



# Seasonal Changes in Central England Temperatures

# Tommaso Proietti and Eric Hillebrand

# **CREATES Research Paper 2015-28**

Department of Economics and Business Aarhus University Fuglesangs Allé 4 DK-8210 Aarhus V Denmark Email: oekonomi@au.dk Tel: +45 8716 5515

# Seasonal Changes in Central England Temperatures

Tommaso Proietti<sup>\*</sup>

Università di Roma "Tor Vergata" and CREATES

Eric Hillebrand<sup>†</sup>

University of Aarhus and CREATES

June 1, 2015

#### Abstract

The aim of this paper is to assess how climate change is reflected in the variation of the seasonal patterns of the monthly Central England Temperature time series between 1772 and 2013. In particular, we model changes in the amplitude and phase of the seasonal cycle. Starting from the seminal work by Thomson ("The Seasons, Global Temperature and Precession", Science, 7 April 1995, vol 268, p. 59–68), a number of studies have documented a shift in the phase of the annual cycle implying an earlier onset of the spring season at various European locations. A significant reduction in the amplitude of the seasonal cycle is also documented. The literature so far has concentrated on the measurement of this phenomenon by various methods, among which complex demodulation and wavelet decompositions are prominent. We offer new insight by considering a model that allows for seasonally varying deterministic and stochastic trends, as well as seasonally varying autocorrelation and residual variances. The model can be summarized as containing a permanent and a transitory component, where global warming is captured in the permanent component, on which the seasons load differentially. The phase of the seasonal cycle, on the other hand, seems to follow Earth's precession in a stable manner, and the reported fluctuations are identified as transitory.

*Keywords*: Global Warming, Seasonal Models, Structural Change, Amplitude and Phase Shifts.

JEL codes: C22.

<sup>\*</sup>Corresponding Author. Dipartimento di Economia e Finanza, via Columbia 2, 00133 Rome, Italy. Email: tommaso.proietti@uniroma2.it Tommaso Proietti gratefully acknowledges financial support by the Italian Ministry of Education, University and Research (MIUR), PRIN Research Project 2010-2011 - prot. 2010J3LZEN, Forecasting economic and financial time series.

<sup>&</sup>lt;sup>†</sup>Aarhus University, Department of Economics and Business CREATES - Center for Research in Econometric Analysis of Time Series, Fuglesangs Allé, 4 8210 Aarhus V, Denmark. Email: EHillebrand@creates.au.dk. Both authors gratefully acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation under grant no. DNRF78.

#### 1 Introduction

In this paper we investigate whether climate change is accompanied by a systematic variation of the phase of the annual cycle in temperatures, which would be responsible for the change in the seasonal timing of events, such as the onset of spring in the continental areas of the northern hemisphere. This theme was investigated by Thomson (1995), who examined the Central England Temperature (CET) series, along with other series, and identified a structural break in the phase of the fundamental seasonal cycle in temperatures (one cycle per year) taking place around 1940; he further showed that the phase shift is correlated with atmospheric  $CO_2$  concentrations. He argued that unpredictability of the phase can represent a more serious problem for biological processes and agriculture than amplitude changes or an average temperature rise.

Stine et al. (2009) estimate the phase shift to have been 1.7 days between 1954 and 2007 over extra-tropical land, whereas Thackeray et al. (2010) found that the temperature cycle advanced by the rate of 2.3 to 5.5 days per decade. Paluvs et al. (2005) confirm the advancement of the spring season in the 1990s, using daily mean near-surface temperature series from seven European locations at mid-latitudes. However, they regard this as a natural phenomenon taking place also in the more remote past. Phenological records seem to support the earlier onset of spring; see, among others, Schwartz and Reiter (2000) and Parmesan (2007).

In contrast to the phase, the amplitude of the seasonal cycle has been decreasing in continental areas of the northern hemisphere. The negative trend is related to the fact that winter months are warming more rapidly than the summer months (Stine et al., 2009).

The principal aim of our research is the characterisation of seasonality in the CET series, along with its global trend. There are several methods for exploring the variation in amplitude and phase, for example complex demodulation (Bloomfield, 2004). We propose a structural time series model as an alternative methodology for estimating changes in amplitude and phase. For validation of our methodology, we rely on the capability of our model to describe the data and to explain most of the variation of temperature, leaving out uninformative residual variation.

We focus on the CET series of monthly mean surface air temperatures for a location in the Midlands region. The series is expressed in degrees Celsius and is available for the period from January 1659 to the present. It can be downloaded at

#### http://www.metoffice.gov.uk/hadobs/ hadcet/data/download.html.

The series was originally constructed by Manley (1974), and is currently updated by the Hadley

Centre; see Parker et al. (1992). There is a quality issue relating to the data prior to 1772, and thus we consider the series from January 1772 to December 2013, for a total of 2904 observations. See also Thomson (1995) for a brief discussion of this issue. The CET series provides one of the longest and most investigated temperature records available at the monthly frequency of observation. As such, its 242 replications of the annual cycle are an invaluable source of information on the the long run trends in the amplitude and the phase of the seasonal cycle.

The series is plotted in the first panel of Figure 1. As is evident from the plot, the series is highly seasonal, with the fundamental trigonometric cycle playing a major role. As stated by Pezzulli et al. (2005), seasonal variation is the most prominent source of climate variability, so that climate change is reflected in the seasonal patterns. As a consequence, the study of seasonal shifts should play an important role in assessing the direction of climate change. The bottom panel of Figure 1 is a plot of the 12 annual series for each month (a monthplot); it highlights that there is some specificity in the way the series pertaining to different months behave. We will return to this feature.

Figure 1: Time series plot of the CET series and its monthplot.



Let us consider the cyclical model

$$y(t) = A(t)\cos\left(\omega t - \varphi(t)\right) + \epsilon(t),$$

where t denotes time. In this model, y(t) has a systematic part, a cycle with period  $2\pi/\omega$ ,  $\omega \in [0, \pi]$  representing the frequency of the cycle in radians (e.g. 12 months if t is expressed in months and  $\omega = \pi/6$ ), amplitude A(t), and phase  $\varphi(t) \in [0, \pi]$ . The systematic part is contaminated by an error term,  $\epsilon(t)$ , which is a stationary random process. The possibly timevarying amplitude regulates the displacement of the cycle along the vertical axis, whereas the phase determines a horizontal displacement (along the time axis). Assuming that there is no noise contamination ( $\epsilon(t) \equiv 0$ , for all t), and that  $A(t) = A, \varphi(t) = \varphi$ , then y(t) is a deterministic cycle. While the role of A is self-evident, that of  $\varphi$  can be understood by evaluating the times  $t_0$  at which the cycle crosses the zero line, that is  $y(t_0) = 0$ . The solutions to  $\cos(\omega t_0 - \varphi) = 0$ are of the form

$$t_0 = \frac{1}{\omega} \left[ \varphi + (2k+1)\frac{\pi}{2} \right],\tag{1}$$

where k is an integer. Equation (1) highlights the role of the phase in regulating the shift along the time axis; if the phase changes by an amount  $d\varphi$ , this will result in a displacement of the cycle wave along the time axis of an amount  $dt_0 = \omega^{-1} d\varphi$ .

The research question posed by the literature reviewed above is whether the global temperature trend has affected the amplitude and the phase of the annual cycle ( $\omega = \pi/6$ ).

The plan of the paper is the following: in the next section we study the preliminary empirical evidence on the time-varying amplitude and phase of the temperatures seasonal cycle arising from the application of complex demodulation and trigonometric regression (Section 2). In Section 3 stationarity tests are applied to the 12 annual time series for the individual months, as well as multivariate stationarity and common trend tests. These provide an important aid to the specification of a stochastic model for temperatures that addresses the main stylized facts. Our baseline model, formulated in Section 4, features a permanent component and a periodic transitory component. The permanent component is driven by a single common stochastic trend as well as by season-specific deterministic drifts. The estimation results and specification diagnostics are presented in Section 5. In Section 6 we draw our conclusions.

## 2 Complex demodulation and trigonometric regression

Complex demodulation (CD) has been used for assessing the changes in the seasonal pattern and in particular for estimating the local amplitude and phase of the fundamental seasonal cycle. Applications to temperature series can be found in Thomson (1995) and Thompson and Clark (2008), for example.

