The Annals of Statistics
1996, Vol. 24, No. 6, 2655–2678

ON DISTRIBUTION-FREE INFERENCE FOR RECORD-VALUE DATA WITH TREND[1]

BY ANDREY FEUERVERGER AND PETER HALL

University of Toronto and
University of Toronto and Australian National University

Maximum likelihood estimators for record-value data with a linear trend are quite sensitive to misspecification of the error distribution. Indeed, incorrect choice of that distribution can lead to inconsistent estimation of the intercept parameter and produce estimators of slope that do not enjoy the asymptotic convergence rate prescribed by the information matrix. These properties and the importance of linearly trended records lead us to suggest a distribution-free approach to inference. We show that the slope and intercept parameters and the entire error distribution can be estimated consistently, and that bootstrap methods are available. The latter may be employed to estimate the variance of estimators of slope, intercept and error distributions. The case of trends that increase faster than linearly is also considered, but is shown to be relatively uninteresting in the sense that the natural estimators have rather predictable properties.

1. Introduction. In some data-recording contexts the values that are of greatest interest are the extremes, be they minima or maxima. Sometimes those particular data are recorded in a very accessible form, and more complete, detailed data are available only at the cost of searching more extensively. This can be the case with records in sporting events, for example. In other contexts, such as on-line data recording by machine, where the capacity for data retention may not be great, data that are not very recent and are not record values may be automatically deleted, for all time.

Often, as a result of improvements in technology or technique, record-value data may not be adequately modelled as successive extremes of independent and *identically distributed* sequences. Incorporating a trend is often the simplest way of allowing for a degree of nonstationarity. Smith (1988) argued cogently and persuasively, on the basis of athletic data, that the case where the trend is linear is often of greatest interest. Ballerini and Resnick (1985, 1987) also studied record-value models with linear trends, although in their case the stochastic errors were modelled as stationary time series rather than independent and identically distributed random variables. See too the work of De Haan and Verkade (1987), where it was shown that the large-sample properties of linearly trended data may be similar to those of data without trend, either because of a heavy-tailed error distribution or a particularly slow trend.

Received January 1995; revised February 1996.
[1]Supported in part by grants from the Natural Sciences and Engineering Research Council of Canada.
AMS 1991 *subject classifications*. Primary 60G70, 62G05; secondary 62G09.
Key words and phrases. Bootstrap, dependent data, linear regression, polynomial regression, record values, stationary sequence, time series, trend.

Smith (1988) also examined trends that are expressible as increasing quadratics or other increasing functions. He conducted statistical analysis by fitting models, such as the Normal, to the error distribution and estimating all parameters (those of the trend function and those of the distribution) by maximum likelihood. This approach to analysis is very powerful, not least because of the opportunities that it offers for analyzing a particularly wide range of trends and error types. The method has some unusual properties, however. In particular, it may be shown (see Section 5.6) that the parameter estimators which arise out of fitting in this way may not all be consistent if the error distribution is chosen incorrectly. For example, if the trend is linear and if the errors are assumed to be normally distributed but actually have another distribution (such as a Gumbel or generalized extreme-value distribution, two of the other error types studied by Smith), then the estimator of slope in the fitted trend will be consistent, but the estimator of intercept will not. This property is in stark contrast to the more traditional behavior that is observed in model fitting under the assumption of specific error distributions, where the choice of error distribution typically affects only second-order asymptotic properties of estimators of trend parameters. In such classical cases, misspecification of the error distribution does not even influence the weak limit of the estimators; it certainly does not interfere with consistency properties.

Furthermore, in the case of a linear trend, while the estimator of slope will be consistent, it will not necessarily converge at the rate predicted by the information matrix under a model for the error distribution, if that model is incorrect. Thus, the asymptotic properties of estimators of both intercept and slope can be rather different, should the model be incorrect, from those for which one might hope. This leads to obvious difficulties in confidence procedures based on record values with linear trend. By way of contrast, the problem does not arise in cases where the trend increases faster than linear. There, convergence rates of estimators and also their limiting distributions are typically unaffected by correctness of the choice of model for the error distribution.

While elementary (although tedious) to derive, these properties are so strikingly different from those that would usually be expected that they argue persuasively for the development of a distribution-free theory of inference for record-value data with trend. That is the approach adopted in the present paper. We take a general viewpoint, considering models where the trend is faster than linear (such as quadratic regression) as well as those where it is only linear. Nevertheless, since the faster-than-linear context is relatively uninteresting from the viewpoint of asymptotic theory, we do not emphasize this case and instead devote almost all our attention to linear trends.

In that setting we develop distribution-free methods for estimating the gradient and the intercept of the trend, the asymptotic variances of these estimators and the entire error distribution. Our approach to estimation of the latter employs a novel conditioning argument and produces root-n consistency. (Here, n denotes sample size.) That fact gives rise to exciting opportunities for using resampling methods in this highly nonstandard problem. Our estimator of slope is based on least-squares fitting and, like our estimator of intercept, enjoys the same convergence rate as estimators obtained by maximum likelihood methods under correctly specified models. In the case of slope, that rate is $n^{-3/2}$, owing to the high degree of leverage offered by a linear trend. The techniques that we suggest employ only record values, since we have found it quite awkward to use information about record times in a truly distribution-free approach.

Like the properties of maximum likelihood estimators discussed earlier, our general results divide sharply between the cases of linear trends and trends which increase more rapidly than linear. In both, bootstrap methods may be used to estimate the variance of our estimators. In the case of a linear trend, the variance of our root-n consistent estimator of intercept depends directly on the density of the sampling distribution and may be estimated consistently but not with convergence rate $n^{-1/2}$.

There is an extensive applied probability literature concerning the theory of records from random samples. It includes particularly the work of Shorrock (1972–1975) and Resnick [(1973a, b, c), (1975), (1987), Chapter 4]. Smith (1988) gives an excellent introduction to work on statistical inference for record values, particularly in the context of athletic events. See also Chatterjee and Chatterjee (1982), Tryfos and Blackmore (1985) and Berred (1982). The paper by Miller and Halpern (1982) also is of interest in the present context.

Section 2 develops our methodology, with particular reference to the case of a linear trend. Numerical properties of our techniques are explored in Section 3, and their theoretical foundation is laid down in Section 4. All technical arguments are placed together in Section 5. They include an outline of the theory behind our earlier assertions about maximum likelihood estimation.

2. Methodology.

2.1. *Summary.* Section 2.2 develops methods for inference under general models for the trend, giving particular emphasis to the more interesting case where the trend is linear. The latter context is taken up in detail in Section 2.3. Bootstrap methods are developed in Section 2.4. Section 2.5 treats the special case of a symmetric error distribution. There, nonparametric estimators of the lower tail of the error distribution, which are generally inferior in quality to those of the upper tail, may be replaced by estimators in the context of the latter.

Assume that the underlying data, of which only the record values and record times are available, are generated by the model

$$Y_i = t(i, \beta) + X_i, \qquad i \geq 1, \tag{2.1}$$

where $X_1, X_2, \ldots$ are independent random variables with an identical distribution F and the trend in the mean, $t(i, \beta)$, depends only on i and the estimable parameter β, most likely a vector. It is convenient to prevent confounding in the definition of location by asking that $E(X_i) = 0$. The value of the record at time i is

$$Z_i = \max(Y_1, \ldots, Y_i), \qquad i \geq 1, \tag{2.2}$$

and this stochastic process is observed.

2.2. *General trend functions.* If the trend function t increases sufficiently rapidly (which condition should be interpreted relative to the rate of decrease of the upper tail of F), then, as we shall show in Section 4, the process $e_i \equiv Z_i - t(i, \beta)$ is asymptotically stationary, for large values of i, in the following sense. There exists a stationary process $\{\varepsilon_j,\ j \geq 1\}$, such that as $i \to \infty$ the joint distribution of $(e_{i+1}, \ldots, e_{i+n})$ converges to that of $(\varepsilon_1, \ldots, \varepsilon_n)$, for any $n \geq 1$. This means that, except for determination of the location or intercept term in the definition of t, inference about the trend function based on data $\{Z_i\}$ is not dissimilar to that about a regression "mean" in a problem with time series errors, such as

$$Y'_i = t(i, \beta) + \varepsilon_i, \qquad i \geq 1.$$

[The fact that $E(\varepsilon_i)$ is nonzero, so that $t(i, \beta)$ is not equal to the mean of Y'_i, is the root of problems in estimating intercept in a linear trend.] In principle we can apply the bootstrap to enhance asymptotic methods in this regression approach to the problem. There are at least two different ways of using the bootstrap. Either we could employ the block bootstrap, which is problematical unless the data are extensive, or we could try to simulate the time series $\{\varepsilon_i\}$, or even the more basic nonstationary process $\{e_i\}$, in a structural manner. We shall call this the "structural bootstrap" for dependent data. Of course, we know how the process was generated, in terms of maxima of independent and identically distributed disturbances plus an estimable trend, but in general we do not have direct access to the distribution F of those disturbances. If the trend function t increases relatively rapidly (faster than linear) and if the upper tail of F is sufficiently light, then the ε_i's are independent and identically distributed and $Z_i = Y_i$ with relatively high probability. Here, application of the structural bootstrap is relatively straightforward, being similar to that of the ordinary bootstrap in the case of regression with independent errors. However, this context is somewhat uninteresting, at least from a theoretical viewpoint, since the unusual character of record values disappears when they occur as frequently as they do in the case of a rapidly increasing trend.

