CYCLICAL TIME SERIES: AN EMPIRICAL ANALYSIS OF TEMPERATURES IN CENTRAL ENGLAND OVER THREE CENTURIES

By

Liudas Giraitis, Fulvia Marotta and Peter C. B. Phillips

September 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2409

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Cyclical time series: an empirical analysis of temperatures in central England over three centuries

Liudas Giraitis[1], Fulvia Marotta[2], and Peter C B Phillips[3]

[1]Queen Mary University of London
[2]De Nederlandsche Bank, University of Oxford
[3]Yale University, University of Auckland, Singapore Management University

September 17, 2024.

Abstract

This paper builds on methodology that corrects for irregular spacing between realizations of unevenly spaced time series and provides appropriately corrected estimates of autoregressive model parameters. Using these methods for dealing with missing data, we develop time series tools for forecasting and estimation of autoregressions with cyclically varying parameters in which periodicity is assumed. To illustrate the robustness and flexibility of the methodology, an application is conducted to model daily temperature data. The approach helps to uncover cyclical (daily as well as annual) patterns in the data without imposing restrictive assumptions. Using the Central England Temperature (CET) time series (1772 - present) we find with a high level of accuracy that temperature intra-year averages and persistence have increased in the later sample 1850-2020 compared to 1772 - 1850, especially for the winter months, whereas the estimated variance of the random shocks in the autoregression seems to have decreased over time.

JEL Classification: C01, C13, C22, C51, Q54
Keywords: Autoregression, cyclical time series models, daily temperatures, missing data

1 Introduction

Climate change is defined as a long-term change in the average weather patterns that have come to define Earth's local, regional and global climates. There is now general scientific acceptance that changes observed in Earth's climate since the beginning of the 20th century are primarily driven by human activity, especially from fossil fuel burning, methane emissions in agriculture, industrialization, and global commercial activity, all leading to a growing accumulation of atmospheric greenhouse gases. Carbon dioxide levels in the atmosphere have risen over 50% and methane over 150% since the start of the Industrial Revolution, raising heat retention capacity and in turn Earth's average surface temperature. Natural processes themselves, including the carbon cycle linking the atmosphere, oceans and continents, can be influenced by human activities and further contribute to changes in climate. Various external forces such as volcanic activity, fluctuations in energy output from the sun, and longer term orbital variations also effect changes in climate.

To monitor and study these phenomena climate scientists use terrestrial, atmospheric, oceanic, and satelite data in modeling to capture evidence of climate change indicators in land and ocean temperatures, rising sea levels, glacial and polar ice loss. Other data is also employed, including the frequency and severity of extreme weather such as hurricanes, heatwaves, wildfires, droughts, floods, and precipitation, as well as global changes in vegetation cover. This paper focuses on aspects of climate change that relate to the rising global mean surface temperature since the pre-industrial period (between 1850 and 1900)[1] and changes in temperature variability that are more regionally dependent. It is less clear whether these changes include the degree of persistence in surface temperature. Persistence is important because the lasting effects of weather on ecosystems and society can depend on the duration, intensity and regularity of extreme weather events.

In the scientific literature there is growing recognition that, for climate variables and events, statistical properties beyond the mean play significant roles in the various impacts of climate change on natural ecosystems (Di Cecco and Gouhier, 2018). Present evidence shows that climate system responses to rising atmospheric greenhouse gases involve distributional shape, not just shifts in mean temperature (Rivas and Gonzalo, 2020). Increases in the mean temperature alter the likelihood of temperature events by shifting the probability distribution towards higher values, whereas changes in temperature variance and skewness alter shape in ways that influence the incidence of temperature extremes. Both are essential in understanding the full effects of climate change. Present evidence shows warming of at least 1.5°C is inevitable within the next two decades (IPCC, 2021). Evidence supporting other distributional changes is less extensive, although there are good scientific reasons for natural consequences (Li and Thompson, 2021).

[1] Earth's average surface temperature has risen about 2°F since the late 19th century, a change largely driven by increased greenhouse gas emissions according to scientific consensus (Lenssen et al., 2019).

An example is provided by rapid arctic warming which raises persistence of weather patterns in the Northern Hemisphere, prolonging and repeating weather events such as droughts, heat waves, polar fronts, and cyclones (Francis and Vavrus, 2015).

As such, persistence can play a key role in understanding longer term climate impact. Temperature and in turn weather persistence is linked to sea surface temperatures, soil-moisture content and large-scale circulation patterns and factors that can alter past and future climate change (Pfleiderer and Coumou, 2018). Extended weather events accumulate negative shocks but also produce an overall impact that exceeds the sum of its parts. For instance, lengthy heatwaves magnify human mortality compared with intermittent individual hot days (Gasparrini and Armstrong, 2011), and multi-year droughts magnify the agricultural, economic, human and social impact compared with intermittent individual years of drought (Lenton et al. (2017)).

Few studies in the literature have explored the role persistence has played in global warming and how it may have changed over time. Existing work assessing temperature persistence suggests that inter-annual temporal variability has increased in some regions but decreased in others. Trends in the persistence of temperature variability, as measured by lag 1 autocorrelation, are generally less well studied in time series econometric modeling. Lenton et al. (2017) is an important exception, finding an increase in autocorrelation and variance in the PDO (Pacific Decadal Oscillation)[2] and AMO (Atlantic Multidecadal Oscillation.)[3] indexes of surface temperatures in the mid-latitude Pacific Ocean and North Atlantic Ocean. In other regions, such as Siberian Russia, the results are consistent with a decrease in both autocorrelation and variance. These findings indicate that trends in autocorrelation and variance are not globally related in a systematic way that is presently understood. Similarly, Graham et al. (2017) found evidence of a positive trend in the overall duration of winter warming events for both the North Pole region and the Pacific Central Arctic. The most rapid Arctic warming has been recorded during winter months.

Beyond these findings there are only scattered results in the literature concerning persistence, including how it can be characterized, e.g., in terms of lag 1 autocorrelations, or in Francis et al. (2018) as conditions that persist for at least four consecutive days, a threshold that was selected based on the observed distribution of duration lengths. Li and Thompson (2021), for instance, estimate surface temperature persistence using lag 10-day autocorrelations of daily mean temperatures. Changes in persistence under climate change conditions are estimated as the percent changes in persistence between two periods of 30 years: the 'historical' period (1970 - 1999) and 'future' period 2070-2099. This study found widespread increases in persistence across the middle and high altitudes of the Northern Hemisphere. Such changes might derive from a seemingly broad array of physical factors. In other work Pfleiderer and Coumou (2018) present a

[2]The Pacific Decadal Oscillation (PDO) is often described as a long-lived El Niño-like pattern that affects Pacific climate variability (Mantua et al., 1997).

[3]AMO is an index of variability in sea surface temperature (SST) of the North Atlantic Ocean measured on the timescale of several decades (McCarthy et al., 2015)

systematic analysis of temperature persistence for the northern hemisphere land area and define persistence as the length of consecutive warm or cold days. A warm (cold) day is defined as a day with a temperature above (below) the seasonal and grid point specific median temperature. The sample employed data from 1954 to 2014. This study showed that the probability of a hot or cold period persisting increases for longer periods, implying a potentially important role for memory in the physical system. The results reported above are mainly produced using simulation exercises with global climate models, for which there is considerable regional and inter-model variability concerning the role of extreme events and the definition of persistence. Some of the assumptions and implications involved in these simulations may not be reproducible in theoretical research or in an observational study. The present paper is concerned to provide such a study.

The goal of our empirical work is to develop a time series model with cyclically varying parameters suited to climate data that can shed light on several key issues on climate change. We seek to address three main questions: (1) Is it possible to estimate the yearly cycle of temperature persistence using a simple time series model and a standard definition of time dependency as persistence? (2) What is the historical path of temperature persistence under climate change conditions? (3) Can we statistically test whether there has been a significant change in the temperature persistence over time?

To answer these questions we fit the model described in Section 3 to a time series of daily Central England Temperatures from 1772 to the present. The model allows for cyclical time variation in all its parameters. Once estimated the fitted model allows us to extract estimates of parameter cycles that may be present over the course of a year. Daily changes in the intra-year behaviour of the intercept and autoregressive persistence in the temperature data are estimated. Overall, the findings indicate that a first order autoregression fits the data well, persistence in temperature is generally high and is higher during the winter and autumn months than in summer, and the cycle of persistence has significantly increased after the industrial revolution, especially during winter and autumn.

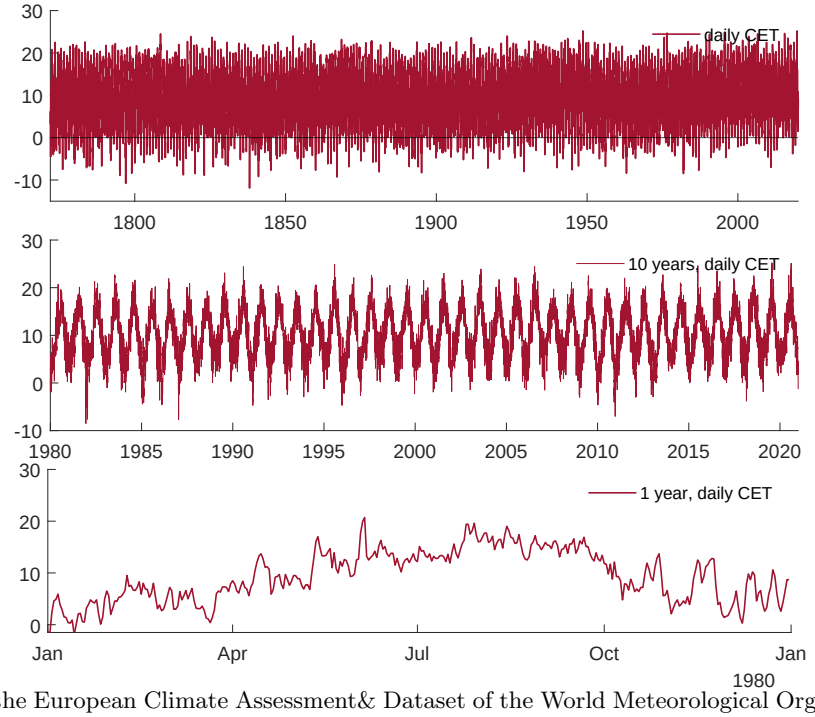
The paper is structured as follows: Section 2 describes the data employed and gives a selective survey of the literature. Section 3 outlines the methodology and proposes the time series model used in the empirical study. Section 4 reports estimation results and Section 5 explores implications of the main findings with a simulation study. Section 6 concludes. Additional computational details and graphics displaying the evolution of the parameter estimates based on different window widths are included in Appendix A.

2 Data

In the following analysis, we utilize the daily Central England Temperature (CET) time series to assess the applicability and effectiveness of our time series model in capturing cyclical variations,

particularly annual and monthly changes, in model parameters. The CET data, sourced from the European Climate Assessment & Dataset of the World Meteorological Organization, provides a comprehensive historical record for this analysis.

Figure 1: Daily Central England Temperatures.



Note: Data from the European Climate Assessment& Dataset of the World Meteorological Organization, covering the years 1772-2020, 1989-2020, and 01/01/1980-31/12/1980, with daily observations.

The Central England Temperature (CET) time series is renowned for its extensive duration, starting in 1772 and extending to the present day.[4] This long-term dataset, initially compiled by Manley (1974) for the period 1659-1973 and subsequently revised by Parker et al. (1992), enables an in-depth examination of climate patterns across both annual and seasonal scales.

Previous research has employed various methodologies to analyze this dataset. For instance, Harvey and Mills (2003) aggregated the data to an annual level, observing no significant warming trend between 1723 and 1999. In contrast, Rho and Vogelsang (2019) used an autoregressive model with a linear trend and trend-break, identifying a positive trend from October to April, though no such trend was observed during the summer months. More recently, Proietti and Hillebrand (2017) applied a structural time series model to decompose the CET data into permanent and transitory components, finding that the deterministic trend was most pronounced in the winter months, while the stochastic trend was more significant in the summer months and during April and May. Similarly, He et al. (2019) developed a Seasonal Shifting Mean Autoregressive model (SSM-AR) and found an increase in temperatures from July to March, with the most substantial warming occurring in the winter months.

[4]Data available at https://hadleyserver.metoffice.gov.uk/hadcet/data/download.html

Our approach brings two notable advances to the existing body of work. In terms of methodology, we introduce a model that is tailored for periodically correlated time series and enables robust inference concerning cyclical parameters. In terms of data analysis with this model, temperature persistence may directly be estimated to reveal how persistence has evolved over time, while not compromising the accuracy of estimating the cyclical means in daily temperatures. Application of this modeling framework to the CET data provides new insights into the cyclical patterns of temperature changes and their implications in understanding climate variability and trends.