Let  $y_t$ , t = 1, ..., n, denote the monthly temperature series, characterised by an annual cycle, with period equal to p = 12 months. A shift in the frequency of the cycle is performed by multiplying the series with the complex exponential at the fundamental seasonal frequency,  $\omega = \pi/6$ , corresponding to one cycle per year, yielding  $x_t = y_t \exp(-i\omega t)$ , where *i* is the imaginary unit,  $i^2 = -1$ . This aims at shifting the spectrum of  $y_t$  by  $-\omega$  along the frequency range. Subsequently, a low-pass filter is applied to the series  $x_t$  and the amplitude and phase of the resulting complex series are computed.

Referring to Bloomfield (2004) for more details, CD operationally works as follows:

• Transform the original time series (complex demodulation) by computing  $x_t = y_t \exp(-\iota \omega t)$ . Let  $x_t = c_t - \iota s_t$ , where

$$c_t = y_t \cos(\omega t), \quad s_t = y_t \sin(\omega t).$$

• Smooth the series  $c_t$  and  $s_t$  by applying a low-pass filter with cutoff frequency  $\omega_c < \omega$ . A possibility is using the moving average filter

$$w(L) = \frac{1}{12} \left( \frac{1}{2} L^{-6} + L^{-5} + \dots L^{-1} + 1 + L + \dots + L^{5} + \frac{1}{2} L^{6} \right),$$

adopted in seasonal adjustment procedures such as the US Census Bureau X-12 procedure (Findley, Monsell, Bell, Otto, and Chen, 1998), where L is the lag operator,  $L^j y_t = y_{t-j}$ , or its convolutions  $[w(L)]^k$ , k = 1, 2, 3, ... Denote  $\tilde{c}_t = w(L)c_t$ ,  $\tilde{s}_t = w(L)s_t$ .

• Compute the amplitude and phase

$$\tilde{A}_t = (\tilde{c}_t^2 + \tilde{s}_t^2)^{1/2}, \quad \tilde{\varphi}_t = \arctan\left(-\frac{\tilde{s}_t}{\tilde{c}_t}\right).$$

An alternative approach to estimating the phase and the amplitude is based on the following trigonometric regression model:

$$y_{t+ms} = \mu_{M+m} + \beta_{M+m} \cos(\omega t) + \beta^*_{M+m} \sin(\omega t) + \epsilon_{t+ms}, \ t = 1, 2, \dots, Ms,$$

where  $\epsilon_t \sim WN(0, \sigma^2)$ , where WN denotes *white noise*, a sequence of uncorrelated zero mean random variables with constant variance. The model is estimated by ordinary least squares (OLS) using *Ms* consecutive observations referring to a span of *M* years, for t = 1, 2, ..., Ms. The regression is carried out for  $m = 0, 1, ..., \lfloor n/s \rfloor - M$  and provides an estimate of the varying amplitude and phase for year M + m:

$$\hat{A}_{M+m} = (\hat{\beta}_{M+m}^2 + \hat{\beta}_{M+m}^*)^{1/2}, \hat{\varphi}_{M+m} = \arctan\left(\frac{\hat{\beta}_{M+m}^*}{\hat{\beta}_{M+m}}\right), m = 0, 1, \dots, \lfloor n/s \rfloor - M.$$

This method is equivalent to that used by Stine et al. (2009), who consider M = 1 in the methodological appendix and M = 21 in the supplement.<sup>1</sup>

The upper plots of Figure 2 display the amplitude and phase estimated by applying CD using respectively k = 3 and k = 20 convolutions of the low-pass filter w(L), so as to achieve two different levels of smoothness. The phase is expressed in days, that is, in light of Equation (1), we plot the estimate of  $30.4375 \times (\varphi + \pi/2)/\omega$ , which converts the phase shift expressed in months into days. The lower plots display the patterns of  $\hat{A}_{M+m}$  and  $\hat{\varphi}_{M+m}$  for M = 3 and M = 21, which are naturally much smoother for M = 21 (21 years of consecutive data are used to conduct the above trigonometric regression).

The phase plot is directly comparable with figure S8 in the supplement to Stine et al. (2009) and displays a sizable drop during the 1960s and 70s, which possibly marks a significant change in seasonality entailing an earlier onset of spring. The amplitude is characterised by a turning point taking place around 1920, when the decreasing trend is discontinued, if not reversed, for a period before 1950. These facts are also discernible from complex demodulation. The main differences are related to the fact that CD yields noisier time series that are affected by intrayearly volatility, which could be reduced or eliminated by averaging or subsampling. Also, the presence of a trend in mean temperatures affects the estimated phases and amplitudes differently. In trigonometric regression, the variation in the intercept,  $\mu_{M+m}$ , captures the trend in the mean.

In conclusion, complex demodulation can be implemented in various ways so that the estimates of the amplitude and the phase of the seasonal cycle can be characterised by different degrees of smoothness. However, there is broad agreement on the following stylized facts. Firstly, the seasonal cycle has been characterised by a time-varying amplitude with a distinctive trend towards a reduction. Secondly, the phase oscillates around a linear trend with negative slope and features a local rapid decline during the 1960s, possibly followed by a subsequent increase.

<sup>&</sup>lt;sup>1</sup>Note that Stine et al. (2009) use  $\cos(\omega(t-0.5))$  and  $\sin(\omega(t-0.5))$  as explanatory variables.



Figure 2: Complex Demodulation: time varying amplitude and phase.

## 3 Are temperatures stationary or unstable?

Let  $y_{i\tau}$ ,  $\tau = 1, ..., T$ , denote a generic annual temperature series relating to a particular season, i = 1, ..., 12 (January, February, and so forth). For each month i we are interested in testing the stationarity of  $y_{i\tau}$  with respect to  $\tau$  against the alternative that  $y_{i\tau}$  has a stochastic trend component. The reference model is the following:

$$y_{i\tau} = \mu_{i\tau} + \epsilon_{i\tau}, \qquad \epsilon_{i\tau} \sim \text{IID N}(0, \sigma_{\epsilon i}^2),$$
  

$$\mu_{i\tau} = \mu_{i,\tau-1} + \beta_i + \eta_{i\tau}, \quad \eta_{i\tau} \sim \text{IID N}(0, \sigma_{ni}^2),$$
(2)

where we assume that  $\epsilon_{i\tau}$  and  $\eta_{i\tau}$  are independent and the process  $\mu_{i\tau}$  has started at time 0 with value  $\mu_0$ . The single equation form of (2) is then

$$y_{i\tau} = \mu_{0i} + \beta_i \tau + \sum_{j=0}^{\tau-1} \eta_{i,\tau-j} + \epsilon_{i\tau}.$$

The null hypothesis is  $H_0: \sigma_{\eta i}^2 = 0$  for each individual *i*, implying that the series is trend stationary, that is  $y_{i\tau} = \mu_{0i} + \beta_i \tau + \epsilon_{i\tau}$ . The alternative is  $H_1: \sigma_{\eta i}^2 > 0$ , in which case the series displays a stochastic trend. The locally best invariant (LBI) test statistic for this testing problem is (Nyblom, 1986):

$$\xi_{i\tau} = \frac{1}{Ts_i^2} \sum_{j=1}^T \left[ \sum_{\tau=1}^j e_{i\tau} \right]^2,$$

where  $e_{i\tau}$  is the residual from a linear trend fitted by ordinary least squares (OLS),  $e_{i\tau} = y_{i\tau} - \hat{\mu}_{0i} - \hat{\beta}_i \tau$ , and  $s_i^2 = T^{-1} \sum_{\tau} e_{i\tau}^2$ . The rejection region is defined by  $\xi_{i\tau} > c_{\tau}$ , with critical values tabulated by Nyblom (1986). The asymptotic distribution is a second-level Cramer-von Mises distribution, for which the 5% critical value is 0.149.

If  $\beta_i = 0$  both under the null and the alternative, we are testing level stationarity,  $y_{i\tau} = \mu_{0i} + \epsilon_{i\tau}$ , against a pure random walk trend. The LBI test statistic (Nyblom and Mäkeläinen, 1983), denoted  $\xi_{i\mu}$ , takes the same form as  $\xi_{i\tau}$ , with  $e_{i\tau} = y_{i\tau} - \bar{y}_i$ ; the asymptotic distribution is a first-level Cramer-von Mises distribution, for which the 5% critical value is 0.461.

