INFERENCE ABOUT TRENDS IN GLOBAL TEMPERATURE DATA*

JOHN W. GALBRAITH[1] and CHRISTOPHER GREEN[2]
[1]*Department of Economics, McGill University, Montreal, Quebec, H3A 2T7, Canada*
[2]*Department of Economics and Centre for Climate and Global Change Research, McGill University, Montreal, Quebec, H3A 2T7, Canada*

Abstract. Interpretation of the effects of increasing atmospheric carbon dioxide on temperature is made more difficult by the fact that it is unclear whether sufficient global warming has taken place to allow a statistically significant finding of any upward trend in the temperature series. We add to the few existing statistical results by reporting tests for both deterministic and stochastic non-stationarity (trends) in time series of global average temperature. We conclude that the statistical evidence is sufficient to reject the hypothesis of a stochastic trend; however, there is evidence of a trend which could be approximated by a deterministic linear model.

1. Introduction

The increasing interest in global warming has focused attention on two time series: the monotonically-increasing (in annual measurements) concentration of atmospheric CO_2 and the highly-variable global-average temperature series. While there is considerable scientific evidence (*e.g.* IPCC, 1990) for the prediction that continued accumulation of greenhouse gases will eventually cause global warming, it is less clear that substantial warming has already taken place. In particular, there has not been agreement on the empirical question of whether or not temperature data contain an upward trend beyond what could reasonably be attributed to sampling fluctuation, much less on the question of a link between the two time series. In this paper we focus on this question of whether there is a genuine trend in temperature, or merely a set of random fluctuations within the range that could be expected from a stationary series. That is, we consider the problem of *inference* about underlying changes in climate, rather than on documentation of the actual increase in average temperature that has recently been observed.

Among the methods that have been used to investigate a possible link between atmospheric carbon dioxide concentration and global average temperatures, there have been a number of statistical studies aimed at finding a statistically significant relationship in long time series of the two variables. Because one hypothesis of interest (that of a global trend toward warming, whether or not related to atmospheric CO_2) implies non-stationary temperature data, it is important that statistical

* The authors are grateful to the SSHRC (Green) and FCAR (Galbraith) for financial support under grants 10-89-0205 and NC-0047.

Climatic Change **22:** 209–221, 1992.

work take account of the voluminous recent literature on the treatment of non-stationary time series data; see in particular Fuller (1976), Dickey and Fuller (1979), Said and Dickey (1984), Phillips (1986, 1987). We will discuss some of these methods briefly in section 2.

An examination of atmospheric CO_2 data (see, *e.g.*, Keeling *et al.*, 1989) clearly suggests a fairly regular seasonal cycle around a marked increasing trend. In the case of global temperature data, however, any underlying pattern is obscured by more irregular fluctuation (see Figures 1–3). It is therefore not clear whether or not the observed changes reflect a statistically significant trend or other change. Moreover, the nature of the stochastic process generating temperature data is crucial to the statistical examination of a possible link between that variable and atmospheric CO_2, or other greenhouse gases. If fluctuations in temperature are simply the random fluctuations of a stationary time series, then there is no genuine global warming *trend* to be explained, by CO_2 concentrations or by any other cause. If there is statistical evidence of an increasing trend in global temperature, however, then there are a number of methods by which to investigate a possible relationship between two non-stationary series which may be applied. There are also well-known pitfalls in attempting to identify such relationships (beginning with Yule (1926)). For example, Phillips (1986) points out the inadequacy of deterministic de-trending methods if the non-stationarity springs from a stochastic trend.[1]

The question of the stationarity or non-stationarity of the global average temperature series has been addressed by, *inter alia*, Solow (1987), Solow and Broadus (1989), Tsonis and Elsner (1989), and is surveyed by Wigley and Barnett (1990). The first of these studies uses a two-phase regression model to test for a possible break in trend; Solow is unable to reject the hypothesis of no change in the trend in the temperature process, but does not address the question of significance of the measured trend itself. Solow and Broadus reach a similar conclusion based on a less formal examination of the temperature data. Tsonis and Elsner (1989) use a Monte Carlo technique to assess the significance of recent large temperature deviations from long-term averages and conclude that a significant change has occurred; they do not address the form of the possible trend or step-change in average temperatures. Finally, it is interesting to note that Kuo, Lindberg and Thomson (1990) use a deterministic trend model to look for evidence of coherence between the spectra of the CO_2 concentration and global average temperature series. The latter authors do not consider the possibility of a stochastic, rather than deterministic, trend in the underlying series. The deterministic linear de-trending which they apply can produce misleading results if the underlying trend is stochastic (again see, *e.g.*, Phillips (1986), but the comparison of the two models in the present paper

[1] By 'stochastic trend' we mean a process containing a latent root in its autoregressive polynomial which lies outside the unit circle. A series with a 'deterministic trend' is one which could be expressed as the sum of a stationary series and a deterministic part which is tending to increase (or decrease) over time. Both are *non-stationary* processes.

suggest that their modelling strategy was the appropriate one.

