particular when $p = 1$, we can write $\mu_i = \beta x_i$ and find that the asymptotic variances computed for Ω_n are very close to the Monte Carlo variances for β in $[-3.6/(\Sigma x_i^2)^{1/2}, 3.6/(\Sigma x_i^2)^{1/2}]$.

We will see that Ω_n is the set where β can be estimated as well as if h were known, i.e., where β is adaptive. Thus Ω_n is a *domain of adaptability* for β.

4.2. *The Hájek–Šidák likelihood approximation.* Let f be absolutely continuous with derivative f'. Define

$$\phi_F(v) = \frac{-f'(v)}{f(v)}, \qquad v \in R,$$

$$a_F(k) = E\big(\phi_F(V^{(k)}\big), \qquad k = 1, \ldots, n.$$

We will assume

$$0 < I(f) = \int_{-\infty}^{\infty} \phi_F^2(v) f(v)\, dv < \infty. \tag{4.2}$$

Next we consider a rank based approximation to the likelihood introduced by Hájek and Šidák ((1967), Chapter 7) to establish the asymptotic sufficiency of ranks. They show, using Le Cam's (1960) contiguity lemmas, that an approximation to the likelihood $\mathscr{L}_{F,h}(\beta)$ is given by

$$\tilde{p}_F(\beta) = (n!)^{-1} \exp\left\{ S_F(\beta) - \tfrac{1}{2} I(f) \sum_{i=1}^{n} (\beta \mathbf{x}_i)^2 \right\},$$

where $S_F(\beta) = \sum_{i=1}^n (\beta \mathbf{x}_i) a_F(R_i)$.

Let $p_F(\beta) = \tilde{p}_F(\beta)/\tilde{p}_F(\mathbf{0}) = n!\tilde{p}_F(\beta)$; then:

THEOREM 4.1 (Hájek and Šidák). *For the model* (3.1), *if* (4.2) *is satisfied, then there exists a sequence* $\{c_n\}$ *depending on the* Y*'s only through the ranks such that* (i) $\operatorname{plim}_{n\to\infty} c_n = 1$, (ii) $c_n \mathscr{L}_{F,h}(\mathbf{0}) p_F(\beta)$ *is a density on* R^n *and*

$$\lim_{n\to\infty} \sup_{\beta \in \Omega_n} E_{P_0} |\mathscr{L}_{F,h}(\beta) - c_n p_F(\beta)| = 0. \tag{iii}$$

Since $L_F(\beta)$ is the closest possible (in the $L_2(P_0)$ sense) rank approximation to $\mathscr{L}_{F,h}(\beta)$, and Theorem 4.1 shows that $p_F(\beta)$ is a local rank approximation to $\mathscr{L}_{F,h}(\beta)$, then we regard $p_F(\beta)$ as a local rank approximation to $L_F(\beta)$.

Let $\mathbf{C} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ and $\mathbf{a} = \mathbf{a}_F = (a_F(r_1), \ldots, a_F(r_n))$; then $p_F(\beta)$ is maximized by

$$\tilde{\beta} = \tilde{\beta}_F = \mathbf{Ca}/I(f).$$

Next it will be shown that one can use the rank statistic theory of Hájek (1962) and Hájek and Šidák (1967) to analyze transformed data in an efficient and adaptive fashion.

We introduce $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}$ and assume

$$\max_{1 \le i \le n} \{\mathbf{x}_i'\mathbf{B}\mathbf{x}_i\} \to 0 \quad \text{as } n \to \infty. \tag{4.3}$$

PROPOSITION 4.1. *If the conditions* (4.2) *and* (4.3) *are satisfied, then for* $\beta \in \Omega_n$, $(\tilde{\beta} - \beta)$ *has asymptotically the* p *variate normal distribution* $N(\mathbf{0}, \mathbf{B}/I(f))$.

PROOF. Let (c_{ji}) denote C'; then we can write $\tilde{\beta}_j = \sum_{i=1}^n c_{ji}a(R_i)$, $j = 1, \ldots, p$. By the Cramér–Wold device, it is enough to show that every linear combination $\sum_{j=1}^p d_j\tilde{\beta}_j$ is asymptotically normal. Let $c_i' = \sum_{j=1}^p d_j c_{ji}$; then,

$$\sum_{j=1}^p d_j\tilde{\beta}_j = \sum_{j=1}^p d_j \sum_{i=1}^n c_{ji}a(R_i) = \sum_{i=1}^n a(R_i) \sum_{j=1}^p d_j c_{ji} = \sum_{i=1}^n c_i'a(R_i).$$

Since $\mathbf{B}$ is positive definite, we can write $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2}$, where $\mathbf{B}^{1/2}$ is a positive definite square symmetric matrix. For $\mathbf{d} = (d_1, \ldots, d_n)$ and $\mathbf{b}_j = j$th row of $\mathbf{B}$ we can write

$$c_i' = \sum_{j=1}^p d_j c_{ji} = \sum_{j=1}^p d_j b_j\mathbf{x}_i = \mathbf{d}'\mathbf{B}\mathbf{x}_i = \mathbf{d}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}_i.$$

It follows from the Cauchy–Schwarz inequality that

$$(c_i')^2 \le (\mathbf{d}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{d})(\mathbf{x}_i'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}_i) = (\mathbf{d}'\mathbf{B}\mathbf{d})(\mathbf{x}_i'\mathbf{B}\mathbf{x}_i).$$