When  $\epsilon_{i\tau}$  is any stationary process, Kwiatkowski *et al.* (1992) show that the same asymptotic distribution is attained under the null, if  $s_i^2$  is replaced by an estimate of the long-run variance such as the Newey and West (1987) estimate

$$s_{iL}^2 = \hat{\gamma}_i^e(0) + 2\sum_{k=1}^l \left(1 - \frac{k}{l+1}\right) \hat{\gamma}_i^e(k) \tag{3}$$

with truncation parameter l, where  $\hat{\gamma}_i^e(k)$  is the autocovariance at lag k of the residuals  $e_{i\tau}$ .

The tests  $\xi_{i\mu}$  and  $\xi_{i\tau}$  can be applied to the annual contrasts  $y_{i\tau} - \frac{1}{12} \sum_{i=1}^{12} y_{i\tau}$ . This provides a test of the null of bivariate cointegration of the temperatures of a particular season with the yearly average temperature series, with cointegration vector given by [1, -1].

Table 1 reports the values of the stationarity tests  $\xi_{i\mu}$  and  $\xi_{i\tau}$  for the annual temperature series for the 12 months and for the annual temperature contrasts, obtained as the difference between the series for a particular month with the yearly average temperature. The tests  $\xi_{i\mu}$ conducted on the individual time series, using different values of the truncation parameter, l = 0, 4, 14, lead to rejecting stationarity in several occurrences. The evidence is likely to be unchanged if we adjusted the critical values so as to account for the multiplicity of the testing problem. Trend stationarity is rejected for May, July, August, September and October ( $\xi_{i\tau} > c_{\tau}$ ). For January, May, June, October and November, we reject the null that the series for month *i* and the average annual temperature series are cointegrated with cointegrating vector [1, -1] according to the test  $\xi_{i\mu}$ . For January, August and October, we reject the same null cointegration in the presence of a linear trend.

Table 1:	Stationarity	tests of	temperature	series
			1	

	Annual Temperature Series						
	$\xi_{i\mu}$	test stati	$istic^{a)}$	$\xi_{i au}$	$istic^{b}$		
	l = 0	l = 4	l = 14	l = 0	l = 4	l = 14	
January	2.371	1.762	1.097	0.074	0.077	0.082	
February	0.096	0.102	0.116	0.043	0.046	0.054	
March	1.568	1.183	0.819	0.121	0.112	0.119	
April	0.408	0.288	0.227	0.134	0.098	0.081	
May	0.247	0.203	0.174	0.203	0.168	0.144	
June	0.193	0.192	0.168	0.114	0.115	0.102	
July	0.533	0.449	0.368	0.214	0.188	0.169	
August	0.961	0.670	0.447	0.405	0.299	0.221	
September	1.483	1.221	0.808	0.244	0.240	0.208	
October	2.545	1.665	0.878	0.393	0.335	0.227	
November	3.031	2.081	1.080	0.102	0.108	0.089	
December	1.299	1.152	0.796	0.043	0.047	0.049	

a) Stationarity test against a random walk. The 5% and 1% critical values are 0.461 and 0.743, respectively.
b) Stationarity test against a random walk with drift. The 5% and 1% critical values are 0.149 and 0.218, respectively.

Table 2:	Stationarity	tests of	temperature	series	con-
trasts wit	h average year	ly tempe	ratures		

	Annual Temperature Contrasts							
	$\xi_{i\mu}$	test stati	$istic^{a)}$	$\xi_{i au}$	$(stic^{b})$			
	l = 0	l = 4	l = 14	l = 0	l = 4	l = 14		
January	1.161	0.959	0.680	0.190	0.181	0.158		
February	0.344	0.355	0.340	0.088	0.095	0.101		
March	0.230	0.247	0.314	0.028	0.031	0.048		
April	0.349	0.296	0.285	0.040	0.035	0.039		
May	0.843	0.701	0.602	0.054	0.051	0.059		
June	2.233	1.755	1.078	0.033	0.036	0.037		
July	0.349	0.391	0.402	0.062	0.074	0.089		
August	0.352	0.331	0.285	0.185	0.179	0.161		
September	0.061	0.075	0.096	0.061	0.075	0.097		
October	0.556	0.521	0.395	0.180	0.179	0.150		
November	0.880	0.852	0.568	0.038	0.043	0.036		
December	0.318	0.355	0.295	0.103	0.120	0.109		

 $^{\rm a)}$  Stationarity test against a random walk. The 5% and 1% critical

<sup>b)</sup> Stationarity test against a random wark. The 5% and 1% critical values are 0.461 and 0.743, respectively.
<sup>b)</sup> Stationarity test against a random walk with drift. The 5% and 1% critical values are 0.149 and 0.218, respectively.

The stationarity test has been extended to the multivariate case by Nyblom and Harvey (2000). Let us consider the multivariate annual temperature series resulting from stacking the series for the individual months,  $\mathbf{y}_{\tau} = \{y_{i\tau}, i = 1, ..., N\}$ , with N = 12, and consider the random walk plus noise model:

$$\begin{aligned} \mathbf{y}_{\tau} &= \boldsymbol{\mu}_{\tau} + \boldsymbol{\epsilon}_{\tau}, \qquad \boldsymbol{\epsilon}_{\tau} \sim \text{IID N}(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon}), \\ \boldsymbol{\mu}_{\tau} &= \boldsymbol{\mu}_{\tau-1} + \boldsymbol{\beta} + \boldsymbol{\eta}_{\tau}, \quad \boldsymbol{\eta}_{\tau} \sim \text{IID N}(\mathbf{0}, \boldsymbol{\Sigma}_{\eta}). \end{aligned}$$

$$(4)$$

Nyblom and Harvey (2000) consider the LBI test of the hypothesis  $H_0$ :  $\Sigma_{\eta} = 0$  against the homogeneous alternative  $H_1$ :  $\Sigma_{\eta} = q\Sigma_{\epsilon}$ , and show that this has rejection region:

$$\xi_{N,\tau} = \operatorname{trace}[\mathbf{S}^{-1}\mathbf{C}] > c,$$

where

$$\mathbf{C} = \frac{1}{T^2} \sum_{\tau=1}^{T} \left[ \sum_{j=1}^{\tau} \mathbf{e}_j \right] \left[ \sum_{j=1}^{\tau} \mathbf{e}_j \right]', \quad \mathbf{S} = \frac{1}{T} \sum_{\tau=1}^{T} \mathbf{e}_\tau \mathbf{e}_\tau' = \hat{\mathbf{\Gamma}}_e(0),$$

 $\mathbf{e}_{\tau}$  results from OLS detrending of the individual series and  $\hat{\mathbf{\Gamma}}_{e}(k)$  is the sample cross-covariance matrix at lag k. They provide the critical values, c, for  $N \leq 4$  and show that the asymptotic distribution is the second-level Cramer-von Mises with N degrees of freedom. When  $\boldsymbol{\beta} = 0$ , the test statistic is denoted  $\xi_{N,\mu}$  and is computed on  $\mathbf{e}_{t} = \mathbf{y}_{t} - \bar{\mathbf{y}}$ .

A non-parametric adjustment, along the lines of Kwiatkowski et al. (1992), can be made when  $\epsilon_{\tau}$  is a general serially correlated and heteroscedastic process; one possibility is to replace **S** by the estimator of the long-run covariance matrix:

$$\mathbf{S}_L = \hat{\mathbf{\Gamma}}_e(0) + \sum_{k=1}^l \left(1 - \frac{k}{l+1}\right) [\hat{\mathbf{\Gamma}}_e(k) + \hat{\mathbf{\Gamma}}_e(k)'].$$

If the null is rejected, it is interesting to test the null that the nonstationarity is due to the presence of a common stochastic trend. Nyblom and Harvey (2000) show that a test for a specified number of common trends,  $H_0$ : rank( $\Sigma_\eta$ ) = K, against the alternative that there are more,  $H_1$ : rank( $\Sigma_\eta$ ) > K, is given by the sum of the N - K smallest eigenvalues of the matrix  $\mathbf{S}_L^{-1}\mathbf{C}$ .

