Inference About Trends in Temperature Data After Controlling for Serial Correlation and Heteroskedastic Variance

Ross McKitrick

Associate Professor
Department of Economics,
The University of Guelph,
Guelph ON Canada N1G 2W1.
e-mail: rmckitri@uoguelph.ca
Telephone: (519) 824-4120 x2532
Fax: (519) 763-8497

November 7, 2001

Abstract

Global surface temperature anomalies post-1958 are examined for evidence of significant trends. The routine application of a linear least-squares method is critiqued. Testing for model misspecification in the form of serial correlation and heteroskedasticity in the residuals leads to consideration of a piecewise trend with lagged anomalies and a linear model of the error variance. While the naï ve least-squares model suggests a highly significant trend post-1979, correcting the known specification errors leads to the finding of a reduced trend magnitude and evidence that the trend is insignificant, i.e. within the bounds of random noise. This points to the need for careful treatment of time series variables in order to avoid spurious inference.

Acknowledgments

I thank Kirill Kondratiev for his invitation for me to address this topic.

Inference About Trends in Temperature Data After Controlling for Serial Correlation and Heteroskedastic Variance

1. Introduction: Trends as a Statistical Model

Denote a time series of temperature anomalies with a_t. We are often interested in seeing if the mean of a_t is evolving over time. The wide availability of statistical software has made the basic linear trend regression:

$$a_t = b_0 + b_1 t + e_t \tag{1}$$

a popular technique, but it is subject to much abuse. We must always remember that it is a statistical model and the results printed by a computer package are based on assumptions which the researcher is expected to know. It is not uncommon to see the parameters of (1) estimated using a technique like ordinary least squares (OLS) then conclusions drawn about the "significance" of the estimated slope parameter $\hat{b}_1$ based on the printed *t*- or *F*-statistics. Such inferences are only valid if (1) satisfies some restrictive conditions which in the case of time series data rarely hold. The statistical model takes the form $a_t - b_0 - b_1 t = e_t$ and the printed *t* and *F* statistics take for granted that e_t is a random variable independently drawn from a Gaussian distribution with mean zero and a constant variance s^2. In effect, this assumes that there is no relationship among the a_t's between adjacent periods and that the variance of the unexplained portion of a_t remains constant over the whole sample. These assumptions can be tested. The OLS procedure generates an estimated standard error for $\hat{b}_1$, here denoted $\hat{s}_1$ and if the assumptions of the statistical model hold true the ratio $\hat{b}_1 / \hat{s}_1$ follows an exact *t* distribution and inference about significance can be made using standard tables. If the assumptions about the distribution of e_t are violated then the ratio $\hat{b}_1/\hat{s}_1$ does not follow a *t* distribution except asymptotically, hence for finite samples *t* tables are not guaranteed to be reliable for inference about significance. Furthermore, since *t* is not stationary, the finite sample bias of $\hat{b}_1$ converges to zero at a different rate than that of $\hat{b}_0$, making application of standard asymptotic statistical methods complex in a trend model. On these points see Hamilton (1994, ch. 16).

If model assumptions are found not to hold some attention must be paid to the reasons why. There are purely statistical corrections that can be used to isolate a Gaussian component of e_t, allowing inferences about $\hat{b}_1/\hat{s}_1$. The term for this class of models is *Generalized Least Squares* (GLS). But the non-normality of the residuals often points to an error in the specification of the trend equation. A time series model that makes use of information about the physical process being modeled, and which yields independent and/or homoskedastic residuals without applying special statistical corrections, would be considered a better procedure for detecting any trends in the data.

In the sections to follow we discuss issues in the specification of a trend model. After considering what the term "trend" means we look at the kinds of tests that must be applied to justify a model like (1). Since this equation usually does not apply to temperature anomaly data we discuss alternate strategies and then show how they affect empirical estimates of the magnitude and significance of the trend coefficient in surface and tropospheric anomaly data. It turns out that the better the fit of the trend model, and the closer we get to Gaussian residuals, the smaller is the estimated trend and the less significant is the coefficient.

