Economics Letters 244 (2024) 111992

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Trends in temperature data: Micro-foundations of their nature

Maria Dolores Gadea-Rivas [a], Jesús Gonzalo [b,*], Andrey Ramos [b]

[a] Department of Applied Economics, University of Zaragoza, Spain
[b] Department of Economics, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

JEL classification:
C22
Q54

Keywords:
Trends
Unit roots
Structural breaks
Temperature
Aggregation

ABSTRACT

Determining whether Global Average Temperature (GAT) is an integrated process of order 1, I(1), or a stationary process around a trend function is crucial for detection, attribution, impact, and forecasting studies of climate change. In this paper, we investigate the nature of trends in GAT building on the analysis of individual temperature grids. Our micro-founded evidence suggests that GAT is stationary around a non-linear deterministic trend in the form of a linear function with one structural break. This break can be attributed to a combination of breaks on individual grids and the standard aggregation method under acceleration in global warming.

1. Introduction

Global Average Temperature (GAT) observed since the late 1800s exhibits an upward trend widely interpreted as evidence of global warming (GW) (Mann et al., 1998; Gadea and Gonzalo, 2020; AR6-IPCC, 2021). The specific nature of the trend is an open question in the empirical literature with a non-trivial answer. Neither theoretical climate nor economic models with climatic variables help to identify a particular trend specification.

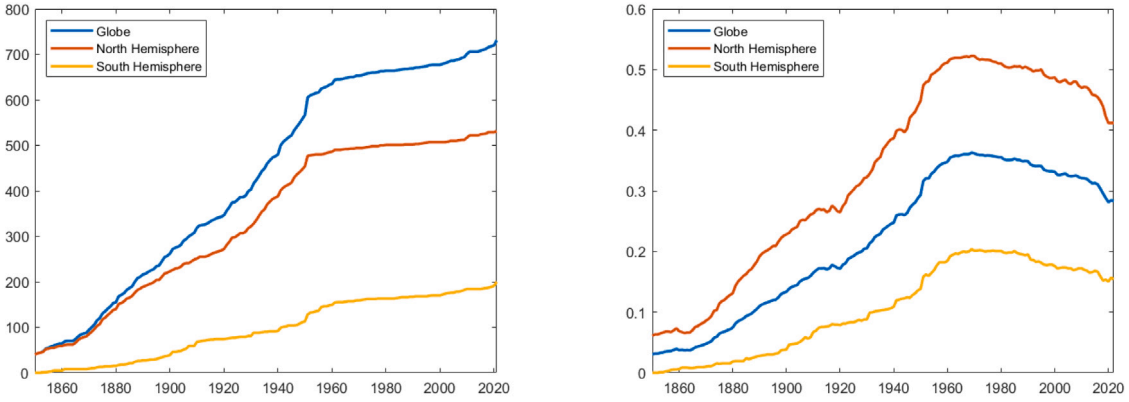
From the quantitative climate perspective, two dominant strands of literature debate. A first group of authors assume stochastic trends. Gordon (1991) and Yan and Wu (2010) use a random walk model to show that the trend in GAT can be attributed to random fluctuations rather than to specific physical drivers. Similarly, Woodward and Gray (1995) concludes that GAT has a unit root instead of a deterministic trend and suggests that temperature forecasts will not predict the observed trend to continue. The literature using cointegration analysis (Kaufmann et al., 2006, 2010; Dergiades et al., 2016; Bruns et al., 2020; Chang et al., 2020; Pretis, 2020) also assumes stochastic trends, but in contrast, their attribution argument is that the I(1) nature of GAT is inherited from its association with the anthropogenic forcing from CO2 and other greenhouse gases.[1] Other recent references using I(1) models for GAT and climate drivers include Turasie and Coelho (2016), Reid (2017), and Cummins et al. (2022). A second group of authors like Seidel and Lanzante (2004), Mudelsee (2019) and Gay-Garcia et al. (2009), or Estrada and Perron (2017), focus more on the statistical properties of the GAT series and conclude that the process follows a clear deterministic trend with possible structural breaks. The assumption of deterministic trends is also implicit in the statistical procedures implemented by the Intergovernmental Panel on Climate Change (IPCC) on its Sixth Assessment (AR6-IPCC, 2021) and previous reports. A mixture of both models for GAT is proposed in McKitrick et al. (2023) who conclude that if there is an I(1) component it must be very small.

Determining whether GAT is an I(1) process or a stationary process around a non-linear trend is crucial for detection, attribution, and impact studies of climate change (see McKitrick et al. 2023). If GAT is assumed to be I(1), attribution of GW using cointegration models involves demonstrating that the stochastic trends in temperature and the radiative forcing from CO2 and other greenhouse gases are common. However, the I(1) assumption also implies that exogenous temporary shocks like solar flares or volcanic eruptions generate long-lasting effects on temperature, which does not seem to be the case in the observed record. If trends are detected to be deterministic, detection of GW is possible through traditional trend-tests. But this assumption poses challenges for attribution and impact studies due to the problem of 'unbalanced' relations. In attribution studies, the mismatch in the order of integration between GAT and anthropogenic forcing hinders the estimation of the climate sensitivity (for a review, see Rohling et al. (2018)) using regression analysis. This is also the case for the impact studies that rely on dynamic growth equations linking the growth

* Corresponding author.
E-mail addresses: lgadea@unizar.es (M.D. Gadea-Rivas), jesus.gonzalo@uc3m.es (J. Gonzalo), anramosr@eco.uc3m.es (A. Ramos).