Also, the test statistics  $\xi_{N-1,\mu}$  and  $\xi_{N-1,\tau}$  can be used to test the null of cointegration on the 11 contrasts obtained as the deviation of the annual temperatures of month *i* with a reference month, e.g. December. Denoting  $\mathbf{A} = [\mathbf{I}_{N-1} - \mathbf{i}_N]$ , the multivariate stationarity test computed on  $\mathbf{z}_t = \mathbf{A}\mathbf{y}_t$  is a test of the null that the nonstationary series  $\mathbf{y}_{\tau}$  share the same stochastic trend, i.e. not only there is a single common trend, but the trend enters the series with exactly the same loading.

Both multivariate stationarity tests  $\xi_{N\mu}$  and  $\xi_{N\tau}$ , reported in Table 3 for truncation lags 0, 1, and 4, lead to rejection of the null of stationarity. (The optimal truncation lag, determined according to the method proposed by Hobijn and Franses (2000), is equal to 1). However, if we allow for a linear trend plus drift, the hypothesis that  $\Sigma_{\eta}$  has rank 1 is accepted. We may tentatively conclude that the nonstationarity is due to the presence of a single common trend.

The research question is not only if the monthly and annual temperature series have a common trend, but also if the trend is the same for all the seasons. If the seasons have different trends, then a phase shift of the seasonal cycle may take place. We thus focus on testing the stationarity of  $\mathbf{Ay}_{\tau}$ , which is accepted when we allow for a linear trend in the mean (test  $\xi_{N-1,\tau}$  in the table). This evidence is in contrast with the rejection of the same trend hypotheses rejected in the univariate setup; see the discussion concerning Table 2.

Annual Temperature Series								
	$\xi_{N\mu}$ test stat	$tistic^{a)}$		$\xi_{N\tau}$ test statistic <sup>b)</sup>				
l = 0	l = 1	l = 4	l = 0	l = 1	l = 4			
$8.573^{*}$	$6.779^{*}$	$4.471^{*}$	$1.336^{*}$	$1.242^{*}$	1.040			
$H_0$ :	$H_0: \operatorname{rank}(\boldsymbol{\Sigma}_{\eta}) = 1: \text{ constant, no trend}^{c)}$			$H_0: \operatorname{rank}(\boldsymbol{\Sigma}_{\eta}) = 1: \text{ linear trend}^{\mathrm{d})}$				
l = 0	l = 1	l = 4	l = 0	l = 1	l = 4			
$1.144^{*}$	$1.115^{*}$	$0.974^*$	0.388	0.386	0.384			
	Annu	al Temperature Con	trasts $\mathbf{A}\mathbf{y}_{\tau}, \mathbf{A}$	$= [\mathbf{I}_{N-1} - \mathbf{i}_N]$				
	$\xi_{N-1,\mu}$ test sta	$atistic^{e)}$		$\xi_{N-1\tau}$ test sta	$tistic^{f}$			
l = 0	l = 1	l = 4	l = 0	l = 1	l = 4			
5 596*	5 123*	3 860*	0.850	0.847	0.868			

Table 3: Multivariate Stationarity tests of temperature series and their contrasts

<sup>a)</sup> The 5% and 1% critical values are 2.901 and 3.396, respectively.

<sup>b)</sup> The 5% and 1% critical values are 1.059 and 1.180, respectively.

 $^{\rm c)}$  The 5% and 1% critical values are 0.941 and 1.065, respectively.

 $^{\rm d)}$  The 5% and 1% critical values are 0.552 and 0.609, respectively.

 $^{\rm f)}$  The 5% and 1% critical values are 2.696 and 3.221, respectively.

 $^{\rm g)}$  The 5% and 1% critical values are 0.981 and 1.114, respectively.

The main conclusions that we draw from the analysis of the stability of the temperature series is that trend stationarity is rejected. The most likely source for the rejection is the presence of a nonstationary random component driving all the 12 series pertaining to each individual months. In the appendix we report the evidence arising from the application of seasonal stability tests, which point to the presence of stochastic trend component at the long run frequency. Whether the common stochastic trend affects the months in the same way or in different ways remains an open issue. We will return to this issue in Section 5.

#### 4 Model specification

Our preliminary analysis suggests that we need to model the following characteristics:

- Seasonal heteroscedasticity and periodicity: the temperature in the winter months is more variable and the serial dependence varies with the season (see Figure 1).
- The seasons may drift at a different rate (Figure 1).
- A global (warming) trend affects all the seasons, possibly with a heterogenous response (Section 3).

We formulate a specification which decomposes monthly temperatures series into a permanent and a transitory component. Using the trigonometric regression method outlined in Section 2, we can investigate the changes in phase and amplitude in the permanent component. The permanent component, denoted  $\mu_t$ , has a deterministic linear trend component with seasonspecific intercept and slope, and a stochastic component that results from the amplification or contraction of a global stochastic trend. In symbols, for season *i*, the permanent component is

$$\mu_t = \mu_{0i} + \beta_i t + \theta_i \delta_t,$$

where  $\delta_t$  is the common trend affecting all the seasons and  $\theta_i$  is the loading for the *i*-th season. We expect a large part of the movements in the series to be transitory: a winter colder than usual does not affect the seasonal pattern permanently. We model the transitory component,  $\psi_t$ , as a periodic and heteroscedastic AR(1) process.

In summary, the model has the following formulation:

$$y_t = \mu_t + \psi_t, \qquad t = 1, 2, \dots, T,$$
  

$$\mu_t = \mathbf{x}'_t(\boldsymbol{\mu}_0 + \boldsymbol{\beta}t + \boldsymbol{\theta}\delta_t), \qquad (5)$$
  

$$\delta_t = \delta_{t-1} + \eta_t, \qquad \eta_t \sim N(0, 1),$$
  

$$\psi_t = (\mathbf{x}'_t \boldsymbol{\phi})\psi_{t-1} + \zeta_t, \qquad \zeta_t \sim N(0, (\mathbf{x}'_t \boldsymbol{\nu})^2).$$

All the disturbances are serially and mutually uncorrelated. Here,  $\mathbf{x}_t$  denotes a 12 × 1 selection vector, taking value 1 in the position corresponding to the current season and zero elsewhere,

so that if the *t*-th observation refers to month j,  $\mathbf{x}'_t \boldsymbol{\beta} = \beta_j$ ,  $\mathbf{x}'_t \boldsymbol{\nu} = \nu_j$ , and so forth. The vector is characterised by the periodic property  $\mathbf{x}_t = \mathbf{x}_{t-12}$ .

The common global trend  $\delta_t$  is a random walk with no drift, assumed to have started at time 0 with  $\delta_0 = 0$  (any initial effect  $\theta \delta_0$  would be incorporated in the intercept  $\mu_0$ , and thus cannot be identified separately). Hence, the permanent component features a stochastic trend and a deterministic trend, and both trends have season-specific aspects, captured by the vectors  $\beta$  and  $\theta$ .

The transitory component is modelled as a periodically stationary AR(1) process, with seasonally heterogeneous AR coefficients  $(\mathbf{x}'_t \boldsymbol{\phi})$  resulting from the selection of the relevant component of the 12 × 1 vector  $\boldsymbol{\phi}$ .

Let  $\boldsymbol{\mu}_0 = \mathbf{i}\bar{\mu}_0 + \boldsymbol{\mu}_0^*$ ,  $\boldsymbol{\beta} = \mathbf{i}\bar{\beta} + \boldsymbol{\beta}^*$ ,  $\boldsymbol{\theta} = \mathbf{i}\bar{\theta} + \boldsymbol{\theta}^*$ , where  $\mathbf{i}$  is a vector of 1s and the scalars  $\bar{\mu}_0, \bar{\beta}, \bar{\theta}$ are the average of the corresponding vectors, e.g.  $\bar{\mu}_0 = \mathbf{i}'\boldsymbol{\mu}_0/12$ , and  $\boldsymbol{\mu}_0^* = \mathbf{N}\boldsymbol{\mu}_0, \boldsymbol{\beta}^* = \mathbf{N}\boldsymbol{\beta}$ , and  $\boldsymbol{\theta}^* = \mathbf{N}\boldsymbol{\theta}$ , with  $\mathbf{N} = \mathbf{I} - \mathbf{i}\mathbf{i}'/12$ . Then, the permanent component can be expressed as

$$\mu_t = \bar{\mu}_t + \gamma_t^D + \gamma_t^S,$$
  

$$\bar{\mu}_t = \bar{\mu}_0 + \bar{\beta}t + \bar{\theta}\delta_t,$$
  

$$\gamma_t^D = \mathbf{x}_t'(\boldsymbol{\mu}_0^* + \boldsymbol{\beta}^*t),$$
  

$$\gamma_t^S = \mathbf{x}_t'\boldsymbol{\theta}^*\delta_t.$$

In other words, the permanent component has a global stochastic trend component,  $\bar{\mu}_t$ , represented by a random walk with drift  $\bar{\beta}$ , a deterministic seasonal component,  $\gamma_t^D$ , featuring seasonal drifts, and a stochastic seasonal component driven by the global temperature trend. This is a systematic sample from a random walk, with weights given by the elements of  $\theta^*$ .

