

NOTES AND CORRESPONDENCE

The Application of Size-Robust Trend Statistics to Global-Warming Temperature Series

THOMAS B. FOMBY

Department of Economics, Southern Methodist University, Dallas, Texas

TIMOTHY J. VOGELSANG

Department of Economics, Cornell University, Ithaca, New York

7 November 2000 and 24 July 2001

ABSTRACT

In this note, new evidence is provided confirming that global temperature series spanning back to the mid-1800s have statistically significant positive trends. Although there is a growing consensus that global temperatures are on the rise systematically, some recent studies have pointed out that strong serial correlation (or a unit root) in global temperature data could, in theory, generate spurious evidence of a significant positive trend. In other words, strong serially correlated data can mimic trending behavior over fixed periods of time. A serial-correlation-robust trend test recently was proposed that controls for the possibility of spurious evidence due to strong serial correlation. This new test is valid whether the errors are stationary or have a unit root (strong serial correlation). This test also has the attractive feature that it does not require estimates of serial correlation nuisance parameters. The test is applied to six annual global temperature series, and it provides strong evidence that global temperature series have positive trends that are statistically significant even when controlling for the possibility of strong serial correlation. The point estimates of the rate of increase in the trend suggest that temperatures have risen about 0.5°C (1.0°F) 100 yr^{-1} . If the analysis is restricted to twentieth-century data, many of the point estimates are closer to 0.6°C .

1. Introduction

In December of 1997, 160 nations having more than 1400 delegates met in Kyoto, Japan, at the United Nations Framework Convention on Climate Change to discuss the “global-warming problem.” The participation of such a large number of countries and delegates emphasizes the growing international concern about the possibility of global warming. By now, there is a growing consensus that global temperatures have been on the rise systematically for the past 100–150 yr. This consensus has grown in part out of statistical studies of global temperature time series that have indicated positive, systematic, and statistically significant trends in these series. These include, for example, Bloomfield (1992), Bloomfield and Nychka (1992), and Zheng and Basher (1999). On the other hand, several authors, including Gordon (1991) and Woodward and Gray (1993,

1995) have pointed out that a statistically significant trend can be spuriously generated when the data have a unit root in the random component. They argue that global temperature data might be well represented by time series models with no trend but with a unit root (or near unit root) in the random component, in which case the recent warming would not be expected to continue systematically into the future.

In this paper we apply a test recently proposed by Vogelsang (1998) that directly controls for the possibility of a spurious trend resulting from a unit root or near-unit root in the data. Although the tests proposed by Vogelsang (1998) were originally designed with economic time series data in mind, they are natural candidates for application to global temperature data. Our empirical results indicate that there is considerable *robust* evidence suggesting that a positive and statistically significant trend is present in global temperature series. Note that we do not attempt to explain the causes of these trends in the temperature series. Our modest goal is simply to help to resolve some of the remaining debate as to whether the recent increase in global temperatures is systematic or purely random.

Corresponding author address: Dr. Timothy J. Vogelsang, Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14853-7601.
E-mail: tjv2@cornell.edu

The basic statistical question as to whether global temperature series have a significant positive trend has nothing to do, *per se*, with the serial correlation properties of the temperature data. A time series with a positive deterministic trend could have errors (i) without serial correlation, (ii) with mild serial correlation (autoregressive roots not close to one), (iii) with strong serial correlation (at least one autoregressive root close to one), or (iv) with a unit root. While the four cases generate very different looking time series, all of them would be consistent with systematic global warming. Uncertainty regarding the serial correlation structure boils down to a statistical “nuisance parameter” issue that affects inference regarding the deterministic trend parameters. Trend tests constructed for case (i) are invalid for cases (ii)–(iv). Trend tests constructed under the assumption of stationary errors remain invalid for case (iv) but are valid (in large samples) for cases (i)–(iii). Unfortunately, in finite samples, such tests may give spurious evidence in case (iii). Trend tests designed explicitly for case (iv) usually do not give spurious evidence in cases (i)–(iii), but they often suffer from low power. On the other hand, the trend test used in this note is valid for all four cases and is therefore very robust.

Some authors have argued that global temperature data may exhibit long memory and that fractionally integrated errors are appropriate. These include, for example, Bloomfield (1992) and Bloomfield and Nychka (1992). Trend tests designed for case (ii) can also give spurious results when errors have long memory. Although not designed explicitly for long-memory errors, the trend test used here remains fairly robust to long-memory errors.