[1] Bennedsen et al. (2023) studies the statistical properties of radiative forcing from different components (including CO2) and find that the series follows an I(1), close to an I(2) process. Stochastic trends in radiative forcing are imparted by economic activity and atmospheric lifetimes.

https://doi.org/10.1016/j.econlet.2024.111992
Received 1 December 2023; Received in revised form 16 September 2024; Accepted 20 September 2024
Available online 26 September 2024



(a) Number of grids observed continuously (b) Proportion of non-missing grids

Fig. 1. Observation process for gridded temperatures.

rate of per-capita output with temperatures (see Dell et al., 2012 and references therein). Beyond this literature, understanding the nature of the trends is helpful for producing more accurate long-term forecasts as stressed by Kaufmann et al. (2010).

The existing evidence regarding the nature of trends in GAT relies on the analysis of aggregated series. This paper contributes to the debate by offering a micro-founded explanation for the observed trends based on the study of the trend dynamics of individual units used in computing those averages. Consistent with the hypothesis of Seidel and Lanzante (2004) and Gay-Garcia et al. (2009), our findings suggest that GAT is trend-stationary with a structural break in the trend function. Unit root tests implemented on individual units provide evidence in the same direction. We also discuss the effects of the aggregation method on generating spurious structural breaks or accentuating the existing breaks in the data. Concretely, our main hypothesis is that the break in GAT can be attributed to a combination of individual grid breaks and the standard aggregation method under warming acceleration. Depending on the strength of the break's signal, these situations may bias standard unit root tests towards the non-rejection zone. We illustrate this hypothesis through a set of Monte-Carlo simulations assuming linear and broken-trend individual processes and emulating the standard aggregation methods used to compute the GAT.

The rest of the paper is organized as follows. Section 2 provides empirical evidence of the nature of trends in aggregated and individual temperature series. Section 3 presents the simulation exercises. Finally, Section 4 concludes.

2. Empirical evidence

2.1. Data

Temperature data is obtained from the latest version of the HadCRUT5 dataset[2] jointly developed by the Climatic Research Unit (CRU) at the University of East Anglia and the Hadley Centre at the UK Met Office. For a more detailed information about the dataset, see Morice et al. (2021). In our empirical analysis we use the series of gridded temperature anomalies from the period 1961–1990 at a resolution of $5° \times 5°$.[3] A feature of the dataset is that the number of grids with non-missing data is relatively low during the early part of the record and gradually increases over time. Panel (a) of Fig. 1 presents the number of grids that are continuously observed from the given year onwards, while Panel (b) plots the proportion of non-missing grids (out of 2592) each year.

[2] Accessible at https://crudata.uea.ac.uk/cru/data/temperature/.

[3] Similar results are obtained if we use raw-stations data.

2.2. Computing average temperature

Two alternative methods to compute average temperatures are considered. Method A aggregates all grids with non-missing data each year. For each t, the average $\bar{T}_t^A$ is obtained as:

$$\bar{T}_t^A = \frac{1}{N_t} \sum_{i=1}^{N} I_{it} \times T_{it}, \quad t = 1, \dots, T, \tag{1}$$

where T_{it} is the temperature in grid i at year t, I_{it} is an indicator for T_{it} non-missing, and $N_t = \sum_{i=1}^{N} I_{it}$. Due to the non-uniform observation process for grids, N_t grows with t. Method A closely resembles the standard aggregation procedure adopted by CRU.

Method B, on the other hand, uses the set of grids with non-missing data throughout the entire sample period, ensuring a stable number of grids on the computation. For each t, $\bar{T}_t^B$ is obtained as:

$$\bar{T}_t^B = \frac{1}{|S|} \sum_{i \in S} T_{it}, \quad t = 1, \dots, T, \tag{2}$$

where $S = \{i : I_{it} = 1, \quad \forall t\}$ and $|S|$ is the cardinality of S. Method B is the approach adopted by Gadea and Gonzalo (2020) to estimate the mean and any other distributional characteristic of temperature.

