## HAC-Robust Trend Comparisons Among Climate Series With Possible Intercept Shifts

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Environmetrics</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>env-13-0078.R2</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>McKitrick, Ross; University of Guelph, Department of Economics Vogelsang, Tim; Michigan State University, Economics</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Autocorrelation, Trend comparisons, HAC methods, Level shifts, Climate models, Tropical Troposphere</td>
</tr>
</tbody>
</table>
HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS

Ross McKitrick*
Department of Economics
University of Guelph
Guelph ON Canada N1G 2W1
rmckitri@uoguelph.ca
519-824-4120 x52532

Timothy J. Vogelsang
Department of Economics
Michigan State University
tjv@msu.edu

June 10, 2014
*Corresponding author

Under 3rd review, Environmetrics
Not for citation

Abstract: Comparisons of trends across climatic data sets are complicated by the presence of serial correlation and possible step-changes in the mean. We build on heteroskedasticity and autocorrelation (HAC) robust methods, specifically the Vogelsang-Franses (VF) nonparametric testing approach, to allow for a step-change in the mean (level shift) at a known or unknown date. The VF method provides a powerful multivariate trend estimator robust to unknown serial correlation up to but not including unit roots. We show that the critical values change when the level shift occurs at a known or unknown date. We derive an asymptotic approximation that can be used to simulate critical values, and we outline a simple bootstrap procedure that generates valid critical values and p-values. Our application builds on the literature comparing simulated and observed trends in the tropical lower- and mid-troposphere since 1958. The method identifies a shift in observations around 1977, coinciding with the Pacific Climate Shift. Allowing for a level shift causes apparently significant observed trends to become statistically insignificant. Model over-estimation of warming is significant whether or not we account for a level shift, although null rejections are much stronger when the level shift is included.

Acknowledgments: We thank numerous seminar and conference participants, as well as the referees, for helpful comments. RM thanks the Institute for New Economic Thinking for financial support.

Keywords: Autocorrelation; trend estimation; mean shift; HAC methods; climate models
HAC-Robust Trend Comparisons Among Climate Series With Possible Level Shifts

1 INTRODUCTION

Many empirical applications involve comparisons of linear trend magnitudes across different time series with autocorrelation and/or heteroskedasticity of unknown form. Vogelsang and Franses (2005, herein VF) derived a class of heteroskedasticity and autocorrelation robust (HAC) tests for this purpose. The VF statistic is similar in form to the familiar regression $F$-type statistics but remains valid under serial dependence up to but not including unit roots in the time series. For treatments of the theory behind HAC estimation and inference see Andrews (1991), Kiefer and Vogelsang (2005), Newey and West (1987), Sun, Phillips and Jin (2008) and White and Domowitz (1984) among others. Like many HAC approaches, the VF approach is nonparametric with respect to the serial dependence structure and does not require a specific model of serial correlation or heteroskedasticity to be implemented. Unlike most nonparametric approaches, the VF approach avoids sensitivity to bandwidth selection by setting the bandwidth equal to the entire sample.

Here we extend the VF approach to the case in which one or more of the series has a possible level shift. Our assumption throughout is that a researcher considers a one-time level shift as a fundamentally different process than a continuous trend. Consequently, if the null hypothesis is that two series have the same trend, and one series exhibits a trend while the other exhibits a level shift and no trend, a rejection of the null would be considered valid since the two phenomena are distinct and a prediction of one is not confirmed by observing the other.

Accounting for level shifts does not necessarily increase the likelihood of rejecting a null of trend equivalence. In the top panel of Figure 1, a comparison of the linear trend coefficients would suggest they are similar, but clearly $y_1$ differs from $y_2$ in that the former is steadily trending while the latter is trendless with a single discrete level shift at the break point $T_b$. By contrast, in the bottom panel a failure to account for the shift would overstate the difference between the trend slopes. In each case, the influence of the shift term is highlighted by the fact that if the trend slope comparisons were conducted over the pre-shift or post-shift
intervals, they might indicate opposite results to those based on the entire sample (with the shift term omitted).

The basic linear trend model is written as

\[ y_{it} = a_i + b_i t + u_{it}, \]  

where \( i = 1, \ldots, n \) denotes a particular time series and \( t = 1, \ldots, T \) denotes the time period. The random part of \( y_{it} \) is given by \( u_{it} \) which is assumed to be covariance-stationary (in which case \( y_{it} \) is labeled a trend stationary series, that is, stationary around a linear time trend, if one is present). For a series of length \( T \), we parameterize the break point by denoting the fraction of the sample occurring before it as \( \lambda = T_b / T \).

The following issues must be addressed in order to derive an HAC-robust trend comparison test in the presence of a possible level shift. (i) If \( \lambda \) is known, and specifically is known to be in the \((0,1)\) interval, the VF test score can be generalized, as we show in Section 3, but the distribution is shown to depend on \( \lambda \) and the critical values change. It will turn out that the form of the VF statistic and its critical values are the same whether one is testing hypotheses involving the trend coefficients or other parameters in the trend function.

(ii) If \( \lambda \) is unknown it must be estimated along with the magnitude of the associated shift term. But this gives rise to a problem of non-identification if we want to allow for the possibility that the true value of the level shift parameter is zero.

The regression model with level shift takes the form

\[ y_{it} = a_i + g_i D U_i(\lambda) + b_i t + u_{it} \]  

where the dummy variable \( D U_i(\lambda) = 0 \) if \( t \leq \lambda T \) and 1 otherwise (we will typically suppress the \( \lambda \) term where it is convenient to do so). Hence, for series \( i \), estimation of (2) by OLS yields an estimated intercept of \( \hat{a}_i \) up to \( T_b \) and \( \hat{a}_i + \hat{g}_i \) thereafter. In our empirical application we are primarily interested in testing hypotheses about the trend slope parameters, \( b_i \), while controlling for the possibility of a level shift. If it is reasonable to view \( \lambda \) as known, then inference about the trend slopes will proceed in a straightforward way with \( D U_i(\lambda) \) included in the model even in the case where the true value of \( g_i \) is zero. However, if it is more reasonable to treat \( \lambda \) as unknown and we want to be robust to the possibility that \( g_i \) is non-zero, then inference about the trend slopes \( (b_i) \) becomes more delicate because \( \lambda \) is not identified when \( g_i \) is zero. We
would face a similar identification problem if we wanted to test the null hypothesis that $g_i$ itself is zero and $\lambda$ is unknown.

There is now a well-established literature in statistics and econometrics for carrying out inference where a parameter is not identified under the null hypothesis but is identified under the alternative hypothesis. See for example Davies (1987), Andrews (1993), Andrews and Ploberger (1994), and Hansen (1996) among others. One solution to this identification problem involves the use of a supremum function, which is akin to a data-mining approach. In the present case we can compute the $VF$ statistic for equality of trends for a range of $\lambda$ allowed to vary across $(0, 1)$ and find the largest $VF$ statistic, the sup-$VF$ statistic. To be robust to the possibility that there are no level shifts in the data, i.e. to be robust to the critique that the date of the level shift was chosen to data-mine an outcome for the equality of trends test, we work out the null distribution of the sup-$VF$ statistic for equality of trends for the case where $g_i = 0$. This yields a trend equality test that is very robust to the possibility and location of potential level shifts.

Although our focus is on the problem of trend inference allowing for possible unknown level shifts, our extension of the $VF$ approach is general enough to include tests of the null hypothesis of no level shift and we report some limited results in the paper for these tests. A potential application of tests for a level shift is the homogenization of weather data. Many long observational records are believed to have been affected by possible equipment and/or sampling changes, changes to the area around monitoring locations and so forth (see Hansen et al. 1999, Brohan et al. 2006 for examples in the land record; Folland and Parker 1995, Thompson et al. 2008 for examples in sea surface data). A typical method for detecting and removing level shifts is to construct a reference series which is not expected to exhibit the discontinuity, such as the mean of other weather station records in the vicinity, and then look for one or more jumps in a record relative to its reference series.

While the application of the $VF$ approach to testing for a level shift is potentially quite useful in many empirical settings, the problem of testing for a level shift in a trending series with a known or unknown shift date has already received some attention in the econometrics literature (see Vogelsang (1997) and Sayginsoy and Vogelsang (2011)) and the empirical climate literature (see Gallagher, Lund and Robbins (2013) and references therein). Each proposed method has inherent strengths and weaknesses. A complete comparison
of the $VF$ approach to existing tests for level a shift would be a substantial undertaking and is beyond the scope of this paper, but we draw some contrasts in Sections 4 and 5.

The question of whether or not a level shift is present in trending data can strongly affect the resulting trend calculations and tests of equality of trend slopes. If a change point $\lambda$ is known, the analysis in Section 3 applies, and if a change point is suspected but the date is unknown, the analysis in Section 4 applies. In our application we focus on the case where there is at most one level shift in each series. In other applications, such as those involving very long weather series, one might suspect there are multiple shifts. If they occur at known dates, then our extension of the $VF$ approach is general enough to apply. However, should it be more reasonable to model the shift dates as unknown, or should there be uncertainty regarding the number of shifts, this greatly complicates the analysis, especially from a computational perspective, and is beyond the scope of this paper. In addition, if one thinks level shifts occur frequently and with randomness, then there is the additional difficulty that the range of possible specifications could, in principle, include the case in which the level changes by a random amount at each time period, which is equivalent to having a random walk, or unit root component in $u_t$. If $y_t$ has a unit root component, inference in models (1) and (2) becomes more complicated. More importantly, it is difficult to give a physical interpretation to a unit root component of a temperature series. See Mills (2010) for a discussion of temperature trend estimation when a random walk is a possible element of the specification.

In our application we think it is reasonable to assume that the observed series are well characterized by a trend and at most one level shift at a known date, and that the errors are covariance stationary. We focus on the prediction of climate warming in the troposphere over the tropics. As shown in Section 6, climate models predict a steady warming trend in this region due to rising atmospheric greenhouse gas levels, but none predict a step-change, so trends and shifts can be regarded as distinct phenomena. A number of studies (summarized below) have shown that models likely overstate the warming trend, but there is disagreement as to whether the bias is statistically significant. McKitrick et al. (2010) used the original $VF$ test to examine this issue over the 1979-2009 interval, coinciding with the record available from weather satellites. We extend their analysis to the 1958-2012 interval using data from weather balloons. This long span encompasses a date at which a known climatic event caused a level shift in many observed temperature
series. If the shift is nontrivial in magnitude, the comparison would thus be akin to that in Figure 1, such that failure to take it into account could bias the comparison either towards over-stating or understating the difference in trend slopes.

The event in question occurred around 1978 and is called the Pacific Climate Shift (PCS). This manifested itself as an oceanic circulatory system change during which basin-wide wind stress and sea surface temperature anomaly patterns reversed, causing an abrupt step-like change (level shift) in many weather observations, including in the troposphere, as well as in other indicators such as fisheries catch records (see Seidel and Lanzante 2004, Tsonis et al. 2007, Powell Jr. and Xu 2011, and extensive references therein). For our purposes we do not try to present a specific physical explanation of the PCS or even evidence that its origin was exogenous to the climate system, only that it was a large event at an approximately known date, the existence of which has been documented and studied extensively and which resulted in a shift in the mean of the temperature data. We first present results based on assumption that the PCS occurred at a known date (Section 6.2) and then based on the assumption that the PCS is not known to have occurred or that the date of occurrence is unknown (Section 6.3). We find, in some cases, that the shift term is significant at the 5% or 10% level, confirming the overall importance of controlling for this possibility when comparing trends.

If the date of the PCS is taken as given and exogenous, we find that the models project significantly more warming in both the lower- and mid-troposphere than are found in weather balloon records over the interval. This finding remains robust if we treat the date of the PCS as unknown and apply the conservative data-mining approach. In fact this finding is robust whether or not we include a level shift in the regression model: we reject equivalence of the trend slopes between the observed and model-generated temperature series either way. The evidence against equivalence is simply stronger when we control for a level shift and this is true whether we treat the date of the shift to be known or unknown.

We also find that if the date of the PCS is assumed to be known then: a) the appearance of positive and significant trend slopes in the individual observed temperature series vanishes once we control for the effect of the level shift, and b) we find statistical evidence for a level shift in some but not all observed temperature series. If the date of the PCS is assumed to be unknown, statistical evidence remains for a level shift in the mean of the observed mid-troposphere series but is weak in the lower troposphere series. This is not
surprising given that we use the data-mining robust critical value which decreases the power of detecting such a shift.

2 BASIC SET-UP WITH NO SHIFT OR KNOWN SHIFT DATE

2.1 TREND MODELS

The literature on estimation and inference in model (1) is by now well established, and it may hardly seem possible that there is something new to be said on the subject. In fact the last decade or so has seen some very useful methodological innovations for the purpose of computing robust confidence intervals, trend significance and trend comparisons in the presence of autocorrelation of unknown form. Many of these robust estimators use the nonparametric HAC approach which is now widely used in econometrics and empirical finance literatures. In contrast the nonparametric HAC approach is used less in applied climatic or geophysical papers although nonparametric approaches have been proposed by Bloomfield and Nychka (1992) and further examined by Woodward and Gray (1993) and Fomby and Vogelsang (2002) for the univariate case. As far as we know, McKitrick et al. (2010) is the first empirical climate paper to apply nonparametric HAC methods in multivariate settings.

It will be convenient to define a general deterministic trend model which contains (1) and (2) as special cases:

\[ y_{it} = \beta_t d_{0t} + \delta_t d_{1t} + u_{it} \]  

where \( d_{0t} \) is a single deterministic regressor (typically the time trend in our applications) and \( d_{1t} \) is a \( k \times 1 \) vector of additional deterministic regressors (typically the intercept and shift terms) and \( \delta_t \) is the corresponding \( k \times 1 \) vector of parameters. Model (1) is thus obtained for \( d_{0t} = t, \beta_t = b_l, \) and \( d_{1t} = 1, \delta_t = a_i, \) and model (2) is obtained for \( d_{0t} = t, \beta_t = b_l, \) and \( d_{1t} = (1, DU_t)' \), \( \delta_t = (a_i, g_l)' \). Notice that we are assuming that each time series \( y_{it} \) has the same deterministic regressors. This is needed for the \( VF \) statistic to be robust to unknown conditional heteroskedasticity and serial correlation. In some applications it might be reasonable to model some of the series as having different trend functions. For example, we know that the climate model series in the application do not have level shifts because level shifts are not part of the climate
model structures. When we think series could have different functional forms for the trend, we can simply include in \( d_{it} \) the union of trend regressors across all the series. While this will result in a loss of degrees of freedom, in many applications the regressors will be similar across series, so the loss in degrees of freedom will often be small. We view this loss of degrees of freedom as a small price to pay for robustness to unknown forms of conditional heteroskedasticity and autocorrelation.

We estimate model (3) using OLS equation by equation. OLS has some nice properties in our set up. Because the regressors are the same for each equation, we have the well-known exact equivalence between OLS and generalized least squares (GLS) estimators that account for cross series correlation. Because we have covariance stationary errors, the well-known Grenander and Rosenblatt (1957) result applies in which case OLS is also asymptotically equivalent to GLS estimators that account for serial dependence in the data.

Defining the \( n \times 1 \) vector \( \beta = (\beta_1, \beta_2, \ldots, \beta_k)' \), and the \( k \times n \) matrix \( \delta = (\delta_1, \delta_2, \ldots, \delta_n) \), model (3) can be written in vector notation as

\[
y_t = \beta d_{ot} + \delta' d_{1t} + u_t. \tag{4}
\]

The parameters of interest are in the vector \( \beta \) so it is convenient to express the OLS estimator using the “partialling out” result for linear regression, aka the Frisch-Waugh-Lovell result (see Davidson and MacKinnon, 2004 and Wooldridge, 2005) as follows. Let \( d_{ot} \) and \( \tilde{y}_t \) denote respectively the OLS residuals from the regression of \( d_{ot} \) on \( d_{1t} \) and the regression of \( y_t \) on \( d_{1t} \). The OLS estimator of \( \beta \) can be written as

\[
\hat{\beta} = \left( \sum_{t=1}^{T} \tilde{d}_{ot}^2 \right)^{-1} \sum_{t=1}^{T} \tilde{d}_{ot} \tilde{y}_t \tag{5}
\]

and it directly follows that

\[
\hat{\beta} - \beta = \left( \sum_{t=1}^{T} \tilde{d}_{ot}^2 \right)^{-1} \sum_{t=1}^{T} \tilde{d}_{ot} u_t. \tag{6}
\]

Note that the form of \( \hat{\beta} \) in equation (5) would be unchanged if we redefined \( d_{ot} \) to be the shift term and \( d_{1t} \) to be the intercept and trend terms, however the definition of the \( \sim \) variables would be adjusted accordingly. This implies that the test statistic on hypotheses about the shift term will take the same form as those for the trend terms when the shift date is known.
2.2 THE VF TEST

We are interested in testing null hypotheses of the form

\[ H_0: R\beta = r \]  \hspace{1cm} (7)

against alternatives \( H_0: R\beta \neq r \) where \( R \) and \( r \) are known restriction matrices of dimension \( q \times n \) and \( q \times 1 \) respectively where \( q \) denotes the number of restrictions being tested. The matrix \( R \) is assumed to have full row rank. Robust tests of \( H_0 \) need to account for correlation across time, correlation across series, and conditional heteroskedasticity as summarized by the long run variance of \( u_t \). This is defined as

\[ \Omega = \Gamma + \sum \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma \Gamma

The VF statistic is constructed using the following estimator of \( \Omega \):

\[ \hat{\Omega}_T = \hat{\Gamma}_0 + \sum_{j=1}^{T} \left( 1 - \frac{j}{T} \right) (\hat{\Gamma}_j + \hat{\Gamma}'_j), \]  \hspace{1cm} (8)

where \( \hat{\Gamma}_j = E(u_t u_{t-j}') \) is the matrix autocovariance function of \( u_t \). Those familiar with the time series literature will notice that \( \Omega \) is proportional to the spectral density matrix of \( u_t \) evaluated at frequency zero.

In the Supplementary Information (SI) we provide a finite sample motivation for the form of \( \hat{\Omega}_T \). Also note that (8) was originally proposed by Kiefer, Vogelsang and Bunzel (2000, 2001) although in the different but computationally identical form:

\[ \Omega_T = 2T^{-2} \sum \hat{S}_t \hat{S}_t', \]  \hspace{1cm} (10)

where \( \hat{S}_t = \sum_{j=1}^{T} \hat{u}_j \). See Kiefer and Vogelsang (2002) for a formal derivation of the exact equivalence between (8) and (10).

When only one restriction is being tested \( (q = 1) \), we can define a \( t \)-statistic version of \( VF \) as
2.3 Asymptotic Limit of the \( VF \) Statistic

We now provide sufficient conditions for obtaining an asymptotic approximation to the sampling distribution of the \( VF \) statistic. A formal proof is given in the SI. The fraction \( c \in (0,1) \) of the sample is \( cT \) and we denote the integer portion of this quantity as \( \lfloor cT \rfloor \). The symbols \( \Rightarrow \) and \( \rightarrow \) denote weak convergence and convergence in distribution, \( \Lambda \) is the matrix square root of \( \Omega \) (i.e. \( \Omega = \Lambda \Lambda' \)) and \( -\text{8}^\Omega^\text{32} \) denotes a \( -\text{8}^\text{2} \times 1 \) vector of independent standard Wiener processes where \( -\text{8} \) is a positive integer.