The remainder of the note is organized as follows. In the next section we describe the basic statistical model and review well-known inference methods. A small Monte Carlo experiment is used to illustrate some of the potential pitfalls of the standard tests. In section 3 we introduce and briefly discuss the robust trend test proposed by Vogelsang (1998). Some additional Monte Carlo results are provided to illustrate the robustness properties. In section 4 we describe the global temperature data used in the note and report the empirical results. Section 5 gives some concluding remarks.

2. The statistical model and inference

In testing for a significant trend in time series data, a popular stylized model is the simple linear model

$$y_t = \beta_1 + \beta_2 t + u_t, \quad t = 1, 2, 3, \dots, T. \quad (2.1)$$

The linear trend function $\beta_1 + \beta_2 t$ approximates the systematic component of global temperatures, whereas the random error term u_t approximates the natural variation in temperatures over time. Statistical evidence of global warming is indicated by a positive value of β_2 , and β_2 measures the average change in temperatures per

year (assuming annual data). We take the null hypothesis to be no global warming and the alternative hypothesis to be global warming, which can be written as

$$H_0: \beta_2 \leq 0 \quad H_1: \beta_2 > 0. \quad (2.2)$$

Clearly, aggregate temperature data are serially correlated over time. Therefore, it must be assumed that u_t has serial correlation. Note, however, that the serial correlation structure of u_t is not directly relevant to the null hypothesis being tested. Whether u_t follows an autoregressive moving-average (ARMA) process with or without an autoregressive root close to one, or alternatively, a long-memory process, the null and alternative hypotheses of interest remain Eq. (2.2). Where the serial correlation structure of u_t matters is in the construction of test statistics for H_0 . This is true because the sampling distributions of typical estimators of β_2 depend on the serial correlation structure of u_t . See Bloomfield and Nychka (1992) for examples.

Suppose that u_t is a zero-mean second-order stationary time series with autocovariance function $\gamma_j = \text{cov}(u_t, u_{t-j})$. In this case, it follows from the classic results of Grenander and Rosenblatt (1957) that the ordinary least squares (OLS) estimates of β_1 and β_2 based on Eq. (2.1) are asymptotically equivalent to the generalized least squares (GLS) estimates assuming the form of the serial correlation is known.

To see an example of a GLS estimator, suppose u_t follows the ARMA model $A(L)u_t = B(L)\xi_t$, where ξ_t is an independent and identically distributed (iid) error process with mean zero and variance σ_ξ^2 , L is the lag operator, $Lu_t = u_{t-1}$, and $A(L) = 1 - a_1L - a_2L^2 - \dots - a_pL^p$, $B(L) = 1 + b_1L + b_2L^2 + \dots + b_qL^q$. Provided that $B(L)^{-1}$ exists, the GLS estimate of β_2 would be obtained from nonlinear least squares estimation of the regression

$$B(L)^{-1}A(L)y_t = \beta_1 B(L)^{-1}A(L)1 + \beta_2 B(L)^{-1}A(L)t + \xi_t. \quad (2.3)$$

If, in addition it is reasonable to assume that ξ_t is normally distributed, then the GLS estimate of β_2 from Eq. (2.3) is the maximum likelihood estimator (MLE). The advantage of OLS over GLS or MLE is that OLS does not require specification of the lag polynomials, $A(L)$ and $B(L)$. Because of the large sample equivalence of OLS and GLS/MLE, there is no loss in efficiency if the OLS estimate of β_2 is used for testing Eq. (2.2).

The OLS estimate of β_2 based on Eq. (2.1) is given by the usual formula

$$\hat{\beta}_2 = \frac{\sum_{t=1}^T (t - \bar{t})y_t}{\sum_{t=1}^T (t - \bar{t})^2},$$

where $\bar{t} = T^{-1} \sum_{t=1}^T t$. It is well known from Grenander and Rosenblatt (1957) that in large samples

$$(\hat{\beta}_2 - \beta_2) \approx N\left[0, \sigma^2 \left[\sum_{t=1}^T (t - \bar{t})^2 \right]^{-1}\right] \approx N\left(0, \frac{12\sigma^2}{T^3}\right),$$

where $\sigma^2 = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j < \infty$. As long as an estimate of σ^2 can be obtained, asymptotically valid standard errors can be computed for $\hat{\beta}_2$ and a t statistic can be constructed.

