



# Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate<sup>☆</sup>

Yoosoon Chang<sup>a</sup>, Robert K. Kaufmann<sup>b</sup>, Chang Sik Kim<sup>c</sup>, J. Isaac Miller<sup>d,\*</sup>,  
Joon Y. Park<sup>a,c</sup>, Sungkeun Park<sup>c,1</sup>

<sup>a</sup> Department of Economics, Indiana University, USA

<sup>b</sup> Department of Earth and Environment, Boston University, USA

<sup>c</sup> Department of Economics, Sungkyunkwan University, South Korea

<sup>d</sup> Department of Economics, University of Missouri, USA

## ARTICLE INFO

### Article history:

Available online 23 May 2019

### JEL classification:

C14  
C23  
C33  
Q54

### Keywords:

Climate change  
Temperature distribution  
Global temperature trends  
Functional unit roots

## ABSTRACT

We analyze a time series of global temperature anomaly distributions to identify and estimate persistent features in climate change. We employ a formal test for the existence of functional unit roots in the time series of these densities, and we develop a new test to distinguish functional unit roots from functional deterministic trends or explosive behavior. Results suggest that temperature anomalies contain stochastic trends (as opposed to deterministic trends or explosive roots), two trends are present in the Northern Hemisphere while one stochastic trend is present in the Southern Hemisphere, and the probabilities of observing moderately positive anomalies have increased. We postulate that differences in the pattern and number of unit roots in each hemisphere may be due to a natural experiment which causes human emissions of greenhouse gases and sulfur to be greater in the Northern Hemisphere, decreasing the mean temperature anomaly but increasing the spatial variance relative to the Southern Hemisphere. Together, these results are consistent with the theory of anthropogenic climate change.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Though it may seem like a statistical minutia, the question of whether the time series of historical temperatures contain a stochastic and/or deterministic trend is important. The properties of this time series are critical for both the detection and attribution of climate change. Identifying the presence of such trends is a key step in testing hypotheses about the physical principles that are postulated to drive climate change. Moreover, these time series properties affect the statistical techniques that are appropriate for analyzing the observational record and simulation results.

<sup>☆</sup> The authors are grateful for useful comments from William A. “Buz” Brock, David Hendry, Jim Stock, participants of the 2016 Conference on Econometric Models of Climate Change, 2016 SNDE Symposium, 2015 INET-Cambridge Workshop, 2014 NBER-NSF Conference, and 2014 SETA, and seminar attendees at University of Carlos III, University of Pompeu Fabra, CEMFI, Carleton University, Korea University, Seoul National University, Vienna University of Economics and Business, CORE, USC, University of Notre Dame, Hitotsubashi University, Maastricht University, and University of Missouri. An earlier version of this paper was circulated under the title “Time Series Analysis of Global Temperature Distributions: Identifying and Estimating Persistent Features in Temperature Anomalies.”

\* Correspondence to: Department of Economics, University of Missouri, 118 Professional Building, Columbia, MO 65211, USA.

E-mail address: [millerjisaac@missouri.edu](mailto:millerjisaac@missouri.edu) (J.I. Miller).

<sup>1</sup> Present address: Korea Institute for Industrial Economics and Trade, South Korea.

First principles suggest that the highly persistent movements in the radiative forcing of greenhouse gases and sulfur emissions are caused by the long-lived nature of capital stock, which emits these gases and the relatively long residence time of well-mixed greenhouse gases, which allows the atmosphere to integrate emissions into concentrations (Kaufmann et al., 2013). The transmission of this persistence in radiative forcing to temperature is consistent with the basic physics that are embodied in climate models. A zero dimensional energy balance model can be rewritten in the form of an error correction model typically used to analyze relations among time series with common stochastic trends (Kaufmann et al., 2013). Hence, the statistical identification of and distinction between stochastic and deterministic trends lies at the heart of efforts to test the physical mechanisms hypothesized to drive climate change.

Given the underlying importance of the time series properties, a significant literature focuses on detecting a trend in temperature and distinguishing a linear trend from lower-order unit root-type persistence (i.e. a stochastic trend). To date, the evidence is mixed. Many studies generate results that are consistent with the presence of a stochastic trend (Gordon, 1991; Woodward and Gray, 1993, 1995; Gordon et al., 1996; Kärner, 1996). Conversely, many other studies generate results that are consistent with the presence of a deterministic trend with possibly highly persistent noise (Bloomfield, 1992; Bloomfield and Nychka, 1992; Baillie and Chung, 2002; Fomby and Vogelsang, 2002).

Although Bloomfield (1992) tests for a linear trend, he emphasizes the importance of using a model-based nonlinear deterministic component. Accordingly, the notion of a deterministic trend also includes nonlinearities in the form of a quadratic trend (Woodward and Gray, 1995; Zheng and Basher, 1999), an exponential trend (Zheng et al., 1997), and breaks in an otherwise linear trend (Zheng et al., 1997; Zheng and Basher, 1999; Gay-Garcia et al., 2009; Estrada et al., 2010, 2013; Estrada and Perron, 2012, 2014; McKittrick and Vogelsang, 2014). Nonlinearity also is investigated by estimating a general deterministic trend nonparametrically (Gao and Hawthorne, 2006). Their results suggest that the estimated trend contains high degrees of nonlinearity and variability, which can be approximated by a stochastic trend.

Beyond tests on individual time series, the presence of stochastic trends is examined by testing whether temperature cointegrates with radiative forcing. If these variables cointegrate, the shared stochastic trend would be consistent with the hypothesis that economic activity and atmospheric lifetimes impart a stochastic trend to radiative forcing and this trend is communicated to temperature. Many studies find evidence of this cointegration (Kaufmann and Stern, 2002; Kaufmann et al., 2006, 2011; Mills, 2009; Dergiades et al., 2016).<sup>2</sup>

Previous studies analyze means but often ignore the distributions of temperatures (Ballester et al., 2010; Donat and Alexander, 2012). Brock et al. (2013) underscore the importance of spatial heterogeneity – temperature anomalies increase with latitude (Hansen et al., 2010). Zheng and Basher (1999) argue that stronger variability in high latitudes of the Northern Hemisphere make it difficult to detect a trend in local temperature anomalies.

The effects of heterogeneity can be better understood by considering higher-order moments of the spatial distribution of the anomalies. New techniques make it possible to evaluate whether cross-sectional distributions, such as distributions of global temperature anomalies, are stationary or nonstationary (Bosq, 2000; Park and Qian, 2012; Hormann et al., 2015; Chang et al., 2016; Beare et al., 2017, *inter alia*). Using these tools, analysts can evaluate the persistence in the mean and the higher-order moments of the global distributions of temperature anomalies.

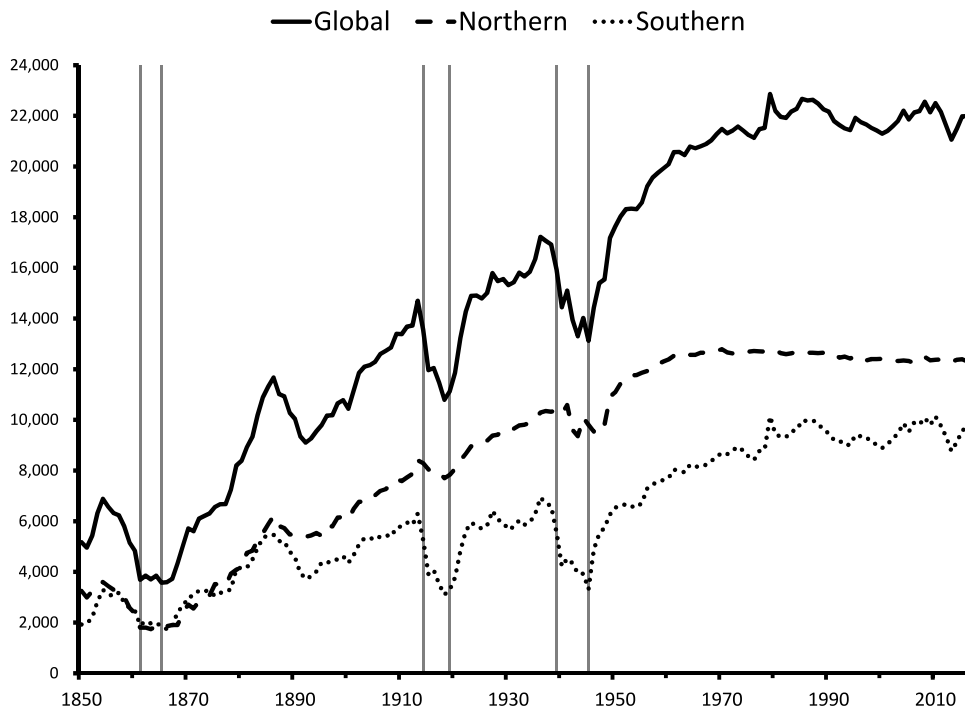
Building on these capabilities, we extend the work of Chang et al. (2016) to distinguish between persistence that is induced by unit root-type nonstationarity (a stochastic trend) from that induced by a deterministic trend or an explosive root in the distributions of temperature anomalies (global, Northern Hemisphere and Southern Hemisphere) during the instrumental record (1850–2016). Further, we analyze the test proposed by Chang et al. (2016) when such higher-order nonstationarity is present under the null.

Our approach may be compared to recent spatiotemporal approaches to climate modeling, such as that by Castruccio and Stein (2013). Typically, these parametric models are used to emulate output from general circulation models or assess policy. They are carefully constructed to allow for deviations from stationarity both temporally (Castruccio et al., 2014; Leeds et al., 2015) and spatially (Jun and Stein, 2008). Our approach is designed to detect (and distinguish between types of) temporal nonstationarity. Although we do not explicitly model spatial nonstationarity, the approach features a nonparametric kernel density estimator of spatially distributed temperature anomaly data. The kernel estimator is known to be quite robust even when the data are mildly nonstationary across space (Hallin et al., 2004).

Our results identify substantial nonstationarity in the moments of the distributions – primarily in the mean (i.e., global warming) and in the second moment. We postulate that a natural experiment, in which anthropogenic forcings differ between hemispheres, generates hemispheric differences in the pattern and number of nonstationary coordinate processes. These results suggest that stochastic trends in the spatial distributions of temperature anomalies are metaphorical fingerprints left by anthropogenic emissions of radiatively active gases that vary over space and time due to human activities.

Conversely, none of the nonstationarity that we detect is more persistent than that of a stochastic trend. Such evidence casts doubt on a deterministic trend in temperature anomalies, which would imply that changes in the moments – in particular, an increasing mean – are inevitable. As such, these results are inconsistent with the hypotheses that temperature can be modeled using a deterministic trend or that warming is being accelerated by a so-called runaway greenhouse effect.

<sup>2</sup> These results are disputed by those who argue that cointegration is a statistical artifact of a broken deterministic trend (Gay-Garcia et al., 2009; Estrada et al., 2013; *inter alia*). However, Kaufmann et al. (2010) argue that the appearance of a broken deterministic temperature trend is inherited from the forcing variables, which may suggest a break in the 1980s due to legislation limiting acid deposition.



**Fig. 1.** Number of annual temperature observations. Observations for the globe, NH, and SH, based on  $5^\circ$  by  $5^\circ$  grid boxes. Total possible annual observations for the globe is  $36 \times 72 \times 12 = 31,104$  ( $36 \times 5^\circ$  along each meridian,  $72 \times 5^\circ$  around the Equator, 12 months per year) and 15,552 for each hemisphere. American Civil War (1861–65), World War I (1914–19), and World War II (1939–45) indicated.

Our results and the methods used to obtain them are described in the following three sections. In Section 2, we introduce the global temperature anomaly data, and we discuss the time series framework for analyzing state distributions, testing procedures for nonstationarity of those distributions, and estimation of the proportions of variations in cross-sectional moments attributable to nonstationarity. A finite-sample analysis tailored to temperature anomaly data shows good size and power properties of the tests. We discuss step-by-step implementation of the tests and present our empirical results in Section 3, and we discuss these results in the context of the extant literature in Section 4. Section 5 concludes, and an appendix contains proofs of the theoretical results.

## 2. Data and methodology

### 2.1. Global temperature distributions

We use the HadCRUT4 data set (Morice et al., 2012), which combines marine temperature data compiled by the Met Office Hadley Centre with land temperature data compiled by the Climatic Research Unit of the University of East Anglia. These data are expressed in degrees Celsius over 1850 to 2016 as anomalies from the monthly average over the base period 1961–90. Specifically, deviations are calculated for each land station, and then the deviations are averaged across all stations in a given grid box that is  $5^\circ$  latitude by  $5^\circ$  longitude. For marine data, the measurements are taken from ships or buoys and the anomaly is calculated based on the monthly average over 1961–90 for each grid box.

The maximum number of observations of temperature anomalies in each month is 2592, which is the product of 36 increments of  $5^\circ$  latitude and 72 increments of  $5^\circ$  longitude. We create an annual distribution of temperature anomaly observations from the monthly HadCRUT4 data, providing a maximum number of  $2592 \times 12 = 31,104$  annual observations. Observations per year, shown in Fig. 1, generally increase from about 5000 at the beginning of the sample to about 22,000 by the late 1970s and through the end of the sample. Three noticeable dips in the sample size correspond roughly with the American Civil War (1861–65), World War I (1914–19), and World War II (1939–45).

Hemispheric means often are analyzed separately in studies on climate change, because more land in the Northern Hemisphere (NH) translates into more error from station and other types of biases, but less land in the Southern Hemisphere (SH) translates into more small-sample and coverage errors from fewer non-missing grid box observations. Global means are estimated by averaging the hemispheric means. (See Brohan et al., 2006)

Working with densities requires a more complicated averaging strategy. We obtain the temperature distributions from the monthly temperature anomaly data pooled over each year in the NH and SH. We estimate the densities of temperature anomalies for the NH and SH separately. Then, for each year and at each temperature, we average the estimated NH and SH density functions to obtain an estimate of the global density function at that temperature. The global density is described by the density function over a compact subset of temperatures. Each hemisphere receives an equal weight to avoid giving too much weight to the NH, where there are more non-empty grid boxes. In order to make the supports of these densities compact, we omit approximately 1% extreme outliers.<sup>3</sup> Specifically, we set the supports  $[-5.14, 5.18]$ ,  $[-6.24, 6.26]$ , and  $[-3.24, 3.07]$  for the global, NH, and SH distributions, respectively. We utilize the typical nonparametric density estimator with the Epanechnikov kernel and Silverman bandwidth.