Two assumptions are sufficient for obtaining the limit of \( VF \). Define the partial sums of \( \text{38} \) as

\[
S_t = \sum_{i=1}^{t} u_i.
\]

The first assumption is that a functional central limit theorem (FCLT) holds for \( S_T \). As \( T \to \infty \),

\[
T^{-1/2} S_{\lfloor cT \rfloor} \Rightarrow \Lambda W_n(c). \tag{11}
\]

The second assumption is related to the deterministic regressors in the model. Assume that there is a scalar \( \tau_{0T} \), and a \( \text{38} \times \text{38} \) matrix \( \tau_{1T} \), such that

\[
T^{-1} \sum_{t=1}^{\lfloor cT \rfloor} \tau_{0T} d_{0t} \to \int_0^c f_0(s) ds \quad \text{and} \quad T^{-1} \sum_{t=1}^{\lfloor cT \rfloor} \tau_{1T} d_{1t} \to \int_0^c f_1(s) ds. \tag{12}
\]

For example, in model (2), \( d_{0t} = t, d_{1t} = (1, DU_t)' \), \( \tau_{0T} = T \), \( \tau_{1T} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), \( f_0(s) = s \) and \( f_1(s) = (1, DU(\lambda > s))' \) where, in this case, \( DU \) denotes a continuous indicator function taking the value 1 if \( \lambda > s \) and 0 otherwise.

Stack \( f_0(s) \) and \( f_1(s) \) into a vector \( f(s)' = (f_0(s), f_1(s)') \) and define the stochastic process

\[
\mathbf{B}_q'\mathbf{f}(c) = \int_0^c dW_q(s) - \left( \int_0^1 dW_q(s) f(s)' \right) \left( \int_0^1 f(s)f(s)' ds \right)^{-1} \int_0^c f(s)ds, \tag{13}
\]

In the SI we show that under assumptions (11) and (12), the limit of \( VF \) under the null hypothesis (7) is given by

\[
VF \rightarrow \mathcal{N} \left[ 2 \int_0^1 \mathbf{B}_q'(c) B_q'(c)' dc \right]^{-1} Z_q / q \equiv VF_{q}^{\infty}, \tag{14}
\]
where $Z_q \sim N(0, I_q)$ and is independent of the random matrix $2 \int_0^1 B_q^f(c)B_q^f(c)' dc$. The limit of the $VF$ statistic can therefore be seen to be similar to an $F$ random variable, however it follows a nonstandard distribution that depends on the deterministic regressors in the model via the stochastic process $B_q^f(c)$. The critical values of $VF_q^\infty$ thus depend on the regressors in $d_{it}$ and by extension depend on the value of $\lambda$ when a level shift dummy variable is included in the model. It is important to note that the critical values do not depend on which regressors are placed in $d_{ot}$ (the regressor of interest for hypothesis testing). For a given value of $\lambda$, one uses the same critical values for testing the equality of trend slopes or testing hypotheses about the intercepts or testing hypotheses about level shifts in model (2).

Obtaining the critical values of the nonstandard asymptotic random variable defined by (14) is straightforward using Monte Carlo simulation methods that are widely used in the econometrics and statistics literatures. In the application, when we take the date of the Pacific Climate Shift to be exogenously given at 1977:12, this yields a value of $\lambda = 0.3636$. For model (2) with $\lambda = 0.3636$ we simulated the asymptotic critical values of $VF$ for testing one restriction ($q = 1$) which we tabulate in Table 1a. The Wiener process that appears in the limiting distribution is approximated by the scaled partial sums of 1,000 i.i.d. $N(0,1)$ random deviates. The vector $f(s)$ is approximated using $\int_0^1 f(t)BT(t > 0.3774T), t/T)'$ for $t=1,2,\ldots,T$. The integrals are approximated by simple averages. 50,000 replications were used. We see from Table 1a that the right tail of the $VF$ statistic is fatter than that of a $\chi_1^2$ random variable.

### 2.4 Bootstrap Critical Values and $p$-Values

If carrying out simulations of the asymptotic distributions is not easily accomplished using standard statistical packages, an alternative is to use a simple bootstrap which is described in detail in the SI. Residuals from Equation (4) can be resampled and used to compute $\hat{\Omega}$ from Equation (8) and $VF$ from Equation (9); then the percentiles of the bootstrapped $VF$ statistic in many repetitions can provide critical values and $p$ values. A particular advantage of the $VF$ method is that its asymptotic null distributions do not depend on unknown correlation parameters, and it falls within the general framework considered by Gonçalves and Vogelsang (2011), where it was shown that the simple, or naïve, i.i.d. bootstrap will generate valid critical values. No special methods, such as blocking, are required here, and the bootstrap critical values are asymptotically equivalent to the distribution given by (14).
3 TREATING THE SHIFT DATE AS UNKNOWN

Many previous authors (e.g. Seidel and Lanzante 2004) have treated the date of the level shift as known because the PCS was an exogenous event observed across many different climatic data series. As a robustness check we also report results where we treat the date of the level shift as unknown. We take a "data-mining" approach which has a long history in the change point literature. For a given hypothesis, we compute the $VF$ statistic for a grid of possible shift dates and determine the one that gives the largest $VF$ statistic. In other words, we search for the shift date that gives the strongest evidence against the null hypothesis. The effect of on critical values of searching over shift dates must be taken into account, otherwise this approach would be a “data-mining” exercise that could give potentially misleading inference. The level of the test will be inflated above the nominal level compared to the case where the shift date is assumed to be known. Fortunately, it is easy to obtain critical values that take into account the search over shift dates.

For a given potential shift date $T_p$, let $VF(\lambda)$ denote the $VF$ statistic for testing a given null hypothesis. The limiting random variable given by (14) depends on $\lambda$ through the level shift regressor and we now label the limit by $VF^\omega_q(\lambda)$ to make explicit the dependence on the shift date used to estimate the model. For technical reasons (see Andrews (1993)) we need to “trim” the fraction $v$ from each end of the sample, leaving a grid of potential shift dates given by $vT + 1, vT + 2, ..., T - vT$ (in our application we set $v=0.1$). Define the “data-mined” $VF$ statistic as

$$supVF = sup_{\lambda} VF(\lambda) \text{ for } \lambda \in (v, 1 - v)$$

Under the null hypothesis (7) and under the assumption there is no level shift in the data, we have

$$supVF \rightarrow sup_{AE(v, 1-v)} VF^\omega_q(\lambda)$$

(15)

where the limit follows from (14) and application of the continuous mapping theorem. Using simulation methods identical to those used for the known shift date case, we computed asymptotic critical values for $supVF$ for $v=0.1$ and $q=1$ for testing hypotheses about the trend slope parameters in model (2). These critical values are given in Table 1b. Using the $supVF$ statistic along with the critical values given by (15) provides a very conservative test with regard to the shift date.
4 TESTING FOR A LEVEL SHIFT IN A UNIVARIATE TIME SERIES

As part of the empirical application we provide visual evidence that the observed temperature series exhibit level shifts around the time of the PCS. Some formal statistical evidence regarding these level shifts can be provided by application of the $VF$ statistic to an individual time series. Consider model (2) for the case of $n=1$ and place the model in the general framework (3) with $d_{0t} = DU_t$, $d_{1t} = (1, t)', \beta_t = g_t$ and $\delta_t = (a_t, b_t)'$. If we take the shift date as known, then the $VF$ statistic for testing for no level shift ($H_0: g_t = 0$) can be computed as before using (9) with $R=1$ and $r=0$. The asymptotic null critical values are still given by Table 1a.

If we treat the shift date as unknown, we can apply the sup$VF$ statistic although the asymptotic critical values depend on which regressor is placed in $d_{0t}$. While it is true that for a given value of $\lambda$, the distribution of $VF_q^\omega(\lambda)$ is the same regardless of the regressor placed in $d_{0t}$, the covariance structure of $VF_q^\omega(\lambda)$ across $\lambda$ depends on which regressor is placed in $d_{0t}$. Therefore, the sup$VF$ statistic when testing for a zero trend slope has different asymptotic critical values than the sup$VF$ statistic for testing a zero level shift. We simulated the asymptotic critical values of sup$VF$ for testing for a zero level shift for the case of $\nu=0.1$ and $q=1$ and provide those critical values in Table 1b.

Other tests for a level shift at an unknown date of a trending time series have been proposed in the empirical climate literature. Reeves et al (2007) provides a review of change point detection methods developed in the climate literature but the review focuses on tests designed for time series variables that do not have serial correlation. In contrast Lund et al (2007) propose a test for a level shift that allows a specific form of autocorrelation - the first order periodic autoregressive model. We prefer the $VF$ approach for two reasons. First, the $VF$ approach is robust to more general forms of autocorrelation. Second, we formally derive and characterize the limiting null distribution of the sup statistic and this allows us to tabulate null critical values. Lund et al (2007) also use a sup-type statistic but they do not provide any asymptotic theory that can be used to generate valid approximate critical values. A recent paper by Gallagher, Lund and Robbins (2013) develops asymptotic theory for a level shift test that treats the shift date as unknown but their analysis is confined to trend models where $u_{it}$ is assumed to be uncorrelated over time. What seems to be missing from the empirical climate literature are level shift tests that allow the shift date to be unknown and permit serial
correlation in $u_t$. Fortunately there are several papers in the econometrics literature that propose level shift tests for trending series that have these properties, see Ploberger and Krämer (1996), Vogelsang (1997) and Sayginsoy and Vogelsang (2011). While clearly well outside the scope of this paper, it would be interesting to compare the sup-$VF$ test for a shift in trend at unknown date with the other tests proposed in the literature.

5 Finite Sample Performance of the VF Statistic

In this section we report some results from a small Monte Carlo simulation study that demonstrates the finite sample performance of the $VF$ statistic. We compare the performance of the $VF$ statistic with a traditional Wald statistic. The Wald statistic is configured to be robust to heteroskedasticity and serial correlation over time and to be robust to correlation across series. We use two established methods for constructing the Wald statistic. The first method is based on the parametric estimator of $\Omega$ given by

$$\hat{\Omega}_M = \hat{\gamma}_0 + \sum_{j=1}^{M-1} (1 - \frac{j}{M}) (\hat{\gamma}_j - \hat{\gamma}_{j+1}),$$

which is the Bartlett kernel estimator. The value of the bandwidth, $M$, was chosen by the data dependent method proposed by Andrews (1991) based on the AR(1) plug-in method. This data dependent method tends to choose values of $M$ that are small relative to $T$ although $M$ tends to be larger when serial correlation in the data is strong compared to cases where serial correlation is weak. The second method also uses (16) but with prewhitening based on autoregression models with lag 1 (AR(1)) fit to the components of $\hat{u}_t$. Prewhitening was explored by Andrews and Monahan (1992). We set $M=1$ in the prewhitening case which makes the estimator of $\Omega$ an AR(1) parametric estimator. We compute Wald-type statistics based on these two estimators of $\Omega$ using (9) with $\hat{\Omega}_T$ replaced with either $\hat{\Omega}_M$ or the prewhitened estimator. The resulting Wald statistics are denoted by $W_{B\text{art}}$ and $W_{PW}$ respectively. Under the assumptions used to derive the limiting null distribution of the $VF$ statistic, it is well known that $W_{B\text{art}}$ and $W_{PW}$ have limiting null distributions given by $\chi^2_q$ where $\chi^2_q$ denotes a chi-square random variable with $q$ degrees of freedom.

We use the following data generating process:

$$y_{1t} = g_1 D_{U_t}(\lambda) + b_1 t + u_{1t},$$

$$y_{2t} = b_2 t + u_{2t},$$

We use the following data generating process:

$$y_{1t} = g_1 D_{U_t}(\lambda) + b_1 t + u_{1t},$$

$$y_{2t} = b_2 t + u_{2t},$$

John Wiley & Sons
where $u_{1t} = \frac{1}{\kappa}u_{1t}^*,$ $u_{2t} = \frac{1}{\sqrt{1+\eta^2}}(\frac{1}{\kappa}u_{2t}^* + \eta u_{1t}^*), \quad u_{it}^* = \rho_1 u_{i,t-1}^* + \rho_2 u_{i,t-2}^* + \epsilon_{it}, \epsilon_{it} \sim \text{iid} \mathcal{N}(0,1), \text{cov}(\epsilon_{1t}, \epsilon_{2t}) = 0,$ and $u_{i0}^* = 0,$ $\kappa = \frac{1-\rho_2}{\sqrt{[1-\rho_2]^2-\rho_1^2][1+\rho_2]}.$ The errors, $u_{it}^*$, are configured to have unit variances with $\text{cov}(u_{1t}, u_{2t}) = \frac{\eta}{\sqrt{1+\eta^2}}.$ When $\eta = 0,$ $y_{1t}$ and $y_{2t}$ are uncorrelated with each other.

We report two sets of results. The first set of results focus on empirical null rejection probabilities. For $y_{1t}$ we set $b_1 = .01$ and $g_1 = 0$ so that there is no level shift in $y_{1t}.$ For $y_{2t}$ we set $b_2 = .01$ so that the null hypothesis of equal trend slopes holds. We set $\eta = 0$ so that the two series are uncorrelated. We report results for $T=120, 240, 636$ and a selection of values of $\rho_1$ and $\rho_2.$ In all cases $50,000$ replications were used and we computed empirical rejection probabilities for the $W_{\text{Bart}}, W_{\text{PW}}$ and $VF$ statistics for testing $H_0: b_1 = b_2$ using the appropriate asymptotic critical values. The simulation results for this configuration highlight the impact of serial correlation structure on null rejection probabilities relative to the sample size.

The results are tabulated in Table 2a. There are two sets of results reported for each of the three statistics. The first set of results corresponds to the case where no level shift dummy variable is included in the estimated model. The second set of results corresponds to the case where the level shift dummy is included in the estimated model. In this case we also report results for the sup$VF$ statistic. Results are organized into three blocks corresponding to the three sample sizes. Within a block results are given for seven configurations of the autoregressive parameters ranging from no serial correlation to strong serial correlation.

If the asymptotic approximations were working perfectly for the statistics, we would see rejections of 0.05 in all cases. When the serial correlation is absent, all statistics have empirical rejection probabilities close to 0.05 regardless of the sample size. Once there is serial correlation is the model, over-rejections can occur depending on the strength of the serial correlation relative to the sample size. First focus on the case of AR(1) errors ($\rho_2 = 0$). Rejections tend to be close to 0.05 when $\rho_1$ is small but as $\rho_1$ increases in value, rejections tend to increase. This is especially true for the $W_{\text{Bart}}$ statistic where rejections exceed 0.25 when $\rho_1 = 0.9.$ In contrast, $W_{\text{PW}}$ and $VF$ suffer from less severe over-rejection problems although they tend to be over-sized when $T=120$ and $\rho_1 = 0.9.$ For a given value of $\rho_1,$ as $T$ increases, over-rejections tend to fall for all three statistics but slowest for $W_{\text{Bart}}.$ Overall for the AR(1) error case, $W_{\text{PW}}$ and $VF$ have similar rejections to

John Wiley & Sons
each other and outperform $W_{Bart}$. The sup$V_F$ statistic tends to over-reject more than $V_F$ when serial correlation is strong although the differences between sup$V_F$ and $V_F$ decrease as the sample size increases. It is a common finding that supremum statistics tend to have more over-rejection problems than statistics that treat break dates as known.

One of the reasons that $W_{pw}$ performs relatively well with AR(1) errors is that $W_{pw}$ is explicitly designed for AR(1) error structures. But, when the errors are not AR(1), $W_{pw}$ can suffer from over-rejection and under-rejection problems. Consider the case $\rho_1 = 0.3, \rho_2 = 0.3$ where $W_{pw}$ shows substantial over-rejections that are larger than $W_{Bart}$ and $V_F$. These over-rejections tend to persist as $T$ increases. In contrast $V_F$ is much less distorted and rejections approach 0.05 as $T$ increases. For the case of $\rho_1 = 0.9, \rho_2 = -0.3, W_{pw}$ under-rejects and the under-rejection problem becomes more severe as $T$ increases whereas $V_F$ has rejections close to 0.05 for all sample sizes. The $W_{Bart}$ statistic tends to over-reject mildly in this case.

In general, Table 2a indicates that the $V_F$ statistic has the least over-rejection problems and is the better statistic with regard to control of type 1 error.

In the second set of results we use $T=660$ to match the empirical application. We now include a level shift in $y_{1t}$ with $\lambda = 0.3636$ and we set $b_1 = 0.01$ and $g_1 = 0.25$. For $y_{2t}$ we set $b_2 = 0.01, 0.0105, 0.011, 0.0116, 0.0121$. We report results for $\eta = 0, 0.5$. While we ran simulations for a wide range of values for $\rho_1$ and $\rho_2$, we only report results for $\rho_1 = 0, 0.9$ and $\rho_2 = 0$ given that results for other serial correlation configurations have similar patterns to what is reported in Table 2a.

The results are given in Table 2b. The first block of 20 rows gives results for $\eta = 0$ whereas the second block of 20 rows gives results for $\eta = 0.5$. Within each $\eta$ block, results are first given for $g_1 = 0$ followed by results for $g_1 = 0.25$. For each value of $g_1$, results are given for $\rho = 0$ followed by results for $\rho = 0.9$. When $b_2 = 0.01$, we are observing null rejection probabilities whereas for other values of $b_2$ we are observing power.

First focus on the results when the null hypothesis is true, i.e. $b_2 = 0.01$. For $\rho_1 = 0$ we see that when the level shift dummy is included, we have rejections close to 0.05 for all statistics. However, when the level shift dummy is not included and $g_1 = 0.25$, we observe severe over-rejections which range from 0.292 to 0.442.

John Wiley & Sons
The statistic with the least severe over-rejection problem is $VF$. When $\rho_1 = 0.9$, we have relatively strong autocorrelation in the data. When either $g_1 = 0$ or the level shift dummy is included in the model, there are some mild over-rejection problems ranging from 0.069 for $VF$ to 0.132 for $W_{\text{start}}$ with $W_{\text{PW}}$ in between. Over-rejections are slightly worse when $\eta = 0.5$ compared to $\eta = 0$. As we saw in Table 2a, sup$VF$ tends to over-reject slightly more than $VF$ when autocorrelation is strong. In addition, when there is a level shift in the data, sup$VF$ tends to have rejections above 0.05. This happens because sup$VF$ nests the null hypotheses of equal trend slopes and no level shift. A rejection using sup$VF$ indicates a level shift and/or differences in trend slopes.

Now focus on the cases where $b_2 > 0.01$. In these cases $y_{2t}$ has a bigger trend slope than $y_{2t}$ and we should be rejecting the null of equal trend slopes. When $g_1 = 0$, we see that all statistics have good power when $\rho_1 = 0$ and power is higher for $\eta = 0.5$ compared to $\eta = 0$. Power increases as expected as $b_2$ increases. Across the three statistics, $VF$ tends to have lower power than other two statistics. This illustrates the well known tradeoff between over-rejection problems and power. Note that while power of $VF$ is lowest, its power is still relatively good in an absolute sense. If we include the level shift dummy in the model even though it is not needed ($g_1 = 0$), all three tests show a reduction in power as one would expect. An unexpected finding is that the sup$VF$ statistic has higher power than $VF$ and the two Wald statistics when the level shift regressor is included but there is no shift in the data. In contrast and as expected, power of sup$VF$ is lower than the tests for the case where the level shift regressor is not included in the estimated model.

The most interesting power results occur for $g_1 = 0.25$ and $b_2 = 0.0105$ when the level shift dummy is left out of the model. In this case the estimator of $b_1$ is biased up and one can show that the probability limit of the estimator of $b_1$ exactly equals 0.0105. For the case of $\rho_1 = 0$, rejections of all three statistics are close to the nominal level of 0.050. This shows that an omitted level shift variable can cripple the power of the tests to detect a difference in trend slopes between two series. For larger values of $b_2$ the tests have power even if the level shift variable is not included. When $b_2 = 0.011$, power is higher if the level shift dummy is included whereas for $b_2 = 0.012$ power is higher when the level shift dummy is left out. When a level shift is present in the data, sup$VF$ has less power overall than $VF$ as expected given that sup$VF$ treats the break date as unknown and uses conservative critical values.
These simulation results show that: i) the $VF$ statistic has type 1 errors closest to the nominal level, ii) the $VF$ statistic has lower power which is the price paid for more accurate type 1 error, however the power of $VF$ is still reasonably good, iii) including a level shift dummy when there is no level shift in the data lowers power, iv) failure to include a level shift dummy when there is a level shift in the data causes type 1 errors to be excessively larger than the nominal level and, depending on the magnitude/direction of the level shift, can make it difficult to detect slopes that are different, v) positive correlation across series ($\eta = .5$) tends to increase power and vi) stronger serial correlation tends to inflate over-rejections under the null while reducing power.