3 Methodology

This section employs methodology for estimation and inference with a first order Periodic Autoregression (PAR(1)) where the observed data may be unevenly spaced, effectively leading to missing data. The PAR framework is constructed to model data X_i, where the index $i = 1, \ldots, N$ takes the form $i = jT + t$ to capture P cycles indexed by $j = 0, \ldots, P-1$ with the period T, giving $N = PT$ observations in total. Specifically, the model is defined as follows In this section, we consider estimation of a Periodic Autoregressive (PAR(1)) time series denoted by x_i, where $i = 1, \ldots, N$, with a period of T. The model is defined as follows:

$$x_{jT+t} = \alpha_t + \phi_t x_{jT+t-1} + \varepsilon_{jT+t}, \quad t = 1, \ldots, T, \quad j = 1, \ldots, P. \tag{1}$$

Here, P represents the number of cycles, $N = PT$ is the total number of observations, and α_t and ϕ_t, $t = 1, ..., T$ are time-varying parameters that change periodically with period T. The data $\{x_1, ..., x_N\}$ is divided into P cycles

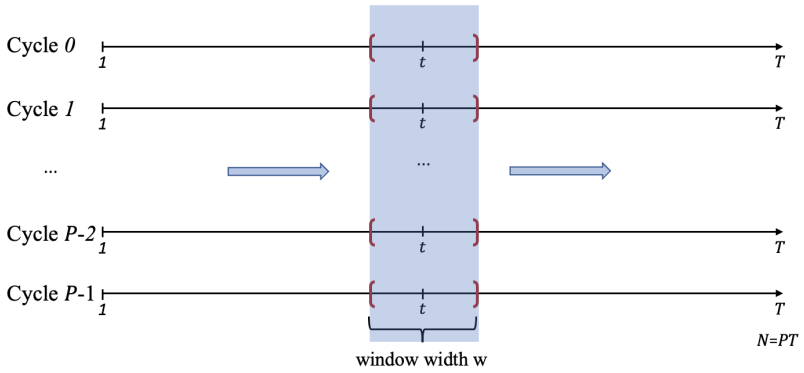
$$\{x_{jT+1}, ..., x_{jT+T}\}, \quad j = 0, ..., P-1,$$

organized as displayed in Figure 2. To estimate the parameters α_t, ϕ_t of the model at time points $t = 1, ..., T$, we use a rolling window approach introduced in Giraitis et al. (2024) for estimation of cyclical time-varying parameters in a PAR(p) model. Estimation of α_t, ϕ_t at time t is based on the block of data x_i, $i \in J_{t,w}$ where the block

$$J_{t,w} = \{i = jT + s : \, j = 0, ..., P-1, \, s = 1, ..., T, \, |s - t| \leq w/2\},$$

is centred at time t and has a fixed width w. To allow for unevenly spaced observations in each block $J_{t,w}$, $t = 1, ..., T$, we include in the observed data specification a missing data indicator

Figure 2: Rolling window estimation.



h_i, $i = 1, ..., N$ defined as follows

$$h_i \equiv h_{i,tw} = \begin{cases} 1 & \text{for } i \in J_{t,w}, \\ 0 & \text{for } i \notin J_{t,w}, \end{cases} \tag{2}$$

giving a partially observed time series

$$y_i = h_i x_i, \quad i = 1, ..., N. \tag{3}$$

To employ an asymptotic theory of estimation and inference in our empirical work we assume that x_i within the block $J_{t,w}$ follows a stationary AR(1) model,

$$x_i = \alpha_t + \phi_t x_{i-1} + \varepsilon_i \quad \text{for } i \in J_{t,w} \tag{4}$$

with parameters α_t, ϕ_t. The parameters α_t, ϕ_t are then estimated using the following formulae

$$\begin{aligned} \widehat{\phi}_t &= \frac{\sum_{j=2}^{N} h_j h_{j-1}(y_j - \bar{y}_t)(y_{j-1} - \bar{y}_t)}{\sum_{j=2}^{N}(y_{j-1} - \bar{y}_t)^2 h_j h_{j-1}}, \qquad \bar{y}_t = (\sum_{j=1}^{N} h_j)^{-1} \sum_{j=1}^{N} y_j, \\ \widehat{\alpha}_t &= N_2^{-1} \sum_{j=2}^{N} h_j h_{j-1}(y_j - \widehat{\phi}_t y_{j-1}), \quad N_2 = \sum_{j=2}^{N} h_j h_{j-1}. \end{aligned} \tag{5}$$

As the number of observations in the block grows, specifically as the number of operational observations for AR(1) regression $N_2 \to \infty$, the following asymptotic results established in

Giraitis et al. (2024) hold

$$\sqrt{\frac{N_2}{1-\phi^2}}\,(\hat{\phi}_t-\phi_t)\to\mathcal{N}(0,1),\qquad \sqrt{\frac{N_2}{1-\phi^2}}\,\frac{(\hat{\alpha}_t-\alpha_t)}{\sqrt{\mathbb{E}x_t^2}}\to\mathcal{N}(0,1), \tag{6}$$

$$(\sum_{j=1}^{N}h_j)^{-1}\sum_{j=1}^{N}y_j^2=\mathbb{E}x_t^2+o_p(1). \tag{7}$$

These results allow us to construct precise confidence intervals for the estimates of α_t, ϕ_t, $t=1,....,T$. The missing data indicators $\{h_i=h_{i,tw},\ \ i=1,...,N\}$ in (5) and (6) depend on $\{t,w\}$ and need to be recomputed for each t. Further details and proofs of (6) and (7) are provided in Giraitis et al. (2024).

To apply this methodology to the Central England Temperature (CET) dataset, which spans 248 years of daily temperature observations from 1772 to 2020, we first define the cycles of interest. Annual cycles are the focus here, with each cycle consisting of 365 daily observations (excluding leap years). So the cycle length is $T=365$ and the total number of cycles is $P=248$, giving a total of $N=90{,}520$ possible observations.

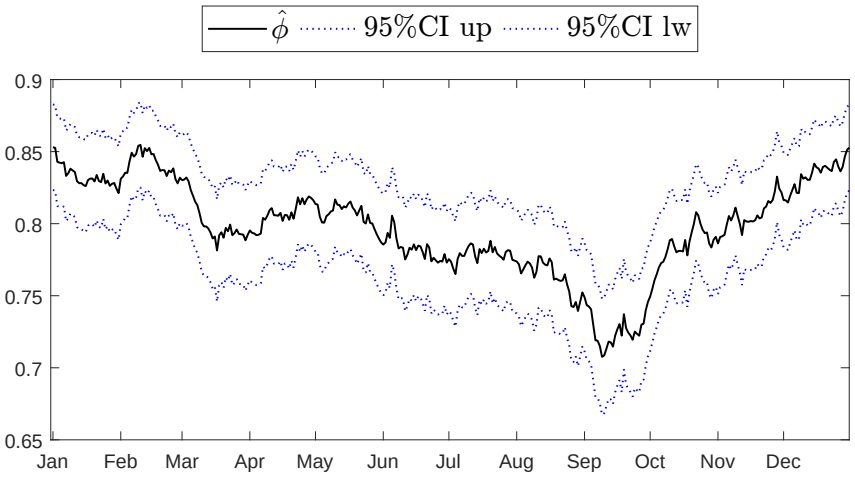
In our rolling window analysis, we consider neighboring years (e.g., 20 or 40 years) as replications of the same cycle. This allows us to fit the model to data blocks where the assumption of local stationarity is reasonable. By grouping consecutive years where temperature parameters are assumed to be stable, we estimate the annual variations of the parameters by applying the model to these blocks and shifting the estimation window across all time points $t=1,\ldots,365$.

While our primary focus is on the annual variation of the mean temperature (with higher temperatures expected during summer months and lower temperatures during winter), we must account for the long-term trend in temperature increases over the years. The global temperature has been rising steadily, which affects the mean temperature across different years (IPCC, 2021). This rising trend means that the assumption that each year is a perfect replication of the same cycle does not hold uniformly across the entire dataset. As temperatures have increased, the trend and cycle characteristics of the data may have shifted, especially in more recent years.

To address this issue, we divide the data into smaller groups of years where we assume the change in temperature is minimal within each group. This approach helps mitigate the impact of the long-term trend on our estimates. The exact behavior of the temperature changes over time is not fully known, but we understand that the rate of change has accelerated in recent years. Consequently, for more recent data, we use smaller groups to ensure that the assumption of minimal change within each group remains valid.

For each chosen group of years P, where each cycle contains $T=365$ daily observations, we apply the AR(1) model as specified in equation (3) to blocks J of width $w=7$, 15, or 30 days. These window widths correspond to smoothing over periods of one week, two weeks, and one month, respectively. Throughout this process, we do not impose a specific model for the type of

Figure 3: Estimation results for AR(1) parameter ϕ_t.



Note: Years 1980-2020, w=30; 95% confidence interval for parameter ϕ_t.

change involved in parameter evolution. Instead, our goal is to analyze the intra-year variations in the parameters α_t, ϕ_t, the mean $\mu_t = \frac{\alpha_t}{1-\phi_t}$, and the variance of the residuals $\sigma^2_{\varepsilon,t}$.

The model is fitted for various window widths ($w = 7, 15, 30$) to capture changes over one week, two weeks, and one month. Our primary focus remains on exploring the intra-year behavior of the parameter sequences α_t, ϕ_t, $\mu_t = \frac{\alpha_t}{1-\phi_t}$, and the innovation variances $\sigma^2_{\varepsilon,t}$.

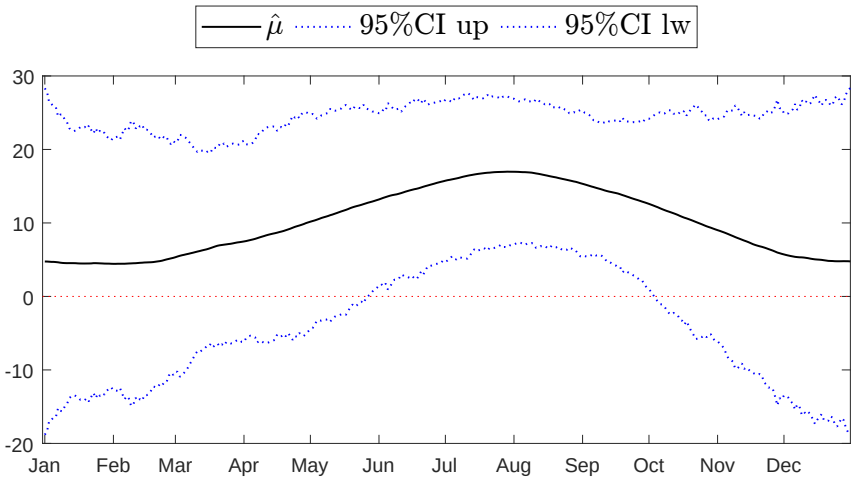
4 Empirical Results

By leveraging data from multiple neighboring years as repetitions of the same annual cycle, the model outlined in Section 3 enables us to estimate cyclical patterns in the model parameters with a high degree of precision. Specifically, the procedure yields not only parameter estimates but also corresponding confidence intervals for the following key quantities: (1) the first-order autoregressive coefficient $\hat{\phi}_t$, which captures the persistence of temperature fluctuations; (2) the intra-year mean $\hat{\mu}_t$; (3) the intercept $\hat{\alpha}_t$; and (4) the variance of the residuals $\hat{\sigma}^2_{\varepsilon_t}$. Findings based on the most recent 40-year period from January 1980 to December 2020 are presented in Figures 3 - 6. For this analysis, the total number of cycles, P, is set to 40, and parameter estimation is performed using a window width of 30 days, resulting in $40 \times 30 = 1200$ observations overall.[5]

Closer examination of the estimated ϕ_t values, depicted in Figure 3, reveals several insights. First, it is evident that the autoregressive coefficient ϕ is consistently estimated with high precision throughout the year, with all values being significantly different from zero and generally high, ranging from 0.6 to just over 0.87. This indicates a strong degree of persistence in daily temperature fluctuations. Moreover, the persistence parameter exhibits considerable seasonal

[5]Results for different window widths w are provided in Appendix A.3.

Figure 4: Estimation results for the local mean μ_t.



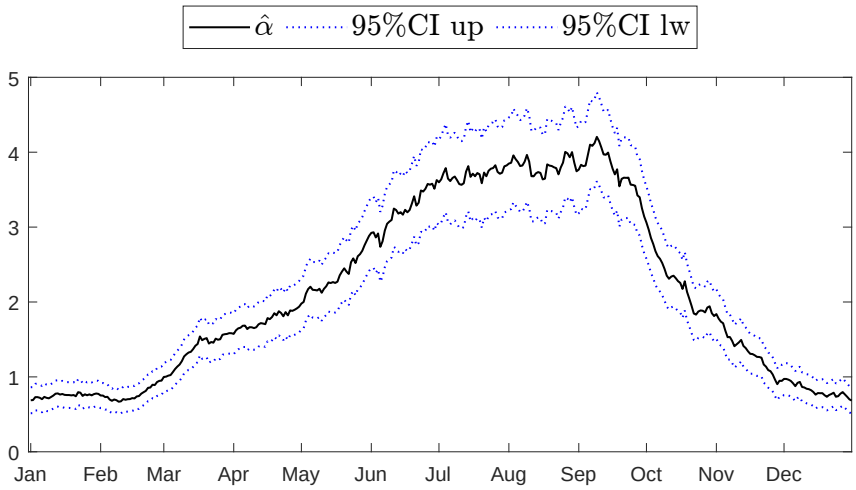
Note: Years 1980-2020, w=30; 95% confidence interval for parameter μ_t.

variation. The highest persistence is observed during the winter months, with peaks in December (0.86) and February (0.84). In contrast, persistence declines steadily during the summer, reaching its lowest point in early September (0.66), before increasing again with the onset of autumn.

