ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Global Temperature Trends

Trevor Breusch and Farshid Vahid

March 2011

Working Paper 04/11

Global Temperature Trends

Trevor Breusch
Crawford School of Economics and Government
Australian National University
E-mail: Trevor.Breusch@anu.edu.au
and

Farshid Vahid
Department of Econometrics and Business Statistics
Faculty of Business and Economics
Monash University
E-mail: Farshid.Vahid@monash.edu

Updated with new data, March 2011

Abstract

Are global temperatures on a warming trend? It is difficult to be certain about trends when there is so much variation in the data and very high correlation from year to year. We investigate the question using statistical time series methods. Our analysis shows that the upward movement over the last 130-160 years is persistent and not explained by the high correlation, so it is best described as a trend. The warming trend becomes steeper after the mid-1970s, but there is no significant evidence for a break in trend in the late 1990s. Viewed from the perspective of 30 or 50 years ago, the temperatures recorded in most of the last decade lie above the confidence band of forecasts produced by a model that does not allow for a warming trend.

1 Introduction

Records show that temperatures have increased globally over the last 100 to 150 years. An interesting question is whether this rise is really an upward *trend*, that is a systematic or persistent tendency to move in the one direction. It appears that the warming trend has become steeper in the last 30 to 50 years or so, although it has also been suggested in the public debate that the trend has all but disappeared in the last decade.

This paper examines such questions using the tools of time series analysis. There are several series of global temperatures available from different sources, none more authoritative than the others. One labelled *T3GL* is compiled by the Climatic Research Unit at the University of East Anglia and the Hadley Centre of the UK Met Office. It is a global average of combined land and sea surface temperatures over widely dispersed locations, in a time series from 1850 to date, expressed as the deviation from the average of the period 1961-1990. These deviations are called "temperature anomalies", but we simply refer to them as "temperature data". We only look at the annual (Jan to Dec) average, although a monthly series is also available. The other two series we use are labelled *NCDC* and *LOTI*, which are temperature anomalies over land and ocean compiled by the National Climatic Data Center of the US Department of Commerce, and by the Goddard Institute for Space Studies at NASA respectively. The latter two series cover the shorter period from 1880 to date.[1] The plots of the three series are provided in Figure 1.[2] While the range of the series (i.e. their maxima and minima) may not be comparable because the averages on which the anomalies are based are not the same for all three series, it can be seen that their movements are quite similar.

We examine a temperature record as a time series of observations, but with scant regard for the nature and source of the data. Thus nothing in our analysis is informed by the science of climate change, nor by other theories that may explain the movements in the series, nor by other data such as fossil or geological records. We do use the fact that the records were made sequentially at annual intervals since 1850 or 1880. We also recognise that there are two interesting states to be considered: one in which the variable has a tendency to revert to its long-run average, perhaps after many years, and another where there is some persistent upward movement, possibly varying in intensity. A finding that confirms a warming

[1] Data and documentations for the three series *T3GL, NCDC and LOTI* are available at http://www.cru.uea.ac.uk/cru/data/temperature/, http://www.ncdc.noaa.gov/cmb-faq/anomalies.html and http://data.giss.nasa.gov/gistemp/ respectively. Our copies were downloaded on 28 February 2011.

[2] All figures and estimation results reported in this paper are produced using Eviews version 7. An Eviews program file containing all command lines that reproduce the estimated equations and graphs reported in this paper is available from the authors upon request.