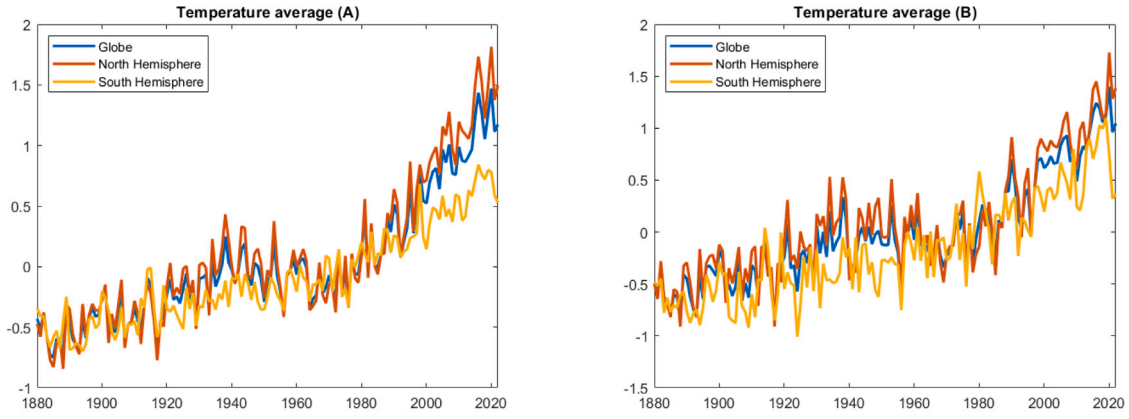
Separate averages are calculated for the Northern (NH) and Southern Hemisphere (SH) using data from 1880 to 2022.[4] Global temperature is obtained as a weighted average of both hemispheres, with the weights accounting for the difference in land areas. The estimated series under each method are presented in Fig. 2.

2.3. Trends in average temperature

The nature of the trends in average temperatures is studied. The test-statistics of the Augmented Dickey Fuller (ADF) tests (Dickey and Fuller, 1981) reported in Table 1 indicate that unit roots cannot be rejected for the global and NH averages obtained under method A.[5] For method B series, the unit root is rejected in all cases. A similar result is obtained if, instead, the more efficient (Elliott et al., 1996) (ERS) test is implemented. In addition, we implement the Kim and Perron (2009) (KP) test allowing for a break in the trend function under both the null and alternative hypothesis. In this case, the unit root is rejected in all

[4] Only 155 grids are observed over the full sample period, with the majority of them located in the NH.

[5] This is the main evidence for Kaufmann et al. (2010) and related literature to suggest stochastic trends in temperature.



(a) Method A (b) Method B

Fig. 2. Temperature averages (1880–2022).

series including the method A global and NH averages.[6] The Perron and Yabu (2009) (PY) procedure to estimate deterministic trends with either integrated or stationary noise components detect structural breaks in all aggregated series.[7] The breaks in the trend occurs around 1964–65 and is consistent with the "onset of a sustained global warming" also described in Estrada et al. (2013), a period characterized by a large increase in the rate of growth of temperatures and radiative forcing as a consequence of the post-World-War-II economic expansion and the consequent acceleration in the emissions of greenhouse gases.[8]

Our aggregated evidence aligns with the literature defending deterministic trends in temperature. Accounting for structural breaks, specially in method A aggregates, is crucial for the conclusions about unit roots. In method B aggregates, the signal of the break is not strong enough to drive the ADF test towards the non-rejection area.

2.4. Trends in individual grids

Unit root tests implemented on individual grids strongly reject the presence of unit roots. Results reported in Table 2 indicate that the proportion of rejections in the ADF test is high across the three sample periods considered. This proportion increases further when implementing the KP test.[9]

Assuming that individual trends are deterministic, we first model each individual grid through a linear-trend model:

$$T_{it} = \beta_{0i} + \beta_{1i} t + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{3}$$

where $e_{it} = \rho_i e_{it-1} + \upsilon_{it}$, $|\rho_i| < 1$, and υ_{it} are uncorrelated. For each sample period, we estimate Eq. (3) and generate density plots of $\hat{\beta}_1$. Panel (a) in Fig. 3 shows that the density shifts to the right in more recent samples, reflecting the well-known acceleration in the GW process (Mann et al., 1998; AR6-IPCC, 2021; Hansen et al., 2023).

Table 1
Test-statistics of the unit root tests and structural break detection analysis in average temperatures (1880–2022).

Temperature series	ADF test	KP test	PY-SB
Aggregation method A			
Global	−1.619	−5.514***	YES
Northern Hemisphere	−0.863	−4.323***	YES
Southern Hemisphere	−6.972***	−9.189***	YES
Aggregation method B			
Global	−4.336***	−7.569***	YES
Northern Hemisphere	−4.413***	−7.680***	YES
Southern Hemisphere	−8.5615***	−9.739***	YES

Notes: The table contains the test-statistics of the ADF and KP unit root tests implemented on each average temperature series. ADF-test equation includes an intercept and a linear trend. Number of lags selected based on the BIC. The KP tests allows for one break in the intercept and the linear slope under both the null and alternative hypotheses. Critical values at the 1% of significance are −4.028 for the ADF-test and −3.445 for the KP-test. ***, **, and * denotes rejection of the null hypothesis at 1, 5 and 10% level of significance respectively. Column PY-SB reports the conclusion about the existence of structural breaks in the Perron and Yabu (2009) (PY) procedure.

Table 2
Proportion of rejections in the unit root tests and structural break detection analysis on individual grids.

