# BREAKING RECENT GLOBAL TEMPERATURE RECORDS*

GILBERT W. BASSETT JR.**

*Department of Economics (m/c144), University of Illinois at Chicago, Box 802451, Chicago, Il 60680-2451, U.S.A.*

**Abstract.** Global surface temperature was a record in 1988. What is the probability that this record will be surpassed in the next few years? Answers are provided given a variety of simple statistical models for temperature. The answers illustrate how record breaking is influenced by alternative model specifications. Estimates for the probability of a record are shown to range widely. If annual temperature is independent and identically distributed then a new record is unlikely. But probabilities increase rapidly if there is a trend or autocorrelation. Estimates of the probability of a record using data on global temperature suggest that another record in the next few years would not be a rare event.

## 1. Introduction

A record high global temperature occurred in 1988. What is the probability that this record will be broken in the next few years? This question is considered here under alternative models for temperature. Annual global temperature is taken to be the realization of a random variable and a model describes the temperature random variable. Attention is restricted to models in which temperature depends only on statistical parameters; the models make no reference to causal mechanisms that affect global temperature. Further, the models are all simple; they only have a few parameters. The objective is to see how the probability of recently set records varies under alternative specifications.

The specific question that will be considered was motivated by the 'Hansen bet',[1]

> Climate expert James Hansen ... told a group of climatologists last week that his confidence that the greenhouse effect has arrived is even higher than it was in 1988, when he testified before Congress that he believed the global warming of recent decades was driven by gases produced by human activity. So sure is he now of this conclusion that he said he'd bet even money that one of the next 3 years will be the hottest in 100 years. '... People aren't going to believe such an "incredible" and scientifically outrageous prediction', Hansen said...

* After the original version of this paper was completed it was reported that a new temperature record was set in 1990; *Science* (1991).

** This research was done while visiting at Argonne National Laboratory. I have benefitted from discussions with Mark Fernau and Mo Yin Tam. I would like to thank S. Lebedeff for providing the temperature data. I also want to thank a referee for numerous helpful suggestions.

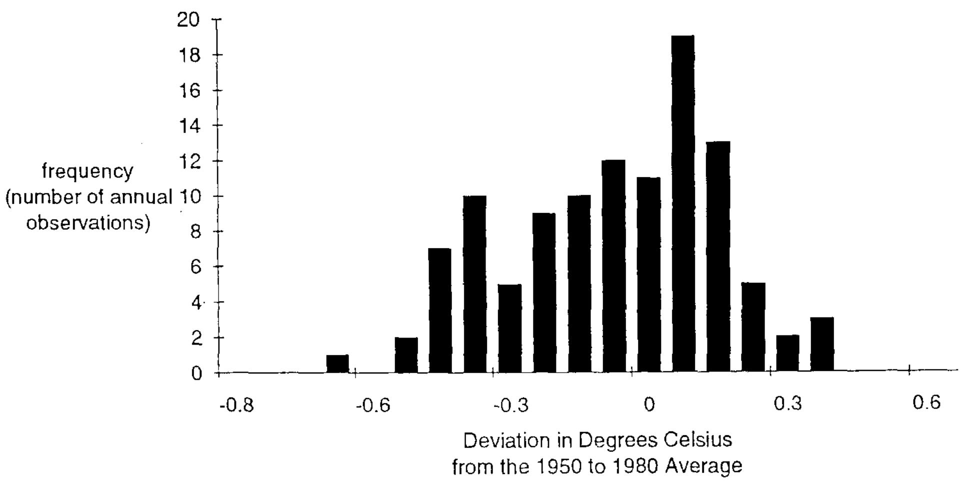
[1] Science (1990b).

*Climatic Change* **21**: 303–315, 1992.

Is the prediction really outrageous? Should the 'Hansen bet' be accepted?

Answers to these questions will be presented for several alternative temperature models. Since the models do not refer to factors that impact global climate (sunspots, volcanic activity, atmospheric concentrations of greenhouse gases, etc.) they cannot be used to infer the impact of human activities on global climate; see Wigley *et al.* (1985) and Solow and Broadus (1989) for a discussion of statistical issues relating to global climate.