Overall, these results provide strong evidence that temperatures during the colder months exhibit higher persistence. Given that persistence is consistently high at every point t in the cycle, $t = 1, \ldots, 365$, this has significant implications for the model's capacity to accurately estimate the 95% confidence intervals around the annual mean temperature. The results for the mean, as shown in Figure 4, follow the expected seasonal pattern: the average temperature is lower during the winter months, gradually rising and reaching its peak in the summer, specifically around the beginning of August. However, the confidence bands surrounding the estimated mean are quite large. This wide interval is directly related to the high level of persistence in the data. Specifically, as ϕ_t increases, the confidence bands for the mean widen, making it challenging to draw precise conclusions about the behavior and variation of the mean temperature across different times of the year.

Figure 5 presents the estimated annual pattern of the intercept parameter α_t. The model is capable of producing relatively precise confidence intervals for α_t. The parameter estimates are statistically significant, consistently differing from zero across all points in the cycle. The estimation suggests that α_t increases during the spring months, though with some degree of volatility, before peaking around mid-September. After this peak, there is a sharp decline in α_t as the year progresses into the autumn months. The observed pattern in α_t corresponds with seasonal temperature variations, where intercept values rise in response to increasing temperatures in spring and summer and then fall as temperatures decrease in autumn.

Figure 5: Estimation results for the intercept α_t.



Note: Years 1980-2020, w=30; 95% confidence interval for parameter α_t.

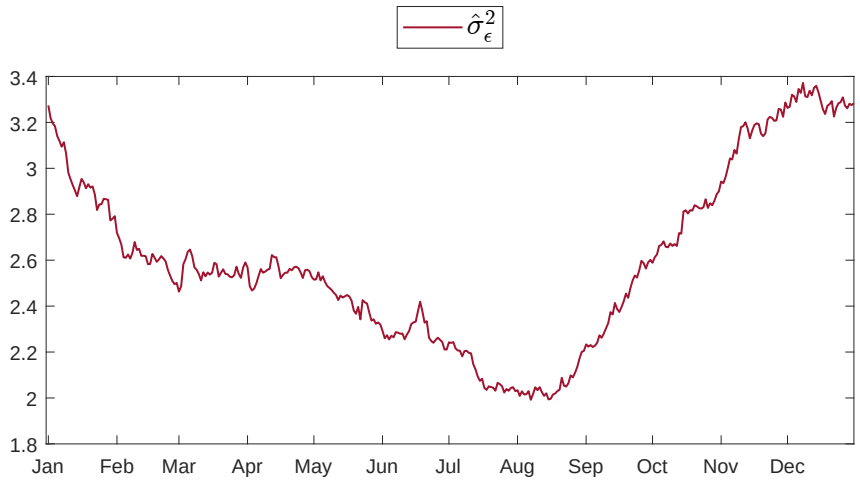
Once the model is fitted to the data, we can compute the residuals[6] $\hat{\varepsilon}_t$ for each day t of the year and subsequently estimate the variance of these residuals, $\sigma^2_{\varepsilon,t}$. This allows us to examine the changes in the volatility of exogenous shocks to temperature across the annual cycle. The results, depicted in Figure 6, align with findings in the climate literature, confirming that the variance of the residuals also exhibits cyclical variation throughout the year. Specifically, we observe that the estimated variance $\hat{\sigma}^2_{\varepsilon,t}$ is particularly high during the winter months, peaking in December with values just below 3.5. In contrast, the variance decreases during the summer months, reaching 2.

To evaluate the adequacy of the fitted AR(1) model, defined by equations (1 - 3), we conducted several diagnostic checks on the residuals $\hat{\varepsilon}_t$. The first two panels of Figure 7 offer a visual examination of the residuals: the first panel displays the histogram of the residuals, while the second panel presents a Q–Q (quantile-quantile) plot. Together, these plots suggest that the residuals approximate a normal distribution. The Q–Q plot specifically compares the quantiles of the residuals with those of a standard normal distribution, revealing that the residuals closely align with normality. This visual evidence, combined with the findings from Figure 6, which highlights the local changes in residual variance across seasons, indicates that the residuals are normally distributed but exhibit changing variance.

Interestingly, this result contradicts the intuitive expectation that global warming might induce greater randomness into the model's exogenous component, leading to heavy-tailed residuals. However, the diagnostic checks suggest that the distribution of the exogenous shocks does not have heavy tails. This observation provides partial evidence that the observed changes in

[6]For details on the computation of residuals, refer to Section A.1. In the following discussion, we present results for the global residuals scaled by their variance.

Figure 6: Estimation results for the variance $\sigma^2_{\varepsilon,t}$.



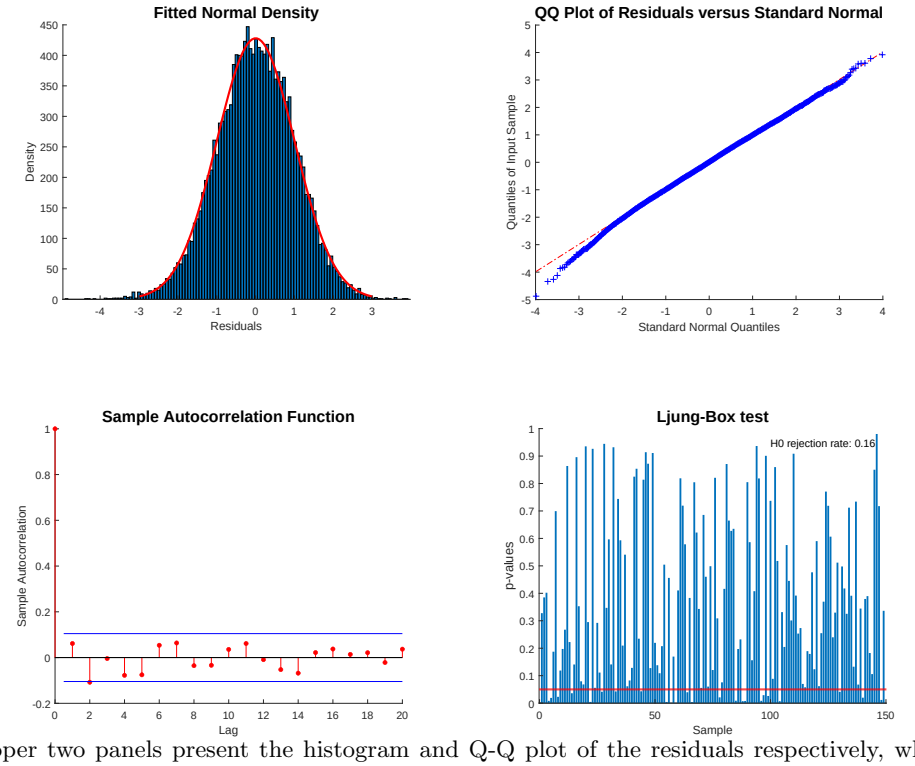
Note: Years 1980-2020, w=30.

temperature, including extreme events like heatwaves or cold spells, are primarily driven by changes in the model's parameters, such as persistence, rather than by a rise in randomness.

The lower two panels of Figure 7 further investigate the presence of autocorrelation in the residuals at different lags. The third panel displays the correlogram of the residuals, focusing on the year 2000 as a representative case, although similar results are observed across the entire 40-year sample. The correlogram shows that the residuals exhibit no significant autocorrelation at any lag, suggesting a good fit of the model.

To systematically test for autocorrelation across all residuals and lags, we applied the Ljung-Box Q-test on consecutive groups of residuals, each containing 100 observations. The results, shown in the fourth panel of Figure 7, display the *p-values* of the test across the 150 samples, alongside the 5% significance level. The null hypothesis—that the residuals are uncorrelated from lag 1 to lag 20—is rejected only when the *p-value* is below 0.05. In this case, the test rejects the null hypothesis in only 16% of the samples, confirming that the residuals are mostly uncorrelated at these lags. Overall, these findings indicate that the AR(1) model provides a satisfactory fit to the data.

Figure 7: Testing for normality and autocorrelation in the residuals $\hat{\varepsilon}_t$



Note: The upper two panels present the histogram and Q-Q plot of the residuals respectively, while the lower two panels present the sample autocorrelation function of the residuals together with the Lijung-Box test for the entire sample. Years 1980-2020, w=30.

4.1 The change in the cycle over the years

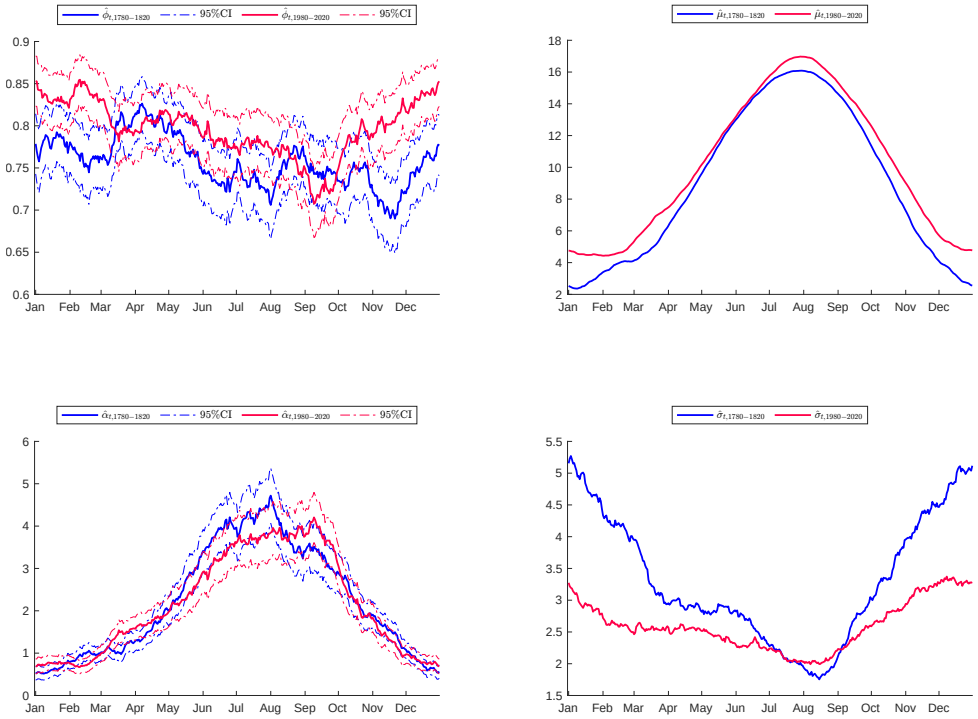
The previous section showed the methodology is effective in estimating cyclically varying parameter, giving well-resolved parameter estimates with normally distributed, uncorrelated residuals. Building on this approach we turn to examine how the parameters have evolved over time, seeking to answer two key questions: (i) how have the model parameters, which vary over the year, changed over the past two centuries? and (ii) have there been significant changes in these parameters across different seasons?

To address these questions we utilize our complete dataset to estimate parameter changes over the past 200 years. By comparing parameter estimates from different time periods we can assess whether significant changes have occurred in the model parameters, such as the mean or persistence. These comparisons help to determine whether the underlying temperature model has shifted over time.

Figure 8 displays the main results of this analysis[7]. The model parameters were estimated and compared for two distinct periods: the first 40 years (1780-1820) and the most recent 40 years (1980-2020). This comparison provides the means to evaluate whether there have been significant changes in temperature parameters from the 18th century to recent times.

[7]Detailed results for different window widths, $w = 7, 15$, are provided in Appendix A.3

Figure 8: Comparing past and present of parameters ϕ_t, α_t, μ_t, $\sigma^2_{\varepsilon,t}$.



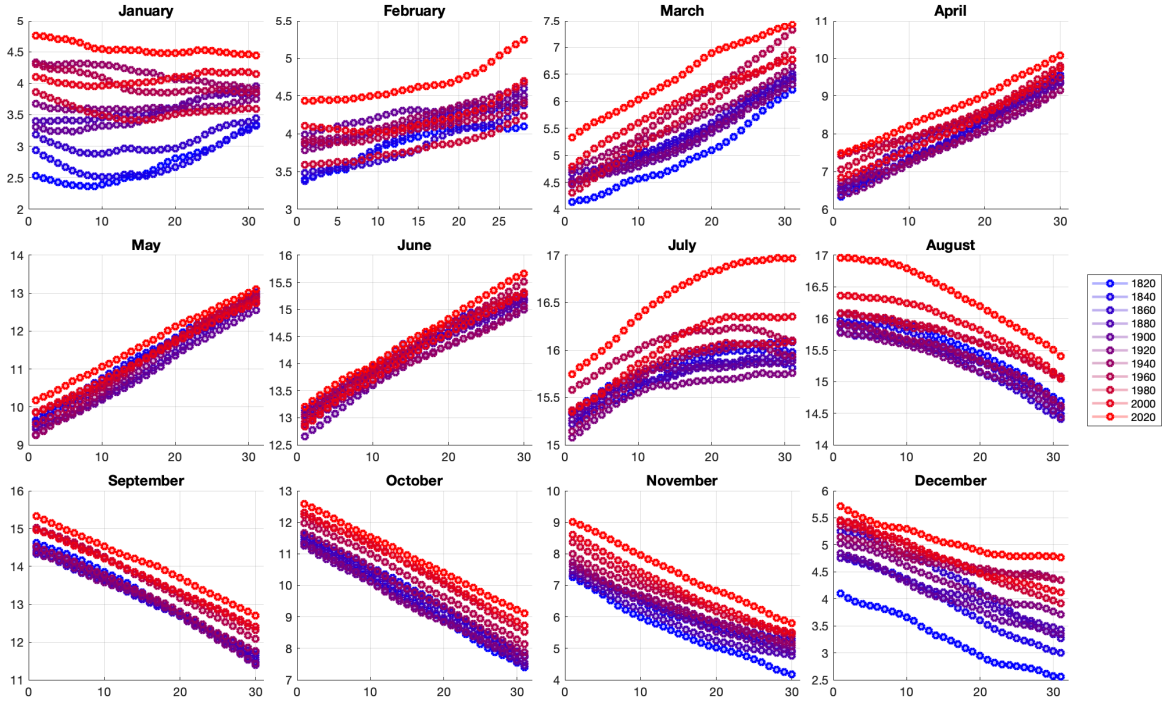
Note: Years 1780-2020, group of 40 years, w=30.