The model nests several interesting special cases:

- The monthly series feature the same stochastic trend: this arises when  $\theta$  is a constant vector, that is  $\theta = \theta \mathbf{i}_N$ . Permanent shocks do not have any effect on the seasonal pattern, but they modify the long run path only. However, the seasons drift at a different rate.
- The monthly series feature the same stochastic trend and the same deterministic drift. This arises when  $\boldsymbol{\theta}$  and  $\boldsymbol{\beta}$  are both constant vectors:  $\boldsymbol{\theta} = \boldsymbol{\theta} \mathbf{i}_N$  and  $\boldsymbol{\beta} = \boldsymbol{\beta} \mathbf{i}_N$ . In this case we can rewrite the permanent component as  $\mu_t = \mathbf{x}'_t \boldsymbol{\mu}_0 + \delta^*_t, \delta^*_t = \delta^*_{t-1} + \boldsymbol{\beta}^* + \eta^*_t, \eta^*_t \sim N(0, \sigma^2_\eta)$ , i.e. a season-specific constant plus a random walk with drift. The seasons have the same drift.

In the unrestricted model, the coefficient vectors  $\mu_0, \beta, \theta, \nu, \phi$  can be functionally related to the season index i = 1, ..., 12, i.e. they vary smoothly across the seasons. This would allow for parsimonious modelling of the coefficients: we can let  $\beta = \mathbf{W}\beta^{\dagger}$ , where  $\mathbf{W}$  is a matrix of spline weights and  $\beta^{\dagger}$  is a small dimensional vector containing the value of the spline at the knots. In particular, we use a periodic spline for modelling a pattern in x = [0, 12] with knots located at the points (0 1 3 7 10 12), see Poirier (1976), so that  $\beta^{\dagger}$  has five elements and

$$\mathbf{W} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.8312 & 0.4694 & -0.0493 & 0.0477 & -0.2989 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.4604 & 1.0853 & 0.2633 & -0.1244 & 0.2362 \\ -0.4910 & 0.8002 & 0.6296 & -0.2368 & 0.2980 \\ -0.2761 & 0.3650 & 0.9311 & -0.2308 & 0.2108 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.1769 & -0.1312 & 0.7378 & 0.4839 & -0.2674 \\ 0.1903 & -0.0914 & 0.3222 & 0.9377 & -0.3588 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 \\ -0.2901 & 0.0410 & -0.0598 & 0.4964 & 0.8124 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

Similarly, we write  $\boldsymbol{\mu}_0 = \mathbf{W} \boldsymbol{\mu}_0^{\dagger}$ ,  $\boldsymbol{\theta}_0 = \mathbf{W} \boldsymbol{\theta}^{\dagger}$ ,  $\boldsymbol{\nu} = \mathbf{W} \boldsymbol{\nu}^{\dagger}$ ,  $\boldsymbol{\phi} = \mathbf{W} \boldsymbol{\phi}^{\dagger}$ . As a result, our baseline specification has 25 parameters. We discuss the robustness of the selection of knots in Section 5.3.

### 5 Estimation results

#### 5.1 Parameter estimates

Our model is represented in state space form and the unknown parameters,

$$m{\xi} = \left( m{\mu}_{0}^{*'}, m{eta}^{*'}, m{ heta}^{*'}, m{\phi}^{*'}, m{
u}^{*'} 
ight)',$$

are estimated by maximum likelihood, with the support of the Kalman filter, see Durbin and Koopman (2012). The maximised likelihood is  $\ell_M = -2208.08$ . The parameter estimates are reported in Table 4 along with their estimation standard errors, which are multiplied by the factor  $10^5$  for convenience.

The first column displays the initial level of temperature for the different months. The main highlights of the estimation results are:

- The estimated slopes,  $\beta$ , are higher in the winter months (January and December, in particular) and November. The parameter estimates are around 10 times the standard error for these seasons. On the contrary, the drift is not significantly different from zero for June and July.
- The estimated average drift,  $\bar{\beta}$ , implies a rise in temperatures amounting to 0.37 degrees Celsius over a century.
- The loadings  $\boldsymbol{\theta}$  on the global stochastic trend,  $\delta_t$ , are higher (in absolute value) for April and May, and August, September and October. The loadings are close to zero for January and December.
- The transitory component has higher variability in the winter months: both the variance  $\nu_i$  and the periodic autoregressive coefficients are higher. The periodic autoregressive coefficients are also higher in the summer months.

Month	$oldsymbol{\mu}_0$	s.e.(*)	β	s.e.(*)	θ	s.e.(*)	$\phi$	s.e.(*)	ν	s.e.(*)
January	3.06	0.02	0.00045	5.75699	0.0021	0.0022	0.35	0.03	2.81	0.10
February	3.73	0.02	0.00037	5.20530	0.0100	0.0059	0.26	0.03	1.94	0.05
March	5.28	0.01	0.00029	4.11056	0.0186	0.0156	0.09	0.01	1.47	0.06
April	8.14	0.01	0.00019	4.21173	0.0218	0.0184	0.07	0.03	1.23	0.11
May	11.61	0.01	0.00010	3.85030	0.0209	0.0148	0.17	0.02	1.13	0.13
June	14.65	0.01	0.00004	3.89015	0.0191	0.0086	0.29	0.01	1.10	0.13
July	16.25	0.01	0.00007	4.16420	0.0193	0.0052	0.34	0.02	1.12	0.10
August	15.65	0.00	0.00020	4.17306	0.0234	0.0086	0.28	0.02	1.17	0.05
September	13.23	0.01	0.00037	4.58062	0.0270	0.0121	0.16	0.03	1.34	0.05
October	9.63	0.02	0.00051	4.28350	0.0243	0.0114	0.08	0.02	1.85	0.09
November	5.70	0.03	0.00056	4.01393	0.0128	0.0054	0.11	0.03	3.00	0.11
December	3.16	0.02	0.00053	4.66956	0.0021	0.0015	0.24	0.01	3.79	0.13
Mean	9.17		0.00031		0.0168		0.20		1.83	
Variance	23.01		0.00000		0.0001		0.01		0.74	

**Table 4:** Parameter estimates and their standard errors (multiplied by  $10^5$ ).

(\*) Multiplied by  $10^5$ .