The estimated densities are regarded as the data that we subsequently analyze. We expect that estimation errors in the temperature anomaly densities have a negligible effect on our analysis, because the number of cross-sectional observations each year is very large relative to the number of years. The estimation errors decrease with the cross-sectional dimension, but they are expected to accumulate as the time dimension increases.

Specifically, the rate of consistency of the kernel density estimator with Silverman bandwidth is well-known to be  $N^{2/5}$ . For most years of our analysis, we have about 22,000 observations, so  $N^{2/5}$  is about 55 and the density estimates differ from the density by a term proportional to  $1/55$ . The discriminatory power of the test statistics come from a test statistic that is  $O_p(1)$  under the unit root null but  $O_p(1/T)$  under the stationary alternative and  $O_p(T)$  or higher under the high-order nonstationary alternative. Supposing that the density estimation errors are uncorrelated over time – or at most stationary – the additional term is proportional to  $1/(167 \times 55)$ , which is too small to materially affect the size or power of the tests.

Let  $f_t(s)$  denote the value of a temperature anomaly density at time  $t$  and ordinate  $s$  (temperature anomaly), for  $t = 1, \dots, T$  and  $s \in \mathbb{R}$ . We define the *temporal mean* of a time series  $(f_t)$  of temperature anomaly densities as  $\bar{f}_T(s) = T^{-1} \sum_{t=1}^T f_t(s)$  for  $s \in \mathbb{R}$ , and the *cross-sectional mean* as  $\mu_t = \int s f_t(s) ds$  for  $t = 1, \dots, T$ .<sup>4</sup> The top left panel of each of Figs. 2–4 shows the annual temperature anomaly densities ( $f_t(s)$ ). Specifically, Fig. 2 shows the global densities, Fig. 3 shows those for the NH, and Fig. 4 shows those for the SH. We interpret the temporally demeaned temperature anomaly densities  $- (w_t(s))$  in our subsequent notation – as deviations from the average probability of observing a temperature anomaly over the sample time span. For example, in all of the figures, the probability of observing a  $+1^\circ\text{C}$  temperature anomaly appears to be below average in 1850 but above average in 2012, whereas the probability of observing  $-1^\circ\text{C}$  appears to be the reverse. Clearly, these are neither constant over time, as a flat graph would imply, nor do they appear to be generated by random noise.

The remaining panels of Figs. 2–4 show the time paths of the estimated *cross-sectional moments* of the distributions ( $f_t$ ). Specifically, the means (middle left panels), variances (middle right panels), skewnesses (bottom left panels), and kurtoses (bottom right panels) are plotted. The cross-sectional variance is given by  $\sigma_t^2 = \int (s - \mu_t)^2 f_t(s) ds$ , the cross-sectional skewness is given by  $\tau_t^3 = \int (s - \mu_t)^3 f_t(s) ds / \sigma_t^3$  and the cross-sectional kurtosis is given by  $\kappa_t^4 = \int (s - \mu_t)^4 f_t(s) ds / \sigma_t^4$  for  $t = 1, \dots, T$ .

Casual inspection suggests that the means have been increasing since about 1975 and perhaps since as early as 1910 in the SH, roughly consistent with the break dates identified by Gay-Garcia et al. (2009). While the means have increased, the skewnesses of the globe and NH are stable or slightly increasing while that of the SH appears to have decreased from positive to negative. Taken together, these casual observations suggest that the probabilities of observing moderately positive temperature anomalies have increased, while the probabilities of observing extremely positive temperature anomalies (up to the maxima of our supports) may have decreased in the SH, in spite of the recent record-high temperatures of Australia during the summer of 2017–18, which are not in our data. The variances of the distributions appear to have decreased on the whole, which could be due to improved measurement, increasing spatial coverage, and more observations. However, the variances have increased in recent decades. Note that the spatial variances may decrease even while the persistent means imply that the temporal variances must increase.

## 2.2. Basic framework for time series analysis

We analyze the temperature densities described above as a time series of functional observations. The underlying theory in this section is covered in more detail by Chang et al. (2016), but because this framework may be new to many readers, we briefly outline the key elements so that our discussion is self-contained.

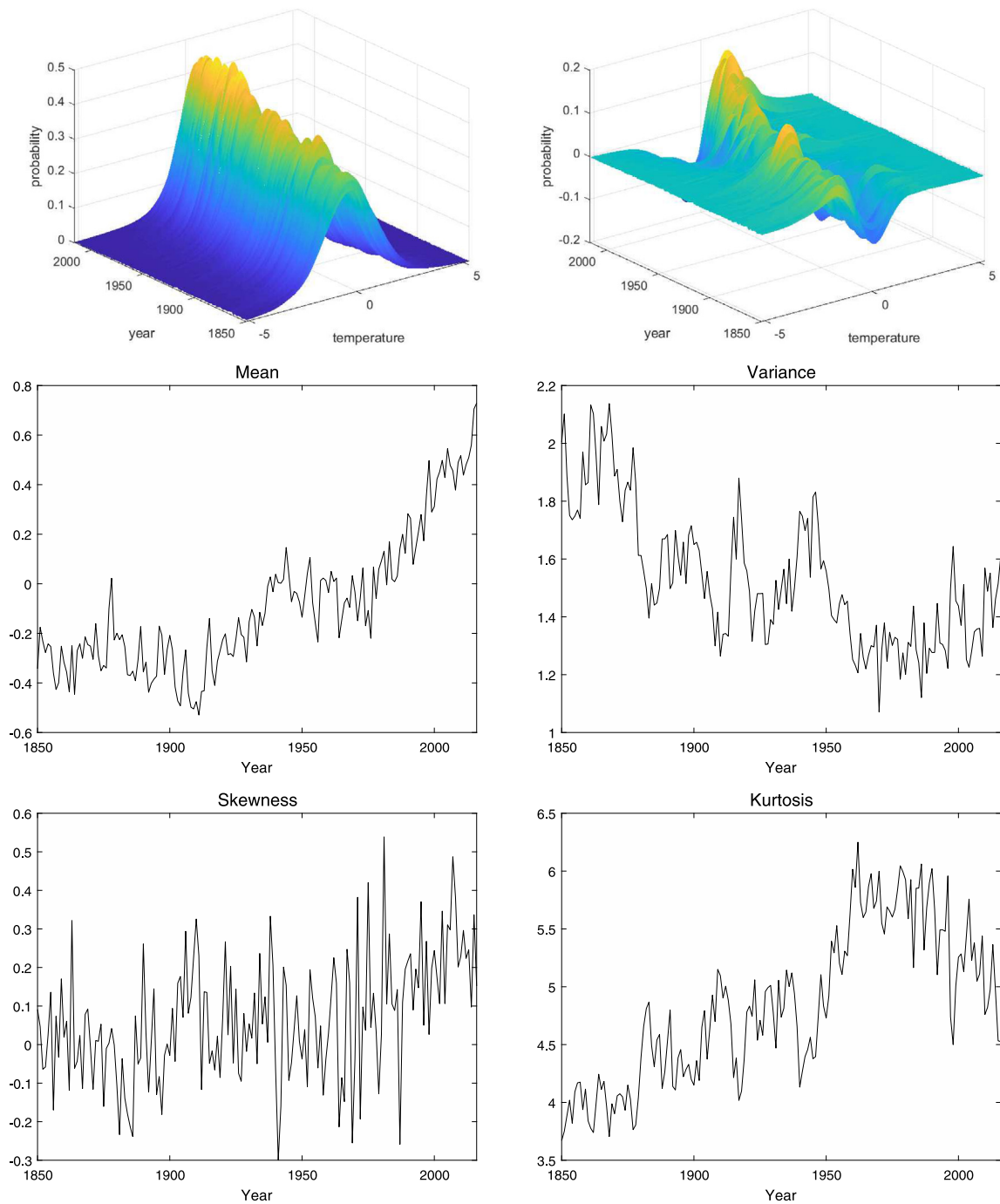
As defined above,  $(f_t)$  denotes the temperature anomaly density at time  $t$ , and we define

$$w_t(s) = f_t(s) - \bar{f}_T(s) \quad (1)$$

to be the temporally demeaned temperature density for  $t = 1, \dots, T$  and  $s \in K$ , where  $K$  is a compact subset of  $\mathbb{R}$ . Clearly, we have  $\int_K f_t(s) ds = 1$  for all  $t = 1, 2, \dots$ , and therefore,  $(w_t)$  may be regarded as elements in the Hilbert space  $H$  given

<sup>3</sup> The compact supports avoid the well-known empty bin problem in nonparametric density estimation. Our omission of extreme outliers should not substantially affect our conclusions about the nonstationarity of the distributions.

<sup>4</sup> In practice, integrals are evaluated using Riemann sums over arbitrarily fine partitions of the supports of the densities.

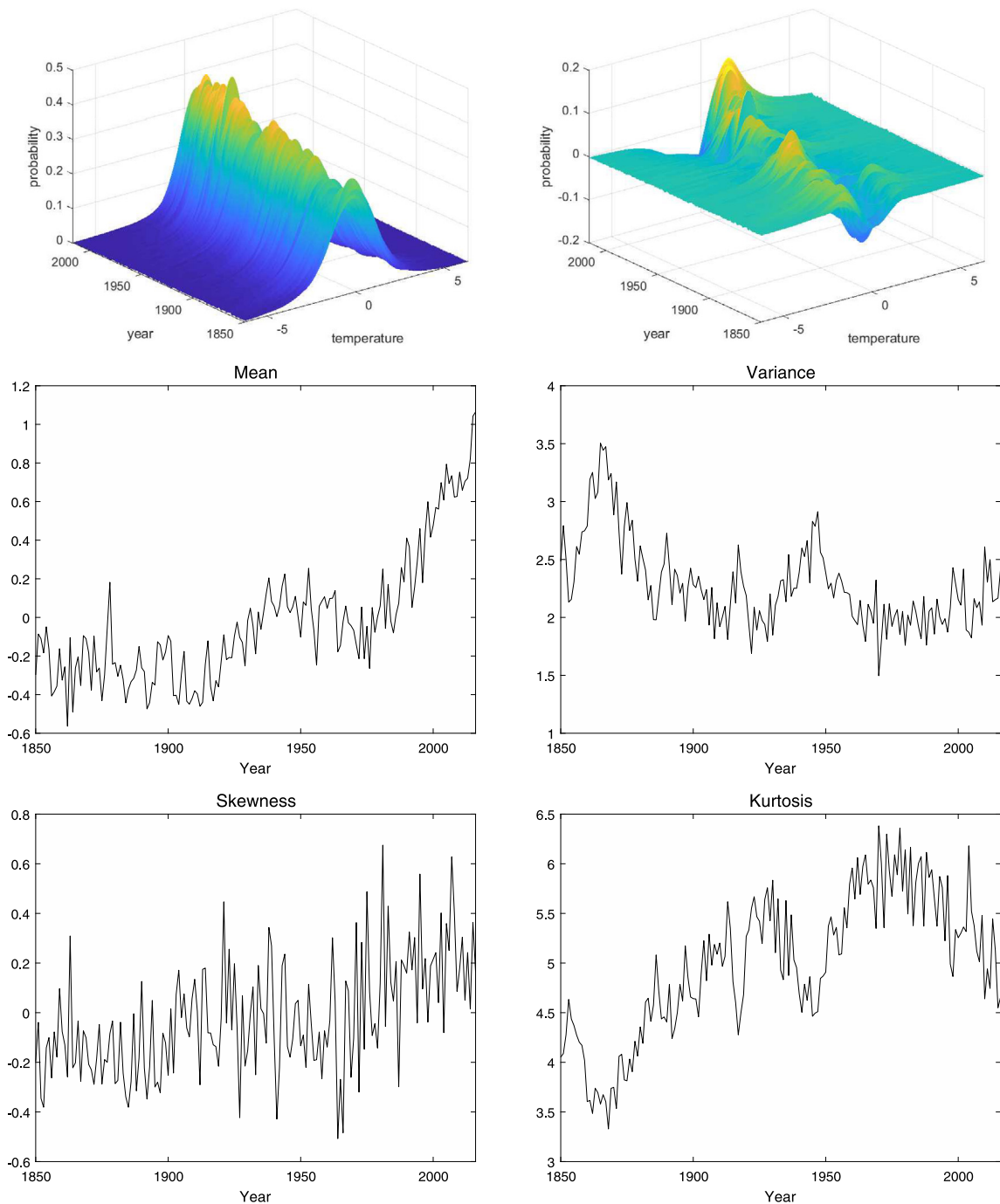


**Fig. 2.** Global temperature anomaly densities and moments. Annual temperature anomalies measured on a  $5^\circ$  by  $5^\circ$  grid box. Undemeaned densities (top left panel) and temporally demeaned global densities (top right panel). Cross-sectional mean (middle left panel), variance (middle right panel), skewness (bottom left panel), and kurtosis (bottom right panel) of anomalies over time.

by

$$H = \left\{ w \left| \int_K w(s) ds = 0, \int_K w^2(s) ds < \infty \right. \right\}, \quad (2)$$