Moreover,

$$\sum_{i=1}^n (c_i')^2 = \sum_{i=1}^n (\mathbf{d}'\mathbf{B}\mathbf{x}_i)^2 = \sum_{i=1}^n (\mathbf{d}'\mathbf{B}\mathbf{x}_i)(\mathbf{x}_i'\mathbf{B}\mathbf{d})$$

$$= \mathbf{d}'\mathbf{B}\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'\right)\mathbf{B}\mathbf{d} = \mathbf{d}'\mathbf{B}(\mathbf{X}'\mathbf{X})\mathbf{B}\mathbf{d} = \mathbf{d}'\mathbf{B}\mathbf{d},$$

$$\frac{\max_i\{c_i'\}^2}{\Sigma(c_i')^2} \le \max_i \{\mathbf{x}_i'\mathbf{B}\mathbf{x}_i\} \to 0 \quad \text{as } n \to \infty$$

It follows from Hájek and Šidák ((1967), page 216) that $\sum_{j=1}^p d_j\tilde{\beta}_j$ is asymptotically normal. □

Note that under the conditions of Proposition 4.1, $\tilde{\beta}$ is asymptotically optimal and adaptive in the sense of having the same asymptotic distribution as the maximum likelihood estimate $\hat{\beta}_h$ with h assumed known.

EXAMPLE 4.1 (Normal errors). When $F = \Phi$, the approximate partial likelihood estimate is $\tilde{\boldsymbol{\beta}} = \mathbf{Ca}$ with $a_\Phi(k)$ equal to the normal scores $E(Z^{(k)})$, $k = 1, \ldots, n$. This is the same as the normal theory MLE based on $h(Y_i)$, h known, except $h(Y_i)$ is replaced by $a_\Phi(r_i)$. In fact, $\tilde{\boldsymbol{\beta}}$ is the normal scores estimate proposed by Fisher and Yates (1938) for experiments where only the ranks of the Y's are available. Now for $\boldsymbol{\beta} \in \Omega_n$, $\tilde{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix $[\mathbf{X}'\mathbf{X}]^{-1}$. In other words, all the asymptotic optimality properties that the least squares estimate enjoy in the normal linear model, $\tilde{\boldsymbol{\beta}}$ has in the model (3.1) with $F = \Phi$.

EXAMPLE 4.2. In the proportional hazard model, $F(x) = 1 - \exp(-e^t)$, the approximate partial likelihood estimate is $\tilde{\boldsymbol{\beta}} = \mathbf{Ca}$ where $a_F(k) = a(k) = \sum_{j=N+1-k}^{N} 1/j$, $k = 1, \ldots, n$, are the familiar exponential or Savage scores. For $\boldsymbol{\beta} \in \Omega_n$, $\tilde{\boldsymbol{\beta}}$ is asymptotically normal $N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1})$ and it is asymptotically optimal in the proportional hazard model with unknown baseline hazard function.

Suppose that the true model has error distribution F_0, but, we (incorrectly) use the estimate $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_F$ corresponding to the error distribution $F \neq F_0$. The asymptotic bias is then $b(F, F_0)\boldsymbol{\beta}$ where

$$b(F, F_0) = 1 - \left[\int_0^1 \phi_F\left(F^{-1}(u)\right)\phi_{F_0}\left(F_0^{-1}(u)\right) du / I(f)\right].$$

Using the same arguments that led to Proposition 4.1, we find

PROPOSITION 4.2. *Under conditions* (4.2) *and* (4.3), *for* $\boldsymbol{\beta} \in \Omega_n$, $[\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} - b(F, F_0)\boldsymbol{\beta}]$ *has asymptotically the* $N(\mathbf{0}, \mathbf{B}/I(f))$ *distribution.*

It follows from Proposition 4.2 that the MSE of $\hat{\beta}_j$ can be approximated as

$$\mathrm{MSE}\left(\tilde{\beta}_j\right) \simeq \left(\sum_{i=1}^{n} c_{ji}^2\right)/I(f) + b^2(F, F_0)\beta_j^2.$$

EXAMPLE 4.3. Suppose we use the normal scores estimate of Example 4.1 when in fact the true distribution is the standardized logistic $F_0(t) = 1/[1 + \exp(-\pi t/\sqrt{3})]$ with variance one; then $F = \Phi$, $I(f) = 1$, and $1 - b(F, F_0) = \pi/\sqrt{3}\int_0^1 \Phi^{-1}(u)(2u - 1)\, du = \sqrt{\pi/3}$. Thus, in transformed simple linear regression, $\mathrm{MSE}(\tilde{\beta}_j) \simeq (\Sigma x_{1i}^2)^{-1} + \left(1 - \sqrt{\pi/3}\right)^2\beta^2$. The approximate squared bias is of the same order as the approximate variance for $\beta \in \Omega_n$. Since $\left(1 - \sqrt{\pi/3}\right)^2 = 0.00054$, the variance will dominate the squared bias for small to moderate sample sizes and $|\beta| \leq 1$.

REMARK 4.1. Pettitt (1982, 1983) considers an approximation to $L_F(\boldsymbol{\beta})$ obtained by using a Taylor expansion about $\boldsymbol{\beta} = \mathbf{0}$. The resulting estimates are different from $\tilde{\boldsymbol{\beta}}$, but asymptotically they are equivalent to $\tilde{\boldsymbol{\beta}}$ and enjoy all the properties of $\tilde{\boldsymbol{\beta}}$ stated in this section.