The present study investigates the question of the stationarity or non-stationarity of global temperature data by applying tests for *stochastic* non-stationarity (in the form of a latent root of the auto regressive polynomial which lies on the unit circle), and for a particular deterministic non-stationarity. The statistical test for the former is that of Dickey and Fuller (1979), a t-type test using a Monte Carlo tabulation of the appropriate non-standard distribution. By examining two parametric forms, the results offer some information about the nature of the trend, as well as its significance or insignificance.

It is important to bear in mind from the outset the limitations of statistical tests such as these. The tests are of use in answering the question of whether any changes in temperature over the sample period (in this case 1880–1988 inclusive) imply a statistically significant change in mean, or whether instead the observed fluctuations may legitimately be ascribed to sampling error. This question, while of considerable interest, is necessarily a narrow one. The tests cannot tell us anything about fluctuations with a very long period, such that the sample of data available to us fail to cover a full cycle. In general, of course, we cannot expect to be able to detect patterns which repeat only over a span of time greatly exceeding the span of our data. Instead we hope only to be able to determine whether or not sampling fluctuations are sufficient to account for such changes as we do see, or whether instead we should recognize that some underlying change has taken place, bearing in mind the possibility that the change could later be reversed as part of a longer cycle. Wigley and Raper (1990), using simulation models of 'internal' climate variability, suggest that low-frequency fluctuations such as those produced by oceans could account for a natural change of up to about 0.3 °C/century. While this is not sufficient to account for the observed increase over the last century, the mechanism may explain the stationary deviations from the linear trend model below.

The results of this study may be viewed as complementary to those of Solow (1987), who tests a related deterministic-trend hypothesis. Again, Solow tests for a change in a deterministic trend, and finds no evidence of such change, but this result leaves unanswered the question of whether the (presumably unbroken) trend is positive to a statistically significant degree, or whether the series is stationary. We attempt to answer this question in section 3, with the result of section 2 as a precursor.

2. Data and Testing for a Stochastic Trend

The data used here were kindly provided by James Hansen and Jeffrey Jonas of the Goddard Space Flight Center, Institute for Space Studies, and are described in Hansen and Lebedeff (1987). The series of global temperature 'changes', or deviations, runs in monthly increments from 1880 to 1988, yielding a total of 1308 observations. Hansen and Lebedeff use the symbol ΔT for these changes, but the series should *not* be confused with one of first differences, $T_\tau - T_{\tau-1}$. Each entry represents the deviation of an average temperature in the region from the mean for that month over the interval 1951–1980, measured in Celsius degrees; see Hansen and Lebedeff, pp. 13,348–350. Hence the series is a level series, non-stationary under the hypothesis that temperature contains a trend. Because the deviation is taken from the monthly mean, a constant seasonal effect is removed from the data. If seasonal patterns are changing over the sample period, however, these changes will remain in the data.

The data are also available for particular regions, and we report statistics for the Northern and Southern hemisphere averages as well; Figures 1 to 3 represent these data graphically, and it is important to note that unlike most graphical representations, these figures represent each data point. The lines are therefore much less smooth than those in Hansen and Lebedeff, and give some indication of the inferential problem involved in separating sample fluctuation from trend. We use monthly rather than annual data because, while monthly data contain little extra information with respect to long-term (low-frequency) changes in average temperatures, they may provide detail useful in investigating the form and magnitude of any trend found to be present.

Our first null hypothesis is that of a stochastic trend in the temperature deviations,[2] and in particular that the series can be described by a model with an autoregressive polynomial containing a root of unity. While we do not wish to suggest that the stochastic trend hypothesis is especially plausible as a characterization of a very long span of temperature data, the model may fit well over a restricted sample.[3] It will be useful to investigate this form of trend before considering a deterministic trend, because data well characterized by a stochastic trend may nonetheless appear to contain a deterministic trend if the deterministic model is fitted. It is therefore useful to eliminate the stochastic trend hypothesis first; if we were unable to do so, we would have little confidence in the value of a linear trend model. Moreover, even if the temperature process has an autoregressive root which is less than, but close to, unity ('borderline non-stationarity'), we may need extremely large samples before being able to rely on asymptotic distributional results in a test for a deterministic trend.