The analysis reveals several key findings. First, there is substantial evidence of a significant change in the persistence coefficient ϕ_t. The data indicates that the shape of the cycle for this parameter has evolved over time, particularly in colder months such as January, February, November, and December, where the estimates from the two periods are significantly different at the 95% confidence level. In contrast, persistence in summer months, such as June and July, shows some variation, though it is not statistically significant. Spring and autumn months exhibit stability in persistence, with overlapping confidence bands suggesting minimal change.

Second, the intercept parameter α_t shows a similar seasonal pattern. While changes in α_t are influenced by both persistence and average temperature, the results suggest an increase in the average temperature, especially in winter months. However, the large confidence bands and limited data availability make it challenging to draw definitive conclusions about statistical significance. Finally, the sample variances of the residuals $\hat{\sigma}^2_{\varepsilon,t}$, as depicted in the lower-right panel of Figure 8, has decreased significantly in winter months from the early period (1780-1820) to the recent period (1980-2020). In contrast, the variance in summer months has remained relatively stable.

In summary, our findings suggest a complex evolution in temperature parameters over time. During spring and early autumn, the temperature model appears to have remained relatively stable. However, significant changes are evident in late fall and winter, where both average temperatures and persistence have increased. This increased persistence implies that recent

Figure 9: Progressive change of the mean μ_t over time: from 1772 to 2020.



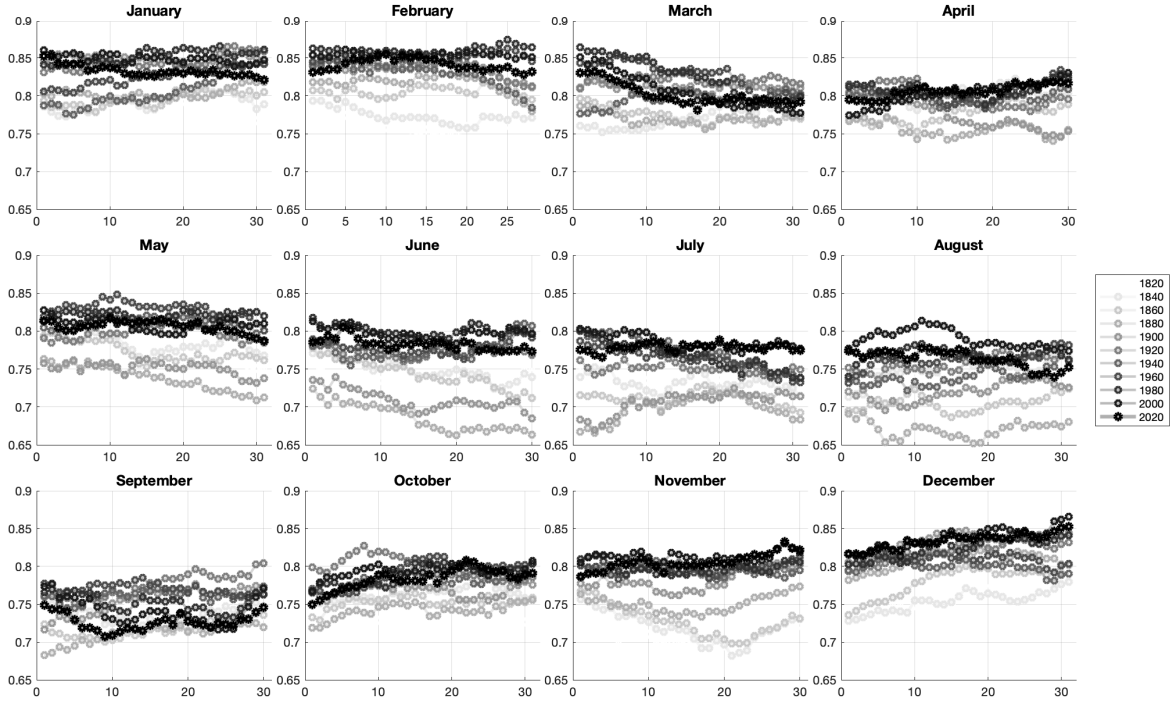
Note: Years 1772-2020, $w = 30$. The lines show estimation of μ_t for consecutive groups of 40 years: 1780-1920,. . . ,1980-2020.

temperatures with significant deviations from the mean are slower to revert to the average compared to those from two centuries ago. Additionally, the decrease in residual variation indicates reduced overall volatility in temperatures. These changes suggest a trend towards higher, less volatile, and more persistent temperatures, particularly evident in winter. Consequently, there is an increased likelihood of observing prolonged temperature extremes such as heatwaves or cold spells in recent years compared to the past.

Building on the same approach we now explore progressive changes in the model parameters using the entire dataset. The model is now estimated using consecutive 40-year intervals (viz., 1780-1820, 1820-1860, ..., 1980-2020) and parameter estimates are plotted, documenting parameter evolution over the past two centuries. Figures 9 and 10 illustrate the results for the mean μ_t and the persistence parameter ϕ_t. Each panel in these figures represents a different month, with colors indicating estimates from the consecutive time periods: from blue (representing the earliest period 1780-1800) through to red (representing the most recent years 2000-2020).

Analysis of transitions in the mean parameter μ_t reveals several key trends: (1) For the months of January, February, November, and December, the change in the mean temperature over time is more pronounced compared to other months. (2) January exhibits a gradual increase in mean temperature, while for summer months like July and August, the average temperatures

Figure 10: Progressive change of AR(1) parameter ϕ_t over time: from 1772 to 2020.



Note: Years 1780-2020, $w = 30$. The lines show estimation of ϕ_t for consecutive groups of 40 years: 1780-1920,..., 1980-2020.

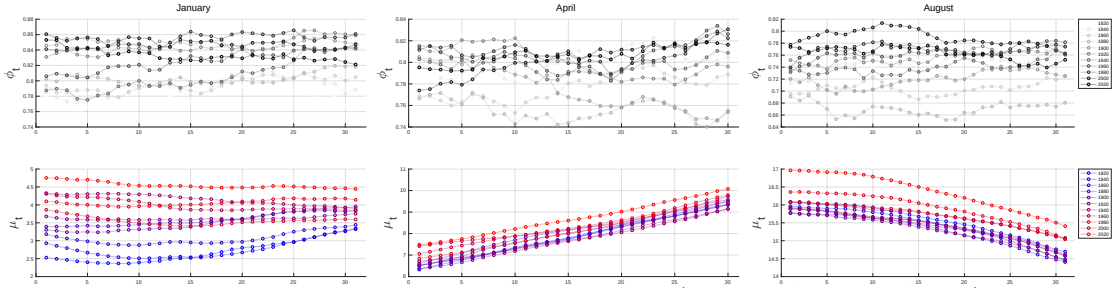
have risen notably in recent years. (3) For spring months such as April, May, and June, the data suggests that there has been little to no change in the intra-year cycle of average temperatures.

Turning to the persistence parameter ϕ_t, Figure 10 shows that significant changes are observed in January, February, November, and December. For these months, persistence appears to have changed early on, as indicated by the shift from light shaded grey lines (representing earlier periods) towards the black lines (representing the most recent years). In contrast, while there is evidence of increased persistence in summer months, this change does not follow a linear trajectory. Some of the dark shaded lines fall below the light shaded lines, suggesting that persistence initially decreased before rising above earlier levels. Figure 11 zooms into the daily variation of parameters for January, April and August.

Overall, these figures provide a detailed view of how both mean temperature and persistence parameters have evolved across different seasons and decades, highlighting significant changes in colder months and varying trends in warmer months.

In previous analyses a window width of $w = 30$ was used to estimate the parameters. To illustrate how the choice of window width affects results, we examined two specific months (February and May) where cycles in the parameters ϕ_t and μ_t exhibit distinct changes. Figure 12 provides a detailed comparison of these changes, showing results with different sample sizes

Figure 11: Zooming in: progressive change for parameters ϕ_t and μ_t for selected months.



Note: Years 1772-2020. The figure reports for each selected month (January, April and August) estimation of ϕ_t and μ_t for window width w=30. Estimates are calculated over consecutive groups of 40 years.

and window widths. Group sizes of 20 years are shown in panels (a) and (b) and 40 years in panel (c). Within each horizontal panel the window width is varied so that $w = 7$ in panel (a), $w = 15$ in panel (b), and $w = 30$ in panel (c). The lines in the figures are color-coded to represent different time periods, with cooler colors indicating earlier samples (e.g., dark blue starting in 1772) and warmer colors representing more recent samples (e.g., dark red ending in 2020).

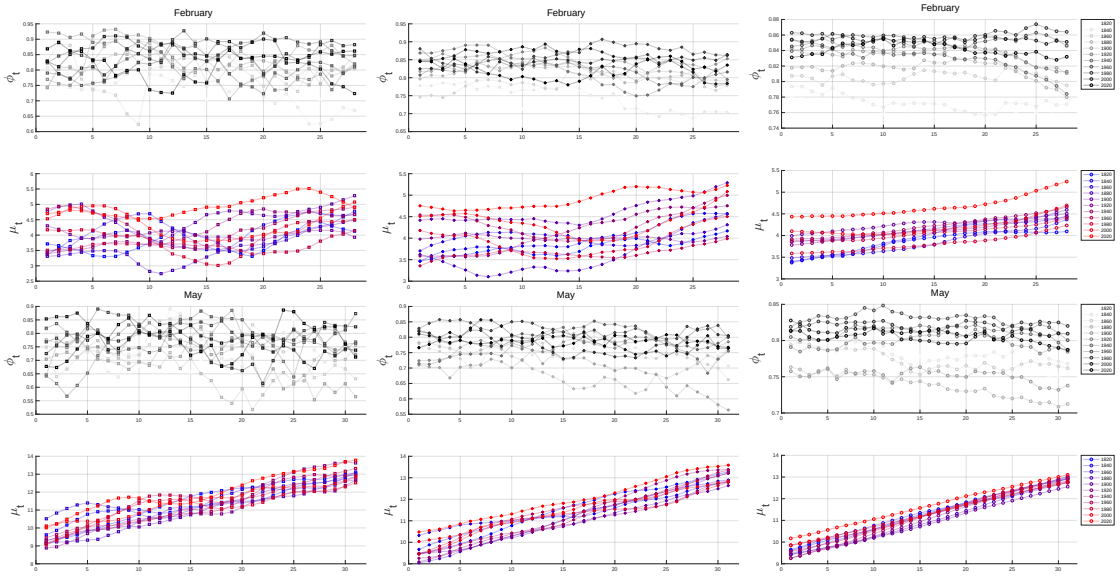
The figure illustrates how the choice of window width and sample size affects the estimates. Smaller windows ($w = 7$ and $w = 15$) lead to more variability in the estimates, resulting in curves that are less smooth. This increased variability arises because smaller windows use fewer data points, making the estimation results more stochastic. The larger windows instead ($w = 30$) provide smoother estimates due to the increased sample size. Smoothness comes at the cost of reduced precision, as the estimates become less sensitive to short-term changes. In sum, selecting an appropriate window width depends on a trade-off between accuracy and smoothness in estimation. The choice should be guided by measures of goodness of fit and the specific temporal resolution of interest. whether the focus is on monthly, weekly, or daily changes.

5 Simulating daily temperatures with the fitted cyclical model

Using the empirical findings and diagnostics a simulation study was designed and conducted to further explore certain aspects of temperature dynamics. The study aimed to: (1) analyze changes in the distributions of maximum and minimum temperatures; (2) project these changes into the future; and (3) assess the prevalence of cold spells and heat waves. To achieve these objectives, we employed the fitted parameters from two distinct historical periods: 1780-1820 and 1980-2020, as shown in Figure 8. By simulating data from models parameterized for these periods we investigated how temperature distributions might have evolved and project future trends.

1,000 replications of daily temperature cycles (365 days) were obtained using the following

Figure 12: February and May: progressive change for parameters ϕ_t and μ_t over time for different window widths w.