2. Trends and Other Changes in the Mean

According to equation (1), a trend implies a steady pattern of increase in the conditional mean of a_t by the amount b_1 each period. In other words, using equation (1) to summarize the evolution of the mean forces upon the data the interpretation that there is a continuing tendency for the mean to rise by b_1 each period. In Figure 1 there are some examples of problematic data configurations that might lead to detection of a false “trend” (shown by the heavy dashed line) based on a naï ve application of (1). The step function in panel A is not a trend in the sense that there is not a continuing, steady pattern of increase in the mean of a_t. Instead the variable is constant over time, with a single step arising from a permanent one-time change in the mean. The data in panel B show a decreasing tendency interrupted by a single upward step. What, then, is the “trend”? It is reasonable to argue that the trend is the downward slope, since that is the continuous tendency, and the step is a one-time shock, which changes the mean of the data series but not the tendency of the mean to decline over time. In panel C the solid part represents the data for which we have observations, and the dotted line represents the unobserved past and future. Again model (1) would imply an upward trend, but the data are more properly interpreted as cyclical.

When discussing trends in temperature anomaly data it is usually against the background of concern about enhanced radiative forcing due to carbon dioxide concentrations. Since these have been smoothly trending upward since the measurements became available from the Mauna Loa observatory the detection of a smooth upward trend in the global average temperature anomaly would be suggestive of a relationship between temperature and CO_2 concentrations. If temperatures follow a pattern like that shown in Figure 1B this would not suggest a relationship between the two, since the smooth portion of the trend is downwards and the upward movement is a single large event, not a continuous sequence of small events. If we had an alternative explanation for the discontinuity in temperatures (such as the changing mode of a major internal climate oscillation) this effect should be removed when estimating the trend magnitude. The simplest way to do so is to break the sample up into “before” and “after” segments.

In terms of global temperature data, the 1978 Pacific Climate Shift has been identified as an important discontinuity that produced an upward step in many temperature indicators (Ebbesmeyer *et. al.* 1991). The event was driven by a change in the circulation pattern of the Pacific ocean, not by a change in the optical depth of the atmosphere. However it may create the appearance of a trend if the data are not properly analysed. Table 2.3 in the Third Assessment Report of Working Group I of the Intergovernmental

Panel on Climate Change (IPCC 2001 p. 122) shows a group of trend estimators for surface and atmospheric temperatures in which the trends are separately computed over 1958-2000, 1958-1978 and 1979-2000. The lower tropospheric weather balloon record produced by the US National Oceanic and Atmospheric Administration (NOAA) has a cooling trend of –0.08 ^{o}C per decade from 1958 to 1978 and a continued cooling trend of –0.03 ^{o}C per decade from 1979 to 2000. Yet they find a warming trend over the entire 1958 to 2000 period of +0.07 ^{o}C per decade! Similarly the combined balloon-satellite record produced by the UK Met Office (denoted UKMO) has a trend from 1958 to 1978 of –0.03 ^{o}C per decade and from 1979 to 2000 of +0.03 ^{o}C per decade. Yet over the whole sample the estimated "trend" is +0.11 ^{o}C per decade. Clearly the lower tropospheric data follow a pattern like that in Figure 1A or B. The event in 1979 is likely the Pacific Climate Shift and the "trend" in the data is zero or negative.

Unfortunately this crucial information is only available to a careful reader of the Table in the body of the report. It is left out of the Summary for Policy Makers, which says only: "Since the late 1950s (the period of adequate observations from weather balloons) the overall global temperature increases in the lowest 8 kilometres of the atmosphere and in surface temperatures have been similar at 0.1 ^{o}C per decade." (p.4) The statement is, to say the least, misleading. The data show a tropospheric cooling trend interrupted by discrete step in 1978. Clearly one must be careful in defining what is meant by a "trend" in data that follows patterns as in Figure 1.