The test for stochastic non-stationarity that we apply is described in Fuller (1976) and Dickey and Fuller (1979). The null hypothesis is of a root of unity in the autoregressive representation of the time series; in AR(1) form,[4]

$$Y_t = Y_{t-1} + u_t\,, \quad \alpha(L)u_t = \theta(L)\varepsilon_t\,, \tag{1}$$

where ε_t is a white-noise process and the latent roots of the autoregressive poly-

[2] It is appropriate to take non-stationarity as the null here since a clear rejection of this null will allow us to conduct inference more easily in the second stage.

[3] However Gordon (1991) does argue in favour of the random walk model as a characterization of these data.

[4] Test statistics are reported for the data in level form. The logarithmic transformation makes only minimal changes in test statistics (typically in the second decimal place).

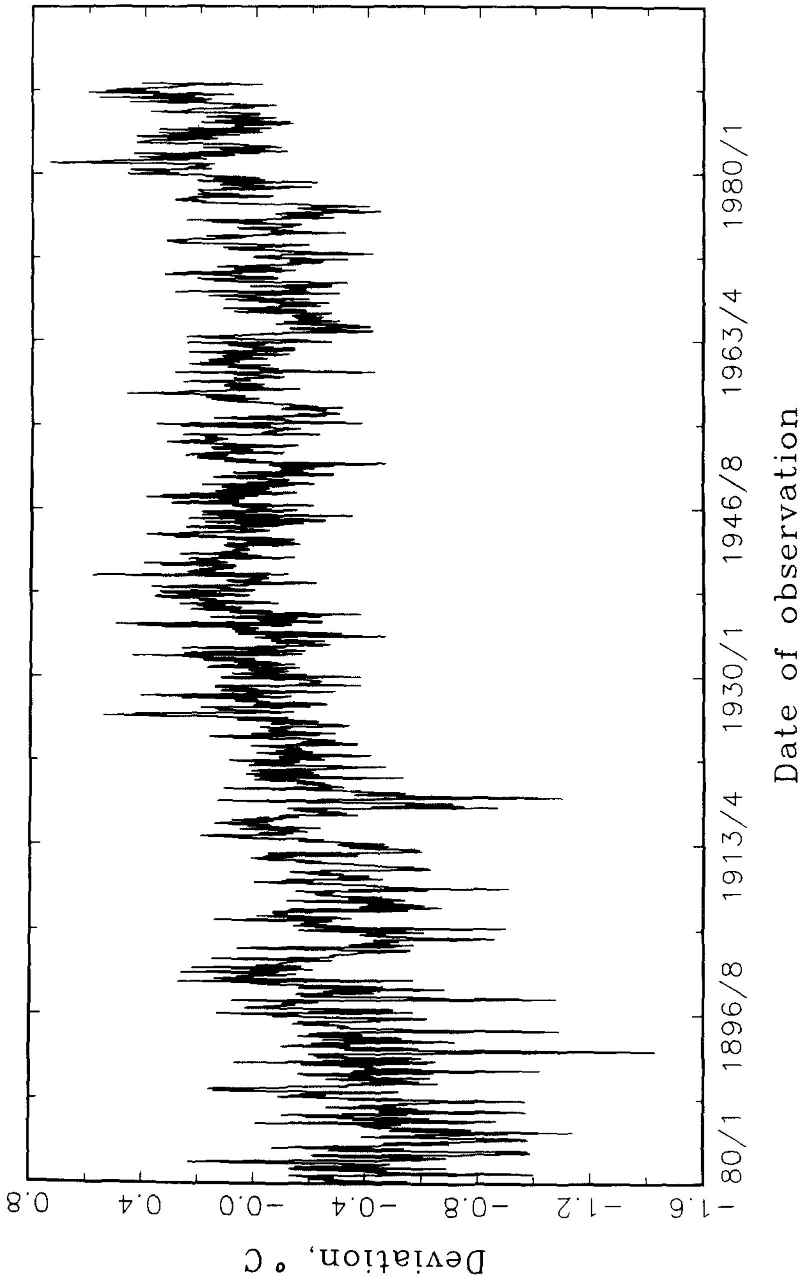

Fig. 1. Temperature deviations from 1951–1980 mean (Global).