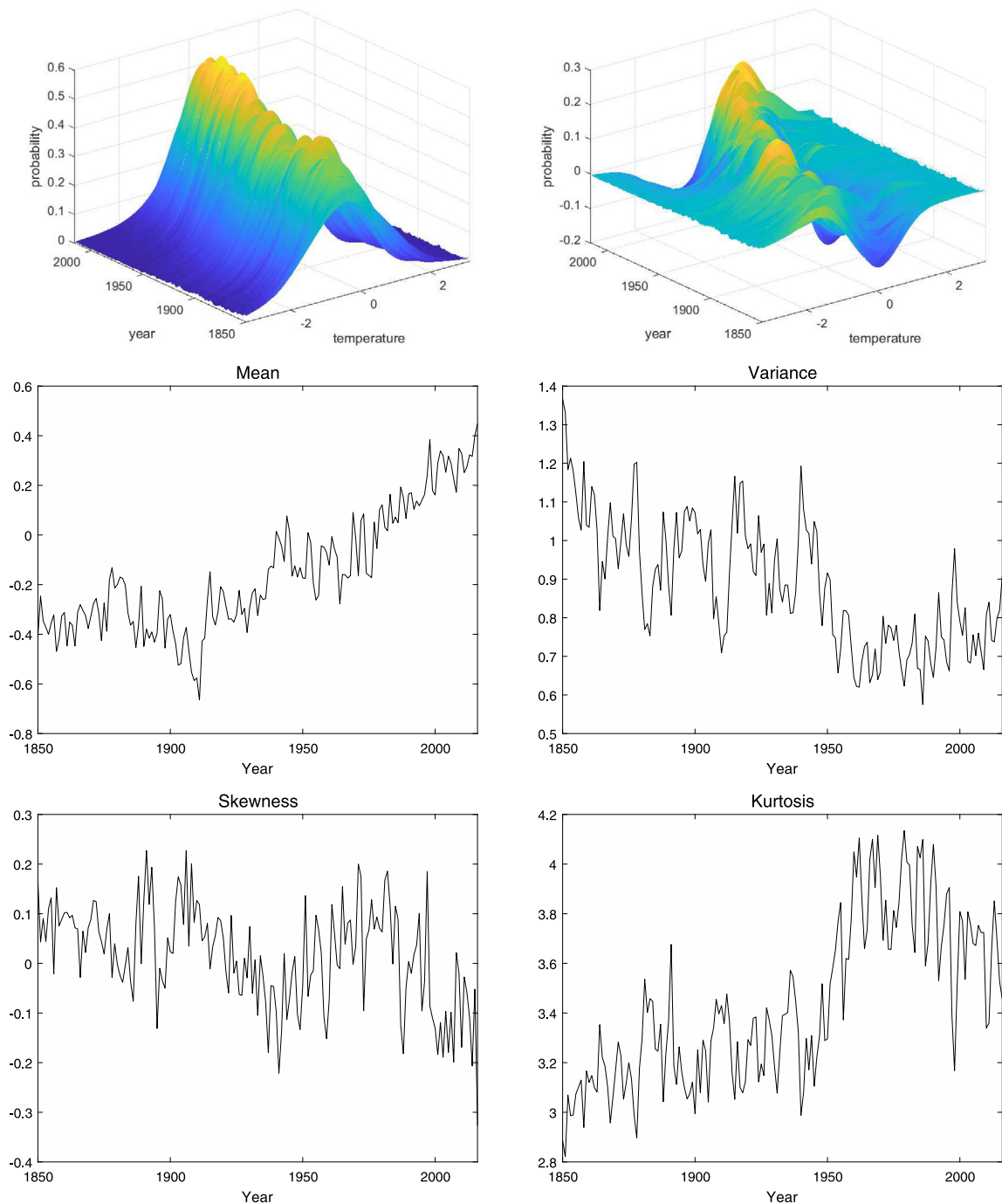
with inner product  $\langle v, w \rangle = \int_K v(s)w(s)ds$  for  $v, w \in H$ .



**Fig. 3.** NH temperature anomaly densities and moments. Annual temperature anomalies measured on a 5° by 5° grid box. Undemeaned densities (top left panel) and temporally demeaned global densities (top right panel). Cross-sectional mean (middle left panel), variance (middle right panel), skewness (bottom left panel), and kurtosis (bottom right panel) of anomalies over time.

In our analysis, we assume that the global temperature densities ( $f_t$ ) are random, not deterministic, and consequently, the centered global temperatures densities ( $w_t$ ) defined in (1) become random elements taking values in the Hilbert space  $H$ , or  $H$ -valued random elements. For an introduction to random elements taking values in a Hilbert space, the reader is referred to Bosq (2000). For each  $t = 1, \dots, T$ ,  $f_t$  is a random function and we may define its moments. In particular, its mean is given by the expectation  $\mathbb{E}f_t$ , and its variance is given by the expected tensor product  $\mathbb{E}(f_t - \mathbb{E}f_t) \otimes (f_t - \mathbb{E}f_t)$  of





**Fig. 4.** SH temperature anomaly densities and moments. Annual temperature anomalies measured on a  $5^\circ$  by  $5^\circ$  grid box. Undemeaned densities (top left panel) and temporally demeaned global densities (top right panel). Cross-sectional mean (middle left panel), variance (middle right panel), skewness (bottom left panel), and kurtosis (bottom right panel) of anomalies over time.

the demeaned  $f_t$  with itself.<sup>5</sup> The mean and variance of  $f_t$  therefore become an element of  $H$  and an operator acting on  $H$  respectively for  $t = 1, \dots, T$ . On the other hand, because each element  $f_t$  of the sequence  $(f_t)$  represents a density, we

<sup>5</sup> Essentially, tensor products of finite dimensional vectors yield matrices. In contrast, tensor products of functions are infinite dimensional. In some sense, they are like infinite dimensional matrices, but they are formally interpreted as operators in a Hilbert space of functions.

may also define its moments. We have already defined these as *cross-sectional moments*  $\mu_t, \sigma_t^2$ , etc., of  $f_t$ . Note that the cross-sectional moments of  $f_t$  are random variables for each  $t = 1, \dots, T$ .

We assume that there exists an orthonormal basis  $(v_i)$  of  $H$  such that the  $i$ th coordinate process  $\langle v_i, w_t \rangle$  is nonstationary, having a stochastic or deterministic trend, for each  $i = 1, \dots, n$ , while it is stationary for each  $i \geq n+1$ .<sup>6</sup> Using the symbol  $\bigvee$  to denote span, we may write  $H = H_N \oplus H_S$  with

$$H_N = \bigvee_{i=1}^n v_i \quad \text{and} \quad H_S = \bigvee_{i=n+1}^{\infty} v_i,$$

which we call the *nonstationary and stationary subspaces* of  $H$  respectively. Subsequently, we define  $\Pi_N$  and  $\Pi_S$  to be the projections on  $H_N$  and  $H_S$ , and let  $w_t^N = \Pi_N w_t$  and  $w_t^S = \Pi_S w_t$ , where  $(w_t^N)$  and  $(w_t^S)$  signify the nonstationary and stationary components of  $(w_t)$  respectively. Because  $\Pi_N + \Pi_S$  equals the identity operator in  $H$ , we have  $w_t = w_t^N + w_t^S$ .

We say that  $(f_t)$  is (weakly) stationary if its mean and variance are time invariant, finite, and well defined. In this case,  $n = 0$ , because all the coordinate processes are stationary. Under stationarity and ergodicity, we may expect that  $\bar{f}_T(s) \approx \mathbb{E}f_t(s)$  and  $w_t(s) \approx f_t(s) - \mathbb{E}f_t(s)$  for all  $t = 1, \dots, T$  and  $s \in K$  if  $T$  is large. Consequently, we may effectively let

$$w_t(s) = f_t(s) - \mathbb{E}f_t(s) \quad (3)$$

if  $T$  is large, in place of our definition in (1). In our subsequent analysis, we do not distinguish between any stationary time series defined from  $(w_t)$  in (3) and  $(w_t)$  in (1).

### 2.3. Testing for nonstationarity

We use a test for the nonstationarity of the global temperature anomaly distributions based on the sample variance operator

$$Q_T = \sum_{t=1}^T w_t \otimes w_t, \quad (4)$$

which yields the quadratic form

$$\langle v, Q_T v \rangle = \sum_{t=1}^T \langle v, w_t \rangle^2 \quad (5)$$

for any  $v \in H$ .

The magnitude of quadratic form (5) in  $v \in H$  defined by  $Q_T$  differs primarily depending upon whether  $v$  is in  $H_N$  or in  $H_S$ . For  $v \in H_S$ , the coordinate process  $(\langle v, w_t \rangle)$  becomes stationary and  $T^{-1} \sum_{t=1}^T \langle v, w_t \rangle^2 \rightarrow_p \mathbb{E} \langle v, w_t \rangle^2$ , and the quadratic form is of order  $T$ . In contrast, the order of magnitude of the quadratic form in  $v \in H_N$  is of order bigger than  $T$ , because we assume that for all  $v \in H_N$  the coordinate process  $(\langle v, w_t \rangle)$  has a stochastic or deterministic trend. We may therefore extract the principal components of  $Q_T$  in (4) and use them to test for nonstationarity in the temperature anomaly distributions.<sup>7</sup>

The exact order of magnitude of the quadratic form in  $v \in H_N$  defined by  $Q_T$  further depends on the type of nonstationarity exhibited by the coordinate process  $(\langle v, w_t \rangle)$ . To identify these different types of nonstationarity in the temperature anomaly distributions, we define the unit root subspace  $H_U$  of  $H$  such that  $(\langle v, w_t \rangle)$  is a unit root process for all  $v \in H_U$ , and the deterministic and explosive subspace  $H_X$  of  $H$  such that  $(\langle v, w_t \rangle)$  has a deterministic trend or explosive root for all  $v \in H_X$ . We assume that  $(\langle v, w_t \rangle)$  for  $v \in H_X$  is roughly an order bigger than that of a unit root process. The precise specifications and examples of deterministic trends and explosive unit roots allowed are given in Assumption 2.1.

More explicitly, we assume that there is an orthonormal basis  $(v_i)$  such that

$$H_X = \bigvee_{i=1}^m v_i, \quad H_U = \bigvee_{i=m+1}^n v_i, \quad \text{and} \quad H_S = \bigvee_{i=n+1}^{\infty} v_i, \quad (6)$$

in which case we have in particular that  $H = H_N \oplus H_S = H_X \oplus H_U \oplus H_S$ , and  $H_X$  and  $H_U$  are  $m$  and  $(n-m)$ -dimensional respectively. If there is no deterministic/explosive nonstationarity, then  $m = 0$  and all nonstationarity is unit root nonstationarity. In fact,  $m = 0$  for the temperature anomalies below. However, we also consider the case of  $m \neq 0$  to introduce our test to distinguish between these cases. If  $m = 0$ , the representation in (6) follows immediately from that

<sup>6</sup> Of course, there exists a wide variety of nonstationary processes that do not have any trends, stochastic or deterministic. However, we only consider nonstationary processes with such trends.

<sup>7</sup> As noted by Hormann et al. (2015), static principal components does not achieve sufficient dimension reduction when considering only stationary coordinate processes, and they propose dynamic functional principal components to overcome the deficiency. Because the discriminatory powers of our tests rely on different orders of magnitude depending on the stationarity or nonstationarity of the coordinate processes, the use of static functional principal components is appropriate in this context.



in Chang et al. (2016), our representation extends theirs to allow for  $m \neq 0$ . If, for instance,  $(w_t)$  is given by  $w_t = ut + w_t^\circ$  for  $t = 1, 2, \dots$ , where  $u \in H$  and  $(w_t^\circ)$  is a purely stochastic time series having only unit root and stationary components, then we may set  $H_X = \text{span}\{v_1\}$  with  $v_1 = u/\|u\|$ , and represent  $H_X^\perp$  as  $H_X^\perp = H_U \oplus H_S$  as in Chang et al. (2016) to obtain (6). In this case, of course, we have  $m = 1$ .

Denote by  $v_1(Q_T), v_2(Q_T), \dots$  the orthonormal eigenvectors of operator  $Q_T$  in (4) with associated eigenvalues  $\lambda_1(Q_T) \geq \lambda_2(Q_T) \geq \dots$ . It follows that

$$\lambda_i(Q_T) = \langle v_i(Q_T), Q_T v_i(Q_T) \rangle = \sum_{t=1}^T \langle v_i(Q_T), w_t \rangle^2.$$

Therefore, it is natural to estimate  $H_N$  by the span of  $v_1(Q_T), \dots, v_n(Q_T)$  – i.e.,  $n$  orthonormal eigenvectors of  $Q_T$  associated with the  $n$  largest eigenvalues of  $Q_T$ . Chang et al. (2016) establish the consistency of the estimator for the case in which we only have unit root nonstationarity. Extending their proof to allow for more general types of nonstationarity is straightforward. In our setup, if normalized by  $T^2$ ,  $\lambda_{m+1}(Q_T), \dots, \lambda_n(Q_T)$  have well-defined limit distributions as  $T \rightarrow \infty$ , while  $\lambda_1(Q_T), \dots, \lambda_m(Q_T)$  diverge faster than the rate  $T^2$ . In particular, the unit root subspace  $H_U$  can be estimated consistently by the span of  $(n-m)$ -orthonormal eigenvectors  $v_{m+1}(Q_T), \dots, v_n(Q_T)$  of  $Q_T$ .

We find the values of  $n$  and  $m$  by successive testing procedures for the null hypothesis of unit root nonstationarity against the alternative hypothesis of stationarity and against the alternative hypothesis of deterministic/explosive nonstationarity. We expect that the eigenvalues  $(\lambda_i(Q_T))$  have discriminatory powers for such tests. However, these eigenvalues cannot be used directly, because their limit distributions depend on nuisance parameters. In particular, they are sensitive to serial correlation in  $(\Delta w_t)$ . Therefore, we construct tests based on eigenvalues with limit distributions that are free of nuisance parameters.