Record breaking problems are often analyzed under the assumption of independent and identically distributed (IID) random variables; for an excellent review of the theory and applications of record breaking see Glick (1978). In this case the probability of breaking a record decreases quickly from the start of the series. For global temperature the distribution of values over the past 100 years is shown in Figure 1.[2] When this data is used to estimate the parameters of an IID temperature model the probability of a record in the next three years is found to be only about 0.05.




Case I

Average Temperature= -.11°
$\hat{\sigma}(v)$= .227

Fig. 1. Annual temperature.

[2] The temperature data is from Hansen and Lebedeff (1987). Annual global surface temperature is expressed as the deviation in degrees Celsius from the 1951–1980 average. In this paper the Hansen-Lebedeff data will be taken as the measure for annual global temperature. Another frequently cited data set has been developed by Jones *et al.* (1990) – its correlation with the Hansen-Lebedeff data is 0.94. Measurement of global temperature is not straightforward; comparability over time is difficult because of (i) heat island effects near cities, (ii) changing instrumentation over time, and (iii) until recently, sparse global coverage.

The IID situation however need not be accurate. Figure 2 depicts the time series of global temperature. It gives the impression of an upward drift in global temperatures. If this trend is a genuine feature of climate then the identically distributed assumption is wrong. The probability of a record when there is an increasing temperature is calculated below and shown to be about 0.37.

The IID assumption also could fail because independence is wrong. The simplest possibility in this regard is that successive temperatures are correlated. Figure 3 depicts temperature plotted against its previous value. The appearance of a positive slope suggests dependence between successive temperatures. Serial correlation is also suggested by a variety of mechanisms that could plausibly affect year-to-year changes in global climate (the relation however is likely to be much more complex than the simple 'one-period' serial correlation considered here). Simple reasons for serial correlations are thermal feedback between the ocean and atmosphere, sunspots, and volcanic activity. Oceans dissipate heat slowly so warm years, which heat the oceans, will likely be followed by warmer than average years; if sunspots affect climate then their own correlations will begin to show up in the temperature series; and volcanic activity – though itself possibly independent – will induce correlations if volcanic ash stays in the atmosphere for long periods. None of these mechanisms need imply a change in long term climate.

With a serial correlation model the probability of a record depends not only on the current record value, but also on how long ago the record was set. If the record was recent, as is true with global temperature, then the probability of a new record is higher than when the record was set in the distant past. It is shown below that with parameter values estimated from the historical data the probability of a record in the next three years is actually greater than $\frac{1}{2}$. Serial correlation alone suffices to make a new record a likely event.

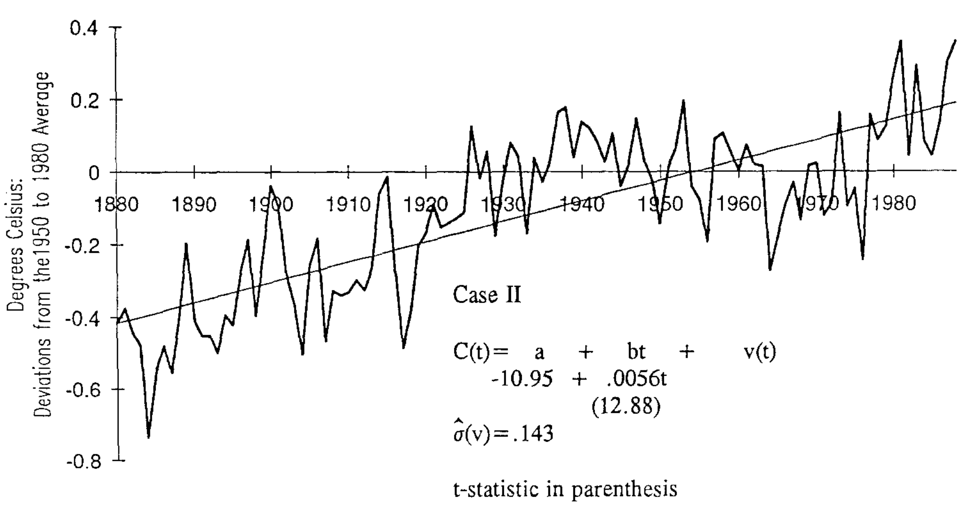



Fig. 2. Annual temperature time series.