#### 5.2 The global temperature trend, amplitude, and phase changes

Let  $\tilde{\boldsymbol{\xi}}$  denote the maximum likelihood estimator (MLE) of  $\boldsymbol{\xi}$ , and denote its probability density function with  $f(\tilde{\boldsymbol{\xi}}_s)$ . Let  $\alpha_t$  denote a statistic of interest, e.g. the phase of the permanent component at time t, obtained by the trigonometric regression model outlined in Section 2. We are interested in estimating the conditional density

$$f(\alpha_t | Y_n) = \int f(\alpha_t | \tilde{\boldsymbol{\xi}}, Y_n) f(\tilde{\boldsymbol{\xi}}) \mathrm{d}\tilde{\boldsymbol{\xi}}, \tag{6}$$

where  $Y_n = \{y_1, \ldots, y_n\}$  denotes the available temperature series, or characteristic values, such as the mean, the standard deviation, or a particular quantile. The method of composition (Tanner, 1996) can be used to obtain a Monte Carlo estimate of the conditional density of interest. In particular, we draw a sample  $\tilde{\alpha}_t^{(i)}$ ,  $i = 1, \ldots, R$ , from (6) by the following algorithm:

- 1. Obtain a bootstrap sample  $\tilde{\boldsymbol{\xi}}^{(i)} \sim f(\tilde{\boldsymbol{\xi}})$ : generate a simulated series  $\{y_t^{(i)}, t = 1, \ldots, n\}$  from model (5), setting the parameters equal to  $\tilde{\boldsymbol{\xi}}$ . Estimate model (5) for the generated series to get the required draw,  $\tilde{\boldsymbol{\xi}}^{(i)}$ .
- 2. Simulate  $\tilde{\alpha}_t^{(i)} \sim f(\alpha_t | \tilde{\boldsymbol{\xi}}^{(i)}, Y_n)$ , using the simulation smoother proposed by Durbin and Koopman (2002). In practice, denoting  $Y_n^{(i)} = \{y_1^{(i)}, y_2^{(i)}, \dots, y_n^{(i)}\},\$

$$\tilde{\alpha}_t^{(i)} = \mathbf{E}\left(\alpha_t | Y_n, \tilde{\boldsymbol{\xi}}^{(i)}\right) + \alpha_t^{(i)} - \mathbf{E}\left(\alpha_t^{(i)} | Y_n^{(i)}, \tilde{\boldsymbol{\xi}}^{(i)}\right).$$

The above expectations are evaluated by the Kalman filter and smoother (Durbin and Koopman, 2012).

Setting  $\alpha_t = \bar{\mu}_t$ , we can obtain point and interval estimates by computing the average and the standard deviation of the draws  $\bar{\mu}_t^{(i)}$ , i = 1, ..., R. These are displayed in the upper left plot of Figure 3. The estimates are based on R = 1000 draws and provide an interesting characterisation of the global warming trend in temperatures: the second decade of the 20th century marks the inception of an upward trend that increases the level by one degree Celsius.

The estimates of the transitory component,  $E(\psi_t|Y_n)$ , and 95% bootstrap confidence intervals are plotted for the period from January 2005 to December 2013, with the purpose of illustrating the range and the variability of the periodically stationary component in temperatures. December 2010, which has been labelled "the coldest December in a century," is clearly visible, as well as a positive and high transitory component in 2006, which has been considered as one the warmest recent years (see http://news.bbc.co.uk/2/hi/science/nature/6177663.stm).

The permanent component results from a deterministic cycle with seasonal drifts and from a random component due to the global temperature trend,  $\delta_t$ , affecting the seasons in different ways (due to the differences in the loadings  $\boldsymbol{\theta}$ ). We can assess the dynamics of the amplitude Figure 3: Point and interval estimates of the global trend component  $\bar{\mu}_t = \bar{\mu}_0 + \bar{\beta}t + \bar{\theta}\delta_t$ , and the transitory component (subperiod 2015-2013), conditional on the full sample. Conditional mean and approximate 95% confidence bands for the amplitude and the phase of the permanent component,  $\mu_t$ , according to the trigonometric regression method with M = 3.



and the phase of the permanent component by simulating from the conditional distribution of  $\mu_t$ , given the observed time series, and performing a trigonometric regression or a complex demodulation analysis on the draws. The bottom panels of Figure 3 display the point and interval estimates of the amplitude and phase  $\hat{A}_{M+m}$ , and  $\hat{\varphi}_{M+m}$ , respectively (see Section 2), using M = 3 years of consecutive monthly data. Heuristically, the point estimates can be interpreted as the trend behaviour of the two characteristics, abstracting from the transitory dynamics.

As can be seen in the bottom right panel of Figure 3, the phase of the permanent component closely follows a linear function of time with a negative slope implying a shift of the annual cycle of about minus -4.62 days in the 242 years making up our sample time series, i.e. about minus one month every 1600 years. Therefore, in about 13,000 years the seasons will be reversed. This is the effect of precession of Earth's axis of rotation. Thus, interestingly, the phase shows a

steady downward trend that does not seem to be affected by global warming. Our results do not seem to suggest the presence of a shift in the seasons over and above precession.

Obviously, the phase angle does not admit an additive decomposition into a permanent and a transitory component. However, a weighted additive decomposition exists for the tangent of the phase angle, which is the ratio of the coefficients attached to the sine and the cosine components of the trigonometric representation. Setting  $\tan(\varphi_y) = \beta_y^*/\beta_y$  for  $y_t = c + \beta_y \cos(\omega t) + \beta_y^* \sin(\omega t)$ ,  $\tan(\varphi_{\mu}) = \beta_{\mu}^*/\beta_{\mu}$  for  $\mu_t = c_{\mu} + \beta_{\mu} \cos(\omega t) + \beta_{\mu}^* \sin(\omega t)$ , and  $\tan(\varphi_{\psi}) = \beta_{\psi}^*/\beta_{\psi}$  for  $\psi_t = c_{\psi} + \beta_{\psi} \cos(\omega t) + \beta_{\psi}^* \sin(\omega t)$ , it holds

$$\tan(\varphi_y) = w_\mu \tan(\varphi_\mu) + (1 - w_\mu) \tan(\varphi_\psi),$$

for  $w_{\mu} = \beta_{\mu}/(\beta_{\mu} + \beta_{\psi})$ . The decomposition is shown in Figure 4, which is based on the trigonometric regression of  $y_t$ , and the estimated  $\mu_t$  and  $\psi_t$ , using M = 3. The plot shows that the downward movements in  $\tan(\varphi_y)$  between 1950 and 1980 are transitory. The component  $\tan(\varphi_{\mu})$ behaves like a straight line and multiplication by the weight  $w_{\mu}$  does not introduce a systematic reduction in the 1950's. These results lend support to the notion that the changes in the phase of the seasonal cycle could be ascribed to the transitory component.





#### 5.3 Specification diagnostics

The estimation results presented above refer to our selected specification, using a parsimonious parameterisation of the monthly features (intercept, drift, loading, AR coefficient, periodic variance). For each of these features, the 12 effects are expressed as linear combinations, with weights given by the matrix  $\mathbf{W}$ , of five representative points (January, March, July, October and December). The selected specification represents a point in the model space, which consists of all possible numbers and locations of the knots of the periodic spline. It was obtained starting from a configuration in which we had a knot for each season (spring, summer, autumn, winter), selected a representative knot within each season, and later added the most relevant additional knot, which turned out to be January.

In Table 5 we report the maximised log-likelihood,  $\ell(\hat{\boldsymbol{\xi}})$ , associated with alternative, more general specifications, which differ from our selected model by allowing unrestricted estimation of one feature at a time, e.g. the intercept  $\mu$ , while constraining all the remaining parameters to lie on a periodic spline. All these specifications lead to an increase in the Bayesian Information Criterion (BIC), reported in the 5th column,

$$BIC = -2\ell(\hat{\boldsymbol{\xi}}) + \ln(n)\#(\hat{\boldsymbol{\xi}}),$$

where  $\#(\hat{\boldsymbol{\xi}})$  denotes the number of parameters, given in the 3rd column. Leaving the intercepts, the drifts, or the loadings unrestricted reduces the Akaike Information Criterion, AIC =  $-2\ell(\hat{\boldsymbol{\xi}}) + 2\#(\hat{\boldsymbol{\xi}})$ . We also considered the possibility of increasing the complexity of the model by adding an additional knot, for a location corresponding to June, which is the best additional knot for augmenting the spline model.

The second part of Table 5 presents the results of reducing the complexity of the selected specification along meaninful directions. We first consider the case of no seasonal drifts,  $H_0$ :  $\boldsymbol{\beta} = \mathbf{0}$ . This restriction is strongly rejected according to the likelihood ratio (LR) test reported in the 6th column, whose p-value is given in the last column of the table. The second specification restricts the drift to be the same for all the months:  $H_0: \boldsymbol{\beta} = \boldsymbol{\beta} \mathbf{i}$ , for a scalar drift  $\boldsymbol{\beta}$ ; the LR test has a *p*-value close to 5%. Also, we reject the null that the global trend component enters all the seasons with the same weight,  $H_0: \boldsymbol{\theta} = \bar{\boldsymbol{\theta}} \mathbf{i}$ , as well as the null that  $H_0: \boldsymbol{\beta} = \boldsymbol{\beta} \mathbf{i}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}} \mathbf{i}$ .