6 APPLICATION: DATA AND METHODS

6.1 OBSERVATIONS AND MODEL DATA

Our empirical application uses data from the tropical lower- and mid-troposphere (LT, MT respectively), where we will compare trends from a large suite of general circulation models (GCMs) to those observed in three monthly radiosonde records over the 1958-2012 interval ($T = 660$). Held and Soden (2000), Karl et al. (2006) and Thorne et al. (2011) provide discussions of the importance of this region for assessing climate models. In response to rising greenhouse gas levels, models predict a maximum warming trend will occur in the lower- to mid-troposphere over the tropics. Karl et al. (2006, p. 11) noted that weather balloons had not detected such a trend and deemed the mismatch a “potentially serious inconsistency” with models. Douglass et al. (2007) argued that the inconsistency was real and statistically significant, while Santer et al. (2008) countered that the difference was not significant if autocorrelation was taken into account, though they only used data from 1979 to 1999 and a simple AR1 correction. McKitrick et al. (2010) extended the data to 2009 and employed the $VF$ approach, concluding the trend differences were significant. Fu et al. (2011) also found climate models significantly exaggerate the gain in the warming trend with altitude throughout the tropics. Additionally, Bengtsson and Hodges (2011) and Po-Chedley and Fu (2012) found that even models constrained to match observed post-1979 sea-surface temperatures overestimated warming trends in the topical mid-troposphere.
While the trend discrepancy is now well-established, Santer et al. (2011) emphasize the need for multidecadal comparisons to identify whether it is structural or temporary. By using the half century-length weather balloon records we meet this concern, but we also span the 1977-78 Pacific Climate Shift. All climate models predict a steady trend in response to rising greenhouse gases, and none predict a large one-time jump preceded and followed by decades of static temperatures. Hence we satisfy the conditions described in the introduction, namely that the underlying phenomenon implies a trend as opposed to a jump, and our null hypothesis of trend equivalence requires controlling for a potential step-change at a potentially unknown date.

Our application mainly focuses on trend slopes and comparisons of trend slopes across series in which case we set $d_{ti} = t$ and therefore $\beta_i = b_i$. For model (2) $d_{it} = (1, DU_t)'$ with the level shift set at 1977:12, implying $\lambda = 0.3636$. We also provide some results on the level shift parameters themselves in which case $d_{0t} = DU_t, \beta_i = g_i$ and $d_{it} = (1, t)'$. Let $\hat{\beta}_i$ denote the OLS estimator of $\beta_i$ for a given parameter of interest for a given time series using either model (1) or model (2). If only one restriction is being tested ($q=1$) of the form $H_0: \beta_i = 0$ then we can write $V F_t$ as

$$V F_t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$$

where

$$se(\hat{\beta}_i) = \sqrt{\sum_{t=1}^{T} \frac{d_{0t}^2}{\hat{\Omega}_T}},$$

and $\hat{\Omega}_T$ is computed using (8) or (10) using $\hat{u}_{it}$ from the respective models. Let $cv_{0.025}$ denote the 2.5% right tail critical value of the asymptotic distribution of $V F_t$. For model (1) $cv_{0.025} = 6.482$ (see Table 1 of Vogelsang and Franses, 2005; their $t^*_c$ statistic) and for model (2) $cv_{0.025} = 7.028$ (see Table 1a). A 95% confidence interval (CI) is computed as $\hat{\beta}_i \pm se(\hat{\beta}_i) \cdot cv_{0.025}$.

The tropics are defined as 20N to 20S. The GCM runs are the same as those used in McKitrick et al. (2010), and were the ones used for the Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC 2007). There were 57 runs from 23 models for each of LT and MT layers. Each model uses prescribed forcing inputs up to the end of the 20th century climate experiment (see Santer et al. 2005), and
most models include at least one extra forcing such as volcanoes or land use. Projections forward after 2000 use the A1B emission scenario. Tables 3 and 4 report, for the LT and MT layers respectively, the climate models, the extra forcings, the number of runs in each ensemble mean, estimated trend slopes in the cases with and without level shifts, and $VF$ standard errors. All series had a significant AR1 coefficient, but as reported in McKitrick et al. (2010), over two-thirds also have significant higher-order AR terms as well, which motivates the use of a HAC estimator as opposed to a simple AR(1) treatment as in Santer et al. (2008) and Fu et al. (2011).

We used three observational temperature series. The HadAT radiosonde series is a set of MSU-equivalent layer averages published on the Hadley Centre website\(^1\) (Thorne et al. 2005) spanning 1958 to 2012. We use the 2LT layer to represent the GCM LT-equivalent and the T2 layer to represent the GCM MT-equivalent. The other two series are denoted RAdiosonde Observation Bias Correction using Reanalyses (RAOBCORE, Haimberger 2005) and Radiosonde Innovation Composite Homogenization (RICH, Haimberger et al. 2008). Both were obtained from the Institute for Meteorology and Geophysics at the University of Vienna,\(^2\) however this site does not provide the data in LT- and MT-equivalent forms so MSU-equivalent layer averages of the tropical latitudes were computed for us by John Christy (pers. comm.) using weighting functions that match those used for the HadAT series. We did not use the RATPAC series published by the National Oceanic and Atmospheric Administration (NOAA\(^3\)) since the zonal averages are only available in quarterly or annual form. Another series, called IUK-radiosonde from the University of New South Wales,\(^4\) only goes up to 2005.

The HadAT and RICH series deal with the problem of homogenizing short data segments by comparing series at suspected breakpoints to reference series formed using nearby observations to detect if shift terms are needed. The RAOBCORE series uses reference series generated by nearby weather forecasting systems (called "reanalysis data"). Production of these series is therefore an application of the shift-detection methods developed in this paper, but for our current purposes we will take the data as given and apply it to the model-observation comparison.

\(^1\) http://www.metoffice.gov.uk/hadobs/hadat/msu/anomalies/hadat_msu_tropical.txt.
\(^2\) http://www.univie.ac.at/theoret-met/research/raobcore/index.html.
\(^3\) Available at http://www1.ncdc.noaa.gov/pub/data/ratpac/
\(^4\) Available at http://www.crcr.unsw.edu.au/staff/profiles/sherwood/radproj/index.html
The last three lines of Tables 3 and 4 report the estimated trend slopes and VF standard errors for the observed temperature series. Each panel of Figure 2 displays the observed (LT and MT) model-simulated data (red dots), with a least squares trend line through the model mean, allowing for a break at 1977:12, shown in dark red. The estimated LT trends for, respectively, HadAT, RICH and RAOB CORE are 0.135, 0.126 and 0.149 °C/decade. The MT trends are, respectively, 0.086, 0.090 and 0.117 °C/decade. The effect of allowing for a level shift (step-change) is shown in Figure 3. The trend through the average of the three radiosonde series, allowing for a break, is shown in blue. Using a shift date of 1977:12 the observed LT trends fall to 0.070, 0.090 and 0.054 °C/decade and the MT trends fall to 0.004, 0.044 and 0.025 °C/decade. Thus about half of the positive LT trend in Figure 2 can be attributed to the one-time change at 1977:12 and essentially all the MT change is accounted for by the step-change.

The shift terms are insignificant in all model runs. In the observational series, one of three is significant at 5% in the LT layer, and in the MT layer one is significant at 10% and one at 5%. It might seem surprising that the effect of the shift dummy is so dramatic on the balloon series trend slope parameters, yet the shift coefficients themselves are not more strongly significant. But in general, trend slopes are estimated more efficiently than level shifts (or intercepts) and therefore it is more difficult to conduct inference about level shifts than about trend slopes. While there is sufficient noise in the data to make inference about level shifts difficult, it is nonetheless clear that the possibility of a level shift must be taken into account, and the noise is not so large as to mask information about the trend slopes. Unmodeled level shifts also induce spurious noise into the model when conducting inference about trend slopes. Therefore, controlling for a possible level shift makes inferences about trend slopes more informative.

Figure 4 plots all the estimated trend slopes along with their 95% CIs. The top row leaves out the level shift and the bottom row includes it. The model-generated trends are grouped on the left in red. The trends are ranked from smallest to largest and the numbers above each marker refer to the GCM number (see Table 2 for names). The three blue trends on the right edge are, respectively, the Hadley, RICH and RAOB CORE series. With or without the shift term the range of model runs and their associated CI's overlaps with those of the observations. In that sense we could say there is a visual consistency between the models and observations. However, that is too weak a test for the present purpose, since the range of model runs can be
made arbitrarily wide through choice of parameters and internal dynamical schemes, and even if the reasonable range of parameters or schemes is taken to be constrained on empirical or physical grounds, the spread of trends in Figure 4 (spanning roughly 0.1 to 0.4 °C/decade in each layer) indicates that it is still sufficiently wide as to be effectively unfalsifiable. Also, if we base the comparison on the range of model runs rather than some measure of central tendency it is impossible to draw any conclusions about the models as a group, or as an implied physical theory. Using a range comparison, the fact that, in Figure 4, models 8, 5 and 16 are reasonably close to the observational series does not provide any support for models 2, 3 and 4, which are far away. We want to pose the trend comparison in a form that tells us something about the behaviour of the models as a group, or as a methodological genre, and this requires using the multivariate testing framework.

6.2 Multivariate Trend Comparisons: No-shift and Known Shift Date Cases

For each layer we now treat the 23 climate model generated series and the 3 observational series as an \( n=26 \) panel of temperature series. We estimate models (1) and (2) using the methods described in Section 3. The parameters of interest are the trend slopes. We are interested in testing the null hypothesis that the weighted average of the trend slopes in the 23 climate model generated series is the same as the average trend slope of the observed series. The weight coefficient \( w_i \) equals the number of runs in model \( i \)’s ensemble mean, to adjust for the reduction in variance in multi-run ensemble means. Placing the observed series in positions \( i=24,25,26 \) the restriction matrices for this null hypothesis are

\[
R = \begin{bmatrix}
    w_1/57, \ldots, w_{23}/57, -1/3, -1/3, -1/3
\end{bmatrix}, \quad r = 0
\]

where the \( w_i \) terms sum to 57.

Table 5 presents the \( VF \) statistics for the test of trend equivalence between the climate models and observed data. Also reported are the \( VF \) statistics for testing the significance of the individual trends of the observed temperature series, the magnitudes of which (°C/decade) are indicated in parentheses beside the series name. Asymptotic critical values are provided in the table captions and significance is indicated as described in the table. We also compute bootstrap p-values for the tests using the method outlined in Section 2.4 using 10,000 bootstrap replications.
In the trend model without level shifts (top panel of Table 5), the zero trend-hypothesis is rejected at the 1% significance level for all 6 observed series, apparently indicating strong evidence of a significant warming trend in the tropical troposphere over the 1958-2012 interval. A test that the climate models, on average, predict the same trend as the observational data sets is rejected in both the LT and MT layers at 1% significance. Table 6 repeats the model-observation trend equivalence test for each of the 23 models individually. In the LT layer the differences are significant at 5% or lower in 8 cases and in 20 cases in the MT layer. (Not reported are the single-model tests of trend significance, though these can be inferred from Tables 2 and 3 and Figure 4. 22 of 23 models have significant trends (at 5%) in both layers without allowing for a break, and 21 of 23 have significant trends in both layers allowing for a break.) So while, on average, the model trends are significantly different from observations, in the LT layer it can at least be said that if we ignore the step-change at 1977:12, almost two-thirds of the models have trends that individually do not significantly differ from the observations.

When we add the level shift dummy at 1977:12 (middle block of Table 5), the trend magnitudes and values of the $VF$ statistics for testing the zero trend-hypothesis drop considerably. The critical values for $VF$ are larger than in the case without the mean-shift dummy. Now none of the observed series has a significant trend. Hence when the level shift is left out, the increase in the series is spuriously associated with a trend slope, whereas the trend is explained by a jump in the data around 1977. The average shift term is not significant at either level (Table 5 rows 17 and 18), though as mentioned above, in three of six individual balloon series, the shift term is significant at the 10% or 5% level.

The $VF$ test of equivalence of average trends between the climate models and observed data is more strongly rejected when the level shift dummy is included. Notice that bootstrap p-values drop to essentially zero in this case. This finding is not surprising because, as is clear in Tables 3 and 4, while the estimated trend slopes decrease for the observed series when the level shift dummy is included, the estimated trend slopes of the climate model series are not systematically affected by the level shift dummy. Therefore, there is a

---

The climate-models do not explicitly model the Pacific Climate Shift and so the level shift coefficient has no special meaning for the climate model data. Not surprisingly, the estimated level shift coefficients were positive in 11 cases and negative in 12 of the climate model series.
greater discrepancy between the climate model trends and the observed trends. Table 6 shows model-by-model comparisons. When a shift with known date is included, the number of 5% rejections in the LT layer rises from 8 to 13 out of 23, but in the MT layer it drops from 20 to 16. This latter change be attributed in part to the increase in the critical values, emphasizing the importance of taking into account the dependence of the test on the introduction of the shift term.

### 6.3 Multivariate Trend Comparisons: Unknown-Shift Date Case

As a robustness check regarding the assumption that the shift date of the PCS is known, we report results where we treat the shift date as unknown and use the supVF statistic. The supVF statistic is calculated by computing the VF score with the shift term $T_b$ sequentially set across the middle 80% of the data set, then selecting the maximum value.

For the tests of model-observational equivalence allowing for a level shift, Figure 5 shows the sequence of VF scores, with 10%, 5% and 1% critical values shown. The supVF occurs at date 1979:6 in the LT layer and at 1979:5 in the MT layer. These shift dates are close to, though not the same as, 1977:12 but all of the VF scores for dates near 1977:12 time interval far exceed the 1% critical values for the supVF statistic. Since the supVF test is conservative regarding the choice of shift date, this provides strong additional support for the results in Section 6.2.

The supVF scores of tests of model-observational equivalence are reported in Table 5 for model averages and in Table 6 for individual models. A pattern emerges particularly clearly in Table 5 that when we apply the data mining approach, the test scores get larger, as expected, but so do the critical values. The net effect is to reduce the significance of the test scores when the shift date is treated as unknown. Had we not used the critical values as given by (15), we might have spuriously inflated the significance of our findings in the event the level shifts were not present in the data. This is a useful lesson in the perils of naïve data-mining, in which a specification is selected that maximizes the chance of rejecting some null hypothesis, without taking into account the effect of the data mining process on the null distribution of the test. The price one pays for being
honest and using the conservative critical values implied by (15) is lower power in detecting a deviation from
the null hypothesis. Even with lower power, we see in the third panel of Table 5 that the average model is still
clearly rejected against the data in both the LT and MT layers. Since this result is not dependent on choosing a
particular shift date it provides strong confirmation of the empirical finding assuming a known shift date.

Regarding the trend magnitudes, we do not report the supVF score for the test of a zero trend in the 3rd
block of Table 5. Recall that the search process looks for the location that maximizes the chance of rejecting a
null hypothesis. In this case the significance of the trend would be maximized simply by leaving the shift term
out altogether, which corresponds to the test scores in the first block of Table 5.

The supVF scores for the test of whether the average shift term is zero are shown in the bottom two rows
of Table 5. Compared to the known-date case the VF scores are larger, as expected, and the increase exceeds
that in the critical values, yielding marginal significance in the LT layer and significance in the MT layer. Note
that even if we were to choose a specification with no level shifts in the observed temperature series, we
would still reject model-observational equivalence, as shown in the first block of Table 5. The supVF scores
provide additional rationale for the importance of controlling for a possible break at an unknown date.

7 CONCLUSIONS
Heteroskedasticity and autocorrelation robust (HAC) covariance matrix estimators have been adapted to
the linear trend model, permitting robust inferences about trend significance and trend comparisons in data
sets with complex and unknown autocorrelation characteristics. Here we extend the multivariate HAC
approach of Vogelsang and Franses (2005) to allow more general deterministic regressors in the model. We
show that the asymptotic (approximating) critical values of the test statistics of Vogelsang and Franses
(2005) are nonstandard and depend on the specific deterministic regressors included in the model. These
critical values can be simulated directly. Alternatively, a simple bootstrap method is available for obtaining
valid critical values and p-values.

The empirical focus of the paper is a comparison of trends in climate model-generated temperature data
and corresponding observed temperature data in the tropical troposphere. Our empirical innovation is to
make the trend model robust to the possibility of a level shift in the observed data corresponding to the
Pacific Climate Shift that occurred around 1978. With respect to the Vogelsang and Franses (2005) approach, this amounts to adding a level shift dummy to the model which requires a new set of critical values which we provide.

As our empirical findings show, the detection of a trend in the tropical lower- and mid-troposphere data over the 1958-2012 interval is contingent on the decision of whether or not to control for a level shift coinciding with the PCS. If the term is included, a time trend regression with autocorrelation-robust error terms indicates that the trend is small and not statistically different from zero in either the LT or MT layers. Also most climate models predict a significantly larger trend over this interval than is observed in either layer. We find a statistically significant discrepancy between the average climate model trend and observational trends whether or not the mean-shift term is included. However, with the shift term included the null hypothesis of trend equivalence is rejected much more strongly (at much smaller significance levels).

Regarding the question of preferred specification (that is, whether to include a shift or not), where the researcher suspects a break has occurred, results ought to be robust to controlling for the possibility. In the multivariate tests when we fix the break at 1977:12 the shift terms are not significant in either level, but when we use the grid search method the shift is significant at 10% in the LT layer and at 5% in the MT layer. Since breaks are harder to identify than trends, these findings indicate the importance of controlling for the possibility that one is present.

The testing method used herein is both powerful and relatively robust to over-rejections under the null hypothesis caused by strong serial correlation. The power of the test is indicated by the span of test scores in Table 6 in which relatively small changes in modeled trends translate into smaller p-values. Using the data-mining method provides a check on the extent to which the results depend on the assumption of a known shift date.

As such our empirical approach has many other potential applications on climatic and other data sets in which level shifts are believed to have occurred. Examples could include stratospheric temperature trends which are subject to level shifts coinciding with major volcanic eruptions, and land surface trends where it is believed that the measuring equipment changed or was moved. Generalizing the approach to allow more than one unknown break point is left for subsequent work.
8 REFERENCES


Davies, R.B. (1987) “Hypothesis testing when a nuisance parameter is not identified under the null hypothesis,” *Biometrika*, 74, 117-123.


## Tables

<table>
<thead>
<tr>
<th>%</th>
<th>$V_{F_t}$</th>
<th>VF</th>
</tr>
</thead>
<tbody>
<tr>
<td>.700</td>
<td>1.681</td>
<td>11.676</td>
</tr>
<tr>
<td>.750</td>
<td>2.185</td>
<td>14.685</td>
</tr>
<tr>
<td>.800</td>
<td>2.737</td>
<td>18.543</td>
</tr>
<tr>
<td>.850</td>
<td>3.417</td>
<td>23.798</td>
</tr>
<tr>
<td>.900</td>
<td>4.306</td>
<td>32.353</td>
</tr>
<tr>
<td>.950</td>
<td>5.688</td>
<td>49.399</td>
</tr>
<tr>
<td>.975</td>
<td>7.028</td>
<td>70.255</td>
</tr>
<tr>
<td>.990</td>
<td>8.779</td>
<td>99.978</td>
</tr>
<tr>
<td>.995</td>
<td>9.999</td>
<td>125.809</td>
</tr>
</tbody>
</table>

**Table 1a**: Asymptotic Critical Values. Model (2), known shift date with $\lambda = 0.3636, q = 1$. The value in the first column shows the percentage in the upper (right) tail exceeding the indicated values of $V_{F_t}$ and $V_F$. Left tail critical values of $V_{F_t}$ follow by symmetry around zero.