Because σ^2 is proportional to the spectral density of u_t at frequency zero, there are many possible estimators of σ^2 to choose from. See Bloomfield and Nychka (1992) or Woodward and Gray (1993) for examples used in the global-warming literature. Here we consider the class of nonparametric estimators (see Priestley 1981) of σ^2 given by

$$\hat{\sigma}^2 = \sum_{j=-(T-1)}^{T-1} k(j/M) \hat{\gamma}_j, \quad \hat{\gamma}_j = T^{-1} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j},$$

where \hat{u}_t are the OLS residuals from Eq. (2.1), $k(x)$ is a kernel or weighting function, and M is the truncation lag or bandwidth. In what follows, we use the Bartlett kernel defined as $k(x) = 1 - |x|$ for $|x| \leq 1$ and $k(x) = 0$ otherwise. The truncation lag M is chosen using the data-dependent autoregressive lag-one [AR(1)] plug-in formula given by Andrews (1991). Using $\hat{\sigma}^2$, a standard error for $\hat{\beta}_2$ can be computed using the formula

$$\text{se}(\hat{\beta}_2) = \left\{ \hat{\sigma}^2 \left[\sum_{t=1}^T (t - \bar{t})^2 \right]^{-1} \right\}^{1/2},$$

which leads to the t statistic

$$t_{\text{HAC}} = \hat{\beta}_2 / \text{se}(\hat{\beta}_2).$$

Borrowing from the econometric literature, the HAC subscript denotes the fact that this t statistic is valid for errors with conditional heteroscedasticity and autocorrelation of unknown form. Under standard regularity conditions, including stationarity of u_t , it follows that $t_{\text{HAC}} \xrightarrow{d} N(0, 1)$. The null hypothesis given by Eq. (2.2) can be tested in the usual way using t_{HAC} .

As long as the serial correlation in the model is not too strong, statistics like t_{HAC} perform well in practice. However, as pointed out by Woodward and Gray (1993), Vogelsang (1998), and others, statistics like t_{HAC} can perform poorly in practice when the serial correlation in the errors is strong. More specifically, if the errors can be modeled as an ARMA process, then under the null hypothesis t_{HAC} will tend to overreject when one of the autoregressive roots of u_t is close to or equal to 1. In fact, the distribution theory underpinning the t_{HAC} statistic breaks down when there is a unit root in the errors.

The distribution theory for t_{HAC} also breaks down if the errors have long memory. Consider the model

$$(1 - L)^d u_t = v_t, \quad (2.4)$$

where v_t is a stationary process with finite spectral density at frequency zero. The process given by Eq. (2.4)

is called a fractionally integrated process and is often labeled an $I(d)$ process. When $0 < d < 0.5$, u_t is stationary but has long memory because the autocovariances γ_j slowly decay as j increases. Because of this slow decay, u_t has an unbounded spectral density at frequency zero. It is for this reason that the distribution theory for t_{HAC} breaks down when the errors have long memory. Of course, one could estimate the trend model under the assumption of long-memory errors and construct appropriate tests (see Bloomfield and Nychka 1992). We do not follow this approach in this paper given that the focus is on robust tests that do not require a fully parametric specification of the error model.

To illustrate the problems caused by strong serial correlation and to motivate the robust statistic considered in the next section, we conducted a very simple finite sample Monte Carlo simulation study. We modeled u_t according to the two simple processes

$$u_t = \alpha u_{t-1} + e_t, \quad (2.5)$$

$$(1 - L)^d u_t = e_t, \quad (2.6)$$

where $u_0 = 0$, and e_t is iid $N(0, 1)$. Equation (2.5) is an AR(1) error model whereas Eq. (2.6) is a simple $I(d)$ error model. The errors are stationary for $|\alpha| < 1$ and $d < 0.5$, respectively.

Using Eqs. (2.4) and (2.5) as the error models, we generated data from Eq. (2.1) with $\beta_1 = \beta_2 = 0$. We can set $\beta_1 = 0$ because t_{HAC} does not depend on the true value of β_1 (t_{HAC} is exactly invariant to β_1). We used a sample size of $T = 129$ that is similar to the length of annual global temperature series (see section 4). We report results for $\alpha = 0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 1.0$, and $d = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0$. We used 10 000 replications and computed empirical rejection probabilities at the 5% nominal level (for a one-tailed test) using the standard normal critical value. The results are tabulated in Table 1. For comparison we also report results for a t statistic based on the MLE estimator of β_2 assuming the errors follow Eq. (2.4). We denote this statistic by $t_{\text{AR}(1)}$. The MLE estimator was obtained using nonlinear least squares estimation of the regression:

$$y_t = \beta_1(1 - \alpha) + \beta_2\alpha + \beta_2(1 - \alpha)t + \alpha y_{t-1} + e_t.$$

The table clearly illustrates that for small values of α , the t_{HAC} and $t_{\text{AR}(1)}$ statistics perform reasonably well with rejection probabilities close to 0.05. However, as α increases and approaches 1, rejection probabilities increase to well above 0.05. In a similar fashion, if the errors are $I(d)$ and d is not close to zero, both tests tend to overreject. The overrejection problem becomes more pronounced as d increases. If the serial correlation in the global temperature data is strong (α near 1 or $d > 0$), then it is theoretically possible that one could obtain spurious evidence of a significant positive trend.

