

# Time Series Analysis of Global Temperature Distributions: Identifying and Estimating Persistent Features in Temperature Anomalies\*

Yoosoon Chang<sup>†</sup>, Chang Sik Kim<sup>‡</sup>, J. Isaac Miller<sup>§</sup>, Joon Y. Park<sup>¶</sup>, Sungkeun Park<sup>||</sup>

## Abstract

We analyze a time series of global temperature anomaly distributions to identify and estimate persistent features in climate change. In our study, temperature densities, obtained from globally distributed data over the period from 1850 to 2012, are regarded as a time series of functional observations that are changing over time. We employ a formal test for the existence of functional unit roots in the time series of these densities. Further, we develop a new test to distinguish functional unit roots from functional deterministic trends or explosive behavior. We find some persistent features in global temperature anomalies, which are attributed in particular to significant portions of mean and variance changes in their cross-sectional distributions. We detect persistence that is characteristic of a unit root process, but none of the persistence appears to be deterministic or explosive.

This Version: June 4, 2016

JEL Classification: C13, C23, Q54

*Key words and phrases:* climate change, temperature distribution, global temperature trends, functional unit roots

---

\*The authors are grateful for useful comments to William A. “Buz” Brock, Jim Stock, and the participants of 2015 INET-Cambridge Workshop, 2014 NBER-NSF Conference, 2014 SETA, and the attendants of the seminars 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. This work was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2014S1A5B8060964). The usual caveat applies.

<sup>†</sup>Department of Economics, Indiana University

<sup>‡</sup>Department of Economics, Sungkyunkwan University

<sup>§</sup>Corresponding author. Address correspondence to J. Isaac Miller, Department of Economics, University of Missouri, 118 Professional Building, Columbia, MO 65211, or to millerjisaac@missouri.edu.

<sup>¶</sup>Department of Economics, Indiana University and Sungkyunkwan University

<sup>||</sup>Korea Institute for Industrial Economics and Trade

# 1 Introduction

Although the complicated mechanisms underpinning climate change and by which the effects of climate change are propagated lies in the realm of the physical sciences, William Nordhaus recently underscored the importance of the social sciences in studying climate change, writing that “global warming begins and ends with human activities” (2013, pg. 15). As evidenced by a proliferation of recent studies, econometricians and statisticians have a skill set that is well-suited to the statistical analysis of temporal and spatio-temporal climate data.<sup>1</sup>

Statistical models of climate change may be of interest by themselves, or they may be a natural component of a larger model to study the economic effects of climate change, as in Weitzman (2009) or Brock *et al.* (2013), for example. Questions about the permanence or transience of global temperature changes are of central importance in measuring the possible effects of economic activity on climate change or of climate change on economic activity.

Analyses of persistent time series often revolve around nonstationarity and the degree of persistence. Indeed, the evident increasing trend (nonstationarity) in the cross-sectional means of global temperature distributions over time has given rise to the popular term “global warming” to reflect climate change more generally. A large number of studies have focused on trend detection and on distinguishing a linear trend from lower-order unit root-type persistence (a stochastic trend), in particular. Studies supportive of a stochastic trend include those by Gordon (1991), Woodward and Gray (1993, 1995), Gordon *et al.* (1996), and Kärner (1996), while studies supportive of a deterministic trend with possibly highly persistent noise include those by Bloomfield (1992), Bloomfield and Nychka (1992), Baillie and Chung (2002), and Fomby and Vogelsang (2002).

The linearity of such a trend has been called into question, and nonlinearity has been introduced in the form of a quadratic trend (Woodward and Gray, 1995; Zheng and Basher, 1999), 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, 2013). Although Bloomfield (1992) uses a linear trend for testing, he emphasizes the importance of using a model-based nonlinear deterministic component. Gao and Hawthorne (2006) take nonlinearity in the deterministic trend a step further by estimating a general deterministic trend nonpara-

---

<sup>1</sup>Notable works in this area to which econometricians have contributed include Baillie and Chung (2002), Fomby and Vogelsang (2002), Gao and Hawthorne (2006), Kaufmann *et al.* (2006a, 2006b, 2010, 2013), Dong *et al.* (2008), Estrada and Perron (2012, 2014), Brock *et al.* (2013), Estrada *et al.* (2013), and McKittrick and Vogelsang (2013), to name a few.

metrically. Their estimated trend appears to have such high degrees of nonlinearity and variability that it may be approximated by a stochastic trend.

The possibility of a stochastic trend has also become important in determining a long-run relationship between global temperature and forcing variables – particularly relating to anthropogenic emissions of carbon dioxide and other gases. Kaufmann and Stern (2002) find that global or hemispheric temperature series have stochastic trends (possibly in addition to deterministic trends) and therefore they use cointegration analysis to analyze the relationship. (See also Kaufmann *et al.*, 2006a, 2006b.) Gay-Garcia *et al.* (2009) argue that the earlier findings of a stochastic trend in temperature is a statistical artifact of a broken deterministic trend instead. In contrast, Kaufmann *et al.* (2010) argue that the appearance of a broken deterministic temperature trend is inherited from the forcing variables, which they suggest may exhibit the appearance of a break due to legislation limiting air pollution in the 1980’s. The addition of weather variability makes the stochastic trend difficult to detect (Kaufmann *et al.*, 2013). Estrada *et al.* (2013) use simulated temperatures to eliminate weather variability, and they find a break.

While examination of cross-sectional means is useful, it ignores the global distributions of temperatures, as Ballester *et al.* (2010), Donat and Alexander (2012), *inter alia* point out. Moreover, Brock *et al.* (2013) underscore the importance in a broader economic model of allowing for spatial heterogeneity – specifically for the observation that anomalies have been higher in higher latitudes than near the equator (Hansen *et al.*, 2010, and consistent with the findings of Zheng and Basher, 1999, who attribute a failure to detect a deterministic trend in the northern part of the Northern Hemisphere to stronger variability in that area). Considering higher-order moments of the spatial distribution of the anomalies better allows for such heterogeneity than only looking at the means over time.

Although persistence in higher-order moments has certainly drawn less attention than the increase in the mean associated with global warming, the potential effects of such changes may be no less severe. For example, suppose that a local anomaly of a certain magnitude will melt an ice pack. Even if the global mean does not increase, an increasing kurtosis may imply that this magnitude is more likely to occur. The resulting *local* melt may have a *global* impact – for example, through higher albedo (reflection of light) or through perturbations of ocean conveyor belt circulation. These “tipping points,” as Nordhaus (2013) calls them, or “indirect carbon-cycle feedback-forcing effects,” as Weitzman (2009) calls them, may occur through climate change even without *global* warming specifically.

Recent advances in time series modeling (Bosq, 2000; Park and Qian, 2012; Chang *et al.*, 2015) allow for the analysis and detection of stationarity or nonstationarity of cross-sectional distributions, such as distributions of global temperature anomalies. Using these

tools, we may detect persistence not only in the mean, but also in higher-order moments of the global distributions of temperature anomalies.

Our tests allow for much richer temporal dynamics than recent spatio-temporal climate models that assume temporal stationarity (e.g., Castruccio and Stein, 2013) or allow for nonstationarity only in the forms of “modest” dependence (Castruccio *et al.*, 2014) or seasonal variations (Leeds *et al.*, 2015). However, because we do not model any spatial covariances, our assumptions on the spatial dimension are much stricter than the sophisticated and possibly nonstationary spatial covariances in the spatial models of Jun and Stein (2008) *inter alia*, and the recent spatio-temporal model of Castruccio and Stein (2013).

In this paper, we use a new methodology proposed by Chang *et al.* (2015) to test for nonstationarity in distributions of global industrial-era (1850-2012) temperature anomalies. Indeed, we extend their methodology to distinguish between persistence induced by unit root-type nonstationarity (a stochastic trend) from that induced by a deterministic trend or an explosive root. We find substantial nonstationarity in the first four moments of the distributions – primarily in the mean (i.e., global warming) and in the (decreasing) variance. However, none of the nonstationarity that we detect is more persistent than that of a stochastic trend.

In the absence of some kind of underlying nonlinearity in the evolution of the distributions, our evidence suggests the statistical possibility of reversion in the moments of these anomalies. As Woodward and Gray (1993) emphasize, such evidence has little to say about whether or not the increasing trend will continue for the foreseeable future. We claim only that the evidence casts doubt on the type of (deterministic) trend in which certain changes in the moments – e.g., an increasing mean – are inevitable. We note that our evidence makes no claim on the anthropogenic nature of the observed climate change. Rather our statistical evidence may be useful in further specifying models or tests of anthropogenic forcing, as do some of the aforementioned studies, such as Kaufmann *et al.* (2013) and Estrada *et al.* (2013).