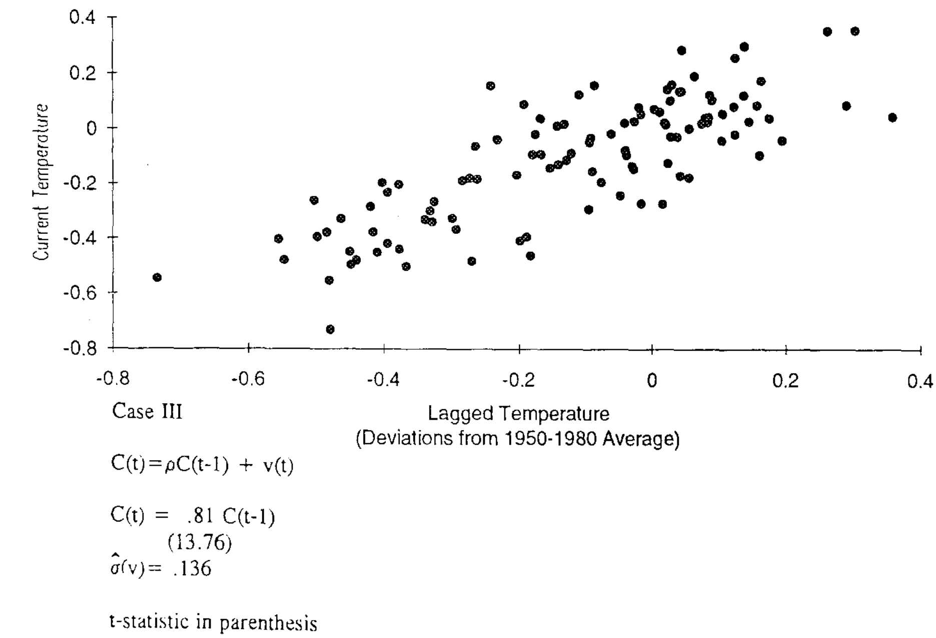



Fig. 3. Current vs lagged temperature.

The combination of a trend and serial correlation is also considered. When estimated values are substituted for parameters the probability of a record is also greater than $\frac{1}{2}$.

It should be emphasized that the analysis here does not bear directly on the question of whether increased concentrations of greenhouse gases in the atmosphere have already caused changes in global climate. The demonstration that the probability of a record may not be as small as might have been intuitively suspected and that the probability can be quite high even if there is no upward temperature trend suggests caution in inferring climate change from a succession of record temperatures. But the existing temperature data leave room for views that are consistent with a wide range of beliefs regarding global warming.

The next section describes the models and derives expressions for record probabilities. Section 3 shows how the probabilities change as a function of the parameters. It also provides numerical values for the probabilities by using estimated values for parameters. Results are summarized in Section 4.

## 2. Record Probabilities for Alternative Temperature Models

Global temperature values are denoted by $c(t)$ where $t = 1880, \ldots, 1991$. The unit for $c(t)$ is degrees Celsius and it is expressed as deviations from the 1950 to 1980 average temperature. The temperature values are taken to be realizations of random variables $C(t)$; uppercase $C$ denotes the temperature random variable with the actual realized temperature being denoted by the lower case $c$. The temperature values for $t = 1880, \ldots, t'$, where $t' = 1988$, are known. The maximum temperature – denoted by $r$ – occurred in the most recent year $t'$ so that $c(t') = r$.

The probability of a new record within three years, conditional on realizations up to $t'$, is given by

$$\begin{aligned}\text{Record Probability} &= \\ &= P[\max(C(t'+i) > r; \quad i = 1, 2, 3 \,|\, c(1880), \ldots, c(t')] \\ &= 1 - P[C(t'+i) < r; \quad i = 1, 2, 3 \,|\, c(1880), \ldots, c(t')]\end{aligned}$$

This expresses the record probability in terms of the joint distribution of the next three temperature values, conditional on values observed up to $t'$.

The conditional joint distribution of the next three values is derived using the following specification for the temperature time series,

$$C(t) = m(t) + e(t); \quad t = 1880, \ldots, 1991 \tag{1}$$

where $m(t)$ is the mean of $C(t)$ and $e(t)$ is a random variable that represents deviations from the mean.