<table>
<thead>
<tr>
<th>%</th>
<th>supVF trend slope</th>
<th>supVF level shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>.700</td>
<td>79.765</td>
<td>95.455</td>
</tr>
<tr>
<td>.750</td>
<td>88.184</td>
<td>109.94</td>
</tr>
<tr>
<td>.800</td>
<td>98.532</td>
<td>116.20</td>
</tr>
<tr>
<td>.850</td>
<td>111.78</td>
<td>130.76</td>
</tr>
<tr>
<td>.900</td>
<td>131.92</td>
<td>150.99</td>
</tr>
<tr>
<td>.950</td>
<td>166.41</td>
<td>188.68</td>
</tr>
<tr>
<td>.975</td>
<td>205.15</td>
<td>225.78</td>
</tr>
<tr>
<td>.990</td>
<td>261.39</td>
<td>279.85</td>
</tr>
<tr>
<td>.995</td>
<td>301.94</td>
<td>322.48</td>
</tr>
</tbody>
</table>

**Table 1b**: Asymptotic Critical Values. Model (2), unknown shift date, $q = 1$. 10% Trimming ($\lambda^* = 0.1$). The value in the first column shows the percentage in the upper (right) tail exceeding the indicated values of the supVF statistics for, respectively, the trend slope and level shift coefficients.
Table 2a: Empirical Null Rejections with AR(2) Errors. $H_0: b_1 = b_2$. 5% Nominal Level, $b_1 = b_2 = .01, \eta = 0, g_1 = 0$. The data generating process is given by (17) and (18).
<table>
<thead>
<tr>
<th>$\eta$</th>
<th>$g_1$</th>
<th>$\rho_1$</th>
<th>$b_2$</th>
<th>Without Level Shift</th>
<th>With Level Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$W_{PW}$</td>
<td>$W_{Bart}$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.052</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.452</td>
<td>.451</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.954</td>
<td>.955</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>.9</td>
<td>.01</td>
<td></td>
<td></td>
<td>.068</td>
<td>.123</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.093</td>
<td>.155</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.166</td>
<td>.247</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.284</td>
<td>.386</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>.435</td>
<td>.551</td>
</tr>
<tr>
<td>0</td>
<td>.25</td>
<td>0</td>
<td>.01</td>
<td>.442</td>
<td>.441</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.051</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.450</td>
<td>.449</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.954</td>
<td>.954</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>.9</td>
<td>.01</td>
<td></td>
<td></td>
<td>.092</td>
<td>.153</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.068</td>
<td>.122</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.092</td>
<td>.154</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.165</td>
<td>.245</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>.280</td>
<td>.384</td>
</tr>
<tr>
<td>.5</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.051</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.691</td>
<td>.691</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.998</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>.9</td>
<td>.01</td>
<td></td>
<td></td>
<td>.068</td>
<td>.123</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.112</td>
<td>.182</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.242</td>
<td>.341</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.439</td>
<td>.556</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>.656</td>
<td>.758</td>
</tr>
<tr>
<td>.5</td>
<td>.25</td>
<td>0</td>
<td>.01</td>
<td>.683</td>
<td>.684</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.050</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.687</td>
<td>.689</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.998</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>.9</td>
<td>.01</td>
<td></td>
<td></td>
<td>.111</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>.0105</td>
<td></td>
<td></td>
<td>.067</td>
<td>.121</td>
</tr>
<tr>
<td></td>
<td>.011</td>
<td></td>
<td></td>
<td>.111</td>
<td>.179</td>
</tr>
<tr>
<td></td>
<td>.0116</td>
<td></td>
<td></td>
<td>.241</td>
<td>.338</td>
</tr>
<tr>
<td></td>
<td>.0121</td>
<td></td>
<td></td>
<td>.436</td>
<td>.552</td>
</tr>
</tbody>
</table>

Table 2b: Empirical Null Rejections and Empirical Power with AR(1) Errors. $H_0: b_1 = b_2$, $T=636$, 5% Nominal Level, $b_1 = .01$, $\rho_2 = 0$. The data generating process is given by (17) and (18).
<table>
<thead>
<tr>
<th>Data Series</th>
<th>Model/ Obs Name Extra Forcings; No. runs</th>
<th>Simple Trend</th>
<th>Trend + Level shift</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Trend (C/decade)</td>
<td>95% CI ± width</td>
</tr>
<tr>
<td>1</td>
<td>BCCR BCM2.0 O; 2</td>
<td>0.168</td>
<td>0.039</td>
</tr>
<tr>
<td>2</td>
<td>CCCMA3.1-T47 NA; 5</td>
<td>0.347</td>
<td>0.027</td>
</tr>
<tr>
<td>3</td>
<td>CCCMA3.1-T63 NA; 1</td>
<td>0.370</td>
<td>0.053</td>
</tr>
<tr>
<td>4</td>
<td>CNRM3.0 O; 1</td>
<td>0.234</td>
<td>0.048</td>
</tr>
<tr>
<td>5</td>
<td>CSIRO3.0 1</td>
<td>0.152</td>
<td>0.046</td>
</tr>
<tr>
<td>6</td>
<td>CSIRO3.5 1</td>
<td>0.251</td>
<td>0.062</td>
</tr>
<tr>
<td>7</td>
<td>GFDL2.0 O, LU, SO, V; 1</td>
<td>0.174</td>
<td>0.071</td>
</tr>
<tr>
<td>8</td>
<td>GFDL2.1 O, LU, SO, V; 1</td>
<td>0.102</td>
<td>0.104</td>
</tr>
<tr>
<td>9</td>
<td>GISS_AOM 2</td>
<td>0.178</td>
<td>0.048</td>
</tr>
<tr>
<td>10</td>
<td>GISS_EH O, LU, SO, V; 6</td>
<td>0.202</td>
<td>0.079</td>
</tr>
<tr>
<td>11</td>
<td>GISS_ER O, LU, SO, V; 5</td>
<td>0.180</td>
<td>0.083</td>
</tr>
<tr>
<td>12</td>
<td>IAP_FGOALS1.0 3</td>
<td>0.204</td>
<td>0.085</td>
</tr>
<tr>
<td>13</td>
<td>ECHAM4 1</td>
<td>0.218</td>
<td>0.091</td>
</tr>
<tr>
<td>14</td>
<td>INMCM3.0 SO, V; 1</td>
<td>0.184</td>
<td>0.051</td>
</tr>
<tr>
<td>15</td>
<td>IPSL_CM4 1</td>
<td>0.177</td>
<td>0.053</td>
</tr>
<tr>
<td>16</td>
<td>MIROC3.2_T106 O, LU, SO, V; 1</td>
<td>0.155</td>
<td>0.053</td>
</tr>
<tr>
<td>17</td>
<td>MIROC3.2_T42 O, LU, SO, V; 3</td>
<td>0.216</td>
<td>0.085</td>
</tr>
</tbody>
</table>
Table 3: Summary of Lower Troposphere data series. Notes: Each row refers to model ensemble mean (rows 1—23) or observational series (rows 24—26). All models forced with 20\textsuperscript{th} century greenhouse gases and direct sulfate effects. Rows 10, 11, 19, 22 and 23 also include indirect sulfate effects. ‘Extra forcing’ indicates which models included other forcings: ozone depletion (O), solar changes (SO), land use (LU), volcanic eruptions (V). NA: information not supplied to PCMDI. No. runs: indicates number of individual realizations in the ensemble mean. Trend slopes estimated using OLS, 95\% CI is trend ± number shown, which is computed using VF method (see Section 4). For instance, the RAOB CORE simple trend (bottom row first entry) is 0.147±0.052 °C/decade.
<table>
<thead>
<tr>
<th>Data Series</th>
<th>Model / Obs Name Extra Forcings; No. runs</th>
<th>Simple Trend</th>
<th>Trend (°C/decade)</th>
<th>95% CI ± width</th>
<th>Trend + Level shift</th>
<th>Trend (°C/decade)</th>
<th>95% CI ± width</th>
<th>Level Shift (°C/decade)</th>
<th>95% CI ± width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCCR BCM2.0 O; 2</td>
<td></td>
<td>0.173</td>
<td>0.034</td>
<td></td>
<td>0.176</td>
<td>0.065</td>
<td>-0.013</td>
<td>0.215</td>
</tr>
<tr>
<td>2</td>
<td>CCCMA3.1-T47 NA; 5</td>
<td></td>
<td>0.370</td>
<td>0.028</td>
<td></td>
<td>0.363</td>
<td>0.053</td>
<td>0.030</td>
<td>0.176</td>
</tr>
<tr>
<td>3</td>
<td>CCCMA3.1-T63 NA; 1</td>
<td></td>
<td>0.395</td>
<td>0.052</td>
<td></td>
<td>0.407</td>
<td>0.092</td>
<td>-0.048</td>
<td>0.303</td>
</tr>
<tr>
<td>4</td>
<td>CNRM3.0 O; 1</td>
<td></td>
<td>0.292</td>
<td>0.057</td>
<td></td>
<td>0.260</td>
<td>0.088</td>
<td>0.128</td>
<td>0.291</td>
</tr>
<tr>
<td>5</td>
<td>CSIRO3.0 1</td>
<td></td>
<td>0.121</td>
<td>0.046</td>
<td></td>
<td>0.162</td>
<td>0.104</td>
<td>-0.165</td>
<td>0.344</td>
</tr>
<tr>
<td>6</td>
<td>CSIRO3.5 1</td>
<td></td>
<td>0.240</td>
<td>0.067</td>
<td></td>
<td>0.302</td>
<td>0.092</td>
<td>-0.244</td>
<td>0.304</td>
</tr>
<tr>
<td>7</td>
<td>GFDL2.0 O, LU, SO, V; 1</td>
<td></td>
<td>0.162</td>
<td>0.069</td>
<td></td>
<td>0.114</td>
<td>0.139</td>
<td>0.193</td>
<td>0.457</td>
</tr>
<tr>
<td>8</td>
<td>GFDL2.1 O, LU, SO, V; 1</td>
<td></td>
<td>0.097</td>
<td>0.113</td>
<td></td>
<td>0.126</td>
<td>0.205</td>
<td>-0.116</td>
<td>0.675</td>
</tr>
<tr>
<td>9</td>
<td>GISS_AOM 2</td>
<td></td>
<td>0.170</td>
<td>0.047</td>
<td></td>
<td>0.170</td>
<td>0.091</td>
<td>-0.001</td>
<td>0.301</td>
</tr>
<tr>
<td>10</td>
<td>GISS_EH O, LU, SO, V; 6</td>
<td></td>
<td>0.189</td>
<td>0.078</td>
<td></td>
<td>0.226</td>
<td>0.117</td>
<td>-0.144</td>
<td>0.385</td>
</tr>
<tr>
<td>11</td>
<td>GISS_ER O, LU, SO, V; 5</td>
<td></td>
<td>0.165</td>
<td>0.079</td>
<td></td>
<td>0.186</td>
<td>0.135</td>
<td>-0.084</td>
<td>0.447</td>
</tr>
<tr>
<td>12</td>
<td>IAP_FGOALS1.0 3</td>
<td></td>
<td>0.191</td>
<td>0.081</td>
<td></td>
<td>0.232</td>
<td>0.119</td>
<td>-0.164</td>
<td>0.392</td>
</tr>
<tr>
<td>13</td>
<td>ECHAM4 1</td>
<td></td>
<td>0.208</td>
<td>0.085</td>
<td></td>
<td>0.232</td>
<td>0.146</td>
<td>-0.096</td>
<td>0.480</td>
</tr>
<tr>
<td>14</td>
<td>INMCM3.0 SO, V; 1</td>
<td></td>
<td>0.189</td>
<td>0.054</td>
<td></td>
<td>0.186</td>
<td>0.105</td>
<td>0.008</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Variable(s)</td>
<td>N</td>
<td>Trend Slope ± 95% CI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------</td>
<td>-------------</td>
<td>---</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>IPSL_CM4;</td>
<td></td>
<td></td>
<td>0.181 ± 0.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>MIROC3.2_T106</td>
<td>O, LU, SO, V; 1</td>
<td>0.162 ± 0.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>MIROC3.2_T42</td>
<td>O, LU, SO, V; 3</td>
<td>0.218 ± 0.093</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>MPI2.3.2a</td>
<td>SO, V; 5</td>
<td>0.191 ± 0.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ECHAM5</td>
<td>O; 4</td>
<td>0.204 ± 0.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>CCSM3.0</td>
<td>O, SO, V; 7</td>
<td>0.209 ± 0.088</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>PCM_B06.57</td>
<td>O, SO, V; 4</td>
<td>0.164 ± 0.029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>HADCM3</td>
<td>O; 1</td>
<td>0.165 ± 0.036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>HADGEM1</td>
<td>O, LU, SO, V; 1</td>
<td>0.221 ± 0.064</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>HadAT</td>
<td></td>
<td></td>
<td>0.086 ± 0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>RICH</td>
<td></td>
<td></td>
<td>0.090 ± 0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>RAOBCORE</td>
<td></td>
<td></td>
<td>0.117 ± 0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Summary of Mid-Troposphere data series. Notes same as for Table 3. Trend slopes estimated using OLS, 95% CI is trend ± number shown, which is computed using VF method (see Section 4). For instance, the RAOBCORE simple trend (bottom row first entry) is 0.132 ± 0.053 °C/decade.
<table>
<thead>
<tr>
<th>Test</th>
<th>Null Hypothesis</th>
<th>Test Score</th>
<th>Bootstrap p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend in: No Level shift</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hadley LT (0.135)</td>
<td>trend = 0</td>
<td>260.4***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>RICH LT (0.126)</td>
<td>trend = 0</td>
<td>212.0***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>RAOBCORE LT (0.149)</td>
<td>trend = 0</td>
<td>281.4***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Hadley MT (0.086)</td>
<td>trend = 0</td>
<td>117.0***</td>
<td>&lt; 0.004</td>
</tr>
<tr>
<td>RICH MT (0.090)</td>
<td>trend = 0</td>
<td>111.4***</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>RAOBCORE MT (0.117)</td>
<td>trend = 0</td>
<td>197.4***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>LT average</td>
<td>Models = Observed</td>
<td>97.2**</td>
<td>&lt; 0.007</td>
</tr>
<tr>
<td>MT average</td>
<td>Models = Observed</td>
<td>167.0***</td>
<td>&lt; 0.002</td>
</tr>
<tr>
<td>Trend in: With Level shift at Date (Assumed Known): December 1977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hadley LT (0.064)</td>
<td>trend = 0</td>
<td>13.2</td>
<td>0.273</td>
</tr>
<tr>
<td>RICH LT (0.093)</td>
<td>trend = 0</td>
<td>22.6</td>
<td>0.160</td>
</tr>
<tr>
<td>RAOBCORE LT (0.065)</td>
<td>trend = 0</td>
<td>11.2</td>
<td>0.311</td>
</tr>
<tr>
<td>Hadley MT (-0.001)</td>
<td>trend = 0</td>
<td>0.1</td>
<td>0.925</td>
</tr>
<tr>
<td>RICH MT (0.048)</td>
<td>trend = 0</td>
<td>5.1</td>
<td>0.485</td>
</tr>
<tr>
<td>RAOBCORE MT (0.042)</td>
<td>trend = 0</td>
<td>3.4</td>
<td>0.563</td>
</tr>
<tr>
<td>LT average</td>
<td>Models = Observed</td>
<td>354.5***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>MT average</td>
<td>Models = Observed</td>
<td>685.9***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Shift Term in:</td>
<td>Avg Obs shift term = 0</td>
<td>19.4</td>
<td>0.189</td>
</tr>
<tr>
<td>MT average</td>
<td>Avg Obs shift term = 0</td>
<td>30.8</td>
<td>0.107</td>
</tr>
<tr>
<td>Trend in: With Level Shift at Date Assumed Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT average</td>
<td>Models = Observed</td>
<td>495.6***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>MT average</td>
<td>Models = Observed</td>
<td>937.7***</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Shift Term in:</td>
<td>Avg Obs shift term = 0</td>
<td>180.4*</td>
<td>0.061</td>
</tr>
<tr>
<td>MT average</td>
<td>Avg Obs shift term = 0</td>
<td>259.8**</td>
<td>&lt; 0.025</td>
</tr>
</tbody>
</table>

Table 5: Results of hypothesis tests using VF statistic. Notes: sample period (monthly): January 1958 to December 2012. The bootstrap p-value* is computed using the method described in Section 3.3 using 10000 bootstrap replications. VF Critical Values: Without level shift, 27.14 (10%, denoted *) 41.53 (5%, denoted **), 83.96 (1%, denoted ***). With level shift at known date, 32.35 (10%, denoted *), 49.40 (5%, denoted **), 99.98 (1%, denoted ***). With level shift at unknown date, 150.99 (10%, denoted *), 188.68 (5%, denoted **), 279.85 (1%, denoted ***).

1Interpolated using critical values in Table 1b.
<table>
<thead>
<tr>
<th>Model</th>
<th>LT Layer</th>
<th>MT Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known</td>
<td>Unknown</td>
</tr>
<tr>
<td>No Shift</td>
<td>Shift Date</td>
<td>Shift Date</td>
</tr>
<tr>
<td>1</td>
<td>22.53</td>
<td>46.39*</td>
</tr>
<tr>
<td>2</td>
<td>959.34***</td>
<td>280.48***</td>
</tr>
<tr>
<td>3</td>
<td>751.09***</td>
<td>932.99***</td>
</tr>
<tr>
<td>4</td>
<td>78.72**</td>
<td>33.14</td>
</tr>
<tr>
<td>5</td>
<td>1.52</td>
<td>17.02</td>
</tr>
<tr>
<td>6</td>
<td>93.92***</td>
<td>154.70**</td>
</tr>
<tr>
<td>7</td>
<td>22.90</td>
<td>11.55</td>
</tr>
<tr>
<td>8</td>
<td>6.87</td>
<td>5.93</td>
</tr>
<tr>
<td>9</td>
<td>11.25</td>
<td>13.73</td>
</tr>
<tr>
<td>10</td>
<td>31.94*</td>
<td>397.95***</td>
</tr>
<tr>
<td>11</td>
<td>14.17</td>
<td>194.10***</td>
</tr>
<tr>
<td>12</td>
<td>29.78*</td>
<td>413.78***</td>
</tr>
<tr>
<td>13</td>
<td>46.21**</td>
<td>521.81***</td>
</tr>
<tr>
<td>14</td>
<td>12.52</td>
<td>15.12</td>
</tr>
<tr>
<td>15</td>
<td>16.13</td>
<td>10.10</td>
</tr>
<tr>
<td>16</td>
<td>3.30</td>
<td>10.56</td>
</tr>
<tr>
<td>17</td>
<td>40.62*</td>
<td>415.17***</td>
</tr>
<tr>
<td>18</td>
<td>32.22*</td>
<td>232.33***</td>
</tr>
<tr>
<td>19</td>
<td>83.54**</td>
<td>55.32**</td>
</tr>
<tr>
<td>20</td>
<td>32.85*</td>
<td>330.88***</td>
</tr>
<tr>
<td>21</td>
<td>44.83**</td>
<td>74.18**</td>
</tr>
<tr>
<td>22</td>
<td>38.97*</td>
<td>36.35*</td>
</tr>
<tr>
<td>23</td>
<td>156.67***</td>
<td>126.96***</td>
</tr>
</tbody>
</table>

Last row: fraction of models exhibiting difference from observations significant at <=5%.