Sample	ADF test	KP test	PY-SB
1880–2022	92.90%	97.42%	81.94%
1920–2022	88.15%	98.84%	67.05%
1960–2022	90.87%	97.01%	25.98%

Notes: The table reports the proportion of units for which the null of unit root is rejected in the ADF and KP tests. The number of grids are 155, 346, and 635 for the sample periods 1880-2022, 1920–2022, and 1960–2022, respectively. ADF-test equation includes an intercept and a linear trend. Number of lags selected based on the BIC. The KP tests allows for one break in the intercept and the linear slope under both the null and alternative hypotheses. Tests at 1% of significance. Column PY-SB reports the proportion of grids with structural breaks detected by the Perron and Yabu (2009) (PY) procedure.

Focusing solely on the estimated slopes for the sample period 1960–2022, the mean of the coefficient is higher for the set of grids that appear later in the record.

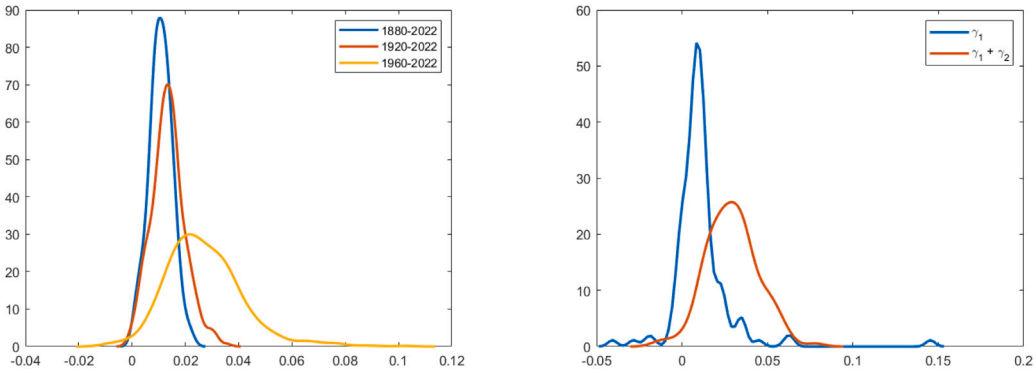
Acceleration in GW, combined with the standard aggregation method A, can trivially generate the breaks in aggregated series even when the individual units follow a linear model. Consider the following extreme case. During the initial part of the record, a set of N_1 units with an average trend-slope $\bar{\beta}_1$ are observed. Assume that at a certain period, T^*, a different set of N_2 units start to be observed with average slope $\bar{\beta}_2$, with $\bar{\beta}_2 > \bar{\beta}_1$. The slope of the average series is $\bar{\beta}_1$ before T^*, and

[6] Following the suggestion of a Referee, we implemented the unit root test in Otto (2021) allowing for a general slowly-varying deterministic trend component. The results for the aggregated series align with the KP tests.

[7] This result is robust even when we allow for polynomial trends. For instance, the Vogelsang (1997) test assuming linear and quadratic trends suggests a structural break that occurs in the linear term for both method A and method B aggregates.

[8] We study the possibility of a second break in the aggregated series following the sequential procedure of Kejriwal and Perron (2010). No evidence of a second break in either of the aggregated series is obtained.

[9] The null of unit root is rejected in 100% of the cases with the Otto (2021) test.



(a) Density of the slopes in the linear trend model (b) Density of the slopes in the broken-trend model

Fig. 3. Density of individual estimates.

$(N_1\bar{\beta}_1 + N_2\bar{\beta}_2)/(N_1 + N_2)$ after. For the average computed using Method B the slope is $\bar{\beta}_1$ the full period. Depending on the weights, the ADF test can be biased towards non-rejection.

Alternatively, we explore trend-breaks within individual grids:

$$T_{it} = \alpha_{0i} + \alpha_{1i}D_{ui} + \gamma_{1i}t + \gamma_{2i}D_{ti} + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \quad (4)$$

where $D_{ui} = 1\{t > TB_i\}$, $D_{ti} = 1\{t > TB_i\} \times (t - TB_i)$, TB_i is the period in which the structural break occurs for unit i, and e_{it} follows the same structure as before. This corresponds to model A3 in Kim and Perron (2009) where γ_{1i} is the trend-slope before the break date, and $\gamma_{1i} + \gamma_{2i}$ is the trend-slope after the break. According to the PY results in Table 2, out of the 155 grids observed continuously from 1880 to 2022, approximately 82% contain a break in the linear trend, with this break occurring around 1965 as in the aggregated series. The densities of $\hat{\gamma}_1$ and $\hat{\gamma}_1 + \hat{\gamma}_2$ plotted in Panel (b) of Fig. 3 provide micro-level evidence of the warming acceleration phenomenon.