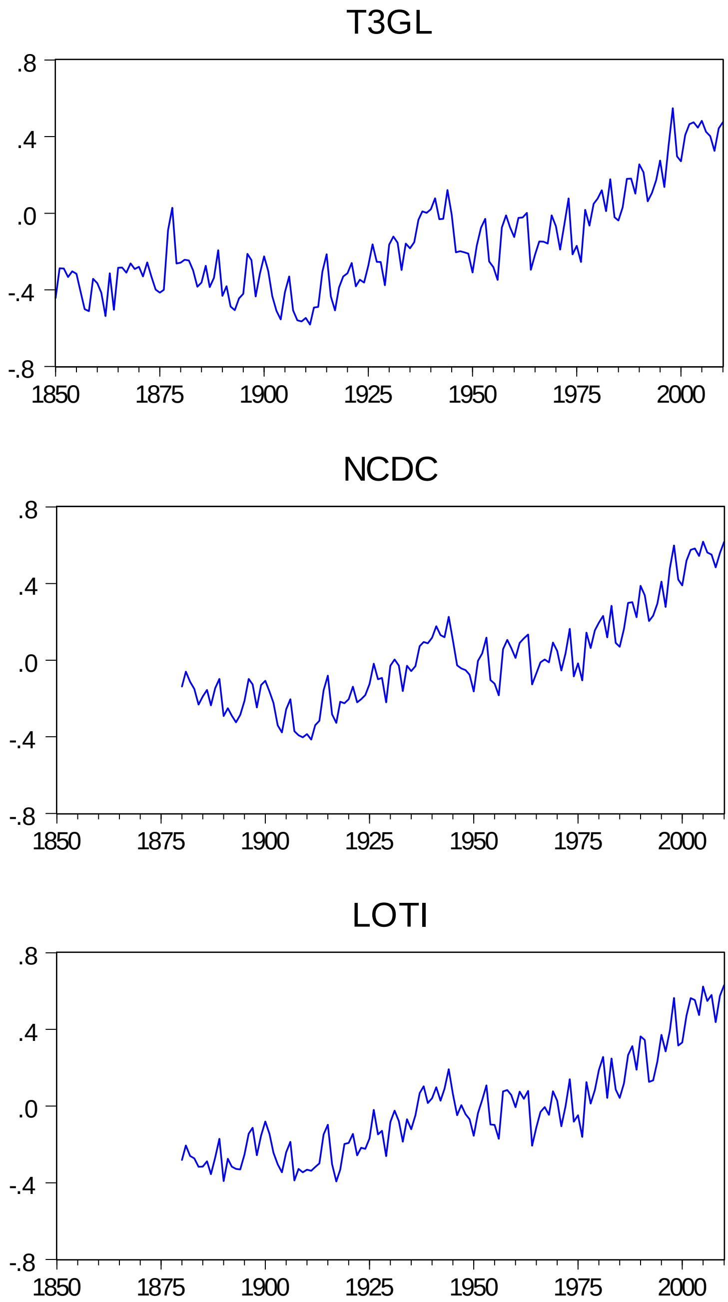

Figure 1: Global Temperature Anomalies

trend may be cause to seek an external explanation, but this investigation of the existence of a warming trend relies on the temperature data alone.

2 Is there a warming trend in the data?

2.1 Interpreting the question

A seemingly simple question such as "Is there a warming trend in global temperature data?" may be clear enough in ordinary discourse, but the analyst who does not have any knowledge of possible underlying mechanisms that govern the dynamics of temperatures and needs to formulate this question in terms of an hypothesis about a purely statistical model of the data would find this question ambiguous. Technically, trend is a periodic component that takes infinite time to complete one revolution. This definition makes two things apparent. First, it is much harder to learn about trends than to estimate periodic components that repeat themselves frequently, say every 4 to 8 years, from a finite sample of observations. This is because a modest size sample may contain several re-occurrences of 4 to 8 year cycles, but a finite sample can only contain one observation of the trend – and moreover an incomplete observation at that! Second, it is impossible to distinguish trend from a periodic component of large but finite period based on a finite sample. For example, with only 150 observations, cycles that take 300 or more years to complete are indistinguishable from trend.

The technical definition of a trend as a cycle with infinite period has another implication which highlights its difference with our ordinary understanding of a trend. A time series with wandering behaviour, like a random-walk, which at each point has equal probability of going up or going down, also satisfies the technical definition of a trend. In fact, such a component is often called a "stochastic trend." However, a variable that is equally likely to go above its current level and stay above it for a long period, or to go below its current level and stay below it for a long period is not deemed to have a trend in the ordinary sense. In particular, in the context of the temperature data given that the question is concerned with "a *warming* trend," it is obvious that a purely wandering behaviour around the current level is not of interest. Hence, we interpret the question as "Is there a systematic or persistent tendency in the temperature to move above its current level at every period?"