The mean is assumed to depend linearly on time

$$m(t) = a + bt \tag{2}$$

where $a$ and $b$ are parameters. The parameter $b$ is the annual change in temperature trend in degrees Celsius. ($b$ is a model parameter, not to be confused with $\hat{b}$, which estimates the trend using sample data). The specification $b = 0$ means there is no trend.

The random variables $e(t)$ are assumed to be generated by the first-order autoregressive model

$$e(t) = \rho e(t-1) + v(t) \quad -1 < \rho < 1 \tag{3}$$

where the $\rho$ parameter is the correlation between successive $e(t)$'s, and the $v(t)$'s are independent with a common Gaussian distribution with mean zero and variance $\sigma^2(v)$; $v(t) \sim G(0, \sigma^2(v))$. The variance of $C(t)$ is $\sigma^2(v)/[1 - \rho^2]$.

The following cases will be considered separately.

Case I: $m(t) = a$ and $\rho = 0$.

This says that temperature varies around a mean value that does not change over time, $b = 0$, and that annual temperature is independent of past values.

Case II: $m(t) = a + bt; \quad \rho = 0$.

This says temperature increases as a linear function of time, but each year's temperature is independent of previous years.

Case III: $m(t) = a; \quad -1 < \rho < 1.$

This says there is no trend, $b = 0$, but temperature values are influenced by the previous year. This case will be considered in detail below.

Case IV: $m(t) = a + bt; \quad -1 < \rho < 1.$

This allows for both trend and first-order serial correlation.

Remarks:
1. Each specification is considered a separate model for temperature. The intent is not to select one model as best, but to see how they differ in their implied record probabilities.
2. Attention is restricted to simple models. More realistic univariate models would allow for time-varying parameters, higher order autoregressive specifications, moving average errors, first and higher order differences of $C(t)$, and so on. Multivariate models would introduce explanatory variables, like atmospheric concentrations of $CO_2$, as possible determinants of $m(t)$; for a recent example see Kuo *et al.* (1990) and also Barnett (1990). Conditional quantile models also could be used to represent and estimate, not only the mean, but also the quantiles and extremes of the $C(t)$ distribution; see Koenker and Bassett (1978) for discussion of quantile estimation and modeling.

The autoregressive model (3) implies that the distribution of temperature in the next three years, conditional on past values, depends only on $e(t')$, whose realized value is $c(t') - m(t')$. Substituting this into (1) gives an expression for temperature in the next three years,

$$\begin{aligned}
C(t'+1) &= m(t'+1) + \rho(c(t') - m(t')) + v(t'+1) \\
C(t'+2) &= m(t'+2) + \rho^2(c(t') - m(t')) + \rho v(t'+1) + v(t'+2) \\
C(t'+3) &= m(t'+3) + \rho^3(c(t') - m(t')) + \rho^2 v(t'+1) + \\
&\qquad \rho v(t'+2) + v(t'+3)
\end{aligned}$$

The random variables for the next three years are seen to depend on:

i the IID random variables; $v(t'+1)$, $v(t'+2)$, $v(t'+3)$,
ii the $m(t)$ and $\rho$ parameters, and
iii the realized value in 1988, $c(t')$.

The mean of $(C(t'+1), C(t'+2), C(t'+3))$ is given by

$$E[C(t'+1)] = m(t'+1) + \rho(c(t') - m(t'))$$
$$E[C(t'+2)] = m(t'+2) + \rho^2(c(t') - m(t'))$$
$$E(C(t'+3)] = m(t'+3) + \rho^3(c(t') - m(t'))$$

and the covariance matrix is given by $\sigma^2(v)\Omega(\rho)$, where

$$\Omega(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1+\rho^2 & \rho+\rho^3 \\ \rho^2 & \rho+\rho^3 & 1+\rho^2+\rho^4 \end{bmatrix}$$

The probability of a record in terms of the unknown parameters can now be derived for each case.