It is important to note that the problem here is not isolated to the case where α is exactly 1 or $d \geq 0.5$. If

TABLE 1. Empirical null rejection probabilities in finite samples.*

AR(1) errors					$I(d)$ errors				
α	t_{HAC}	$t_{\text{AR}(1)}$	t_z^*	$t - \text{PS}_T$	d	t_{HAC}	$t_{\text{AR}(1)}$	t_z^*	$t - \text{PS}_T$
0.00	0.053	0.053	0.050	0.043	0.0	0.053	0.053	0.050	0.043
0.20	0.067	0.055	0.052	0.043	0.1	0.103	0.095	0.069	0.058
0.40	0.077	0.057	0.055	0.042	0.2	0.152	0.141	0.095	0.075
0.60	0.091	0.062	0.061	0.041	0.3	0.196	0.184	0.116	0.087
0.80	0.122	0.074	0.082	0.033	0.4	0.227	0.224	0.141	0.092
0.90	0.163	0.089	0.114	0.026	0.5	0.258	0.257	0.172	0.099
0.95	0.212	0.114	0.150	0.024	0.8	0.331	0.303	0.245	0.076
1.00	0.351	0.314	0.290	0.051	1.0	0.351	0.314	0.290	0.051

* Data were generated according to Eq. (2.1). For AR(1), errors $u_t = \alpha u_{t-1} + e_t$, $e_t \sim \text{iid } N(0, 1)$. For $I(d)$ errors, $(1 - L)^d u_t = e_t$, $e_t \sim \text{iid } N(0, 1)$. In all cases $u_0 = 0$. The null and alternative hypotheses are given by $H_0: \beta_2 \leq 0$ and $H_1: \beta_2 > 0$. Here $T = 129$ and 10 000 replications were used. Asymptotic critical values were used with nominal level 0.05.

problems only occurred in the nonstationary cases, they could be deemed irrelevant for temperature data since nonstationarity can be ruled out on scientific grounds. However, the spurious rejection problem is still relevant for temperature data because it can occur for values of α near 0.8 or 0.9 or $0.2 \leq d \leq 0.4$, which are clearly in the stationary region.

So, while there is considerable empirical evidence that suggests global temperature series have a significant positive trend, skeptics can point to the overrejection problem caused by strong serial correlation to cast some doubt. In the next section, we introduce a test statistic that does not have the overrejection problem in the presence of strong serial correlation or even a unit root in the data. This test also remains fairly robust to $I(d)$ errors.

3. A robust trend test

In this section we review the robust trend test statistic, labeled $t - \text{PS}_T$, recently proposed by Vogelsang (1998). The statistic is defined shortly. Like the t_{HAC} statistic and similar statistics, $t - \text{PS}_T$ is robust to the form of serial correlation in u_t . Unlike t_{HAC} , $t - \text{PS}_T$ does not overreject when the serial correlation is strong or if there is a unit root in the errors. In addition, the $t - \text{PS}_T$ statistic has the very practical and useful property that it does not require an estimate of σ^2 . Thus, the subjective choices such as weighting kernel and truncation lag do not have to be made by the empirical researcher.

The $t - \text{PS}_T$ test statistic is based on the following regression that is obtained by computing partial sums of the original data, y_t :

$$z_t = \beta_1 t + \beta_2 \left[\frac{1}{2}(t^2 + t) \right] + S_t, \quad (3.1)$$

where $z_t = \sum_{j=1}^t y_j$, $S_t = \sum_{j=1}^t u_j$, and the regressors are obtained from the formulas $t = \sum_{j=1}^t 1$ and $\frac{1}{2}(t^2 + t) = \sum_{j=1}^t j$. Suppose Eq. (3.1) is estimated by OLS. Let $\hat{\beta}_2$ denote the OLS estimate of β_2 obtained from Eq. (3.1). Because S_t has a unit root by construction, $\hat{\beta}_2$ has a larger sampling variability than $\hat{\beta}_2$ since the Grenander–

Rosenblatt result does not apply to $\tilde{\beta}_2$. The advantage of $\tilde{\beta}_2$ over $\hat{\beta}_2$ for testing is that more robust tests can be constructed using $\tilde{\beta}_2$. Therefore, efficiency is sacrificed for robustness.