REMARK 4.2. The results of this section would remain valid if we replaced $a_F(k) = E\phi_F(V^{(k)})$ by $\phi_F(V^{(k)})$. See Bell and Doksum (1964). Similarly, we could have used $\phi_F(F^{-1}(k/(n+1)))$ instead of $a_F(k)$ (Chernoff and Savage (1958); Hájek and Šidák (1967)).

5. Adaptive estimation for unknown F. We first construct adaptive estimates for the case where h is known and F is unknown; then, in the power transformation case, we extend the results to obtain adaptive estimates for symmetric F. For the case where h is known, the adaptive estimates of Dionne (1981), Bickel (1982), Koul and Susarla (1983) and Ritov (1984) apply. Here we construct other adaptive estimates that can be extended to the case of an unknown power transformation. We start by estimating the score function $\phi_0(u) = \phi_F(F^{-1}(u))$.

DEFINITION 5.1. An estimate $\hat{\phi}$ of ϕ_0 is *consistent* if $\int(\hat{\phi}(u) - \phi_0(u))^2\,du$ converges in probability to zero as $n \to \infty$ when $\beta = \mathbf{0}$.

Hájek and Šidák (1967), page 260) give the following consistent estimate of ϕ_F: Let $\{\delta_n\}$ be a sequence of numbers such that $\delta_n \to 0$, $n^{1/4}\delta_n^3 \to \infty$. Put $r_n = n^{3/4}\delta_n^{-2}$, $s_n = n^{1/4}\delta_n^3$ and $t_j = [j_n/(s_n+1)]$, $1 \le j \le s_n$, where [ ] is the greatest integer function. Now for $j = 1, \ldots, s_n$, define

$$\tilde{\phi}_n(u) = \frac{2r_n s_n}{n+1}\left[D(t_j) - D(t_{j+1})\right], \qquad \frac{t_j}{n} \le u \le \frac{t_{j+1}}{n},$$

$$= 0, \quad \text{otherwise,}$$

where $D(t) = [h(Y^{(t+r_n)}) - h(Y^{(t-r_n)})]^{-1}$.

Our adaptive (h known, F unknown) estimate of β is of the form $\beta^* = \mathbf{C}\mathbf{a}^*$, where $(\mathbf{a}^*)' = (a^*(r_i), \ldots, a^*(r_n))\hat{I}^{-1}$, $a^*(k) = \hat{\phi}_n(k/(n+1))$ and $\hat{I}$ is an estimate of $I(f)$. Let $\tilde{I}$ be the estimate of $I(f)$ defined by $\tilde{I} = \int_0^1 \tilde{\phi}_n^2(u)\,du$.

Using the arguments of Hájek and Šidák ((1967), pages 259–266), we find:

PROPOSITION 5.1. *Under assumption* (4.2), $\tilde{\phi}_n$ *is a consistent estimate of* ϕ_0 *and* $\tilde{I}$ *is a consistent estimate of* $I(f)$ *when* $\beta = \mathbf{0}$.

PROPOSITION 5.2. *Assume conditions* (4.2) *and* (4.3). *Suppose that* $\hat{\phi}$ *and* $\hat{I}$ *are two estimates of* ϕ_0 *and* I *that are consistent when* $\beta = \mathbf{0}$ *and suppose that* $\hat{\phi}$ *is independent of the rank vector* $(R_1, \ldots, R_n)$ *when* $\beta = \mathbf{0}$; *then for* $\beta \in \Omega_n$, $\beta^* - \beta$ *is asymptotically normal* $N(\mathbf{0}, B/I(f))$.

PROOF. First use the Cramér–Slutsky theorem to conclude that the asymptotic distribution will be unchanged if in β^*, $\hat{I}$ is replaced by $I(f)$. Next use the arguments of Hájek and Šidák ((1967), Section 7, 1.6).

Since $\tilde{\phi}_n$ and $\tilde{I}$ satisfy the conditions of Proposition 5.2, we have constructed an estimate β^* that is asymptotically optimal when h is known, F is unknown.

Now consider the case where both h and F are unknown and F is symmetric. We then need to estimate $\phi_0(u) = \phi_F(F^{-1}(u))$ on the basis of $Y_1, \ldots, Y_n$ for $\beta = 0$. If we used Hájek and Šidák's estimate $\tilde{\phi}_n$ with $h(y) = y$, we would end up with an estimate of

$$\phi_Y(u) = \phi_G(G^{-1}(u)) = -\frac{g'(G^{-1}(u))}{g(G^{-1}(u))},$$

where G is the distribution of Y. Since $G(y) = P(Y \le y) = P(h(Y) \le h(y)) = F(h(y))$,

$$\phi_0(u) = \left[\phi_Y(u) - \frac{h''(G^{-1}(u))}{h'(G^{-1}(u))}\right]\left[h'(G^{-1}(u))\right]^{-1}.$$

Thus, since we have an estimate of $\phi_Y(u)$, we need estimates of $h'(G^{-1}(u))$ and $J(u) = h''(G^{-1}(u))/[h'(G^{-1}(u))]^2$.