*Record Probabilities*

Case I: Temperature values are independent with the common distribution $G(a, \sigma^2(v))$. Hence, a record in the next three years has probability

$$1 - P[G(a, \sigma^2(v)) < r]^3. \tag{4}$$

Case II: Temperature values are independent, but the mean of the annual distribution changes over time. The record probability is

$$1 - P_1P_2P_3 \tag{5}$$

where $P_1 = P[G(a + b(t'+1),\ \sigma^2(v)) < r]$, $P_2 = P[G(a + b(t'+2),\ \sigma^2(v)) < r]$, and $P_3 = P[G(a + b(t'+3), \sigma^2(v)) < r]$

Case III: The record probability is

$$1 - P[G_3(\mu, \sigma^2(v)\Omega(\rho)) < r\iota] \tag{6}$$

where $\iota$ is a three-vector of ones and $G_3$ is a three-dimensional Gaussian random vector whose mean $\mu$ is,

$$\mu = \begin{bmatrix} a + \rho_2(c(t') - a) \\ a + \rho_3(c(t') - a) \\ a + \rho(c(t') - a) \end{bmatrix}$$

Case IV: The expression for the record probability is

$$1 - P[G_3(\mu, \sigma^2(v)\Omega(\rho)) < r\iota] \tag{7}$$

where $\mu$ is now,

$$\mu = \begin{bmatrix} a + b(t' + 1) + \rho_2(c(t') - a - b(t')) \\ a + b(t' + 2) + \rho_3(c(t') - a - b(t')) \\ a + b(t' + 3) + \rho(c(t') - a - b(t')) \end{bmatrix}$$

Remarks:

1. The record breaking problem is different from forecasting future values. For the forecasting problem $C(t' + i)$ is estimated as the mean of the conditional distribution; that is, if $C^p$ denotes the predicted value then

$$\begin{aligned} C^p(t' + 1) &= m(t' + 1) + \rho(c(t') - m(t')) \\ C^p(t' + 2) &= m(t' + 2) + \rho^2(c(t') - m(t')) \\ C^p(t' + 3) &= m(t' + 3) + \rho^3(c(t') - m(t')) \end{aligned}$$

   This differs from the 'record breaking probability' problem, which depends on both the mean and the dispersion of future values. The variance term, $\sigma^2(v)$, affects the probability of a record, but plays no role in constructing a point estimate for a future value.
2. The variances of future observations, conditional on the present value of $c(t')$, are seen (on the diagonal of $\sigma^2\Omega(\rho)$) to be nonconstant when $\rho \neq 0$. This differs from the unconditional variances of $C(t)$, which are constant over time.
3. The covariance matrix $\sigma^2\Omega(\rho)$ also differs from the 'first-order' form in its off-diagonal elements. Considered unconditionally, covariances are equal for any two adjacent time periods. But for the conditional the covariance between this year and next year is $\sigma^2(v)\rho$, but between next year and the year after it is $\sigma^2(v)$ $(\rho + \rho^3)$.

## 3. Record Probabilities

Record probability values are derived below for each of the four cases.

*Case I*

Figure 1 shows the frequency plot for the global temperature for 1880 to 1988. The sample mean is −0.11° and the sample standard deviation is 0.227. The 1988 record temperature was $c(1988) = 0.35°$. (Recall that temperature is measured in degrees Celsius and expressed in terms of the deviation from average temperature over 1950 to 1980).

For the IID model the predicted value for temperature for each of the next three years would be the sample mean of −0.11°.

The probability that a record will occur is found by substituting the estimated parameter values into (4),

$$\text{Record probability} = 1 - Pr[G(-0.11, 0.227^2) < 0.35]^3$$

$$= 1 - Pr[G(0,1) < (0.46/0.227)]^3$$
$$= 1 - Pr[G(0,1) < 2.03] = 1 - 0.979^3$$
$$= 0.061.$$

In this case the Hansen bet should be accepted (that is, bet against a new record) when offered at even odds.

*Case II*

Figure 2 shows the temperature data plotted against time. It also shows the estimated trend and least squares estimates for the trend values. The predicted temperatures for the next three years are given by the trend estimates,

$$C^p(1989) = -10.95 + 0.00562 * 1989 = 0.195°$$
$$C^p(1990) = -10.95 + 0.00562 * 1990 = 0.200°$$
$$C^p(1991) = -10.95 + 0.00562 * 1991 = 0.206°$$

The probability of a record can be estimated by substituting the estimates for $a$, $b$, and $\sigma^2(v)$ into (5). It is found that the probability of record next year is 0.134, for the year after it is 0.143, and for three years from now it is 0.152. With temperature values assumed to be independent this gives 0.37 as the probability for a record within three years. The upward temperature trend therefore leads to an increase in the record probability from 0.05 to 0.37.