Let b be a nonrandom number. The choice of b is discussed below. The $t - \text{PS}_T$ test is defined as

$$t - \text{PS}_T = t_z^* \exp(-bJ_T),$$

where $t_z^* = T^{-1/2}t_z$ with t_z being the standard OLS t statistic for testing $\beta_2 = 0$, in regression Eq. (3.1) (i.e., t_z is the t statistic that would be automatically computed by standard regression packages under the implicit assumption of iid errors), and

$$J_T = (\text{RSS}_y - \text{RSS}_J)/\text{RSS}_J,$$

where RSS_y denotes the OLS residual sum of squares from regression (2.1), and RSS_J denotes the OLS residual sum of squares from the regression

$$y_t = \beta_1 + \beta_2 t + \sum_{i=2}^9 \beta_{i+1} t^i + u_t. \quad (3.2)$$

Note that $[(T - 10)/8]J_T$ is the F test for testing the hypothesis $\beta_3 = \beta_4 = \dots = \beta_{10} = 0$ in regression Eq. (3.2). Vogelsang (1998) recommended that the polynomial order be 9 in Eq. (3.2) because power of the $t - \text{PS}_T$ test is an increasing function of the polynomial order, but the increase in power is negligible for polynomial orders greater than 9.

When computing the $t - \text{PS}_T$ test, the value used for b depends on the significance level. Vogelsang (1998) showed that, for a given significance level, b can be found such that the asymptotic critical value of $t - \text{PS}_T$ is the same when the errors are stationary and when the errors have a unit root. This property ensures that $t - \text{PS}_T$ will not tend to overreject when the errors have strong serial correlation. Vogelsang (1998) derived the asymptotic distribution of $t - \text{PS}_T$ and computed asymptotic critical values and the corresponding b values using Monte Carlo simulation methods. For percentage points 1%, 2.5%, 5%, and 10% the asymptotic critical values and b s (in the parentheses) are 2.647 (1.501), 2.152 (0.995), 1.720 (0.716), and 1.331 (0.494).

TABLE 2. Empirical results for global temperature series.*

Series	Years	$\hat{\beta}_2$	t_{HAC}	$\tilde{\beta}_2$	$t - \text{PS}_T$ 5%	$t - \text{PS}_T$ 2.5%	$t - \text{PS}_T$ 1%	J_T
JWB	1854–1993	0.004155	10.12 ^a	0.003732	—	3.111 ^b	2.524	0.4128 ^f
JWB	1900–93	0.005840	9.253 ^a	0.005613	—	3.271 ^b	2.485	0.5428 ^c
JWBE	1854–1993	0.004429	9.478 ^a	0.004114	—	2.994 ^b	2.271	0.5465 ^c
JWBE	1900–93	0.005714	9.292 ^a	0.005524	—	2.714 ^b	1.898	0.7066 ^d
WH	1880–1993	0.005359	9.688 ^a	0.005277	—	3.104 ^b	2.432	0.4819 ^f
WH	1900–93	0.004928	6.388 ^a	0.004461	1.936 ^c	1.597	1.127	0.6896 ^d
VGL	1881–1993	0.005184	8.497 ^a	0.005065	—	2.757 ^b	2.154	0.4883 ^f
VGL	1900–93	0.004723	5.559 ^a	0.004204	1.6885	1.380	0.957	0.7235 ^d
JP	1860–2000	0.004334	6.903 ^a	0.003653	1.852 ^c	1.425	0.886	0.9390
JP	1900–2000	0.006497	8.328 ^a	0.005937	—	2.151 ^b	1.436	0.7990 ^d
TaveGL	1856–2000	0.004268	7.342 ^a	0.003564	1.881 ^c	1.465	0.930	0.8971 ^d
TaveGL	1856–2000	0.006573	8.818 ^a	0.005978	—	2.387 ^b	1.632	0.7508 ^d

Critical values for test statistics

%	t_{HAC}	$t - \text{PS}_T$	b
0.950	1.645	1.720	0.716
0.975	1.960	2.152	0.995
0.990	2.330	2.647	1.501

* The superscripts a, b, and c denote rejection of the null hypothesis of no global warming at the 1%, 2.5%, and 5% levels of significance, respectively. The superscripts d, e, and f denote rejection of a unit root in the errors at the 1%, 2.5%, and 5% levels of significance, respectively. The J_T is a left-tail test with 1%, 2.5%, and 5% critical values of 0.488, 0.678, and 0.908, respectively.