In Section 2, we introduce the global temperature anomaly data, and we discuss the time series framework for analyzing state distributions and testing procedures for nonstationarity of those distributions. We discuss step-by-step implementation of the tests and our empirical results in Section 3, and we conclude with Section 4.

## 2 Data and Methodology

We first present the data set of global temperature anomalies used in our analysis. We then review the basic time series framework and testing procedures used by Chang *et al.* (2015)

to test for nonstationarity of state distributions. Since this procedure may be new to many readers, our discussion is self-contained but necessarily abbreviated, and interested readers are referred to that paper for additional technical details.

Although the methodology and theory of our analysis are largely based on Chang *et al.* (2015), our procedure contains a novel aspect. While they consider a test for nonstationarity against only a stationary left-hand-sided alternative, we extend their test to an explosive or deterministically trending right-hand-sided alternative. The extension will be critical in discerning persistence characteristic of a unit root process from much stronger persistence in temperature anomalies.

## 2.1 Global Temperature Distributions

We employ the HadCRUT3 data set, well-known to climate researchers and described in detail by Brohan *et al.* (2006). The data set 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 monthly measurements extend from 1850 to 2012 and aim to cover as much of the globe as possible.

The HadCRUT3 data show temperature anomalies in degrees Celsius from the monthly average over the period 1961-90. Specifically, deviations are calculated for each land station (110-4,098 stations per month throughout the sample), and then the deviations are averaged across all stations in a given grid box of  $5^\circ$  latitude and  $5^\circ$  longitude. For marine data, the measurements are taken from ships or buoys (1,495-1,648,815 marine observations per month throughout the sample), and the anomaly is calculated based on the monthly average over 1961-90 for each grid box. The interested reader is referred to Brohan *et al.* (2006) for a very detailed discussion of data construction and known limitations, such as warming effects from urbanization and technological changes in temperature measurement over more than a century and a half.

The maximum number of temperature anomaly observations in each month is given by 2,592, 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 HadCRUT3 data, providing a maximum number of  $2,592 \times 12 = 31,104$  annual observations. Figure 1 shows annual time series of the number of non-empty box-months for the globe and for each hemisphere. The observations each year generally increase from about 5,000 at the beginning of the sample to about 22,000 in the mid-1990's, leveling out at about 21,000 by the end of the sample. There are three obvious dips, corresponding roughly with the American Civil War (1861-65), World War I (1914-19), and World War II (1939-45).

Hemispheric means are often analyzed separately in studies on climate change, since

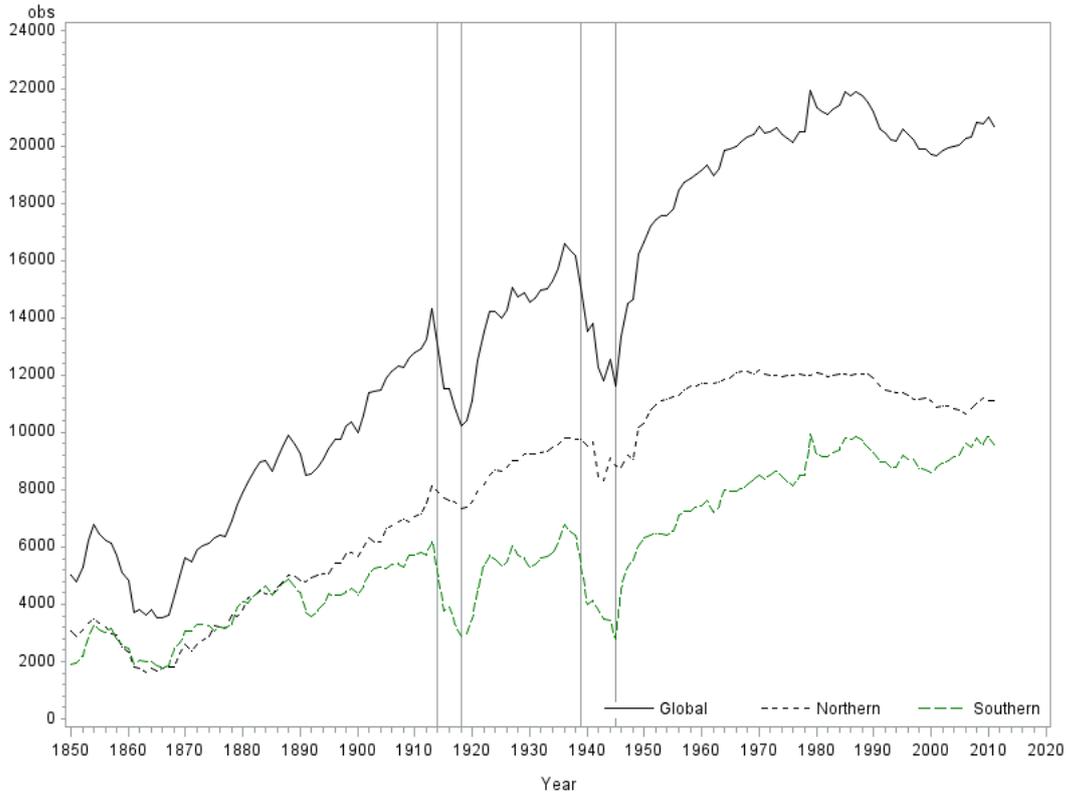


Figure 1: **Number of Annual Temperature Observations.** Observations for the globe, NH, and SH, based on  $5^\circ$  by  $5^\circ$  grid boxes. Total possible annual observations for the globe is  $36 \times 72 \times 12 = 31,104$  ( $36 \times 5^\circ$  along each meridian,  $72 \times 5^\circ$  around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

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. (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 each temperature, we average the estimated NH and SH densities to obtain a global density. Each hemisphere receives an equal weight to avoid giving too much weight to the NH, where there are more non-empty grid boxes. We omit approximately 5% extreme outliers and make the supports of these densities compact.<sup>2</sup>

<sup>2</sup>The compact supports avoid the well-known empty bin problem in nonparametric density estimation. We note that the HadCRUT3 data already omits extreme temperature anomalies in its construction (Brohan

Specifically, we set the supports  $[-2.62, 2.50]$ ,  $[-3.46, 3.06]$  and  $[-2.26, 1.78]$  for the global distribution, the NH distribution, and the SH distribution, respectively. We utilize the typical nonparametric density estimator with a normal kernel and Silverman bandwidth to estimate the densities.

Figures 2-4 show the densities and time series plots of the first four moments of the temperature anomalies in each year. Specifically, Figure 2 shows the global densities and sample moments, Figure 3 shows those for the NH, and Figure 4 shows those for the SH. The estimated densities are regarded as the data that we subsequently analyze. We may well expect that estimation errors in the temperature anomaly densities have a negligible effect on our analysis, since 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. Therefore, we treat the densities as being observable in our subsequent discussions.

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>3</sup> The top left panel of each of Figures 2-4 shows the annual temperature anomaly densities  $(f_t(s))$ . The temporally demeaned temperature anomaly densities  $-(w_t(s))$  in our subsequent notation – are shown in the top right panels. We may interpret the latter 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.