To this end, for  $\kappa \geq 1$  fixed, we define  $(z_t)$  by

$$z_t = (\langle v_1(Q_T), w_t \rangle, \dots, \langle v_\kappa(Q_T), w_t \rangle)' \quad (7)$$

for  $t = 1, \dots, T$ , and the product sample moment  $Q_\kappa^T = \sum_{t=1}^T z_t z_t'$ . In order to account for serial correlation in  $(\Delta w_t)$ , we define the long-run variance estimator  $\Omega_\kappa^T = \sum_{|k| \leq \ell_T} \varpi(k) \Gamma_T(k)$  of  $(z_t)$ , where  $\varpi$  is the weight function,  $\ell_T$  is the bandwidth parameter and  $\Gamma_T$  is the sample autocovariance function defined as  $\Gamma_T(k) = T^{-1} \sum_t \Delta z_t \Delta z_{t-k}'$ . Our test statistics are given by

$$\tau_\kappa^T = T^{-2} \lambda_{\min}(Q_\kappa^T, \Omega_\kappa^T) \quad (8)$$

and

$$\sigma_\kappa^T = T^{-2} \lambda_{\max}(Q_\kappa^T, \Omega_\kappa^T), \quad (9)$$

where  $\lambda_{\min}(Q_\kappa^T, \Omega_\kappa^T)$  and  $\lambda_{\max}(Q_\kappa^T, \Omega_\kappa^T)$  are respectively the smallest and the largest generalized eigenvalues of  $Q_\kappa^T$  with respect to  $\Omega_\kappa^T$ .

We use the statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$  to investigate the dimension of nonstationarity and the presence of deterministic/explosive nonstationarity. More precisely, the statistic  $\tau_\kappa^T$  is used to test the null hypothesis  $H_0^\tau$  against the alternative hypothesis  $H_1^\tau$ , which are given by

$$H_0^\tau : n = \kappa \quad \text{and} \quad H_1^\tau : n \leq \kappa - 1, \quad (10)$$

respectively. On the other hand, the statistic  $\sigma_\kappa^T$  is used to test the null hypothesis  $H_0^\sigma$  against the alternative hypothesis  $H_1^\sigma$ , which are given by

$$H_0^\sigma : n \leq \kappa, \quad m = 0 \quad \text{and} \quad H_1^\sigma : m \geq 1, \quad (11)$$

respectively. Recall that  $n$  is the dimension of nonstationarity, and  $m = 0$  implies that there exists no deterministic/explosive nonstationarity.

To develop the asymptotic theory for test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ , we define  $(w_t^\circ)$  as

$$w_t^\circ = \sum_{i=1}^{\infty} \langle v_{m+i}, w_t \rangle v_{m+i} \quad (12)$$

for  $t = 1, 2, \dots$ , we let  $Q_T$  denote the sample variance operator of  $(w_t)$  in (4), and we let  $\Omega_T$  denote the long-run variance estimator of  $(\Delta w_t)$  defined with bandwidth parameter  $\ell_T$ .

**Assumption 2.1.** We assume that (a)  $(w_t^\circ)$  satisfies Assumption 2.1 of Chang et al. (2016), and (b)  $T^{-2} \langle v_i, Q_T v_i \rangle \rightarrow_p \infty$  and  $\langle v_i, \Omega_T v_i \rangle \rightarrow \infty$  for all  $i = 1, \dots, m$  as  $T \rightarrow \infty$ .

Part (a) of Assumption 2.1 is mild. As discussed in Chang et al. (2016), the conditions are weak and are expected to hold widely. Part (b) of Assumption 2.1 also is expected to hold generally. Note that the first condition in part (b) implies

**Table 1**One-sided critical values for the test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ .

$\tau_\kappa^T$					
$\kappa$	1	2	3	4	5
1%	0.0248	0.0162	0.0123	0.0100	0.0084
5%	0.0366	0.0216	0.0156	0.0122	0.0101
10%	0.0460	0.0254	0.0177	0.0137	0.0111
$\sigma_\kappa^T$					
$\kappa$	1	2	3	4	5
1%	0.7452	1.0041	1.2232	1.4172	1.5987
5%	0.4613	0.6797	0.8649	1.0346	1.1954
10%	0.3471	0.5399	0.7064	0.8624	1.0101

$T^{-2} \sum_{t=1}^T \langle v_i, w_t \rangle^2 \rightarrow_p \infty$  for all  $i = 1, \dots, m$ . Note that, for  $i = m + 1, \dots, n$ ,  $(\langle v_i, w_t \rangle)$  has a unit root, and we have  $\langle v_i, Q_T v_i \rangle = \sum_{t=1}^T \langle v_i, w_t \rangle^2 = O_p(T^2)$ . If  $(\langle v_i, w_t \rangle)$  has a linear time trend, we have  $\langle v_i, Q_T v_i \rangle = \sum_{t=1}^T \langle v_i, w_t \rangle^2 = O_p(T^3)$  and the condition holds. On the other hand, if  $(\langle v_i, w_t \rangle)$  is an explosive AR process with an AR coefficient  $\rho > 1$ , then  $\langle v_i, Q_T v_i \rangle = \sum_{t=1}^T \langle v_i, w_t \rangle^2 = O_p(\rho^{2T})$  and the condition is obviously satisfied. The second condition in part (b) implies that the long-run variance estimate of  $(\langle v_i, \Delta w_t \rangle)$  diverges to infinity for  $i = 1, \dots, m$ . This condition holds if  $(\langle v_i, w_t \rangle)$  has a time trend of order  $r > 1/2$ , as required by the first condition  $T^{-2} \langle v_i, Q_T v_i \rangle \rightarrow_p \infty$ . In this case,  $(\langle v_i, \Delta w_t \rangle)$  has a time trend of order  $r - 1$  and its long-run variance estimate becomes of order  $\ell_T T^{2(r-1)}$ , and therefore, we may choose  $\ell_T \rightarrow \infty$  such that  $\ell_T/T \rightarrow 0$  and  $\ell_T T^{2(r-1)} \rightarrow \infty$  as  $T \rightarrow \infty$ .

The following theorems and corollaries summarize the asymptotics of  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ . For the presentation of our asymptotic analysis, it is useful to introduce

$$Q_\kappa = \int_0^1 W_\kappa(r) W_\kappa(r)' dr - \left( \int_0^1 W_\kappa(r) dr \right) \left( \int_0^1 W_\kappa(r) dr \right)', \quad (13)$$

where  $W_\kappa$  denotes  $\kappa$ -dimensional standard Brownian motion for any fixed  $\kappa \geq 1$ . Due to temporal demeaning in (1), this distribution is analogous to many of those used for tests of unit roots and cointegration in models with non-zero means under the null.

**Theorem 2.2.** Let  $n = \kappa$  and  $m = 0$ . Then  $\tau_\kappa^T \rightarrow_d \lambda_{\min}(Q_\kappa)$  and  $\sigma_\kappa^T \rightarrow_d \lambda_{\max}(Q_\kappa)$  jointly as  $T \rightarrow \infty$ .

**Corollary 2.3.** Let  $n = \kappa$  and  $m \geq 1$ . (a) If  $m \leq \kappa - 1$ , then  $\tau_\kappa^T \rightarrow_d \lambda_{\min}(Q_{\kappa-m})$  as  $T \rightarrow \infty$ . (b) If  $m = \kappa$ , then  $\tau_\kappa^T \rightarrow_p \infty$  as  $T \rightarrow \infty$ .

**Corollary 2.4.** Let  $m = 0$  and  $n \leq \kappa - 1$ . (a) If  $n \geq 1$ , then  $\sigma_\kappa^T \rightarrow_d \lambda_{\max}(Q_n)$  as  $T \rightarrow \infty$ . (b) If  $n = 0$ , then  $\sigma_\kappa^T \rightarrow_p 0$  as  $T \rightarrow \infty$ .

**Theorem 2.5.** (a) If  $n \leq \kappa - 1$ , then  $\tau_\kappa^T \rightarrow_p 0$  as  $T \rightarrow \infty$ . (b) If  $m \geq 1$ , then  $\sigma_\kappa^T \rightarrow_p \infty$  as  $T \rightarrow \infty$ .

The critical values of the test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$  are computed assuming  $n = \kappa$  and  $m = 0$ , as in Theorem 2.2 and similarly to those of Chang et al. (2016), and they are tabulated for the values of  $\kappa = 1, \dots, 5$  (Table 1). Although the critical values are computed using  $n = \kappa$  and  $m = 0$ , Table 2 shows the limits in other cases which fall under the respective nulls of the tests.

For the test based on  $\tau_\kappa^T$ , we reject the null hypothesis  $H_0^T$  in favor of the alternative hypothesis  $H_1^T$  in (10) if  $\tau_\kappa^T$  takes values smaller than the critical values in Table 1. It is clear from our asymptotics in Theorem 2.2 and those in Corollary 2.3 summarized in Table 2 that the test based on  $\tau_\kappa^T$  is asymptotically valid as long as  $n = \kappa$  regardless of whether  $m = 0$  or  $m \geq 1$ . If  $m = 0$ , the asymptotic rejection probability of the test becomes identical to the size. The asymptotic rejection probability is smaller if  $1 \leq m \leq \kappa - 1$ , because the dimension of the limiting Brownian motion in (13) is reduced by  $m$  so that the critical value based on  $Q_\kappa$  is too extreme. If  $1 \leq m = \kappa$ , the dimension  $\kappa - m$  is zero, the test statistic diverges so that the asymptotic rejection probability is zero. This case is conceptually similar to that resulting from a deterministic trend in a univariate unit root test. In the latter two cases, the test becomes conservative with the asymptotic rejection probability strictly smaller than the size.

The test based on  $\tau_\kappa^T$  clearly is consistent, because the statistic converges to 0 for all  $n \leq \kappa - 1$ , due to part (a) of Theorem 2.5. Therefore, we may use  $\tau_\kappa^T$  to determine  $n$  as follows.<sup>8</sup> We start from a value of  $\kappa$  large enough to be bigger than  $n$  and test the null hypothesis  $H_0^T$  against the alternative hypothesis  $H_1^T$  successively downward, until we reach  $\kappa = 1$ .

<sup>8</sup> Our testing procedure here is entirely analogous to the sequential procedure in Johansen (1995), which is commonly used to determine the cointegration ranks in error correction models.

**Table 2**

Limits under the assumptions of the corollaries. Limits shown for the scenarios considered in Corollaries 2.3 and 2.4. The inequalities signify the relative magnitudes of the critical values.  $\tau_\kappa^T$  is less likely to reject (conservative) with a smaller critical value, while  $\sigma_\kappa^T$  is less likely to reject (conservative) with a larger critical value.

$\tau_\kappa^T$ with $n = \kappa$							
	$m = 0$		$m = 1$		$m = 2$		$m = 3$
$\kappa = 1$	$\lambda_{\min}(Q_1)$	<	$\infty$				
$\kappa = 2$	$\lambda_{\min}(Q_2)$	<	$\lambda_{\min}(Q_1)$	<	$\infty$		
$\kappa = 3$	$\lambda_{\min}(Q_3)$	<	$\lambda_{\min}(Q_2)$	<	$\lambda_{\min}(Q_1)$	<	$\infty$
$\kappa = 4$	$\lambda_{\min}(Q_4)$	<	$\lambda_{\min}(Q_3)$	<	$\lambda_{\min}(Q_2)$	<	$\lambda_{\min}(Q_1)$
$\sigma_\kappa^T$ with $m = 0$							
	$n = 0$		$n = 1$		$n = 2$		$n = 3$
$\kappa = 1$	0	<	$\lambda_{\max}(Q_1)$				
$\kappa = 2$	0	<	$\lambda_{\max}(Q_1)$	<	$\lambda_{\max}(Q_2)$		
$\kappa = 3$	0	<	$\lambda_{\max}(Q_1)$	<	$\lambda_{\max}(Q_2)$	<	$\lambda_{\max}(Q_3)$
$\kappa = 4$	0	<	$\lambda_{\max}(Q_1)$	<	$\lambda_{\max}(Q_2)$	<	$\lambda_{\max}(Q_3)$

We proceed as long as we reject the null hypothesis in favor of the alternative hypothesis, and set our estimate for  $n$  to be the largest value  $\kappa$  for which we fail to reject the null hypothesis. Because this successive testing procedure employs a consistent test, it allows us to find the true value of  $n$  with asymptotic probability of virtually one by making the size of the test small enough.

The test based on  $\sigma_\kappa^T$  can be used to test for the presence of deterministic/explosive nonstationarity. For this test, we reject the null hypothesis  $H_0^\sigma$  in favor of the alternative hypothesis  $H_1^\sigma$  in (11) if the value of  $\sigma_\kappa^T$  is greater than the appropriate critical value in Table 1. Asymptotic validity of the test based on  $\sigma_\kappa^T$  can be established from Theorem 2.2 and part (a) of Corollary 2.4, and its consistency follows directly from part (b) of Theorem 2.5. The test is valid as long as we construct  $\sigma_\kappa^T$  with a value of  $\kappa$  such that  $\kappa \geq n$ , as may be gleaned from Table 2. The statistic  $\sigma_\kappa^T$  does not diverge for  $\kappa < n$  as long as  $m = 0$ . However, the test based on  $\sigma_\kappa^T$  generally does not yield correct asymptotic sizes in this case.

#### 2.4. Nonstationarity in cross-sectional moments

Once we determine  $n$  and estimate the nonstationary subspace  $H_N$ , we are able to determine the nonstationary proportion of each cross-sectional moment. Similarly to Chang et al. (2016), we define a function

$$\mu_i(s) = s^i - \frac{1}{|K|} \int_K s^i ds$$

for  $i = 1, 2, \dots$  and Lebesgue measure  $|K|$  of  $K$ , and note that  $\langle \mu_i, w_t \rangle = \langle \mu_i, f_t \rangle - \mathbb{E} \langle \mu_i, f_t \rangle$  represents the fluctuations over time of the  $i$ th moments of the distributions with densities ( $f_t$ ) around their expected values.

The function  $\mu_i$  may be decomposed as  $\mu_i = \Pi_N \mu_i + \Pi_S \mu_i$  with  $\Pi_N$  and  $\Pi_S$  defined as projections respectively on the nonstationary and stationary subspaces  $H_N$  and  $H_S$ , so that

$$\|\mu_i\|^2 = \|\Pi_N \mu_i\|^2 + \|\Pi_S \mu_i\|^2 = \sum_{j=1}^n \langle \mu_i, v_j \rangle^2 + \sum_{j=n+1}^{\infty} \langle \mu_i, v_j \rangle^2, \quad (14)$$

where  $(v_j)$  for  $j = 1, 2, \dots$  is an orthonormal basis of  $H$  such that  $(v_j)_{1 \leq j \leq n}$  spans  $H_N$  and  $(v_j)_{j \geq n+1}$  spans  $H_S$ .