Vogelsang (1998) provided extensive theoretical and simulation evidence showing that the $t - \text{PS}_T$ statistic does not suffer from overrejection problems as the serial correlation in the errors becomes strong. He also showed $t - \text{PS}_T$ has good power. To illustrate this robustness of $t - \text{PS}_T$, empirical rejection probabilities for $t - \text{PS}_T$ were computed in the same simulations discussed above in section 2. Rejections were calculated using the asymptotic 5% critical of 1.720 and $b = 0.716$. We also report rejection probabilities for the t_z^* statistic that is the $t - \text{PS}_T$ statistic implemented using $b = 0$.

These results are reported in Table 1. In the case of AR(1) errors, empirical rejection probabilities for $t - \text{PS}_T$ are always near or below 0.05 even when α is close to or equal to 1. In fact, $t - \text{PS}_T$ tends to be conservative when α is close to 1. On the other hand, if t_z^* were used, substantial overrejections occur for α close to 1. This illustrates the importance of the $\exp(-bJ_T)$ component for the performance of the $t - \text{PS}_T$ statistic. If the errors are fractionally integrated, $t - \text{PS}_T$ does tend to overreject somewhat. However, the tendency to overreject is much less severe than for the t_{HAC} , $t_{\text{AR}(1)}$, and t_z^* statistics. Clearly, if the $t - \text{PS}_T$ statistic indicates that a global temperature series has a positive and significant trend, then this would be more robust evidence than has been previously obtained in the climate literature.

Because the $t - \text{PS}_T$ test is relatively new and not widely known, some readers may benefit from the following brief discussion of some of the properties of the test and the rationale for its components. Vogelsang (1998) should be consulted for additional technical details. Readers not interested in this discussion can jump to section 4 for the empirical results.

The first component of $t - \text{PS}_T$, t_z^* , directly tests null hypothesis [Eq. (2.2)] and could be used as a trend test

on its own. The t_z^* statistic has some nice properties. As long as the errors in Eq. (2.1) are stationary and $\sigma^2 < \infty$, t_z^* has a well-defined asymptotic distribution (although it is not standard normal) that does not depend on any unknown nuisance parameters. Therefore, t_z^* does not require an estimate of σ^2 . However, as illustrated above, t_z^* tends to overreject in small samples as the serial correlation in the errors becomes strong. This overrejection problem occurs because the asymptotic distribution of t_z^* is different when the errors have a unit root and its critical values are much larger. In contrast to the t_{HAC} statistic, t_z^* continues to have a well-defined asymptotic distribution that does not depend on nuisance parameters when the errors have a unit root. This fact allows for correction of the overrejection problem using $\exp(-bJ_T)$.

The J_T statistic was proposed by Park and Choi (1988) and Park (1990) for testing the null hypothesis that the errors in Eq. (2.1) have an autoregressive unit root. The J_T test is a left-tailed test where a unit root in the errors is rejected for small values of J_T . Park and Choi (1988) showed that, when the errors have a unit root, J_T has a well-defined asymptotic distribution free of nuisance parameters that is similar to a chi-square distribution. And, they showed that, when the errors are stationary, J_T converges to zero in which case $\exp(-bJ_T)$ converges to 1. Thus, $t - \text{PS}_T$ and t_z^* are equivalent tests in large samples when the errors are stationary.

When the errors have a unit root, J_T takes on large positive values and thus $\exp(-bJ_T)$ takes on small positive values provided $b > 0$. Therefore, $\exp(-bJ_T)$ can be used to shrink the distribution of $t - \text{PS}_T$ when the errors have a unit root. There does not exist a value of b so that the asymptotic distribution of $t - \text{PS}_T$ is the same for both stationary errors and unit root errors.