2. Statistical Corrections for Autocorrelated Errors and Heteroskedastic Variances

If (1) is applied to the data in panels A, B or C, the residual e_t will no longer be independent over time. The realized values will follow a systematic pattern represented by the vertical distance between the data line and the heavy dashed "trend" line. Across most of the data set, an observation of one period's value of e_t plus a Gaussian noise term e_t could provide a good predictor of e_{t+1}, using $r e_t + e_t$, where 0< r <1. Therefore an improvement to (1) is the GLS model

$$a_t = b_0 + b_1 t + r e_{t-1} + e_t \qquad (2)$$

where $e_{t-1} = a_{t-1} - b_0 - b_1(t-1)$ and e_t is assumed to be independent and identically distributed with a zero mean and constant variance. If (2) is the "true" model and 0< r <1, then use of (1) will generate variance estimates on $\hat{b}_1$ that are biased too small, creating a tendency towards a false finding of "significance" in the trend (Davidson and MacKinnon 1993 Ch. 10). Model (2) is referred to in econometrics as an AR(1) model, denoting the first-order autoregressive process which replaces e_t from (1).

A test of (2) against the alternative (1) can be carried out using a Lagrange Multiplier (LM) test. The residuals from one model are regressed on the right hand side variables plus themselves lagged once. A *t*-statistic on the lagged residuals is an asymptotic test of the null hypothesis that there is no serial correlation of residuals. For instance, if (1) is estimated, yielding residuals $\hat{e}_t$, then from the regression

$$\hat{e}_t = f_0 + f_1 t + f_2 \hat{e}_{t-1} \qquad (3)$$

the *t*-statistic on the OLS estimator $\hat{f}_2$ tests the null hypothesis that the residuals are not serially correlated.

Another generalization of (1) can be used to treat the problem of autoregressive-conditional heteroskedasticity (ARCH) in the residuals e_t. ARCH implies that the variance of the error terms changes over the sample, which violates another of the assumptions on which inference in the linear trend model is based. A test for ARCH processes in the error terms uses the Gauss-Newton Regression (GNR) procedure (Davidson and MacKinnon 1993). The residuals from (1) are squared then regressed on themselves lagged g times:

$$\hat{e}_t^2 = q_0 + q_1\hat{e}_{t-1}^2 + \cdots + q_g\hat{e}_{t-g}^2 \qquad (4)$$

where *g* represents the order of the ARCH process. An *F* test of the joint significance of the estimated q_i parameters is a test of the absence of ARCH in the error process. If the residuals of (1) do contain an ARCH process, the variance estimator of OLS is biased and *t* tests again are potentially unreliable for determining the significance of the trend.

GLS corrects statistical problems related to the fact that (1) fails to use all the information contained in the residuals to guide inference about $\hat{b}_1$. But detection of autoregressive and/or ARCH errors often points to a specification error, in this case that a linear trend model imposes the wrong structure on the variable being analysed. If this is true, application of an improved specification may solve the problem of autocorrelated residuals and/or ARCH errors without resorting to GLS.

3. Alternate Time-Series Specifications

The atmospheric system that gives rise to globally-averaged anomalies is subject to periodic shocks denoted v_t that integrate over time and have impacts distributed across multiple periods. There may also be a tendency for the mean to evolve over time according to some function $f(t)$. Therefore we can use a model in which a weighted sum of present and future monthly anomalies follows $f(t)$+ v_t, which we write as $\Sigma_p y_i a_{t+i} = v_t + f(t)$. Using a linear specification of *f* and rearranging slightly we can get the alternate form