*Case III*

The forecasted values for the next three years under this model decay toward the mean of −0.11° (and away from the record value of 0.35°)

$$C^p(1989) = -0.11 + 0.8 * 0.46 = 0.266°$$
$$C^p(1990) = -0.11 + 0.8^2 * 0.46 = 0.193°$$
$$C^p(1991) = -0.11 + 0.8^3 * 0.46 = 0.135°$$

Figure 3 depicts temperature plotted against its lagged value. It also shows the least squares estimates for $\rho$ and $\sigma^2(v)$. The fitted and actual values are presented in Figure 4. Figure 3 shows the estimated correlation is high and Figure 4 shows the model fits the data fairly well.

It is important to observe that the appearance of a trend in Figure 2 is spurious given the correlated but identically distributed temperature model of Case III. With high enough correlation there can be long term 'trends' in temperature records that are induced by the serial dependence alone.

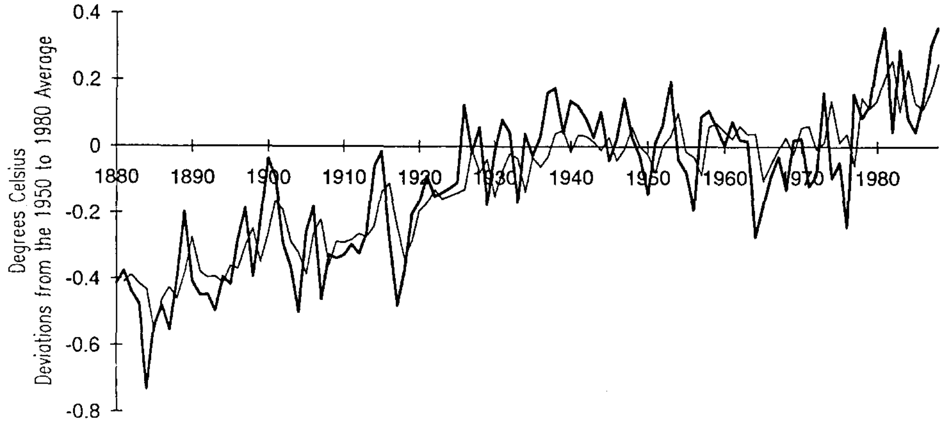



Fig. 4. Fitted and actual temperature Case III.

The expression for a record probability is given by equation (6).[3] Figure 5 depicts this probability as a function of $\rho$. In the figure the value for $e(t')$ is taken to be the 1988 estimated value, $c(1988) - \hat{a} = 0.35 - (-0.11) = 0.46°$.[4] The probabilities are presented for two bounding values for $\sigma^2(v)$; the $0.12^2$ value is based on deviations from trend, whereas the $0.24^2$ is based on deviations from the overall mean. Recall that the variance of $C(t)$ is given by $\sigma^2(v)/[1-\rho^2]$ and therefore the variance of temperature varies with both $\sigma^2(v)$ and $\rho$. The probabilities are computed for $\rho$ values in the interval [0, 1] in increments of 0.1.

The Figure shows the probability of a record increases rapidly as $\rho$ increases. When $\rho$ is zero the probability of a record is near zero; the 0.061 estimate of Case I corresponds to $\rho = 0$ and $\sigma^2(v) = 0.227^2$, the point labeled B in the figure.

The probabilities reach a maximum at $\rho = 1$ where the probability of a record is 7/8. The case $\rho = 1$ means that temperature is a random walk, with a $\frac{1}{2}$ probability that the temperature next year will be higher or lower than in the current year. When the record has just been set the probability of exceeding the record, the current value, in at least one of the next three years is therefore, $1 - (\frac{1}{2})^3$ or 7/8.