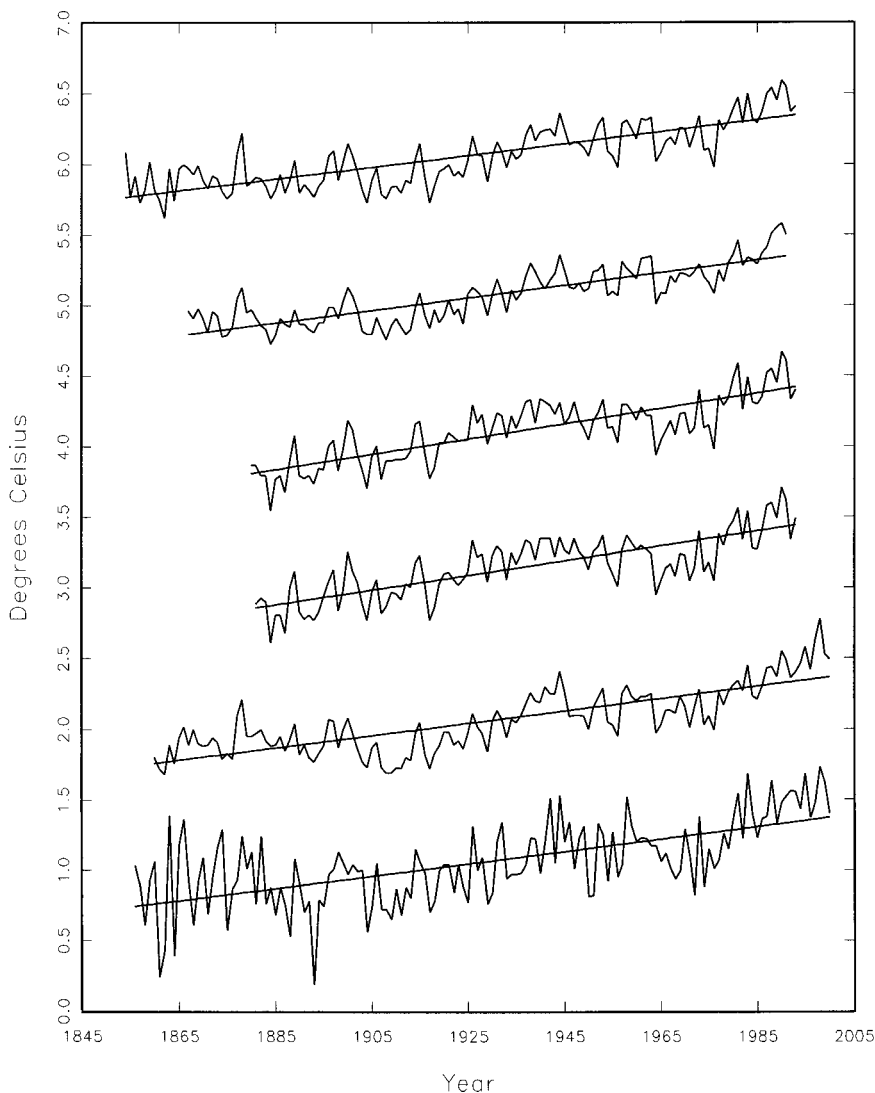


FIG. 1. Global temperature ($^{\circ}\text{C}$) series (1854–2000) and fitted OLS trend lines. Labels from top to bottom: JWB, JWBE, WH, VGL, JP, and TaveGL.

However, for a given percentage point, it is possible to compute the value of b such that the asymptotic critical values of $t - \text{PS}_T$ are the same for both stationary and unit root errors. Because the critical value is the same whether the errors are stationary or have a unit root, the overrejection problem does not occur.

4. Analysis of the global warming data

In this section we test the hypotheses in Eq. (2.2) for six annual global temperature series. The starting and ending dates of the series vary but all series start in the late 1800s and end in the late 1900s. All series are measured in degrees Celsius relative to a reference period average. For each series the labels, spans, and sources are as follows:

- JWB: 1854–1993 (relative to 1950–79 average), Jones et al. (1994);
- JWBE: The JWB series adjusted for the influence of El Niño–Southern Oscillation events as reported by Jones et al. (1994);
- WH: 1880–1993 (relative to 1951–80 average), Wilson and Hansen (1994);
- VGL: 1881–1993 (relative to 1951–75 average), Vinnikov et al. (1994);
- JP: 1860–2000 (relative to 1961–90 average), Jones et al. (1999) and Parker et al. (1995), downloaded from the Web site online at www.meto.govt.uk/research/hadleycentre/obsdata/ (HadCRUG series);
- TaveGL: 1856–2000 (relative to 1961–90 average), downloaded from the Web site online at www.cru.uea.ac.uk/ftpdata/tavegl.dat maintained by P. Jones.

The six series are plotted in Fig. 1 along with fitted trend lines obtained from the OLS estimates of Eq. (2.1). OLS estimates of β_2 obtained from Eqs. (2.1) and (3.1) along with the t_{HAC} and $t - \text{PS}_T$ statistics are reported in Table 2. There are potentially three entries for the $t - \text{PS}_T$ statistic corresponding to b values appropriate for tests at the 5%, 2.5%, and 1% levels of significance. If the $t - \text{PS}_T$ statistic is significant at some significance level (e.g., 2.5%), then it is significant at higher significance levels (e.g., 5%) and is not reported for those higher significance levels. Because it is sometimes argued that most of the global warming has occurred in the twentieth century, we also report results using 1900 as the starting date for each series.