$$(a_t - f_1 a_{t-1} - \cdots - f_p a_{t-p}) = m_0 + m_1 t + e_t \qquad (5)$$

where $e_t = v_t / y_0$. In this case an autoregressive (AR) process in a_t with *p* lags is driven by a linear trend plus a constant (which may be zero) plus a Gaussian error. The instantaneous effect of *t* on a_t is given by the estimate $\hat{m}_1$ while the long-run trend ($d\Sigma a_{t+i} / dt$ for $i \geq 0$) is given by

$$\hat{m}_1 /(1 - \Sigma_p \hat{f}_i) . \tag{6}$$

If we have reason to believe that the process generating the temperature anomalies has undergone a structural change we can replace $m_1 t$ with a piecewise linear function ($dt + aDt$) where *D* is a dummy variable taking the value 0 in the period up to the structural break and 1 thereafter. Introducing this form allows the trend to differ between periods: in the first, where *D*=0 it is simply d and in the second where *D*=1 it is $d + a$. In addition the constant term can be augmented by replacing it with $m_0 + lD$ so that the constant in the second period is $m_0 + l$.

Combining these and rearranging slightly yields the augmented ARMA model, denoted ARMAX(*p,0*):

$$a_t = m_0 + lD + f_1 a_{t-1} + \cdots + f_p a_{t-p} + dt + aDt + e_t \qquad (7).$$

Note that with appropriate restrictions on parameters this reduces to (1), and these restrictions can be tested. It is also possible to test the residuals e_t from (7) for the presence of ARCH processes, and if need be apply corrections. This is discussed in the results section.

4. Empirical Results

Model (1) was estimated on global surface air temperature anomaly data from January 1959 to December 2000. The data were obtained from the UK Hadley Centre (2001). A graph of the data is in Figure 2. All estimations were done using the econometrics software SHAZAM (White 1993). Table 1 shows that a simple linear model yields a decadal trend of 0.11 ^{o}C/decade and the ratio of the trend coefficient to its OLS-estimated standard deviation is 19.6. If this ratio follows a *t* distribution it would be highly significant. Application of the LM test (equation 3) yields a *t* statistic of 26.5, decisively rejecting the null hypothesis of no serial correlation in the errors. Application of the GNR method (equation 4) with one lag on the squared residuals yields a *t*-statistic of 12.6, rejecting the null hypothesis of homoskedastic errors. Hence the *t*-test from (1) is unreliable in this case and the model must be respecified.

Adding a break point at January 1979 to make the time trend piecewise linear improved the fit of the model, raising the R^2 from 0.43 to 0.57 and the log-likelihood value from –2245 to –2169. As shown in Table 2, the pre-1979 trend is negative and apparently significant (–0.06 ^{o}C/decade) and the post-1979 trend is positive and apparently significant at 0.16 ^{o}C/decade with a *t*-statistic of 12.5. However the LM test of no serial correlation has a *t*-statistic of 20.8 and the GNR test of no homoskedasticity has a *t*-statistic of 9.3, indicating that the model remains misspecified and inference on the trend estimator is still unreliable.

Equation (7), the ARMA model with piecewise-linear trend, was then estimated using lags on a_t at 1, 2, 8, 24 and 27 months. These lags were chosen by gradually extending the lag length while deleting some insignificant terms and observing the Akaike Information Criterion (AIC). The AIC is $\ln(\hat{s}^2) + 2K/N$ where $\hat{s}^2$ is the variance of the model estimate, *K* is the number of estimated parameters and *N* is the sample length. The AIC declines as the fit improves and rises as more parameters are added or the sample declines, so a minimum-AIC rule is applied for model selection. The results are shown in Table 3. The R^2 value rises substantially to 0.80 and the log likelihood value rises from –2169.5 to –1876.1. The coefficient on the pre-1979 trend variable loses significance, however the post-1979 trend differential just remains significant at 5%. The instantaneous trend term has a small but apparently significant *t*-statistic

Testing the significance of the long-run trend term

$$\frac{d + a}{f_1 + f_2 + f_8 + f_{24} + f_{27}}$$

is not straightforward since $(d + a)/(\Sigma f_i) = 0$ is not a linear restriction. SHAZAM uses an asymptotic Wald statistic which yields a *P*-value based on comparison with c^2 tables. However the finite sample distribution of this statistic is not known, so this test may be unreliable. Chebychev's inequality (Rice 1988) yields an upper bound on the *P*-value which is valid for all sample sizes regardless of the distribution of the Wald statistic, so this is also reported. The *P*-value for the significance of the long-run trend is <0.001 according to the assumption of a c^2 distribution, and no higher than 0.084 regardless of the distribution.