2.2 Towards answering the question

2.2.1 Simple linear trend models

If we start with the simplest model of a trend as a straight line on this untransformed scale, we could fit regression equations to the series to see if there are statistically significant trends. The estimated equations for the three series are (with standard errors in parentheses below the estimated parameters):

$$\begin{aligned} T3GL_t &= \underset{(0.144)}{-0.651} + \underset{(0.0014)}{0.0058}t + u_t \\ u_t &= \underset{(0.062)}{0.606}u_{t-1} + \underset{(0.063)}{0.249}u_{t-4} + \varepsilon_t \\ R^2 &= 0.85, \quad \hat{\sigma} = 0.103, \quad \hat{\lambda}_1 = 0.92 \end{aligned} \tag{1}$$

$$\begin{aligned} NCDC_t &= \underset{(0.125)}{-0.513} + \underset{(0.0014)}{0.0075}t + u_t \\ u_t &= \underset{(0.071)}{0.585}u_{t-1} + \underset{(0.069)}{0.243}u_{t-4} + \varepsilon_t \\ R^2 &= 0.89, \quad \hat{\sigma} = 0.088, \quad \hat{\lambda}_1 = 0.91 \end{aligned} \tag{2}$$

$$\begin{aligned} LOTI_t &= \underset{(0.108)}{-0.502} + \underset{(0.0013)}{0.0072}t + u_t \\ u_t &= \underset{(0.075)}{0.502}u_{t-1} + \underset{(0.076)}{0.296}u_{t-4} + \varepsilon_t \\ R^2 &= 0.87, \quad \hat{\sigma} = 0.095, \quad \hat{\lambda}_1 = 0.90 \end{aligned} \tag{3}$$

Interestingly, a model comprising a linear trend plus an autoregressive cyclical component with the same lag structure is chosen for all three temperature series by the methodology adopted here. This methodology, a version of the procedure of Hannan and Rissanen (1982), starts with finding a pure autoregressive model for the cyclical component of the series that can whiten the correlogram of the residuals. Then it considers all possible autoregressive and moving-average models with less or equal lag structure than the initial autoregressive model, using a model selection criterion that penalises large models. Here we use the Schwarz criterion or BIC for this purpose. Inspection of correlograms and histograms of the residuals and formal specification tests such as a Breusch-Godfrey serial correlation test, a Jarque-Bera normality test and a Ramsey linearity test all support these models. In all three estimated models, the coefficient of linear trend is positive and statistically significant, thus indicating a warming trend. Such models of a stationary autoregressive process about a linear trend are called "trend stationary."

However, the statistic $\hat{\lambda}_1$ reported below each of the three equations suggests that there may be a problem. This statistic is the largest inverted root of the estimated autoregressive process and it is a measure of persistence in the deviations of each series from its linear trend. The closer to one $\hat{\lambda}_1$ is, the more persistent these deviations become. The extreme case of wandering behaviour corresponds to this root being equal to one, which is known as the series having a "unit root". Since the estimate of this parameter from a finite sample is biased downwards, it may well be that the above temperature series each has a unit root. Very strong dependence from one period to the next implied by a unit root has the effect of letting the series wander far and wide, with diversions that can look like trends at least for a short time. This means that if the series indeed has a unit root, but we have modelled it as stationary deviations from a linear trend, the confidence implied in the estimated models above about the linear trend may indeed be false.

One may ask if we could rule out wandering behaviour (i.e. unit roots) in temperatures on a-priori grounds. After all, it is unsettling to think about temperatures the same way we think about stock prices. However, if we remember that a unit root process is observationally equivalent to a process with a dominant stochastic cycle with periodicity that exceeds the span of our sample, then a unit root is just a proxy for long cycles. It is a well-established principle in time series analysis that statistical inference is more accurate if we approximate a near unit root by a unit root than by a stationary process. Looking at it this way, a unit root model is just a vehicle for accounting for the strong persistence in the series and getting better measures of confidence.