The case where $t(i, \beta)$ is linear in i, say $t(i, \beta) = a + bi$, where $\beta = (a, b)$, is also of more practical value; see Smith (1988). For a linear trend the following alternative approach will form the basis for our structural bootstrap algorithm. Note that the probability $F_i(x)$ of the event $Z_i - (a + bi) \leq x$, conditional on $Z_i - \{a + b(i+1)\} \leq x$, is an asymptotic (for large i) approximation to $F(x)$. To appreciate why, note that

$$P\{Z_i - (a + bi) \leq x\} = \prod_{j=1}^{i} F\{x + b(i - j)\}$$

and so $F_i(x) = F(x)/F(x + bi)$. The probability in the denominator here converges to 1 as i increases and hence $F_i \to F$ as $i \to \infty$. Of course, the accuracy of the approximation depends on factors such as the size of b (it is more accurate for larger b) and the weight of the upper tails of the distribution F (it is more accurate for lighter tails).

Quite generally, for distributions of X_i with sufficiently light upper tails, the unknown parameter β in the trend function t may be estimated consistently by least squares, at least up to the value of an intercept. In particular, if $\nu \geq 2$, $\beta = (\beta_1, \ldots, \beta_\nu)$ and

$$t(i, \beta) = \sum_{j=1}^{\nu} \beta_j i^{j-1}, \tag{2.3}$$

then $\beta_2, \ldots, \beta_\nu$, but not necessarily β_1, may be estimated by minimizing

$$S(\beta) = \sum_{i=1}^{n} \{Z_i - t(i, \beta)\}^2 w(i) \tag{2.4}$$

for suitable weights $w(i)$. Except for the case $\nu = 2$, the estimator of β_1 is also consistent. Moreover, if $\nu \geq 3$, then the in-probability convergence rate of *any* estimator $\hat{\beta}_k$ of any component β_k of β is identical in the cases where the data used to construct $\hat{\beta}_k$ are the "ideal" values $Y_1, \ldots, Y_n$ and where they are the record values $Z_1, \ldots, Z_n$. Details will be given in Section 4.

2.3. *Linear trend.* Assume that $\beta = (a, b)$ and $t(i, \beta) = a + bi$. To ensure identifiability of the parameter a we ask that $E(X) = 0$. Let us begin by treating the problem as one of linear regression and define

$$\hat{b} = \left\{\sum_{i=1}^{n} Z_i(i - \bar{i})\right\} \Big/ \left\{\sum_{i=1}^{n} (i - \bar{i})^2\right\}, \qquad \hat{a} = \bar{Z} - \hat{b}\bar{i}, \tag{2.5}$$

where $\bar{Z} = n^{-1} \sum_{i=1}^{n} Z_i$ and $\bar{i} = \frac{1}{2}(n + 1)$. We shall show in Section 4 that if the tails of F are sufficiently light, then $\hat{a}$ and $\hat{b}$ both enjoy asymptotic Normal distributions, but only the latter estimator is consistent. The asymptotic variances are, respectively, $4n^{-1}\sigma^2$ and $12n^{-3}\sigma^2$, where

$$\sigma^2 = \sum_{-\infty < i < \infty} \operatorname{cov}(\varepsilon_0, \varepsilon_i) = \operatorname{var}(\varepsilon_0) + 2\sum_{i=1}^{\infty} \operatorname{cov}(\varepsilon_0, \varepsilon_i) \tag{2.6}$$

and, as noted in Section 2.2, the stochastic process $\{\varepsilon_j\}$ may be regarded as the limit of the process $\{e_{i+j}\}$ as $i \to \infty$.

These properties motivate the problems of estimating σ^2 and a consistently, which we consider next. The solutions are relatively straightforward and elegant if the errors X_i have a bounded distribution. There, if $E(X) = 0$ and $P(|X| < x_0) = 1$, then the process $\{\varepsilon_j\}$ is m dependent, where $m - 1$ denotes the integer part of $2x_0/b$. Then σ^2 may be estimated as a finite series. Indeed, in view of (2.6) we may write $\sigma^2 = \gamma(0) + 2\sum_{i=1}^{m} \gamma(i)$, where $\gamma(i) = \mathrm{cov}(\varepsilon_0, \varepsilon_i)$. A root-$n$ consistent estimator of $\gamma(i)$ is

$$\hat{\gamma}(i) = (n-i)^{-1} \sum_{j=1}^{n-i} \{Z_{j+i} - \hat{b}(j+i)\}(Z_j - \hat{b}j) - (\bar{Z} - \hat{b}\bar{i})^2$$

and so a root-n consistent estimator of σ^2 is

$$\hat{\sigma}^2 = \hat{\gamma}(0) + 2\sum_{i=1}^{\widehat{m}} \hat{\gamma}(i),$$

where $\widehat{m}$ is an empirical approximation (generally somewhat ad hoc and an overestimate) to m. Methodology and theory for estimating σ^2 may be developed without the assumption of a finite error distribution. It could based, for example, on estimating the spectrum of the process $\{\varepsilon_i\}$, although that requires choice of smoothing parameter to be addressed. Alternatively, bootstrap methods may be employed; see Section 2.4.

Consistent estimation of a may be regarded as a special case of estimation of the moments of $X + a$, which in turn may be treated as an application of estimation of the distribution G of $X + a$. We shall address the latter problem first. A consistent estimator of $G(x)$ is

$$\widehat{G}(x) = \left\{\sum_{i=1}^{n} I(Z_i - \hat{b}i \le x)\right\} \Big/ \left[\sum_{i=1}^{n} I\{Z_i - \hat{b}(i+1) \le x\}\right], \tag{2.7}$$

where $\hat{b}$ is defined as in (2.5). Bias can be somewhat reduced, at the expense of an increase in variance, by taking the series in (2.7) over only larger values of i. This does not have any effect on first-order performance of the estimators, however. By convention we take the ratio on the right-hand side of (2.7) to be zero if the denominator vanishes.

The rth moment of $X + a$, for positive integers r, is given by

$$\mu_r = E\{(X+a)^r\} = r\int_0^\infty x^{r-1}\{1 - G(x) + (-1)^r G(-x)\}\,dx,$$

of which an estimator is

$$\hat{\mu}_r = r\int_0^\infty x^{r-1}\{1 - \widehat{G}(x) + (-1)^r \widehat{G}(-x)\}\,dx. \tag{2.8}$$

Since, assuming that $E(X) = 0$, we have $\mu_1 = a$, then a consistent estimator of a is

$$\tilde{a} = \hat{\mu}_1 = \int_0^\infty \{1 - \widehat{G}(x) - \widehat{G}(-x)\}\,dx. \tag{2.9}$$

We shall show in Section 4 that under appropriate regularity conditions, $\widehat{G}$ and $\hat{\mu}_r$ are root-n consistent for G and μ_r, respectively. Their asymptotic variances depend on $h = H'$, the density of the distribution H of $\varepsilon_i + a$, and may be estimated either directly, by kernel methods applied to the data $\{Z_i\}$, or by using the bootstrap.

A predictor of the mean $m_i = a + bi$ of Y_i, perhaps for $i > n$, is $\widehat{m}_i = \tilde{a} + \hat{b}i$. This estimator is root-n consistent, provided that $i = O(n)$ as $n \to \infty$.

2.4. *Bootstrap.* There are at least two approaches to the bootstrap, appropriate in cases where the trend increases at a polynomial rate faster than linear and where it is linear, respectively. The former is simpler, since there (for samples of sufficiently large size) the Z_i's are equal to the Y_i's with high probability, as noted in Section 2.2. Hence, one may for many intents and purposes consider the problem to be one of regression with independent errors. The more important and more interesting case is that where the trend is linear; then we suggest resampling via the estimator $\widehat{G}$ of the distribution of $X + a$. Specifically, calculate $\hat{b}$ as in (2.5), calculate $\widehat{G}$ as in (2.9), put $\tilde{G}(x) = \inf_{y \geq x} \widehat{G}(y)$ in order to overcome any failure of monotonicity of $\widehat{G}$ (this can be a problem in the lower tail), let $(X_i + a)^*, \dots, (X_n + a)^*$ be derived by resampling independently and at random from the distribution with distribution function $\tilde{G}$, conditional on the original data $\{Z_i\}$, and define

$$Y_i^* = (X_i + a)^* + \hat{b}i \quad \text{and} \quad Z_i^* = \max_{1 \leq j \leq i} Y_j^* \quad \text{for} \quad 1 \leq i \leq n.$$

Bootstrap inference may be based on the Z_i^*'s. In particular, the variances of $\tilde{a}$, $\hat{b}$ and $\widehat{m}_i$ may be estimated in the obvious way. Note that the estimators of $\text{var}(\tilde{a})$ and $\text{var}(\widehat{m}_i)$ will not be root-n consistent, since the asymptotic variances depend on h, for which root-n consistent estimators do not exist (unless a parametric model is available for H). However the estimators of variance are consistent if h is sufficiently smooth, as we shall show in Section 4.