The proportion of the component of  $\mu_i$  lying in  $H_N$  is given by

$$\pi_i^N = \frac{\|\Pi_N \mu_i\|}{\|\mu_i\|} = \sqrt{\frac{\sum_{j=1}^n \langle \mu_i, v_j \rangle^2}{\sum_{j=1}^{\infty} \langle \mu_i, v_j \rangle^2}} \quad (15)$$

with the convention that  $\pi_i^N = 0$  when  $n = 0$  ( $\mu_i$  is entirely in  $H_S$ ). On the other hand,  $\mu_i$  is entirely in  $H_N$  if  $\pi_i^N = 1$ .  $\pi_i^N$  represents the proportion of the nonstationary component in the  $i$ th moment, which we call the *nonstationary proportion* of the  $i$ th moment. As  $\pi_i$  approaches zero, the  $i$ th moment is predominantly stationary, but it is predominantly nonstationary as  $\pi_i$  tends to unity.

To supplement  $\pi_i^N$ , we propose a new ratio given by

$$\pi_i^U = \frac{\|\Pi_U \mu_i\|}{\|\mu_i\|} = \sqrt{\frac{\sum_{j=m+1}^n \langle \mu_i, v_j \rangle^2}{\sum_{j=1}^{\infty} \langle \mu_i, v_j \rangle^2}}, \quad (16)$$

where  $\Pi_U$  is the projection on the unit root subspace  $H_U$ , with the convention that  $\pi_i^U = 0$  when  $m = n$ . When  $m = 0$ ,  $\pi_i^U = \pi_i^N$  so that the component of  $\mu_i$  in  $H_N$  is entirely in  $H_U$ . Alternatively, when  $m = n$  and  $\pi_i^U = 0$ , all of the proportion

in  $H_N$  is in the deterministic and explosive subspace  $H_X$ . We call  $\pi_i^U$  the *unit root proportion* of the  $i$ th moment. Generally, it is more difficult to predict the  $i$ th moment if  $\pi_i^U$  is closer to unity. In contrast, the  $i$ th moment is easier to predict if  $\pi_i^U$  is small — either because  $\|\Pi_S \mu_i\|$  is relatively large due to stationarity or because  $\|(\Pi_N - \Pi_U) \mu_i\|$  is relatively large due to a deterministic trend.

To illustrate, suppose that  $n = 1$  and  $m = 0$  so the spatial distribution contains only one unit root process, which is the series of time-varying coefficients on the first basis function  $v_1$ . The numerator of  $\pi_i^U$  is given by

$$\int_K \left( r^i - \frac{1}{|K|} \int_K s^i ds \right) v_1(r) dr$$

which is the square root of the variation over time in the  $i$ th spatial moment (the mean, if  $i = 1$ ) due to the unit root in this one series of coefficients given by  $(\langle v_1, w_t \rangle)$ . The proportion  $\pi_i^U$  tracks the proportion of this variation over time compared to the total variation over time due to the series of coefficients on all of the basis functions.

## 2.5. Finite-sample evaluation

We perform Monte Carlo simulations to investigate the finite-sample performance of the tests based on  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ . Chang et al. (2016) conducted simulations to explore the finite-sample properties of  $\tau_\kappa^T$ , but tailored to a different empirical application. We use the time series of estimated distributions of global temperature anomalies so that our simulations for both test statistics are relevant to the present application.

As discussed above, the large numbers of cross-sectional observations imply that the approximation errors due to the estimation of state densities are quite small and so we ignore them. We consider time dimensions of  $T = 100, 200, 300$  and a nominal significance level of 5%, and all the reported results are based on 5000 iterations.

We generate  $(w_t)$  as

$$w_t = \sum_{i=1}^T c_{it} v_i, \quad (17)$$

where  $(c_{it})$ , for  $i = 1, \dots, I$  and  $t = 1, \dots, T$ , are scalar stochastic processes that differ based on the hypotheses as described below, and  $(v_i)$ , for  $i = 1, \dots, I$ , are nonrandom orthonormal vectors in the Hilbert space.  $(c_{it} = \langle v_i, w_t \rangle)$  is the  $i$ th coordinate process of  $(w_t)$  for  $i = 1, \dots, I$ . Note that the simulated sample  $(w_t)$  is generated from the  $I$ -dimensional subspace of  $H$  spanned by  $(v_i)$ ,  $i = 1, \dots, I$ . We set  $I = T$ , and let  $(v_i)$  in (17) be the orthonormal eigenvectors associated with nonzero eigenvalues of the estimated variance operator of  $(w_t)$  obtained from global temperature anomalies.

We consider two sets of data-generating processes for each test. For the test based on  $\tau_\kappa^T$ , we consider  $H_0^\tau : n = 2$  vs.  $H_1^\tau : n = 1$  (two vs. one stochastic trend in  $(w_t)$ ) and  $H_0^\tau : n = 1$  vs.  $H_1^\tau : n = 0$  (one vs. no stochastic trend in  $(w_t)$ ), and we set  $m = 0$  in both cases (no higher-order nonstationarity in  $(w_t)$ ). This exercise is particularly meaningful for the analysis of the SH, in which we subsequently need to distinguish between one and two unit roots. For the test based on  $\sigma_\kappa^T$ , we consider two cases with the same null and alternative,  $H_0^\sigma : m = 0$  vs.  $H_1^\sigma : m = 1$  (no vs. some higher-order nonstationarity in  $(w_t)$ ), but with  $n = 2$  and  $n = 1$ . The number of stochastic trends in  $(w_t)$ ,  $n - m$ , is 2 vs. 1 in the first set and 1 vs. 0 in the second set. In both of the latter exercises, the null and alternative differ based on the existence of a higher-order trend, but the null with  $n = 2$  reflects more closely the globe and NH, while the null with  $n = 1$  reflects more closely the SH.

In order to implement the case of  $H_0^\tau : n = 2$  vs.  $H_1^\tau : n = 1$ , we employ a data-generating process given by

$$\begin{aligned} i = 1 : \Delta c_{1t} &= \beta_1 \Delta c_{1,t-1} + \eta_{1t} \\ i = 2 : (c_{2t} - (1 - T^{-1}b)c_{2,t-1}) &= \beta_2 (c_{2,t-1} - (1 - T^{-1}b)c_{2,t-2}) + \eta_{2t} \\ i \geq 3 : c_{it} &= \alpha_i c_{i,t-1} + \eta_{it} \end{aligned}$$

where  $(\eta_{it})$  are independent normal random variates with a mean of zero and a variance given by that of the estimated coordinate processes of the global temperature anomaly distribution, with the null imposed for  $i = 2$ . The coefficients  $\beta_1$ ,  $\beta_2$ , and  $\alpha_i$  for  $i \geq 3$  are those from fitting the estimated coordinate processes to the respective equations above, again with the null imposed ( $b = 0$ ) for  $i = 2$ .<sup>9</sup> In order to obtain plausible values for  $b$ , we first estimate  $\alpha_2$  from the second coordinate process of the temperature anomaly distribution using  $c_{2t} = \alpha_2 c_{2,t-1} + e_{2t}$  and then set  $b \approx \hat{\alpha}_2 T$ . Specifically, with  $T = 100, 200, 300$ , we set  $b \in [10, 50]$  under the alternative.

The case of  $H_0^\tau : n = 1$  vs.  $H_1^\tau : n = 0$  is implemented using

$$\begin{aligned} i = 1 : (c_{1t} - (1 - T^{-1}b)c_{1,t-1}) &= \beta_1 (c_{1,t-1} - (1 - T^{-1}b)c_{1,t-2}) + \eta_{1t} \\ i \geq 2 : c_{it} &= \alpha_i c_{i,t-1} + \eta_{it} \end{aligned}$$

in a similar manner.

<sup>9</sup> The results are robust to the choice of different autoregressive orders.

**Table 3**

Rejection rates for the test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ . Rejections rates in the columns in which  $a = 0$  or  $b = 0$  compare with a nominal asymptotic size of 5%, while those in the remaining columns show power of the tests against the stated alternatives.

$\tau_\kappa^T$	$T$	$b$					
		$H_0^T : n = \kappa$			$H_1^T : n = \kappa - 1$		
		0	10	20	30	40	50
$\kappa = 2$	100	0.0200	0.0616	0.2570	0.6054	0.8424	0.9288
	200	0.0376	0.1200	0.4768	0.8558	0.9774	0.9992
	300	0.0408	0.1570	0.5524	0.9062	0.9916	0.9998
$\kappa = 1$	100	0.0422	0.3888	0.8564	0.9842	0.9984	0.9996
	200	0.0462	0.4648	0.9326	0.9986	1.0000	1.0000
	300	0.0548	0.4938	0.9440	0.9990	1.0000	1.0000
$\sigma_\kappa^T$	$T$	$a$					
		$H_0^T : m = 0$		$H_1^T : m = 1$			
		0	0.1	0.5	1	1.5	2
$\kappa = 2$ ( $= n$ )	100	0.0426	0.0450	0.1332	0.4652	0.8188	0.9670
	200	0.0440	0.0468	0.1270	0.4618	0.8190	0.9776
	300	0.0448	0.0534	0.1298	0.4570	0.8196	0.9748
$\kappa = 1$ ( $= n$ )	100	0.0442	0.0508	0.1754	0.5366	0.8814	0.9894
	200	0.0504	0.0584	0.1874	0.5690	0.8932	0.9894
	300	0.0514	0.0526	0.1788	0.5726	0.8974	0.9880

For the case of  $H_0^T : m = 0$  vs.  $H_1^T : m = 1$  with  $n = 2$ , we utilize

$$i = 1 : \Delta(c_{1t} - T^{-1/2}a) = \beta_1 \Delta(c_{1,t-1} - T^{-1/2}a) + \eta_{1t}$$

$$i = 2 : \Delta c_{2t} = \beta_2 \Delta c_{2,t-1} + \eta_{2t}$$

$$i \geq 3 : c_{it} = \alpha_i c_{i,t-1} + \eta_{it}$$

similarly. That is, we impose the null of  $a = 0$  to estimate  $\beta_1$ ,  $\beta_2$ , and  $\alpha_i$  for  $i \geq 3$ . We estimate  $\alpha_1$  from the first coordinate process with  $\Delta c_{1t} = \alpha_1 \Delta c_{1,t-1} + e_{2t}$  and then set  $a \approx \hat{\alpha}_1 T^{1/2}$ . In this case, with  $T = 100, 200, 300$ , we set  $a \in [0.1, 2]$  under the alternative.

The case of  $H_0^T : m = 0$  vs.  $H_1^T : m = 1$  with  $n = 1$ , is implemented using

$$i = 1 : \Delta(c_{1t} - T^{-1/2}a) = \beta_1 \Delta(c_{1,t-1} - T^{-1/2}a) + \eta_{1t}$$

$$i \geq 2 : c_{it} = \alpha_i c_{i,t-1} + \eta_{it}$$

similarly to that of  $H_0^T : m = 0$  vs.  $H_1^T : m = 1$  with  $n = 2$ .

Similarly to the finite-sample results of Chang et al. (2016), we find that the test based on  $\tau_\kappa^T$  has good size properties and is conservative in smaller samples (Table 3). As expected, power increases with  $a$  and is excellent for all but the closest local alternatives. Note that we set  $\kappa = n$  for  $\sigma_\kappa^T$ , with the idea that it may be (correctly) selected using the test based on  $\tau_\kappa^T$ . As discussed above, we expect the size to be asymptotically correct as long as  $\kappa \geq n$ . The results are also favorable for the test based on  $\sigma_\kappa^T$ . In particular, because of the rapid divergence under the alternative – part (b) of Theorem 2.5 – the power of the test is not affected very much by the sample size. In sum, the tests based on  $\tau_\kappa^T$  and  $\sigma_\kappa^T$  perform very well even in relatively small samples.

Simulations under the alternative hypothesis  $H_1^T : n = 0$  are informative about estimation error of the unit root proportions in the moments. Table 4 shows the mean from estimating the unit root proportions of the first four moments for  $T = 100, 200, 300$  and  $b \in [10, 50]$ , in which case the true unit root proportion is zero. The biases are approximately 0.06, 0.19, 0.02, and 0.07 for unit root proportions of the first four moments using time series with spans of  $T = 100$  and  $T = 200$  similar to the span ( $T = 167$ ) of our temperature anomaly series. Because the estimators are consistent, we may interpret these means as measures of small sample biases. Estimation of the unit root proportion in the second moment appears to be more biased than that in the other moments, but by using local alternatives in the form of  $(1 - T^{-1}b)$ , we expect these to be upper bounds on the bias.

### 3. Persistent features in temperature anomalies

In this section, we discuss how to implement the tests and create the proportions discussed above using actual data, and we present the results for the temperature anomaly distributions. We then show unit root proportions and graphical representations of the stationary and nonstationary components.

**Table 4**

Means of the estimated unit root proportions. The simulated unit root proportion is zero in all cases, and the means are drawn from the estimates of the proportions using 5000 simulated samples.

	$T$	$b$				
		10	20	30	40	50
$\hat{\pi}_{1T}^N = 0$	100	0.0668	0.0670	0.0688	0.0692	0.0688
	200	0.0463	0.0460	0.0472	0.0468	0.0463
	300	0.0425	0.0422	0.0426	0.0420	0.0421
$\hat{\pi}_{2T}^N = 0$	100	0.1806	0.1811	0.1801	0.1780	0.1791
	200	0.1893	0.1895	0.1893	0.1891	0.1894
	300	0.1904	0.1904	0.1905	0.1907	0.1907
$\hat{\pi}_{3T}^N = 0$	100	0.0176	0.0169	0.0187	0.0186	0.0179
	200	0.0106	0.0107	0.0108	0.0106	0.0107
	300	0.0077	0.0078	0.0078	0.0079	0.0076
$\hat{\pi}_{4T}^N = 0$	100	0.0668	0.0678	0.0672	0.0654	0.0665
	200	0.0714	0.0715	0.0714	0.0713	0.0714
	300	0.0718	0.0719	0.0719	0.0720	0.0720

### 3.1. Empirical implementation of the tests and proportions

To implement our methodology, we use the cross-sectional densities that we regard as functional observations on the Hilbert space  $H$  introduced in (2). In our analysis,  $H$  has a countable basis. This implies that any  $w \in H$  can be represented as an infinite linear combination of the basis elements, and that the representation is unique. Therefore, there is a one-to-one correspondence between  $H$  and  $\ell^2(\mathbb{R})$  — i.e., the set of all infinite sequences of real numbers that are square summable, and the correspondence is uniquely defined, once the basis elements are fixed.