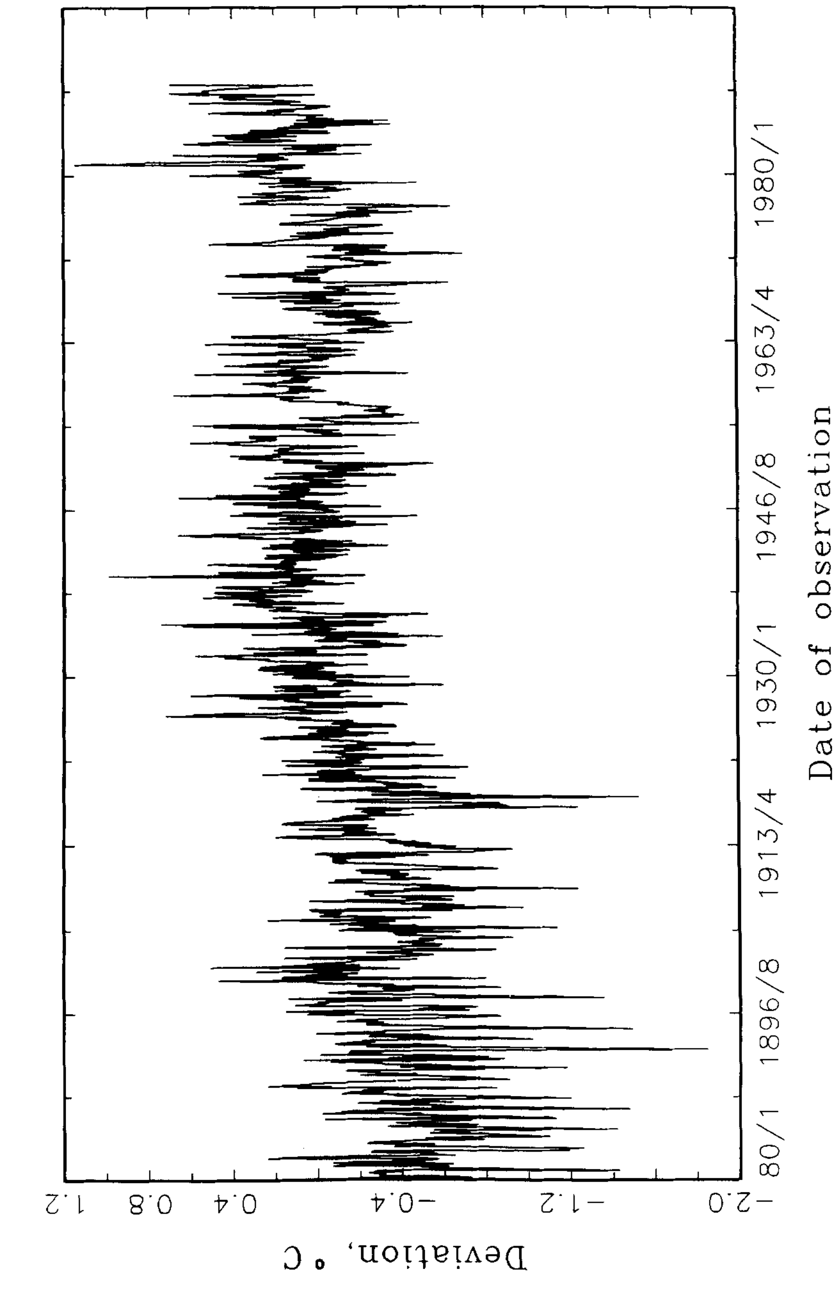

Fig. 2. Temperature deviations from 1951–1980 mean (Northern hemisphere).