Specification	Log-likelihood	N. parameters	AIC	BIC			
Selected model	-2208.08	25	4466.16	4615.51			
Unrestricted intercepts	-2194.82	32	4453.64	4644.80			
Unrestricted drifts	-2196.32	32	4456.64	4647.80			
Unrestricted loadings	-2196.42	32	4456.84	4648.00			
Unrestricted AR	-2206.32	32	4476.64	4667.80			
Unrestricted variances	-2208.46	32	4480.92	4672.08			
Additional knot (June)	-2200.20	30	4460.40	4639.62			
Restrictions	Log-likelihood	N. parameters	AIC	BIC	LR	df	p-value
No drifts	-2220.04	20	4480.08	4599.56	23.92	5.00	0.0002
Same drift	-2212.69	21	4467.38	4592.83	9.22	4.00	0.0558
Same loading	-2222.73	21	4487.46	4612.91	29.30	4.00	0.0000
Same drift and loading	-2228.26	17	4490.52	4592.08	40.36	8.00	0.0000

 Table 5:
 Analysis of model specification.

Diagnostic checking for the selected specification can be performed on the standardised innovations of the model,  $r_t = (y_t - E(y_t|Y_{t-1}))/\sqrt{\operatorname{Var}(y_t|Y_{t-1})}$ , computed by the Kalman filter, conditional on the maximum likelihood estimate of the parameters. For a correctly specified model,  $r_t$  is IID normal with zero mean and unit variance. The time series plot is available in the upper panel of Figure 5. The distribution of  $r_t$  does not deviate from normality, as can be visually appreciated from the estimated probability density function, reported in the central panel of Figure 5, and from the value of the Jarque-Bera test statistic (Jarque and Bera, 1987), 0.998, which does not lead to reject the null of normality. The sample autocorrelation function of  $r_t$  is displayed in the bottom panel, for all the lags from 1 to 36. The largest sample autocorrelation occurs at lag 14 and equals -0.055 and the value of the Ljung-Box statistic using 12 lags amounts to 20.350, with a *p*-value of 0.061. We may conclude that the selected specification provides a data-coherent representation of the serial dependence structure of the CET series.

#### 6 Conclusions

The Central England Temperature series provides a very useful dataset for assessing whether the global trend in temperatures, or global warming, has affected the annual cycle. In particular, we have investigated whether a shift of the seasonal cycle towards an earlier inception of the spring season has occurred. The paper has attempted to distinguish between the permanent and the

Figure 5: Standardised innovations: time series plot, density estimate and sample autocorrelation function.



transitory components of the series. The permanent component is driven by a single stochastic trend to which the seasons seem to have responded in a differential way. It is characterised by an upward movement since the beginning of the twentieth century, which is usually identified as a global warming trend. The transitory component is a stationary periodic autoregressive process with seasonally heteroscedastic disturbances. The most important changes in the seasonal cycle concern the amplitude, which appears to be more sensitive to the variations in the global trend. It has been decreasing until around 1920, and later it stabilises. The phase of the seasonal cycle is much less affected by the global warming trend, and it is characterised by a steady downward trend, consistent with precession of Earth's axis of rotation. The rapid movements in the phase taking place around 1950 can be ascribed to the transitory component in temperatures.

# Appendix

In this appendix we report the results of seasonal stability tests and some details on the fit provided by a well known model for seasonal time series, the so called basic structural model, see Harvey (1989).

#### A Seasonal stability test

The econometric literature has provided tests of the stability of the seasonal pattern that are relevant to our discussion. Let us assume that the temperature series admits the decomposition:

$$y_t = \mu_t + \gamma_t + \epsilon_t,$$

where  $\mu_t$  is the trend component,  $\gamma_t$  is the seasonal component, and  $\epsilon_t \sim \text{IID N}(0, \sigma_{\epsilon}^2)$ . Assume further that

$$\mu_t = \mu_t + \beta + \eta_t, \quad \eta_t \sim \text{IID N}(0, \sigma_\eta^2),$$
  

$$\gamma_t = \gamma_t^D + \gamma_t^S.$$
(7)

The seasonal component is decomposed into a deterministic term,  $\gamma_t^D$ , a linear combination with fixed coefficients of sines and cosines defined at the seasonal frequencies  $\omega_j = 2\pi j/12$ , j = 1..., 6, plus a nonstationary stochastic term,  $\gamma_t^S$ , arising as a linear combination of the same explanatory variables with random coefficients:

Defining

$$\mathbf{z}_t = \left[\cos\omega_1 t, \sin\omega_1 t, ..., \cos\omega_5 t, \sin\omega_5 t, ..., \cos\omega_6 t\right]',$$

 $\gamma_t^D = \mathbf{z}_t' \boldsymbol{\gamma}_0$ , where  $\boldsymbol{\gamma}_0$  is a vector of 11 fixed coefficients. The stochastic component is  $\gamma_t^S = \mathbf{z}_t' \sum_{i=1}^t \boldsymbol{\kappa}_i$  where  $\boldsymbol{\kappa}_t$  is a vector of serially independent disturbances with zero mean and covariance matrix

$$\mathbf{\Omega} = \operatorname{diag}(\sigma_1^2, \sigma_1^2, \dots, \sigma_5^2, \sigma_5^2, \sigma_6^2),$$

independently of  $\epsilon_t$  and  $\eta_t$ . The remaining components are defined as before.

Canova and Hansen (1995) and Busetti and Harvey (2003) have derived the locally best invariant test of the null that seasonality is stable versus the alternative that it is stochastically evolving. The null hypothesis that the trigonometric cycle at frequency  $\omega_j$  is deterministic is then formulated as  $H_0: \sigma_j^2 = 0$ , versus  $H_1: \gamma_0 = 0, \sigma_j^2 > 0$  (stochastic seasonality). The test statistic is

$$\varpi_j = \frac{a_j}{T^2 \sigma^2} \sum_{t=1}^T \left[ \sum_{i=1}^t (e_i \cos \omega_j i)^2 + \sum_{i=1}^t (e_i \sin \omega_j i)^2 \right],$$
(8)

where  $a_j = 1, j = 6$  and  $a_j = 2$  otherwise, and  $e_t$  are the OLS residuals obtained from the regression of  $y_t$  on a set of explanatory variables  $x_t = [1, t, \mathbf{z}'_t]'$  accounting for a deterministic trend and deterministic seasonals. Under the null  $\varpi_j$  is asymptotically distributed according to a Cramér-von Mises distribution with 2 degrees of freedom for j = 1, 2, 3, 4, 5 and 1 degree of freedom for j = 6. The test statistic for  $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_6^2$ , is  $\varpi = \sum_j^6 \varpi_j$ . We may also wish to test the stability of the trend component,  $H_0: \sigma_\eta^2 = 0$  (versus  $H_1: \sigma_\eta^2 > 0$ . The test statistic is  $\xi_\tau = \frac{1}{T^2 \sigma^2} \sum_{t=1}^T \left[ \sum_{i=1}^t e_i^2 \right]$ .

Given that we are interested in testing the stability of the seasonal trigonometric cycles in the presence of seasonal trends, we also consider a more general model for the deterministic seasonal component, namely  $\gamma_t^D = \mathbf{z}_t'(\gamma_0 + \gamma_1 t)$ , where  $\gamma_1$  is a vector of seasonal drifts. The test statistic for  $H_0: \sigma_j = 0$ , takes the same form as (8), with  $e_t$  representing the residual from the regression of  $y_t$  on a set of explanatory variables  $x_t = [1, t, \mathbf{z}_t', t\mathbf{z}_t']'$ . We label the test  $\varpi_j^+$ . Its asymptotic null distribution is CvM with time trend; the critical values are available in Table I(b) of Harvey (2001).

The test statistics in (8) require an estimate of  $\sigma^2$ . A nonparametric estimate is obtained by rescaling the estimate of the spectrum of the sequence  $e_t$  at the frequency  $\omega_j$ , by  $2\pi$ , using a Bartlett window.

Table 6 presents the test statistics using a truncation lag l = 24. The Busetti-Harvey Canova-Hansen test leads to reject the null at the zero frequency and at the fundamental frequency. However, when we consider the possibility of seasonal trends, we accept the null of stability.