(a) (b) (c)

Note: Years 1772-2020. The figure reports for each month (February and May) estimation of ϕ_t and μ_t for window widths: $w = 7$ in panel (a), $w = 15$ in panel (b) and w=30 in panel (c). Estimates in (a) and (b) are calculated over consecutive groups of 20 years, while for (c) over groups of 40 years

models for the two historical periods:

1. $X_{jT+t} = \hat{\alpha}(t)_{1780-1820} + \sum_{k=1}^{p} \hat{\phi}(t)_{1780-1820} X_{jT+t-k} + \varepsilon_{jT+t}, \quad j = 0, \ldots, N,$
2. $X_{jT+t} = \hat{\alpha}(t)_{1980-2020} + \sum_{k=1}^{p} \hat{\phi}(t)_{1980-2020} X_{jT+t-k} + \varepsilon^{*}_{jT+t}, \quad j = 0, \ldots, N,$

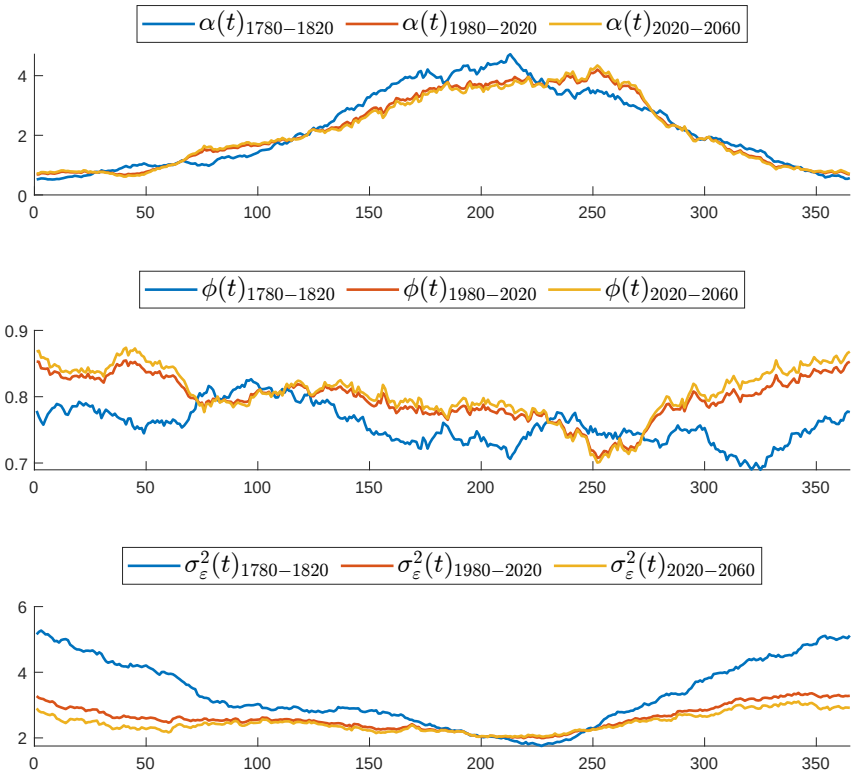
where $\hat{\alpha}(t)$ and $\hat{\phi}(t)$ are the parameter estimates from Figure 8, and $N = 1000$ represents the number of cycle replications. The Gaussian errors ε_{jT+t} and ε^{*}_{jT+t} have variances matching those estimated for the respective periods. For each replication, we extract information such as average temperature and distributions of maximum and minimum temperatures by month. Performing this exercise across all simulations, we estimated empirical distributions of these temperature metrics.

Further, to create possible future scenarios we projected the estimated parameter trends forwards to the next 40 years. This involved extending the parameters linearly from the past trends, resulting in:

3. $X_{jT+t} = \hat{\alpha}(t)_{2020-2060} + \sum_{k=1}^{p} \hat{\phi}(t)_{2020-2060} X_{jT+t-k} + \varepsilon_{jT+t}, \quad j = 0, \ldots, N,$

where the projected parameters $\hat{\alpha}(t)_{2020-2060}$ and $\hat{\phi}(t)_{2020-2060}$ are computed by extrapolating the differences between the 1780-1820 and 1980-2020 periods.

Figure 13: Projected change in the parameters α_t, ϕ_t, $\sigma^2_{\varepsilon,t}$ for Model 1 - 3.



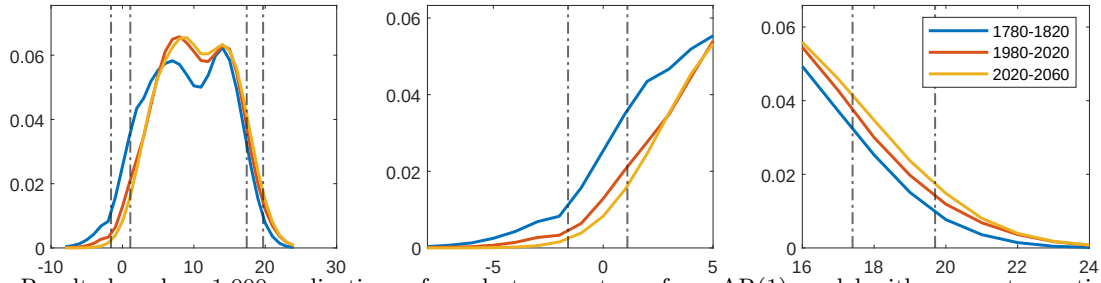
w=30.

$$\begin{aligned}\hat{\alpha}(t)_{2020-2060} &= \hat{\alpha}(t)_{1980-2020} + 0.2\left(\hat{\alpha}(t)_{1980-2020} - \hat{\alpha}(t)_{1780-1820}\right), \\ \hat{\phi}(t)_{2020-2060} &= \hat{\phi}(t)_{1980-2020} + 0.2\left(\hat{\phi}(t)_{1980-2020} - \hat{\phi}(t)_{1780-1820}\right).\end{aligned} \tag{8}$$

Figure 13 shows the projected parameter cycles for the future model, represented by the yellow lines. For simplicity, we use the error variance estimated for the 1980-2020 period in the future simulations. This approach provides one mechanism for exploring potential future changes. Actual developments in the future may, of course, differ considerably amd exhibit characteristics such as non-linear trends or external temperature drivers not considered in this exercise. Here we focus simply on showing how a model of this type, based on observational data, can be utilized for simulating projections, rather than directly forecasting future parameter changes or climate evolution.

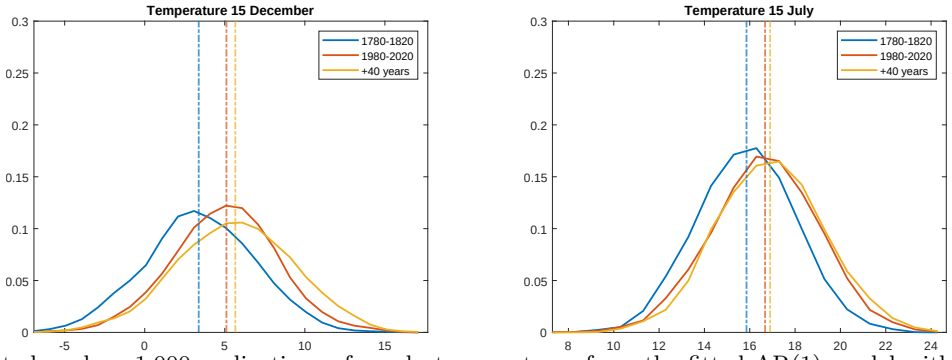
The distributions of observed temperatures for the periods 1770-1820 and 1980-2022, along with the simulated temperatures for 2040-2060, are depicted in Figure 14. The empirical distributions are estimated by fitting kernel probability density functions to the data, using the Epanechnikov kernel smoother for nonparametric density estimation. The first panel of the figure illustrates the seasonal temperature density with two prominent peaks. The second and

Figure 14: Probability densities of real temperatures for years 1770-1820, 1980-2022 and simulated temperatures for 2040-2060. A comparison of the upper (1%, 5%) and lower (95%, 99%) quantiles.



Note: Results based on 1,000 replications of yearly temperatures from AR(1) model with parameters estimated in years 1780-1820 (in blue), 1980-2020 (in red) and projected change in 40 years (in yellow).

Figure 15: Simulated average temperatures for a single day.



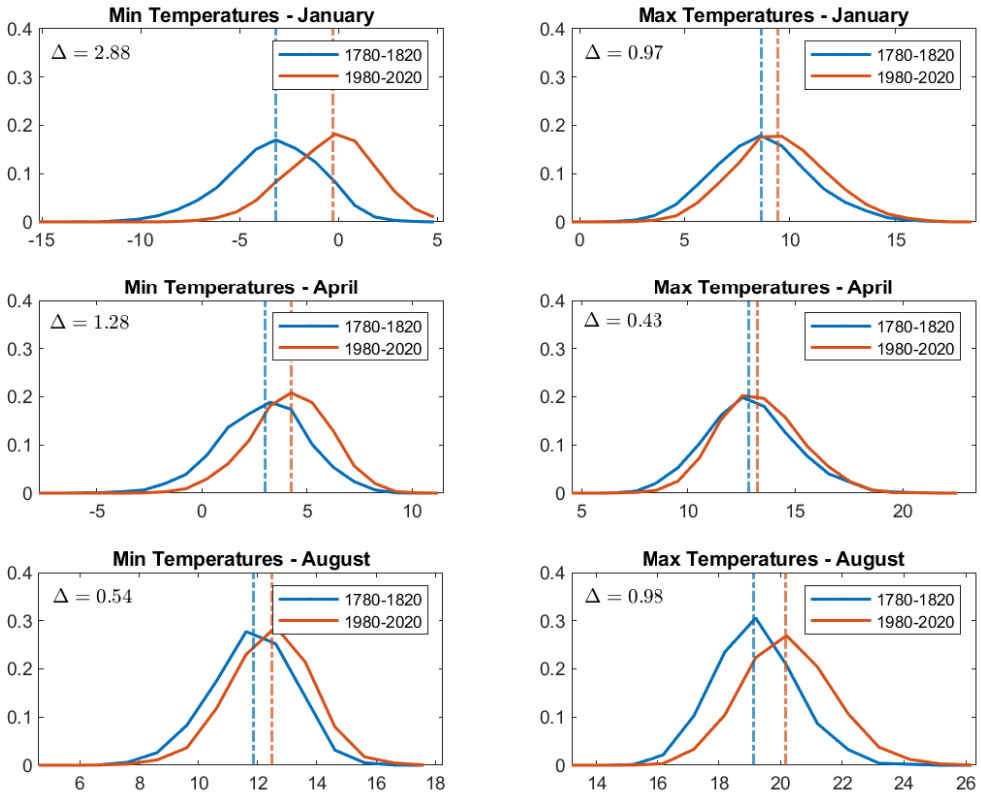
Note: Results based on 1,000 replications of yearly temperatures from the fitted AR(1) model with parameters estimated in years 1780-1820 (in blue), 1980-2020 (in red) and projected change in 40 years (in yellow).

third panels provide a closer look at the quantiles: the lower quantiles (1% and 5%) in the second panel and the upper quantiles (95% and 99%) in the third panel. In panel two, we observe that the density associated with lower temperatures decreases when comparing recent and future distributions to those of the past. Conversely, panel three shows an increase in the probability of higher temperatures, with the red and yellow lines indicating a shift towards warmer temperatures. Overall, the distribution has shifted to the right, indicating a higher likelihood of warmer temperatures.

Figure 15 examines the temperature distributions for two specific days: July 15th (a warm day) and December 15th (a cold day). This figure confirms the previous results, showing that the distribution of daily average temperatures has shifted to the right for both days, with a more pronounced increase for the cold day in December.

In a further analysis we focus on the distributions of maximum and minimum temperatures, which are critical for understanding extreme weather events such as heatwaves and cold spells. Figure 16 displays the simulated distributions of minimum and maximum temperatures, using

Figure 16: Distributions of minimum and maximum annual temperatures: past and present.



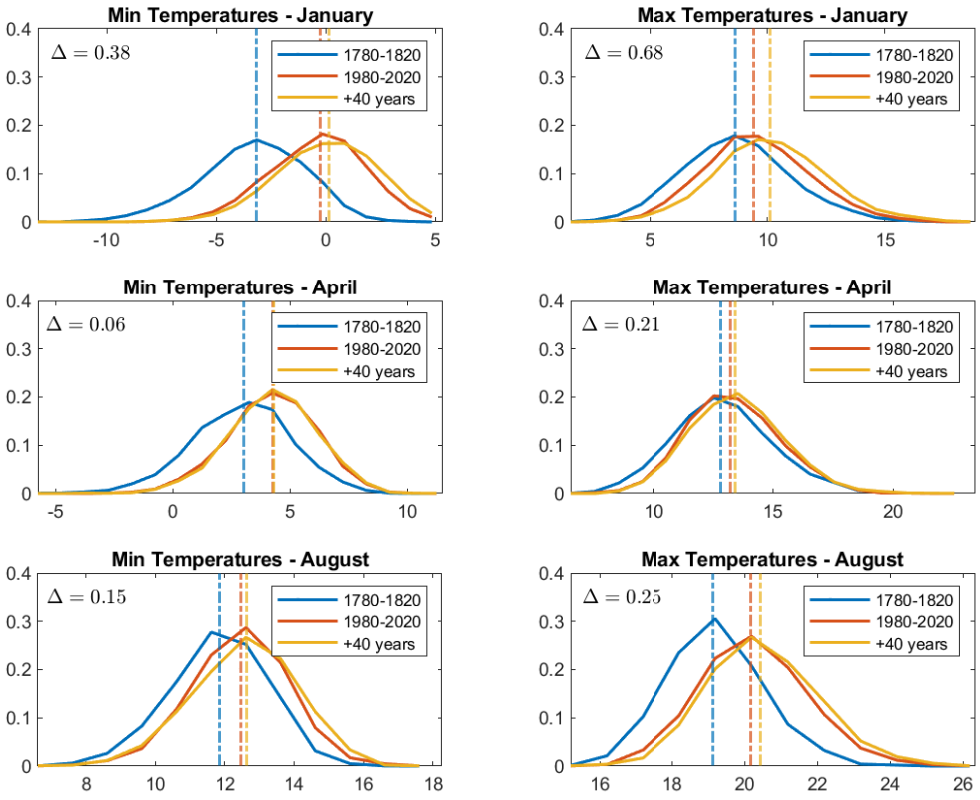
Note: Based on 1,000 replications of yearly temperatures from AR(1) model with parameters estimated in years 1780-1820 (in blue), 1980-2020 (in red).