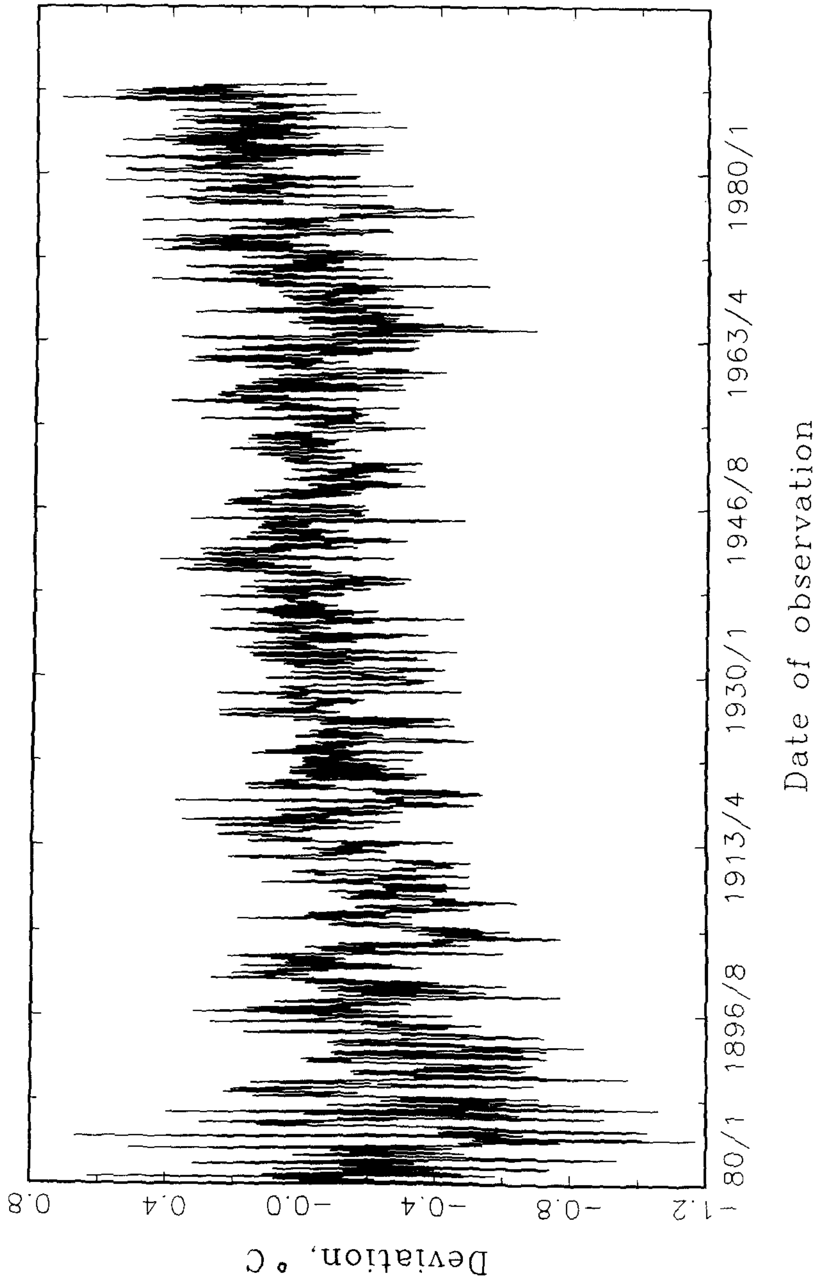

Fig. 3. Temperature deviations from 1951–1980 mean (Southern hemisphere).

nomial $\alpha(L)$ lie within the unit circle. The test statistic is the conventional 't'-statistic on γ_0 in the regression

$$\Delta Y_t = \alpha + \gamma_0 Y_{t-l} + \sum_{i=1}^{k} \gamma_i \Delta Y_{t-i} + v_t \,, \tag{2}$$

or equivalently

$$Y_t \;=\; \alpha + (1+\gamma_0) Y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta Y_{t-i} + v_t \,.$$

Under the null hypothesis, $\gamma_0 = 0$. However the distribution of the statistic is *not* the standard t-distribution; Fuller (1976) tabulates the percentiles of the distribution for the model (2) and for the variant (3) below. The number of lagged differences, k, is chosen to render $\{v_t\}_1^T$ a white noise process; Said and Dickey (1984) provide bounds on the appropriate rate of increase of k with the sample size when these lagged values are used to capture moving-average, as well as autoregressive, components in the underlying error process $\{u_t\}_1^T$.

Table I reports test statistics and critical values for the model (2) applied to the

TABLE I: Augmented Dickey-Fuller Statistics; $H_0 : \gamma_0 = 0$

Model: (2)	Percentile of distribution[5]			Test statistics[6]		
k	5%	2.5%	1%	North	South	Global
2	−2.86	−3.12	−3.43	−9.78	−10.60	−8.27
4	−2.86	−3.12	−3.43	−8.09	−8.28	−6.78
8	−2.86	−3.12	−3.43	−5.12	−5.50	−4.45
Model: (3)	**Percentile of distribution**			**Test statistics**		
k	5%	2.5%	1%	North	South	Global
2	−3.41	−3.66	−3.96	−12.90	−13.70	−11.68
4	−3.41	−3.66	−3.96	−11.23	−11.24	−10.10
8	−3.41	−3.66	−3.96	−7.70	−8.10	−7.18
	Estimates of γ_0					
	Model: (2)			Model: (3)		
k	North	South	Global	North	South	Global
2	−0.22	−0.28	−0.16	−0.36	−0.44	−0.31
4	−0.19	−0.23	−0.14	−0.35	−0.41	−0.29
8	−0.13	−0.17	−0.10	−0.28	−0.34	−0.23