2.2.2 Can unit root tests give us some information?

In a series x_t that has a unit root, its *differences* $\Delta x_t = x_t - x_{t-1}$ follow a stationary process. Hence a process with a unit root is also called a *difference-stationary* or an *integrated* process. In the 1980s several statistical tests were designed to distinguish a difference-stationary from a trend-stationary process. A good summary of these tests with some discussions that are particularly relevant for our analysis is in Stock (1994). If it is found that apparent trend can be ascribed to a unit root, then what seems to be trend might be just high-correlation wandering that is just as likely to go down as go up. An external explanation of the apparent trend would no longer be needed.

Most of the available unit root tests consider the null hypothesis of a unit root against a stationary or a trend stationary alternative. Given our discussions above about the possibility of observational similarity of a unit root process and a process with a deterministic trend in finite samples, it is not surprising to know that the finite sample properties of unit root tests are poor. The performance of these tests depend crucially on the type of trend and the period of cycles in the data. Stock (1994) emphasises the importance of properly specifying the deterministic trends before proceeding with unit root tests, and advises that "this is an area which one should bring economic theory to bear to the maximum extent possible." In our context, the responsibility falls on the shoulder of climate theory rather than economic theory, an area that we know nothing about. Here, we proceed with simple assumptions about trends.

The available unit root tests differ by their treatment of residual serial correlation at higher frequency. In the absence of an assumed trend, both the Dickey-Fuller test and the Phillips-Perron test indicate strongly the presence of a unit root, as might be expected from the data plot. However when a linear trend is allowed, these tests give conflicting results. The augmented Dickey-Fuller test indicates that a unit root cannot be rejected at any of the usual decision levels (with 3 or 4 augmentation lags). The GLS version of the Dickey-Fuller test agrees. In contrast, the Phillips-Perron test for the same situation rejects the unit root in the presence of a trend (with 3 or 4 lags). The results are the same for all three series. This apparent conflict among unit root tests may be attributed to the severe size distortion of the Phillips-Perron test in the presence of moderate negative MA roots, also discussed in Stock (1994).

The question we are trying to answer though is not about a unit root in the temperature data, it is about a tendency of the data to drift upwards. Hence, the unit root tests by themselves do not answer our question. If we trust the Phillips-Perron test, then we can trust equations (1), (2) and (3), which clearly show a positive and highly significant trend in the temperature data. However, if we dismiss the result of the Phillips-Perron test because of its size distortion in finite samples and trust the result of the augmented Dickey-Fuller test, the presence of a unit root does not exclude the possibility that there may be a deterministic trend in the data as well. So we need to do further analysis.

2.2.3 Controlling for extremely high persistence, is there a warming trend?

For a difference-stationary process, the absence or presence of a deterministic linear trend is a question about the mean of the differenced series. A non-zero mean in a differenced series becomes a constant amount of drift added to the series each period, i.e. a deterministic trend component. This can be seen algebraically because we can write a difference stationary process as $\Delta x_t = \alpha + z_t$ where α is the mean of Δx_t and z_t is a zero mean stationary process, and this implies

$$\begin{aligned}
\Delta x_t &= \alpha + z_t \Longrightarrow x_t = x_{t-1} + \alpha + z_t \\
&\Longrightarrow x_t = x_{t-2} + \alpha + z_{t-1} + \alpha + z_t \\
&\quad \dots \\
&\Longrightarrow x_t = x_0 + \alpha t + \sum_{j=1}^{t} z_j \ ,
\end{aligned}$$

which shows that x_t has a deterministic trend with coefficient α, the mean of Δx_t. The last term in the expression for x_t is the integrated component that embodies the unit root in x_t. Therefore to answer the question about the warming trend we need to answer if the mean of the first difference of the temperature series is zero or