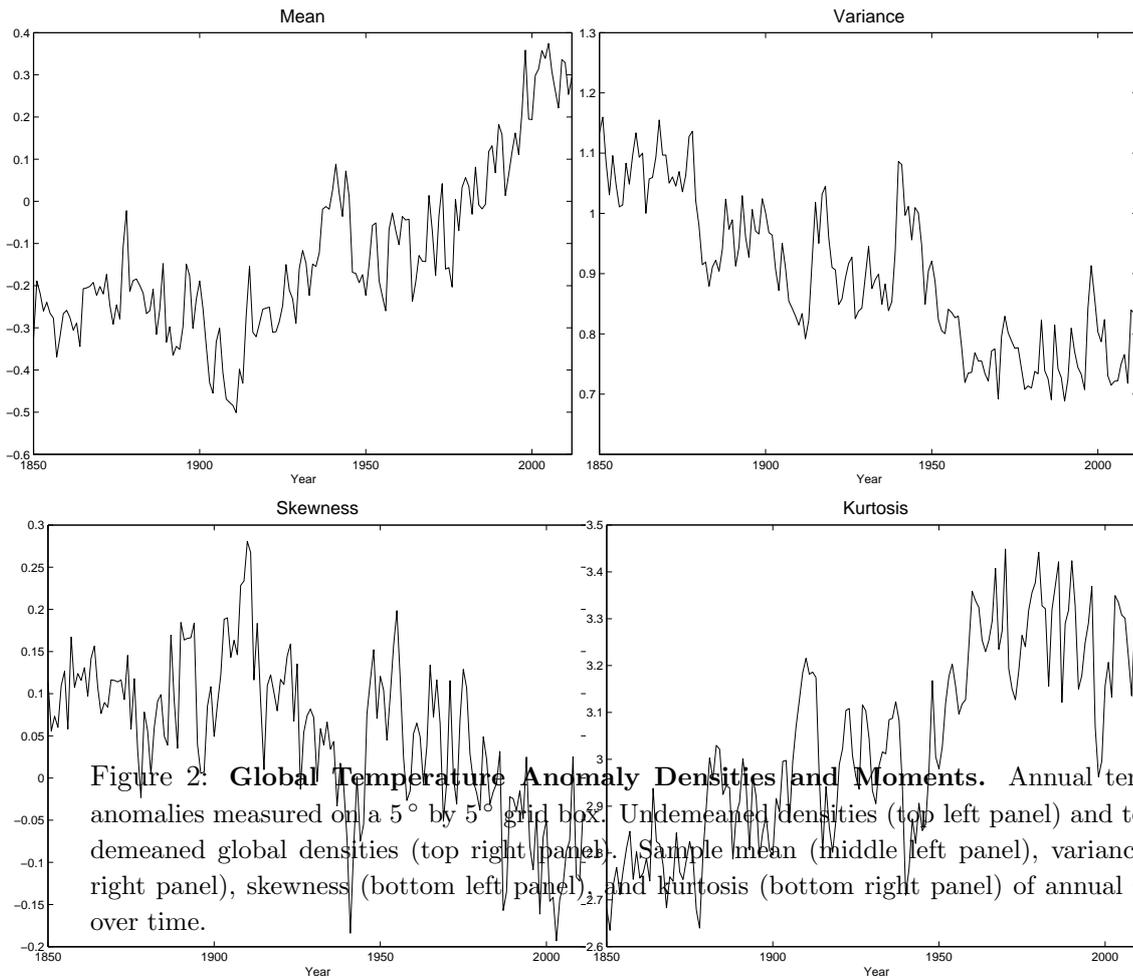
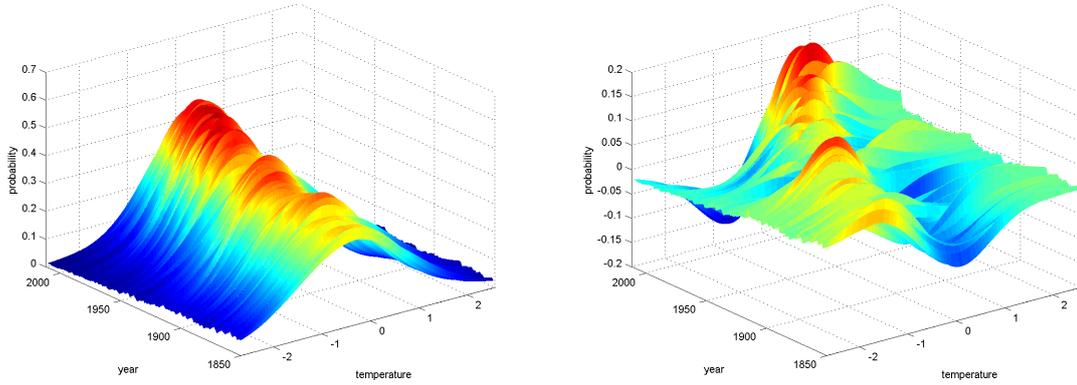
The remaining panels of Figures 2-4 show the time paths of the estimated *cross-sectional moments* of the distributions  $(f_t)$ . Specifically, the means (middle left panels), variances (middle right panels), skewnesses (bottom left panels), and kurtoses (bottom right panels) are plotted. The cross-sectional mean is defined above as  $\mu_t = \int s f_t(s) ds$ . Furthermore, 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  $t = 1, \dots, T$ .

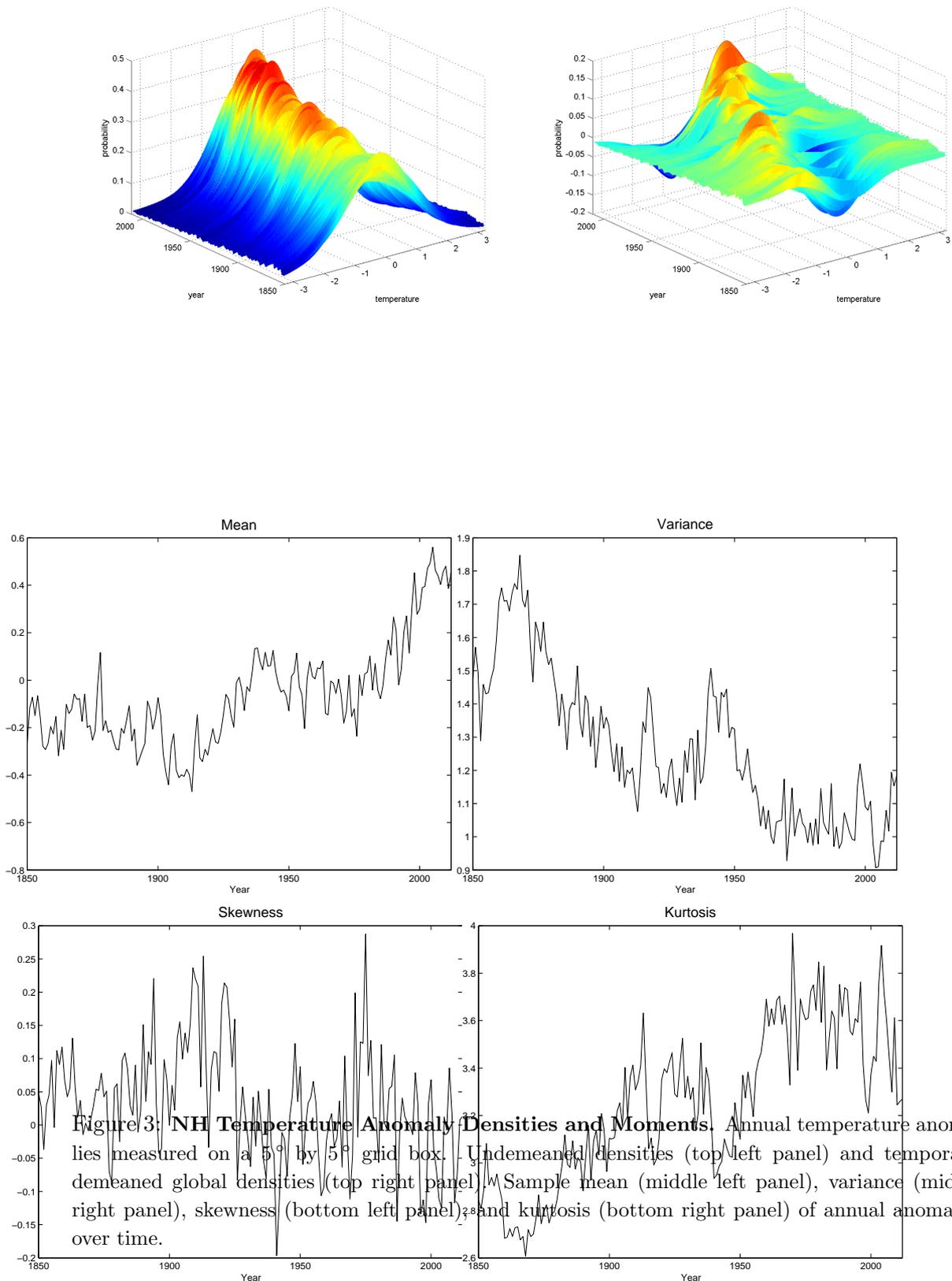
Casual inspection suggests that the means have been increasing since about 1975 and