Bootstrap methods may also be used in the obvious way to construct confidence intervals for $E(Y_i) = a + bi$ and prediction intervals for future values of Y_i or Z_i. (We suggest the percentile method, possibly calibrated by application of the double bootstrap, since the slow convergence rate of estimators of $\text{var}\, \tilde{a}$ makes percentile-t problematical here.) Assuming that $i = o(n^{3/2})$ and also the regularity conditions introduced in Section 4, these intervals have asymptotically correct coverage.

It should be noted that likelihood-based confidence intervals for $E(Z_i)$, founded on a model for the distribution of X, are generally asymptotically correct [provided $i = o(n^{3/2})$], even if the model is misspecified. However, confidence intervals for $E(Y_i)$ derived by similar means are typically asymptotically in error.

2.5. *Symmetric error distribution.* One would expect there to be difficulties with any distribution-free procedure applied to record-value data if either the sample size or the value of b was too small. In such cases the manner in which data are recorded will result in there being insufficient information about the lower tail of the distribution of X, with consequent inaccuracies in estimation of a. However, if we suppose that X has a symmetric distribution, then we may replace the nonparametric estimator of the lower tail by that of the upper tail, which is of course considerably more accurate.

This procedure is very simple to implement in practice. Under the hypothesis of symmetry, take the median of the distribution estimator $\widehat{G}$ as the estimator of a; call this $\check{a}$. Put $\check{G}(x) = \tilde{G}(x)$ if $x > \check{a}$ and $\check{G}(x) = 1 - \tilde{G}(2\check{a} - x)$ otherwise. Then implement the bootstrap procedure described in Section 2.4 with $\check{G}$ replacing $\tilde{G}$.

In principle it is possible, under the assumption of symmetry, to use information from both tails to estimate either one. In practice, however, there is generally so much less information in the lower tail compared with the upper that it does not seem worthwhile to attempt empirical calculation of the weights with which these should be combined.

3. Numerical properties. The procedures discussed in Section 2 were implemented on an SGI Challenge computer using the S statistical software package; see Becker, Chambers and Wilks (1988). In particular, we implemented the least-squares estimators $\hat{b}, \hat{a}$ as defined in (2.5) for the slope and intercept parameters of the model for record values data under the linear trend $t(i, \beta) = a + bi$. We also implemented several variants of the intercept estimator $\tilde{a}$ given in (2.9), which is based on the nonparametric estimator (2.7) for the distribution function of $X + a$. These estimators were studied in detail for a variety of distributions for the error terms (the X's) and for a variety of values for the parameters n, a and b using Monte Carlo methods. We also conducted Monte Carlo trials to study the efficacy of bootstrap methods for estimating the variances of these estimators. Finally, we applied our methods to carry out a detailed analysis of data for the (yearly) world record times in the men's one mile run.

Tables 1 and 2 provide a selected summary of our Monte Carlo results for the performance of the estimators $\hat{b}$ and $\tilde{a}$ for samples of size $n = 100, 200, 300$ and values for slope of $b = 0.25, 0.5, 0.75$. In Table 1 the underlying distribution of the error terms $(X_1, \dots, X_n)$ was taken to be $N(0, 1)$, while in Table 2 the underlying distribution of the error terms was taken to be the standardized uniform distribution $12^{1/2}U[-0.5, +0.5]$. [The uniform distribution is used in the record-values context, for example, in Tryfos and Blackmore (1985).] Due to the invariance of our procedures in that parameter, the true value of the intercept a was set equal to 0 in all trials reported here. For each combination of sample size n and slope b we generated 200 data sets, and for each data set we computed, in turn, the corresponding sequences $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$, and finally $(Z_1, \dots, Z_N)$ in accordance with the model given in (2.1) and (2.2) using the linear trend function $t(i, \beta) = a + bi$. For each set of values for n and b, Tables 1 and 2 provide the mean, standard deviation and root mean square error of the 200 computed values for each of the three estimators. As expected—and readily apparent from the tables—the slope parameter b is estimated with exceptional accuracy, in full accordance with the asymptotic theory.

Likewise (but not included in the tables) the least-squares estimator $\hat{a}$ was found to be subject to very severe bias which does not diminish as the sample size increases, although the bias does decrease slowly as increasing values of

TABLE 1
Summary of Monte Carlo trials for estimation of the slope and intercept of the linear trend $t(i, \beta) = a + bi$ for records data with $N(0, 1)$ distributed X's

n	**b**	**Est.**	**Mean**	**Std. dev.**	**RMSE**	**Bootstrap**
100	0.25	$\hat{b}$	0.2508	0.0052	0.0053	0.0045
		$\tilde{a}$	0.136	0.332	0.358	0.29
	0.50	$\hat{b}$	0.5009	0.0038	0.0039	0.004
		$\tilde{a}$	0.058	0.250	0.257	0.25
	0.75	$\hat{b}$	0.7498	0.0043	0.0043	0.004
		$\tilde{a}$	0.067	0.261	0.270	0.23
200	0.25	$\hat{b}$	0.2503	0.0018	0.0018	0.0018
		$\tilde{a}$	0.119	0.253	0.279	0.20
	0.50	$\hat{b}$	0.5002	0.0017	0.0017	0.0015
		$\tilde{a}$	0.044	0.198	0.203	0.18
	0.75	$\hat{b}$	0.7503	0.0013	0.0013	0.0015
		$\tilde{a}$	0.008	0.158	0.158	0.18
300	0.25	$\hat{b}$	0.2501	0.0010	0.0010	0.0010
		$\tilde{a}$	0.102	0.229	0.251	0.18
	0.50	$\hat{b}$	0.4999	0.0008	0.0008	0.0009
		$\tilde{a}$	0.072	0.150	0.167	0.16
	0.75	$\hat{b}$	0.7501	0.0009	0.0009	0.0008
		$\tilde{a}$	0.025	0.156	0.158	0.13

TABLE 2
Summary of Monte Carlo trials for estimation of the slope and intercept of the linear trend $t(i, \beta) = a + bi$ for records data with standard uniform $12^{1/2}U[-0.5, 0.5]$ distributed X's

n	**b**	**Est.**	**Mean**	**Std. dev.**	**RMSE**	**Bootstrap**
100	0.25	$\hat{b}$	0.2507	0.0033	0.0034	0.0037
		$\tilde{a}$	0.163	0.282	0.325	0.25
	0.50	$\hat{b}$	0.5003	0.0038	0.0038	0.0037
		$\tilde{a}$	0.062	0.238	0.246	0.23
	0.75	$\hat{b}$	0.7502	0.0037	0.0037	0.0034
		$\tilde{a}$	0.014	0.219	0.219	0.20
200	0.25	$\hat{b}$	0.2502	0.0013	0.0013	0.0012
		$\tilde{a}$	0.147	0.209	0.256	0.17
	0.50	$\hat{b}$	0.5000	0.0013	0.0013	0.0013
		$\tilde{a}$	0.055	0.163	0.172	0.16
	0.75	$\hat{b}$	0.7502	0.0014	0.0014	0.0013
		$\tilde{a}$	0.024	0.161	0.163	0.16
300	0.25	$\hat{b}$	0.2500	0.0007	0.0007	0.0007
		$\tilde{a}$	0.139	0.184	0.230	0.16
	0.50	$\hat{b}$	0.5000	0.0008	0.0008	0.0007
		$\tilde{a}$	0.055	0.148	0.158	0.12
	0.75	$\hat{b}$	0.7500	0.0007	0.0007	0.0007
		$\tilde{a}$	0.032	0.134	0.138	0.13

b result in a larger percentage of record values. It is worth remarking here that maximum likelihood estimation under a misspecified error distribution would result in essentially similar bias characteristics for the estimator of the intercept; see Section 5.6.

The bias of the estimator $\tilde{a}$ is very considerably less than that of $\hat{a}$ and decreases (quickly) with b and (less quickly) with increasing sample size. The variance of $\tilde{a}$ was found generally to be slightly higher than that of $\hat{a}$, but this is more than amply compensated by the large reduction in bias, resulting in a mean-squared error substantially below that of the least-squares estimator.