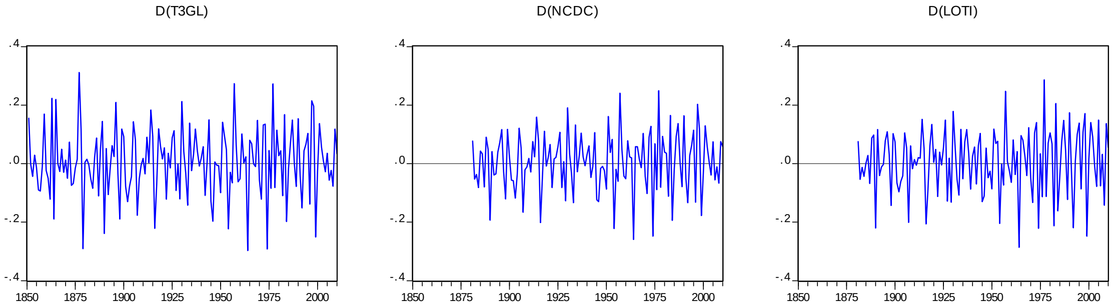

Figure 2: First differences of temperature series

not. A natural approach is to build time series models for the serial correlation in the differenced series, and to consider the significant of intercepts in those models. First differences of the three temperature series are plotted in Figure 2.

Again, using the Hannan-Rissanen method, we find that an ARMA(0,2) fits all three differenced series. This implies ARIMA(0,1,2) models for the actual series. The estimated models are (with standard errors in parentheses below the estimated parameters):

$$\Delta T3GL_t = \underset{(0.0028)}{0.0051} + \varepsilon_t \underset{(0.077)}{-0.396}\,\varepsilon_{t-1} - \underset{(0.077)}{0.274}\varepsilon_{t-2}\ , \qquad \hat{\sigma} = 0.104 \qquad (4)$$

$$\Delta NCDC_t = \underset{(0.0032)}{0.0057} + \varepsilon_t \underset{(0.086)}{-0.385}\,\varepsilon_{t-1} - \underset{(0.086)}{0.224}\varepsilon_{t-2}\ , \qquad \hat{\sigma} = 0.091 \qquad (5)$$

$$\Delta LOTI_t = \underset{(0.0025)}{0.0067} + \varepsilon_t \underset{(0.087)}{-0.507}\,\varepsilon_{t-1} - \underset{(0.087)}{0.202}\varepsilon_{t-2}\ , \qquad \hat{\sigma} = 0.096 \qquad (6)$$

Diagnostic tests for linearity, adequacy of dynamics and normality support the three equations. Comparing these results with those in equations (1), (2) and (3), we see that the evidence for a positive linear trend is much weaker in the presence of a unit root. While the t-statistics of the coefficient of the trend in equations (1), (2) and (3) were all over 4, implying a highly significant and precisely estimated positive trend, the t-statistics for the drift terms (the intercepts) in equations (4), (5) and (6) are 1.83,1.78 and 2.65 respectively. While these smaller values show that there is not as much information about the linear trend when a unit root is assumed, they are sufficient to reject the null hypothesis of zero drift against a one-sided alternative of a positive drift at the 5% level of significance. Thus the data indicate the existence of a warming trend also in this context. It is important to note that, even if the drift was found to be zero, the unit root in this model implies wandering (non-stationary) behaviour with no tendency to revert to average or typical temperatures in the long run.

3 Has the warming trend become steeper?

With our lack of knowledge about the underlying mechanisms to explain trends in temperatures, and with the generic difficulty of uncovering a trend from a finite sample of observations, the question of determining breaks in the trend becomes an enormously harder question to answer. This is because it will involve not only uncovering trends from smaller segments of the sample, but also comparing these trends and determining if they are the same or statistically different. Moreover, if distinguishing unit roots from linear trends was difficult because unit roots could sometimes behave similar to a linear trend in a finite sample, distinguishing them from a piecewise linear trend is much more difficult.

Even if one assumes that the trend in data is piecewise linear and continuous, testing the hypothesis that the trend has become steeper since a specific year that is chosen *after looking at the data* is not as straightforward as performing a simple t-test or an F-test of significance of a parameter or a set of parameters (sometimes called a break-point test or a Chow test). We explain this problem below.