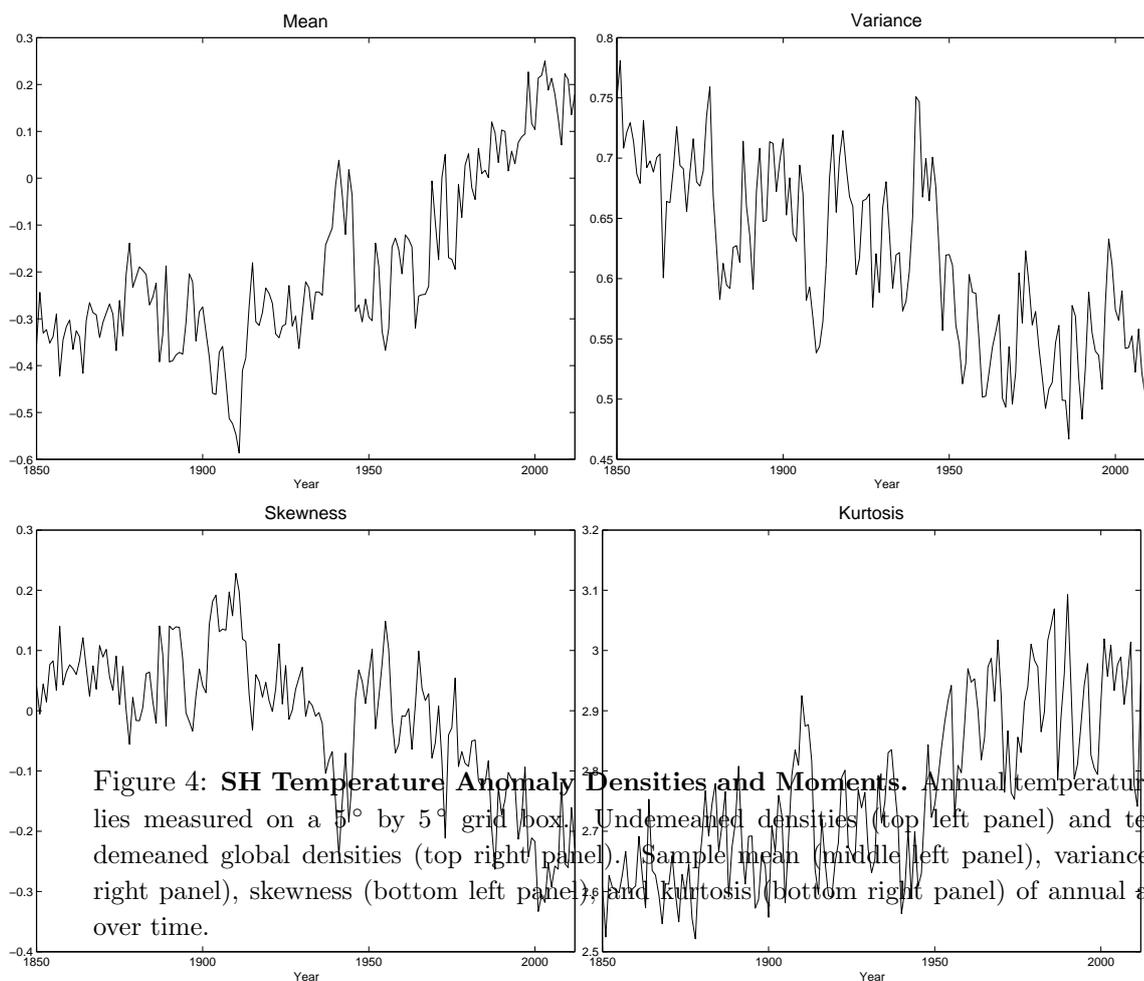
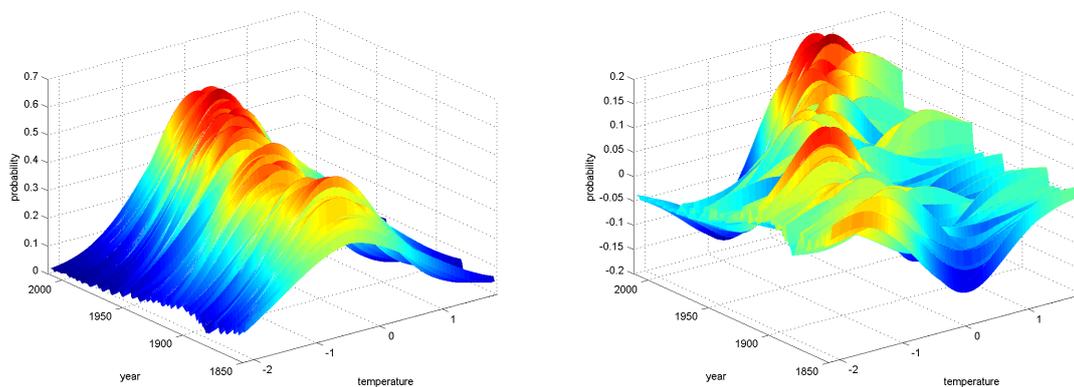
---

*et al.*, 2006). We do not believe that our omission should substantially affect our qualitative results, since our aim is to describe global rather than local anomalies.

<sup>3</sup>We may, of course, compute the cross-sectional mean as a Riemann sum using a fine enough partition over the support of the given density function.







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 appear to have decreased from positive to negative, suggesting that although the probabilities of observing moderately positive temperature anomalies have increased, the probabilities of observing extremely positive temperature anomalies (up to the maxima of our supports) may have decreased. The variances appear to have decreased, while the kurtoses have increased. Such movement suggests that the distributions have become more peaked around their (increasing) means, but without associated decreases in the probabilities of outliers. Instead, the probabilities of observing moderate temperature anomalies may have decreased.

In order to explore the persistence of the moments, we now turn to a more formal analysis of the stationary and nonstationary spaces of the temporally demeaned temperature anomaly densities.

## 2.2 Basic Framework for Time Series Analysis

We analyze the temperature densities obtained above as a time series of functional observations. 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 by

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

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

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, we let its mean be given by the expectation  $\mathbb{E}f_t$ , and define its variance to be the expected tensor product  $\mathbb{E}(f_t - \mathbb{E}f_t) \otimes (f_t - \mathbb{E}f_t)$  of the demeaned  $f_t$  with itself.<sup>4</sup> The mean and variance of  $f_t$  therefore become a function and

---

<sup>4</sup>Essentially, tensor products of finite dimensional vectors yield matrices. In contrast, tensor products of functions become infinite dimensional and they are formally interpreted as operators in a Hilbert space of functions.

an operator respectively for  $t = 1, \dots, T$ . On the other hand, since each element  $f_t$  of the sequence  $(f_t)$  represents a density, we 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>5</sup> By convention, we let  $n = 0$  if all of the coordinate processes are stationary. 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 will respectively be referred to as the nonstationarity and stationarity subspaces of  $H$ . Subsequently, we define  $\Pi_N$  and  $\Pi_S$  to be the projections on  $H_N$  and  $H_S$ , and let

$$w_t^N = \Pi_N w_t \quad \text{and} \quad w_t^S = \Pi_S w_t,$$

where  $(w_t^N)$  and  $(w_t^S)$  signify respectively the nonstationary and stationary components of  $(w_t)$ . Since  $\Pi_N + \Pi_S$  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 it has time invariant mean and variance that are finite and well defined. In this case, we have  $n = 0$ , since the coordinate processes are all stationary. Under stationarity, 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

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

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).

Once we fix an arbitrary orthonormal basis  $(\phi_i)$  of  $H$ , we may write any function  $w$  in  $H$  as a linear combination of  $(\phi_i)$  as in  $w = \sum_{i=1}^{\infty} c_i \phi_i$  with a numerical sequence  $(c_i)$ . In implementing our approach, we use an orthonormal wavelet basis  $(\phi_i)$  to represent vectors in  $H$  as their finite linear combinations of  $M$  leading basis elements for some large  $M$ . This yields the correspondence  $w \leftrightarrow (c_1, \dots, c_M)'$  between  $w \in H$  and  $(c_1, \dots, c_M)' \in \mathbb{R}^M$ , which allows us to regard a function in  $H$  essentially as a large dimensional vector in Euclidean space. Under this convention, the inner product  $\langle v, w \rangle$  becomes the usual Euclidean inner

---

<sup>5</sup>Of course, there exists a wide variety of nonstationary processes that do not have any trends, stochastic or deterministic. In the paper, however, we only consider nonstationary processes with trends increasing either stochastically or deterministically.

product of two vectors in  $\mathbb{R}^M$  corresponding respectively to  $v$  and  $w$  in  $H$ , and the tensor product  $v \otimes w$  reduces to the conventional Euclidean outer product of two vectors in  $\mathbb{R}^M$  corresponding respectively to  $v$  and  $w$  in  $H$ .

### 2.3 Testing for Nonstationarity

The test for nonstationarity of the global temperature anomaly distributions we use in the paper is based on the sample 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 magnitude of the quadratic form in  $v \in H_N$  is of order bigger than  $T$ , since 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 principle components of  $Q_T$  in (4) and use them to test for nonstationarity in the temperature anomaly distributions.

The exact magnitude of the quadratic form in  $v \in H_N$  defined by  $Q_T$  further depends on what type of nonstationarity the coordinate process  $(\langle v, w_t \rangle)$  exhibits. The quadratic form is of order  $T^2$  if the coordinate process has unit root nonstationarity (a stochastic trend). On the other hand, it is of order  $T^3$  if the coordinate process has a linear deterministic trend, and it diverges at an exponential rate if the coordinate process has an explosive root.

To separately identify these different types of nonstationarity in the global temperature distributions, we define the unit root subspace  $H_U$  of  $H$  to be the  $m$ -dimensional subspace of an  $n$ -dimensional subspace  $H_N$  such that  $(\langle v, w_t \rangle)$  is a unit root process for all  $v \in H_U$ . For completeness, we also defined the deterministic and explosive subspace  $H_X$  of  $H$  such that  $H_N = H_U \oplus H_X$  and  $H = H_S \oplus H_U \oplus H_X$ . There is no unit root nonstationarity if  $m = 0$ , whereas the entire nonstationarity is unit root nonstationarity if  $m = n$ . In fact, we find that  $m = n$  in our empirical results on temperature anomalies below. However, we also consider the case  $m < n$  here to introduce our testing procedure to determine  $m$  as well as  $n$ .