Consider the power transformation

$$h(y) = \frac{y^\lambda - 1}{\lambda}, \qquad \lambda \neq 0,\ h(y) = \log y,\ \lambda = 0. \tag{5.1}$$

In this case $h'(y) = y^{\lambda-1}$, $h''(y)/[h'(y)]^2 = (\lambda - 1)y^{-\lambda}$ and natural estimates of $h'(G^{-1}(u))$ and $J(u)$ are

$$\left[G_n^{-1}(u)\right]^{\hat{\lambda}-1} \text{ and } \hat{J}(u) = (\hat{\lambda} - 1)\left[G_n^{-1}(u)\right]^{-\hat{\lambda}}, \qquad \frac{1}{n} < u < 1 - \frac{1}{n}, \tag{5.2}$$

where G_n is the empirical distribution of the Y's, $G_n^{-1}(u) = \inf\{x\colon G_n(x) \ge u\}$ and $\hat{\lambda}$ is a consistent estimate of λ. When F is symmetric, one convenient estimate is the one defined by Hinkley (1975): Fix $0 < p < \frac{1}{2}$, let $Y(i)$ denote the ith ordered Y and consider the equation

$$\tfrac{1}{2}\left[Y^\lambda([np]) + Y^\lambda([n(1-p)])\right] = \tilde{Y}^\lambda,$$

where $\tilde{Y}$ is the median of the Y's. This equation has either only one solution $\lambda = 0$, or two solutions one of which is $\lambda = 0$. In the first case, set $\hat{\lambda} = 0$, and in the second case set $\hat{\lambda}$ equal to the solution different from zero. Now $\hat{\lambda}$ converges in probability to λ (Hinkley (1975)).

Let $i = [nu]$, and let $U(i) = G(Y(i))$, $i = 1, \ldots, n$, be uniform order statistics; then, when $\beta = 0$,

$$\begin{aligned} G_n^{-1}(u) &= Y(i) = G^{-1}(U(i)) \\ &= G^{-1}\left(\frac{i}{n+1}\right) + \frac{1}{g(G^{-1}(b))}\left(U(i) - \left(\frac{i}{n+1}\right)\right), \end{aligned} \tag{5.3}$$

where

$$\left|b - \frac{i}{n+1}\right| \le \left|U(i) - \frac{i}{n+1}\right|.$$

Now we let $\hat{\phi}_Y$ denote the estimate of ϕ_Y obtained by applying the Hájek–Šidák estimate $\tilde{\phi}_n$ with $h(y) = y$, and we let $\hat{J}$ be as previously defined and set

$$\hat{\phi}_0(u) = \hat{\phi}_Y(u)\big[G_n^{-1}(u)\big]^{1-\hat{\lambda}} + (\hat{\lambda} - 1)\big[G_n^{-1}(u)\big]^{-\hat{\lambda}}.$$

Then since $\hat{\phi}_Y$ is consistent, it follows from (5.3) and the uniform convergence of the uniform empirical process that $\hat{\phi}_0$ is consistent. Moreover, note that $\hat{\phi}_0$ is a function only of the order statistics and therefore is independent of ranks when $\boldsymbol{\beta} = \mathbf{0}$.

Let

$$a^{**}(k) = \hat{\phi}_0(k/(n+1)),\ I^{**} = \int \hat{\phi}_0^2(u)\,du,$$
$$(\mathbf{a}^{**})' = (a^{**}(r_1), \ldots, a^{**}(r_n))(I^{**})^{-1}$$

and $\boldsymbol{\beta}^{**} = \mathbf{C}\mathbf{a}^{**}$; then:

THEOREM 5.1. *If F and G satisfy assumption* (4.2), *if F is symmetric about* 0, *if h is given by* (5.1) *and if condition* (4.3) *is satisfied, then for $\boldsymbol{\beta} \in \Omega_n$, $\boldsymbol{\beta}^{**}$ is multivariate asymptotically normal $N(\boldsymbol{\beta}, \mathbf{B}/I(f))$.*

6. Estimation of a general transformation h. We write

$$h(Y_i) = \mu_i + \varepsilon_i, \qquad \mu_i = \sum_{j=1}^{p} x_{ij}\beta_j,$$

and let G_i denote the distribution of Y_i and F the distribution of ε_i. Note that $G_i(y) = P(Y_i \le y) = P(h(Y_i) \le h(y)) = F(h(y) - \mu_i)$. Thus, we can write $h(y) = F^{-1}G_i(y) + \mu_i$. In our parametrization $\sum_{i=1}^{n}\mu_i = 0$; thus,

$$h(y) = \frac{1}{n}\sum_{i=1}^{n} F^{-1}G_i(y). \tag{6.1}$$

We assume that the μ's are not all zero.

6.1. *Fixed parameters. ANOVA models.* We consider the nonlocal case with β_j and μ_i fixed as sample size increases. The distribution F is assumed to be known. From (6.1) we see that if we can estimate the G_i, then we can estimate h. This can be done in analysis of variance models with several observations per cell. These models can be written as

$$h(Y_{jk}) = \theta_j + \varepsilon_{jk}, \qquad k = 1, \ldots, n_j,\ j = 1, \ldots, p,$$

where θ_j and n_j are the mean and sample size in cell j, respectively. The usefulness of such models has been discussed by Box and Cox (1964) and Cox (1984). Now define $\lambda_{jn} = n_j/n$, let G_j denote the distribution of Y_{jk} and let $\hat{G}_j$ be the empirical distribution function in cell j. Assume that

$$\lim_{n\to\infty} \lambda_{jn} = \lambda_j \quad \text{exists and satisfies } 0 < \lambda_j < 1,\ j = 1, \ldots, p. \tag{6.2}$$

Now we can write $h(y) = \sum_{j=1}^{p}\lambda_{jn}F^{-1}G_j(y)$ and our estimate of $h(y)$ is defined

by

$$\hat{h}(y) = \sum_{j=1}^{p} \lambda_{jn} F^{-1} \hat{G}_j(y).$$

Let $[a, b]$ be any set contained in the support of each G_j, $j = 1, \ldots, p$, and let $c_j = F^{-1}G_j(a)$, $d_j = F^{-1}G_j(b)$. We can now establish weak process convergence on the space $D[a, b]$ of functions on $[a, b]$ that are right continuous and have left-hand limits.