3.1 The statistical significance of data-dependent questions

The statistical theory of hypothesis testing is based on the assumption that the question that is confronted with the data (i.e. the null hypothesis) is formulated independently of the test sample. If the question is determined endogenously from the test sample, then that sample has no independent information to validate or reject the hypothesis. For example, if after looking at the results of equation (1) one asks "Is the coefficient of trend 0.0058?" obviously we cannot use this sample to test this hypothesis with a t-test. A similar problem arises when we look at the plots in Figure **??** and then ask questions such as: "Was there a cooling trend between 1880 and 1910?" or "Was there a warming trend between 1910 and 1945?" or "Was there a cooling trend between 1945 and 1975?" or "Has the trend become steeper since 1975?" or "Has the trend disappeared since the extremely hot year of 1998?". The standard statistical tests will overstate the apparent significance of these hypothesised events because these questions are not posed exogenously, as assumed by the statistical theory, but are instead endogenous results of the local behaviour of the very data that are used for testing them. Endogenous questions arise often in practice. For example, the question of "Are business cycles dead?" became popular a couple years ago among economists precisely because of a 15 year experience of positive growth. It is natural to ask such questions and they may lead us to new discoveries. But to test the statistical significance of such hypothesis with standard statistical procedures using the same data that gave rise to the question leads to incorrect inference. In particular, when a time series has extremely high persistence, getting too excited about a run of movements in one direction and extrapolating that into the future can be dangerous.

3.2 Data-determined trend breaks

Rather than peeking at the data and incorrectly evaluating a question about trend as if the question was exogenous (as assumed in the standard assessments of statistical significance) we might ask: When does the most remarkable break in the trend in this series appear to occur? Can we determine the statistical significance of this break taking into account that it has been data-determined?

The simplest approach would be to fit a model that accounts for the serial correlation while allowing one break point where the trend can change. This approach could be adopted with a model that is stationary around the deterministic trend, such as models (1)-(3) with an additional trend variable to allow the trend coefficient to change. Alternatively, we could specify a model with a unit root component, such as models (4)-(6) with an additional dummy variable to indicate the break point. In either case we can search over all possible break points (excluding the extreme ends of the series) to see where a break might best be located. One criterion might be the model fit, as measured by the value of the maximized likelihood. In this situation where every model has the same degree of parameterization, this approach is equivalent to choosing the result by minimum AIC or BIC, or by most any other selection criterion. Another criterion for judging the most remarkable break point is the magnitude of the usual test statistic (t or F or asymptotic χ^2) for testing the statistical significance of the coefficient of the variable that indicates the break.

Irrespective of the criterion used to judge the break point, and for all three of the data series, the most remarkable break point in the trend stationary models is in the mid-1970s. The next most notable candidate for a single break in the trend, disjoint in time from that region in the 1970s, is located in 1909-1913. There is nothing to suggest that anything remarkable has happened around 1998. The evidence from the unit root models tells the same story. The magnitudes of the most extreme t-test statistics are shown in Table 1.

Data Set˙·.Model	Trend stationary	Unit root
T3GL	3.74	2.66
NCDC	3.39	2.68
LOTI	4.51	4.02

Table 1: Maximum values of the break point t-statistics

The exercise of searching for the most extreme evidence of a break in trend along the series shows that the usual criteria for assessing statistical significance do not apply. In the case of the *T3GL* series for instance, there are 141 estimation runs being examined (each taking the first four observations fixed and allowing a trend break at one year in the remaining 157 observations except where subsamples of eight or fewer observations would remain at either end of the sample). At one extreme, if we had 141 independent random draws of the test statistic, the probability that at least one such draw will exceed the usual 1% significance point is not 1%, but instead 76%. And at the nominal 5% level of significance, one would almost certainly find at least one draw to reject the no breaks hypothesis. Even with a nominal significance level of 0.1% the actual probability of finding some 'significant' extreme value is 13%. To preserve actual significance levels of 1% and 5% requires setting nominal significance levels of 0.007% and 0.036% respectively. In the case of the standard normal distribution used as an asymptotic approximation to the distribution of a t-statistic, the 5% significance level corresponds to a two-sided critical value of 3.57 (not the nominal value 1.96) and the 1% level corresponds to a two-sided critical value of 3.97 (not the nominal value 2.58).