Extensive numerical experiments were conducted to explore the behavior of the estimator $\tilde{a}$. Examination of many plots of the estimators $\hat{G}$ and $\tilde{G}$ for the cdf of $X+a$ revealed that estimation of G was very accurate above the median for typical values of b and, to a lesser extent, somewhat below the median as well for the larger values of b. However, the estimates for G below the median deteriorated very rapidly, as would be expected. For this reason, the positive and negative components of the expectation estimator defined in (2.9) were examined separately. This was done in conjunction with two particular options: namely, the choice of whether to use the raw (nonmonotone) $\hat{G}$ as defined in (2.7) or to use the version $\tilde{G}$ that has been rendered monotone; second, the choice of whether or not to drop lower values of the index i from the sums appearing in (2.7). In using the monotone version $\tilde{G}$, we found quite generally (and as expected) that the source of bias resulting in $\tilde{a}$ originated almost entirely from within its negative component. When the non-monotone version $\hat{G}$ was used, this bias was for all intents and purposes eliminated, but at the cost of increased variance (especially in the negative component of $\tilde{a}$) resulting overall in a deterioration in terms of mean-squared error (or at best no improvement). For this reason, all results reported here are in terms of the monotone version $\tilde{G}$. Our findings concerning the dropping of lower values in i from the sums in (2.7) were likewise problematical, generally showing little potential for gain, with the increases in variance being accompanied by only modest reductions in bias. For this reason as well, all results reported here involve untruncated data, that is, no dropping of initial values. Finally, in the case where the distribution of X is assumed to be symmetric, the value of a may alternately be estimated as the median of the distribution function $\tilde{G}$. It turned out, however, that this estimator did not in general outperform $\tilde{a}$. [This appears to be due to deterioration in the quality of estimate of $G(x)$ as x decreases.] In all the above cases, the quality of estimation depends substantially upon the number of record values attained. We note here that for the values $b = 0.25$, 0.5 and 0.75 the percentages of record values observed were approximately 33, 50 and 64% for $N(0,1)$ errors, and very similar for the standard uniform.

The final columns in both Tables 1 and 2 summarize the results of some bootstrap experiments for estimation of the variances of the estimators $\hat{b}$ and $\tilde{a}$. From each combination of n and b reported above, 10 samples were selected and 100 bootstrap subsamples were taken from each of these by drawing $(X_i+a)^*$ values from the distribution $\tilde{G}$ that was computed for each of the samples. The variances of the estimators from each of the bootstrap samples were computed and then averaged across the 10 trials and are reported as the standard deviations shown in the final columns. The individual bootstrap-estimated variances were found to vary over only a very narrow range across the 10 trials in every case that we examined, and, as may be seen in the tables, they provide excellent estimates for the sampling variability of the estimates involved.

Finally, in Figure 1 we provide a summary of an analysis of the world record times for the men's one mile run using the methods that have been presented in this paper. Our starting point was a data set of the 121 yearly best times for the years 1860–1980 which was kindly supplied to us by Sidney Resnick. With the value for 1860 included, this data set comprises 34 record values, and these 34 values (together with the years in which they occurred) are the data on which our analysis was based. Although there is some evidence of slight nonlinearities in the trend function of this series and some evidence of a modest decline in variability over the years, our analysis was based on the full data set mentioned. The line drawn in the figure is based on the least-squares estimator $\hat{b}$ for the slope and the estimator $\tilde{a}$ of (2.9) for the intercept computed without dropping any initial values and using the monotone distribution function $\tilde{G}$. (Actually, the computations were first carried out us-

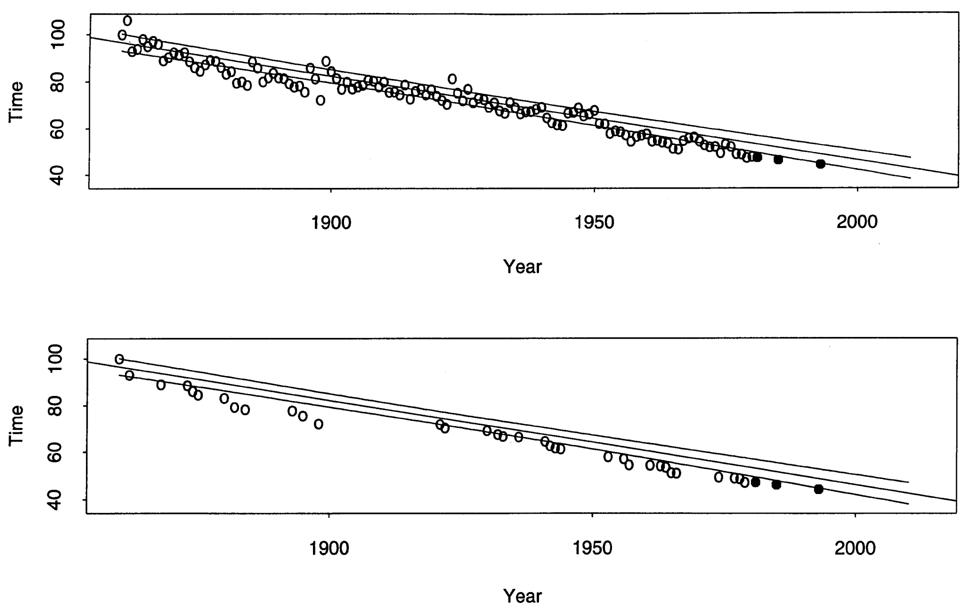

FIG. 1. *Linear trend function fitted to the 34 (yearly) record values for men's one mile run, 1860–1980. The quadratic curves are bootstrap-determined (pointwise) confidence bands for the trend function. The bottom graph shows only the (yearly) record values on which the linear fit is based. The top graph is identical but shows all yearly best times. The vertical scale is actual run time (in seconds) less 3 min. Filled dots are subsequent (yearly) records, 1981–1994.*

ing linearly transformed data, mainly so as to render the trend positive, and then transformed back for presentation in our figure.) We have also drawn the upper and lower "2-sigma" confidence bands for the value of the linear trend function "$a+bi$" up to the year 2010. These quadratic confidence bands are based on 250 bootstrap resamples and were computed in the usual way, based on the bootstrap-estimated sampling variances (and covariance) for $\tilde{a}$ and $\hat{b}$. Finally, with the kind assistance of Cecil Smith of the Ontario Track and Field Association, the three additional (yearly) record values that have occurred between 1981 and December 1994 were determined. (These occurred in 1981, 1985 and 1993.) These values are plotted in distinct (filled) points in the figure and are all seen to lie just below the lower quadratic band. This location is very reasonable for these records, considering that the straight line shown is intended to estimate the mean for both the record and the nonrecord (yearly) values combined.

4. Theoretical justification. Our first result describes the asymptotic stationarity of the process $\{e_i\}$, introduced in Section 2.2. To simplify notation we suppress the parameter β. Assume that $t(i)$ is nondecreasing in i, that for each $j \geq 0$ there exists $s(j) \in [0, \infty]$ such that $t(i+j)-t(i) \to s(j)$ as $i \to \infty$, that for constants $C, \xi > 0$ and all integers i and $j \geq 1$, $t(i+j)-t(i) \geq Cj^{\xi}$ and that $E\{\max(X,0)^{1/\xi}\} < \infty$, where X has the distribution of an arbitrary X_i. We collectively call these conditions (C_1). Note particularly that the value $s(j) = \infty$ is allowed and in fact will occur for each $j \neq 0$ if t is a nonlinear polynomial trend.

Without loss of generality the sequence $\{X_i\}$ is doubly infinite. In this notation, let $\varepsilon_1, \varepsilon_2, \ldots$ denote random variables with joint distribution given by

$$P(\varepsilon_1 \leq x_1, \ldots, \varepsilon_n \leq x_n) = P\Big[\max_{0 \leq j < \infty}\{X_{k-j} - s(j)\} \leq x_k,\ 1 \leq k \leq n\Big].$$

It may be proved that under the assumption that for constants $C, \xi > 0$, $s(j) \geq Cj^{\xi}$ and $E\{\max(X,0)^{1/\xi}\} < \infty$ [both implied by conditions (C_1)], the quantity $\max_{0 \leq j < \infty}\{X_j - s(j)\}$ is a.s. finite. Therefore, the stochastic process $\{\varepsilon_j\}$ is well defined.

THEOREM 4.1. *Under conditions* (C_1) *the finite-dimensional joint distributions of the stochastic process* $\{e_{i+j},\ j \geq 1\}$ *converge to those of* $\{\varepsilon_j,\ j \geq 1\}$ *as* $i \to \infty$.

The simplest examples of the application of Theorem 4.1 are furnished by the case where t is a polynomial. If the polynomial is of precise degree p, then we may take $\xi = p$. If $p \geq 2$, then the function s is degenerate, satisfying $s(j) = \infty$ for $j \geq 1$ and $s(0) = 0$. This means that the process $\{\varepsilon_j\}$ has the same distribution as $\{X_j\}$ and, in particular, is a sequence of independent random variables. However, when $p = 1$, the process $\{\varepsilon_j\}$ is genuinely dependent.