The LM test now shows no serial correlation in the residuals (t = 0.90) but the GNR test shows at least first-order ARCH (t = 2.02). Since there is no serial correlation a GLS correction addressing only the heteroskedasticity can be applied. If e_t is distributed $N(0, h_t)$ where h_t is a function that varies over time, then the GLS correction involves replacing the residual from (7) with

$$e_t = \frac{e_t}{\sqrt{h_t}}. \qquad (8)$$

Then since $e_t \sim N(0,1)$ inference based on the parameters and their computed standard errors is valid. A number of specifications for h_t were examined, and the lowest AIC value was found for

$$h_t = w_0 + w_1 e_{t-1}^2 + w_2 e_{t-2}^2 + \Sigma_j b_j a_{t-j} \qquad (9)$$

where j = (1, 2, 8, 24, 27). This also eliminates all remaining ARCH. Time trend terms were never significant in (9).

Equation (7), augmented with (8) and (9) was estimated using a maximum likelihood routine. The results are in Table 4. The pre-1979 trend is insignificant, as are the post-1979 trend differential and the 1979-2000 instantaneous trend. The post-1979 long-run trend is 0.1261 ^{o}C/decade, and has a *P*-value of between 0.005 and 0.126. Since the trend coefficients in this model are insignificant and the evidence for the transformed trend is ambiguous, significance was also tested using a likelihood ratio (LR) statistic. The LR test is $-2(LLF_R - LLF_U)$ where LLF_R denotes the log likelihood function value with restrictions imposed ($d + a = 0$) and LLF_U denotes that in the original, unrestricted form. The statistic asymptotically follows a c^2 distribution with 2 degrees of freedom (the number of restrictions). The value of the test is 3.506 which has a *P* value of 0.173 (the 5% significance level is 5.991). The model with no time trend (but including a separate intercept for the post-1979 period) is not rejected against the alternative with trend.

5. Conclusions

The naïve model (1) yields an estimated decadal “trend” of 0.11 ^{o}C/decade in the surface data since 1958 which is considered highly significant on the basis of a *t* ratio of 19.58. However tests for independence and homoskedasticity of the residuals strongly reject, leading us to search for improvements in specification. A piecewise-linear model suggests a post-1979 trend of 0.16 ^{o}C/decade with a *t* ratio of 12.5, again highly significant. However, evidence of misspecification persists. Moving eventually to an ARMA model with correction for ARCH residuals resolves the specification problems, reduces the post-1979 measured surface warming rate to 0.13 ^{o}C/decade. A Wald test suggests significance may be maintained but the upper bound is in the insignificant range. A likelihood ratio test does not reject the hypothesis of no trend in the surface data pre- or post-1979. That is, any observed upward movement is consistent with random noise in the temperature data.

An important conclusion of this paper is that properties of time-series variables must be taken into account when testing for trends. Inference based on naï ve modeling strategies can easily lead to unreliable conclusions. In the case of surface temperatures analyzed here, the conventional methods based on OLS lead to high trend estimates and artificially small standard errors. Improved model specification and corrected statistical modeling yields a lower trend magnitude and evidence against statistical significance. This echoes the concerns of Zheng and Basher (1999) who have also warned against the dangers of erroneous trend detection based on improper use of time series climatological data.

References

Davidson, Russell and James MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford.