The actual test statistics obtained here for the individual tests are not independent because of the overlapping samples. The correct critical values will be somewhere between the nominal values and the adjusted values obtained by assuming independence. A better approximation to critical values in this case is given in Bai and Perron (2003), where the two-sided 10% critical value is 2.98, the 5% critical value is 3.21 and the 1% critical value is 3.66. Against these benchmarks, the extreme test outcomes in 1974-1978 as shown in Table 1 would be judged statistically significant breaks in trend if we assumed that the series were stationary around their piecewise linear trends. (The estimated trend function for the *T3GL* series is graphed in Figure 3.) However, if we assumed that the series had unit roots, only the break in the *LOTI* series would be judged significant. In either of the trend stationary or the unit root models, using the critical values reported in Bai and Perron (2003) to test the hypothesis of a single break against the alternative of two breaks, no other data-determined break in these series is judged to be statistically significant.

4 An alternative way to address the question of a warming trend

Another way to address the question of a warming trend is to put ourselves in the place of an analyst 50 years ago and produce confidence bands for the future given information available at that time from models that assume no warming trend. We can then ask whether the realised history in the last 50 years would fit within these bands. In particular, we ask if one assumed that there was no drift in the data, and we attributed all movement in the data to its strongly persistent

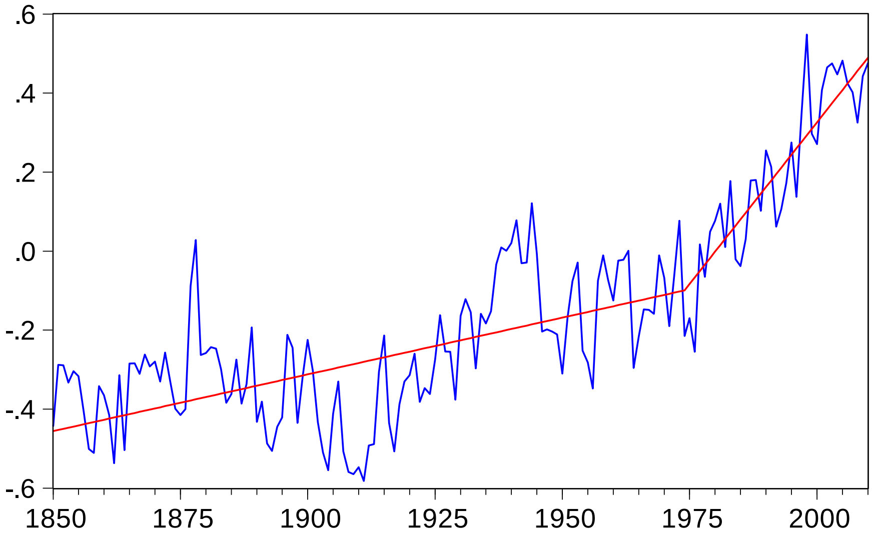

Figure 3: The most dramatic data-determined broken trend for the *T3GL* series based on the trend-stationary model

stochastic component (an assumption that is not rejected by the pre-1961 data for all three series), would the realised temperatures in 1961-2010 be considered extremely unlikely and surprisingly hot? It is important to remember that while a stochastic process with a unit root can wander anywhere in infinite time, in any *finite* period of time its movements relative to its starting point are bound by a probability law. This allows us to provide a 95% interval forecast for 1961-2010 and examine if the realised 1961-2010 temperature history fits within this band. Figure 4 presents this for the three series.[3] We can see that the realised temperatures in most of the past 10 years lie above the no-drift forecast intervals. This re-iterates the results of the previous section: there is sufficient evidence in all three temperature series to reject the hypothesis of no drift in favour of a warming trend in global temperatures. If we look at this question from the perspective of 30 years ago, we reach the same conclusion.