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  $n$  largest eigenvalues of  $Q_T$ . Chang *et al.* (2015) establish the consistency of the estimator for the case in which we only have unit root nonstationarity. The extension of their proof to allow for more general types of nonstationarity is straightforward. Under our setup, if normalized by  $T^2$ ,  $\lambda_{n-m+1}(Q_T), \dots, \lambda_n(Q_T)$  have well defined limit distributions as  $T \rightarrow \infty$ , while  $\lambda_1(Q_T), \dots, \lambda_{n-m}(Q_T)$  diverge faster than the rate  $T^2$ . In particular, the unit root subspace  $H_U$  can be consistently estimated by the span of  $m$ -orthonormal eigenvectors  $v_{n-m+1}(Q_T), \dots, v_n(Q_T)$  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 hypotheses of stationarity, and then against the alternative hypothesis of deterministic/explosive nonstationarity. We expect the eigenvalues ( $\lambda_i(Q_T)$ ) to have discriminatory powers for such tests. However, they cannot be used directly, since their limit distributions are dependent upon nuisance parameters. Therefore, we need to construct tests based on eigenvalues with limit distributions free of nuisance parameters.

To this end, we define  $(z_t)$  by either

$$z_t = (\langle v_1(Q_T), w_t \rangle, \dots, \langle v_p(Q_T), w_t \rangle)' \quad (6)$$

( $v_p$  is the eigenvector associated with the  $p$ -th largest eigenvalue) or

$$z_t = (\langle v_{n-q+1}(Q_T), w_t \rangle, \dots, \langle v_n(Q_T), w_t \rangle)' \quad (7)$$

( $v_n$  is the eigenvector in  $H_N$  associated with the smallest eigenvalue) for  $t = 1, \dots, T$ , and we use the index  $r$  to denote  $p$  or  $q$  depending upon whether  $(z_t)$  is given by (6) or (7). Moreover, we define the product sample moment  $Q_r^T = \sum_{t=1}^T z_t z_t'$ , and the long-run variance estimator  $\Omega_r^T = \sum_{|k| \leq \ell} \varpi_\ell(k) \Gamma_T(k)$  of  $(z_t)$ , where  $\varpi_\ell$  is the weight function with bandwidth parameter  $\ell$  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

$$\tau_p^T = T^{-2} \lambda_{\min}(Q_p^T, \Omega_p^T) \quad (8)$$

$\tau_p^T$	$p = 1$	2	3	4	5
1%	0.0274	0.0175	0.0118	0.0103	0.0085
5%	0.0385	0.0223	0.0154	0.0127	0.0101
10%	0.0478	0.0267	0.0175	0.0139	0.0111
$\sigma_q^T$	$q = 1$	2	3	4	5
99%	0.7487	1.0073	1.2295	1.4078	1.5952
95%	0.4660	0.6787	0.8645	1.0336	1.1892
90%	0.3494	0.5399	0.7066	0.8574	1.0092

Table 1: **One-sided Critical Values for the Test Statistics  $\tau_p^T$  and  $\sigma_q^T$ .**

and

$$\sigma_q^T = T^{-2} \lambda_{\max}(Q_q^T, \Omega_q^T), \quad (9)$$

where  $\lambda_{\min}(Q_p^T, \Omega_p^T)$  and  $\lambda_{\max}(Q_q^T, \Omega_q^T)$  are respectively the smallest and the largest generalized eigenvalues of  $Q_r^T$  with respect to  $\Omega_r^T$  for  $r = p$  or  $q$ .

The test statistics  $\tau_p^T$  and  $\sigma_q^T$  introduced in (8) and (9) are used with the critical values obtained under the null hypothesis that  $(z_t)$  defined in (6) or (7) is a unit root process in order to determine  $n$  and  $m$ . Under very general conditions, Chang *et al.* (2015) show that the statistic  $\tau_p^T$  has a well-defined nondegenerate limit distribution that is free of nuisance parameters and depends only on  $p$ , as long as  $n - m + 1 \leq p \leq n$  (for  $m, n \geq 1$ ). We may extend their result and establish that it is also true for the statistic  $\sigma_q^T$  under the same conditions if  $1 \leq q \leq m$  (for  $m, n \geq 1$ ). We compute the critical values of the statistic  $\sigma_q^T$  up to  $q = 5$  and tabulate them in Table 1 together with the critical values of the statistic  $\tau_p^T$  for easy reference.

Note that the statistic  $\tau_p^T$  converges to 0 for all  $p > n$ . Therefore, we may use  $\tau_p^T$  to determine  $n$  as follows.<sup>6</sup> We start from a value of  $p$  large enough to be bigger than  $n$  and test the null hypothesis  $H_0 : \dim(H_N) = p$  against the alternative hypothesis  $H_1 : \dim(H_N) \leq p - 1$  successively downward, until we reach  $p = 1$ . For each test, we reject the null hypothesis if the value of  $\tau_p^T$  is *smaller* than the respective critical values provided in Table 1. 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  $p_{\max}$ , for which we fail to reject the null hypothesis. By such a successive testing procedure employing a consistent test, we may find the true value of  $n$  with asymptotic probability of virtually one by making the size of the test small enough.

Once  $n$  is found, we may use the statistic  $\sigma_q^T$  to determine  $m$ . Note that the statistic

<sup>6</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.

$\sigma_q^T$  diverges to infinity for all  $m < q \leq n$ . We start from  $q = n$  and test the null hypothesis  $H_0 : \dim(H_U) = q$  against the alternative hypothesis  $H_1 : \dim(H_U) \leq q - 1$  successively downward, until we reach  $q = 1$ . For the test, we reject the null hypothesis if  $\sigma_q^T$  takes a value *larger* than the respective critical value reported in Table 1, in contrast to the test based on  $\tau_p^T$ . As above, we proceed as long as the null hypothesis is rejected in favor of the alternative hypothesis and set our estimate for  $m$  to be the largest value  $q_{\max}$  of  $q$ , for which we fail to reject the null hypothesis. Again, we may find the true value of  $m$  with asymptotic probability arbitrarily close to one.

## 2.4 Nonstationarity in Cross-Sectional Moments

Once we determine  $n$  and estimate the nonstationary subspace  $H_N$ , we may determine the nonstationary proportion of each cross-sectional moment. Similarly to Chang *et al.* (2015), 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

$$\|\mu_i\|^2 = \|\Pi_N \mu_i\|^2 + \|\Pi_S \mu_i\|^2 = \sum_{j=1}^n \langle \mu_i, v_j \rangle^2 + \sum_{j=n+1}^{\infty} \langle \mu_i, v_j \rangle^2, \quad (10)$$

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

$$\pi_i^N = \frac{\|\Pi_N \mu_i\|}{\|\mu_i\|} = \sqrt{\frac{\sum_{j=1}^n \langle \mu_i, v_j \rangle^2}{\sum_{j=1}^{\infty} \langle \mu_i, v_j \rangle^2}} \quad (11)$$

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 the addition of a new ratio given by

$$\pi_i^U = \frac{\|\Pi_U \mu_i\|}{\|\mu_i\|} = \sqrt{\frac{\sum_{j=n-m+1}^n \langle \mu_i, v_j \rangle^2}{\sum_{j=1}^{\infty} \langle \mu_i, v_j \rangle^2}}, \quad (12)$$

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

### 3 Persistent Features in Temperature Anomalies

We now 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, and we interpret the evident persistent features in the temperature anomaly data.

#### 3.1 Empirical Implementation of the Tests and Proportions

To implement our methodology, we need to deal with cross-sectional densities that we regard as functional observations on the Hilbert space  $H$  introduced in (2). In our analysis,  $H$  is assumed to have 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  $\mathbb{R}^\infty$  and the correspondence is uniquely defined, once the basis elements are fixed.

For instance, once a basis  $(\phi_1, \phi_2, \dots)$  is given, we may write any  $w \in H$  as  $w = c_1\phi_1 + c_2\phi_2 + \dots$  and the correspondence becomes  $w \leftrightarrow (c_1, c_2, \dots)$ . We use this correspondence in our analysis of functional observations. Of course, the correspondence becomes operational only if we replace  $\mathbb{R}^\infty$  by  $\mathbb{R}^M$  for some large  $M$ . Subsequently, we let  $[w] = (c_1, \dots, c_M)'$  and define a correspondence

$$w \leftrightarrow [w] \quad (13)$$

between  $H$  and  $\mathbb{R}^M$ , in place of  $\mathbb{R}^\infty$ . In our analysis, we use a Daubechies wavelet basis and set  $M = 1,037$ , which we believe to be sufficiently large.