PROPOSITION 6.1. *Suppose that F has a continuous derivative f bounded away from 0 and ∞ on each $[c_j, d_j]$, $j = 1, \ldots, p$, suppose that each G_j is continuous on $[a, b]$ and suppose that (6.2) holds. Then the process $\sqrt{n}[\hat{h}(\cdot) - h(\cdot)]$ converges weakly on $D[a, b]$ to the Gaussian process*

$$\sum_{j=1}^{p} \sqrt{\lambda_j}\, W_j\big(G_j(\cdot)\big)/fF^{-1}G_j(\cdot),$$

where $W_1, \ldots, W_p$ are independent Brownian bridges on $[0, 1]$.

PROOF. Write $u_n = \hat{G}_j(y)$, $u = G_j(y)$ and

$$D_{jn}(y) = \sqrt{n}\left[F^{-1}\hat{G}_j(y) - F^{-1}\big(G_j(y)\big)\right] = \frac{F^{-1}(u_n) - F^{-1}(u)}{u_n - u}\sqrt{n}\,(u_n - u).$$

By the arguments in Doksum ((1974), pages 272–273) D_{jn} converges weakly to $W(G_j)/\sqrt{\lambda_j} fF^{-1}G_j$. The result follows. □

Note that $F^{-1}G_j(y) = h(y) - \mu_j$; thus, $\hat{h}(y)$ is approximately normal with mean $h(y)$ and variance $n^{-1}\sum_{j=1}^{p}\lambda_j G_j(y)[1 - G_j(y)]/f^2(h(y) - \mu_j)$.

6.2. *Local parameter set.* We return to the general transformed linear model with h unknown, F known and $\beta \in \Omega_n$. Now we define

$$\tilde{h}(y) = F^{-1}G_n(y),$$

where $G_n(y) = n^{-1}\#[Y_i \le y] = n^{-1}\#[h(Y_i) \le h(y)] = F_n(h(y), \mu)$ and $F_n(t, \mu) = n^{-1}\sum I[\varepsilon_i \le t - \mu_i]$.

Let $[a', b']$ be a set contained in the support of each G_i for $i = 1, \ldots, n$ and let $[c_i', d_i'] = [F^{-1}G_i(a'), F^{-1}(G_i(b'))]$.

PROPOSITION 6.2. *Suppose that F has a uniformly continuous and bounded density f that is bounded away from 0 on each $[c_i', d_i']$, $i = 1, \ldots, n$, and suppose that each G_i is continuous; then, $\sqrt{n}[\tilde{h}(\cdot) - h(\cdot)]$ converges weakly on $D[a', b']$ to the Gaussian process $W(F(h(\cdot)))/f(h(\cdot))$, where W is a Brownian bridge.*

PROOF. Let $v_n = G_n(y)$, $v = Fh(y)$ and write

$$D_n(y) = \sqrt{n}\left[\tilde{h}(y) - h(y)\right] = \frac{F^{-1}(v_n) - F^{-1}(v)}{v_n - v}\sqrt{n}\,(v_n - v).$$

Let $F_n(t) = F_n(t, \mathbf{0})$ be the empirical d.f. of $\varepsilon_1, \ldots, \varepsilon_n$; then,

$$\sqrt{n}(v_n - v) = \sqrt{n}[F_n(h(y), \mu) - Fh(y)]$$
$$= \sqrt{n}[F_n(h(y)) - F(h(y))] + \sqrt{n}[F_n(h(y), \mu) - F_n(h(y))].$$

The first term converges to $W(F(h(y)))$. The second term has expected value $n^{-1/2}\Sigma_i p_{in}$ and variance $(1/n)\Sigma_i p_{in}(1 - p_{in})$, where

$$p_{in} = |F(h(y) - \mu_i) - F(h(y))|.$$

By a Taylor expansion, $p_{in} = f(h(y_0))\mu_i$ with $|h(y) - h(y_0)| \leq |\mu_i|$. It follows from this that the second term converges uniformly in probability to zero. Finally, note that $[F^{-1}(v_n) - F^{-1}(v)]/(v_n - v)$ converges appropriately to $1/f(h(y))$ as in the proof of Proposition 6.1. □

7. Monte Carlo results.

7.1. *Transformed linear regression.* Consider the transformed regression model with $p = 1$, $n = 15$, $x_i = (i - 8)/7$, and $F = N(0, 1)$. Even though n is small, the Monte Carlo results will be checked against the asymptotic results for Ω_n. According to these asymptotic results and the ball park rule of Section 4.4, the mean squared error (MSE) of the MPLE $\hat{\beta}$ and the normal scores estimate $\tilde{\beta}$ should be approximately $1/\Sigma x_i^2 = 49/280 = 0.175$ for $\beta \in [-3.6/(\Sigma x_i^2)^{1/2}, 3.6/(\Sigma x_i^2)^{1/2}] = [-1.5, 1.5]$. Table 3 where $\hat{\beta}$ is computed using the likelihood sampler with $M = 100$ and where the number of Monte Carlo runs is 500, indicates that in this case the Ω_n asymptotic is roughly in effect even for $n = 15$. Note that we have included the optimal (UMVU) estimate assuming h is known; it is denoted by LS(h) since in this case, it is also the least squares estimate. The Monte Carlo standard errors of the MSE's are given in parentheses below the MSE's.