Table 6. VF tests of equivalent trends between individual model (ensemble mean in cases of multiple runs) and average of balloon series. Column 1: model. Column 2: LT layer, No Shift case. Columns 3 and 4 (Shift1 and Shift2, respectively): with shift assumed known at 1977:12 and with shift at date assumed unknown. Columns 5-7: same for MT layer. VF Critical Values: Without level shift, 27.14 (10%, denoted *), 41.53 (5%, denoted **), 83.96 (1%, denoted ***). With level shift known to be at 1977:12, 33.12 (10%, denoted *), 51.20 (5%, denoted **), 98.46 (1%, denoted ***). With level shift at unknown date, 131.92 (10%, denoted *), 166.41 (5%, denoted **), 261.39 (1%, denoted ***). Last row: fraction of models exhibiting difference from observations significant at <=5%.
Figure 1. Schematics of two series to be compared.
135x132mm (96 x 96 DPI)
HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS

SUPPLEMENTARY INFORMATION

Ross McKitrick*
Department of Economics
University of Guelph
Guelph ON Canada N1G 2W1
rmckitri@uoguelph.ca
519-824-4120 x52532

Timothy J. Vogelsang
Department of Economics
Michigan State University
tjv@msu.edu

Contents
1 Motivation and Background of VF Approach 2
2 Derivation of Null Limit of VF 5
3 Bootstrap Critical Values and p-Values 7
4 Main Computational Code For Empirical Results 9
5 Code For Simulation of Asymptotic Critical Values 24
6 Code For Finite Sample Simulations 27
1 Motivation and Background of VF Approach

Motivation for the form of the VF statistic can be developed by considering the very simple model

\[ y_{it} = a_i + u_{it}, \quad (A1) \]

Model (A1) can written in terms of the general model (3) setting \( d_{it} = 1, \beta_i = a_i \) and \( d_{it} = 0 \). Because \( u_{it} \) is a mean zero time series, it follows that \( a_i = E(y_{it}) \). For the purpose of matrix representations the natural organization is to denote rows by the time index \( t \) and columns by the data source index \( i \). However the matrix representation of the statistical theory becomes easier if we transpose the data so that the columns represent time. We can then refer to time series of column vectors:

\[ y_t = (y_{1t}, \ldots, y_{nt})', \quad a = (a_1, a_2, \ldots, a_n)' \]

and \( u_t = (u_{1t}, \ldots, u_{nt})' \).

Rewrite the model as

\[ y_t = a + u_t \]

and suppose we are interested in testing \( q \) linear restrictions about the means \( a \), of the form:

\[ H_0: Ra = r, \quad H_1: Ra \neq r \]

where \( R \) and \( r \) are, respectively, \( q \times n \) and \( q \times 1 \) matrices of known constants. We require that \( q \leq n \) and that \( R \) have full rank (\( rank(R) = q \)). The natural estimator of \( a \) is the vector of sample averages, i.e. the OLS estimator given by \( \hat{a} = \bar{y} = T^{-1} \Sigma_{t=1}^{T} y_t \).

The variance of \( \hat{a} \), given the covariance stationarity assumption for \( u_t \), is given by

\[ Var(\hat{a}) = T^{-2}E[(\Sigma_{t=1}^{T} u_t)(\Sigma_{t=1}^{T} u_t)'] = T^{-1}[\Sigma_0 + \Sigma_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) (\Gamma_j + \Gamma_j')] \]

Letting \( \Omega_T = \Sigma_0 + \Sigma_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) (\Gamma_j + \Gamma_j') \) we have the more compact expression

\[ Var(\hat{a}) = T^{-1} \Omega_T. \quad (A2) \]

An \( F \)-statistic constructed using (A2) is infeasible because \( \Omega_T \) is unknown. A natural estimator of \( \Omega_T \) is given by
\[ \hat{\Omega}_T = \hat{f}_0 + \Sigma_{j=1}^{T-1} \left( 1 - \frac{j}{T} \right) \left( \hat{f}_j + \hat{f}_j' \right), \hat{f}_j' = T^{-1} \Sigma_{t=j+1}^T \hat{u}_t \hat{u}'_{t-j}, \]

where \( \hat{u}_t = y_t - \hat{\mu} \). Using \( \hat{\Omega}_T \) in place of \( \Omega_T \) leads to the \( VF \) statistic:

\[ VF = \frac{(\hat{R} \hat{\mu} - r) \left( T^{-1} \hat{R} \hat{\Omega}_T R \right)^{-1} (\hat{R} \hat{\mu} - r)}{q}. \]

Because \( \hat{\Omega}_T \) is constructed without assuming a specific model of serial correlation, it is in the class of nonparametric spectral estimators of \( \Omega \).

Asymptotic theory is used to generate an approximation for \( \hat{\Omega}_T \) and the null distribution of \( VF \) using the FCLT given by (12). If it were the case that \( \hat{\Omega}_T \) were a consistent estimator of \( \Omega \), then \( VF \) would converge in distribution to a \( \chi^2/q \) random variable. It turns out that \( \hat{\Omega}_T \) is not a consistent estimator of \( \Omega \); however, it is relatively easy to show that \( \hat{\Omega}_T \) does converge in distribution to a random matrix that is proportional to \( \Omega \) but otherwise does not depend on unknown quantities. This property of \( \hat{\Omega}_T \) means that the \( VF \) statistic can be used to test \( H_0 \) because \( VF \) can be approximated by a random variable that does not depend on unknown parameters.

Recall the partial sum of \( u_t \) given by \( S_t = \Sigma_{j=1}^T u_j \). Evaluating \( S_t \) at \( t = [cT] \) gives \( S_{[cT]} = \Sigma_{t=1}^{[cT]} u_t \) which is the sum of the first \( c^{th} \) proportion of the data. For a given value of \( c \), the quantity \( [cT] \to \infty \) as \( T \to \infty \).

Therefore, if we scale by \( T^{-1/2} \) we obtain the result

\[ T^{-1/2} S_{[cT]} = \left( \frac{[cT]}{T} \right)^{1/2} \left[ cT \right]^{1/2} S_{[cT]} \xrightarrow{d} c^{1/2} N(0, \Omega) = N(0, c\Omega). \]

For a given value of \( c \), the scaled partial sums of \( u_t \) satisfy a Central Limit Theorem (CLT). These limits hold pointwise in \( c \). The FCLT given by (11) is a stronger statement that says this collection of CLTs, as indexed by \( c \), hold jointly and uniformly in \( c \) and that the family of limiting normal random variables are a Wiener process (or standard Brownian motion). Not surprisingly, the FCLT requires slightly stronger assumptions for \( u_t \) than a CLT. For example, the condition \( \Sigma_{j=0}^\infty |j_{ij}| < \infty \), where \( j_{ij} \) is the \( i,j \) element of the matrix \( \Gamma_j \), is strengthened to \( \Sigma_{j=0}^\infty \|j_{ij}\| < \infty \) which requires the autocovariances to shrink faster to zero as \( j \) increases.

Using the FCLT given by (11) it immediately follows that

\[ \text{John Wiley & Sons} \]
\[ \sqrt{T}(\bar{\alpha} - \alpha) = \sqrt{T}\bar{u} = T^{-1/2}\Sigma_{t=1}^{T} u_t = T^{-1/2}S_T \Rightarrow \Lambda W_n(1) \sim N(0, \Lambda \Lambda') = N(0, \Omega). \]

Using the FCLT, it is straightforward to determine the asymptotic behavior of \( \tilde{\Omega}_T \). The first step is to write \( \tilde{\Omega}_T \) as a function of \( \tilde{\delta}_T = \Sigma_{j=1}^{T} \tilde{u}_j \) using (10):

\[ \tilde{\Omega}_T = \tilde{\delta}_0 + \Sigma_{j=1}^{T-1} (1 - \frac{1}{T})(\tilde{\delta}_j + \tilde{\delta}_j') = 2T^{-2} \Sigma_{j=1}^{T-1} \tilde{\delta}_j \tilde{\delta}_j'. \]

Using the FCLT, the limit of \( T^{-\frac{1}{2}} \tilde{\delta}_{cT} \) is easy to derive:

\[ T^{-\frac{1}{2}} \tilde{\delta}_{cT} = T^{-\frac{1}{2}} \Sigma_{i=1}^{cT} \tilde{u}_i = T^{-\frac{1}{2}} \Sigma_{i=1}^{cT} (y_i - \hat{\alpha}) = T^{-\frac{1}{2}} \Sigma_{i=1}^{cT} (a + u_i - \hat{\alpha}) \]

\[ = T^{-\frac{1}{2}} \Sigma_{i=1}^{cT} u_i - T^{-\frac{1}{2}} [cT](\hat{\alpha} - \alpha) = T^{-\frac{1}{2}} S_{cT} - (\frac{1}{T})\sqrt{T}(\hat{\alpha} - \alpha) \]

\[ \Rightarrow \Lambda W_n(c) - c\Lambda W_n(1) = \Lambda(W_n(c) - cW_n(1)) \equiv \Lambda B_n(c). \]

The stochastic process, \( B_n(c) \), is the well known Brownian bridge. Using this result for \( T^{-\frac{1}{2}} \tilde{\delta}_{cT} \) and the continuous mapping theorem, it follows that

\[ \tilde{\Omega}_T = 2T^{-1} \Sigma_{i=1}^{T-1} (T^{-\frac{1}{2}} \tilde{\delta}_i) (T^{-\frac{1}{2}} \tilde{\delta}_i') \Rightarrow 2 \Lambda \int_0^1 B_n(c)(c)\prime dc\Lambda'. \]

We see that while \( \tilde{\Omega}_T \) does not converge to \( \Omega = \Lambda \Lambda' \), it does converge to a random matrix that is proportional to \( \Lambda \Lambda' \).

Establishing the limit of \( VF \) is now simple:

\[ VF = (R\tilde{u} - r')[T^{-1}R\tilde{\Omega}_T R']^{-1}(R\tilde{u} - r)/q = \sqrt{T}(R\tilde{u} - r)'[R\tilde{\Omega}_T R']^{-1}\sqrt{T}(R\tilde{u} - r)/q \]

\[ = (R\sqrt{T}\tilde{u})'[R\tilde{\Omega}_T R']^{-1}R\sqrt{T}\tilde{u}/q \Rightarrow (R\Lambda W_n(1))' [R2\Lambda \int_0^1 B_n(c)(c)' dc\Lambda']^{-1} R\Lambda W_n(1)/q. \]

While not obvious at first glance, the restriction matrix, \( R \), drops from the limit. Because Wiener processes are Gaussian (normally distributed), linear combinations of Wiener processes are also Wiener processes.

Therefore, \( R\Lambda W_n(c) \) is a \( q \times 1 \) vector of Wiener processes and we can rewrite \( R\Lambda W_n(c) \) as \( \Lambda' W_q(c) \) where \( \Lambda' \) is the \( q \times q \) matrix square root of \( R\Lambda\Lambda' R' \), i.e. \( \Lambda'\Lambda' = R\Lambda\Lambda R' \). Similarly, we can rewrite \( R\Lambda B_n(c) \) as \( \Lambda' B_q(c) \) where \( B_q(c) = W_q(c) - cW_q(1) \). Because \( R \) is assumed to be full rank, it follows that \( \Lambda' \) is full rank and is therefore invertible. We have
\[ VF \Rightarrow (R \Lambda W_n(1))^\top \left[ R 2 \Lambda \int_0^1 B_n(c) B_n(c)' dc \Lambda' \right]^{-1} R \Lambda W_n(1)/q \]

\[ = (\Lambda^* W_q(1))^\top \left[ 2 \Lambda^* \int_0^1 B_q(c) B_q(c)' dc \Lambda' \right]^{-1} \Lambda^* W_q(1)/q \]

\[ = W_q(1)^\top \left[ 2 \int_0^1 B_q(c) B_q(c)' dc \right]^{-1} W_q(1)/q, \]

and the \( \Lambda^* \) matrices drop out because \( \Lambda^* \) is invertible.

The limit of \( VF \) does not depend on unknown parameters. The limit is a quadratic form involving a vector of independent standard normal random variables, \( W_q(1) \) and the inverse of the random matrix

\[ 2 \int_0^1 B_q(c) B_q(c)' dc. \]

Because \( W_q(1) \) is independent of \( B_q(c) \) for all \( c \), \( W_q(1) \) is independent of

\[ 2 \int_0^1 B_q(c) B_q(c)' dc \]

and the limit of \( VF \) is similar in spirit to an \( F \) random variable but its distribution is nonstandard. The random matrix \( 2 \int_0^1 B_q(c) B_q(c)' dc \) can be viewed as an approximation to the randomness of \( R \tilde{\Omega} R' \) whereas \( W_q(1) \) approximates the randomness of \( \sqrt{T}(R \tilde{\alpha} - r) \).

\section{Derivation of Null Limit of VF}

The first step of the derivation is to obtain the asymptotic behavior of \( R \hat{\beta} - r \) under \( H_0 \). Under \( H_0 \) we have \( R \beta = r \) and it follows that \( R \hat{\beta} - r = R \hat{\beta} - R \beta = R(\hat{\beta} - \beta) \). Using (6) we have

\[ T^{-1} \tau_0^R R(\hat{\beta} - \beta) = \left( T^{-1} \sum_{t=1}^T \tau_{0t}^2 \bar{d}_{0t}^2 \right)^{-1} R T^{-1/2} \sum_{t=1}^T \tau_{0t} \bar{d}_{0t} u_t \]

Using simple algebra, (11) and (12), we have

\[ T^{-1} \sum_{t=1}^T \tau_{0t}^2 \bar{d}_{0t}^2 \to \int_0^1 \tilde{f}_0^2(c) dc, \]  \( (A3) \)

\[ RT^{-1/2} \sum_{t=1}^T \tau_{0t} \bar{d}_{0t} u_t \Rightarrow R \int_0^1 \tilde{f}_0(c) \Lambda dW_n(c), \]  \( (A4) \)

where

\[ \tilde{f}_0(c) = f_0(c) - \left( \int_0^1 f_0(s) f_1(s)' ds \right) \left( \int_0^1 f_1(s) f_1(s)' ds \right)^{-1} f_1(c). \]
The limit in (A4) can be more compactly written as follows:

\[ R \int_0^1 \tilde{f}_0(c) dW_n(c) = \int_0^1 \tilde{f}_0(c) R dW_n(c) = \Lambda^* \int_0^1 \tilde{f}_0(c) dW_q(c), \]

(A5)

where \( \Lambda^* \) is the \( q \times q \) matrix square root of \( R \Lambda \Lambda^T \), i.e. \( \Lambda^* \Lambda = R \Lambda \Lambda^T \). The equivalence in distribution indicated by (A5) follows because Wiener processes are mean zero and Gaussian. Using (A3), (A4) and (A5) it follows that

\[
\sqrt{T} T_{0t}^{-1} (R \tilde{\beta} - \tau) \xrightarrow{d} \Lambda^* \left( \int_0^T \tilde{f}_0(c)^2 dc \right)^{-1} \left( \int_0^T \tilde{f}_0(c) dW_q(c) \right). \]

(A6)

We next derive the limit of \( \hat{\Omega}_T \). In deriving this limit it is convenient to stack the deterministic regressors into a single column vector \( d_t \) where \( d_t' = (d_{0t}, d_{1t}) \). Define the combined scaling matrix

\[
\tau_{T(k+1) \times (k+1)} = \begin{bmatrix} \tau_{0t} & 0_{1 \times k} \\ 0_{k \times 1} & \tau_{1T} \end{bmatrix}.
\]

It immediately follows from (11) and (12) that

\[
T^{-1} \sum_{t=1}^{[cT]} \tau_t d_t \rightarrow \int_0^c f(s) ds, \quad T^{-1} \sum_{t=1}^{[cT]} \tau_t d_t d_t' \rightarrow \int_0^c f(s)f(s)' ds,
\]

(A7)

\[
RT^{-1/2} \sum_{t=1}^{[cT]} u_t d_t' \tau_t \rightarrow \int_0^c \Lambda' dW_q(s)f(s)f(s)' ds.
\]

(A8)

The next step is to derive the limit of \( RT^{-1/2} \tilde{S}_T(S) \):

\[
RT^{-1/2} \tilde{S}_T(S) = RT^{-1/2} \sum_{t=1}^{[cT]} \tilde{u}_t = RT^{-1/2} \sum_{t=1}^{[cT]} \left( u_t - \sum_{j=1}^{T} \sum_{T} d_j d_j' \right)^{-1} d_t,
\]

\[
= RT^{-1/2} S_T(S) - RT^{-1/2} \sum_{j=1}^{T} u_j d_j' \tau_T \left( T^{-1} \sum_{j=1}^{T} \sum_{T} d_j d_j' \right)^{-1} \sum_{t=1}^{[cT]} \tau_t d_t
\]

\[
\Rightarrow \int_0^c \Lambda' dW_q(s) - \left( \int_0^1 \Lambda' dW_q(s)f(s)f(s)' ds \right)^{-1} \int_0^c f(s) ds \equiv \Lambda' B_q'(c).
\]

(A9)

The limit in (A9) follows from (11), (A7) and (A8). Using (A9) it directly follows that
\[ R \tilde{\omega} R^* = R 2T^{-2} \sum_{t=1}^{T-1} \tilde{S}_t \tilde{S}_t^\top R^* = 2T^{-1} \sum_{t=1}^{T-1} R T^{-1/2} \tilde{S}_t \tilde{S}_t^\top T^{-1/2} R^* \xrightarrow{d} 2 \Lambda^* \int_0^1 B_q^f(c) B_q^f(c)' dc \Lambda^* \]  

(A10)

Combining (A6) and (A10) gives the limit of \( V F \):

\[
VF = \sqrt{T} \tau_{\theta T}^{-1} (R \hat{\beta} - r)' \left[ (T^{-1} \tau_{\theta T}^2 \sum_{t=1}^{T} d_{0t}^2)^{-1} R \tilde{\omega} R^* \right]^{-1} \frac{1}{\sqrt{T}} \tau_{\theta T}^{-1} (R \hat{\beta} - r)/q \\
\xrightarrow{d} \left( \Lambda^* \left( \int_0^1 \tilde{f}_0(c)^2 dc \right)^{-1} \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right) \right)' \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right)/q \\
\times \left( \Lambda^* \left( \int_0^1 \tilde{f}_0(c)^2 dc \right)^{-1} \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right) \right)/q
\]

\[= \left( \int_0^1 \tilde{f}_0(c)^2 dc \right)^{-1/2} \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right) \left( 2 \int_0^1 B_q^f(c) B_q^f(c)' dc \right)^{-1/2} \left( \int_0^1 \tilde{f}_0(c)^2 dc \right)^{-1/2} \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right)/q.
\]

Using well known properties of Wiener processes, it follows that

\[
\left( \int_0^1 \tilde{f}_0(c)^2 dc \right)^{-1/2} \left( \int_0^1 \tilde{f}_0(c) dW_q(c) \right) = Z_q \sim N(0, 1),
\]

which allows us to write

\[VF = Z_q' \left[ 2 \int_0^1 B_q^f(c) B_q^f(c)' dc \right]^{-1} Z_q = V F_q^{\infty},
\]

completing the derivation. Using properties of Wiener processes, it is simple to show that \( B_q^f(c) \) is Gaussian and that \( Z_q \) and \( B_q^f(s) \) are independent. Therefore, it follows that \( Z_q \) and \( 2 \int_0^1 B_q^f(c) B_q^f(c)' dc \) are independent.

### 3 Bootstrap Critical Values and P-Values

If carrying out simulations of the asymptotic distributions is not easily accomplished using standard statistical packages, an alternative is to use a simple bootstrap approach as follows:
1. For each $i$ take the OLS residuals $\hat{u}_{it}$ from Equation (4) and sample with replacement from $\hat{u}_{i1}, \hat{u}_{i2}, \ldots, \hat{u}_{iT}$ to generate a bootstrap series $\hat{u}^*_i, \hat{u}^*_{i2}, \ldots, \hat{u}^*_{iT}$. Let $y^*_i = u^*_i$ denote the bootstrap resampled series to be used as dependent variables.