Finally, the figure can be used to identify the region of $\rho$ and $\sigma^2(v)$ values where

[3] The probability $P[G_3(\mu, \sigma^2(v)\Omega(\rho)) < r\iota]$ was computed by first converting to the probability of independent variables, $P[G_3(0, I_3) < (\sigma^2(v)\Omega(\rho))^{-0.5}(r\iota - \mu)]$. The mathematical software program 'matlab' was then used to compute $(\Omega(\rho))^{-0.5}$ for $\rho$ at ten equally spaced points in the interval in [0, 1]. Multiplication, again using 'matlab', of $(\sigma^2(v)\Omega(\rho))^{-0.5}(r\iota - \mu)$ gave the 3-vector used to look up probabilities in a table of univariate normal probabilities.

[4] The most recent observation in the Hansen-Lebedeff data set is 1988 and it is used as the final value in computing the probability of a record in the next three years. Preliminary data for 1989 indicates that it was a warm year (the fifth hottest in the last 100), but not a record; Science (1990a). The estimated probability values that are reported here would be different for the correlation models if the last year of data was 1989. The probability of a record depends on the most recent deviation from the mean and a lower value for $e(t')$ would tend to lower the probability of a record.

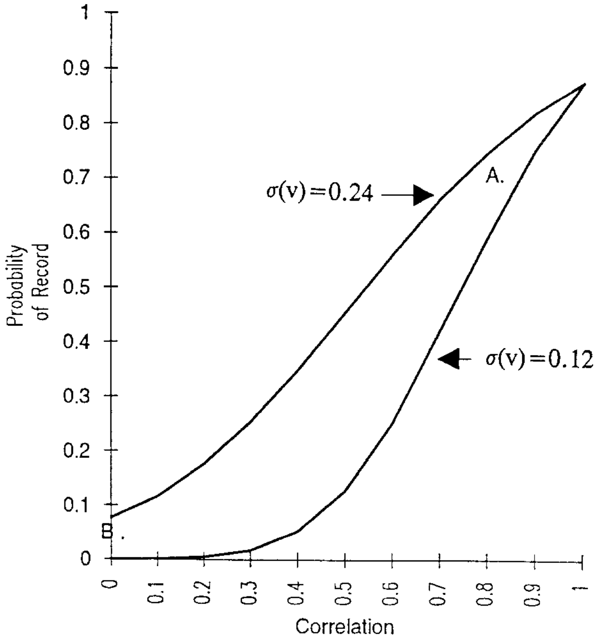



Fig. 5. Record probability as a function of correlation.

the probability of a record is greater than $\frac{1}{2}$. A person who believed this region contained the $\rho$ and $\sigma^2(v)$ parameters would think that the record probability was greater than $\frac{1}{2}$, and hence that the even odds offered by the Hansen bet were not favorable. The estimated values for $\hat{\rho} = 0.8$ and $\hat{\sigma}(v) = 0.136$ (point A in the Figure) are seen to be in this region and the estimated probability of a record is 0.65. Based on the empirical data the Hansen bet looks good for Hansen.

*Case IV*

Figure 6 shows the actual and fitted values for temperature when both the trend and $\rho$ parameters are estimated. The predicted temperature values for the next three years are

$$C_p(1989) = -11.03 + 0.0056 * 1989 + 0.503 * 0.105 = 0.250°$$
$$C_p(1990) = -11.03 + 0.0056 * 1990 + 0.503^2 * 0.105 = 0.230°$$
$$C_p(1991) = -11.03 + 0.0056 * 1991 + 0.503^3 * 0.105 = 0.222°$$

Substituting the estimated values for the parameters in (7) gives an estimated record probability of about 0.75. The estimate for $\rho$ has decreased from 0.81 to 0.50, which – other things equal – lowers the probability of a record. But the trend term increases the probability of a record enough that the net effect shows an increase from the previous case.