Next we show that if the trend function $t(\cdot, \beta)$ increases sufficiently rapidly (in essence, faster than linear) and if the upper tail of the sampling distribution is sufficiently light, then the rate of convergence of any estimator of β under the classical regression model, with independent and identically distributed errors, is preserved when the estimator is applied instead to record-value data. In order to state this result, assume that $t(i, \beta)$ is nondecreasing in i, that there exist constants $C, \eta > 0$ such that $t(i, \beta^0) - t(j, \beta^0) \ge Ci^{\eta}$ for all $1 \le j \le i-1$, where $\beta^0 = (\beta_1^0, \dots, \beta_\nu^0)$ denotes the true value of β; that $E\{\max(X, 0)^{(1/\eta)+1}\} < \infty$ and that the distribution of X is continuous with support either $(-\infty, \infty)$ or $[x_0, \infty)$ for some $x_0 > -\infty$. Collectively we call these conditions (C_2). The restriction on the support of X is imposed merely to exclude highly pathological cases that could not conceivably arise in practice. (For details, see the proof of Theorem 4.2 in Section 5.) This restriction may be removed at the expense of a very mild but rather unattractive condition on the structure of the estimator $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_\nu)$. We may consider the estimator, which we take here to be completely arbitrary, as a function of either independent data $\{Y_i\}$, defined in (2.1) or record-value data $\{Z_i\}$, defined in (2.2). Let $\{\delta_n\}$ denote a sequence of positive constants converging to zero as $n \to \infty$. The assertion $\hat{\beta}_k - \beta_k^0 = O_p(\delta_n)$ means that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P\left(\left|\hat{\beta}_k - \beta_k^0\right| > \lambda\delta_n\right) = 0.$$

THEOREM 4.2. *Assume conditions* (C_2), *and let* $1 \le k \le \nu$. *Then* $\hat{\beta}_k(Y_1, \dots, Y_n) - \beta_k^0 = O_p(\delta_n)$ *if and only if* $\hat{\beta}_k(Z_1, \dots, Z_n) - \beta_k^0 = O_p(\delta_n)$.

To illustrate the application of this result, note that in the increasing polynomial regression model (2.3) and assuming that the coefficient β_ν of the term of highest degree is nonzero, there exists $C > 0$ such that $t(i, \beta^0) - t(j, \beta^0) \ge Ci^{\nu-2}$ for all $1 \le j \le i-1 < \infty$. Therefore, provided $\nu \ge 3$ and $E\{\max(X, 0)^{\nu/(\nu+1)}\} < \infty$, conditions (C_2) hold.

It is straightforward, using methods employed in the proof of Theorem 4.2, to show that for a wide range of specific estimator types and under conditions (C_2), the asymptotic distribution of $\hat{\beta}$ is the same no matter whether the data used are $(Y_1, \dots, Y_n)$ or $(Z_1, \dots, Z_n)$. More specifically, if the trend function consists of a sum of terms of the form bi^α, where $\alpha, b > 0$ and the α's are known, then the parameters b may be estimated consistently by least squares, minimizing the function $S(\beta)$ defined in (2.4). Such estimators satisfy central limit theorems. The asymptotic variance of $\hat{b}$ is identical in the respective cases where the data $\{Y_i\}$ or $\{Z_i\}$ are used if and only if the largest value of α in the series representation of $t(i, \beta)$ is strictly greater than 1. This result also applies to cases where some b's are negative, provided that the trend is eventually strictly increasing in asymptotic proportion to const $\times\, i^\alpha$, where $\alpha > 0$ denotes the largest exponent in the series. If the trend includes an intercept term, then that too may be estimated consistently by least squares if and only if the largest α exceeds 1. The principal regularity condition needed is that the distribution of X have sufficiently many finite moments, their number increasing with decreasing value of the largest α.

Of course, conditions (C_2) specifically exclude the case of a linear trend, which we consider next. Assume that the trend is given by $t(i, \beta) = a + bi$, where $b > 0$, and that $E(X) = 0$ and

$$E\{\max(X, 0)^{4+\delta}\} + E\{\max(-X, 0)^{(3/2)+\delta}\} < \infty$$

for some $\delta > 0$. Collectively we call these conditions (C_3). Let $\hat{a}$, $\hat{b}$ and σ^2 be given by (2.5) and (2.6), and put $\varepsilon_i = \max_{-\infty<j\le i}\{X_j - b(i-j)\}$ for $-\infty < i < \infty$ and $a' = a + E(\varepsilon_0)$. (In the special case of a linear trend, this definition of ε_i coincides with the definition of their distributions given more generally just prior to Theorem 4.1.) Conditions (C_3) are sufficient to ensure absolute convergence of the series used to define σ^2 in (2.6).

THEOREM 4.3. *Under conditions (C_3), $\hat{a} - a'$ and $\hat{b} - b$ are asymptotically normally distributed with zero means and variances $4n^{-1}\sigma^2$ and $12n^{-3}\sigma^2$, respectively.*

It is a little curious that Theorem 4.3 does not require the assumption that the variables X_i have finite second moment. This condition is needed for classical limit theory based on the "ideal" data $\{Y_i\}$, rather than on the records $\{Z_i\}$. That it is not needed here (although an assumption more stringent than finite second moment is required on the upper tail) is a consequence of the manner in which distribution tails are distorted—with weight shifted from the lower to the upper tail—by the operation of taking maxima.

Asymptotic theory for the consistent estimator of a, defined in (2.9), may be developed as a corollary of that for estimating the marginal distribution H of the stationary process $\{\varepsilon_i + a,\ -\infty < i < \infty\}$ and so we treat the latter problem first. Put

$$\widehat{H}(x) = n^{-1}\sum_{i=1}^{n} I(Z_i - \hat{b}i \le x), \tag{4.1}$$

where $\hat{b}$ is defined in (2.5). We claim that, on the space $D(-\lambda, \lambda)$ of right-continuous functions with left-hand limits, mapping the interval $(-\lambda, \lambda)$ (for arbitrary $\lambda > 0$) into the real line, the stochastic process $n^{1/2}(\widehat{H} - H) \in D(-\lambda, \lambda)$ converges weakly to a process $\mathscr{H}$ whose distribution we now define. The conditions that we shall impose are sufficient to ensure that $h \equiv H'$ is well defined, bounded and continuous on $(-\infty, \infty)$, and that for all $x_1, x_2 \in \mathbb{R}$, the series

$$\pi(x_1, x_2) \equiv \sum_{i=-\infty}^{\infty} \{P(\varepsilon_0 \le x_1,\ \varepsilon_i \le x_2) - H(x_1)H(x_2)\}$$

converges absolutely. Let $\zeta_1(\cdot)$ and ζ_2 denote, respectively, a Gaussian process with zero mean whose covariance function is given by π, that is,

$$\operatorname{cov}\{\zeta_1(x_1), \zeta_1(x_2)\} = \pi(x_1, x_2)$$

and a Normal random variable with zero mean and variance $3\sigma^2$, stochastically independent of $\zeta_1(\cdot)$. Define $\mathscr{H}(x) = \zeta_1(x)+\zeta_2 h(x)$. The second term here represents the contribution to the limit of $\widehat{H}$ that arises from employing an estimator of b, rather than the true value, in definition (4.1). Under the conditions below, $|\pi(x_1, x_2) - \pi(x_1, x_1)| = O(|x_1 - x_2|)$ as $x_2 \to x_1$, which ensures continuity of the process $\zeta_1(\cdot)$.

Assume that for some $x_0 > 0$ we have $P(|X| < x_0) = 1$, that the distribution of X is absolutely continuous with a bounded, continuous density and satisfies $E(X) = 0$ and of course that the trend function t is given by $t(i, \beta) = a + bi$, where $b > 0$. We call these conditions (C_4). There is no serious difficulty in substantially weakening the assumption of a bounded X distribution in Theorem 4.4, by first truncating the variables both above and below at levels depending on n and then showing that the chosen truncation does not affect the first-order asymptotic theory. However, such weakening would prove particularly inconvenient in later work. For this reason we have chosen to work here, too, with the more stringent assumption, although it would be enough here to ask that $E(|X|^k) < \infty$ for $k > 0$ sufficiently large.

THEOREM 4.4. *Under conditions* (C_4) *and for each* $\lambda > 0$, *the process* $n^{1/2}(\widehat{H} - H)$ *converges weakly on* $D(-\lambda, \lambda)$ *to* $\mathscr{H}$.

Next we consider properties of the estimators $\widehat{G}$ and $\hat{\mu}_r$, defined in (2.7) and (2.8), respectively, of the distribution and moments of $X+a$. First we define the limiting distributions of these quantities, the former in terms of the stochastic process to which it converges. Let $\mathscr{H}$ be as defined two paragraphs above and put

$$\mathscr{G}(x) = H(x+b)^{-1}\{\mathscr{H}(x) - G(x)\mathscr{H}(x+b)\},$$

which is a Gaussian process with zero mean. Let

$$\tau_r^2 = r^2 \int_0^\infty \int_0^\infty (xy)^{r-1} \operatorname{cov}[\{(-1)^r \mathscr{G}(-x) - \mathscr{G}(x)\},\ \{(-1)^r \mathscr{G}(-y) - \mathscr{G}(y)\}]\, dx\, dy.$$

THEOREM 4.5. *Under conditions* (C_4) *and for each* $\lambda > 0$, *the process* $n^{1/2}(\widehat{G} - G)$ *converges weakly on* $D(-\lambda, \lambda)$ *to* $\mathscr{G}$. *Furthermore*, $n^{1/2}(\hat{\mu}_r - \mu_r)$ *is asymptotically normally distributed with zero mean and variance* τ_r^2.

Taking $r = 1$ in this result we see that $\tilde{a}$ is asymptotically Normal $N(0, n^{-1}\tau_1^2)$.