2. For each $i$ estimate model (3) using $\hat{y}^*_i$ in place of $\hat{y}_i$. Let $\hat{\beta}^*_i$ and $\hat{\delta}^*_i$ denote the OLS estimators and let $\hat{\epsilon}^*_i = y^*_i - \hat{\beta}^*_i d_{it} + \hat{\delta}^*_i d_{it}$ denote the OLS residuals. Let $\hat{\epsilon}^*_i$ denote the $n \times 1$ vector $\hat{\epsilon}^*_i = (\hat{\epsilon}^*_i, \hat{\epsilon}^*_{i2}, \ldots, \hat{\epsilon}^*_{iT})'$ and let $\hat{\beta}^*$ denote the $n \times 1$ vector $\hat{\beta}^* = (\hat{\beta}^*_1, \ldots, \hat{\beta}^*_n)'$.

3. Compute $\hat{\Omega}$ using (8) with $\hat{\epsilon}^*_i$ in place of $\hat{\epsilon}_i$. The equivalent form given by (10) can also be used and is faster to compute.

4. Compute $VF$ using

$$VF^* = (R\hat{\beta}^*)' \left[ \left( \sum_{t=1}^T d_{it}^2 \right)^{-1} R\hat{\Omega}_T R' \right]^{-1} (R\hat{\beta}^*) / q.$$ 

5. Repeat Steps 1 through 4 $N_B$ times where $N_B$ is a relatively large integer. This generates $N_B$ random draws from $VF^*$.

6. Sort the $N_B$ values of $VF^*$ from smallest to largest and let $VF^*[1], VF^*[2], \ldots, VF^*[N_B]$ indicate the sorted values. The right tail critical value for a test with significance level $\alpha$ is given by $VF^*[(1 - \alpha)N_B]$ where the integer part of $(1 - \alpha)N_B$ is used if $(1 - \alpha)N_B$ is not an integer.

7. Bootstrap $p$ values can be computed using the frequency of $VF^*$ values that exceed the value of $VF$ from the actual data.

Note that, by construction, the true value of $\beta^*$ is zero. Therefore $R\beta^* = 0$, i.e. $r=0$ in the bootstrap samples and $VF^*$ should be computed using $r=0$ to ensure that the null holds for $VF^*$. 

---

8

John Wiley & Sons
4 MAIN COMPUTATIONAL CODE FOR EMPIRICAL RESULTS

# R CODE FOR VF TESTS OF TREND EQUIVALENCE CONTROLLING FOR POSSIBLY UNKNOWN BREAK DATE.
# TAKES AS INPUT: FILES lt.txt and mt.txt
# (THESE ARE CREATED USING THE mv.datamaker PROGRAM)
# FULL CODE AND DATA ARCHIVE AVAILABLE AT JOURNAL WEBSITE

rm(list=ls(all=TRUE))  # remove all objects in memory
options(scipen=999)  # forbid printing in scientific notation
#sink("logfiles/mvlog.txt") # print output to a log file
set.seed(5)   # set random number seed
require(plotrix)  # START: SET LOCALS

# set locals
startyear = 1958  # start year
endyear = 2012  # end year
models = 23  # number of models
bal = 3  # number of obs series
tby = 1977  # year of break
tbm = 12  # month of break

bsr = 10000  # of bootstrap reps
pct20 = 0.80*bsr+1  # 80th percentile position
pct10 = 0.90*bsr+1  # 90th percentile position
pct05 = 0.95*bsr+1  # 95th percentile position
pct025 = 0.975*bsr+1  # 97.5th percentile position
pct01 = 0.99*bsr+1  # 99th percentile position
trim = 0.1  # fraction of end to leave off grid search

CVB = 7.028  # .975 VF critical value with 1 break at Dec 1977
ens = c(1,5,1,1,1,1,2,6,5,3,1,1,3,5,4,7,4,1,1)  # Numbers of ensemble members

K = models+bal   # of series
N = 12*(endyear-startyear+1) # sample length in months

omega1 = function(x, K, N, T) {
  d = c(1, rep(0, times=N-1))  # create T-length vector of 1 followed by N-1 zeroes
  x = t(x)  # transpose residual matrix
  S = x[1:K,1]  # S = first col of u
  OMEGA1 = S %*% t(S)  # First accum entry for Omega matrix
  for (i in 2:N)  {  
    d[i]=1  # Vector to add 1st i cols of u (sets next element to 1)
    S = x %*% d  # forms Nx1 vector with sum of 1st i cols of u
    OMEGA1 = OMEGA1 + S %*% t(S)  # accumulates partial sums
  }
  OMEGA1 = 2*OMEGA1 / (N^2)  # Last step
  return(OMEGA1)
}

omega2 = function(x, K, N, T, D) {
  d = c(1, rep(0, times=N-1))  # reset d to T-length vector of 1 followed by N-1 zeroes
  x = t(x)  # transpose residual matrix
  S = x[1:K,1]  # S = first col of u
  OMEGA2 = S %*% t(S)  # First accum entry for Omega matrix
  for (i in 2:N)  {  
    d[i]=1  # Vector to add 1st i cols of u (sets next element to 1)
    S = x %*% d  # forms Nx1 vector with sum of 1st i cols of u
    OMEGA2 = OMEGA2 + S %*% t(S)  # accumulates partial sums
  }
  return(OMEGA2)
}
\begin{verbatim}
  \text{OMEGA2 = 2*OMEGA2 / (N^2)}
  \text{# Last step}
  \text{return(OMEGA2)}
\end{verbatim}

\textbf{# PART I: LT Run}

\text{lt = read.table("data/lt.txt", header=TRUE)}
\text{lt = 120*(as.matrix(lt))}
\text{time = read.table("data/time.txt", header=TRUE)}
\text{attach(time)}
\text{T = time[,1]} \quad \text{# T - simple time trend}
\text{year = time[,2]}
\text{month = time[,3]}
\text{tt = year+(month-1)/12} \quad \text{# monthly time index}
\text{D = as.numeric(tt > (tby+(tbm-1)/12))} \quad \text{# 0-1 dummy for break}
\text{CONS = rep(1, length(T))} \quad \text{# constant term}
\text{beta1 = lm(T ~ 1)}
\text{ttilde1 = as.matrix(beta1$coefficients)} \quad \text{# partial out the t-terms}
\text{eta1 = sum( ttilde1^2 )} \quad \text{# partialling out term (no break)}
\text{beta2 = lm(T ~ 1 + D)}
\text{ttilde2 = as.matrix(beta2$coefficients)} \quad \text{# partialling out term for slope (break)}
\text{eta2 = sum( ttilde2^2 )}
\text{beta3 = lm(D ~ 1 + T)}
\text{ttilde3 = as.matrix(beta3$coefficients)} \quad \text{# partialling out term for shift (break)}
\text{eta3 = sum( ttilde3^2 )}
\text{rm(beta1, beta2, beta3)}

\textbf{# NO BREAK CASE}
\text{reg_t = lm(lt ~ T)} \quad \text{# regress each column of lt on time trend T}
\text{b_t = as.matrix(reg_t$coefficients[2,1:26])} \quad \text{# get slope coeffs}
\text{r_t = as.matrix(reg_t$residuals[,1:26])} \quad \text{# Get residuals}
\text{OMEGA1=omega1(r_t, K, N, T)}
\end{verbatim}
\# BREAK CASE (at D)
reg_td = lm(lt ~ T + D)  # regress each column of lt on T + D
b_td = as.matrix(reg_td$coefficients[2,1:26])  # get slope coeffs
d_td = as.matrix(reg_td$coefficients[3,1:26])  # get shift coeffs (note this is x120 deg C)
r_td = as.matrix(reg_td$residuals[,1:26])  # Get residuals

OMEGA2=omega2(r_td, K, N, T, D)  # ENTRIES FOR TABLES 3, 5, 6 , 7a  # TABLES 3 & 7a
Table3  = matrix(0, nrow = K, ncol = 6)  # Cols: Coeffs & Std Errors
Table7a = matrix(0, nrow = 26, ncol = 2)  # VF scores on shift terms - Not shown in paper
for (i in 1:K)
{
  R=t( c( rep(0,K) ) )  # make R = row of zeroes
  R[1,i] = 1  # set i-th element to 1
  
  VF = eta1 * (R %*% b_t) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_t)
  Table3[i,1] = b_t[i,1]
  Table3[i,2] = 6.482*abs( b_t[i,1]/chol(VF) )
  Table3[i,3] = b_t[i,1]
  Table3[i,4] = cvb*abs( b_t[i,1]/chol(VF) )
  Table3[i,5] = d_t[i,1]/120
  Table3[i,6] = cvb*abs( d_t[i,1]/chol(VF)/120 )
}
cat("Table 3 - LT done", 

# TABLE 5
Table5 = matrix(0, nrow = 22, ncol = 1)  # Trends: Zero tests and Model Equiv
R=t( c( rep(0,K) ) )  # make R = row of zeroes
R[1,24] = 1  # set 24th element (Had-LT) to 1
VF = eta1 * (R %*% b_t) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_t)
Table5[1,1] = VF
R[1,25] = 0  # set 25th element (RICH-LT) to 1
VF = eta1 * (R %*% b_t) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_t)
Table5[2,1] = VF
R[1,26] = 0  # set 26th element (RAOB-LT) to 1
VF = eta1 * (R %*% b_t) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_t)
Table5[3,1] = VF
R=t( as.matrix( c(ens/57,-1/3, -1/3, -1/3 ) ) )  # make R for test of models-obs

Table5[4,1] = VF
Table5[5,1] = VF
Table5[6,1] = VF
VF = \eta_1 \ast \left( R \ast b_t \right) \ast \left( \text{solve} \left[ R \ast \Omega_1 \ast t(R) \right] \right) \ast \left( R \ast b_t \right)

Table5[7,1] = VF

R= t( \text{as.matrix} \left( \text{c}(\text{rep}(0,K)) \right) )

R[1,24] = 1 \quad \# \text{set 24th element (Had-LT) to 1}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

Table5[9,1] = VF

R[1,24] = 0

R[1,25] = 1 \quad \# \text{set 25th element (RICH-LT) to 1}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

Table5[10,1] = VF

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

Table5[11,1] = VF

R[1,26] = 0

cat("Table 5 - LT done", "\n")

CHECK INDIVIDUAL BALLOON RESULTS

R=t( \text{as.matrix} \left( \text{c}(\text{ens}/57,-1, 0, 0) \right) ) \quad \# \text{make R for test of models=obs}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

cat(" Test if HadAT = models in LT", VF,"\n")

R=t( \text{as.matrix} \left( \text{c}(\text{ens}/57,0, -1, 0) \right) ) \quad \# \text{make R for test of models=obs}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

cat(" Test if RICH = models in LT", VF,"\n")

R=t( \text{as.matrix} \left( \text{c}(\text{ens}/57,0, 0, -1) \right) ) \quad \# \text{make R for test of models=obs}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

cat(" Test if RAOBCORE = models in LT", VF,"\n")

COMPUTE BASED ON AVERAGE OF ALL 3

R=t( \text{as.matrix} \left( \text{c}(\text{ens}/57,-1/3, -1/3, -1/3) \right) ) \quad \# \text{make R for test of models=obs}

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

Table5[15,1] = VF

R=t( \text{as.matrix} \left( \text{c}(\text{rep}(0,23), 1/3, 1/3, 1/3) \right) ) \quad \# \text{make R for test of avg shift=0}

VF = \eta_3 \ast \left( R \ast d_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast d_{td} \right)

Table5[17,1] = VF

TABLE 6

Table6 = matrix(0, nrow = models, ncol = 6) \quad \# Model-Obs 1 at a Time

for (i in 1:models) {

R=t( c(\text{rep}(0,models),-1/3, -1/3, -1/3) ) \quad \# \text{make R for test of models=obs}

VF = \eta_1 \ast \left( R \ast b_t \right) \ast \left( \text{solve} \left[ R \ast \Omega_1 \ast t(R) \right] \right) \ast \left( R \ast b_t \right)

Table6[i,1] = VF

VF = \eta_2 \ast \left( R \ast b_{td} \right) \ast \left( \text{solve} \left[ R \ast \Omega_2 \ast t(R) \right] \right) \ast \left( R \ast b_{td} \right)

Table6[i,2] = VF

}

cat("Table 6 - LT done", "\n")
mt = read.table("data/mt.txt", header=TRUE)
mt = 120*(as.matrix(mt))

# PART II: MT Run
#

reg_t  = lm(mt ~ T)     # regress each column of mt on time trend T
b_t = as.matrix(reg_t$coefficients[2,1:26])  # get slope coeffs
r_t = as.matrix(reg_t$residuals[,1:26])   # Get residuals
OMEGA1=omega1(r_t, K, N, T)  # BREAK CASE (at D)

reg_td = lm(mt ~ T + D)     # regress each column of mt on T + D
b_td = as.matrix(reg_td$coefficients[2,1:26])  # get slope coeffs
d_td = as.matrix(reg_td$coefficients[3,1:26])  # get shift coeffs (note this is x120 deg C)
r_td = as.matrix(reg_td$residuals[,1:26])   # Get residuals
OMEGA2=omega2(r_td, K, N, T, D)

# ENTRIES FOR TABLES 4, 5, 6 and 7a
#

for (i in 1:K)    {
R=t(  c( rep(0,K) ) )     # make R = row of zeroes
R[i,1] = 1     # set i-th element to 1
VF - eta1 * (R %*% b_t) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_t)
Table4[i,1] = b_t[i,1]  Table4[i,2] = 6.482*abs( b_t[i,1]/chol(VF) )
VF - eta2 * (R %*% b_td) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% b_td)
Table4[i,3] = b_td[i,1]  Table4[i,4] = cvb*abs( b_td[i,1]/chol(VF) )
VF - eta3 * (R %*% d_td) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% d_td)
Table4[i,5] = d_td[i,1]/120  Table4[i,6] = cvb*abs( d_td[i,1]/chol(VF)/120 )
Table7a[i,2] = VF
}
cat("Table 4 - MT done", "\n")
# TABLE 5

\[ R = t( c( \text{rep}(0,K) ) ) \]  

# make R - row of zeroes

\[ R[1,24] = 1 \]  

# set 24th element (Had-MT) to 1

\[ VF = \eta_1 \times (R \times b_t) \times (\text{solve}( R \times OMEGA1 \times t(R) ) \times (R \times b_t) \]

\[ \text{Table5}[4,1] = VF \]

\[ R[1,24] = 0 \]

\[ R[1,25] = 1 \]  

# set 25th element (RICH-MT) to 1

\[ VF = \eta_1 \times (R \times b_t) \times (\text{solve}( R \times OMEGA1 \times t(R) ) \times (R \times b_t) \]

\[ \text{Table5}[5,1] = VF \]

\[ R[1,26] = 0 \]

\[ R = t( \text{as.matrix}( c(\text{ens}/57,-1/3, -1/3, -1/3) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_1 \times (R \times b_t) \times (\text{solve}( R \times OMEGA1 \times t(R) ) \times (R \times b_t) \]

\[ \text{Table5}[6,1] = VF \]

\[ R[1,26] = 0 \]

\[ R = t( \text{as.matrix}( c(\text{rep}(0,K) ) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{Table5}[12,1] = VF \]

\[ R[1,24] = 0 \]

\[ R[1,25] = 1 \]  

# set 25th element (RICH-MT) to 1

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{Table5}[13,1] = VF \]

\[ R[1,25] = 0 \]

\[ R[1,26] = 1 \]  

# set 26th element (RAOB-MT) to 1

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{Table5}[14,1] = VF \]

\[ R[1,26] = 0 \]

\[ \text{cat("Table 5 - MT done", "\n")} \]

# CHECK INDIVIDUAL BALLOON RESULTS

\[ R = t( \text{as.matrix}( c(\text{ens}/57,-1/3,-1/3,-1/3) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{cat(" Test ifHadAT = models in MT", VF,"\n")} \]

\[ R = t( \text{as.matrix}( c(\text{ens}/57,0, -1/3, -1/3) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{cat(" Test if RICH = models in MT", VF,"\n")} \]

\[ R = t( \text{as.matrix}( c(\text{ens}/57,0, 0, -1) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{cat(" Test if RAOBCORE = models in MT", VF,"\n")} \]

# COMPUTE BASED ON AVERAGE OF ALL 3

\[ R = t( \text{as.matrix}( c(\text{ens}/57,-1, -1/3,-1/3,-1/3) ) ) \]  

# make R for test of models=obs

\[ VF = \eta_2 \times (R \times b_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times b_{td}) \]

\[ \text{Table5}[16,1] = VF \]

\[ R = t( \text{as.matrix}( c(\text{rep}(0, \text{times}=23), 1/3, 1/3, 1/3) ) ) \]  

# make R for test of avg shift=0

\[ VF = \eta_3 \times (R \times d_{td}) \times (\text{solve}( R \times OMEGA2 \times t(R) ) \times (R \times d_{td}) \]

\[ \text{Table5}[18,1] = VF \]

\[ \text{cat("Table 5 - MT done", "\n")} \]
# TABLE 6

for (i in 1:models)
  { 
    R = t(c(rep(0,models), -1/3, -1/3, -1/3))  # make R for test of models=obs
    V = eta1 * (R %*% b_t) %*% (solve(R %*% OMEGA1 %*% t(R))) %*% (R %*% b_t)
    Table6[i,4] = V
    V = eta2 * (R %*% b_td) %*% (solve(R %*% OMEGA2 %*% t(R))) %*% (R %*% b_td)
    Table6[i,5] = V
  }

cat("Table 6 - MT done", "\n")

# # PART III: DRAW FIGURES 2-4 (FIG 5 AT END) # #

# FIGURE 2

hadltmln = lm(lt[,24]/120 ~ tt)  # Compute trend lines
hadmttmln = lm(mt[,24]/120 ~ tt)
ricltmln = lm(lt[,25]/120 ~ tt)
ricmttmln = lm(mt[,25]/120 ~ tt)
raolttmln = lm(lt[,26]/120 ~ tt)
raomttmln = lm(mt[,26]/120 ~ tt)

postscript('graphics/fig2.eps')  # write Figure to EPS file

par(mfrow = c(3, 2))  # create 3x2 panel figure

par(mar=c(1.5,5,2.5,0), yaxt="s", xaxt="n")
plot(lt[,24]/120 ~ tt, ylab = "Hadley deg C anomaly", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)
abline(hadltmln)

par(mar=c(1.5,3,2.5,1.5), yaxt="s", xaxt="n")
plot(mt[,24]/120 ~ tt, ylab = "", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)
abline(hadmttmln)

par(mar=c(2.5,1.5,0), yaxt="s", xaxt="n")
plot(lt[,25]/120 ~ tt, ylab = "RICH deg C anomaly", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)
abline(ricltmln)

par(mar=c(2.5,1.5,1.5), yaxt="s", xaxt="n")
plot(mt[,25]/120 ~ tt, ylab = "", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)
abline(ricmttmln)

par(mar=c(4.5,5,0,0), yaxt="s", xaxt="s")
plot(lt[,26]/120 ~ tt, ylab = "RAOB deg C anomaly", xlab = "LT", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)
abline(raolttmln)
par(mar=c(4.5,3,0,1.5), xaxt="s", xaxt="s")  
plot(mt[,26]/120 ~ tt, ylab = "", xlab = "MT", cex=0.5, lwd=1.5, cex.axis=1.75, cex.lab=1.5)  
abline(raomtlin)  
dev.off()  

cat("Figure 2 done", "\n")  