[5] See Fuller (1976) p. 371 ff. for the full set of percentiles of this distribution under the null hypothesis.
[6] The t-type test statistic is calculated as $\hat{\gamma}_0/\mathrm{SE}(\hat{\gamma}_0)$. In each case the magnitude of the test statistic exceeds the critical value given by any of the top percentiles of the distribution. Hence the probability that the null is false exceeds 99% for each model.

global average temperature series. The addition of a linear (deterministic) time trend to the model (2) requires a change in the critical values applied, which is reflected in the values reported in Table I for the modified model:

$$\Delta Y_t = \alpha + \beta t + \gamma_0 Y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta Y_{t-i} + v_t \,, \tag{3}$$

$t = 1, 2, \ldots T$, where $T = 1308$ is the sample size. Note that it is useful to have the deterministic trend and autoregressive terms present simultaneously in the model, since variation attributable to each of these terms may well be present. Lagrange multiplier tests for autocorrelation (not reported) indicate that $k = 4$ is sufficient in most cases to produce a residual error term not significantly different from white noise.

As Table I indicates, the null hypothesis of a unit root in the autoregressive polynomial is given a very low probability by the test, and the deviation from the null is in the direction of stationarity. Moreover, the parameter γ_0 is substantially less than zero on these samples, indicating that the process is well away from the region of non-stationarity ($1 + \gamma_0 \geqslant 1$). On the existing sample, then, average temperature data are consistent with a fairly strongly autoregressive underlying process, but one which does not retain the effects of all past stochastic shocks indefinitely. This fact is important in interpreting the results of the next section.

3. Testing for a Linear Deterministic Trend

A test for the presence of a linear trend must also account for the fact that the distribution of the 't'-statistic on a trend term in a linear regression has a non-standard distribution[7] when a stochastic trend is present; even in the absence of a stochastic trend the statistic has a non-standard distribution in finite samples if the process is strongly autoregressive. However, the results of section 2 tell us that an underlying stochastic trend is highly unlikely. Furthermore, adding Y_{t-1} to both sides of models (2) and (3) to transform to AR form, the first autoregressive parameter is given by $1 + \gamma_0$ and we see that values of $1 + \gamma_0$ range from 0.56 ($\gamma_0 = -0.44$: Model (3), $k = 2$, Southern hemisphere) to 0.90 ($\gamma_0 = -0.10$: Model (2), $k = 8$, Global average). On a sample of the size available here, this parameter is sufficiently far from unity that we can rely on the asymptotic normality of the t-statistic on the trend term in the model, ignoring finite-sample distortions.[8]

Solow (1987) treated the temperature series as possibly containing a purely deterministic linear trend, and asked whether any break in trend could be detected. While no such break could be found, the possibility remains that an unbroken

[7] The distribution of the statistic on a linear trend term is in fact non-degenerate for a stochastically trending variable which contains no linear trend in the true process: Phillips (1986).

[8] The importance of the small-sample distortion to the distribution is easily established by small Monte Carlo or bootstrap experiments. The effect gains in importance as the largest root of the AR polynomial approaches one, and diminishes as the sample size grows.

trend in temperature affected the entire data series. In this case we would again say that stationarity fails, and the possibility that such a break could be attributed to increasing atmospheric CO_2 concentration remains. To test this possibility, we return to the model (3), or to the transformed (by the addition of Y_{t-1} to each side) version

$$Y_t = \alpha + \beta t + (1 + \gamma_0) Y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta Y_{t-i} + v_t \,. \tag{4}$$

Table II contains the estimates of β on the various samples and the corresponding t-statistics.

Table II indicates that there is sufficient evidence on this sample to reject easily a null hypothesis of no (linear) deterministic trend at conventional significance levels. A natural question to pose in interpreting this result is that of whether or not this trend rate of increase was approximately constant over the sample, an effect which one might attempt to detect through a test for a change in a linear trend at an unknown point in the sample. Solow (1987) provides just such a test statistic which, again, does not offer strong evidence of a change in trend. Further evidence on the stability of the trend is provided in Figure 4, which records recursive estimates of the trend parameter (multiplied by 10^4, and shown with rough confidence bands) over time, indicating the evolution of the parameter estimate as successive data points are incorporated into the sample.[10] There is some indication here of non-constancy in the parameter (the two-standard-error confidence bands for the mid-1940's estimates do not contain the final estimates), but we nonetheless see a reasonable degree of stability as sample evidence accumulates (the recent dip reflects the postwar relative cooling which ended around the mid-1970's). It is also interesting to note that a zero coefficient is outside the bands at all points.