Aggregating individual grids with broken trends using either method A or B results in average series with structural breaks, as those in Fig. 2. Moreover, the aggregation method A under acceleration in GW can deepen the break signal. Imagine that N_1 grids are observed all periods, with average parameters $\bar{\gamma}_1^1$ and $\bar{\gamma}_2^1$ as defined in Eq. (4), with $TB_i = T^*$ being the period of the individual break. Assume that at T^* a different set of N_2 units start to be observed with average parameters $\bar{\gamma}_1^2$ and $\bar{\gamma}_2^2$, with $\bar{\gamma}_1^2 + \bar{\gamma}_2^2 > \bar{\gamma}_1^1 + \bar{\gamma}_2^1$ to allow for warming acceleration. Similar to the linear case, the slope of the method A average series is $\bar{\gamma}_1^1$ before T^*, and $[N_1(\bar{\gamma}_1^1 + \bar{\gamma}_2^1) + N_2(\bar{\gamma}_1^2 + \bar{\gamma}_2^2)]/(N_1 + N_2)$ after. This slope is higher than the method B average, which is $\bar{\gamma}_1^1 + \bar{\gamma}_2^1$ after T^*.

In summary, our individual evidence rejects the existence of unit roots in individual grids. Structural breaks in aggregated series are originated from individual breaks. The aggregation method A under acceleration in GW can deepen the magnitude of such breaks. Due to the small number of grids involved in the computation of method B averages, the signal of the break is not strong enough to wrongly drive the ADF test towards the conclusion of unit roots. For method A averages, the non-rejection of the ADF test can be attributed to both a higher signal, stemming from the utilization of more information each period, and the deepening of the break due to the acceleration in GW that implies the inclusion of more grids in periods of higher trend. These hypotheses are further explored with simulations in the next section.

3. Simulations

The proposed 'micro-founded' explanations for the nature of the trends in aggregate temperatures are validated heuristically using simulations. We consider two alternatives to simulate the non-missing indicator, I_{it}:

- Alternative 1: A fixed number of units are non-missing during the whole sample. A second group of series are missing during the first T^* periods, and non-missing from T^* on. T^* is set to 100 in the simulations.
- Alternative 2: I_{it} follows a Markov Switching (MS) process with two states (missing and non-missing) and transition-probability matrix **P**.

First, we analyze the case where individual units follow a linear-trend model. Two groups of series are simulated:

$$\text{Group 1:} \quad T_{it}^1 = \beta_{0i}^1 + \beta_{1i}^1 t + e_{it}^1, \quad i = 1, \ldots, N_1, \quad t = 1, \ldots, T, \quad (5)$$

$$\text{Group 2:} \quad T_{it}^2 = \beta_{0i}^2 + \beta_{1i}^2 t + e_{it}^2, \quad i = 1, \ldots, N_2, \quad t = 1, \ldots, T, \quad (6)$$

where $\bar{\beta}_1^2 > \bar{\beta}_1^1$ and $e_{it}^j = \rho_i^j e_{it-1}^j + v_{it}^j$, $j = 1, 2$. Observation alternatives 1 and 2 assume that stations entering later in the average computation are chosen randomly from groups 1 or 2. To account for warming acceleration, we define observation alternatives 1* and 2* assuming that the series of group 2 (i.e. those with higher trend-slopes) appear as non-missing later in the record.

The simulation parameters are set based on empirical evidence, aiming to match simulated and observed trends in the aggregated series. Specifically, we construct the empirical distribution of the estimates of a linear model fitted to the individual gridded data over two different samples: 1880–2021 for group 1 and 1960–2021 for group 2. Individual series are simulated using the following calibration: $\beta_{0i}^1 \sim \mathcal{N}(-0.7234, 0.4266^2)$, $\beta_{1i}^1 \sim \mathcal{N}(0.0108, 0.0043^2)$, $\beta_{0i}^2 = -2.36$, $\beta_{1i}^2 \sim \mathcal{N}(0.0271, 0.0145^2)$. β_{0i}^2 is a fixed parameter defined to avoid abrupt changes in the level of the aggregated series. The value of -2.36 is determined by equating the average level of the group 1 and group 2 series after $T^* = 100$ simulation periods, when the second group of series begins to be observed under alternatives 1 and 1*.

Autocorrelation is accounted for by defining $\rho_i^1 \sim \mathcal{N}(0.264, 0.1176^2)$, $\rho_i^2 \sim \mathcal{N}(0.155, 0.1747^2)$, and $v_{it}^j \sim \mathcal{N}(0, 2.9046^2)$, $j = 1, 2$. These autocorrelation parameters are obtained from the empirical distribution of the estimates of an AR(1) model fitted on the residuals of each individual regression. The value of 2.9046 for the standard deviation of the error term is calibrated to match the residual variance of the observed aggregated series with that of the simulated series. Specifically, the standard deviation of the residuals after fitting a broken linear-trend model to the method B average is 0.2333. For an aggregated series calculated with 155 grids, this corresponds to an individual standard deviation of 2.9046.

To specify the transition probability matrix **P**, the number of transitions between missing and non-missing states in the data were counted

Table 3
Proportion of non-rejections in the unit root tests and structural break detection analysis on simulated averages (linear-trend model).