Ebbesmeyer, C.C., D.R. Cayan, D.R. McClain, F.H. Nichols, D.H. Peterson and K.T. Redmond. (1991). "1976 step in Pacific climate: Forty environmental changes between 1968-1975 and 1977-1984." In: J.L. Betancourt and V.L. Tharp, editors. *Proceedings of the 7th Annual Pacific Climate (PACLIM) Workshop*, April 1990. California Department of Water Resources. Interagency Ecological Study Program Technical Report 26. p.115-126.

Hamilton, James D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.

Hadley Centre (2001). "Global Monthly Temperatures." http://www.met-office.gov.uk/research/hadleycentre/CR_data/Monthly/HadCRUG.txt

Intergovernmental Panel on Climate Change (IPCC) Working Group 1 (2001). *Climate Change 2001: The Scientific Basis*. Cambridge: Cambridge University Press.

Rice, John A. (1988) *Mathematical Statistics and Data Analysis*. Belmont; Wadsworth and Brooks-Cole.

White, Ken (1993). SHAZAM User's Manual. Toronto: McGraw-Hill.

Zheng, Xiaogu and Reid E. Basher (1999). "Structural Time Series Models and Trend Detection in Global and Regional Temperature Series." *Journal of Climate* 12, 2347-2358.

Table 1: Empirical Estimates for Equation 1

$$a_t = b_0 + b_1 t + e_t$$

	Estimate (*t* statistic)	Standard Error
trend coefficient ($\hat{b}_1$)	**0.1089** (19.58)	0.0056
R-squared	0.4272	
Log-Likelihood Function	-2245.32	
LM Test for serial correlation (*t*)	**26.54**	
GNR Test for heteroskedasticity (*t*)	**12.63**	

Trend estimate is degrees C per decade

Table 2: Empirical Estimates for Equation 1 Augmented with a Piecewise-Linear Time Trend

$$a_t = b_0 + lD + dt + aDt + e_t$$

	Estimate (*t* statistic)	Standard Error
1958-78 trend coefficient (d)	**-0.0614** (4.39)	0.0140
1979-2000 trend differential (a)	**0.2262** (11.75)	0.0193
1979-2000 trend ($d + a$)	**0.1649** (12.47)	0.0132
R-squared	0.5731	
Log-Likelihood Function	-2169.46	
LM Test for serial correlation (*t*)	**20.81**	
GNR Test for heteroskedasticity (*t*)	**9.35**	

Trend estimates are degrees C per decade

Table 3: Empirical Estimates for Equation 7 (ARMA model)

$$a_t = m_0 + lD + f_1 a_{t-1} + \cdots + f_p a_{t-p} + dt + aDt + e_t$$

	Estimate (t statistic)	Standard Error
1958-78 trend coefficient (d)	-0.0110 (0.89)	0.0125
1979-2000 trend differential (a)	**0.0437** (1.96)	0.0223
1979-2000 instantaneous trend ($d + a$)	**0.0327** (2.14)	0.0153
1979-2000 long-run trend ($(d + a)/(1 - \Sigma f_i)$)	**0.1481**	
Wald *P*-value on long-run trend	0.000	
Chebyshev *P*-value on long-run trend	0.084	
R-squared	0.8028	
Log-Likelihood Function	-1876.05	
LM Test for serial correlation (t)	0.90	
GNR Test for heteroskedasticity (t)	**2.02**	

Trend estimates are degrees C per decade.
Estimates of f_i are not shown.