For instance, once an orthonormal basis  $(\phi_i)$  is given, we may write any  $w \in H$  as  $w = \sum_{i=1}^{\infty} c_i \phi_i$  with  $c_i = \langle \phi_i, w \rangle$  uniquely and establish the correspondence  $w \leftrightarrow (c_i)$ . Of course, the correspondence becomes operational only if we replace  $\ell^2(\mathbb{R})$  by  $\mathbb{R}^M$  for some large  $M$ . Subsequently, we let  $[w] = (c_1, \dots, c_M)'$  and define a correspondence

$$w \leftrightarrow [w] \quad (18)$$

between  $H$  and  $\mathbb{R}^M$  in place of  $\mathbb{R}^\infty$ . In our analysis, we use a Daubechies wavelet basis for  $(\phi_i)$  and set  $M = 1037$ , which we believe to be sufficiently large. Using different bases, such as the Haar wavelet basis, generates results that differ by no more than a few decimal places.

Under the correspondence between  $H$  and  $\mathbb{R}^M$  defined in (18), we have the correspondences  $\langle v, w \rangle \leftrightarrow [v]'[w]$  and  $v \otimes w \leftrightarrow [v][w]'$  for any  $v, w \in H$ . In fact, under the correspondence in (18), the linear operator  $Q$  on  $H$  defined in (4) generally corresponds to a square matrix of dimension  $M$  denoted by  $[Q]$ , and we have in particular  $\langle v, Qw \rangle \leftrightarrow [v]'[Q][w]$  for any  $v, w \in H$ .

For ease of reference and clarity of exposition, we briefly outline seven steps utilized to create the test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$  from a finite sample.

1. **Obtain**  $w_t$ . We represent  $w_t$  as an  $M$ -dimensional vector  $[w_t]$  for each  $t$ , using  $M = 1037$  Daubechies wavelet basis functions.
2. **Create**  $Q_T$ . Implement  $Q_T = \sum_{t=1}^T w_t \otimes w_t$  as  $[Q_T] = \sum_{t=1}^T [w_t][w_t]'$ .
3. **Calculate**  $v_i(Q_T)$ . We identify these as  $[v_i(Q_T)]$ , which are  $M$  orthonormal eigenvectors of the  $M$ -dimensional square matrix  $[Q_T]$ .
4. **Create**  $z_t$  from (7). Inner products  $\langle v_i(Q_T), w_t \rangle$  are computed as  $[v_i(Q_T)]'[w_t]$  for each  $i$  and  $t$ .
5. **Create**  $Q_\kappa^T$  and  $\Omega_\kappa^T$ . Implement  $Q_\kappa^T = \sum_{t=1}^T z_t z_t'$  and  $\Omega_\kappa^T = \sum_{|k| \leq \ell} \varpi_\ell(k) \Gamma_T(k)$  using the Parzen window with Andrews plug-in bandwidth.
6. **Calculate**  $\lambda(Q_\kappa^T, \Omega_\kappa^T)$ . These are generalized eigenvalues of  $Q_\kappa^T$  with respect to  $\Omega_\kappa^T$ .
7. **Calculate Test Statistics**  $\tau_\kappa^T$  from (8) and  $\sigma_\kappa^T$  from (9).

Once these test statistics are calculated, the ranks of the respective spaces are chosen using the sequential procedure described above.

The nonstationary and unit root proportions of the  $i$ th moment defined in (15) and (16) cannot be calculated directly, because  $H_N$  and  $H_U$  are unknown. Instead, we may use the *sample* nonstationary and unit root proportions of the  $i$ th cross-sectional moment

$$\hat{\pi}_{iT}^N = \sqrt{\frac{\sum_{j=1}^n \langle \mu_i, v_j(Q_T) \rangle^2}{\sum_{j=1}^M \langle \mu_i, v_j(Q_T) \rangle^2}} \quad \text{and} \quad \hat{\pi}_{iT}^U = \sqrt{\frac{\sum_{j=m+1}^n \langle \mu_i, v_j(Q_T) \rangle^2}{\sum_{j=1}^M \langle \mu_i, v_j(Q_T) \rangle^2}} \quad (19)$$



**Table 5**  
Test statistics  $\tau_\kappa^T$  and  $\sigma_\kappa^T$ . Global, NH, and SH temperature anomaly distributions.

	$\kappa$	1	2	3	4
Global	$\tau_\kappa^T$	0.0733	0.0365	0.0110	0.0109
	( <i>p-value</i> )	(0.2688)	(0.2995)	(0.0039)	(0.0212)
	$\sigma_\kappa^T$	0.0733	0.0771	0.0993	0.0940
	( <i>p-value</i> )	(0.7314)	(0.9423)	(0.9724)	(0.9964)
NH	$\tau_\kappa^T$	0.0558	0.0521	0.0129	0.0121
	( <i>p-value</i> )	(0.1599)	(0.5662)	(0.0145)	(0.0465)
	$\sigma_\kappa^T$	0.0558	0.0583	0.0690	0.0720
	( <i>p-value</i> )	(0.8403)	(0.9808)	(0.9956)	(0.9995)
SH	$\tau_\kappa^T$	0.0787	0.0233	0.0090	0.0085
	( <i>p-value</i> )	(0.3009)	(0.0702)	(0.0001)	(0.0019)
	$\sigma_\kappa^T$	0.0787	0.0774	0.0913	0.0946
	( <i>p-value</i> )	(0.6992)	(0.9415)	(0.9809)	(0.9963)

to estimate  $\pi_i^N$  and  $\pi_i^U$ . Chang et al. (2016) show that the sample nonstationary proportion  $\hat{\pi}_{it}^N$  is a consistent estimator of the original nonstationary proportion  $\pi_i^N$ , and by extension  $\hat{\pi}_{it}^U$  is a consistent estimator of  $\pi_i^U$ .

### 3.2. Test statistics

Using the  $\tau_\kappa^T$  and  $\sigma_\kappa^T$  test statistics and *p*-values from Table 5 and starting with  $\tau_\kappa^T$  with  $\kappa = 4$  for the global distribution, we reject  $n = 4$  against the alternative  $n \leq 3$  and then reject  $n = 3$  against the alternative  $n \leq 2$ , both with *p*-values less than 5%. Using  $\tau_\kappa^T$  with  $\kappa = 2$ , the *p*-value jumps to about 30%, so we cannot reject  $n = 2$  against  $n \leq 1$ . We obtain the same results for the NH distribution, though the failure to reject  $n = 2$  is even more decisive. For the SH,  $n = 4$  and  $n = 3$  are strongly rejected with *p*-values less than 1%,  $n = 2$  is rejected with a *p*-value of 7%, but  $n = 1$  is not rejected against  $n = 0$ .

We therefore choose the dimension of the nonstationary subspace  $\dim(H_N)$  to be  $n = 2$  for the NH and the globe, but  $n = 1$  for the SH. We may interpret the nondegenerate dimension of the nonstationary subspace to mean that all three series of distributions have some persistence that is strong enough to be *permanent* in the sense that shocks to the temperature anomaly distributions accumulate over time. Changes in the temperature anomaly distributions are not entirely *transitory*.

To determine whether the persistence is of the unit root type or whether it is explosive or deterministic, we examine the test based on  $\sigma_\kappa^T$ . Regardless of the value of  $\kappa$  that we choose for the global distribution,  $m = 0$  is not rejected in favor of  $m \geq 1$  at any reasonable significance level. This is strong evidence for  $m = 0$ . The test yields the same outcome ( $m = 0$ ) again rather decisively, for both the NH and the SH. The test results do not depend upon the choice of the value  $\kappa$ . The fact that  $m = 0$  is not rejected in every case, with *p*-values all exceeding 69% and some exceeding 99%, suggests that all of the nonstationarity is better characterized by unit-root-type persistence than by higher-order persistence, which suggests stochastic trends are present in the moments of the distributions as opposed to explosive roots or linear deterministic trends.

Brohan et al. (2006) note that some types of measurement errors that are expected in the land surface data are strongly correlated over time. Could such correlations be strong enough and large enough to generate the second unit root that we detect in the NH, which has more land surface area? In other words, is there only one unit root in each hemisphere?

To address this question, we repeat the analysis (results not shown) on a sample that starts in 1900 with the idea that uncertainties in the earlier data would be larger (see Brohan et al., 2006, Figure 10, e.g.). Contrary to expectations about the importance of temporally correlated measurement errors, the two unit roots in the NH remain, but the SH contains a second unit root. These findings suggest that the uncertainty from the 19th Century data either obscures or eliminates one of the unit roots in the SH. We favor the latter interpretation – that there is only one unit root in the SH – because including more data increases the ability to reject two unit roots in favor of one.

### 3.3. Estimated proportions

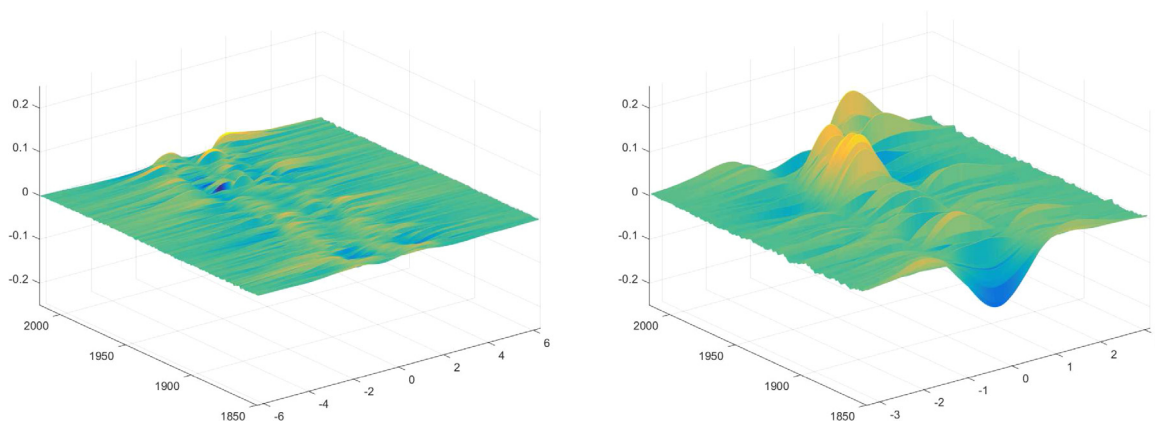
We now turn to the proportions of the subspaces defined above in each cross-sectional moment of temperature anomalies. Note that we set  $\pi_i^U = \pi_i^N$  – i.e., the unit root space spans the entire nonstationary space – because we do not find evidence of any higher-order persistence. Table 6 shows consistent estimates  $\hat{\pi}_{it}^N$  of the proportion of the nonstationary subspace  $\pi_i^N$  in each of the first seven cross-sectional moments of temperature anomalies. The residual proportions of each moment are in the stationary subspace  $\pi_i^S$ .

To explain, a proportion of one third means that a third of the temporal variation in a given moment of the distributions is due to a small number of unit root processes, while the remaining two thirds are due to a large number of stationary processes. Even though the proportion of the variation may be dominated by a large number of stationary processes,

**Table 6**

Sample nonstationary proportions in the first seven moments. Global, NH, and SH temperature anomaly distributions.

	$\hat{\pi}_{1T}^N$	$\hat{\pi}_{2T}^N$	$\hat{\pi}_{3T}^N$	$\hat{\pi}_{4T}^N$	$\hat{\pi}_{5T}^N$	$\hat{\pi}_{6T}^N$	$\hat{\pi}_{7T}^N$
Global	0.3054	0.1933	0.0854	0.0722	0.0442	0.0402	0.0310
NH	0.3127	0.1897	0.1013	0.0741	0.0586	0.0434	0.0433
SH	0.4559	0.3922	0.1788	0.2046	0.1047	0.1327	0.0764

**Fig. 5.** Stationary components. Stationary components ( $w_t^S$ ) for the NH (left) and SH (right).

none of these processes contribute to stochastically trending behavior. They may, however, make it difficult to detect such trends.

Roughly a third (30.5%) of the variation in the mean global mean is strong enough to be unit-root-type persistence. The persistence in the mean appears to be stronger in the SH than the NH, in the sense that 45.6% of the persistence in the Southern mean is of the unit root type, while only 31.3% of that north of the Equator. Recall from Table 4 that for a sample size of between  $T = 100$  and  $T = 200$ , we expect small-sample bias to be no more than 6%.

The global, NH, and SH nonstationary proportions of the second moments are 19.3%, 19.0%, and 39.2% respectively, which suggests that nonstationary it is less evident in the variance than in the mean in both hemispheres. We expect bias to be no more than 19% from Table 4, which suggests that the proportion could be as low as zero in the NH, but could still be as high as 20.2% in the SH.

The proportion of unit root persistence in the third moment for the globe is 8.5%. Like the first two moments, the unit root proportion appears to be less in the NH (10.1%) than in the SH (17.9%). Moreover, like the first moment but unlike the second, our simulation results in Table 4 suggest that the unit root proportion in the third moment is estimated with very little small sample bias.