Alternative	Aggregation method A			Aggregation method B		
	ADF-test	KP-test	PY-SB	ADF-test	KP-test	PY-SB
1	2.20%	0.20%	21.90%	0.10%	0.10%	3.60%
1*	70.60%	3.80%	99.90%	0.00%	0.00%	3.70%
2	1.30%	0.00%	11.30%	0.00%	0.00%	4.60%
2*	59.90%	5.40%	99.90%	0.00%	0.00%	4.20%

Notes: The table reports the proportion of times that the null of unit root is non-rejected for the ADF and KP tests. Tests at 1% of significance. The PY-SB column reports the proportion of times in which the aggregated series contain structural breaks according to the Perron and Yabu (2009) procedure. For alternatives 1 and 1*, we set $N_1 = 150$ and $N_2 = 850$. For alternatives 2 and 2*, the proportion of units observed from the initial period is set to 15%.

Table 4
Proportion of non-rejections in the unit root tests and structural break detection analysis on simulated averages (model with broken-trend).

Alternative	Aggregation method A			Aggregation method B		
	ADF-test	*KP-test*	*PY-SB*	*ADF-test*	*KP-test*	*PY-SB*
1	95.30%	4.40%	99.70%	55.30%	2.40%	100%
1*	93.50%	5.90%	100%	18.30%	0.70%	99.60%
2	99.10%	1.00%	99.90%	52.60%	1.60%	100%
2*	98.30%	1.00%	100%	20.00%	0.60%	99.80%

Notes: The table reports the proportion of times that the null of unit root is non-rejected for the ADF and KP tests. Tests at 1% of significance. The PY-SB column reports the proportion of times in which the aggregated series contain structural breaks according to the Perron and Yabu (2009) procedure. For alternatives 1 and 1*, we set $N_1 = 150$ and $N_2 = 850$. For alternatives 2 and 2*, the proportion of units observed from the initial period is set to 15%.

and the empirical conditional probabilities were computed. We obtained an initial matrix:

$$\mathbf{P}_0 = \begin{bmatrix} 0.9950 & 0.0050 \\ 0.0087 & 0.9913 \end{bmatrix}, \tag{7}$$

where the entrance (i, j) corresponds to the transition probability from state i to state j, and state 1 represents a missing value. We impose that after the first non-missing value, the series remain non-missing for the rest of the record. Therefore, the second row of P is replaced by $[0, 1]$, and the used P becomes:

$$\mathbf{P} = \begin{bmatrix} 0.9950 & 0.0050 \\ 0 & 1 \end{bmatrix}. \tag{8}$$

Sample size is $T = 150$. For observation alternatives 1 and 1*, the number of series in each group are set at $N_1 = 150$ and $N_2 = 850$; for observation alternatives 2 and 2*, we impose that 15% of units are observed from the initial period. In both cases, we aim to approximately reproduce the proportion of continuously observed units from the initial period, that is close to 15% in the North Hemisphere from 1880 onwards. Simulations are replicated $R = 1000$ times.

Table 3 presents the proportion of non-rejections of the ADF and KP tests for each observation alternative and aggregation method. Additionally, it presents the proportion of times in which the aggregated series contain a structural break according to the PY procedure. Consistent with our hypothesis, even though the series are generated without a unit root, the ADF test fails to reject the null hypothesis in 70.60% and 59.90% of cases for observation alternatives 1* and 2*, respectively, using aggregation method A. The KP test, which allows for structural breaks under the null and the alternative hypothesis, detects far fewer unit roots. The generated breaks detected by the PY procedure are entirely attributed to the aggregation process. For aggregation method B, the unit root is rejected in almost all cases using both the ADF and the KP tests. The proportion of series with breaks is small and can be attributed to sample variability.

The analysis of the linear model without breaks serves to demonstrate that under extreme conditions, where individual variables do not exhibit breaks or unit roots, the aggregation process itself can introduce spurious breaks that bias the ADF unit root tests if such breaks are not adequately accounted for.

Next, lets consider the case where individual units contain one break in the trend function. Similar to the previous case, we simulate two groups of series of the form:

$$\text{Group 1:} \quad T_{it}^1 = \alpha_{0i}^1 + \alpha_{1i}^1 D_{ui}^1 + \gamma_{1i}^1 t + \gamma_{2i}^1 D_{ti}^1 + e_{it}^1, \quad i = 1, \ldots, N_1, \quad t = 1, \ldots, T, \tag{9}$$

$$\text{Group 2:} \quad T_{it}^2 = \alpha_{0i}^2 + \alpha_{1i}^2 D_{ui}^1 + \gamma_{1i}^2 t + \gamma_{2i}^2 D_{ti}^2 + e_{it}^2, \quad i = 1, \ldots, N_2, \quad t = 1, \ldots, T, \tag{10}$$