1,000 replications based on Model 1 and Model 2. Figure 17 compares simulations from Models 1, 2, and 3, with the yellow line representing the projected changes for 2040-2060.

We selected January, April, and August for detailed analysis due to their expected variation in temperature distribution patterns. The vertical lines in the figures denote the medians of these distributions. Notably, the median of minimum temperatures in January has risen by almost 3 degrees (2.88°C), while maximum temperatures in January have increased by approximately 1 degree. This finding aligns with earlier results indicating that winter temperatures have risen significantly, with minimum temperatures experiencing the largest increase. For April, minimum temperatures have also risen substantially, but maximum temperatures show a smaller increase of less than half a degree.

In contrast, for August, maximum temperatures have increased the most, with a median increase of nearly 1 degree. When projecting future changes (see Figure 17), the largest increases in median temperatures are observed in January ($\Delta = 0.36$°C) and August ($\Delta = 0.25$°C). The distribution of minimum temperatures has also shifted to the right, with the most significant

Figure 17: Distributions of minimum and maximum annual temperatures: past, present and future.



Note: 1000 replications of yearly temperatures from AR(1) model with parameters estimated in years 1780-1820 (in blue), 1980-2020 (in red) and projected change in 40 years (in yellow).

change occurring in January. In summary, high-temperature events are becoming more frequent at higher temperatures, particularly in winter, while low-temperature events are becoming less severe overall.

Overall, the simulation study reveals the following key findings: (1) The distribution of average temperatures has shifted to the right, increasing the likelihood of higher temperatures. (2) This shift is not uniform; it is more pronounced in colder months. (3) The maximum monthly temperatures projected by the three models show a higher probability of being warmer.

These findings align with results from the estimation phase, which demonstrated that temperature persistence has generally increased (particularly during winter months) and that the variability of the error term has decreased, becoming more uniform across seasons. Given these changes, we expect an increase in both the frequency and duration of heatwaves due to higher temperatures, a greater likelihood of warmer maximum temperatures, increased persistence, and reduced variability.

To corroborate this expectation, we analyzed the frequency of cold spells and heatwaves using simulated distributions from the three models, segmented by season. We defined a heatwave (or cold spell) based on the reference period from 1900-1940. We then compared the number of days in a row (1, 2, 5, 7, or 10) with temperatures exceeding the 90%, 95%, and 99% quantiles (or falling below the 10%, 5%, and 1% quantiles for cold spells) of the 1900-1940 reference distribution.

Tables 1 and 2 present the results for heatwaves and cold spells, respectively. These tables are organized by season and by period (1870-1820, 1980-2022, and 2020-2060), and show the frequency of heatwaves (or cold spells) by duration and intensity. The quantiles of the reference period are provided at the end of both tables.

The results indicate that, using a threshold such as q_{90}, the frequency of heatwaves has increased across all seasons and durations, with both 1-day and 10-day heatwaves becoming more common. Generally, the more recent years show a higher frequency of heatwaves, with autumn months projected to experience an increased number of such events.

In terms of intensity, measured as the frequency of temperatures exceeding the 99% percentile, the model based on years 1780-1820 shows minimal heatwave activity at the q_{99} threshold, with only a small probability of 1-day or 2-day heatwaves. In contrast, intensity increases for the 1980-2020 period, and the frequency of heatwaves at the q_{99} level is even higher for simulations based on 2020-2060. This trend is evident in the frequency of 1-day and 2-day heatwaves across all seasons, with up to 5-day heatwaves in autumn and 7-day heatwaves in winter.

Table 1: Frequency of heatwaves by season and duration (1, 2, 5, 7, 10 days).

		Spring					Summer				
Years	Quantiles	1	2	5	7	10	1	2	5	7	10
1780-1820	$> q_{90}$	6.2	4.1	1.3	0.6	0.2	4.2	2.3	0.5	0.2	0.0
	$> q_{95}$	2.7	1.5	0.3	0.1	0.0	1.6	0.7	0.1	0.0	0.0
	$> q_{99}$	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1980-2020	$> q_{90}$	7.4	5.2	1.9	1.0	0.4	9.5	6.4	2.2	1.2	0.5
	$> q_{95}$	3.3	2.0	0.5	0.2	0.1	4.5	2.7	0.7	0.3	0.1
	$> q_{99}$	0.3	0.2	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0
2020-2060	$> q_{90}$	7.8	5.6	2.3	1.2	0.5	10.9	7.6	2.9	1.6	0.7
	$> q_{95}$	3.5	2.3	0.7	0.3	0.1	5.5	3.5	1.0	0.5	0.2
	$> q_{99}$	0.4	0.2	0.0	0.0	0.0	0.5	0.2	0.0	0.0	0.0
		Autumn					Winter				
		1	2	5	7	10	1	2	5	7	10
1780-1820	$> q_{90}$	5.1	3.4	1.0	0.5	0.1	3.4	1.9	0.4	0.1	0.0
	$> q_{95}$	2.0	1.1	0.2	0.1	0.0	1.9	1.0	0.2	0.0	0.0
	$> q_{99}$	0.2	0.1	0.0	0.0	0.0	0.6	0.3	0.0	0.0	0.0
1980-2020	$> q_{90}$	10.8	8.0	3.5	2.1	0.9	7.8	5.3	1.9	1.0	0.4
	$> q_{95}$	4.8	3.1	1.0	0.5	0.1	4.6	2.9	0.8	0.4	0.1
	$> q_{99}$	0.7	0.3	0.0	0.0	0.0	1.5	0.9	0.2	0.1	0.0
2020-2060	$> q_{90}$	12.8	9.6	4.5	2.7	1.3	10.0	7.1	2.9	1.6	0.7
	$> q_{95}$	6.0	4.0	1.3	0.7	0.2	6.1	4.1	1.4	0.7	0.3
	$> q_{99}$	0.8	0.4	0.1	0.0	0.0	2.2	1.3	0.3	0.1	0.0
Quantiles						q_1	q_5	q_{10}	q_{90}	q_{95}	q_{99}
Temperature reference period 1900-1940						-1.6	1.1	2.6	16.1	17.4	19.7

Table 2: Frequency of coldspells by season and duration (1, 2, 5, 7, 10) days.

		Spring					Summer				
Years	Quantiles	1	2	5	7	10	1	2	5	7	10
1780-1820	$< q_{10}$	7.5	5.5	2.4	1.4	0.6	5.1	3.2	0.8	0.4	0.1
	$< q_5$	4.0	2.7	0.9	0.5	0.1	2.4	1.4	0.3	0.1	0.0
	$< q_1$	0.9	0.5	0.1	0.0	0.0	0.5	0.2	0.0	0.0	0.0
1980-2020	$< q_{10}$	3.5	2.2	0.7	0.3	0.1	2.8	1.5	0.3	0.1	0.0
	$< q_5$	1.4	0.8	0.2	0.1	0.0	1.1	0.5	0.1	0.0	0.0
	$< q_1$	0.1	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0
2020-2060	$< q_{10}$	2.9	1.8	0.5	0.2	0.1	2.4	1.3	0.2	0.1	0.0
	$< q_5$	1.1	0.6	0.1	0.0	0.0	1.0	0.4	0.1	0.0	0.0
	$< q_1$	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
		Autumn					Winter				
		1	2	5	7	10	1	2	5	7	10
1780-1820	$< q_{10}$	7.6	5.0	1.6	0.8	0.3	11.8	8.2	3.1	1.7	0.7
	$< q_5$	3.8	2.3	0.5	0.2	0.1	8.4	5.5	1.8	0.9	0.3
	$< q_1$	0.8	0.4	0.1	0.0	0.0	1.1	0.5	0.1	0.0	0.0
1980-2020	$< q_{10}$	3.6	2.4	0.8	0.4	0.2	4.7	3.1	1.0	0.5	0.2
	$< q_5$	1.6	1.0	0.3	0.1	0.0	2.9	1.8	0.5	0.2	0.1
	$< q_1$	0.2	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0
2020-2060	$< q_{10}$	2.5	1.7	0.5	0.3	0.1	3.3	2.2	0.7	0.3	0.1
	$< q_5$	1.0	0.6	0.2	0.1	0.0	1.9	1.2	0.3	0.2	0.0
	$< q_1$	0.2	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Quantiles						q_1	q_5	q_{10}	q_{90}	q_{95}	q_{99}
Temperature reference period 1900-1940						-1.6	1.1	2.6	16.1	17.4	19.7

6 Conclusions

This paper introduces a novel approach to analyze data using a periodic autoregression that has the capability to capture cyclical changes in parameters in a time varying manner. Empirical application of this modeling framework to Central England Temperature (CET) data showcases how the evolution of both intercept and autoregressive coefficients in a periodic AR(1) model can be captured to reveal new insights on daily temperature changes over long periods of time. The analysis spans daily average temperatures over 1772 to 2020 and yields several notable findings. First, the AR(1) model provides a robust fit for data segments up to 40 years in duration with residuals that exhibit uncorrelated, normally distributed behavior, indicating that the underlying temporal structure of the temperature data are well captured within these periods. Second, significant cyclical variations in the model parameters are revealed in the regression results, pointing to changes in the persistence parameter (ϕ) and the residual variance over time. The findings indicate a statistically significant increase in temperature persistence, the most pronounced changes occurring during the winter and autumn months. This increase in persistence implies that temperature anomalies are more likely to persist over longer periods, particularly in these seasons. Conversely, the residual variance manifests a declining trend, especially in the winter months, suggesting that the amplitude of temperature fluctuations has lessened over time.

These findings helped guide a simulation study aimed at exploring broader implications of our results. The simulation design focused on learning how the estimated cyclical changes in temperature persistence and residual variation impact temperature distributions, considering average temperatures as well as maximum and minimum temperatures. This approach enabled an assessment of the frequency of extreme weather events, such as heatwaves and episodes of extreme cold weather, using the observed trends in temperature persistence. The simulation results therefore provide insights into how such extremes might evolve under prevailing and possible future climatic conditions.

In sum, this paper contributes to climate analysis with observational data using econometric methods for periodically correlated time series that provide a framework for capturing cyclical patterns in temperature data and their implications for understanding climate variability. In addition, simulations with the fitted model provide a view of how changes in temperature persistence and variability can influence various aspects of climate, from daily temperature distributions to the frequency and severity of extreme weather events.

References

Di Cecco, G. J. and Gouhier, T. C. (2018). Increased spatial and temporal autocorrelation of temperature under climate change. *Scientific Reports*, 8(1):1–9.

Francis, J., Skific, N., and Vavrus, S. (2018). North American weather regimes are becoming more persistent: Is Arctic amplification a factor? *Geophysical Research Letters*, 45(20):11–414.

Francis, J. A. and Vavrus, S. J. (2015). Evidence for a wavier jet stream in response to rapid Arctic warming. *Environmental Research Letters*, 10(1):014005.

Gasparrini, A. and Armstrong, B. (2011). The impact of heat waves on mortality. *Epidemiology (Cambridge, Mass.)*, 22(1):68.

Giraitis, L., Marotta, F., and Phillips, P. C. B. (2024). Estimation of partially observed AR(p) time series. Mimeo.

Graham, R. M., Cohen, L., Petty, A. A., Boisvert, L. N., Rinke, A., Hudson, S. R., Nicolaus, M., and Granskog, M. A. (2017). Increasing frequency and duration of Arctic winter warming events. *Geophysical Research Letters*, 44(13):6974–6983.

Harvey, D. I. and Mills, T. C. (2003). Modelling trends in Central England temperatures. *Journal of Forecasting*, 22(1):35–47.

He, C., Kang, J., Teräsvirta, T., and Zhang, S. (2019). The shifting seasonal mean autoregressive model and seasonality in the Central England monthly temperature series, 1772–2016. *Econometrics and Statistics*, 12:1–24.

IPCC (2021). Summary for policymakers. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Lenssen, N. J., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, 124(12):6307–6326.

Lenton, T. M., Dakos, V., Bathiany, S., and Scheffer, M. (2017). Observed trends in the magnitude and persistence of monthly temperature variability. *Scientific Reports*, 7(1):1–10.

Li, J. and Thompson, D. W. (2021). Widespread changes in surface temperature persistence under climate change. *Nature*, 599(7885):425–430.

Manley, G. (1974). Central England temperatures: monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society*, 100(425):389–405.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, 78(6):1069–1080.

McCarthy, G. D., Haigh, I. D., Hirschi, J. J.-M., Grist, J. P., and Smeed, D. A. (2015). Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. *Nature*, 521(7553):508–510.

Parker, D. E., Legg, T. P., and Folland, C. K. (1992). A new daily central England temperature series, 1772–1991. *International Journal of Climatology*, 12(4):317–342.

Pfleiderer, P. and Coumou, D. (2018). Quantification of temperature persistence over the Northern hemisphere land-area. *Climate Dynamics*, 51(1):627–637.

Proietti, T. and Hillebrand, E. (2017). Seasonal changes in Central England temperatures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(3):769–791.