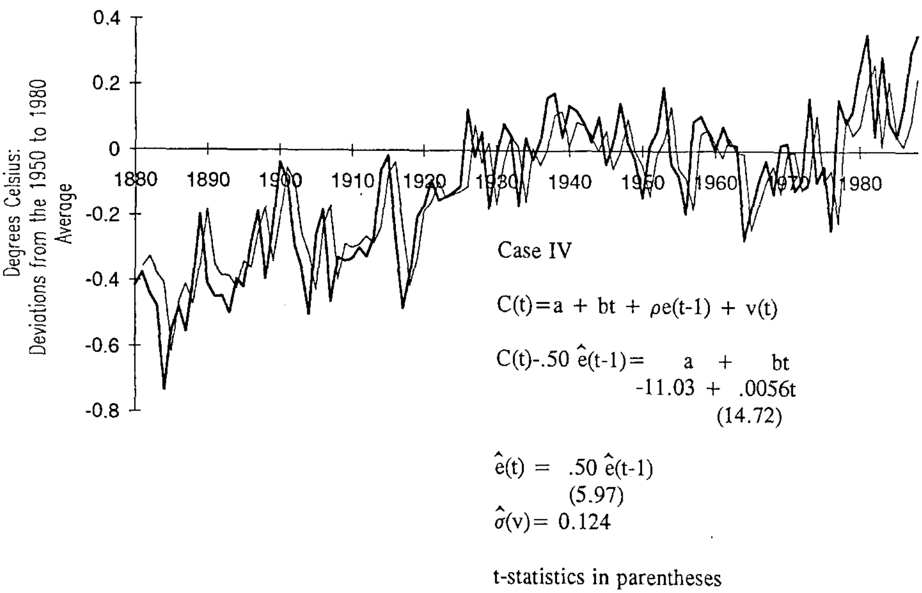



Fig. 6. Fitted and actual temperature Case IV.

## 4. Conclusion

The probabilities for a record under the alternative specifications are summarized in Table I. It shows that a record goes from rare to likely, depending on the temperature model. It is rare when there is no trend and little correlation. It becomes more likely when there is an increasing temperature trend and when annual temperature is positively correlated with past values. Correlation matters because the recent record implies a deviation from trend that is positive and large. With positive correlation this makes future positive deviations more likely, thus increasing the probability of a record. When the probability of a record is estimated allowing for trend and serial correlation the Hansen bet becomes favorable for Hansen at the stated even odds.

TABLE I: Parameter estimates and record probabilities

| | Case I | Case II | Case III | Case IV |
|---|---|---|---|---|
| $\hat{a}$ | −0.1135 | −10.95 | −0.1135 | −11.03 |
| $\hat{b}$ | 0 | 0.005603 | 0 | 0.005647 |
| $\hat{\rho}$ | 0 | 0 | 0.808 | 0.503 |
| $\hat{\sigma}^2(v)$ | $0.227^2$ | $0.143^2$ | $0.136^2$ | $0.124^2$ |
| $\hat{e}(1988)$ | – | – | 0.46 | 0.10 |
| Record probability | 0.06 | 0.37 | 0.65 | 0.75 |

## References

Barnett, T. P.: 1990, 'Beware Greenhouse Confusion', *Nature* **343**, 743.

Glick, N.: 1978, 'Breaking Records and Breaking Boards', *American Mathematical Monthly,* January, 2–26.

Hansen, J. and Lebedeff, S.: 1987, 'Global Trends of Measured Surface Air Temperature', *J. Geophys. Res.* **92**, 13,345–13,372.

Jones, P. D., Wigley, T. M. L., and Wright, P. B.: 1990, *Global Hemispheric Annual Temperature Variations between 1861 and 1988*, ORNL/CDIAC-31, NDP-022/R1, Oak Ridge National Laboratory.

Kerr, R. A.: 1990a, 'Global Warming Continues in 1989, *Science* **247**, 521.

Kerr, R. A.: 1990b, 'Watch Out! Here Comes the Greenhouse', *Science* **248**, 549.

Kerr, R. A.: 1991, 'Global Temperature Hits Record Again', *Science* **251**, 274.

Koenker, R. W. and Bassett, G. W.: 1978, 'Regression Quantiles', *Econometrica* **46**, 33–50.

Kuo, C., Lindberg, C., and Thomson, D. J.: 1990, 'Coherence Established Between Atmospheric Carbon Dioxide and Global Temperature', *Nature* **343**, 709–714.

Solow, A. and Broadus, J. M.: 1989, 'On the Detection of Greenhouse Warming', *Climatic Change* **15**, 449–453.

Wigley, T. M. L., Angell, J. K., and Jones, P. D.: 1985, 'Analysis of the Temperature Record', in *Detecting The Climatic Effects of Increasing Carbon Dioxide*, M. C. MacCracken and F. M. Luther (eds.), Carbon Dioxide Research Division, DOE/ER-0235, December, 57–90.

(Received 5 October 1990; in revised form 1 October 1991)