where the variables and parameters are defined in Eq. (4), $e_{it}^j = \rho_i^j e_{it-1}^j + \upsilon_{it}^j$, $j = 1, 2$, and we impose $\bar{\gamma}_2^2 > \bar{\gamma}_2^1$ to the capture warming acceleration. As in the previous case, alternatives 1* and 2* assume that the series in group 2 start to count later for the average computation. Individual series are simulated using the following calibration: $\alpha_{0i}^1, \alpha_{0i}^2 \sim \mathcal{N}(-0.6061, 0.5642^2)$, $\alpha_{1i}^1, \alpha_{1i}^2 \sim \mathcal{N}(-0.5524, 0.6899^2)$, $\gamma_{1i}^1, \gamma_{1i}^2 \sim \mathcal{N}(0.0110, 0.0174^2)$, $\gamma_{2i}^1, \sim \mathcal{N}(0.0182, 0.0233^2)$, $\gamma_{2i}^2, \sim \mathcal{N}(0.0271, 0.0145^2)$, $\rho_i^j \sim \mathcal{N}(0.1101, 0.0899^2)$, and $\upsilon_{it}^j \sim \mathcal{N}(0, 2.9046^2)$, $j = 1, 2$. These values are based on the estimates of fitting a model to the individual grids allowing for a break in the trend. Other simulation parameters remain as before.

Table 4 presents the proportion of non-rejections of the ADF and KP tests, as well as the proportion of times in which the aggregated series contain structural breaks according to the PY procedure. Consistent with our hypothesis, when averages are computed using method A, the unit root using the ADF test is non-rejected in more than 90% of the cases under all observation alternatives. For alternatives 1 and 2, the breaks in aggregated series detected using the PY procedure are completely attributed to individual breaks, while for alternatives 1* and 2*, those breaks can be generated by a combination of both individual grid breaks and the aggregation process under warming acceleration.

For method B, even though the original series contain a break and it is inherited by the aggregated series (see PY-SB column), the signal of the break is not always strong enough to drive the ADF test towards the non-rejection zone. For alternatives 1* and 2*, that replicates closely the real observation process, the ADF test detects unit roots only 18.30% and 20.00% of the cases, respectively. If the number of N_1 units or the proportion of units that are observed from the beginning of the record is increased, the signal of the break eventually becomes stronger and the ADF test concludes the presence of unit roots more frequently. Notice that for both aggregation methods, the KP-test rejects unit roots in almost all replications, showcasing the importance of accounting for structural breaks under the null and the alternative hypothesis in the unit roots testing procedure.

Once the existence of a break in the trend is established, disentangling its source within the data is difficult, if not impossible, due to the inherent identification problem. In the simulations, an approach is to compare the trends in the method A aggregates before and after the break, between alternatives 1 and 1* (or 2 versus 2*).[10] However, in the data, we observe method A and method B averages under the true non-missing process. If we consider that the non-missing process is well represented by alternative 1* (or 2*), it is required for identification to assume that the observed method B average is a valid counterfactual for the unobserved method A under alternative 1 (or 2), for a meaningful comparison of trends. However, this assumption is very unlikely to

[10] Any break in alternative 1 aggregates is due to individual breaks, while breaks in alternative 1* are attributed to a combination of both individual breaks and the aggregation procedure.

hold in the data because both series are observed under the same observational regime. Moreover, the true observational regime can be more complex than alternative 1* (or 2*), with units entering and going out of the record at any period. In this sense, the purpose of the simulations is to emulate what is actually done in practice and to heuristically demonstrate the potential mechanisms causing breaks in the aggregated series.

4. Conclusions

Aggregate and micro-founded evidence do not support the hypothesis of stochastic trends in temperature. Our evidence suggests that temperatures averages are stationary around a non-linear trend, with the non-linearity being modeled as a one-time break in a linear trend function. The break can be attributed to a combination of individual grid breaks and the standard aggregation method under warming acceleration. The aggregation method is relevant to bias ADF-tests towards the non-rejection zone by amplifying the signal of the break. Our findings carry important empirical implications for studies on the detection, attribution, impact, and forecasting of GW. Radiative forcing from anthropogenic greenhouse gases (such as CO2, methane, and nitrous oxide) is modeled as an I(1), or even I(2) process. It introduces a problem of 'unbalanced' regressions if the aim is to establish human influence on GW using cointegration methods. A similar issue is present on impact studies relating economic growth and temperature data. Alternatives to study the relationship between temperature and greenhouse gases include the use of co-trending methods robust to either type of trends such as Chen et al. (2022), or to adequately transform the forcing series to achieve balanced regressions. Bennedsen et al. (2023), for instance, obtain that the CO2 concentrations series in the first differences can be modeled as a trend stationary process, therefore, a regression of temperature and ΔCO2 concentrations is a valid balanced regression for attribution.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge financial support from the Spanish Ministerio de Ciencia e Innovación through grants PID2019-104960GB-IO0, PID2020-114646RB-C42, TED2021- 129784B-IO0, AEl/10.13039/501100011033-CEX2021-001181-M.

References

AR6-IPCC, 2021. Climate Change 2021: The Physical Science Basis. Cambridge University Press., Cambridge and New York,

Bennedsen, M., Hillebrand., E., Koopman, S.J., 2023. A multivariate dynamic statistical model of the global carbon budget 1959–2020. J. R. Stat. Soc. Ser. A: Stat. Soc. 86 (1), 20–42.

Bruns, S.B., Csereklyei, Z., Stern, D.I., 2020. A multicointegration model of global climate change. J. Econometrics 214 (1), 175–197.