# FIGURE 3  
postscript('graphics/fig3.eps')  
par(mfrow = c(2, 1), xaxt="n", mar=c(1,5,2,2) )  
# create 2-panel figure with tight margins  
mmtavg = rowMeans(lt[,1:23])/120  
omtavg = (lt[,24]+lt[,25]+lt[,26])/360  
# compute model avg  
# balloon average on decadal scale  
yo=predict( lm( omtavg ~ tt + D ) )  
ym=predict( lm( mmtavg ~ tt + D ) )  
# Put broken model trend line on same diagram  
# Put broken obs trend line on same diagram  
yo=yo-(yo[1]-ym[1])  
# Line up values in start years  
par(xaxt="s", mar=c(2,5,0,2) )  
# repeat for MT layer  
plot(mt[,1]/120 ~ tt, cex=.05, col="red", ylab = "MT deg C", ylim=c(-2 , 2), cex.axis=1.75, cex.lab=1.5)  
for (i in 2:23)  
  par(new="T")  
  plot(mt[,i]/120 ~ tt, cex=.05, col="red", axes=F, ylab="", ylim=c(-2 , 2) )  
  # ADD MODEL FIT  
  par(new="T")  
  plot(ym ~ tt, type="l", lwd=6, col="darkred", lty=1, axes=F, ylab="", ylim=c(-2 , 2) )  
  text(2000, 0.6, "MT Models", xpd=TRUE, pos=3, offset=2, cex=2, col="darkred")  
  # ADD OBS FIT  
  par(new="T")  
  plot(yo ~ tt, type="l", lwd=6, col="blue", lty=1, axes=F, ylab="", ylim=c(-2 , 2) )  
  text(2000, -0.2, "MT Observations", xpd=TRUE, pos=1, offset=.5, cex=2, col="blue")  
mltavg = rowMeans(lt[,1:23])/120  
oltavg = (lt[,24]+lt[,25]+lt[,26])/360  
# compute model avg  
# balloon average on decadal scale  
yo=predict( lm( oltavg ~ tt + D ) )  
ym=predict( lm( mltavg ~ tt + D ) )  
# Put broken model trend line on same diagram  
# Put broken obs trend line on same diagram  
yo=yo-(yo[1]-ym[1])  
# Line up values in start years  
plot(lt[,1]/120 ~ tt, cex=.05,col="red", ylab = "LT deg C", ylim=c(-2 , 2), cex.axis=1.75, cex.lab=1.5)  
for (i in 2:23)  
  par(new="T")  
  plot(lt[,i]/120 ~ tt, cex=.05, col="red", axes=F, ylab="", ylim=c(-2 , 2) )  
  # ADD MODEL FIT  
  par(new="T")  
  plot(ym ~ tt, type="l", lwd=6, col="darkred", lty=1, axes=F, ylab="", ylim=c(-2 , 2) )  
  text(2000, 0.6, "LT Models", xpd=TRUE, pos=3, offset=2, cex=2, col="darkred")  
  # ADD OBS FIT  
  par(new="T")  
  plot(yo ~ tt, type="l", lwd=6, col="blue", lty=1, axes=F, ylab="", ylim=c(-2 , 2) )  
  text(2000, -0.2, "LT Observations", xpd=TRUE, pos=1, offset=.5, cex=2, col="blue")
```r
# ADD OBS FIT
par(new="T")
plot(yo ~ tt, type="l", lwd=6, col="blue", lty=1, axes=F, ylab="", ylim=c(-2 , 2) )
text(2005, 0.0, "LT Observations", xpd=TRUE, pos=1, offset=.5, cex=2, col="blue")
dev.off()
cat("Figure 3 done", "\n")

# FIGURE 4

postscript("graphics/fig4.eps") # write figure to EPS file
par(mfrow = c(2, 2)) # create 2-panel figure with tight margins
jcolors=c("red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","red","blue","blue","blue") # top left - MT no break  # sort coefs and cis
mbeta = Table4[1:23,1] # model trends
mciw  = Table4[1:23,2] # model CI width
z = order(mbeta) # vector indicating order low-hi
beta = c(mbeta[z] , Table4[24:26,1]) # ordered model trends then obs trends
ciw  = c(mciw[z] , Table4[24:26,2])
par(mar=c(2,5,2,0), xaxt="n")
plotCI(1:26, beta, ciw, cex=1.5, ylim=c(-0.2 , 0.6), xlab="", ylab="MT Trend in C/decade", cex.axis=1.75, cex.lab=1.5, col=jcolors)
text(1:23, beta[1:23]+ciw[1:23]+.05, z, cex=.8) # show model numbers at top of error bars
abline(h=0, lty=2) # top right - MT with breaks  # sort coefs and cis
mbeta = Table4[1:23,3] # model trends
mciw  = Table4[1:23,4] # model CI width
z = order(mbeta) # vector indicating order low-hi
beta = c(mbeta[z] , Table4[24:26,3]) # ordered model trends then obs trends
ciw  = c(mciw[z] , Table4[24:26,4])
par(mar=c(2,4,2,1), xaxt="n")
plotCI(1:26, beta, ciw, ylim=c(-0.2 , 0.6), xlab="", ylab="", cex.axis=1.75, cex.lab=1.5, col=jcolors)
text(1:23, beta[1:23]+ciw[1:23]+.05, z, cex=.8) # show model numbers at top of error bars
abline(h=0, lty=2) # bottom left - LT no break  # sort coefs and cis
mbeta = Table3[1:23,1] # model trends
mciw  = Table3[1:23,2] # model CI width
z = order(mbeta) # vector indicating order low-hi
beta = c(mbeta[z] , Table3[24:26,1]) # ordered model trends then obs trends
ciw  = c(mciw[z] , Table3[24:26,2])
par(mar=c(2,4,2,1), xaxt="n")
plotCI(1:26, beta, ciw, ylim=c(-0.2 , 0.6), xlab="No Break", ylab="LT Trend in C/decade", cex.axis=1.75, cex.lab=1.5, col=jcolors)
```
\text{(1:23, beta[1:23]+ciw[1:23]+.05, z, cex=.8)} \quad \# \text{show model numbers at top of error bars}
\text{abline(h=0, lty=2)}

\text{\# bottom right - LT with breaks}
\text{\# sort coefs and cis}
\text{mbeta = Table3[1:23,3]} \quad \# \text{model trends}
\text{mciw = Table3[1:23,4]} \quad \# \text{model CI width}
z = order(mbeta) \quad \# \text{vector indicating order low-hi}
beta = c(mbeta[z], Table3[24:26,3]) \quad \# \text{ordered model trends then obs trends}
ciw = c(mciw[z], Table3[24:26,4])
\text{par(mar=c(4,4,0,1), xaxt="s")}
\text{plotCI(1:26, beta, ciw, cex=1.5, ylim=c(-0.2, 0.6), xlab="With Break", ylab="", cex.axis=1.75, cex.lab=1.5, col=jcolors)}
\text{text(1:23, beta[1:23]+ciw[1:23]+.05, z, cex=.8)} \quad \# \text{show model numbers at top of error bars}
\text{abline(h=0, lty=2)}
\text{dev.off()}
cat("Figure 4 done", "\n")

\text{\# PART IV: BOOTSTRAPPING CRITICAL VALUES FOR 1-BREAK CASE}
\text{\# Notes: u is matrix of residuals from regression of mt on T+D}
\text{\# Method: Use u(i) in place of mt(i) in regression}
\text{\# Compute new Omega matrix and VF score for Rb=0, using eta2 scaling term}
\text{\# Take 90th, 95th, 97.5th and 99th percentiles}
\text{\#} 
\text{cat("Beginning bootstrap on", bsr, "reps", "\n")}
\text{vv1 = matrix(0, nrow=bsr, ncol=1)}
\text{vv2 = matrix(0, nrow=bsr, ncol=1)}
\text{reg_t = lm(lt ~ T)} \quad \# \text{Generate set of residuals - no break}
\text{r_t = as.matrix(reg_t$residuals[,1:26])} \quad \# 
\text{reg_td = lm(lt ~ T + D)} \quad \# \text{Generate set of residuals - with break}
\text{r_td = as.matrix(reg_td$residuals[,1:26])} \quad \# 
\text{R=t(as.matrix(c(ens/57,-1/3,-1/3,-1/3)))} \quad \# \text{make R for test of models=obs}
\text{for (i in 1:bsr)}
\text{\# BOOTSTRAP ROUTINE BEGINS}
\text{r1 = apply(r_t, 2, sample, replace=TRUE)} \quad \# \text{generate bootstrap resample of residuals (T)}
\text{bt = lm(r1 ~ T)} \quad \# \text{regress bootstrap sample on T}
\text{b_tb = as.matrix(bt$coefficients[2,1:26])} \# \text{get slope coeffs}
# For Peer Review

```r
r_tb = as.matrix(bt$residuals[,1:26])  # Get residuals
OMEGA1 = omega1(r_tb, K, N, T)
VF = eta1 * (R %*% b_tb) %*% (solve( R %*% OMEGA1 %*% t(R) ) ) %*% (R %*% b_tb)
vv1[i,1] = VF

r2 = apply(r_td, 2, sample, replace=TRUE) # generate bootstrap resample of residuals (T+D)
btd = lm(r2 ~ T+D)    # regress bootstrap sample on T+D
b_tdb = as.matrix(btd$coefficients[2,1:26]) # get slope coeffs
r_tdb = as.matrix(btd$residuals[,1:26])  # Get residuals
OMEGA2 = omega2(r_tdb, K, N, T, D)
VF = eta2 * (R %*% b_tdb) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% b_tdb)
vv2[i,1] = VF

# BOOTSTRAP ROUTINE ENDS

ds1 = sort(vv1)     # sort vv values lowest to highest
d2 = sort(vv2)
count = c(1:bsr)  pvalue = (bsr-count+1)/bsr
d1 = cbind(count,ds1,pvalue)
d2 = cbind(count,d2,pvalue)
cat("VF 5th P-values", "n")
cat("90th percentile (-----)", d1[pct20, "n"]
cat("90th percentile (27.14)", d1[pct10, "n"]
cat("95th percentile (41.53)", d1[pct05, "n"]
cat("97.5th percentile (58.57)", d1[pct025, "n"]
cat("99th percentile (83.96)", d1[pct01, "n", "n"]
cat("VF P-values w/break", "n")
cat("90th percentile (32.35)", d2[pct20, "n"]
cat("95th percentile (49.40)", d2[pct10, "n"]
cat("97.5th percentile (70.26)", d2[pct025, "n"]
cat("99th percentile (99.98)", d2[pct01, "n"]

# Save these results
write(t(d1), file="logfiles/vfbootstrap_nobreak.csv", sep="", ncolumns=3)
write(t(d2), file="logfiles/vfbootstrap_1break.csv", sep="", ncolumns=3)
cat("\n","Bootstrap done", "n")
```

---

## PART V: supVF GRID SEARCH

```r
cat("Beginning grid search", "n")
```
gsl = matrix(0, nrow=N, ncol=models+2) # grid search results, 636 rows, 25 cols

gsm = matrix(0, nrow=N, ncol=models+2) # grid search results, 636 rows, 25 cols
# cols 1-23: MODEL i=AVG OBS        # col 24: AVG MODEL=AVG OBS        # col 25: AVG OBS SHIFT TERM=0

start = round( trim*N ) # search across sample with trim removed
end = round( (1-trim)*N ) # on each end

# GRID SEARCH BEGINS

for (j in start:end) {

D = as.numeric(T > j) # 0-1 dummy for break
beta2 = lm(T ~ 1 + D )
ttilde2 = as.matrix(beta2$residuals) # recompute partialling out term for slope
eta2 = sum( ttilde2^2 )
beta3 = lm(D ~ 1 + T )
ttilde3 = as.matrix(beta3$residuals) # recompute partialling out term for shift
eta3 = sum( ttilde3^2 )
rn(beta2, beta3)

# LT DATA
d = c(1, rep( 0 , times=N-1)) # reset d to T-length vector of 1 followed by N-1 zeroes
reg_td = lm(lt ~ T + D) # regress each column of lt on T +D
b_td = as.matrix(reg_td$coefficients[2,1:26]) # get slope coeffs
d_td = as.matrix(reg_td$coefficients[3,1:26]) # get break coeffs
r_td = as.matrix(reg_td$residuals[,1:26]) # Get residuals
OMEGA2 = omega2(r_td, K, N, T, D) # MODEL i = AVG OBS:  for (i in 1:models) {
R=t( c( rep(0,models), -1/3, -1/3, -1/3 ) ) # make R for test of models=obs
R[1,i] = 1
VF = eta2 * (R %*% b_td) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% b_td)
gsl[j,i] = VF
}

# AVG OBS = AVG OBS TEST:
R=t( c( ens/57,-1/3, -1/3, -1/3 ) ) # make R for test of models-obs
VF = eta2 * (R %*% b_td) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% b_td)
gsl[j,24] = VF

# AVG SHIFT TERM = 0
VF = eta2 * (R %*% d_td) %*% (solve( R %*% OMEGA2 %*% t(R) ) ) %*% (R %*% d_td)
gsl[j,25] = VF
# MT DATA
d = c(1, rep(0, times=N-1))  # reset d to T-length vector of 1 followed by N-1 zeroes
reg_td = lm(mt ~ T + D)  # regress each column of mt on T + D
b_td = as.matrix(reg_td$coefficients[2,1:26])  # get slope coeffs
d_td = as.matrix(reg_td$coefficients[3,1:26])  # get break coeffs
rTd = as.matrix(reg_td$residuals[,1:26])  # Get residuals

OMEGA2 = omega2(rTd, K, N, T, D)  # MODEL i = AVG OBS:
for (i in 1:models) {
  R = t( c(rep(0,models), -1/3, -1/3, -1/3) )  # make R for test of models=obs
  R[1,i] = 1
  VF = eta2 * (R %*% b_td) %*% (solve(R %*% OMEGA2 %*% t(R)) %*% (R %*% b_td))
gsm[,i] = VF
}

# AVG MODELS = AVG OBS TEST:
R = t(as.matrix(c(ens/57,-1/3, -1/3, -1/3)))  # make R for test of models=obs
VF = eta2 * (R %*% b_td) %*% (solve(R %*% OMEGA2 %*% t(R)) %*% (R %*% b_td))
gsm[,24] = VF  # AVG SHIFT TERM = 0
VF = eta3 * (R %*% d_td) %*% (solve(R %*% OMEGA2 %*% t(R)) %*% (R %*% d_td))
gsm[,25] = VF  # GRID SEARCH ENDS
}
cat("Grid search done", 

maxl = apply(gsl, 2, max)  # take supremums of VF scores across grid search
maxm = apply(gsm, 2, max)
write(t(gsl), file="logfiles/gsl.txt", sep="", ncolumns=25)  # write gridsearch matrices. Col 24 = graphed result
write(t(gsm), file="logfiles/gsm.txt", sep="", ncolumns=25)  #

Table5[1,19] = maxl[24]  # Write to Table 5
Table5[2,20] = maxm[24]
Table5[3,21] = maxl[25]
Table5[4,22] = maxm[25]

Table6[,3] = t(maxl[1:23])  # Write to Table 6
Table6[,6] = t(maxm[1:23])

Table6 = rbind( Table6, c( sum(Table6[,1]>41.53), sum(Table6[,2]>49.40), sum(Table6[,3]>166.41), sum(Table6[,4]>41.53), sum(Table6[,5]>49.40), sum(Table6[,6]>166.41))

"\n")

# PART VI: SAVE TABLES AND DRAW FIGURE 5

# Save these results
Table3 = round(Table3, 3)
write(t(Table3), file="logfiles/table3.csv", sep=" ", ncolumns=6)
Table4 = round(Table4, 3)
write(t(Table4), file="logfiles/table4.csv", sep=" ", ncolumns=6)
Table5 = round(Table5, 1)
write(t(Table5), file="logfiles/table5.csv", sep=" ", ncolumns=1)
Table6 = round(Table6, 2)
write(t(Table6), file="logfiles/table6.csv", sep=" ", ncolumns=6)
Table7a = round(Table7a, 2)
write(t(Table7a), file="logfiles/table7a.csv", sep=" ", ncolumns=2)
cat("CSV files written", "\n")

# FIGURE 5
postscript('graphics/fig5.eps')    # write figure to EPS file
par(mfrow = c(2, 1))  par(mar=c(2,5,2,1), xaxt="s")  plot(tt, gsm[,24], type="l", xlab="", ylab = "MT", ylim=c(0,1000), cex.lab=1.5, cex.axis=1.75 )
  abline(h=131.92, lty="solid")
  abline(h=166.41, lty="dashed")
  abline(h=261.39, lty="dotted")
par(mar=c(4,5,0,1), xaxt="s")  plot(tt, gsl[,24], type="l", xlab="Year", ylab = "LT", ylim=c(0,1000), cex.lab=1.5, cex.axis=1.75 )
  abline(h=131.92, lty="solid")
  abline(h=166.41, lty="dashed")
  abline(h=261.39, lty="dotted")
# dev.off()  cat("Figure 5 done", "\n")  cat("End of Program", "\n")
5  CODE FOR SIMULATION OF ASYMPTOTIC CRITICAL VALUES

/* table1a.g  This program simulates asymptotic critical values for Table1a in the McKitrick-Vogelsang paper HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS

Gauss code written and updated by Tim Vogelsang 6/10/2014 */

new;
#lineson;

output file = table1a.out reset;
output off;

T=1000;  N=50000;
fstat=zeros(N,1);
cc=ones(T,1);
t1=seqa(1,1,T);

ppt=0.7|0.75|0.8|0.85|0.9|0.95|0.975|0.99|0.995;
ppti=trunc((1-2*(1-ppt))*N+0.0001);

ppf=0.7|0.75|0.8|0.85|0.9|0.95|0.975|0.99|0.995;
ppfi=trunc(ppf*N+0.0001);

lambda=240/660;
do while lambda<=(240/660);
   TB=trunc(lambda*T+0.0001);
   DU=zeros(TB,1);ones(T-TB,1);
   X=cc-XU-t1;
   XXinv=invpd(X'X);
   l=1;  seed=T;
   do while l<=N;
      z=rndn(1,1,seed);
      y=rndn(T,1,seed);
      bhat=y/X;
      uhat=y-X*bhat;
      Shat=recserar(uhat,uhat(1),1.0);
      omhat=2*Shat'Shat/(T*T);
      fstat[l]=z'*z/omhat;
      if l>1000--0;
         output file = "status.out" on;
         print "l:" 1 "lambda=" lambda;
         output off;
         l=l+1;
      endif;
      l=l+1;
   endo;
ii=1;
do while ii<cols(fstat);
    stats=sortmc(fstat[:,ii],1);
    fstat[:,ii]=stats;
    ii=ii+1;
endo;

output file = table1a.out on;
format 5,4;
print;
print "Asymptotic Critical Values.  FV-t in model with";
print "intercept, intercept shift and trend";
print "Intershift shift occurs at break point lambda = " lambda;
format 6,5;
print "  N       T"; print N-T;print;
format /rd 7,3;
print "  pp      FV-t";
print ppt-sqrt(fstat[ppti,.]); print; print;
format 5,4;
print;
print "Asymptotic Critical Values.  FV in model with";
print "intercept, intercept shift and trend";
print "Intershift shift occurs at break point lambda = " lambda;
format 6,5;
print "  N       T"; print N-T;print;
format /rd 7,3;
print "  pp      FV";
print ppf-fstat[ppfi, .]; print; print;
output off;

lambda=lambda+0.05;
endo;
end;