The MSE's of $\hat{\beta}$ and $\tilde{\beta}$ are rather large for $\beta > 2.8$. In order to check whether this is a weakness of these estimates or a more universal trait of estimates when h is unknown, we also considered the performance of the asymptotically optimal (for all β) MLE (Box–Cox estimate) $\hat{\beta}_{BC}$ for the power transformation model

$$h_\lambda(y) = \frac{\text{sign}(y)|y|^\lambda - 1}{\lambda}, \qquad \lambda \neq 0,\ \lambda_\lambda(y) = \log y,\ \lambda = 0.$$

TABLE 3

Monte Carlo results for regression with $n = 15$, $p = 1$, $x_{i1} = (i - 8)/7$, $F = N(0, 1)$, $M = 100$, *and* 500 *Monte Carlo trials.* $c = 280/49$. *For each* β, *Monte Carlo Bias*$(LS(h)) = 0.000$, *cMSE*$(LS(h)) = 0.95$ (*standard error* $= 0.070$).

β	0	0.35	0.70	1.4	2.1	2.8	3.5	7.0
Bias($\hat{\beta}$)	−0.006	0.007	0.016	−0.076	−0.423	−0.931	−1.533	−4.871
Bias($\tilde{\beta}$)	−0.005	−0.051	−0.117	−0.423	−0.917	−1.513	−2.156	−5.556
cMSE $\hat{\beta}$	1.23	1.30	1.14	0.85	1.41	5.19	13.633	135.74
(std. error)	(0.088)	(0.088)	(0.096)	(0.061)	(0.067)	(0.997)	(0.149)	(0.440)
cMSE $\tilde{\beta}$	0.86	0.85	0.68	1.27	4.90	13.11	26.59	176.44
(std. error)	(0.054)	(0.057)	(0.060)	(0.062)	(0.066)	(0.066)	(0.068)	(0.0060)

TABLE 4

Monte Carlo results for the power transformation model (7.1). $n = 15$, $x_i = (i-1)/7$, $\theta_1 = 1$, $\lambda = 1/2$, $\beta = \theta_2/\sigma$, $F = N(0,1)$ *and* 500 *Monte Carlo trials.* $c = 280/49$. *Bias*$(LS(h)) = -0.006$, *cMSE*$(LS(h)) = 0.94$ (*standard error* $= 0.063$).

σ	β	Bias$(\hat{\beta}_{BC})$	Bias$(\hat{\beta})$	Bias$(\tilde{\beta})$	cMSE$(\hat{\beta}_{BC})$ (std. error)	cMSE$(\hat{\beta})$ (std. error)	cMSE$(\tilde{\beta})$ (std. error)
1	0.7	−0.007	0.016	−0.117	1.79 (0.169)	1.14 (0.080)	0.68 (0.051)
1	7	0.056	−4.87	−5.56	39.18 (3.33)	135.74 (0.440)	176.44 (0.060)
0.1	7	0.119	−4.87	−5.56	133.39 (12.446)	135.74 (0.440)	176.44 (0.060)

We consider the model (see Table 4)

$$h_\lambda(y) = \theta_1 + \theta_2 x_i + \sigma\varepsilon_i, \qquad i = 1, \ldots, n, \tag{7.1}$$

with $n = 15$, $x_i = (i-1)/7$, $\theta_1 = 1$, $\lambda = \frac{1}{2}$ and $F = N(0,1)$.

Let $\hat{\theta}_2$ be the MLE of θ_2; then, because of the reparametrization of Section 3, the appropriate parameter is $\beta = \theta_2/\sigma$ and the appropriate estimate to compare with $\hat{\beta}$ and $\tilde{\beta}$ is $\hat{\beta}_{BC} = \hat{\theta}_2/\sigma$. Note that θ_1, σ and λ are estimated simultaneously with θ_2, but the performance of $\hat{\beta}_{BC}$ is evaluated at $\theta_1 = 1$, $\lambda = \frac{1}{2}$ and $\sigma = 0.1$ and 1. However, σ in $\hat{\beta}_{BC} = \hat{\theta}_2/\sigma$ is not estimated; it is either 0.1 or 1 as indicated in the table.

We see that $\hat{\beta}_{BC}$ has a much smaller bias that $\hat{\beta}$ and $\tilde{\beta}$, but that $\hat{\beta}$ has MSE comparable to those of $\hat{\beta}_{BC}$ except when $\sigma = 1$, $\beta = 7$.

Next, we consider the same model, but with $n = 25$, $x_i = (i-13)/12$. Now $\beta \in \Omega_n$ corresponds to $\beta \in \left[12(-3.6)/\sqrt{1300}\,,\, 12(3.6)/\sqrt{1300}\right] = [-1.2, 1.2]$, and according to the ball park rule, the mean squared errors of $\hat{\beta}$ and $\tilde{\beta}$ should be in the neighborhood of $144/1300$ for β in this interval. Table 5 shows the MSE's of $\hat{\beta}$ and $\tilde{\beta}$ to be at most $(2.53)(144/1300)$ for $\beta \in [-1.8, 1.8]$.