The magnitude of the trend coefficient is about 1.4×10^{-4} °C/month in global data. However this impact coefficient is not the model's unconditional forecast of warming as a function of time. The unconditional expectation of the derivative $\partial y_t / \partial t$ in (3) or (4) is given by $E(\partial y_t / \partial t) = \beta[1 - (1 + \gamma_0)]^{-1} = -\beta \gamma_0^{-1}$. The result

TABLE II: Tests of Significance of Linear Trend and Estimates of Trend Parameter[9]
$H_0 : \beta = 0$

Model: (4)	North		South		Global	
k	$\hat{\beta} \times 10^4$	$t_{(\beta=0)}$	$\hat{\beta} \times 10^4$	$t_{(\beta=0)}$	$\hat{\beta} \times 10^4$	$t_{(\beta=0)}$
2	1.88	8.12	1.53	8.33	1.43	8.04
4	1.85	7.61	1.42	7.41	1.38	7.37
8	1.49	5.70	1.20	5.88	1.11	5.61

[9] The units of β are °C/month.
[10] Higher variability in the earlier part of the figure reflects, of course, the smaller samples sizes used in estimation.

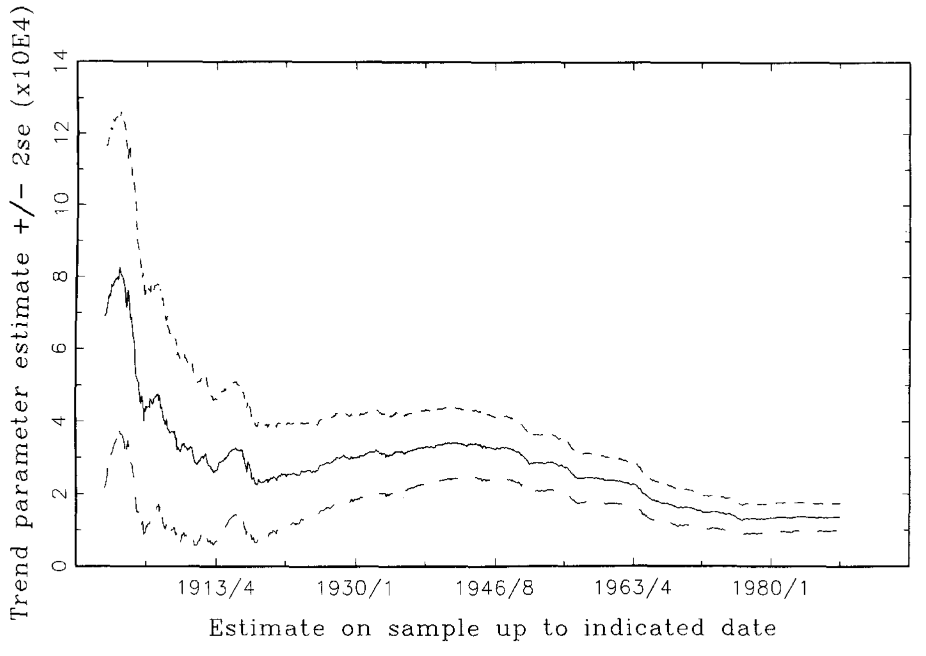

Fig. 4. Recursive trend parameter estimates 1900–1988 (Global sample, k = 4).

depends only on γ_0 since the γ_i, $i \geqslant 1$, appear only on differenced terms and so cancel out of the calculation. For the global sample, this corresponds to approximately 4.7×10^{-4} °C/month.

It is important to bear in mind a number of limitations of this calculation of the expected temperature increase over time. First, the estimate on data up to 1945 is over twice as large; some substantial variation in the estimate is present here (again, see Figure 4). Second, we treat the trend as independent of season, although there is evidence that the winter trend may exceed the summer trend (see, again, Vinnikov *et al.*, 1990). Third, the estimate is tied to the particular parametric (linear) model that we have imposed. A longer span of data may reveal that a non-linear trend will be a better forecaster of future increases in average temperature. (These qualifications also apply, in the latter case even more forcefully, to the standard estimate from the simple trend-only model).

4. Conclusion

The statistical literature on the detection of trends in time series makes an important distinction between a series containing a stochastic trend, for which permanent changes in the distribution function depend upon realizations of a random variable, and one containing a deterministic trend, for which the evolution of the unconditional distribution function over time is predictable. The application of tests for such trends requires, in many instances, the use of non-standard distribution functions for test statistics.