Table 4: Empirical Estimates for Equations 7-9 (ARMA + Heteroskedasticity Model)

$$a_t = m_0 + lD + f_1 a_{t-1} + \cdots + f_p a_{t-p} + dt + aDt + e_t / \sqrt{h_t}$$

	Estimate (t statistic)	Standard Error
1958-78 trend coefficient (d)	-0.0111	0.0117
	(0.95)	
1979-2000 trend differential (a)	0.0382	0.0210
	(1.82)	
1979-2000 instantaneous trend ($d + a$)	0.0271	0.0.0146
	(1.85)	
1979-2000 long-run trend ($(d + a)/(1 - \Sigma f_i)$)	**0.1261**	
Wald *P*-value on long-run trend	0.005	
Chebyshev *P*-value on long-run trend	0.126	
Likelihood Ratio Test (*P*-value) on trend coefficients	0.173	
R-squared	0.8019	
Log-Likelihood Function	-1861.33	

Trend estimates are degrees C per decade.
Estimates of f_i, w_i and b_i are not shown.

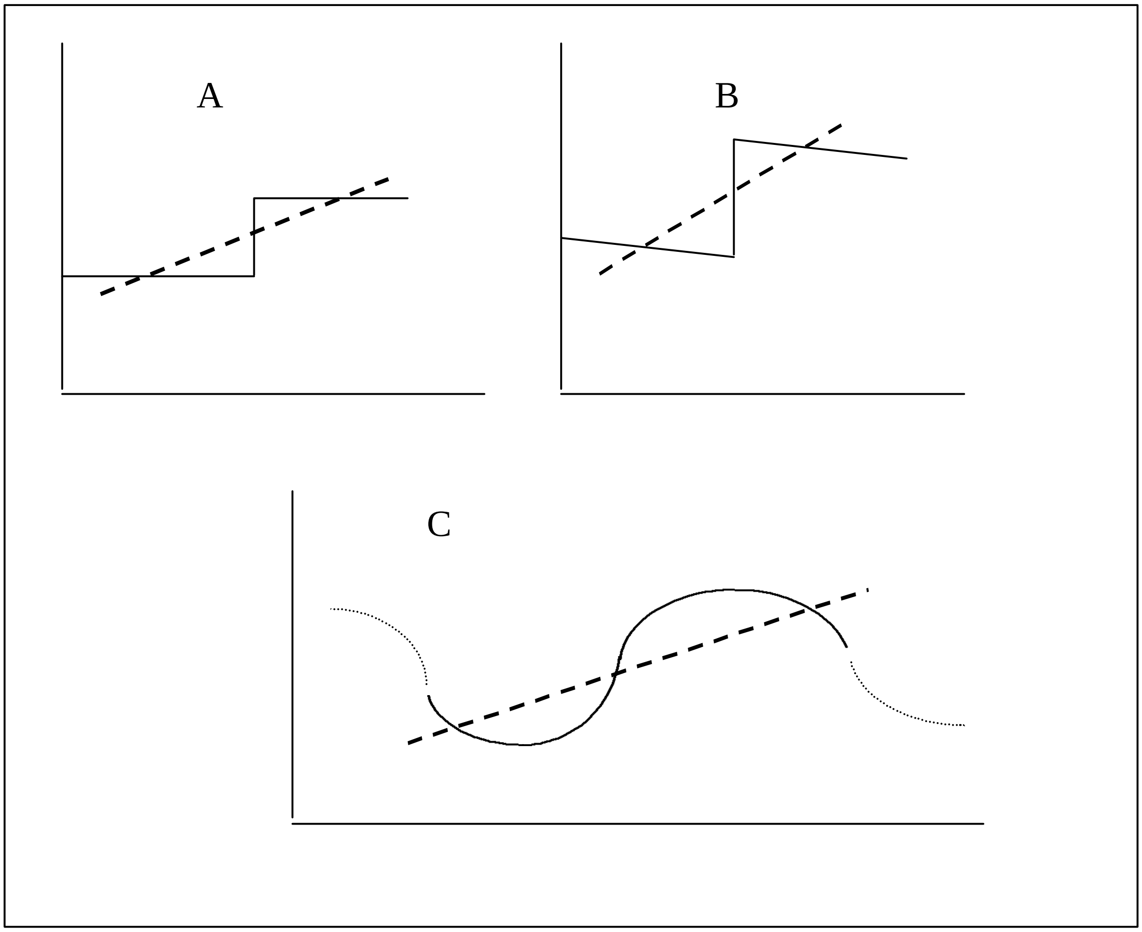

Figure 1. Data patterns that can be inappropriately labeled “positive trends” based on use of linear estimator in equation (1).