[3]The confidence bands in these graphs are calculated using the "dynamic forecast" option in Eviews and they do not incorporate estimation uncertainty.

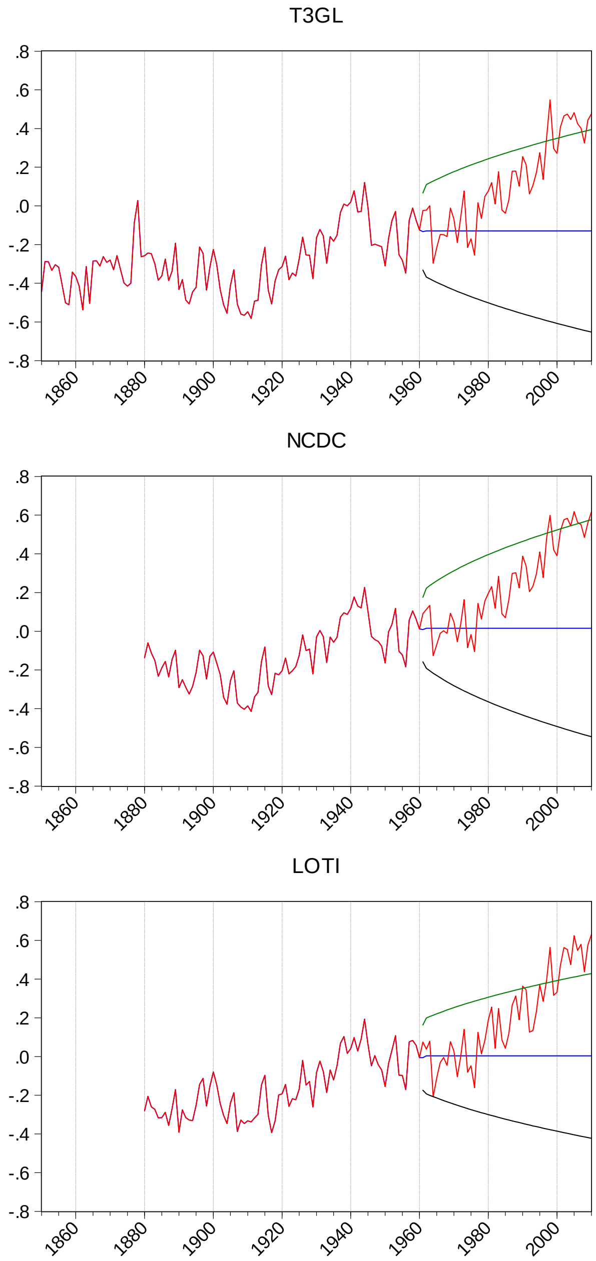

Figure 4: Interval forecasts for 1961-2010 allowing for high persistence but no drift and the realised values

5 Conclusion

We conclude that there is sufficient statistical evidence in the temperature data of the past 130-160 years to conclude that global average temperatures have been on a warming trend. The evidence of a warming trend is present in all three of the temperature series. Although we have used unit roots and linear trends as a coordinate system to approximate the high persistence and the drift in the data in order to answer the questions, we do not claim that we have uncovered the nature of the trend in the temperature data. There are many mechanisms that can generate trends and linear trends are only a first order approximation (see Granger 1988). It is impossible to uncover detailed trend patterns from such temperature records without corroborating data from other sources and close knowledge of the underlying climate system.

References

Bai, J. and Perron, P. (2003), "Critical values for multiple structural change tests," *Econometrics Journal*, 6, 72-78.

Granger, C.W.J. (1988), "Models that generate trends," *Journals of Time Series Analysis*, 9, 329-343.

Hannan, E.J. and Rissanen, J. (1982), "Recursive estimation of autoregressive-moving average order," *Biometrika*, 69, 81-94.

Stock (1994), "Unit roots, structural breaks and trends," Chapter 46 in R.F. Engle and D.L. McFadden (eds) *Handbook of Econometrics, Volume IV*, Elsevier.