Based on Eq. (2.1), point estimates of the trend slopes range from 0.004 155 to 0.006 497 confirming the conventional wisdom that global temperatures have increased at roughly the rate of $0.005^\circ\text{C yr}^{-1}$. Point estimates using twentieth-century data are larger in four cases but are smaller for the WH and VGL series. All of the point estimates are statistically greater than zero according to the t_{HAC} statistic. This kind of finding has been consistently obtained in past studies. A skeptic might not be fully convinced by these results for t_{HAC} given the possibility of overrejection if the errors have strong serial correlation. However, the $t - \text{PS}_T$ test confirms the statistical significance of the trend estimates. In all but one case, the null hypothesis of a nonpositive trend can be rejected at the 5% level using $t - \text{PS}_T$. In a majority of cases, the null hypothesis can be rejected at the 2.5% level. These results strongly suggest that global temperature series have been increasing over time, and the possibility that this conclusion is being spuriously generated by strong serial correlation or a unit root in the data can be effectively ruled out. Last, we also report the J_T statistic used in constructing the $t - \text{PS}_T$ statistics. In all but one case, the null hypothesis that the errors have a unit root can be rejected at the 5% significance level using the J_T test.

5. Conclusions

In this note we applied a recently proposed serial correlation–robust trend function test to six annual global temperature series. Unlike more conventional tests, this new test is robust to strong serial correlation or a unit root in the data. Using this test, we find strong evidence to suggest that typical global temperature series spanning back to the mid-1800s have positive trends that are statistically significant. The robustness of the test effectively rules out the possibility that significance is being spuriously generated by strong serial correlation or a unit root in the data. Therefore, our results confirm and strengthen the growing consensus that average global temperatures are indeed systematically on the rise as

widely believed. The point estimates of the rate of increase in the trend suggest that temperatures have risen about 0.5°C (1.0°F) 100 yr^{-1} . If the analysis is restricted to twentieth-century data, many of the point estimates are closer to 0.6°C .

Acknowledgments. The authors thank an editor and two anonymous referees for constructive comments that lead to improvements of the paper.

REFERENCES

- Andrews, D. W. K., 1991: Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.
- Bloomfield, P., 1992: Trends in global temperatures. *Climatic Change*, **21**, 275–287.
- , and D. Nychka, 1992: Climate spectra and detecting climate change. *Climatic Change*, **21**, 1–16.
- Gordon, A. H., 1991: Global warming as a manifestation of a random walk. *J. Climate*, **4**, 589–597.
- Grenander, U., and M. Rosenblatt, 1957: *Statistical Analysis of Stationary Time Series*. John Wiley and Sons, 300 pp.
- Jones, P. D., T. M. L. Wigley, and K. R. Briffa, 1994: Global and hemispheric temperature anomalies—Land and marine instrumental records. *Trends '93: A Compendium of Data on Global Change*, T. A. Boden et al., Eds., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 603–608.
- , M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199.
- Park, J. Y., 1990: Testing for unit roots and cointegration by variable addition. *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, T. Fomby and G. Rhodes, Eds., JAI Press, 107–134.
- , and B. Choi, 1988: A new approach to testing for a unit root. Center for Analytic Economics Working Paper #88-23, 40 pp. [Available from Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14853-7601.]
- Parker, D. E., C. K. Folland, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Climatic Change*, **31**, 559–600.
- Priestley, M. B., 1981: *Spectral Analysis and Time Series*. Academic Press, 890 pp.
- Vinnikov, K. Y., P. Y. Groisman, and K. M. Lugina, 1994: Global and hemispheric temperature anomalies from instrumental surface air temperature records. *Trends '93: A Compendium of Data on Global Change*, T. A. Boden et al., Eds., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 615–627.
- Vogelsang, T. J., 1998: Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, **66**, 123–148.
- Wilson, H., and J. Hansen, 1994: Global and hemispheric temperature anomalies from instrumental surface air temperature records. *Trends '93: A Compendium of Data on Global Change*, T. A. Boden et al., Eds., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 609–614.
- Woodward, W. A., and H. L. Gray, 1993: Global warming and the problem of testing for trend in time series data. *J. Climate*, **6**, 953–962.
- , and —, 1995: Selecting a model for detecting the presence of a trend. *J. Climate*, **8**, 1929–1937.
- Zheng, X., and R. E. Basher, 1999: Structural time series models and trend detection in global and regional temperature series. *J. Climate*, **12**, 2347–2358.