Chang, Y., Kaufmann, R.K., Kim, C.S., Miller, J.I., Park, J.Y., Park, S., 2020. Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. J. Econometrics 214 (1), 274–294.

Chen, L., Gao, J., Vahid, F., 2022. Global temperatures and greenhouse gases: A common features approach. J. Econometrics 230, 240–254.

Cummins, D.P., Stephenson, D.B., Stott, P.A., 2022. Could detection and attribution of climate change trends be spurious regression? Clim. Dyn. 59 (9), 2785–2799.

Dell, M., Jones, B.F., Olken, B.A., 2012. Temperature shocks and economic growth: Evidence from the last half century. Am. Econ. J.: Macroecon. 4 (3), 66–95.

Dergiades, T., Kaufmann, R.K., Panagiotidis, T., 2016. Long-run changes in radiative forcing and surface temperature: The effect of human activity over the last five centuries. J. Environ. Econ. Manage. 76, 67–85.

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49 (4), 1057–1072.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64 (4), 813–836.

Estrada, F., Perron, P., 2017. Extracting and analyzing the warming trend in global and hemispheric temperatures. J. Time Series Anal. 38, 711–732.

Estrada, F., Perron, P., Martínez-López, B., 2013. Statistically derived contributions of diverse human influences to twentieth-century temperature changes. Nat. Geosci. 6, 1050–1055.

Gadea, M.D., Gonzalo, J., 2020. Trends in distributional characteristics: Existence of global warming. J. Econometrics 214 (1), 153–174.

Gay-Garcia, C., Estrada, F., Sánchez, A., 2009. Global and hemispheric temperatures revisited. Clim. Change 94, 333–349.

Gordon, A.H., 1991. Global warming as a manifestation of a random walk. J. Clim. 4 (6), 589–597.

Hansen, J.E., Sato, M., Simons, L., Nazarenko, L.S., Sangha, I., Kharecha, P., Zachos, J.C., von Schuckmann, K., Loeb, N.G., Osman, M.B., Jin, Q., Tselioudis, G., Jeong, E., Lacis, A., Ruedy, R., Russell, G., Cao, J., Li, J., 2023. Global warming in the pipeline. Oxford Open Clim. Change 3 (1), 1–33.

Kaufmann, R.K., Kauppi, H., Stock, J.H., 2006. Emissions, concentrations, and temperature: A time series analysis. Clim. Change 77 (3), 249–278.

Kaufmann, R.K., Kauppi, H., Stock, J.H., 2010. Does temperature contain a stochastic trend? Evaluating conflicting statistical results. Clim. Change 101, 395–405.

Kejriwal, M., Perron, P., 2010. A sequential procedure to determine the number of breaks in trend with an integrated or stationary noise component. J. Time Series Anal. 31 (5), 305–328.

Kim, D., Perron, P., 2009. Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. J. Econometrics 148 (1), 1–13.

Mann, M.E., Bradley, R.S., Hughes, M.K., 1998. Global-scale temperature patterns and climate forcing over the past six centuries. Nature 29, 779–787.

McKitrick, R., Vogelsang, T., Christy, J., 2023. Temperature trends, climate attribution and the nonstationarity question. Earth Syst. Dyn. Discuss. 1–32, [preprint].

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E., Dunn, R.J.H., Osborn, T.J., Jones, P.D., Simpson, I.R., 2021. An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. J. Geophys. Res. 126 (3), 1–28.

Mudelsee, M., 2019. Trend analysis of climate time series: A review of methods. Earth-Sci. Rev. 190, 310–322.

Otto, S., 2021. Unit root testing with slowly varying trends. J. Time Series Anal. 42 (1), 85–106.

Perron, P., Yabu, T., 2009. Estimating deterministic trends with an integrated or stationary noise component. J. Econometrics 151 (1), 56–69.

Pretis, F., 2020. Econometric modelling of climate systems: The equivalence of energy balance models and cointegrated vector autoregressions. J. Econometrics 214 (1), 256–273.

Reid, J., 2017. There is no significant trend in global average temperature. Energy Environ. 28 (3), 302–315.

Rohling, E.J., Marino, G., Foster, G.L., Goodwin, P.A., von der Heydt, A.S., Köhler, P., 2018. Comparing climate sensitivity, past and present. Annu. Rev. Mar. Sci. 10, 261–288.

Seidel, D.J., Lanzante, J.R., 2004. An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. Clim. Dyn. 109 (D14108), 1–10.

Turasie, A.A., Coelho, C.A.S., 2016. Cointegration modelling for empirical south American seasonal temperature forecasts. Int. J. Climatol. 36 (14), 4523–4533.

Vogelsang, T.J., 1997. Wald-type tests for detecting breaks in the trend function of a dynamic time series. Econometric Theory 13 (6), 818–849.

Woodward, W.A., Gray, H.L., 1995. Selecting a model for detecting the presence of a trend. J. Clim. 8 (8), 1929–1937.

Yan, S., Wu, G., 2010. Application of random walk model to fit temperature in 46 Gamma world cities from 1901 to 1998. Nat. Sci. 2, 1425–1431.