TABLE 5

Monte Carlo results for regression with $n = 25$, $p = 1$, $x_i = (i-13)/12$, $F = N(0,1)$, $M = 100$ *and* 500 *Monte Carlo trials.* $c = 1300/144$. *For each* β, *Monte Carlo Bias*$(LS(h)) = 0.000$, *cMSE*$(LS(h)) = 1.064$ (*standard error* $= 0.067$).

β	0	0.60	1.2	1.8	2.4	3.0	3.6
Bias $\hat{\beta}$	−0.287	−0.118	−0.049	−0.276	−0.634	−1.126	−1.640
Bias $\tilde{\beta}$	−0.006	−0.072	−0.284	−0.649	−1.113	−1.633	−2.184
cMSE$(\hat{\beta})$	2.01	2.53	1.28	1.30	4.24	11.96	24.76
(std. error)	(0.111)	(0.111)	(0.130)	(0.078)	(0.102)	(0.177)	(0.259)
cMSE$(\tilde{\beta})$	0.97	0.95	1.10	2.13	11.26	24.09	43.03
(std. error)	(0.058)	(0.058)	(0.065)	(0.076)	(0.084)	(0.087)	(0.085)

TABLE 6

Two-sample Monte Carlo results, $n_1 = n_2 = 20$, $x_i = \pm \frac{1}{2}$, $M = 100$ *and* 1000 *Monte Carlo trials for all* β *except* $\beta = 2.0$ *and* 2.4, *where* 200 *Monte Carlo trials are used. For each* $\beta \leq 1.6$, *Monte Carlo Bias*$(LS(h)) = 0.001$, $10MSE(LS(h)) = 1.05$ (*standard error* $= 0.047$). *For* $\beta = 2.0$ *and* 2.4, *Monte Carlo Bias*$(LS(h)) = 0.038$, $10MSE(LS(h)) = 1.19$ (*standard error* $= 0.106$).

β	0	0.1	0.5	0.8	1.2	1.6	2.0	2.4
Bias($\hat{\beta}$)	0.001	−0.005	0.012	0.022	−0.012	−0.108	−0.266	−0.535
Bias($\tilde{\beta}$)	0.002	−0.003	−0.036	−0.087	−0.204	−0.395	−0.595	−0.904
10MSE($\hat{\beta}$)	1.14	1.20	1.26	1.27	0.93	0.75	1.11	3.11
(std. error)	(0.052)	(0.053)	(0.052)	(0.052)	(0.038)	(0.035)	(0.086)	(0.117)
10MSE($\tilde{\beta}$)	0.99	1.00	0.98	0.98	0.86	1.83	3.68	8.24
(std. error)	(0.042)	(0.042)	(0.039)	(0.037)	(0.038)	(0.046)	(0.108)	(0.107)

7.2. *The transformed two-sample model.* Here we consider two samples of size $n_1 = 20$, $n_2 = 20$, and $x_i = -\frac{1}{2}$, $i = 1, \ldots, 20$, $x_i = \frac{1}{2}$, $i = 21, \ldots, 40$. Since $\Sigma x_i^2 = n_1 n_2 / n = 10$, we check whether the Ω_n asymptotic is in effect for $\beta \in [-1.14, 1.14]$. Table 6 considers the MPLE $\hat{\beta}$ of Section 3 with $F = N(0, 1)$, and the normal scores (Fisher–Yates) estimate $\tilde{\beta}$ of Section 4 (Example 4.1).

The partial likelihood (rank) estimates do remarkably well. They estimate β almost as well as if h was known for a wide range of β. This is as predicted by the Ω_n asymptotics.

For comparison, we note that the asymptotic MSE (AMSE) of the MLE $\hat{\beta}_{BC}$ for the two-sample power transformation model with fixed β evaluated at $\lambda = 0$ equals the optimal $1/10$ for all β. This follows from Bickel and Doksum ((1981), page 303). For fixed β this AMSE is not known for $\lambda \neq 0$; however, in the Ω_n asymptotics it is $1/10$ for all λ.

Finally, in the spirit of Example 4.3, we checked the performance of the MPLE $\hat{\beta}$ based on the normal likelihood sampler and the normal scores $\tilde{\beta}$ when the true error distribution is logistic. The results in Table 7 show that, as predicted by the Ω_n asymptotics, a logistic error distribution does not diminish the performance of these estimates.

TABLE 7

Two-sample Monte Carlo results, $n_1 = n_2 = 20$, $x_i = \pm \frac{1}{2}$, *estimates derived from normal* F, *logistic error distribution* F_0 *with variance* 1, $M = 100$, 200 *Monte Carlo trials. For each* β, *Monte Carlo Bias*$(LS(h)) = 0.017$, $10MSE(LS(h)) = 0.91$ (*standard error* $= 0.084$).

β	0	0.4	0.8	1.2	1.6	2.0	2.4
Bias($\hat{\beta}$)	0.012	0.041	0.062	0.041	−0.077	−0.284	−0.552
Bias($\tilde{\beta}$)	0.013	0.016	−0.025	−0.144	−0.341	−0.605	3.31
10MSE($\hat{\beta}$)	1.02	1.10	1.13	0.99	0.68	1.22	3.31
(std. error)	(0.102)	(0.117)	(0.106)	(0.098)	(0.071)	(0.098)	(0.134)
10MSE($\tilde{\beta}$)	0.93	0.88	0.69	0.68	1.44	3.82	8.48
(std. error)	(0.089)	(0.085)	(0.065)	(0.068)	(0.094)	(0.116)	(0.125)

Acknowledgments. This paper benefitted from discussions with D. Dabrowska, W. R. van Zwet, S. Dawkins and E. Scott as well as from comments by the referees. In particular, the proof of Proposition 4.1, which replaced an earlier cumbersome proof, is due to one of the referees. I am grateful to P. G. Neville for programming the likelihood sampler and computing its Monte Carlo properties, and to W. C. Navidi and S. Le Roy who computed the Box–Cox Monte Carlo results.