One corollary of the invariance principle for the process $\widehat{G}$ is that, under the hypothesis that the distribution of X is symmetric, the estimator $\check{a}$ discussed in Section 2.5 is root-n consistent for a. Indeed, we may state a central limit theorem as follows.

COROLLARY 4.5.1. *Assume conditions* (C_4), *and that* $G(a+x)+G(a-x) = 1$ *for all* x. *Write* $g = G'$ *for the density corresponding to* G. *Then* $\check{a}$ *is asymptotically normally distributed with zero mean and variance* $n^{-1}g(a)^{-2}\operatorname{var}\{\mathscr{G}(a)\}$.

The assumption that the distribution of X has bounded support is used in a number of ways in the proof of Theorem 4.5 and seems difficult to relax without a substantial amount of additional work. In particular, the assumption implies that the integral in the definition of $\hat{\mu}_r$ may be taken over only a finite range, without affecting the validity of our results. Additionally, it enables us to disregard the possibility that, in the definition of $\mathscr{G}(x)$ and in technical arguments related to that definition, the denominator term $H(x+b)$ is close to zero. To appreciate why, note that $H(x+b)$ is bounded above zero for values of x such that $H(x) > 0$. This property fails if the distribution of X is not bounded below.

Nevertheless, it is possible to establish rates of consistency of our estimators of G and μ_r under weaker conditions than compact support of the X distribution, even though central limit theorems and weak convergence results seem out of reach at present. In particular, if the distribution of X has all moments finite, then it may be shown that the estimators of G, a and μ_r converge at rate $n^{-(1/2)+\delta}$ (the former uniformly on compacts) for each $\delta > 0$.

The asymptotic variance of $\widehat{G}(x)$ equals $n^{-1}\tau(x)^2$, where

$$\begin{aligned}\tau(x)^2 = H(x+b)^{-2}\big[&\pi(x,x) + G(x)^2\pi(x+b,x+b)\\ &- 2G(x)\pi(x,x+b) + 3\sigma^2 h(x)^2\{1-G(x)\}^2\big].\end{aligned}$$

Note particularly that both $\tau(x)^2$ and τ_r^2 depend on the unknown density h of the distribution of $\varepsilon_i + a$. The latter may be estimated by kernel methods, for example, as

$$\hat{h}(x) = (nl)^{-1}\sum_{i=1}^{n} K\{(Z_i - \hat{b}i - x)/l\},$$

where l and K denote a bandwidth and kernel function, respectively. Alternatively, the bootstrap may be used. Our last result, which is stated without proof since the argument is straightforward but somewhat tedious, asserts the consistency of the bootstrap method. We assume conditions (C_4) and that the density $f = F'$ of the distribution of X is Hölder continuous; we shall refer collectively to these assumptions as conditions (C_5).

THEOREM 4.6. *Under conditions* (C_5), *the bootstrap estimators of* $\tau(x)^2$ *and* τ_r^2 *are consistent.*

5. Proofs.

PROOF OF THEOREM 4.1. Observe that for any $j_0 \geq 1$ and $i \geq j_0$,

$$\begin{aligned}p_i(x) &\equiv P(e_{i+1} \leq x_1, \ldots, e_{i+n} \leq x_n)\\ &= P\Big[\max_{0\leq j\leq i+k-1}\{t(i+k-j) - t(i+k) + X_{i+k-j}\} \leq x_k,\ 1 \leq k \leq n\Big]\\ &= p_{ij_01}(x) - p_{ij_02}(x),\end{aligned}$$

where

$$p_{ij_01}(x) = P\Big[\max_{0\le j\le j_0}\{t(i+k-j) - t(i+k) + X_{i+k-j}\} \le x_k,\ 1\le k\le n\Big]$$

and

$$0 \le p_{ij_02}(x) \le \sum_{j=1}^{i+n-j_0} P\{t(j) - t(i+1) + X_j > \min(x_1,\dots,x_n)\}.$$

By hypothesis, $t(i+1) - t(j) \ge C(i+1-j)^{\xi}$ for $j \le i+1$, and so if $j_0 \ge n+1$ is so large that $C(j_0-n)^{\xi} + 2\min(x_1,\dots,x_n) > 0$, then

$$p_{ij_02}(x) \le \sum_{j=j_0-n}^{\infty} \{1 - F(\tfrac{1}{2}Cj^{\xi})\}.$$

The assumption that $E\{\max(X,0)^{1/\xi}\} < \infty$ is sufficient to ensure that the series here converges and in fact that the series may be made less than an arbitrary $\delta > 0$ by choosing $j_0 = j_0(\delta)$ sufficiently large. Therefore,

$$\lim_{j_0\to\infty}\limsup_{i\to\infty} p_{ij_02}(x) = 0. \tag{5.1}$$

Since F is continuous and $t(i+j) - t(i) \to s(j)$ as $i \to \infty$, then

$$p_{ij_01}(x) \to p_{j_01}(x) \equiv P\Big[\max_{0\le j\le j_0}\{X_{k-j} - s(j)\} \le x_k,\ 1\le k\le n\Big]$$

as $i \to \infty$. Now, $p_{j_01}(x) = p_{\infty 1}(x) + p_{j_02}(x)$, where

$$p_{\infty 1}(x) \equiv P\Big[\max_{0\le j<\infty}\{X_{k-j} - s(j)\} \le x_k,\ 1\le k\le n\Big],$$

$$0 \le p_{j_02}(x) \le \sum_{j=j_0+1}^{\infty} P\{X - s(j) > \min(x_1,\dots,x_n)\}$$

$$\le \sum_{j=j_0+1}^{\infty} \{1 - F(\tfrac{1}{2}Cj^{\xi})\},$$

the latter inequality following for sufficiently large j_0 and using the fact that $s(j) \ge Cj^{\xi}$. Once again, the finiteness of $E\{\max(X,0)^{1/\xi}\}$ ensures that the series on the right-hand side converges and may be made arbitrarily small by selecting j_0 sufficiently large. Indeed, $\lim_{j_0\to\infty} p_{j_02}(x) = 0$. Theorem 4.1 follows from this result and (5.1).

PROOF OF THEOREM 4.2. We prove only that $\hat\beta_k(Y_1,\dots,Y_n) - \beta_k^0 = O_p(\delta_n)$ implies $\hat\beta_k(Z_1,\dots,Z_n) - \beta_k^0 = O_p(\delta_n)$. Recall that β^0 denotes the true value of β and observe that for any $m \geq 1$,

$$(5.2)\qquad \begin{aligned} &P(Z_i \neq Y_i \text{ for some } i \geq m) \\ &\quad = P\Big[\max_{1\leq j\leq i-1}\{X_j + t(j,\beta^0) - t(i,\beta^0)\} > X_i \text{ for some } i \geq m\Big] \\ &\quad \leq \sum_{i=m}^{\infty}\sum_{j=1}^{i-1} P\{X_j + t(j,\beta^0) - t(i,\beta^0) > X_i\} \\ &\quad \leq \sum_{i=m}^{\infty}\sum_{j=1}^{i-1} P(X_j - X_i > Ci^{\eta}) = \sum_{i=m}^{\infty}(i-1)P(X_1 - X_2 > Ci^{\eta}) \\ &\quad \leq B\int_{m^{\eta}}^{\infty} x^{(2/\eta)-1}P(X_1 - X_2 > x)\,dx \equiv p(m), \end{aligned}$$

say, where the constant B does not depend on m. Therefore,

$$P\{\hat{\beta}_k(Z_1,\dots,Z_n) \neq \hat{\beta}_k(Z_1,\dots,Z_m,Y_{m+1},\dots,Y_n)\} \leq p(m)$$

for all $1 \leq m \leq n < \infty$. Hence, for each fixed $m \geq 1$,

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} P\{|\hat{\beta}_k(Z_1,\dots,Z_n) - \beta_k^0| > \lambda\delta_n\} \leq p(m) + q(m),$$

where

$$q(m) \equiv \lim_{\lambda\to\infty}\limsup_{n\to\infty} P\{|\hat{\beta}_k(Z_1,\dots,Z_m,Y_{m+1},\dots,Y_n) - \beta_k^0| > \lambda\delta_n\}.$$

In view of the assumption that $E\{\max(X,0)^{(1/\eta)+1}\} < \infty$, the integral defining $p(m)$ converges and $p(m) \to 0$ as $m \to \infty$. Hence, it suffices to prove that for each $m \geq 1$, $q(m) = 0$.