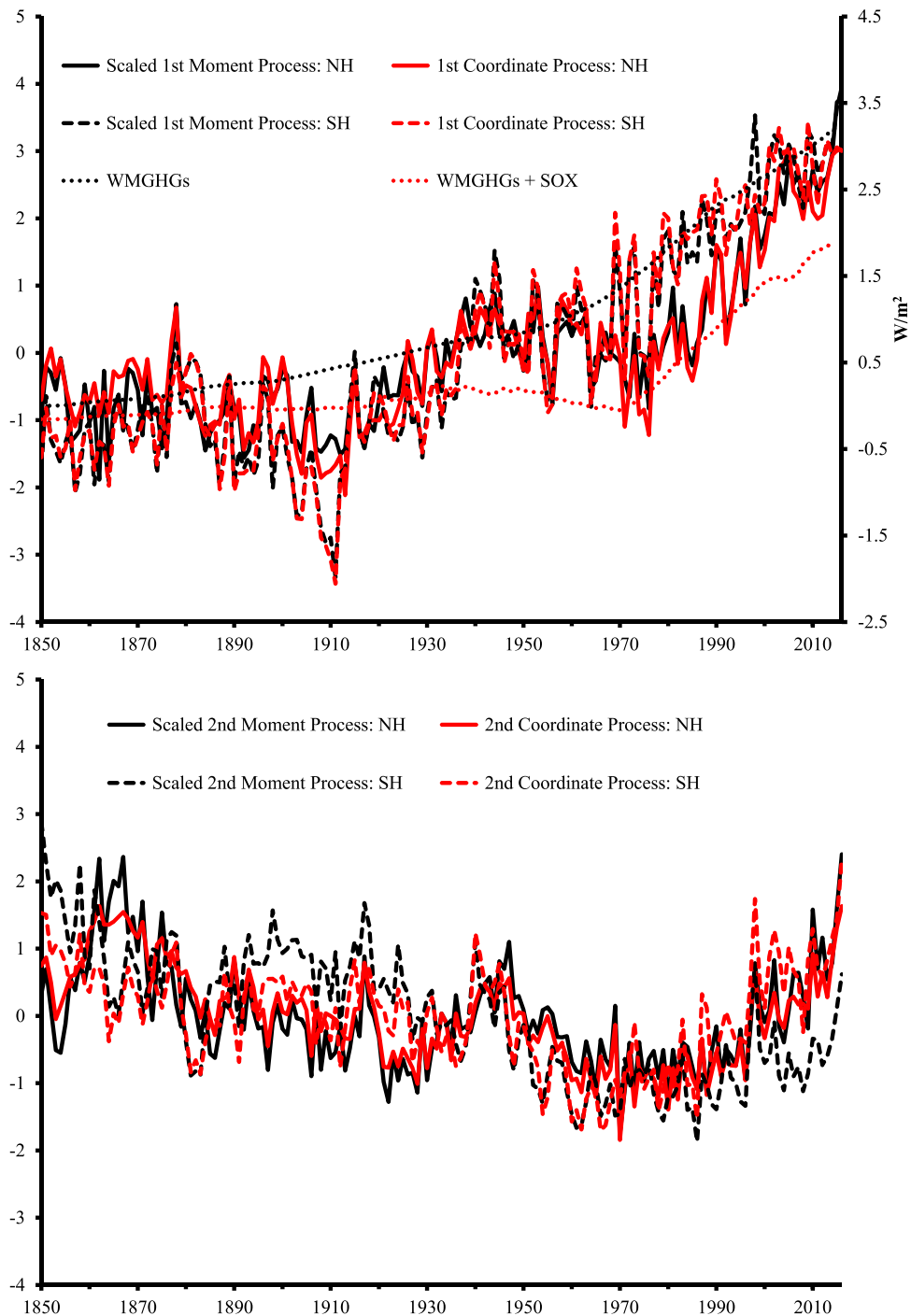
Persistence appears to decline in the variation of the remaining four moments, albeit not monotonically in the SH. The bias associated with the unit root proportion in the fourth moments are qualitatively similar to that in the second: those for the globe and NH could be as low as zero, while that in the SH is evidently nonzero.

Overall, the unit root(s) appear more prominently in the SH than in the NH. This does not suggest that climate change is any less persistent in the NH, but simply that the persistence is harder to detect in the NH, which is roughly consistent with the findings of Zheng and Basher (1999) for the mean.

### 3.4. Estimated components

Because the concepts of stationarity and nonstationarity of densities are quite new, we show the time series of stationary components ( $w_t^S$ ) of the densities ( $w_t$ ) for the NH and SH (Fig. 5). The series for the globe is similar to that of the NH and not shown for brevity. These are calculated by subtracting the estimated nonstationary component ( $w_t^N$ ), calculated as  $w_t^N = \Pi_N w_t$ , from the densities ( $w_t$ ). Recall that the dimension  $n$  is estimated to be two for the globe and NH and one for the SH with  $m = 0$  for both hemispheres. In both hemispheres – but especially in the NH – the stationary components of the densities appear to be more like random noise, showing very little evidence of persistence in any of the moments. This finding suggests that the temporal patterns in the densities in the top right panels of Figs. 2–4 are driven by their nonstationary components rather than by their stationary components.

The concept of nonstationarity is likely to be more familiar to readers in a simple time series context. To this end, we show the first two coordinate processes estimated for each hemisphere:  $\langle v_1, w_t \rangle$  and  $\langle v_2, w_t \rangle$ , or  $\langle c_{1t} \rangle$  and  $\langle c_{2t} \rangle$  using the correspondence in (18) (Fig. 6). We also plot mean and second moment processes  $\langle \mu_1, w_t \rangle$  and  $\langle \mu_2, w_t \rangle$  for each



**Fig. 6.** First two coordinate processes and first two moment processes. First coordinate processes (top) and second coordinate processes (bottom). Normalized first moment process scaled by  $1/\hat{\pi}_{1T}^U$  (top) and normalized second moment process scaled by  $1/\hat{\pi}_{2T}^U$  (bottom). Top panel shows forcings from well-mixed greenhouse gas (WMGHG) emissions (Hansen et al., 2017) and WMGHG emissions plus sulfur emissions (SOX). The latter series is due to Kaufmann and Stern (2002) prior to 1990 and the Joint Global Change Research Institute subsequently.

hemisphere, which are scaled for visual comparison as follows. First, we normalize the processes to unit length with a Euclidean norm, because the eigenvectors used to compute the nonstationary coordinate processes have unit length. Then we scale the normalized mean process of each hemisphere by  $1/\hat{\pi}_{1T}^U$  and the normalized second moment process of each hemisphere by  $1/\hat{\pi}_{2T}^U$ . The figure also shows two series of anthropogenic forcings, that from well-mixed greenhouse

gases and that from well-mixed greenhouse gases net of that from emissions of sulfur, which will be discussed in the next section.

There is a striking similarity between the first coordinate processes, which contain a unit root in both hemispheres, and the scaled mean processes. The nonstationary proportions (Table 6) indicate that 30–45% of the variation in the mean process is due to unit roots, and nearly all of that variation appears to come from the first unit root (Fig. 6).

Furthermore, there is a strong similarity between the second coordinate process and the scaled second moment process in the NH. About 19% of the variation in the second moment process is due to unit roots (Table 6), and again it appears that nearly all of it comes from the second coordinate process. There is a noticeable disparity in the SH. The disparity is not surprising, because an estimated 39% of the variation in the second moment process is estimated to come from the first – not the second – coordinate process. We find the second coordinate processes to be stationary.

Fig. 6 suggests that the first coordinate processes in each hemisphere are cointegrated – i.e., they have common stochastic trends – with the exception of a persistent gap between about 1970 and 2000. If the first coordinate processes in each hemisphere are distinct, we would expect to find three unit roots in the global distributions. But only two are present, which is consistent with some cointegrating linear combination of three nonstationary coordinate processes. This finding is consistent with Kaufmann and Stern (2002), who show that mean temperatures in the NH and SH are similar, but do not cointegrate, which may be associated with this gap.

#### 4. Empirical findings in context

A more in-depth discussion of our empirical results follows. Specifically, we link the detected persistence in the moments of the global and hemispheric distributions to findings in the extant literature on trend detection in the temperature series, and we compare hemispheric differences that are detected by our tests to hemispheric differences in human activities that emit radiatively active gases.

##### 4.1. Anthropogenic fingerprints on the detected trends

The estimated nonstationary coordinate processes clearly illustrate strong persistence in the distributions of both hemispheres. However, our results do not support the hypothesis that temperature contains a deterministic trend. Instead, the close correspondence of the mean processes with unit root processes in each hemisphere supports the hypothesis that changes in mean temperatures follow a stochastic trend.

The presence of a stochastic trend does not mean that temperatures follow a random walk, as Gordon (1991) argues. A random walk is a very special case of a unit root process with increments that are completely unpredictable. Instead, increments with strong but stationary persistence can generate a unit root process, indicating that changes in global mean temperatures have long-lasting effects. Such long-lasting effects may create the appearance of a linear or broken trend.

The presence of a stochastic trend is consistent with the theory of anthropogenic climate change. According to this hypothesis, the stochastic trends in mean temperatures originate from stochastic trends in radiative forcings, which originate from stochastic trends in the capital stock and long residence times in the atmosphere. So-called well-mixed greenhouse gases (WMGHGs), such as CO<sub>2</sub> and methane, have a particularly long residence time: decades to centuries (Ramaswamy et al., 2001). Their long residence time means that the atmosphere integrates WMGHG emissions such that radiative forcing due to these emissions is highly persistent (Stern and Kaufmann, 2000; Kaufmann et al., 2006). The long residence time also means that concentrations are “well-mixed” across the hemispheres.<sup>10</sup>

In addition to the mean processes and first coordinate processes of each hemisphere, the top panel of Fig. 6 also plots radiative forcing due to WMGHGs (Hansen et al., 2017). Visual comparisons strongly suggest that these series contain a common stochastic trend. More formally, Engle–Granger tests reject the hypothesis that the first coordinate process in each hemisphere does not cointegrate with the forcing from WMGHGs.<sup>11</sup> This result is consistent with the hypothesis that the increase in radiative forcing that is associated with anthropogenic emissions of WMGHGs is partially responsible for changes in temperature. However, the statistical finding of cointegration does not identify the direction of causality.

##### 4.2. Hemispheric differences in the first coordinate process

The first coordinate processes in Fig. 6 have a gap between hemispheres from about 1970 to 2000. During this period, the NH is cool relative to the SH, despite the previous results that both temperature series cointegrate with forcing from WMGHGs.

A cooler NH may be associated with anthropogenic sulfur emissions. Most of the human population and its economic activity are located in the NH, which creates a natural experiment in which the atmosphere above the NH receives the greatest emissions of both WMGHGs and sulfur. Anthropogenic sulfur emissions have a relatively short lifetime in the atmosphere, about a week to ten days (Shine et al., 1991), which means that they do not mix thoroughly across the

<sup>10</sup> For example, the average concentration of CO<sub>2</sub> at Barrow, Alaska (71.3230° North, 156.6114° West) in 2014 was 400 ppm; it was 395 ppm at Cape Point, South Africa (34.3523° South, 18.4891° East) (<http://www.esrl.noaa.gov/gmd/>).

<sup>11</sup> The test statistics take values of –4.88 for the NH and –6.46 for the SH, compared to a 1% critical value of –3.96.

hemispheres. The annual mean for the direct radiative forcing of anthropogenic sulfates in the NH is 2–7 times greater than in the SH (Ramasmamy et al., 2001). As a result, the cooling effects of these emissions are stronger in the NH than in the SH.

A plot of forcing from WMGHGs against that from WMGHGs net of forcing from anthropogenic sulfur emissions (Fig. 6, top panel) shows a similar gap. Legislation to reduce acid deposition in North America, Europe, and Japan reduced sulfur emissions in the 1970s, which reduced their cooling effect. Consistent with this timeline, the cooling effect levels off, so that the NH warms faster than that in the SH.

According to this interpretation, persistence in the first coordinate process is more salient in the SH, where 45% of the persistence in the Southern hemisphere is of a unit root type, compared to only 30% in the NH. This smaller percentage implies that there is more stationary variation in the NH. We postulate that the highly persistent warming associated with WMGHGs (and embodied in the first coordinate process of the SH) is partly offset by the cooling from anthropogenic sulfur emissions. Although consistent with hemispheric differences in anthropogenic activity, we recognize alternative explanations are possible.

#### 4.3. Hemispheric differences in the second coordinate process

Beyond the first coordinate process that is present in both hemispheres, the NH contains a second unit root. The bottom panel of Fig. 6 shows both the rescaled second coordinate processes and the second moment processes for each hemisphere. These two processes are quite similar in the NH. The second moment process gently declines until about 1945, after which it abruptly declines until about 1970. We speculate that this pattern may be caused by a well-documented change in instruments used to sample ocean surface temperatures precipitated by World War II, which decreased variations in sampled temperatures. (See Folland and Parker, 1995, especially their Figure 5.)

From about 1970 onward, the second unit root process and second moment process in the NH increase. Such an increase is consistent with spatially disaggregated but not well-mixed sulfur emissions. In some areas of the NH, such as Southeast Asia and parts of Europe, anthropogenic sulfur emissions reduce total forcing even relative to pre-industrial levels (Myhre et al., 2013). These spatial variations in forcing are important because they affect local rates of temperature change (Magnus et al., 2011), which could account for the increase in the second moment process in the NH relative to the SH. Again, alternative explanations are possible.

#### 4.4. Higher-order nonstationarity and a runaway greenhouse gas effect

The fact that we do not detect any higher-order nonstationarity argues strongly against a ‘runaway greenhouse gas effect’ in the historical record. According to this hypothesis, rising temperatures trigger a positive feedback loop that increases greenhouse gas concentrations by reducing the solubility of carbon dioxide in seawater (Woolf et al., 2016), changing the balance between respiration and photosynthesis in terrestrial biota (Davidson and Janssens, 2006), and/or releasing methane from methane hydrates (Archer, 2007) and thereby raising temperatures further. Consistent with this possibility, annual variations in the flux of carbon from soils to the atmosphere are positively correlated with mean annual temperature (Raich et al., 2002). Similarly, statistical analyses indicate that increases in global temperature have a small positive effect on atmospheric CO<sub>2</sub> (Keeling et al., 1989; Kaufmann et al., 2006).