Under the correspondence between  $H$  and  $\mathbb{R}^M$  defined in (13), we have the correspondences

$$\langle v, w \rangle \leftrightarrow [v]'[w] \quad \text{and} \quad v \otimes w \leftrightarrow [v][w]'$$

for any  $v, w \in H$ . In fact, under the correspondence in (13), a linear operator  $Q$  on  $H$  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  $v, w \in H$ . We use these correspondences throughout our analysis.

For ease of reference and clarity of exposition, and because our procedure is new, we briefly outline seven steps utilized to create the test statistics  $\tau_p^T$  and  $\sigma_q^T$  using actual data from a finite sample.

1. **Obtain**  $w_t$ . We regard  $w_t$  as an  $M$ -dimensional vector  $[w_t]$  for each  $t$ .
2. **Create**  $Q_T$ . Implement  $Q_T = \sum_{t=1}^T w_t \otimes w_t$  as  $[Q_T] = \sum_{t=1}^T [w_t][w_t]'$  for each  $t$ .
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** (6) **or** (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**  $Q_r^T$  **and**  $\Omega_q^T$ . Implement  $Q_r^T = \sum_{t=1}^T z_t z_t'$  and  $\Omega_r^T = \sum_{|k| \leq \ell} \varpi_\ell(k) \Gamma_T(k)$  using the Parzen window with Andrews plug-in bandwidth.
6. **Calculate**  $\lambda(Q_r^T, \Omega_r^T)$ . These are generalized eigenvalues of  $Q_r^T$  with respect to  $\Omega_r^T$  for  $r = p$  or  $q$ .
7. **Calculate Test Statistics**  $\tau_p^T$  **and**  $\sigma_q^T$  **from** (8) **and** (9).

Once these test statistics have been 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 (11) and (12) cannot be calculated directly, since  $H_N$  and  $H_U$  are unknown. Instead, we may use the *sample* nonstationary and unit root proportions of the  $i$ -th cross-sectional moment

$$\hat{\pi}_{iT}^N = \sqrt{\frac{\sum_{j=1}^n \langle \mu_i, v_j(Q_T) \rangle^2}{\sum_{j=1}^m \langle \mu_i, v_j(Q_T) \rangle^2}} \quad \text{and} \quad \hat{\pi}_{iT}^U = \sqrt{\frac{\sum_{j=n-m+1}^n \langle \mu_i, v_j(Q_T) \rangle^2}{\sum_{j=1}^m \langle \mu_i, v_j(Q_T) \rangle^2}} \quad (14)$$

	$p, q =$	1	2	3	4
Global	$\tau_p^T$	0.0531	0.0289	0.0105	0.0097
	$\sigma_q^T$	0.0531	0.0536		
NH	$\tau_p^T$	0.0387	0.0379	0.0119	0.0105
	$\sigma_q^T$	0.0387	0.0407		
SH	$\tau_p^T$	0.0611	0.0219	0.0097	0.0089
	$\sigma_q^T$	0.0611			

Table 2: **Test Statistics**  $\tau_p^T$  and  $\sigma_q^T$ . Global, NH, and SH temperature anomaly distributions.

to estimate  $\pi_i^N$  and  $\pi_i^U$ . Chang *et al.* (2015) show that the sample nonstationarity proportion  $\hat{\pi}_{iT}^N$  is a consistent estimator of the original nonstationarity proportion  $\pi_i^N$  and it follows by extension that  $\hat{\pi}_{iT}^U$  consistently estimates  $\pi_i^U$ .

### 3.2 Test Results

Table 2 shows the  $\tau_p^T$  and  $\sigma_q^T$  test statistics for the global, NH, and SH temperature anomalies up to  $p = 4$ . Starting with  $\tau_p^T$  for the global distribution and comparing the statistic with the critical values in Table 1 we reject  $p = 4$  against the alternative  $p \leq 3$ , and then we reject  $p = 3$  against the alternative  $p \leq 2$ , both with a size of 1%. We cannot reject  $p = 2$  against  $p \leq 1$  even with 10%. We obtain the same results for the NH distribution, only with a size of 5%. The NH distribution is actually marginal for  $p = 1$  against  $p = 0$ , but we stop at 2, because we fail to reject  $p = 2$  against  $p = 1$ . The SH distribution strongly rejects  $p = 4$  and  $p = 3$  at 1% size and rejects  $p = 2$  at 5% size, but  $p = 1$  is not rejected against  $p = 0$ .

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 are accumulating over time. Changes in the temperature anomaly distributions are not entirely *transitory*.

Is the persistence of the unit root type, exhibiting mean reversion over a long horizon, or is the persistence explosive or deterministic? To answer this question, we now examine  $\sigma_q^T$ . Looking first at the global distribution,  $q = n = 2$  is not rejected at any reasonable significance level against  $q \leq 1$ , and neither is  $q = 1$  against  $q = 0$ . We thus choose  $\dim(H_U)$  to be  $m = n = 2$  for the global distribution. The same outcome  $m = n = 2$  is obtained for the NH distribution, while that for the SH distribution is similarly  $m = n = 1$ . The fact that  $m = n$  is chosen in every case suggests that all of the nonstationarity is better

	$\hat{\pi}_{1T}^N$	$\hat{\pi}_{2T}^N$	$\hat{\pi}_{3T}^N$	$\hat{\pi}_{4T}^N$	$\hat{\pi}_{5T}^N$	$\hat{\pi}_{6T}^N$	$\hat{\pi}_{7T}^N$
Global	0.516	0.270	0.235	0.188	0.151	0.142	0.117
NH	0.409	0.205	0.160	0.129	0.101	0.094	0.078
SH	0.633	0.199	0.331	0.168	0.212	0.140	0.157

Table 3: **Sample Nonstationary Proportions in the First Seven Moments.** Global, NH, and SH temperature anomaly distributions.

characterized by unit-root-type persistence, suggestive of stochastic trends in the moments of the distributions, than by higher-order persistence associated with explosive roots or linear deterministic trends.

Our results for the mean better agree with those of Gordon (1991) *inter alia*, who found evidence for a unit root in the mean global temperature, than with the linear trend used by many authors. We hasten to add that our results do not imply random walks in the moments. A random walk is a very special case of a unit root process that has completely unpredictable increments. A unit root process may have increments with strong but stationary persistence, meaning that *changes* in the mean global temperature may have long-lasting effects. Such persistent changes yield a stochastic trend that may indeed increase over a long period of time, giving the appearance of a linear trend or a broken linear trend.

### 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 3 shows consistent estimates  $\hat{\pi}_{iT}^N$  for  $i = 1, \dots, 7$  of the proportion of the nonstationary subspace  $\pi_i^N$  in each of the first seven cross-sectional moments of temperature anomalies. The remaining proportions are in the stationary subspace  $\pi_i^S$ .

For the entire globe, just over half (51.6%) of the persistence in the mean is strong enough to be unit-root-type persistence. Curiously, the persistence in the mean appears to be stronger in the SH than the NH, in the sense that 63.3% of the persistence in the Southern mean is of the unit root type, while only 40.9% of that north of the Equator. The global, NH, and SH variances are 27.0%, 20.5%, and 19.9% respectively, suggesting that nonstationary proportions in the variance are lower than those in the mean and roughly similar in both hemispheres.

The proportion of unit root persistence in the skewness for the globe is 23.5%. Like the mean, the skewness appears to be less persistent in the NH (16.0%) than in the SH (33.1%).

The persistence appears to be declining in the remaining four moments for the globe and NH, while it remains roughly 15-20% in the SH.