No	drift	With drift		With drift and seasonal trends		
$\xi_{\mu}$	3.951**	$\xi_{ au}$	$0.324^{**}$	$\xi_{ au}$	$0.324^{**}$	
$\varpi_1$	$1.079^{**}$	$\varpi_1$	$1.416^{**}$	$arpi_1^\dagger$	0.198	
$\varpi_2$	0.072	$\varpi_2$	0.094	$arpi_2^\dagger$	0.053	
$\varpi_3$	0.036	$\varpi_3$	0.047	$arpi_3^\dagger$	0.032	
$arpi_4$	0.105	$arpi_4$	0.139	$arpi_4^\dagger$	0.034	
$\varpi_5$	0.108	$\varpi_5$	0.142	$arpi_5^\dagger$	0.069	
$\varpi_6$	0.066	$\varpi_6$	0.088	$arpi_6^\dagger$	0.039	

 Table 6:
 Busetti-Harvey stability tests

a) The 5% and 1% critical values of the  $\xi_{\mu}$  and  $\varpi_6$  statistics are 0.461 and 0.743, respectively.

<sup>b)</sup> The 5% and 1% critical values of the  $\varpi_j$ , j = 1, 2, 3, 4, 5, statistics are 0.748 and 1.074, respectively.

<sup>c)</sup> The 5% and 1% critical values of the  $\xi_{\tau}$  and  $\varpi_6^{\dagger}$  statistic are 0.149 and 0.218, respectively.

<sup>d)</sup> The 5% and 1% critical values of the  $\varpi_j^{\dagger}$ , j = 1, 2, 3, 4, 5, statistics are 0.247 and 0.329, respectively.

The critical values are obtained from Harvey (2001).

The overall conclusion is that instability seems to characterise only the long-run frequency, once seasonal trends are accounted for.

#### **B** Basic structural model

Representation (7) is often referred to as the basic structural time series model (BSM, see Harvey (1989)). When, consistently with the seasonal stability tests results, it is estimated under the restrictions  $\sigma_2^2 = \sigma_3^2 = \sigma_2^4 = \sigma_5^2 = \sigma_2^6 = 0$  (the trigonometric cycles at the harmonic frequencies are deterministic), we obtain the following maximum likelihood estimates  $\tilde{\sigma}_{\eta}^2 = 0.0047681$ ,  $\tilde{\sigma}_1^2 = 0.0000526$ ,  $\tilde{\epsilon}^2 = 1.8344095$ ,  $\tilde{\beta} = 0.0002287$ . For comparison with the models considered in this paper, we also report the value of the maximised log-likelihood, which equals  $\ell_{BSM} = -2410.38$ . Notice that this is much below the maximised likelihood of the model considered in Section (4),  $\ell_M = -2208.08$ . In particular, the representation of the short-run component, given for the BSM by homoscedastic white noise, appears to be grossly misspecified. This is reflected in the usual diagnostics, based on the standardised Kalman filter innovations, which are periodically autocorrelated and seasonally heteroscedastic.

## References

Bloomfield, P. (2004). Fourier analysis of time series: an introduction. John Wiley & Sons.

- Busetti, F., and Harvey, A. (2003). Seasonality tests. Journal of Business & Economic Statistics, 21(3), 420–436.
- Canova, F., and Hansen, B. E. (1995). Are seasonal patterns constant over time? a test for seasonal stability. Journal of Business & Economic Statistics, 13(3), 237–252.
- Durbin, J., and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3), 603–616.
- Durbin, J., and Koopman, S. J. (2012). Time Series Analysis by State Space Methods (2nd ed.). Oxford: Oxford University Press.
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B.-C. (1998). New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2), 127–152.
- Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.
- Harvey, A. C. (2001). Testing in unobserved components models. Journal of Forecasting, 20(1), 1–19.
- Hobijn, B., and Franses, P. H. (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, 15(1), 59–81.
- Jarque, C. M., and Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. International Statistical Review, 2, 163–172.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1), 159–178.
- Manley, G. (1974). Central england temperatures: monthly means 1659 to 1973. Quarterly Journal of the Royal Meteorological Society, 100(425), 389–405.
- Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, 55(394), 703–708.
- Nyblom, J. (1986). Testing for deterministic linear trend in time series. Journal of the American Statistical Association, 81(394), 545–549.
- Nyblom, J., and Harvey, A. (2000). Tests of common stochastic trends. *Econometric theory*, 16(02), 176–199.

- Nyblom, J., and Mäkeläinen, T. (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, 78(384), 856–864.
- Paluvs, M., Novotná, D., and Tichavský, P. (2005). Shifts of seasons at the european midlatitudes: Natural fluctuations correlated with the north atlantic oscillation. *Geophysical research letters*, 32(12).
- Parker, D. E., Legg, T., and Folland, C. K. (1992). A new daily central england temperature series, 1772–1991. International Journal of Climatology, 12(4), 317–342.
- Parmesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, 13(9), 1860–1872.
- Pezzulli, S., Stephenson, D., and Hannachi, A. (2005). The variability of seasonality. Journal of Climate, 18(1), 71–88.
- Poirier, D. J. (1976). The econometrics of structural change. North Holland, Amsterdam.
- Schwartz, M. D., and Reiter, B. E. (2000). Changes in north american spring. International Journal of Climatology, 20(8), 929–932.
- Stine, A., Huybers, P., and Fung, I. (2009). Changes in the phase of the annual cycle of surface temperature. Nature, 457(7228), 435–440.
- Tanner, M. A. (1996). Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions (3rd ed.). New York: Springer.
- Thackeray, S. J., Sparks, T. H., Frederiksen, M., Burthe, S., Bacon, P. J., Bell, J. R., ... others (2010). Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology*, 16(12), 3304–3313.
- Thompson, R., and Clark, R. (2008). Is spring starting earlier? The Holocene, 18(1), 95–104.
- Thomson, D. J. (1995). The seasons, global temperature, and precession. *Science*, 268(5207), 59–68.

# Research Papers 2013



Analysis of Time Series

- 2015-11: Arianna Agosto, Giuseppe Cavaliere, Dennis Kristensen and Anders Rahbek: Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)
- 2015-12: Tommaso Proietti, Martyna Marczak and Gianluigi Mazzi: EuroMInd-D: A Density Estimate of Monthly Gross Domestic Product for the Euro Area
- 2015-13: Michel van der Wel, Sait R. Ozturk and Dick van Dijk: Dynamic Factor Models for the Volatility Surface
- 2015-14: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting
- 2015-15: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Effects of Macroeconomic Uncertainty upon the Stock and Bond Markets
- 2015-16: Markku Lanne, Mika Meitz and Pentti Saikkonen: Identification and estimation of non-Gaussian structural vector autoregressions
- 2015-17: Nicholas M. Kiefer and C. Erik Larson: Counting Processes for Retail Default Modeling
- 2015-18: Peter Reinhard Hansen: A Martingale Decomposition of Discrete Markov Chains
- 2015-19: Peter Reinhard Hansen, Guillaume Horel, Asger Lunde and Ilya Archakov: A Markov Chain Estimator of Multivariate Volatility from High Frequency Data
- 2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States
- 2015-21: Ulrich Hounyo and Bezirgen Veliyev: Validity of Edgeworth expansions for realized volatility estimators
- 2015-22: Girum D. Abate and Niels Haldrup: Space-time modeling of electricity spot prices
- 2015-23: Eric Hillebrand, Søren Johansen and Torben Schmith: Data revisions and the statistical relation of global mean sea-level and temperature
- 2015-24: Tommaso Proietti and Alessandra Luati: Generalised partial autocorrelations and the mutual information between past and future
- 2015-25: Bent Jesper Christensen and Rasmus T. Varneskov: Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination
- 2015-26: Ulrich Hounyo and Rasmus T. Varneskov: A Local Stable Bootstrap for Power Variations of Pure-Jump Semimartingales and Activity Index Estimation
- 2015-27: Rasmus Søndergaard Pedersen and Anders Rahbek: Nonstationary ARCH and GARCH with t-distributed Innovations
- 2015-28: Tommaso Proietti and Eric Hillebrand: Seasonal Changes in Central England Temperatures