Rho, S.-H. and Vogelsang, T. J. (2019). Heteroskedasticity autocorrelation robust inference in time series regressions with missing data. *Econometric Theory*, Cambridge Vol 35, Iss. 3:601–629.

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

A Appendix

A.1 Computation of residuals

Local residuals. Section 3 outlines the methodology for estimating the cyclical parameters $\alpha_t, \phi_t,\ t = 1, ..., T$ of the periodic autogregressive time series PAR(1) x_i given in (1). The estimators $\widehat{\alpha}_t$ and $\widehat{\phi}_t$ in (5) are based on the block of data x_i, $i \in J_{t,w}$. The 'local' residuals are also computed by block. For every block $J_{t,w}$ around t and of width w, we define them as the residuals of the fitted local AR(1) model

$$\begin{aligned} \widehat{\varepsilon}_{t,i} &= (y_i - \widehat{\alpha}_t - \widehat{\phi}_t y_{i-1}) h_i h_{i-1} \\ &= (x_i - \widehat{\alpha}_t - \widehat{\phi}_t x_{i-1}) h_i h_{i-1}, \quad i = 1, ..., N. \end{aligned}$$

Suppose that the variance $\mathbb{E}\varepsilon_i^2$ of ε_i in the PAR(1) model (1) is also a cyclical parameter, i.e.,

$$\mathbb{E}\varepsilon_{jT+t}^2 = \sigma_t^2, \quad t = 1, \ldots, T, \quad j = 0, \ldots, P-1.$$

Then $\sigma_t^2 = \mathbb{E}\varepsilon_t^2$, $t = 1, ..., T$ can be estimated using local residuals as

$$\hat{\sigma}_t^2 = (\sum_{i=1}^{N} h_i)^{-1} \sum_{i=1}^{N} \widehat{\varepsilon}_{t,i}^2.$$

Recall that the missing data indicator $h_i = h_{i,tw}$ above is given in (2) and depends on t.

Global residuals. Let x_i, $i = 1, \ldots, N$, be a Periodic Autoregressive PAR(1) time series as described in equation (1). The global residuals $\widetilde{\varepsilon}_i$ of the PAR(1) model are computed as follows:

$$\widetilde{\varepsilon}_{jT+t} = x_{jT+t} - \widehat{\alpha}_t - \widehat{\phi}_t x_{jT+t-1}, \qquad j = 1, \ldots, P, \quad t = 1, \ldots, T,$$

where $N = PT$. Here $\widehat{\alpha}_t$ and $\widehat{\phi}_t$, $t = 1, \ldots, T$, are fitted time-varying periodic parameters and T is the length of the cycle.

A.2 Monte Carlo experiment

This section reports the findings of a Monte Carlo experiment to evaluate the performance of the estimators $\widehat{\alpha}_t$ and $\widehat{\phi}_t$ for the cyclically varying parameters α_t and ϕ_t in PAR(1) model (1). The goal is to assess how well the estimation procedure captures the cyclical nature of the data, ensuring that the estimators of parameters are unbiased and that the normal approximation performs well in finite samples.

The model in Eq.(1) was simulated to produce $2,000$ samples. Each simulation generated data for $N = P \times T$ observations, where $P = 100$ represents the number of cycles, and $T = 500$ is the number of observations per cycle. In each cycle the following periodic behavior was assumed for the mean $\mu_t = \alpha_t/(1-\phi_t)$ and the autoregressive parameter ϕ_t:

$$\mu_t = 4\sin\left(2\pi\frac{t}{365}\right), \qquad \phi_t = 0.8\sin\left(2\pi\frac{t}{365}\right), \quad t = 1, ..., 500. \tag{9}$$

The periodicity of the mean and autoregressive parameters was set to reflect the cyclic behavior observed for the actual temperatures data. To illustrate the PAR(1) estimation method we use three different window widths: $w = 2, 6, 10$. The results of the estimation for a single sample are reported Figure A.1. This depicts the true values of the parameters ϕ_t, and α_t (red line) and the estimates of $\widehat{\phi}_t$, and $\widehat{\alpha}_t$ (black line) along with the 95% confidence bands for different window widths, $w = 2, 6, 10$ respectively.

Three key findings emerge from these simulations. First, estimation of ϕ_t is highly accurate even when ϕ_t varies smoothly over time. The estimator $\widehat{\phi}_t$ captures the periodic nature of both the mean and autoregressive parameters well, with the true parameter values ϕ_t consistently falling within the 95% confidence bands. Second, when both the mean μ_t and the autoregressive parameter ϕ_t change smoothly, the rolling window estimator effectively tracks these changes. The estimator for α_t remains highly accurate, with the confidence bands remaining reasonably tight. Third, the results are robust across different window widths. Using a fixed rolling window width of $w = 10$ strikes a balance between capturing smooth periodic changes in the parameters and maintaining precise estimates.

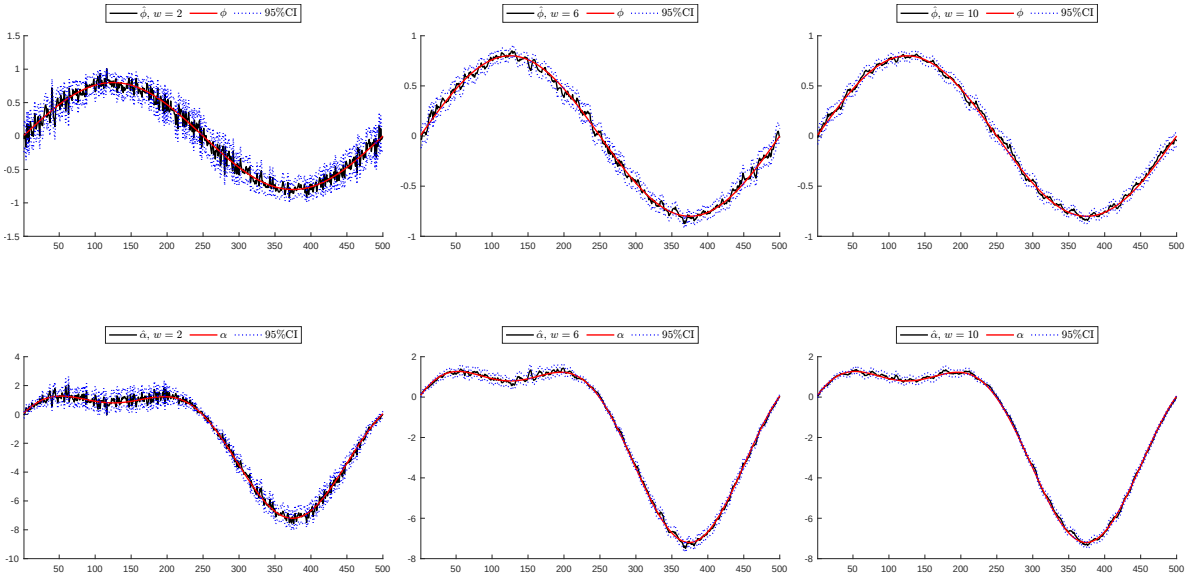
In the second experiment, simulations were conducted to evaluate the bias and coverage probabilities of the rolling window estimators of the AR(1) parameters, $\widehat{\alpha}_t$ and $\widehat{\phi}_t$. $2,000$ samples were generated with $P = 100$ cycles, each with $T = 500$ observations per cycle. The rolling window widths were $w = 2, 6, 10$. The results are presented in Figure A.2, which reports the true value of the cycles for α_t and ϕ_t along with sample average estimates over the 2,000 simulations ($\hat{\mathbb{E}}\widehat{\alpha}_t$ and $\hat{\mathbb{E}}\widehat{\phi}_t$). The second row of the plots shows empirical coverage probabilities for nominal 95% confidence intervals based on the standard normal approximation (6).

The new simulations confirm the results observed in the previous simulation for one sample. Both the intercept and the autoregressive parameter are estimated with high accuracy. The

estimates for $\widehat{\alpha}_t$ and $\widehat{\phi}_t$ closely follow the true periodic patterns. The empirical coverage rate overall is close to the nominal 95% coverage. The normal approximation defined in Eq.6 - 7 holds well overall, providing further evidence that the proposed method satisfactorily captures periodic patterns in the data.

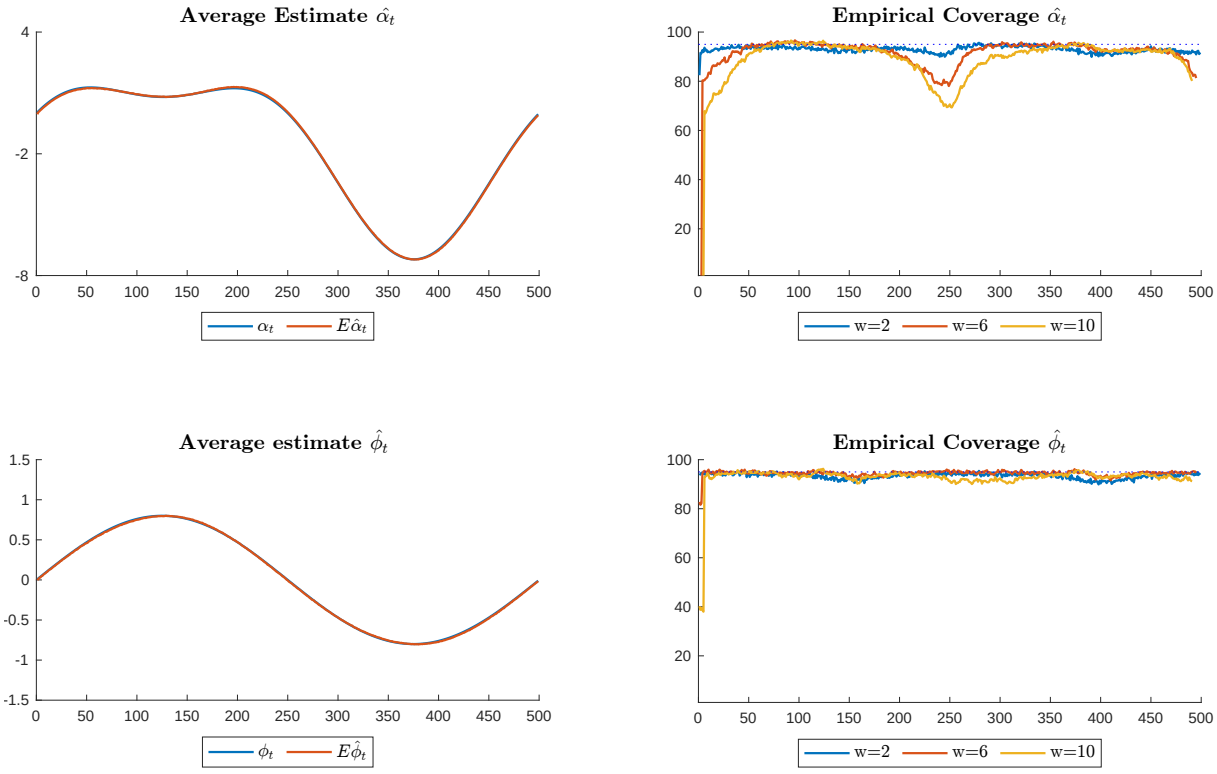
As expected, coverage rates start to be distorted only at those points where sharp changes in the real parameters occur. Estimating with a larger window width helps the accuracy of the estimation at time t but this happens at the expenses of the accurate detection of a sharp change, for which a smaller window width is naturally betted suited. These results highlight the effectiveness of the rolling window estimator in accurately capturing smooth changes in periodic parameters. Overall, the simulations confirm that the approach is reliable in analyzing periodic time series with fixed cycle lengths and autoregressive patterns.

Figure A.1: Estimation results for periodic parameters μ_t and ϕ_t for one sample. Model (9).



Note: μ_t and ϕ_t as in model (9), $P = 100$ cycles, $T = 500$ observations per cycle, window width $w = 2, 6, 10$. The blue dotted line represents the 95% confidence bands.