Triggering a positive feedback loop would increase the persistence of temperature means. But we find no evidence for such higher levels of persistence. Furthermore, hemispheric differences in persistence are inconsistent with some components of the positive feedback loop. Most of the world’s terrestrial biota is located in the NH, which also holds the greatest reservoir of methane hydrates that could melt in response to higher temperatures. If triggered, such changes would increase the persistence of the NH temperature anomalies relative to those for the SH. Yet we find persistence in mean temperatures to be more evident in the SH.

### 5. Conclusion

To increase the ability to detect trends in the global and hemispheric distributions of temperature anomalies, we extend the test for unit roots against stationary roots in time series of distributions proposed by Chang et al. (2016). We not only re-examine the null of their test when the distributions contain a deterministic trend or have an explosive root in addition to the unit root, but we also introduce a new test and testing procedure to distinguish unit roots from these higher-order nonstationarity.

Our empirical results directly support one unit root process (stochastic trend) in the Southern Hemisphere and two unit root processes in the Northern Hemisphere and the globe, with no evidence for higher-order processes in any of the moments of any of these distributions over time. As such, we do not find evidence for a runaway greenhouse gas effect over the span of our data. We acknowledge the difficulty in distinguishing between a stochastic trend and a deterministic trend with breaks using statistical techniques alone.

The absence of higher-order trends is in line with studies that use cointegration methods to relate differences in hemispheric temperature to hemispheric differences in forcing due to differences in anthropogenic emissions. Differences in the number of stochastic trends between the hemispheres are consistent with hemispheric differences in human activity. Generalized warming due to the emission of greenhouse gases in the NH may be countervailed by localized cooling due to sulfur emissions, decreasing the spatial mean but increasing the spatial variance relative to the SH, which lacks the countervailing effect of sulfur emissions.

## Appendix. Proofs of Theorems and Corollaries

**Proof of Theorem 2.2.** Let  $n = \kappa$  and  $m = 0$ . Under Part (a) of [Assumption 2.1](#), we may show exactly as in the proof of Theorem 4.3 of [Chang et al. \(2016\)](#) that

$$T^{-2}\lambda_i(Q_\kappa^T, \Omega_\kappa^T) \rightarrow_d \lambda_i(Q_\kappa)$$

as  $T \rightarrow \infty$ , for  $i = 1, \dots, \kappa$ , and the stated results follow immediately.  $\square$

**Proof of Corollary 2.3.** Let  $n = \kappa$  and  $m \geq 1$ . For  $i = 1, \dots, m$ , we have

$$T^{-2}\lambda_i(Q_\kappa^T, \Omega_\kappa^T) = T^{-2}\lambda_i(Q_\kappa^T) + o_p(1),$$

because  $\langle v_i, \Omega_\kappa^T v_i \rangle = o_p(\langle v_i, Q_\kappa^T v_i \rangle)$  for all  $i = 1, \dots, m$  and

$$\min_{1 \leq i \leq m} \langle v_i, \Omega_\kappa^T v_i \rangle \rightarrow_p \infty$$

as  $T \rightarrow \infty$  by part (b) of [Assumption 2.1](#). However, we may show that

$$T^{-2}\lambda_i(Q_\kappa^T) \rightarrow_p \infty$$

as  $T \rightarrow \infty$  for  $i = 1, \dots, m$ , due to part (b) of [Assumption 2.1](#). This can be shown similarly to the proofs of Lemma 3.1 and Theorem 3.3 of [Chang et al. \(2016\)](#).

On the other hand, for  $i = m+1, \dots, n$ ,  $\lambda_i(Q_\kappa^T, \Omega_\kappa^T)$  asymptotically behaves as  $\lambda_i(Q_\kappa^{\circ T}, \Omega_\kappa^{\circ T})$ , where  $Q_\kappa^{\circ T}$  and  $\Omega_\kappa^{\circ T}$  are defined from  $(w_i^\circ)$  in (12) exactly in the same way as  $Q_\kappa^T$  and  $\Omega_\kappa^T$  are defined from  $(w_t)$ . In fact, we may easily show that

$$T^{-2}\lambda_i(Q_\kappa^T, \Omega_\kappa^T) = T^{-2}\lambda_i(Q_\kappa^{\circ T}, \Omega_\kappa^{\circ T}) + o_p(1).$$

However, for  $i = m+1, \dots, \kappa$ , we have

$$T^{-2}\lambda_i(Q_\kappa^{\circ T}, \Omega_\kappa^{\circ T}) \rightarrow_d \lambda_i(Q_{\kappa-m})$$

as  $T \rightarrow \infty$ , as in the proof of [Theorem 2.2](#). The stated results may therefore be readily deduced.  $\square$

**Proof of Corollary 2.4.** Part (b) is shown by [Chang et al. \(2016\)](#). For part (a), let  $0 < n \leq \kappa - 1$  and  $m = 0$ . Although this case is not considered explicitly by [Chang et al. \(2016\)](#), we may deduce from their asymptotic theory that

$$T^{-2}\lambda_i(Q_\kappa^T, \Omega_\kappa^T) \rightarrow_d \lambda_i(Q_n)$$

as  $T \rightarrow \infty$ , for  $i = n+1, \dots, \kappa$ , from which the stated result follows.  $\square$

**Proof of Theorem 2.5.** Part (a) is shown in Theorem 4.3 of [Chang et al. \(2016\)](#). Part (b) follows from the proof of [Corollary 2.3](#).  $\square$

## References

- Archer, D., 2007. Methane hydrate stability and anthropogenic climate change. *Biogeosciences* 4, 521–544.
- Baillie, R.T., Chung, S.-K., 2002. Modeling and forecasting from trend-stationary long memory models with applications to climatology. *Int. J. Forecast.* 18, 215–226.
- Ballester, J., Giorgi, F., Rodó, X., 2010. Changes in european temperature extremes can be predicted from changes in PDF central statistics. *Clim. Change* 98, 277–284.
- Beare, B.K., Seo, J., Seo, W.K., 2017. Cointegrated linear processes in Hilbert space. *J. Time Series Anal.* 38, 1010–1027.
- Bloomfield, P., 1992. Trends in global temperature. *Clim. Change* 21, 1–16.
- Bloomfield, P., Nychka, D., 1992. Climate spectra and detecting climate change. *Clim. Change* 21, 275–287.
- Bosq, D., 2000. Linear Processes in Function Spaces. In: *Lecture Notes in Statistics*, vol. 149, Springer, New York.
- Brock, W.A., Engström, G., Xepapadeas, A., 2013. Energy balance climate models and general equilibrium optimal mitigation policies. *J. Econ. Dyn. Control* 37, 2371–2396.
- Brohan, P., Kennedy, J.J., Harris, I., Tett, S.F.B., Jones, P.D., 2006. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.: Atmos.* 111, D12106.
- Castruccio, S., McInerney, D.J., Stein, M.L., Crouch, F.L., Jacob, R.L., Moyer, E.J., 2014. Statistical emulation of climate model projections based on precomputed GCM runs. *J. Clim.* 27, 1829–1844.
- Castruccio, S., Stein, M.L., 2013. Global space-time models for climate ensembles. *Ann. Appl. Stat.* 7, 1593–1611.
- Chang, Y., Kim, C.S., Park, J.Y., 2016. Nonstationarity in time series of state densities. *J. Econometrics* 192, 152–167.
- Davidson, E.A., Janssens, I.A., 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* 440, 165–173.
- 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.
- Donat, M.G., Alexander, L.V., 2012. The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.* 39, L14707.
- Estrada, F., Gay, C., Sánchez, A., 2010. A reply to “Does temperature contain a stochastic trend? Evaluating conflicting statistical results” by R. K. Kaufmann et al. *Clim. Change* 101, 407–414.
- Estrada, F., Perron, P., 2012. Breaks, Trends and the Attribution of Climate Change: A Time-Series Analysis. Department of Economics, Boston University, mimeo.



- Estrada, F., Perron, P., 2014. Detection and attribution of climate change through econometric methods. *Bol. Soc. Mat. Mexicana* 20, 107–136.
- Estrada, F., Perron, P., Gay-García, C., Martínez-López, B., 2013. A time-series analysis of the 20th century climate simulations produced for the IPCC's fourth assessment report. *PLoS One* 8, e60017.
- Folland, C.K., Parker, D.E., 1995. Correction of instrumental biases in historical sea surface temperature data. *Q. J. R. Meteorol. Soc.* 121, 319–367.
- Fomby, T.B., Vogelsang, T.J., 2002. The application of size-robust trend statistics to global-warming temperature series. *J. Clim.* 15, 117–123.
- Gao, J., Hawthorne, K., 2006. Semiparametric estimation and testing of the trend of temperature series. *Econom. J.* 9, 332–355.
- Gay-García, 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, 589–597.
- Gordon, A.H., Bye, J.A.T., Byron-Scott, R.A.D., 1996. Is global warming climate change? *Nature* 380, 478.
- Hallin, M., Lu, Z., Tran, L.T., 2004. Kernel density estimation for spatial processes: the L1 theory. *J. Multivariate Anal.* 88, 61–75.
- Hansen, J., Ruedy, R., Lo, K., 2010. Global surface temperature change. *Rev. Geophys.* 48, RG000354.
- Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K., Beerling, D.J., Cao, J., Marcott, S., Masson-Delmotte, V., Prather, M.J., Rohling, E.J., Shakun, J., Smith, P., Lacs, A., Russell, G., Ruedy, R., 2017. Young people's burden: requirement of negative CO<sub>2</sub> emissions. *Earth Syst. Dynam.* 8, 577–616.
- Hormann, S., Kidzinski, L., Hallin, M., 2015. Dynamic functional principal components. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 77, 319–348.
- Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Jun, M., Stein, M.L., 2008. Nonstationary covariance models for global data. *Ann. Appl. Stat.* 2, 1271–1289.
- Kärner, O., 1996. Global temperature deviations as a random walk. *J. Clim.* 9, 656–658.
- Kaufmann, R.K., Kauppi, H., Mann, M.L., Stock, J.H., 2011. Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proc. Natl. Acad. Sci.* 108, 11790–11793.
- Kaufmann, R.K., Kauppi, H., Mann, M.L., Stock, J.H., 2013. Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Clim. Change* 118, 729–743.
- Kaufmann, R.K., Kauppi, H., Stock, J.H., 2006. Emissions, concentrations and temperature: a time series analysis. *Clim. Change* 77, 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.
- Kaufmann, R.K., Stern, D.I., 2002. Cointegration analysis of hemispheric temperature relations. *J. Geophys. Res.* 107, ACL 8-1–ACL 8-10.
- Keeling, C.D., Bacastow, R.B., Carter, A.F., Piper, S.C., Whorf, T.P., Heimann, M., Mook, W.G., Roeloffzen, H., 1989. A three-dimensional model of atmospheric CO<sub>2</sub> transport based on observed winds: 1. Analysis of observational data. In: Peterson, D.H. (Ed.), *Aspects of Climate Variability in the Pacific and Western Americas*. pp. 165–236.
- Leeds, W.B., Moyer, E.J., Stein, M.L., 2015. Simulation of future climate under changing temporal covariance structures. *Adv. Stat. Climatol. Meteorol. Oceanogr.* 1, 1–14.
- Magnus, J.R., Melenberg, B., Muris, C., 2011. Global warming and dimming, the statistical evidence. *J. Amer. Statist. Assoc.* 106, 452–468.
- McKittrick, R., Vogelsang, T.J., 2014. HAC-robust trend comparisons among climate series with possible level shifts. *Environmetrics* 25, 528–547.
- Mills, T.C., 2009. How robust is the long-run relationship between temperature and radiative forcing? *Clim. Change* 94, 351–361.
- Morice, C.P., Kennedy, J.J., Rayner, N.A., Jones, P.D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res.* 117, D08101.
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestad, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., Zhang, H., 2013. Anthropogenic and natural radiative forcing. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp. 659–740.
- Park, J.Y., Qian, J., 2012. Functional regression of continuous state distributions. *J. Econometrics* 167, 397–412.
- Raich, J.W., Potter, C.S., Bhagawati, D., 2002. Interannual variability in global soil respiration, 1980–94. *Global Change Biol.* 8, 800–812.
- Ramaswamy, V., Boucher, O., Haigh, J., Hauglustaine, D., Haywood, J., Myhre, G., Nakajima, T., Shi, G.Y., Solomon, S., 2001. Radiative forcing of climate change. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. (Eds.), *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp. 349–416.
- Shine, K.P., Derwent, R.G., Wuebbles, D.J., Morcrette, J.-J., 1991. Radiative forcing of climate. In: Houghton, J.T., Jenkins, G.J., Ephraim, J.J. (Eds.), *Climate Change: The IPCC Scientific Assessment. Report Prepared for Intergovernmental Panel on Climate Change by Working Group I*. Cambridge University Press, Cambridge, pp. 41–68.
- Stern, D.I., Kaufmann, R.K., 2000. Is there a global warming signal in hemispheric temperature series: a structural time series approach. *Clim. Change* 47, 411–438.
- Woodward, W.A., Gray, H.L., 1993. Global warming and the problem of testing for trend in time series data. *J. Clim.* 6, 953–962.
- Woodward, W.A., Gray, H.L., 1995. Selecting a model for detecting the presence of a trend. *J. Clim.* 8, 1929–1937.
- Woolf, D.K., Land, P.E., Shutler, J.D., Goddijn-Murphy, L.M., Donlon, C.J., 2016. On the calculation of air-sea fluxes of CO<sub>2</sub> in the presence of temperature and salinity gradients. *J. Geophys. Res. Oceans* 121, 1229–1248.
- Zheng, X., Basher, R.E., 1999. Structural time series models and trend detection in global and regional temperature series. *J. Clim.* 12, 2347–2358.
- Zheng, X., Basher, R.E., Thompson, C.S., 1997. Trend detection in regional-mean temperature series: maximum, minimum, mean, diurnal range, and SST. *J. Clim.* 10, 317–326.