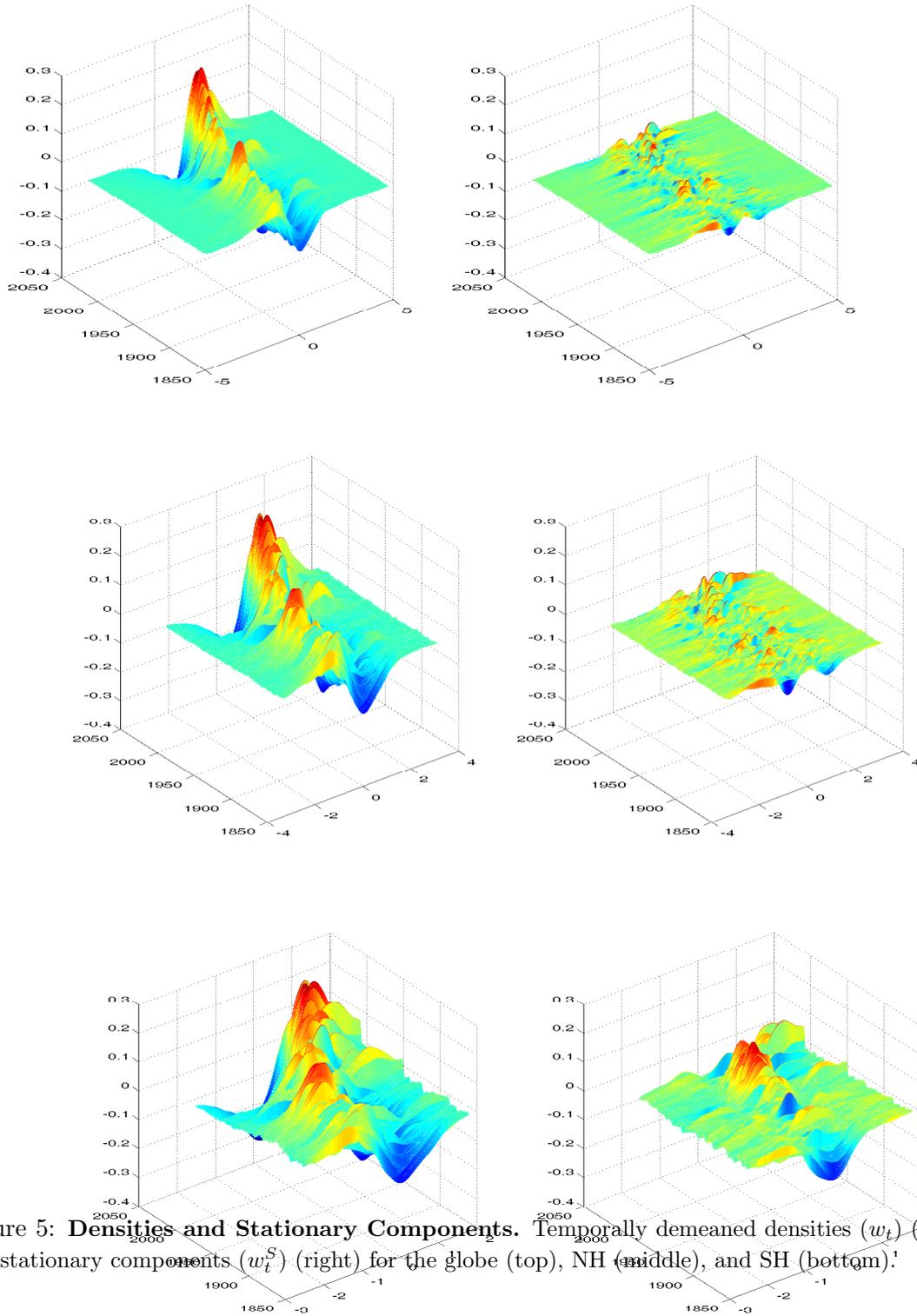
### 3.4 Estimated Components

Because the concepts of stationarity and nonstationarity of densities are quite new, we present some further illustrations of these components. The left panels of Figure 5 show the time series of demeaned densities ( $w_t$ ) (same as the top right panels of Figures 2-4) for the globe, NH, and SH. The right panels show the time series of stationary components of the respective densities ( $w_t^S$ ). These are calculated by subtracting the estimated nonstationary (unit root) components ( $w_t^N$ ), calculated as  $w_t^N = \Pi_N w_t$ , from the densities ( $w_t$ ). Recall that the dimension of  $n$  ( $= m$ ) was estimated to be two for the globe and NH and one for the SH. In all three cases – but especially in the first two – the stationary components of the densities appear to be more like random noise, showing very little evidence for persistence in any of the moments. Evidently, the temporal patterns in the densities are driven by their nonstationary components rather than by their stationary components.

The concept of nonstationarity will be more familiar to readers in a simple time series context. To this end, Figure 6 shows the nonstationary components more clearly. First, we plot the normalized mean process. The mean process is given by  $\mu_t$  above (middle left panels of Figures 2-4), but we normalize the series to unit length with a Euclidean norm since the eigenvectors used to compute the nonstationary coordinate processes have unit length. We then plot the two estimated nonstationary coordinate processes  $\langle v_1, w_t \rangle$  and  $\langle v_2, w_t \rangle$  – which could be written as  $(c_{1t})$  and  $(c_{2t})$  using the correspondence in (13) – for the globe and NH and the one  $\langle v_1, w_t \rangle$  for the SH. Clearly, the estimated nonstationary coordinate processes exhibit more persistence compared to the time series of cross-sectional means that include both stationary and nonstationary components. In other words, the time series plots of the nonstationary coordinate processes better resemble sample paths of unit root processes than those of stationary process – or those of trend stationary processes, for that matter. We see stronger evidence for the globe and NH, but less so for the SH, which is not surprising given that we have only a one-dimensional unit root space for the SH and the unit root proportion of the mean process is over 60%, capturing a substantial portion of nonstationarity in the time series of SH temperature distributions.

## 4 Concluding Remarks

The means and kurtoses of global and hemispheric temperature anomalies appear to be increasing, while the variances and skewnesses appear to be decreasing. Our analysis of the



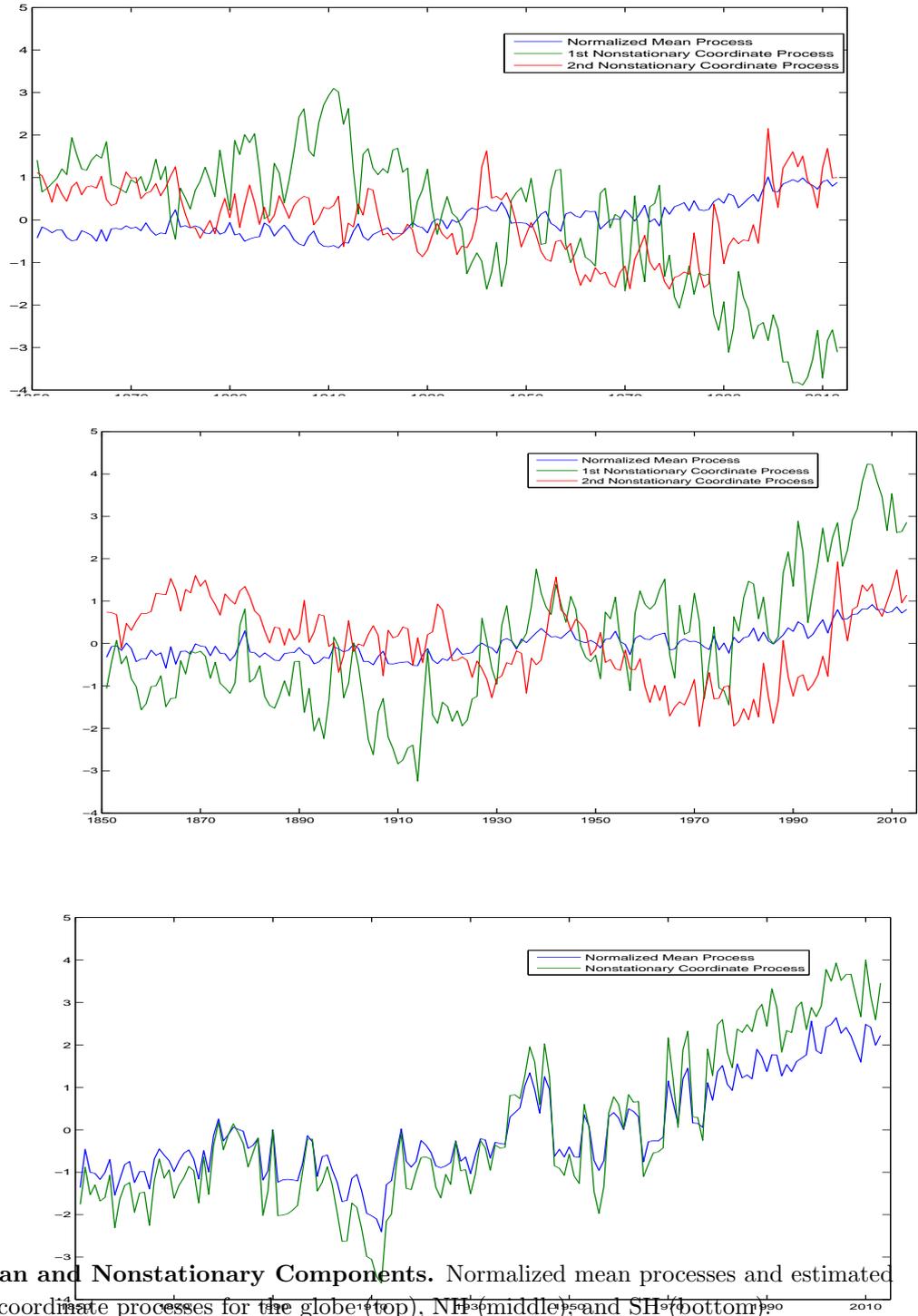


Figure 6: Mean and Nonstationary Components. Normalized mean processes and estimated nonstationary coordinate processes for the globe (top), NH (middle), and SH (bottom).

persistence of these moments suggests stochastically trending movement toward warmer temperatures on average, but with lower dispersion and more outliers. Since the distributions are estimated on the same support over time, more outliers do not mean more extreme temperatures. Rather, more outliers mean that existing extremes are more likely observed. Moreover, the negative skewness means that colder outliers are increasingly less likely, while warmer outliers are more likely than colder. (A decreasing variance has a countervailing effect on warmer outliers, so we cannot say for sure if warmer outliers are more likely over time.)