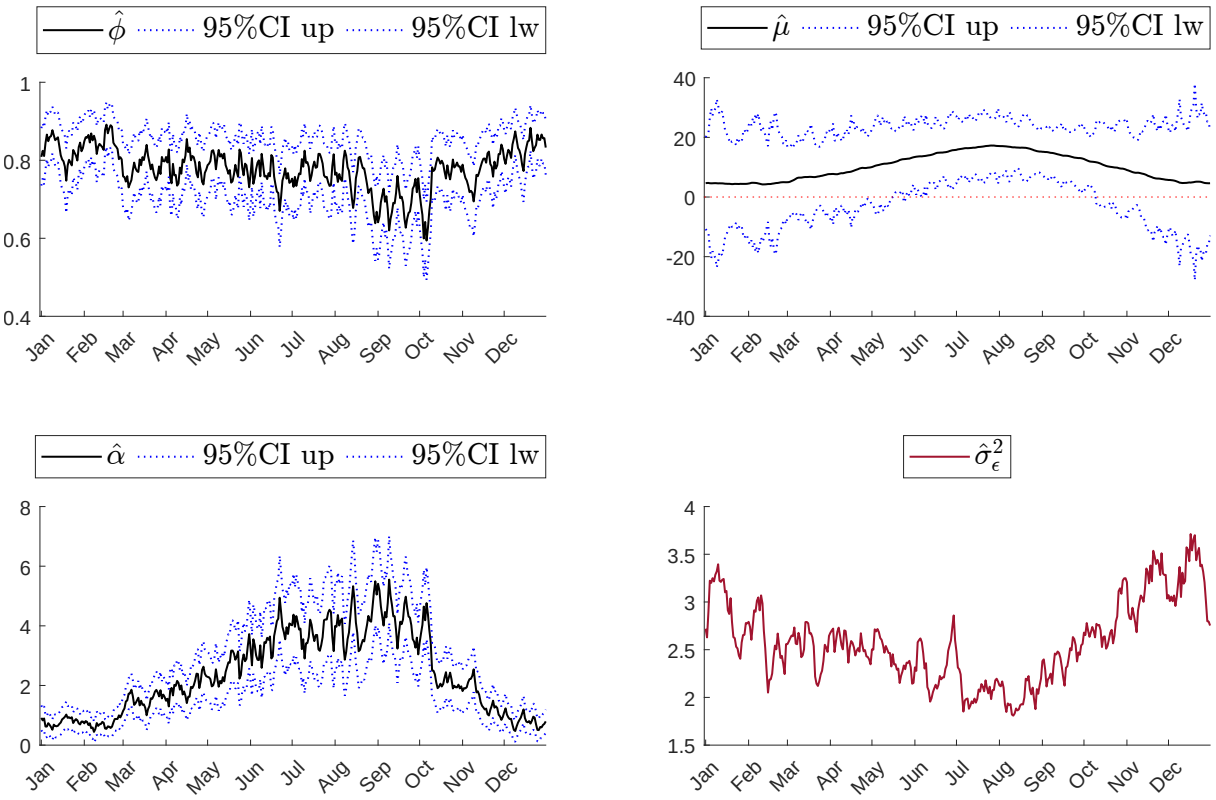
Figure A.2: Monte Carlo results for periodic parameters μ_t and ϕ_t. Model (9).



Note: The four panels show: α_t, sample averages $\hat{\mathbb{E}}\widehat{\alpha}_t$ and $\hat{\mathbb{E}}\widehat{\phi}_t$, empirical coverage probabilities for nominal 95% confidence intervals for α_t and ϕ_t, each reported for window widths $w = 2, 6, 10$. $P = 100$ cycles, $T = 500$ observations per cycle, and $2,000$ replications.

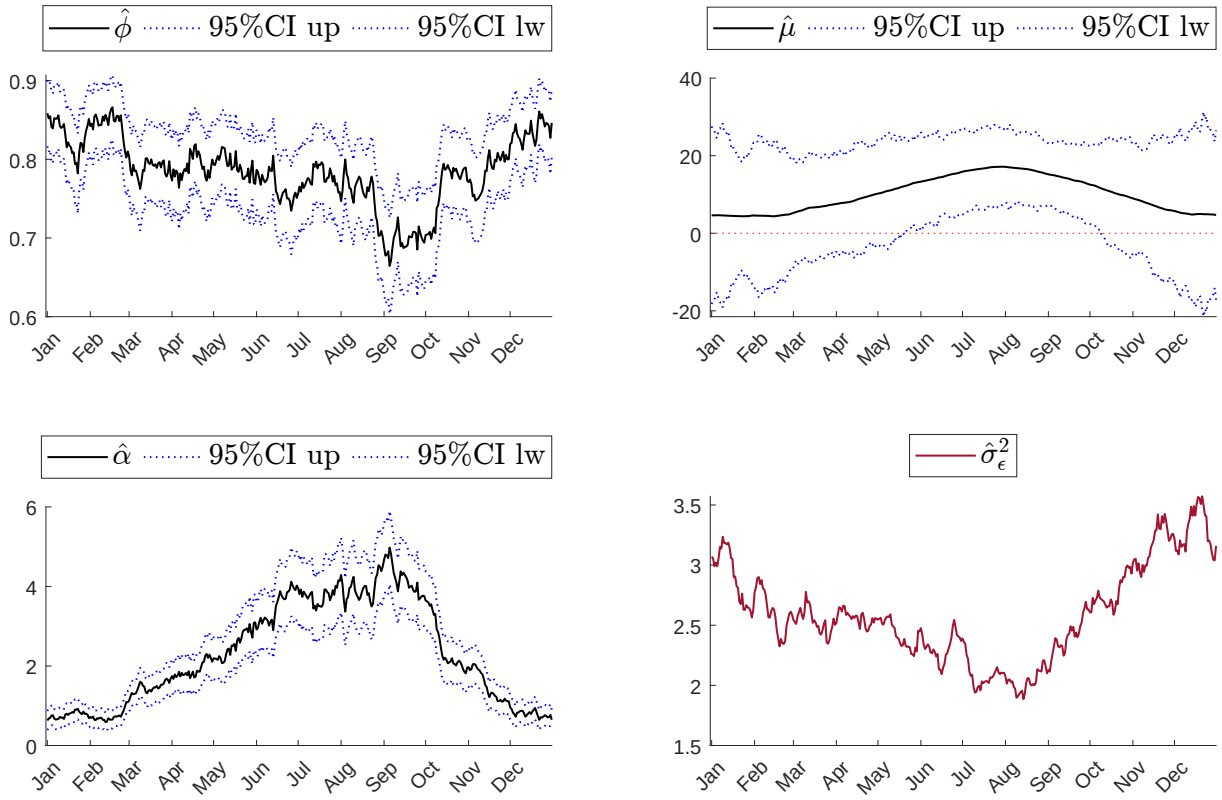
A.3 Further figures

Figure A.3: Estimation results of parameters ϕ_t, α_t, μ_t, $\sigma^2_{\varepsilon,t}$



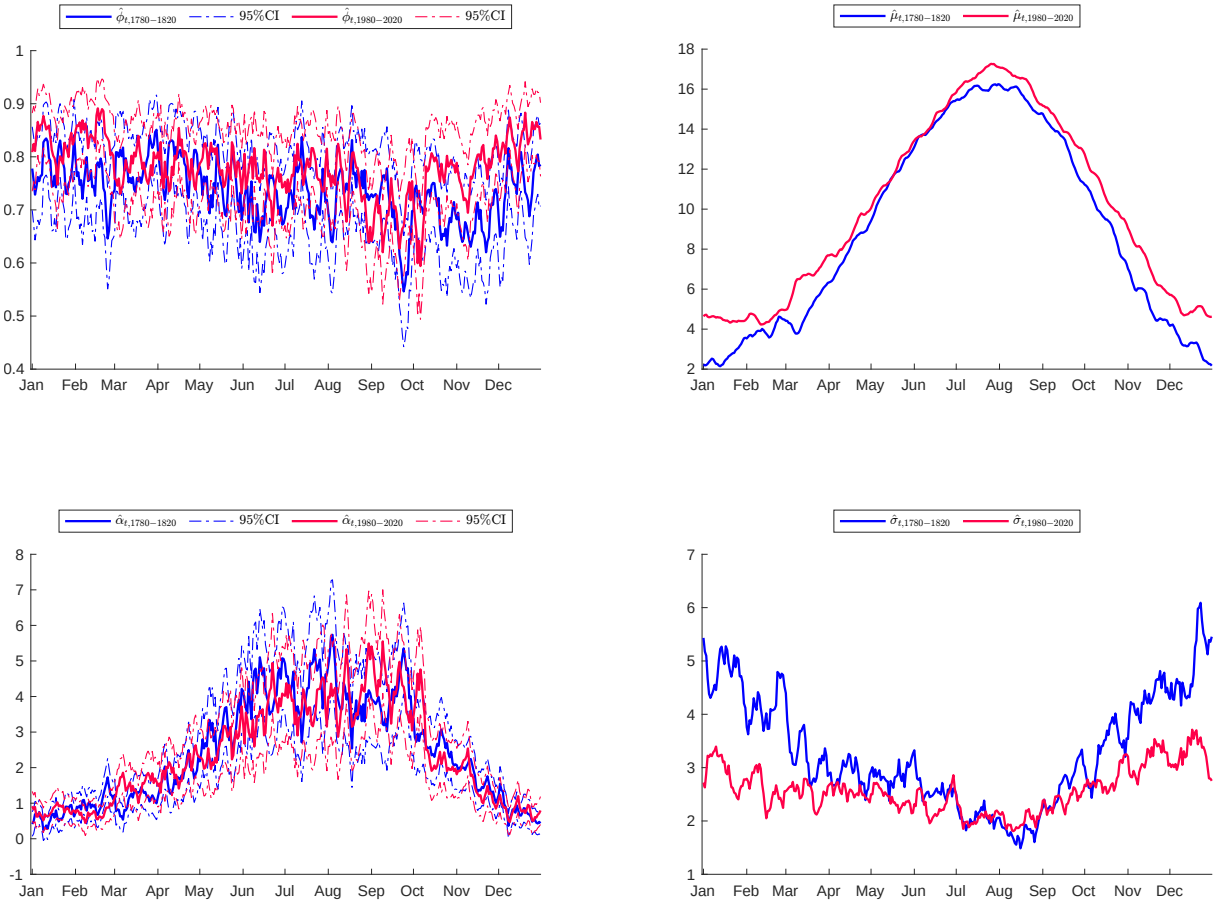
Note: Years 1980-2020, group of 40 years, w=7.

Figure A.4: Estimation results of parameters ϕ_t, α_t, μ_t, $\sigma^2_{\varepsilon,t}$



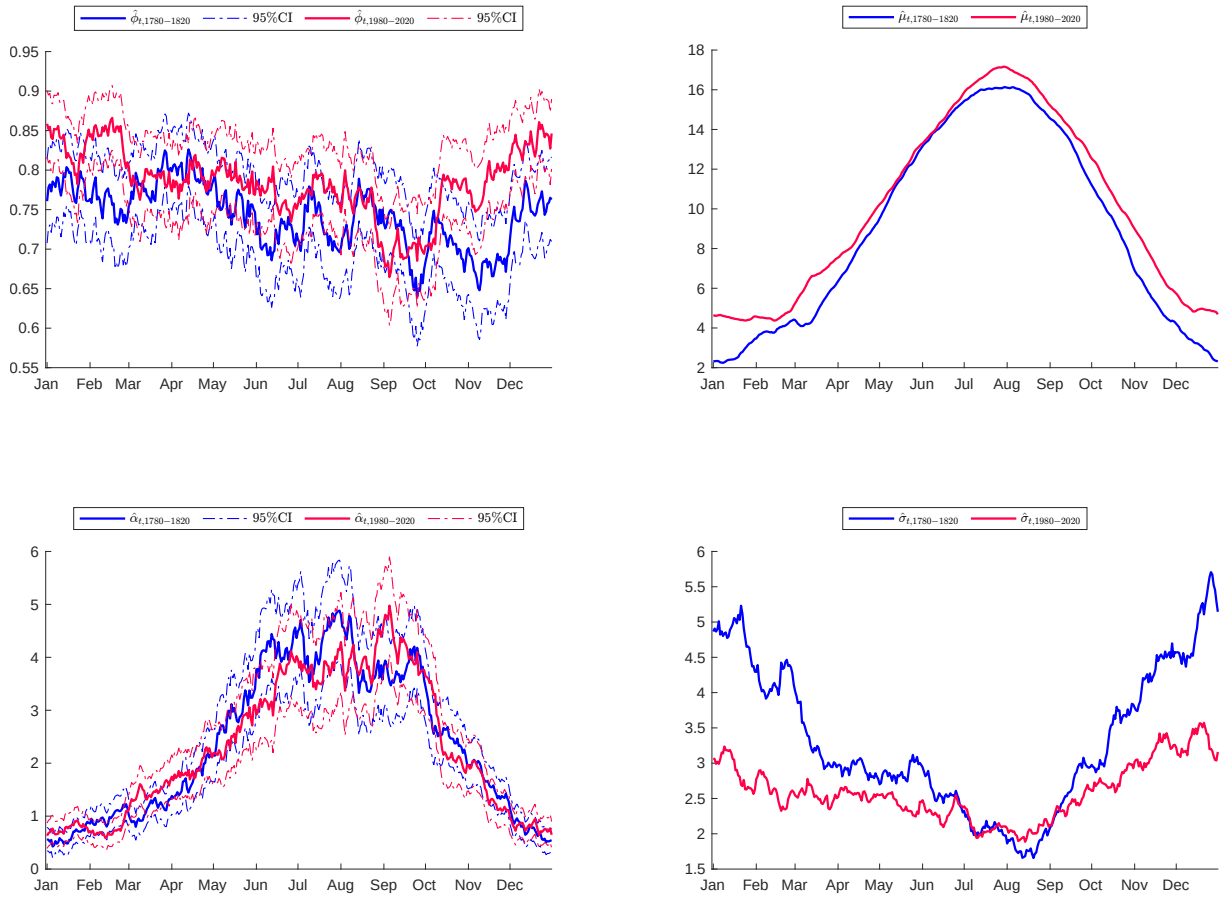
Note: Years 1980-2020, groups of 40 years, w=15 .

Figure A.5: Comparing parameters ϕ_t, α_t, μ_t, $\sigma^2_{\varepsilon,t}$, in the past and present: 1780-1820 and 1980-2020.



Note: Years 1780-1820 and 1980-2020, groups of 40 years, w=7.

Figure A.6: Comparing parameters ϕ, α, μ, $\sigma^2_{\varepsilon,t}$, in the past and present: 1780-1820 and 1980-2020.



Note: Years 1780-1820 and 1980-202, groups of 40 years, w=15.