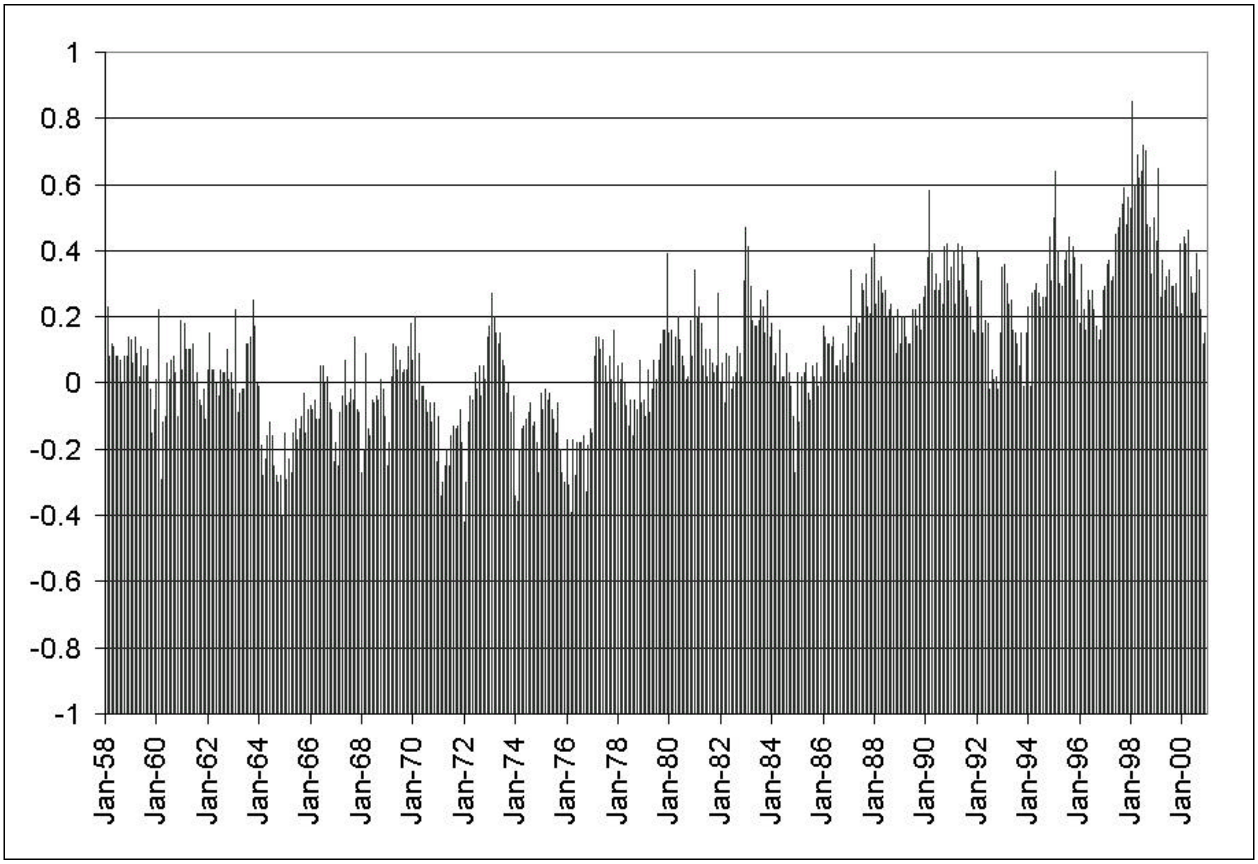

Figure 2. Global Surface Temperature Anomalies January 1958 to December 2000. Source: Hadley Centre (2001).