There are permanent and long-run fluctuations in the moments of the distributions of temperature anomalies over this period – and most notably in the mean and variance of all of the distributions and in the skewness of the SH distributions. The NH distributions appear to be more stable over time, in the sense that even with the unit root persistence in the mean and variance, the proportion of this type of persistence in almost every moment is smaller than that in the SH. From a purely statistical point of view, if the level of persistence is unchanged, we may expect that all of these moments will revert to their long-run averages, but possibly after a *very* long period of time. We should not expect that they will *never* revert.

We acknowledge two important limitations of our approach: data and structure. Brohan *et al.* (2006) extensively discuss limitations of these data. Differences in measuring temperature in the NH and SH could explain some of the variation in our results across the hemispheres. The SH has fewer non-missing grid box observations. But with relatively more marine observations in the SH, these observations have less land bias. In spite of these limitations, our main results on unit root behavior should not be substantially affected, because we do not expect that measurement error associated with these data should create or alter long-run properties of the distributions.

Second, although our approach employs new and sophisticated statistical tools to describe the evolution of temperature anomalies, it certainly has structural limitations. One could envision a plausible scenario in which the slowly increasing mean warms enough polar ice to substantially affect the ocean circulation, which could have a large nonlinear effect on the future mean – an “indirect carbon-cycle feedback-forcing effect.” Such nonlinear feedback could substantially alter the persistence features of the temperature anomaly distributions. In the absence of a richer structural model, a structural break model for higher-order moments of the distributions, along the lines of Gay-Garcia *et al.* (2009) *inter alia* for the means, could provide some useful insight. We leave this for future research.

## References

- Baillie, R.T. and S.-K. Chung (2002). "Modeling and Forecasting from Trend-Stationary Long Memory Models with Applications to Climatology," *International Journal of Forecasting* 18, 215-226.
- Ballester, J., F. Giorgi, and X. Rodó (2010). "Changes in European Temperature Extremes Can Be Predicted from Changes In PDF Central Statistics," *Climatic Change* 98, 277-284.
- Bloomfield, P. (1992). "Trends in Global Temperature," *Climatic Change* 21, 1-16.
- Bloomfield, P. and D. Nychka (1992). "Climate Spectra and Detecting Climate Change," *Climatic Change* 21, 275-287.
- Bosq, D. (2000). *Linear Processes in Function Spaces, Lecture Notes in Statistics 149*. New York: Springer.
- Brock, W.A., G. Engström, D. Grass, A. Xepapadeas (2013). "Energy Balance Climate Models and General Equilibrium Optimal Mitigation Policies," *Journal of Economic Dynamics & Control* 37, 2371-2396.
- Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones (2006). "Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Dataset from 1850," *Journal of Geophysical Research: Atmospheres* 111, D12106.
- Castruccio, S. and M.L. Stein (2013). "Global Space-Time Models for Climate Ensembles," *The Annals of Applied Statistics* 7, 1593-1611.
- Castruccio, S., D.J. McInerney, M.L. Stein, F.L. Crouch, R.L. Jacob, E.J. Moyer (2014). "Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs," *Journal of Climate* 27, 1829-1844.
- Chang, Y., C.S. Kim, and J.Y. Park (2015). "Nonstationarity in Time Series of State Densities," forthcoming in *Journal of Econometrics*.
- Donat, M.G. and L.V. Alexander (2012). "The Shifting Probability Distribution of Global Daytime and Night-Time Temperatures." *Geophysical Research Letters* 39, L14707.
- Dong, L., T.J. Vogelsang, S.J. Colucci (2008). "Interdecadal Trend and ENSO-Related Interannual Variability in Southern Hemisphere Blocking," *Journal of Climate* 21, 3068-3077.

- Estrada, F., C. Gay, and A. Sánchez (2010). “A Reply to ‘Does Temperature Contains a Stochastic Trend? Evaluating Conflicting Statistical Results’ by R. K. Kaufmann et al,” *Climatic Change* 101, 407-414.
- Estrada, F. and P. Perron (2012). “Breaks, Trends and the Attribution of Climate Change: A Time-Series Analysis,” Department of Economics, Boston University, *mimeo*.
- Estrada, F. and P. Perron (2014). “Detection and Attribution of Climate Change Through Econometric Methods,” *Boletín de la Sociedad Matemática Mexicana* 20, 107-136.
- Estrada, F., P. Perron, C. Gay-García, B. Martínez-López (2013). “A Time-Series Analysis of the 20th Century Climate Simulations Produced for the IPCC’s Fourth Assessment Report,” *PLoS ONE* 8, e60017.
- Fomby, T.B. and T.J. Vogelsang (2002). “The Application of Size-Robust Trend Statistics to Global-Warming Temperature Series,” *Journal of Climate* 15, 117-123.
- Gao, J. and K. Hawthorne (2006). “Semiparametric Estimation and Testing of the Trend of Temperature Series.” *The Econometrics Journal* 9, 332-355.
- Gay-Garcia, C., F. Estrada, and A. Sánchez (2009). “Global and Hemispheric Temperatures Revisited,” *Climatic Change* 94, 333-349.
- Gordon, A.H. (1991). “Global Warming as a Manifestation of a Random Walk,” *Journal of Climate* 4, 589-597.
- Gordon, A.H., J.A.T. Bye, and R.A.D. Byron-Scott (1996). “Is Global Warming Climate Change?” *Nature* 380, 478.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010). “Global Surface Temperature Change,” *Reviews of Geophysics* 48, RG000354.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Jun, M. and M.L. Stein (2008). “Nonstationary Covariance Models for Global Data,” *The Annals of Applied Statistics* 2, 1271-1289.
- Kärner, O. (1996). “Global Temperature Deviations as a Random Walk,” *Journal of Climate* 9, 656-658.

- Kaufmann, R.K., H. Kauppi, M.L. Mann, and J.H. Stock (2013). “Does Temperature Contain a Stochastic Trend: Linking Statistical Results to Physical Mechanisms,” *Climatic Change* 118, 729-743.
- Kaufmann, R.K., H. Kauppi, and J.H. Stock (2006a). “Emissions, Concentrations and Temperature: A Time Series Analysis,” *Climatic Change* 77, 249-278.
- Kaufmann, R.K., H. Kauppi, and J.H. Stock (2006b). “The Relationship Between Radiative Forcing and Temperature: What Do Statistical Analyses of the Instrumental Temperature Record Measure?” *Climatic Change* 77, 279-289.
- Kaufmann, R.K., H. Kauppi, and J.H. Stock (2010). “Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results,” *Climatic Change* 101, 395-405.
- Kaufmann, R.K. and D.I. Stern (2002). “Cointegration Analysis of Hemispheric Temperature Relations,” *Journal of Geophysical Research* 107, 10.129/2000JD000174.
- Leeds, W.B., E.J. Moyer, and M.L. Stein (2015). “Simulation of Future Climate Under Changing Temporal Covariance Structures,” *Advances in Statistical Climatology, Meteorology and Oceanography* 1, 1-14.
- McKittrick, R. and T.J. Vogelsang (2013). “HAC-Robust Trend Comparisons Among Climate Series with Possible Level Shifts,” Department of Economics, Michigan State University, *mimeo*.
- Nordhaus, W. (2013). *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. New Haven: Yale University Press.
- Park, J.Y. and J. Qian, (2012). “Functional Regression of Continuous State Distributions,” *Journal of Econometrics* 167, 397-412.
- Weitzman, M.L. (2009). “On Modeling and Interpreting the Economics of Catastrophic Climate Change,” *The Review of Economics and Statistics* 91, 1-19.
- Woodward, W.A. and H.L. Gray (1993). “Global Warming and the Problem of Testing for Trend in Time Series Data,” *Journal of Climate* 6, 953-962.
- Woodward, W.A. and H.L. Gray (1995). “Selecting a Model for Detecting the Presence of a Trend,” *Journal of Climate* 8, 1929-1937.

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

Zheng, X., R.E. Basher, and C.S. Thompson (1997). "Trend Detection in Regional-Mean Temperature Series: Maximum, Minimum, Mean, Diurnal Range, and SST," *Journal of Climate* 10, 317-326.