In the case of global temperature data, we *do not* find evidence of a stochastic trend in the form of a process with a root exceeding unity in an autoregressive lag polynomial. This result allows us to apply a straightforward test for a linear deterministic trend, and in so doing to extend the earlier results of Solow (1987). We do find evidence of a deterministic trend which can be approximated by a linear term. This finding of a statistically significant trend may be of particular significance in light of the well-known observation that the general upward movement in temperatures over the period 1890–1940 was followed by a period of no apparent increase (roughly 1940–1970; see Figures 1–3). Our result implies that the period of relative cooling falls within the range of sample fluctuation consistent with some significant positive trend.

While these results in themselves imply nothing about the link between CO_2 concentration and temperature, we note that an effect of greenhouse gases on temperature, if it were present, would not necessarily show up in the form of a *change* in trend over the sample examined here, but could instead imply a uniform trend over the entire period. While a linear trend seems to be a reasonable approximation to the trend on this sample, further research may eventually reveal that a non-linear deterministic trend will yield a better model of regular temperature change.

References

Dickey, D. A. and Fuller, W. A.: 1979, 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *J. Amer. Statist. Assoc.* **74**, 427–431.

Folland, C. K., Karl, T. R., and Vinnikov, K. Y. A.: 1990, 'Observed Climate Variations and Change', Chapter 7 in the Report of the Intergovernmental Panel on Climate Change, *Climate Change: the IPCC Scientific Assessment*, Cambridge University Press, Cambridge.

Fuller, W. A.: 1976, *Introduction to Statistical Time Series*, John Wiley, New York.

Gordon, A. H.: 1991, 'Global Warming as a Manifestation of a Random Walk', *J. Climate* **4**, 589–597.

Hansen, J. and Lebedeff, S.: 1987, 'Global Trends of Measured Surface Air Temperature', *J. Geophys. Res.* **92**, D11, 13, 345–13, 372.

Intergovernmental Panel on Climate Change: 1990, *Climate Change: The IPCC Scientific Assessment*, Cambridge University Press, Cambridge.

Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P. Heimann, M., Mook, W. G., and Roeloffzen, H.: 1989, 'A Three-Dimensional Model of Atmospheric CO_2 Transport Based on Observed Winds: 1. Analysis of Observational Data', in Peterson, P. H. (ed.), *Aspects of Climate Variability in the Pacific and the Western Americas*, American Geophysical Union, Washington D. C.

Kuo, C., Lindberg, C., and Thomson, D. J.: 1990, 'Coherence Established between Atmospheric Carbon Dioxide and Global Temperature', *Nature* **343**, 709–713.

Phillips, P. C. B.: 1986, 'Understanding Spurious Regressions in Econometrics', *J. Econometr.* **33**, 311–340.

Phillips, P. C. B.: 1987a, 'Time Series Regression with a Unit Root', *Econometrica* **55**, 277–301.

Phillips, P. C. B.: 1987b, 'Towards a Unified Asymptotic Theory of Autoregression', *Biometrika* **74**, 535–548.

Said, S. E. and Dickey, D. A.: 1984, 'Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order', *Biometrika* **71**, 599–607.

Solow, A. R.: 1987, 'Testing for Climate Change: An Application of the Two-Phase Regression Model', *J. Clim. Appl. Meteorol.* **26**, 1401–1405.

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

Tsonis, A. A. and Elsner, J. B.: 1989, 'Testing the Global Warming Hypothesis', *Geophys. Res. Letters* **16**, 795–797.

Vinnikov, K. Ya., Groisman, P. Ya., and Lugina, K. M.: 1990, 'Empirical Data on Contemporary Global Climate Changes (Temperature and Precipitation)', *J. Climate* **4**, 662–677.

Wigley, T. M. L. and Barnett, T. P.: 1990, 'Detection of the Greenhouse Effect in the Observations', Chapter 8 in the Report of the Intergovernmental Panel on Climate Change, *Climate Change: the IPCC Scientific Assessment*, Cambridge University Press, Cambridge.

Wigley, T. M. L. and Raper, S. C. B.: 1990, 'Natural Variability of the Climate System and Detection of the Greenhouse Effect', *Nature* **344**, 324–327.

Yule, G. U.: 1926, 'Why do we Sometimes get Nonsense-Correlations Between Time Series? A Study in Sampling and the Nature of Time-Series', *J. Roy. Statist. Soc.* **89**, 1–64.

(Received 21 December, 1990; in revised form 3 March, 1992)