Suppose the latter result is false. Then there exists $m \geq 1$, $\varepsilon > 0$ and a sequence $\lambda_n \to \infty$ such that, along a subsequence $\{n_i\}$,

$$(5.3)\qquad P\{|\hat{\beta}_k(Z_1,\dots,Z_m,Y_{m+1},\dots,Y_n) - \beta_k^0| > \lambda_n\delta_n\} \geq \varepsilon.$$

For real numbers $x_1,\dots,x_m$, define

$$\pi_n(x_1,\dots,x_m) = P\{|\hat{\beta}_k(x_1,\dots,x_m,Y_{m+1},\dots,Y_n) - \beta_k^0| > \lambda_n\delta_n\}.$$

Since $\hat{\beta}_k(Y_1,\dots,Y_n) - \beta_k^0 = O_p(\delta_n)$, then $E\{\pi_n(Y_1,\dots,Y_m)\} \to 0$. Hence, any given subsequence of $n = 1, 2, \dots$ contains a sub-subsequence such that $\pi_n(Y_1,\dots,Y_m) \to 0$ almost surely as $n \to \infty$ through that subsequence. Take the subsequence to be $\{n_i\}$ and let the sub-subsequence be $\{n_{i(j)}\}$. Now, $\pi_{n_{i(j)}}(Y_1,\dots,Y_m) \to 0$ implies $\pi_{n_{i(j)}}(Z_1,\dots,Z_m) \to 0$, both convergences being almost sure. [Here we need the assumption that the distribution of X is continuous with support $(-\infty,\infty)$ or $[x_0,\infty)$, since it ensures that the support $\mathscr{S}_x$ of $X(m) = (X_1,\dots,X_m)$ equals the support $\mathscr{S}_y$ of $Y(m) = (Y_1,\dots,Y_m)$. Without the assumption we can produce particularly perverse estimators for which the theorem fails. Indeed, if the distribution of X has atoms or is continuous with an appropriate support not equal to $(-\infty,\infty)$ or $[x_0,\infty)$, then, by constructing $\hat{\beta}(u_1, \dots, u_n)$ to take the value $\beta^0 + (1, \dots, 1)$ for particular combinations of $(u_1, \dots, u_m)$ that lie in one of $\mathscr{S}_x$ and $\mathscr{S}_y$ but not in the other, we may ensure that the theorem is violated.] Furthermore, $\pi_n \le 1$. It follows that $E\{\pi_{n_{i(j)}}(Y_1, \dots, Y_n)\} \to 0$ as $j \to \infty$. This contradicts (5.3) and so establishes the theorem. □

PROOF OF THEOREM 4.3. We derive only the central limit theorem for $\hat{b}$, since that for $\hat{a}$ may be established similarly. Without loss of generality, the process $\{X_i\}$ is doubly infinite. The notation below is consistent with that given earlier if we put $a = 0$ and drop the assumption that $E(X) = 0$:

$$e_i = \max_{1\le j\le i}\{X_l - b(i-j)\}, \qquad \varepsilon_i = \max_{-\infty<j\le i}\{X_j - b(i-j)\},$$

$$\Delta_i = \max_{-\infty<j\le 0}\{X_j - b(i-j)\}, \qquad V = \max_{-\infty<j\le 0}\left(X_j + \tfrac{1}{2}bj\right),$$

$$s^2 = \sum_{i=1}^{n}(i-\bar{i})^2, \qquad Y_i' = bi + \varepsilon_i.$$

Let $\hat{b}'$ denote the version of $\hat{b}$ in which $\{Y_i\}$ is replaced by $\{Y_i'\}$. Our first task is to prove that

$$\hat{b}' - \hat{b} = o_p(n^{-3/2}). \tag{5.4}$$

Note that $0 \le \varepsilon_i - e_i = \max(\Delta_i - e_i, 0)$, $\Delta_i \le V - \frac{1}{2}bi$ and $e_i \ge X_i$. Therefore,

$$\begin{aligned} s^2|\hat{b}' - \hat{b}| &= \left|\sum_{i=1}^{n}(i-\bar{i})(\varepsilon_i - e_i)\right| \le \sum_{i=1}^{n}|i-\bar{i}|(\varepsilon_i - e_i) \\ &\le \tfrac{1}{2}(n-1)\sum_{i=1}^{n}\max\left(V - \tfrac{1}{2}bi - X_i, 0\right). \end{aligned} \tag{5.5}$$

The finiteness of $E(|X|^{3/2})$ implies that for each $c, d > 0$,

$$\sum_{i=1}^{n} E\{\max(c - di - X, 0)\} = o(n^{1/2}). \tag{5.6}$$

Result (5.4) follows from (5.5) and (5.6).

In view of (5.4) it suffices to establish the claimed central limit theorem for $\hat{b}'$ rather than $\hat{b}$. Put $\varepsilon_i' = \varepsilon_i - E(\varepsilon_i)$ and $S_i = \sum_{1\le j\le i}\varepsilon_j'$. Let $\theta_n(t)$ denote the stochastic process obtained by interpolating linearly among the points

$$(0,0), (1/n, S_1/n^{1/2}), \dots, (i/n, S_i/n^{1/2}), \dots, (1, S_n/n^{1/2}).$$

Write W for a standard Brownian motion on the interval $[0, 1]$ and note that

$$s^2\hat{b}' = \sum_{i=1}^{n}(i-\bar{i})\varepsilon_i' = \tfrac{1}{2}(n+1)S_n - \sum_{i=1}^{n}S_i.$$

The theorem will follow from this representation if we prove that θ_n converges weakly to σW in the sense of convergence of continuous functions on the interval [0, 1] equipped with the uniform metric. For that purpose it suffices to prove that, with $\mathscr{F}_i$ denoting the sigma field generated by $\{X_{-i}, X_{-i-1}, \ldots\}$,

$$\sum_{i=1}^{n} \left[E\{E(\varepsilon_0'|\mathscr{F}_i)\}^2\right]^{1/2} < \infty, \tag{5.7}$$

$n^{-1}E(S_n^2) \to \sigma^2$ and $n^{-3}s^2 \to 1/12$. See Corollary 5.4 of Hall and Heyde (1980). We shall outline only the derivation of (5.7).

Put

$$\varepsilon_{1i} = \max_{-\infty < j \le -i} (X_j + bj), \qquad \varepsilon_{2i} = \max_{-i+1 \le j \le 0} (X_j + bj).$$

Then $\varepsilon_0 = \max(\varepsilon_{1i}, \varepsilon_{2i})$ and

$$\begin{aligned} 0 \le E(\varepsilon_0|\mathscr{F}_i) - E(\varepsilon_{2i}) &= E\{(\varepsilon_{1i} - \varepsilon_{2i})I(\varepsilon_{1i} > \varepsilon_{2i})|\varepsilon_{1i}\} \\ &\le E\{(\varepsilon_{1i} - X_0)I(\varepsilon_{1i} > X_0)|\varepsilon_{1i}\} = \int_{-\infty}^{\varepsilon_{1i}} F(x)\,dx. \end{aligned}$$

Therefore,

$$\begin{aligned} E\{E(\varepsilon_0'|\mathscr{F}_i)\}^2 &\le E\{E(\varepsilon_0|\mathscr{F}_i) - E(\varepsilon_{2i})\}^2 \\ &\le \int\int P\{\varepsilon_{1i} > \max(x_1, x_2)\}F(x_1)F(x_2)\,dx_1\,dx_2 \\ &\le C_1 \int P(\varepsilon_{1i} > x)F(x)\max\{x, F(x)\}\,dx, \end{aligned} \tag{5.8}$$

where the constants $C_1, C_2, \ldots$ do not depend on i and unqualified integrals are over the entire real line. Now,

$$P(\varepsilon_{1i} > x) \le \sum_{j=-\infty}^{-i} P(X_j + bj > x) = \sum_{j=i}^{\infty} P(X > x + bj).$$

Therefore, since $E\{\max(X, 0)^{4+\delta}\} < \infty$,

$$P(\varepsilon_{1i} > x) \le C_2 \begin{cases} \sum_{j=i}^{\infty} (x + j)^{-(4+\delta)}, & \text{if } x > 0, \\ i^{-(2+\delta)}, & \text{if } -\frac{1}{2}bi < x \le 0, \\ 1, & \text{if } x \le -\frac{1}{2}bi. \end{cases}$$

The assumption $E\{\max(-X, 0)^{(3/2)+\delta}\} < \infty$ implies that

$$F(x) \le C_3\{\max(1, -x)\}^{-\{(3/2)+\delta\}}$$

and so by (5.8),

$$
\begin{aligned}
E\{E(\varepsilon_0'|\mathscr{F}_i)\}^2 \le C_4\bigg[&\sum_{j=i}^{\infty}\int_0^{\infty} x(x+j)^{-(4+\delta)}\,dx \\
&+ i^{-(2+\delta)}\int_{-bi/2}^{0}\{\max(1,-x)\}^{-(3+2\delta)}\,dx \\
&+ \int_{-\infty}^{-bi/2}\{\max(1,-x)\}^{-(3+2\delta)}\,dx\bigg] \\
\le C_5 i^{-(2+\delta)}.&
\end{aligned}
$$