/*/ table1b.g This program simulates asymptotic critical values for Table1b in the McKitrick-Vogelsang paper HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS Gauss code written and updated by Tim Vogelsang 6/10/2014 */
supfstat3=zeros(N,1);
cc=ones(T,1);
t1=seqa(1,1,T);

ppf=0.7|0.75|0.8|0.85|0.90|0.95|0.975|0.99|0.995;
ppfi=trunc(ppf*N+0.0001);

l=1; seed=T;
do while l<=N;
y=rndns(T,1,seed);
TB=trunc(lamstar*T);
do while TB<T-trunc(lamstar*T);
DU=zeros(TB,1)|ones(T-TB,1);
X=cc-DU-t1;
XXinv=invpd(X'X);
bhat=y/X;
uchat=y-X*bhat;
Shat=recserar(uchat,uchat[1],1.0);
omhat=2*Shat'Shat/({T-T});

fstat=(bhat[3]^2)/(omhat*XXinv[3,3]);
if fstat>supfstat3[l]; supfstat3[l]=fstat; endif;
TB=TB+1;
endo;
if l%100==0;
    output file = "status.out" on;
    print "l=
    output off;
endif;
l=l+1;
endo;

ii=1;
do while ii<=cols(supfstat2);
    stats=sortmc(supfstat2[,ii],1);
supfstat2[,ii]=stats;
    stats=sortmc(supfstat3[,ii],1);
supfstat3[,ii]=stats;
    ii=ii+1;
endo;

output file = table1btrend.out on;
format 4,3;
print; print "Asymptotic Critical Values. SupFV in model with";
print "intercept, intercept shift and trend";
print "intercept shift at an unknown data";
print "trimming = " lamstar;
6 Code For Finite Sample Simulations

/* table2a.g This program simulates finite sample results for Table2a in the McKitrick-Vogelsang paper HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS Gauss code written and updated by Tim Vogelsang 6/10/2014 */

new;
#lineson;

Rloop=1;
do while Rloop<3.1;
  if Rloop==1; T=120; endif;
  if Rloop==2; T=240; endif;
  if Rloop==3; T=660; endif;
  TB=trunc((240/660)*T); /* break date corresponding to breakpoint 0.3636 */
  N=50000; /* number of replications */
kernel=1; /* kernel = 1 uses the Bartlett kernel */
cvchi=3.84; /* standard chi-square 5% critical value */
cVF=41.53; /* VF 5% critical value, no level shift in model */
cvVFbrk=49.399; /* VF 5% critical value when level shift is included at breakpoint 0.3636 */
cvsupVF=166.41; /* supVF 5% critical value with 10% trimming */

stat1=zeros(N,1);
stat2=zeros(N,1);
stat3=zeros(N,1);
stat4=zeros(N,1);
stat5=zeros(N,1);
stat6=zeros(N,1);
stat7=zeros(N,1);

cc=ones(T,1); /* intercept variable */
tt=seqa(1,1,T); /* time trend variable */
du=(tt>TB); /* level shift dummy with breakpoint 0.3636 */

dutild=du-meanc(du);
ttild=tt-meanc(tt);

x=ones(T,1)~tt;

beta1=0.01; /* trend slope of series 1, the series that could have level shift */
delta1=0.0; /* intercept shift parameter of series 1, g1 in the paper */
do while delta1<0.0001;

rhomat={0 0, 0.3 0, 0.6 0, 0.9 0.0, 0.3 0.3, 0.6 0.3, 0.9 -0.3};
iii=1;
do while iii<rows(rhomat);

rho=rhomat[iii, .];
if delta1>0; beta2inc=(delta1*(dutild'ttild)/(ttild'ttild))); endif;
if delta1==0; beta2inc=0.0025/(1-rho[1]-rho[2])); endif;

beta2=0.01; /* trend slope of series 2 */
do while beta2<0.010001;

scalvar=sqrt(((1-rho[2])^2-rho[1]^2)*(1+rho[2])/(1-rho[2])); /* scaling needed to normalize variance of AR(2) processes to 1 */
l=1; seed=T; count=zeros(1,7); /* simulation interation loop */
do while l<=N;
e1=rndns(T,1,seed);
e2=rndns(T,1,seed);
e1=recserar(e1,e1[1:2],rho');
e1=scalvar*e1;
e2=recserar(e2,e2[1:2],rho');
e2=scalvar*e2;
e2=(e2*e2coeff*e1)/sqrt(1+e2coeff^2);
y1=beta1*tt+delta1*du+e1; /* series 1 has level shift */
y2=beta2*tt+e2; /* series 2 do NOT have level shift */

/* compute test statistics with no level shift dummy in model */
x0=tt;
x1=cc;
x=x0-x1;
x0tild=x0-x1*(x0/x1); aainv=invpd(x0tild'x0tild);

b1hat=y1/x;
betalhat=b1hat[1];
uhat1=y1-x*b1hat;

b2hat=y2/x;
betahat=b2hat[1];
uhat2=y2-x*b2hat;

uhat=uhat1-uhat2;
Suhat=recserar(uhat,uhat[1,.],ones(1,2));

omhatar1=sig2npm(uhat,1,1,1); /* AR(1) parametric HAC W-pw in paper*/
omhat=sig2npm(uhat,-1,1,0); /* HAC with data dependent bandwidth W-bart in paper*/
omhatT=2*Suhat'Suhat/(T*T); /* VF using M-T */
R=1~-1;
bhat=beta1hat|beta2hat;

stat1[l]=(R*bhat)'invpd(aainv*R*omhatar1*R')*R*bhat;
stat2[l]=(R*bhat)'invpd(aainv*R*omhat*R')*R*bhat;
stat3[l]=(R*bhat)'invpd(aainv*R*omhatT*R')*R*bhat;


/* compute test statistics with level shift dummy in model */

stat7[l]=0;
TBstar=trunc(0.1*T);
ii=TBstar;
do while ii<=T-TBstar;
duest=tt.>ii; /* level shift dummy corresponding to break date ii */
x0=tt;
x1=cc-duest;
x=x0-x1;
x0tild=x0-x1*(x0/x1);
aainv=invpd(x0tild'x0tild);

b1hat=y1/x;
betalhat=b1hat[1];
what1=y1-x*b1hat;

b2hat=y2/x;
betalhat=b2hat[1];
what2=y2-x*b2hat;

what=what1-what2;
Suhat=recsevar(what,what[1,],ones(1,2));
if ii==TB; /* only calculate these variance estimators for the known break date case */
  omhatar1=sig2npm(what,1,1,1); /* AR(1) parametric HAC */
  omhat=sig2npm(what,-1,1,0); /* HAC with data dependent bandwidth */
endif;
omhatT=2*Suhat'Suhat/(T*T); /* VF using M=T */

R=1~-1;
bhat=beta1hat|beta2hat;

a7=(R*bhat)'invpd(aainv*R*omhatT*R')*R*bhat;
if a7>stat7[l]; stat7[l]=a7; endif; /* SupVF calculation */

if ii==TB; /* compute statistics for case where break date is assumed known */
  stat4[l]=(R*bhat)'invpd(aainv*R*omhatar1*R')*R*bhat;
  stat5[l]=(R*bhat)'invpd(aainv*R*omhat*R')*R*bhat;
endif;
stat6[l]=a7;
endif;

ii=ii+1;
endo;


if l%100==0;
    output file = "status.out" on;
    print "l=" l;
    output off;
endif;
l=l+1;
endo;

output file = table2a.out on;

format /rd 5,3;
print delta1-rho-10*beta2~count;
output off;
format 12,10;

beta2=beta2+beta2inc;
endo;
iiii=iiii+1;
endo;
delta1=delta1+0.25;
endo;
e2coeff=e2coeff+0.5;
endo;
tloop=tloop+1;
endo;

do;

/* table2b.g This program simulates finite sample results for Table2b in the McKitrick-Vogelsang paper HAC-ROBUST TREND COMPARISONS AMONG CLIMATE SERIES WITH POSSIBLE LEVEL SHIFTS Gauss code written and updated by Tim Vogelsang 6/10/2014 */
new;
#lineson;

T=660; /* sample size */
TB=trunc((240/660)*T); /* break date corresponding to breakpoint 0.3636 */
N=50000; /* number of replications */
kernel=1; /* kernel = 1 uses the Bartlett kernel */
cvchi=3.84; /*standard chi-square 5% critical value */
cvVF=41.53; /* VF 5% critical value, no level shift in model */
cvVFbrk=49.399; /*VF 5% critical value when level shift is included at breakpoint 0.3636 */
cvsupVF=166.41; /* supVF 5% critical value with 10% trimming */

stat1=zeros(N,1);
stat2=zeros(N,1);
stat3=zeros(N,1);
stat4=zeros(N,1);
stat5=zeros(N,1);
stat6=zeros(N,1);
stat7=zeros(N,1);

cc=ones(T,1); /* intercept variable */
tt=seqa(1,1,T); /* time trend variable */
du=(tt>TB); /* level shift dummy with breakpoint 0.3636 */
du1=du-meanc(du);
ttild=tt-meanc(tt);
sttild2=ttild'ttild;
x=ones(T,1)~tt;
pp=1|2.5|5|10|50|90|95|97.5|99; /* powers */
pp1=pp*0.01*N;
pp2=seqa(1,1,99); pp21=pp2*0.01*N;

e2coeff=0;
do while e2coeff<0.501;

output file = table2b.out on;
format 8,6;
print;
print "Finite Sample Empirical Rejection Probabilites and Power, 5% nominal level";
print "Tests for common trends";
print "Wald test implemented using Bartlett kernel with Andrews VAR(1)";
print "data dependent bandwidth";
print "All series have AR(1) errors with coefficient rhol";
if e2coeff=0; print "Errors of series 2 equals an AR(1)+0.5e1 where e1 is the error for series 1"; endif;
print "Series 1 has trend slope beta1=0.01 and level shift g1";
print "Breakpoint for level shift is 0.3636. SupVF uses 10% Trimming";

format 7,0;
print " N   T"; print N-T; print;
print "g1   rhol 10bet2 W-pw WBart VF W-pw Wbart VF SupVF";
print;
output off;

beta1=0.01; /* trend slope of series 1, the series that could have level shift */
delta1=0.0; /* intercept shift parameter of series 1, \( g_1 \) in the paper */
do while delta1<0.2501;

thetrho={0 0, 0 0.9};
diii=1;
do while iii<rows(thetrho);

theta=thetrho[iii,1];
rho=thetrho[iii,2];
beta2inc=(delta1*(dutild\'ttild)/(ttild\'ttild)); /* set increment for beta2 */
if delta1==0; beta2inc=(.25*(dutild\'ttild)/(ttild\'ttild)); endif;
/* beta2inc=0.0025*(1+theta)/(1-rho); */
beta2=0.01; /* trend slope of series 2 */
do while beta2<(0.01+4*beta2inc+0.00001);

scalarm=sqrt((1-rho^2)/(1+theta^2+2.0*rho*theta)); /* scaling needed to normalize variance of ARMA processes to 1 */
l=1; seed=T; count=zeros(1,7); /* simulation iteration loop */
do while l<=N;
e1=rndns(T,1,seed);
e2=rndns(T,1,seed);
e1=e1+theta*(0|e1[1:T-1]);
e2=e2+theta*(0|e2[1:T-1]);
e1=recserar(e1,e1[1],rho);
e1=scalarm*e1;
e2=recserar(e2,e2[1],rho);
e2=scalarm*e2;
e2=(e2+e2coeff*e1)/sqrt(1+e2coeff^2);
y1=beta1*tt+delta1*du+e1; /* series 1 has level shift */
y2=beta2*tt+e2; /* series 2 do NOT have level shift */
/* compute test statistics with no level shift dummy in model */

x0=tt;
x1=cc;
x=x0-x1;
x0ttild=x0-x1*(x0/x1); aainv=invpd(x0ttild\'x0ttild);

b1hat=y1/x;
beta1hat=b1hat[1];
what1=y1-x*b1hat;

b2hat=y2/x;
beta2hat=b2hat[1];
what2=y2-x*b2hat;
uhat=uhat1~uhat2;
Suhat=recserar(uhat,uhat[1,.],ones(1,2));

omhatar1=sig2npm(uhat,1,1,1); /* AR(1) parametric HAC */
omhat=sig2npm(uhat,-1,1,0);  /* HAC with data dependent bandwidth */
omhatT=2*Suhat'Suhat/(T*T); /* VF using M=T */

R=1--1;
bhat=betal1hat|beta2hat;

stat1[l]=(R*bhat)'invpd(aainv*R*omhatar1*R')*R*bhat;
stat2[l]=(R*bhat)'invpd(aainv*R*omhat*R')*R*bhat;
stat3[l]=(R*bhat)'invpd(aainv*R*omhatT*R')*R*bhat;


/* compute test statistics with level shift dummy in model */
/* compute test statistics with level shift dummy in model */

stat7[l]=0;
TBstar=trunc(0.1*T);
ii=TBstar;
do while ii<=(T-TBstar);
duest=tt.>ii; /* level shift dummy corresponding to break date ii */
x0=tt;
x1=cc~duest;
x=x0-x1;
x0tild=x0-x1*(x0/x1);
aainv=invpd(x0tild'x0tild);

b1hat=y1/x;
betal1hat=b1hat[1];
what1=y1-x*b1hat;

b2hat=y2/x;
betal2hat=b2hat[1];
what2=y2-x*b2hat;

what=what1~what2;
Suhat=recserar(what,what[1,.],ones(1,2));

if ii==TB; /* only calculate these variance estimators for the known break date case */
omhatar1=sig2npm(what,1,1,1); /* AR(1) parametric HAC */
omhat=sig2npm(what,-1,1,0);  /* HAC with data dependent bandwidth */
endif;
omhatT=2*Suhat'Suhat/(T*T); /* VF using M-T */
R=1~1;
bhat=beta1|beta2hat;

a7=(R*bhat)'invpd(aainv*R*omhat*R')*R*bhat;
if a7>stat7[l]; stat7[l]=a7; endif; /* SupVF calculation */

if ii==TB; /* compute statistics for case where break date is assumed known */
  stat4[l]=(R*bhat)'invpd(aainv*R*omhatar1*R')*R*bhat;
  stat5[l]=(R*bhat)'invpd(aainv*R*omhat*R')*R*bhat;
  stat6[l]=a7;
  endif;
  ii=ii+1;
endo;


if l%100==0;
  output file = "status2.out" on;
  print "l=", l;
  output off;
  l=l+1;
endo;

output file = table2b.out on;
format /rd 5,3;
print delt1-rho-10*beta2-count;
output off;
format 12,10;
beta2=beta2+beta2inc;
endo;
iii=iii+1;
endo;
delt1=delt1+0.25;
endo;
e2coeff=e2coeff+0.5;
endo;
end;

/* This procedure computes the spectral density matrix at frequency zero for a vector of time series. It uses the procedure sig2npv(). It is assumed that the first series in the vector corresponds
to a constant in the regression model.

\( v \) - matrix of time series (each column contains one series)

ST - truncation lag, set ST = -1 to use the automatic truncation lag of Andrews (1991) based on the AR(1) plugin method.

kernel - an integer representing the kernel that is used
1 = bartlett, 2 = parzen, 3 = Tukey-Hanning, 4 = quadratic spectral

prewhite = 0 if no prewhitening 1 if VAR(1) prewhitening as in Andrews and Monahan (1992)

Tim Vogelsang 12/15/97 corrected 5/20/98 */

\[
\text{proc}(1) = \text{sig2npm}(v, \text{ST}, \text{kernel}, \text{prewhite});
\]

local k, T, vlag, x, y, vhat, A, Als, B, C, D, va, del, dells, ii1, ii2, jhat, i, j;
local ar1, denom, num1, num2, sear1, ehat, ahat1, ahat2;
local ar1, denom, num1, num2, sear1, ehat, ahat1, ahat2;

T = rows(v); k = cols(v);

vhat = v; jhat = zeros(k, k);

if prewhite == 1;
  x = v[1:T-1, .];
  y = v[2:T, .];
  Als = invpd(x' * x);
  A = Als * Als';
  (va, B) = eigh(B * Als');
  (va, C) = eigh(B * Als');
  dells = B * Als * C;
  ii1 = (dells > 0.97); ii2 = (dells < -0.97);
  del = (1 - ii1) .* (1 - ii2) .* dells + 0.97 * ii1 - 0.97 * ii2;
  A = B * del * C';
  D = inv(eye(k) - A);
  y = x * A';
  T = rows(vhat);
endif;

/* now compute truncation lag */

if ST < 0;
  ar1 = zeros(k, 1); num1 = zeros(k, 1); denom = zeros(k, 1);
  sear1 = zeros(k, 1); num2 = zeros(k, 1);

  i = 1;
  do while i < k;
    ar1[i] = vhat[2:T, i] / vhat[1:T-1, i];
    ehat = vhat[2:T, i] - ar1[i] * vhat[1:T-1, i];
    sear1[i] = ehat' * ehat / (T - 1);
    num1[i] = (4.0 * (ar1[i] * sear1[i])^2);
    num2[i] = (4.0 * (ar1[i] * sear1[i])^2) / ((1.0 - ar1[i])^8);
    denom[i] = (sear1[i]^2) / ((1.0 - ar1[i])^4);
  endwhile;
endif;
i=i+1;
end;

ahat1=sumc(num1)/sumc(denom);
ahat2=sumc(num2)/sumc(denom);

if kernel == 1; ST=1.1447*(T*ahat1)^(1.0/3.0); endif;
if kernel == 2; ST=2.6614*(T*ahat2)^(0.2); endif;
if kernel == 3; ST=1.7462*(T*ahat2)^(0.2); endif;
if kernel == 4; ST=1.3221*(T*ahat2)^(0.2); endif;
endif;

/* now compute the spectral density matrix element by element */
i=1;
do while i<=k;
  j=i;
do while j<=k;
    jhat[i,j]=sig2npv(vhat[.,i],vhat[.,j],ST,kernel);
    jhat[j,i]=jhat[i,j];
    j=j+1;
  endo;
  i=i+1;
endo;

if prewhite==1; jhat=D*jhat*D'; endif;
retp(jhat);
endp;

/* This procedure computes the nonparametric estimator of the long run variance for a vector of time series but only for an element of that matrix. In particular, this procedure computes the ijth element of the matrix. This procedure computes the estimator in the univariate case by simply setting v1=v2. Note that the truncation lag must be given to this procedure.

Inputs are:
v1 = ith time series
v2 = jth time series
ST = the truncation lag using the notation of Andrews (1991)

kernel = an integer representing the kernel that is used
1 = bartlett, 2 = parzen, 3 = Tukey-Hanning, 4 = quadratic spectral

Tim Vogelsang 12/6/97 */
proc(1) = sig2npv(v1,v2,ST,kernel);

local i,R,R0,T,sigma2;

if rows(v1)/=rows(v2);
    print "Error in sig2npv(), the vectors are not same size!";
endif;

T=rows(v1);
R=zeros(T-1,1);

if kernel<0 or kernel>4;
    kernel=4;
    print "using the Quadratic spectral kernel";
endif;

R0=v1'v2;

i=1;
do while i<=T-1;
    R[i]=v1[i+1:T]'v2[i+1:T-1] + v2[i+1:T]'v1[i+1:T];
    i=i+1;
endo;
sigma2=(R0+K(seqa(1,1,T-1)/ST,kernel)'R)/T;

retp(sigma2);
endp;

/* Kernels */

proc(1) = _K(x,kernel);

local y,xx,dd;
xx=((x>0).*x)+(abs(x)<0.0000001);

if kernel=1; /*bartlett*/
    y=(x.<1.0).*((1.0-x));
endif;

if kernel=2; /*parzen*/
    y=(x.>0.0).*((x.<0.5).*((1.0-6.0*x).*x.+(1.0-x)));
    y=y+(x.>0.5).*((x.<1.0)*2.0.*((1.0-x)).^3);
endif;

if kernel=3; /*Tukey-hanning*/
    y=(x.<1.0)*0.5.*((1.0+cos(pi*x)));
endif;

if kernel=4; /*quadratic spectral */
    dd=(25.0./(12.0*pi^2*xx.*xx)) {cos(1.2*pi*xx) - cos(1.2*pi*xx)}/
    y=(x.>0).*dd;
    y=y+(abs(x)<0.00000001);
if kernel==5; /* Daniells */
    y=((x>0).*sin(pi*xx)./(pi*xx))+(abs(x)<0.00000001);
endif;
retp(y);
endp;