REFERENCES

ANSCOMBE, F. J. and TUKEY, J. W. (1954). The criticism of transformation. Unpublished manuscript.

BEALL, G. (1942). The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika* **32** 243–262.

BEGUN, J. M. (1981). A class of rank estimates of relative risk. Unpublished manuscript.

BELL, C. B. and DOKSUM, K. A. (1965). Some new distribution-free statistics. *Ann. Math. Statist.* **36** 203–214.

BERK, R. and SAVAGE, I. R. (1968). The information in a rank order and the stopping time of some associated SPRT's. *Ann. Math. Statist.* **39** 1661–1674.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BICKEL, P. J. (1984). Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **79** 315–316.

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76** 296–311.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252.

BOX, G. E. P. and COX, D. R. (1982). The analysis of transformations revisited, rebutted. *J. Amer. Statist. Assoc.* **77** 209–210.

BREIMAN, L. and FRIEDMAN, J. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–597.

CARROLL, R. J. (1982). Tests for regression parameters in power transformation models. *Scand. J. Statist.* **9** 217–222.

CARROLL, R. J. and RUPPERT, D. (1981). On prediction and the power transformation family. *Biometrika* **68** 609–615.

CARROLL, R. J. and RUPPERT, D. (1984). Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **79** 312–313.

CHERNOFF, H and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–202.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

COX, D. R. (1984). Interaction. *Internat. Statist. Rev.* **52** 1–32.

DIONNE, L. (1981). Efficient nonparametric estimators of parameters in the general linear hypothesis. *Ann. Statist.* **9** 457–460.

DOKSUM, K. A. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277.

DOKSUM, K. A. (1984). The transformation controversy. Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **79** 316–319.

DOKSUM, K. A. and WONG, C.-H. (1983). Statistical tests based on transformed data. *J. Amer. Statist. Assoc.* **78** 411–417.

DRAPER, N. R. and SMITH, H. (1981). *Applied Regression Analysis*. Wiley, New York.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

FISHER, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed., Example 46.2. Oliver and Boyd, Edinburgh.

FISHER, R. A. and YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver and Boyd, Edinburgh.

HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.
HÁJEK, J. (1974). Asymptotic sufficiency of the ranks in the Bahadur sense. *Ann. Statist.* **2** 75–83.
HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.
HINKLEY, D. V. (1975). On power transformations to symmetry. *Biometrika* **62** 101–112.
HINKLEY, D. and RUNGER, G. (1984). The analysis of transformed data. *J. Amer. Statist. Assoc.* **79** 302–309.
HOCKING, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32** 1–44.
HOEFFDING, W. (1951). "Optimum" nonparametric tests. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 83–92. Univ. California Press.
KALBFLEISCH, J. D. (1978). Likelihood methods and nonparametric tests. *J. Amer. Statist. Assoc.* **73** 167–170.
KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika* **60** 267–278.
KOUL, H. L. and SUSARLA, V. (1983). Adaptive estimation in linear regression. *Statist. Decisions* **1** 379–400.
KRUSKAL, J. B. (1965). Analysis of factorial experiments by estimating monotone transformation of the data. *J. Roy. Statist. Soc. Ser. B* **27** 251–263.
LAI, T. L. (1975). On Chernoff–Savage statistics and sequential rank tests. *Ann. Statist.* **3** 825–845.
LE CAM, L. (1960). Locally asymptotically normal families of distribution. *Univ. Calif. Publ. Statist.* **3** 37–98.
LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43.
LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.
MILLER, A. J., SHAW, D. E., VEITCH, L. G. and SMITH, E. J. (1979). Analysing the results of a cloud seeding experiment in Tasmania. *Comm. Statist. A—Theory Methods* **8** 1017–1047.
PETTITT, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. Roy. Statist. Soc. Ser. B* **44** 234–243.
PETTITT, A. N. (1983). Approximate methods using ranks for regression with censored data. *Biometrika* **70** 121–132.
PETTITT, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *J. Roy. Statist. Soc. Ser. C* **33** 169–175.
RITOV, Y. (1984). Efficient and unbiased estimation in nonparametric linear regression with censored data. Unpublished manuscript.
RUBIN, D. B. (1984). Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **79** 309–312.
SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics—the two-sample case. *Ann. Math. Statist.* **27** 590–615.
SAVAGE, I. R. (1957). Contribution to the theory of rank order statistics—the trend case. *Ann. Math. Statist.* **28** 968–977.
SETHURAMAN, J. (1970). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives, II. *Ann. Math. Statist.* **41** 1322–1333.
TAYLOR, J. (1986). The retransformed mean after a fitted power transformation. *J. Amer. Statist. Assoc.* **81** 114–118.
TERRY, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Ann. Math. Statist.* **23** 346–366.
TUKEY, J. W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.* **28** 602–632.
WOODLEY, W. L., SIMPSON, J., BIONDINI, R. and BERKELEY, J. (1977). Rainfall results 1970–1975: Florida area cumulus experiment. *Science* **195** 735–742.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720