This bound leads directly to (5.7). □

PROOF OF THEOREM 4.4. We adopt notation from the proof of Theorem 4.3 and in particular take $a = 0$ and drop the assumption that $E(X) = 0$. Since $|X| \le x_0$, then $Z_i = Y_i'$ for all $i \ge m_0$, where m_0 denotes the integer part of $(2x_0/b) + 1$. This result plays a role here similar to that assumed by (5.2) and allows us to assert that with $\zeta_{2n} = n^{3/2}(\hat{b} - b)$ and

$$\widehat{H}_1(x, u) = n^{-1}\sum_{i=1}^{n} I(\varepsilon_i \le x + in^{-3/2}u)$$

we have

$$|\widehat{H}(x) - \widehat{H}_1(x, U)| \le n^{-1}m_0. \tag{5.9}$$

Put

$$
\begin{aligned}
H_1(x,u) = E\{\widehat{H}_1(x,u)\} &= n^{-1}\sum_{i=1}^{n} P\big(\varepsilon_0 \le x + in^{-3/2}u\big) \\
&= H(x) + \tfrac{1}{2}n^{-1/2}uh(x) + o\big(n^{-1/2}\big),
\end{aligned}
\tag{5.10}
$$

the $o(n^{-1/2})$ term being of that size uniformly in $|x|, |u| \le \lambda$ for each fixed $\lambda > 0$. Let $\zeta_{3n}(x, u) = n^{1/2}\{\widehat{H}_1(x, u) - H(x)\}$. The next step in our proof is to establish a joint invariance principle for the pair $(\zeta_{3n}(\cdot,\cdot), \zeta_{2n})$ on the space $D\{(-\lambda,\lambda)\times(-\lambda,\lambda)\}\times(-\infty,\infty)$, where $D\{(-\lambda,\lambda)\times(-\lambda,\lambda)\}$ denotes the space of right-continuous functions with left-hand limits, from $(-\lambda,\lambda)\times(-\lambda,\lambda)$ to the real line, λ is any positive number and the function space is equipped with the uniform metric.

It is straightforward to prove that $\zeta_{1n}(x, u) \equiv n^{1/2}\{\widehat{H}_1(x, u) - H_1(x, u)\}$ converges weakly, as a stochastic process defined on $D\{(-\lambda,\lambda)\times(-\lambda,\lambda)\}$, to the process $\zeta_1(x)$ defined just prior to Theorem 4.4 (thus, the effect of u disappears in the limit) and that this convergence is joint with that of ζ_{2n} to $2\zeta_2$. The method of proof of convergence of finite-dimensional distributions of ζ_{1n}, jointly with that of the one-dimensional "process" ζ_{2n}, utilizes m_0 dependence of the summands in the definitions of both quantities. Tightness of ζ_{1n} as a stochastic process may be established by decomposing ζ_{1n} in the obvious way into a sum of m_0 distinct processes, each having independent summands, establishing tightness of these component processes by following essentially the standard route for proving tightness in the convergence of an empirical process to its Gaussian limit (the only variant being a minor modification to take care of the appearance of u as well as x) and combining the results to obtain tightness of the original process ζ_{1n}, noting that the maximum of a sum does not exceed the sum of the maxima of individual summands.

This argument establishes weak convergence of the stochastic process $(\zeta_{1n}(x, u), \zeta_{2n})$ to $(\zeta_1(x), 2\zeta_2)$. In view of (5.10), this implies convergence of $\zeta_{3n}(\cdot, \zeta_{2n})$ to $\zeta_1(\cdot) + \zeta_2 h(\cdot)$. The theorem follows directly from that result and (5.9). □

PROOF OF THEOREM 4.5. Define $\tilde{H}(x) = n^{-1} \sum_{1 \le i \le n} I\{Z_i - \hat{b}(i+1) \le x - b\}$. An argument similar to that in the proof of Theorem 4.4 may be employed to show that under conditions (C_4), $n^{1/2}(\widehat{H} - \tilde{H}) \to 0$ in probability, uniformly on the interval $(-\lambda, \lambda)$ for each $\lambda > 0$. It follows that

$$\widehat{G}(x) = \widehat{H}(x)/\tilde{H}(x+b) = G(x) + J(x) + o_p(n^{-1/2}) \tag{5.11}$$

uniformly in $x \in \mathscr{S}_x^\delta$, where $\mathscr{S}_x^\delta$ denotes the set of all points that lie within a sufficiently small distance $\delta > 0$ of at least one point in the support $\mathscr{S}_x$ of X and

$$J(x) = H(x+b)^{-1}[\widehat{H}(x) - H(x) - G(x)\{\widehat{H}(x+b) - H(x+b)\}].$$

Therefore, recalling the definition of $\hat{\mu}_r$ at (2.8) and remembering that (again because of the compact support of the X distribution) the integral there may be taken over only a finite range, we see that

$$\hat{\mu}_r - \mu_r = r \int_0^\infty x^{r-1}\{(-1)^r J(-x) - J(x)\}\, dx + o_p(n^{-1/2}). \tag{5.12}$$

The theorem follows directly from (5.11) and (5.12). □

Maximum likelihood estimation. We treat the case of a linear trend $t(i) = a + bi$, when a model is assumed for the distribution of X. Our proof outline here will show that if the error distribution is misspecified but the assumption of a linear trend is correct, then the maximum likelihood estimator (MLE) of a converges in probability to a number different from a and that the MLE of b is consistent but is so heavily biased that it converges at rate n^{-1}, not the $n^{-3/2}$ suggested by the information matrix. By way of contrast, distribution-free estimators of a and b suggested in Section 2 converge to the true values at rates $n^{-1/2}$ and $n^{-3/2}$, respectively.

Let the assumed density of the error distribution be $f(x) = c \exp\{-l(x)\}$, where $c > 0$ is a constant and l is a symmetric, nonnegative loss function. We particularly have in mind the case where $l(x) = |x|^\alpha$ for some $\alpha > 0$. Let F denote the associated distribution function and put $l_1(x) = l(x) - \log F(x) > 0$. Employing both the record values Z_i and the record indicators I_i (where $I_i = 0$ or 1 according as $Z_i = Z_{i-1}$ or $Z_i > Z_{i-1}$), the negative log likelihood multiplied by n^{-1} equals

$$\mathscr{L}(a, b) = n^{-1}\sum_{i=1}^{n} \{l(Z_i - a - bi) + (1 - I_i)l_1(Z_i - a - bi)\},$$

except for a constant not depending on a or b; see Smith [(1988), equation (2.3)]. Write a_0, b_0 for the true values of a, b and put $\delta_a = a - a_0$, $\delta_b = n(b - b_0)$ and $\varepsilon = \max_{-\infty < j \le 0}(X_j + bj)$. Then $\mathscr{L}(a, b)$ is asymptotic to $\mathscr{L}_1(\delta_a, \delta_b)$ as $n \to \infty$, where

$$\mathscr{L}_1(u, v) = \int_0^1 [E\{l(\varepsilon - u - vx)\} + pE\{l_1(\varepsilon - b - u - vx)\}]\, dx.$$

If (u_0, v_0) denotes the minimizer of $\mathscr{L}_1(u, v)$, then $\hat{a} - a_0 \to u_0$ and $n(\hat{b} - b) \to v_0$ in probability as $n \to \infty$. Since the actual distribution of ε is arbitrary, we may select it so that $\delta_a \ne 0 \ne \delta_b$.

Acknowledgment. We thank Dr. Uri Goldburt of Tel Aviv University for assistance in the athletic records data gathering exercise.

REFERENCES

BALLERINI, R. and RESNICK, S. (1985). Records from improving populations. *J. Appl. Probab.* **22** 487–502.

BALLERINI, R. and RESNICK, S. (1987). Records in the presence of a linear trend. *Adv. in Appl. Probab.* **19** 801–828.

BECKER, R. A., CHAMBERS, J. M. and WILKS, A. R. (1988). *The New S Language*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

BERRED, M. (1992). On record values and the exponent of a distribution with regularly varying upper tail. *J. Appl. Probab.* **29** 575–586.

CHATTERJEE, S. and CHATTERJEE, S. (1982). New lamps for old: an exploratory analysis of running times in Olympic games. *J. Roy. Statist. Soc. Ser. C* **31** 14–22.

DE HAAN, L. and VERKADE, E. (1987). On extreme value theory in the presence of a trend. *J. Appl. Probab.* **24** 62–76.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

MILLER, R. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.

RESNICK, S. I. (1973a). Limit laws for record values. *Stochastic Process. Appl.* **1** 67–82.

RESNICK, S. I. (1973b). Record values and maxima. *Ann. Probab.* **1** 650–662.

RESNICK, S. I. (1973c). Extremal processes and record value times. *J. Appl. Probab.* **10** 864–868.

RESNICK, S. I. (1975). Weak convergence to extremal processes. *Ann. Probab.* **3** 951–960.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

SHORROCK, R. W. (1972). A limit theorem for inter-record times. *J. Appl. Probab.* **9** 219–223.

SHORROCK, R. W. (1973). Record values and inter-record times. *J. Appl. Probab.* **10** 543–545.

SHORROCK, R. W. (1974). On discrete time extremal processes. *Adv. in Appl. Probab.* **6** 580–592.

SHORROCK, R. W. (1975). Extremal processes and random measures. *J. Appl. Probab.* **12** 316–323.
SMITH, R. L. (1988). Forecasting records by maximum likelihood. *J. Amer. Statist. Assoc.* **83** 331–338.
TRYFOS, P. and BLACKMORE, R. (1985). Forecasting records. *J. Amer. Statist. Assoc.* **80** 46–50.

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
CANADA M5S 3G3

CENTRE FOR MATHEMATICS AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA