STATISTICAL FORECASTING
How fast will future warming be?

Terence C. Mills

The Global Warming Policy Foundation

GWPF Technical paper 1

GWPF REPORTS

THE GLOBAL WARMING POLICY FOUNDATION

CREDITS

Cover image Light pollution in Europe, 2002
Public domain artwork.

STATISTICAL FORECASTING
How fast will future warming be?

Terence C. Mills

Contents

Foreword

By Professor Ross McKitrick

Economists have put a lot of effort over the years into devising and running elaborate modelling systems to generate forecasts of macroeconomic indicators, financial markets, resource prices, and other key economic quantities. But the repeated failures of such models to generate accurate predictions has taught the profession a healthy scepticism about the ability of large structural models, regardless of how complex, to provide reliable forecasts. A particularly acute challenge arose when relatively simple statistical time-series methods began yielding better forecasts than massive system-simulation models.

It is difficult not to wonder whether a parallel with modern climatology will arise. Like the economy, the climate is a deeply complex system that defies simple representation. Giant computer modelling systems have been developed to try and simulate its dynamics, but their reliability as forecasting tools is proving to be very weak. The problem is that many important policy decisions are based on climate-model projections of the future, on the assumption that they can be treated as forecasts. If they are not valid for this purpose, we need to know whether there are alternative methods that are.

In this insightful essay, Terence Mills explains how statistical time-series forecasting methods can be applied to climatic processes. The question has direct bearing on policy issues since it provides an independent check on the climate-model projections that underpin calculations of the long-term social costs of greenhouse gas emissions. In this regard, his conclusion that statistical forecasting methods do not corroborate the upward trends seen in climate model projections is highly important and needs to be taken into consideration.

As one of the leading contributors to the academic literature on this subject, Professor Mills writes with great authority, yet he is able to make the technical material accessible to a wide audience. While the details may seem quite mathematical and abstract, the question addressed in this report is of great practical importance not only for improving the science of climate forecasting, but also for the development of sound long-term climate policy.

Ross McKitrick
Guelph, January 2016

Ross McKitrick is Professor of Economics, University of Guelph and Research Chair in Energy, Ecology and Prosperity at the Frontier Centre for Public Policy.

About the author

Terence Mills is Professor of Applied Statistics and Econometrics at Loughborough University.

Summary

The analysis and interpretation of temperature data is clearly of central importance to debates about anthropogenic global warming (AGW). Climatologists currently rely on large-scale general circulation models to project temperature trends over the coming years and decades. Economists used to rely on large-scale macroeconomic models for forecasting, but in the 1970s an increasing divergence between models and reality led practitioners to move away from such macro modelling in favour of relatively simple statistical time-series forecasting tools, which were proving to be more accurate.

In a possible parallel, recent years have seen growing interest in the application of statistical and econometric methods to climatology. This report provides an explanation of the fundamental building blocks of so-called 'ARIMA' models, which are widely used for forecasting economic and financial time series. It then shows how they, and various extensions, can be applied to climatological data. An emphasis throughout is that many different forms of a model might be fitted to the same data set, with each one implying different forecasts or uncertainty levels, so readers should understand the intuition behind the modelling methods. Model selection by the researcher needs to be based on objective grounds.

ARIMA models are fitted to three representative data sets: the HADCRUT4 global surface series, the RSS global lower troposphere series and the Central England Temperature (CET) series. A clear finding presents itself for the two global temperature series. Irrespective of the model fitted, forecasts do not contain any trend, with long-horizon forecasts being flat, albeit with rather large measures of imprecision even from models in which uncertainty is bounded. This is a consequence of two interacting features of the fitted models: the inability to isolate a significant drift or trend parameter and the large amount of overall noise in the observations themselves compared to the fitted 'signals'. The CET exhibits season-specific trends, with evidence of long-term warming in the winter months but not in the summer.

1 Introduction

The analysis and interpretation of temperature data is clearly of central importance to debates about anthropological global warming (AGW) and climate change in general. For the purpose of projecting future climate change, scientists and policymakers rely heavily on large-scale ocean–atmosphere general circulation models, which have grown in size and complexity over recent decades without necessarily becoming more reliable at forecasting. The field of economics spent the post-war decades developing computerised models of the economy that also grew to considerable size and complexity, but by the late 1970s two uncomfortable truths had been realised. First, these models produced generally poor forecasts, and adding more equations or numerical detail did not seem to fix this. Second, relatively simple statistical models that had no obvious basis in economic theory were proving much more reliable at forecasting. It took many years for economists to rationalise statistical forecasting by working out its structural connections to this theory. But before this had happened, economic practitioners were already relying on these models simply because of their relative success.

Is there a parallel with climatology? In recent years, statisticians and econometricians have begun applying the tools of statistical forecasting to climate datasets. As these exercises have become more and more successful, there is a corresponding concern that such models either have no basis in climatological theory, or may even seem to contradict it. In this report we focus on forecasting models in general, and their application to climate data in particular, while leaving aside the potentially interesting question of how such models might or might not be reconciled with the physical theory underpinning climate models.

Data organised as evenly-spaced observations over time are called 'time series'. The analysis of time series has a long and distinguished history, beginning with descriptive examinations but with major technical advances occurring in the early years of the 20th century, following quickly on from the development of the concept of correlation.[1] The publication of the first edition of George Box and Gwilym Jenkins' famous book *Time Series Analysis: Forecasting and Control* in 1970 brought the techniques for modelling and forecasting time series to a wide audience. Their methods have since been extended, refined and applied to many disciplines, notably economics and finance, where they provide the foundations for time series econometrics.[2] Although some of the fundamental developments in time series were made using meteorological data, it is notable that many contributors to the debates concerning AGW and climate change seem unaware of this corpus of theory and practice, although contributions by time series econometricians have now begun to appear, albeit with rather limited influence on such debate.[3,4]

The main purpose of this report is to set out a framework that encompasses a wide range of models for describing the evolution of an individual time series. All such models decompose the data into random and deterministic components, and then use these components to generate forecasts of future observations, accompanied by measures of uncertainty. A central theme of the report is that the choice of model has an important impact on the form of the forecasts and on the behaviour of forecast uncertainty, particularly as the forecast horizon increases. But since some models fit the data better than others, we are able to provide some guidance about which sets of forecasts are more likely to be accurate. The framework is illustrated using three readily available and widely used temperature series. These are:

- the HADCRUT4 global land and sea surface anomaly series, available monthly from January 1850
- the Remote Sensing System (RSS) lower troposphere series, available monthly from January 1979
- Central England temperatures (CET), available monthly from January 1659.

In each case the series are examined up to December 2014.[5]

All computations are performed using commercially available software, so that the analyses should be easily replicable, and hence could be refined and extended by anyone familiar with such software. Indeed, it is taken to be the very essence of statistical modelling that these models, and hence the forecasts computed using them, should be subjected to 'severe testing' and subsequently replaced by superior models if found wanting in any aspect.

2 Basic time-series modelling

The ARIMA framework

We will develop several popular time-series models in some detail, with the technical details provided in the Appendix. Their various implications can best be understood by starting with their underlying structures. The basic model begins with a temperature time series x_t observed over the period $t = 1, 2, \ldots, T$. Our aim is to forecast the values of x_t at future times $T+1, T+2, \ldots$ etc.[6] The simplest decomposition breaks x_t down into the sum of a deterministic term (the level) and a random term (noise):

$$x_t = \mu_t + \varepsilon_t \tag{1}$$

where μ_t and ε_t are the level and noise components, respectively. These are typically assumed to be independent of (or uncorrelated with) each other.

Two essential points about statistical forecasting methods are as follows.

1. The deterministic level component depends only on time and upon invariant parameters that can be estimated from the data. Once they have been estimated, μ_t can be forecast precisely for any period in the future. The validity of

the forecast obviously depends on the validity of the equation used and the quality of the parameter estimates.

2. The random component, if identified correctly, averages out to zero over time and has a stable variance, so model (1) will yield a forecast consisting of a central tendency with uncertainty bounds.[7] The width of the uncertainty bounds will depend upon the behaviour of ε_t .

The basic form of the deterministic component typically consists of a constant term, which we will call μ_0 , and a trend, $\mu_1 t$, although with additional parameters we can also allow for breaks and jumps in the trend, as well as recurring seasonal patterns. A 'non-random' model for x_t could then be defined as

$$x_t = \mu_0 + \mu_1 t$$

If we knew, or could estimate, the values of the μ coefficients we would then be able to forecast future values of x_t perfectly. However, this would not be a very sensible model because, no matter how good our coefficient estimates are, the model would almost certainly never perfectly match the data, since there will always be random noise terms. Now suppose, for the purpose of an example, that all our observations on x_t take the same value. Then we could write down two models that fit the observed data equally well:

$$x_t = \mu_0 + a_t$$

and

$$x_t = x_{t-1} + a_t \tag{2}$$

Here a_t denotes a sequence of zero-mean, constant-variance, independent errors, or 'innovations', typically referred to as 'white noise'. As all values of x_t are the same in the observed sample, all the a_ts in this sample will be zero. Although both models will fit the observed data equally well, they imply different things about the future. If, for some reason, a future value of a_t is non-zero then equation (2) implies that the level of x_t at this time would shift by the amount a_t to a new level and this change would be permanent. In contrast, the first model predicts that x_t will always return to its original value since the shift lasts for just a single period: the change would thus be transitory. Hence it is clear why the form of the model can imply different forecasts. In general we expect to be able to determine which form is likelier from its fit to the historical data.

A basic form of the noise component of the model comprises two processes: an autoregression and a moving average.

- In the autoregression x_t is determined by one or more past values of x_t. If x_t is dependent on only one lag, x_{t-1}, it is an autoregression of order 1, denoted

AR(1). If it depends on two lags, x_{t-1} and x_{t-2}, then this is an autoregression of order 2, denoted AR(2), etc.

- The second process is the moving average of current and past white noise error terms. An example is $\varepsilon_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$, which we call an MA(2) process, and so forth.

Suppose we construct a time-series model using the above combination of building blocks. First, if we are interested in modelling the deviations around the mean our basic model will be:

$$x_t = \mu_0 + (\text{AR and MA terms})$$

It tidies the notation a bit to define the deviation $\hat{x}_t = x_t - \mu_0$ so that we can write out a model as, say,

$$\hat{x}_t = \beta_1 \hat{x}_{t-1} + \beta_2 \hat{x}_{t-2} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} \tag{3}$$

The right-hand side is seen to comprise an AR(2) process along with an MA(3) process. Models such as these are known as 'autoregressive-moving average' processes or ARMA(p, q) for short, where p is the order of the AR process and q is the order of the MA process. Thus equation (3) is an ARMA(2, 3) process.

Much of time-series analysis concerns estimating the coefficients of ARMA models. Since there are $p \times q$ possible forms of equation (3) there can be many potential models to choose from, so statistical techniques and algorithms have been developed to search through and identify the optimal form. In general, the more parameters that get added to a model the less precisely each one is estimated, so the algorithms have to trade-off the quality of fit with the benefit of parsimony.

We can add a trend term by changing the basic model to:

$$x_t = \mu_0 + \mu_1 t + (\text{AR and MA terms})$$

so that now $\hat{x}_t = x_t - \mu_0 - \mu_1 t$. In the same way we can add shift terms and breaks in the trend, so, for instance, the trend might be μ_1 up to some year and μ_2 thereafter. We can also add recurring seasonal patterns, or make the trend nonlinear by combining the linear term with a smoothly varying function of time. Once again, there are many possible functions that can be used, and computational methods can be employed to choose the most appropriate.

Equation (2), $x_t = x_{t-1} + a_t$, is a special case of an autoregressive process known as a 'random walk'. If we add a constant term, so that now $x_t = \delta + x_{t-1} + a_t$, we have a random walk with drift, where δ is the drift parameter. We will introduce the symbol ∇ to denote the difference between the current value and the last period's value, or the first difference. So $\nabla x_t = x_t - x_{t-1}$. Then the random walk with drift can be written as $x_t - x_{t-1} = \nabla x_t = \delta + a_t$. Models which contain differences of x_t are often referred to as 'integrated processes'. If first differences are used then the process will be integrated of order one, denoted I(1). If second differences are used, in other words,

$$\nabla^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$$

then the process will be I(2). Thus the general I(1) process will be of the form

$$\nabla x_t = (\text{AR and MA terms})$$

For example:

$$\nabla x_t = \beta_1 \nabla x_{t-1} + \beta_2 \nabla x_{t-2} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3}$$

is an autoregressive-integrated-moving average model of orders 2, 1 and 3, in other word ARIMA $(2, 1, 3)$.

Stationarity

There remains one critically important concept to explain, namely stationarity (and, by implication, non-stationarity). A stationary process is one that, while subject to random shocks, always returns to its mean value. Also, the variance of a stationary process remains constant over time and the correlation between any two observations spaced k intervals apart remains constant as well. Consider the simple example of an AR(1) process:

$$x_t = \theta x_{t-1} + a_t$$

Substituting $\theta x_{t-2} + a_{t-1}$ for x_{t-1} gives $x_t = \theta^2 x_{t-2} + a_t + \theta a_{t-1}$. By repeated substitution k times we obtain

$$x_t = \theta^k x_{t-k} + a_t + \theta a_{t-1} + \theta^2 a_{t-2} + \ldots + \theta^k a_{t-k}$$

If θ is bounded between −1 and +1 then, as k goes to infinity, θ^k must go to zero so the first term will disappear leaving only a_t and its lags, each of which has a mean of zero. So the expected value of x_t will go to zero as well. The variance of x_t can then be shown to be $\sigma_a^2/(1-\theta^2)$, where σ_a^2 is the variance of a_t, and this will be constant regardless of t.

If $\theta = 1$, so that we have a random walk process, then the series becomes nonstationary and two important things change. First, as k gets larger, x_t will no longer return to zero, but will take the value $x_t = x_0 + \sum_{j=0}^{t-1} a_{t-j}$, in other words the starting value x_0 plus the sum of all random shocks since then. Second, the variance increases with time and will eventually go to infinity.

But it is also clear that while x_t is I(1) and therefore nonstationary, its first difference is I(0) and is said to be 'difference stationary'. Alternatively, if x_t is nonstationary but the deviations around a linear trend are stationary, we say that x_t is 'trend stationary'. The statistical properties of nonstationary series are far more complex than stationary series, so an important part of the model-fitting process involves trying to isolate a stationary random process.

If we end up with a model of the form

$$\nabla x_t = \delta + \text{AR and MA terms}$$

then this is an I(1) process with drift. The h-period ahead forecast, when h is large, is that x_t will change by $h\delta$ units, since the expected values of the AR and MA components will tend to zero. If δ is zero, then the best forecast is that x_t will drift up or down by random amounts but will not trend in any particular direction.

This section has presented some basic time-series modelling concepts that underpin statistical forecasting. The Appendix restates all the same material in formal terms that will be familiar to specialists. We now turn to empirical results from applying these tools to temperature data.

3 Fitting basic models to temperature series

HADCRUT4

For this series standard identification techniques (see Appendix) suggested that an ARIMA $(0,1,3)$ process is the most suitable model within this class, being estimated as[8]

$$\nabla x_t = \underset{(0.0008)}{0.0005} + a_t - \underset{(0.022)}{0.520} a_{t-1} - \underset{(0.025)}{0.080} a_{t-2} - \underset{(0.022)}{0.123} a_{t-3} \qquad \hat{\sigma}_a = 0.1234$$

The standard errors shown in parentheses reveal that the constant, estimated to be just 0.0005, is insignificantly different from zero. Omitting this from the model leaves the moving average coefficients unaltered and reduces the estimate of the innovation standard error σ_a marginally to 0.1233. The model can thus be expressed in the form

$$x_t = x_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t = a_t - 0.520a_{t-1} - 0.080a_{t-2} - 0.123a_{t-3}$$

with the implication that temperatures are non-stationary – I(1) – but without any drift upwards (or, indeed, downwards), so they will wander widely from their initial position x_0. The monthly fluctuations in temperatures are negatively correlated with changes up to three months apart, but uncorrelated with changes more than three months apart.

We mentioned above that a trend broken into different segments can also be fitted. As an example of a segmented trend, a model with regimes with breaks at December 1919, December 1944, December 1975 and December 2001 was also fitted to HADCRUT4.[9] We denote the trend adjustment in segment i as $S_t(i)$ so the trend itself is the sum of the initial trend plus the adjustment terms up to that point. With ε_t specified as an AR(4) process, this model is estimated to be:

$$x_t = \underset{(0.029)}{-0.288} - \underset{(0.000055)}{0.000096}t + \underset{(0.00018)}{0.00136}S_t(1) - \underset{(0.00025)}{0.00157}S_t(2)$$
$$+\underset{(0.00028)}{0.00201}S_t(3) - \underset{(0.00063)}{0.00106}S_t(4) + \varepsilon$$

$$\varepsilon = \underset{(0.022)}{0.458}\varepsilon_{t-1} + \underset{(0.025)}{0.158}\varepsilon_{t-2} + \underset{(0.025)}{0.028}\varepsilon_{t-3} + \underset{(0.022)}{0.074}\varepsilon_{t-4} + a_t \qquad \hat{\sigma}_a = 0.1210$$

Note that the noise component is certainly stationary, with the largest root of the autoregressive polynomial being 0.82. The evolving slopes of the trend function are thus estimated to be as shown in Table 1 (see equation (A10) of the Appendix for the definition of the δ coefficients).

Table 1: HADCRUT4 regimes

Regime	**Period**	**Estimated regime slope**	t**-ratio**
1	1850–1919	$\hat{\beta}_1 = -0.00010 \quad (0.00006)$	1.74
2	1920–1944	$\hat{\beta}_1 + \hat{\delta}_1 = 0.00127 \quad (0.00014)$	8.80
3	1945–1975	$\hat{\beta}_1 + \hat{\delta}_1 + \hat{\delta}_2 = -0.00030 \quad (0.00013)$	2.28
4	1976–2001	$\hat{\beta}_1 + \hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3 = -0.00170 \quad (0.00018)$	9.61
5	2002–2014	$\hat{\beta}_1 + \hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3 + \hat{\delta}_4 = -0.00064 \quad (0.00051)$	1.26

These slopes give the monthly change in temperature in each regime. Scaling up to decadal changes by multiplying by 120 gives 0.01°C, 0.15°C, 0.04°C, 0.20°C and 0.08°C respectively for the five regimes. It is seen, however, that the slope of the final regime is insignificantly different from zero and imposing this restriction does not alter the estimate of σ_a and barely alters the estimates of the slopes. Figure 1 shows the HADCRUT4 series with this restricted segmented trend imposed.

RSS

For this series standard identification techniques suggested that an ARIMA $(0, 1, 1)$ process is the most suitable model within this class, being estimated to be:

$$\nabla x_t = a_t - \underset{(0.044)}{0.410}a_{t-1} \qquad \hat{\sigma}_a = 0.1126$$

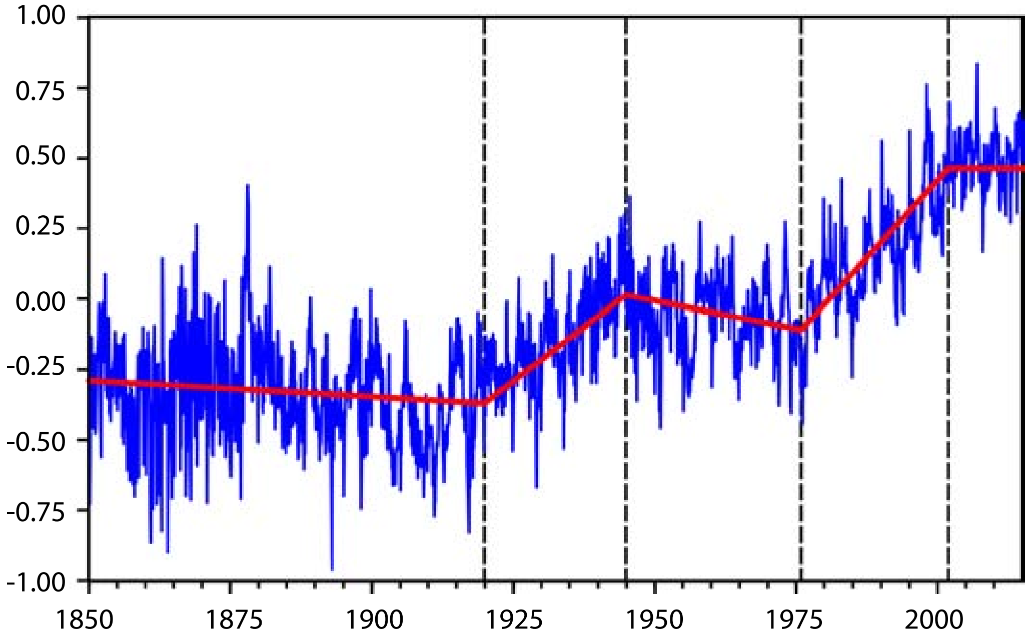

Figure 1: HADCRUT4 with segmented trend superimposed

Monthly data, January 1850–December 2014

This can be expressed as

$$x_t = x_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t = a_t - 0.410 a_{t-1}$$

As an example of a segmented trend, a model with $m = 2$ regimes with breaks at December 1997 and October 1998 was also fitted. With ε_t specified as an AR(2) process, this model is estimated to be:

$$x_t = \underset{(0.076)}{-0.126} + \underset{(0.00055)}{0.00095} \left(t - S_t(2)\right) + \underset{(0.083)}{0.138} \left(S_t(1) - S_t(2)\right) + \varepsilon_t$$

$$\varepsilon_t = \underset{(0.047)}{0.550} \varepsilon_{t-1} + \underset{(0.047)}{0.247} \varepsilon_{t-2} + a_t \qquad \hat{\sigma} = 0.1100$$

In this model the restriction $\beta + \delta_1 + \delta_2 = 0$ has been imposed, with a test of this restriction producing an insignificant statistic, since the unrestricted sum is −0.0001 with standard error 0.0006. Thus the evolving slopes of the trend function are estimated to be as shown in Table 2.

This trend is shown superimposed on the RSS series in Figure 2. An almost identical trend function is obtained by fitting a smooth transition with $\gamma = 0.7$ and the midpoint of the transition being April 1997.

Table 2: RSS regimes

Regime	Period	Estimated regime slope		t-ratio
1	Jan 1979–Dec 1997	$\hat{\beta}_1 = -0.00095$	(0.00055)	1.72
2	Jan 1998–Oct 1999	$\hat{\beta}_1 + \hat{\delta}_1 = -0.0148$	(0.0079)	1.88
3	Nov 1999–Dec 2014	$\hat{\beta}_1 + \hat{\delta}_1 + \hat{\delta}_2 = 0$		−0.11

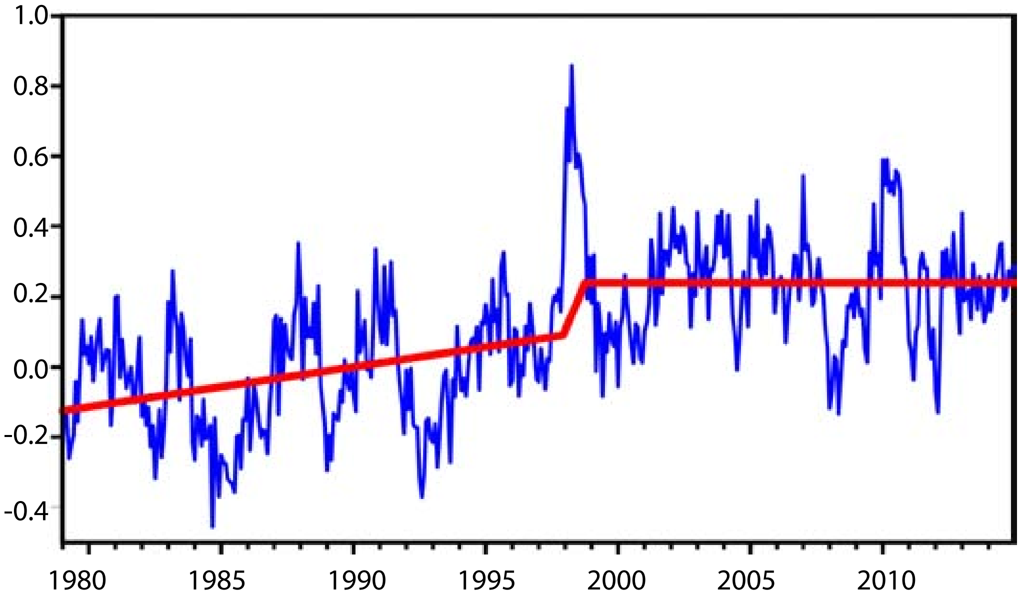

Figure 2: RSS with segmented trend superimposed

Monthly data, January 1979–December 2014

4 Seasonal extensions of the basic model

The temperature series investigated so far are both 'global' and hence contain no seasonal fluctuations. To deal with a regional temperature series, whose evolution will necessarily include a seasonal fluctuation, the level and noise components of equation (1) need to be extended. We will use s to denote the seasonal period: for temperatures recorded at quarterly intervals, $s = 4$, while for monthly data, $s = 12$. The model for μ_t may be extended to:

$$x_t = \sum_{i=1}^{s} (\mu_0 + \mu_1 t) d_i(t) + \text{ARIMA terms}$$

The $d_i(t)$ are seasonal 'dummy' variables defined as:

$$d_i(t) = \begin{cases} 1 & \text{if } t = \text{ interval } i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots, s$$

The level model thus allows for a different deterministic (linear) trend in each season. Extensions to nonlinear seasonal trends, or breaking and segmented seasonal trends, are clearly possible if thought desirable. Further details, including stochastic seasonal patterns and seasonal random walks, are discussed in the Appendix.

5 Fitting seasonal models to temperature series

Central England Temperature Record

The monthly CET has a clear seasonal pattern, as is shown in Figure 3. The following, rather simple, specification of the model of equations (A13) and (A14) is found to produce an adequate fit to the series:

$$x_t = \sum_{i=1}^{12} (\mu_0 + \mu_1 t) d_i(t) + \varepsilon_t \tag{4}$$

where the estimates of the seasonal trend components are shown in Table 3 and the noise component is given by the AR(2) process[10]

$$\varepsilon_t = \underset{(0.015)}{0.261} \varepsilon_{t-1} + \underset{(0.015)}{0.080} \varepsilon_{t-2} + a_t \qquad \hat{\sigma}_a = 1.3211$$

The seasonality in this model contains no random component and thus is completely deterministic, with the seasonal factors for each month remaining constant throughout the entire period. Each month does, however, evolve as a different linear trend. Table 3 expresses these trends at centennial rates: the smallest trend is seen to be for June, which has increased by an insignificant 0.001°C each century. The largest is for January, with a centennial increase of 0.04°C. These trend increases over the entire 350 or so years are illustrated in Figure 4, which shows the fitted monthly temperatures for 1659 and 2014: the January temperature has increased from 2.38°C to 4.11°C, i.e. by 1.73°C, while the June temperature has only increased from 14.31°C to 14.34°C, i.e. by just 0.03°C. It is clear that winters have become progressively warmer but that summers have remained much the same over the entire period, the seasonal increases being winter 1.35°C, spring 0.97°C, summer 0.34°C and autumn 1.11°C, with an overall average increase of 0.94°C. An alternative estimation yielded a driftless random walk with monthly dummies. This model thus differs from the ARIMA-plus-deterministic-seasonal-trends model by containing no trend component whatsoever. The estimated level is shown superimposed on the observed CET series in Figure 3: the mean level for 1659 is 9.02°C, that for 2014 is 10.41°C.

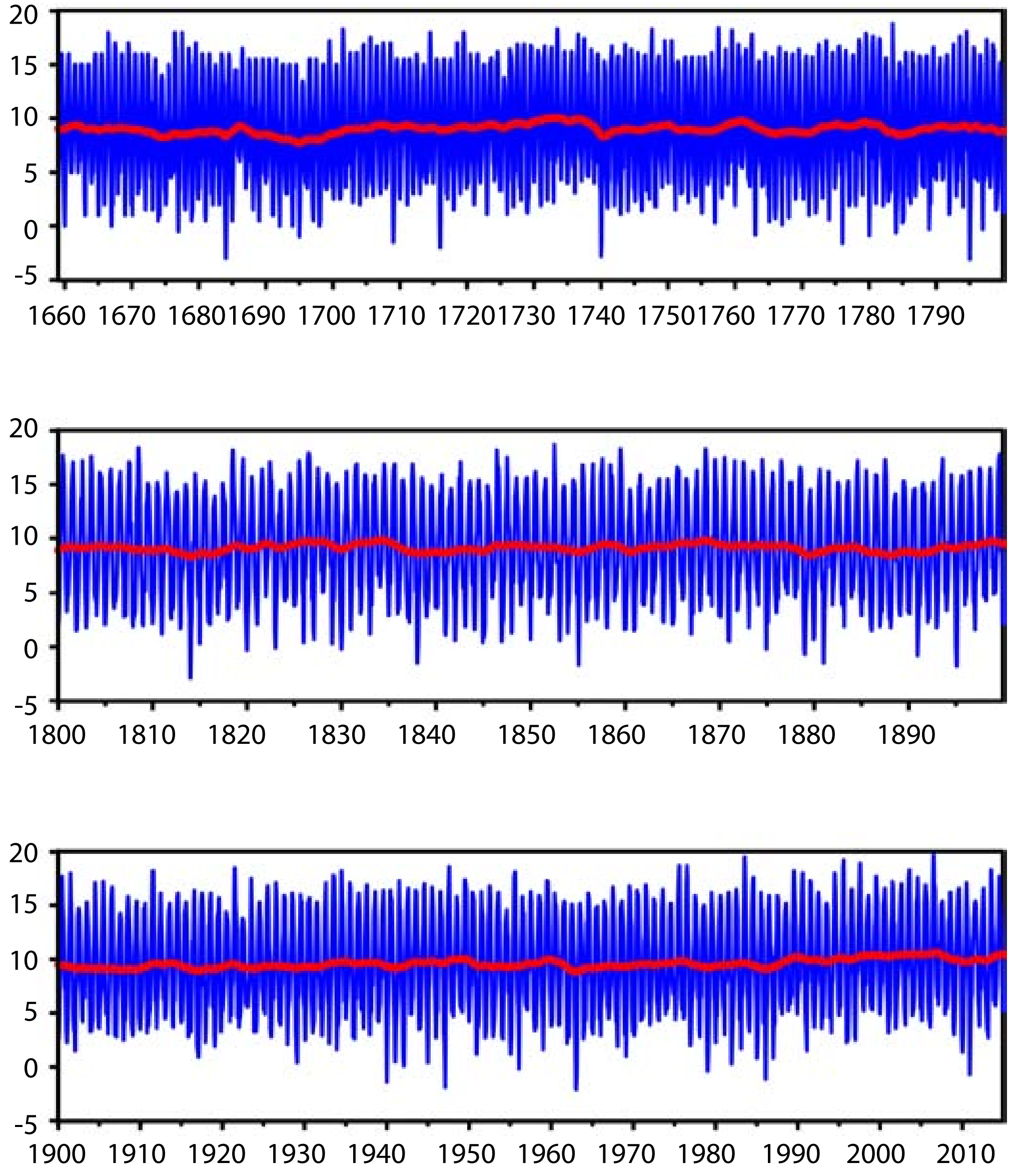

Figure 3: CET with estimated structural level superimposed

Monthly data in °C, January 1659–December 2014

Table 3: Estimates of equation (4) for the CET monthly series

i	$\hat{\alpha}_i$	$100 \times \hat{\beta}_i$
1	2.382	0.0406
2	3.338	0.0248
3	4.591	0.0346
4	7.468	0.0218
5	10.969	0.0117
6	14.307	0.0008
7	15.720	0.0118
8	15.388	0.0115
9	12.972	0.0172
10	9.042	0.0313
11	5.424	0.0299
12	3.447	0.0097
Standard error	0.147	0.0060

6 Forecasting from time series models

Having fitted a model of the type discussed above to an observed series, forecasts of future observations, along with measures of forecast uncertainty, may then be calculated. It is very important to understand, however, that the type of model that has been fitted will both determine the forecast and its associated uncertainty.

The fitted model expresses x_t as a function of past observations (x_{t-1},x_{t-2}, etc.), plus current and past random disturbances (a_t, a_{t-1}, a_{t-2}, etc.). The technique for generating forecasts involves setting the forecasts of the disturbances to their expected value of zero and then using the estimated model to generate x_{t+1} as a function of x_t, x_{t-1}, etc. Then x_{t+2} is generated as a function of the forecast x_{t+1} as well as x_t, x_{t-1} and so forth. This continues recursively as far into the future as we desire to forecast. Once the forecast horizon exceeds the dimension of the AR process the forecast will be entirely a function of earlier forecast values. For this reason we expect the uncertainty of the forecast to increase as the horizon extends into the future, though it does not keep growing unless the process is non-stationary.

If the process is stationary and there is no trend, the forecast of x_t will always converge on the sample mean μ_0. How quickly or slowly it converges will depend on the coefficients of the AR process. The mathematics involves computing what are called the 'roots of the characteristic equation' associated with the autogression. The higher the value of the largest root of the characteristic equation associated with the autoregression, the more slowly is the return to the mean.

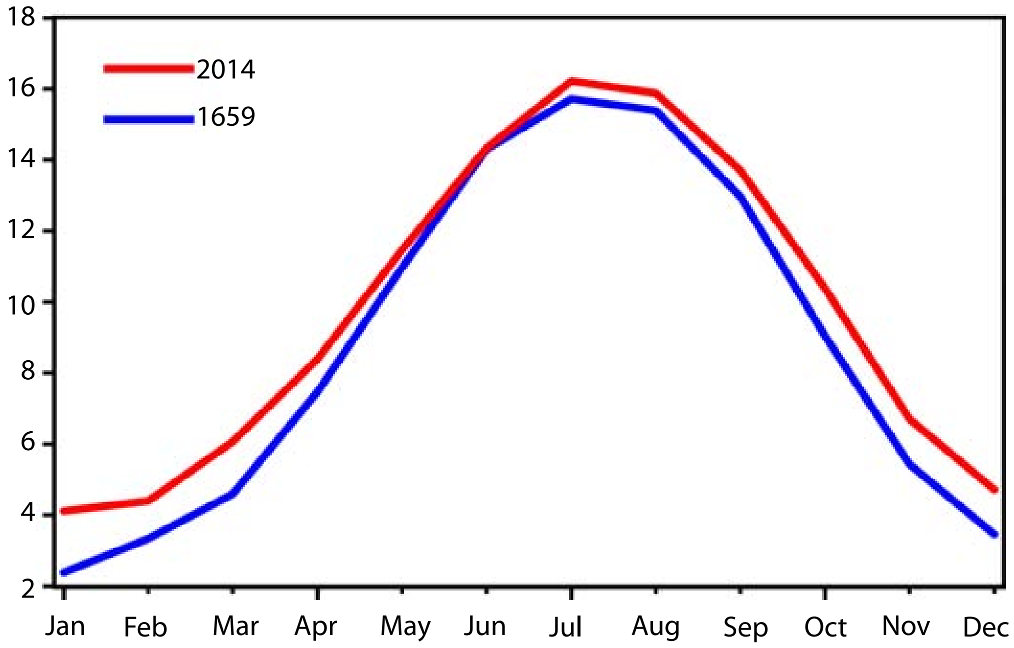

Figure 4: Fitted monthly CET temperatures for 1659 and 2014.

If the process is stationary around a linear trend, the same concept applies but this time the forecast converges to the trend line $\mu_0 + \mu_1 t$. Again, the roots of the characteristic equation of the AR process will determine how quick the convergence is.

If the process is a random walk, the optimal forecast of any future value is the last observed value of the series.

The variance of the forecast depends in a complex way on the coefficients of the ARIMA model. The mathematical details are in the Appendix. If the process is stationary the variance of the forecast converges to a finite maximum value. In the case of a nonstationary process, such as an I(1) or I(2), the forecast variance grows in an unlimited way as the horizon extends into the future.

7 Forecasting temperature series

HADCRUT4

Figure 5 shows the last four years of the HADCRUT4 series and forecasts out to end-2020, accompanied by 95% forecast intervals calculated as:

$$f_{T,h} \pm 1.96\sqrt{V(e_{T,h})}$$

for origin T = December 2014. The forecasts for January and February 2015 are 0.588 and 0.593, respectively, before falling to 0.582 for all longer horizons. The forecast intervals begin at (0.346, 0.830) and then increase in width to (−0.045,1.209) by end-2020, reflecting the more than doubling of the forecast error standard deviation from 0.123 to 0.320 over this period.

Figure 6 shows analogous forecasts from the segmented trend model with the last regime restricted to having a slope of zero, given that the estimated slope for this regime is insignificantly different from this value. Here the stationary autoregressive component has a short-term influence on the forecasts (the January 2015 forecast is 0.557), before dying away to allow the forecasts to reach and maintain a level of 0.463 by mid-2017, these being accompanied by forecast error standard deviations that reach a maximum of 0.157 at around the same time, so that the uncertainty in the forecasts is eventually bounded at 0.463 ± 0.157, in contrast to the ARIMA forecasts shown in Figure 5, in which the forecasts are both higher and much less precise.

Given that the within-sample fits of the two models are much the same, these examples effectively illustrate how alternative models can produce different forecasts having different levels of precision.

RSS

Figures 7 and 8 show forecasts out to December 2020 for RSS computed using the ARIMA and segmented trend models respectively. While the forecasts are rather similar, being 0.269 for all horizons for the ARIMA $(0, 1, 1)$ and converging to 0.237 by mid-2017 for the segmented trend, there are large differences in forecast uncertainty, even though the within-sample fits are again much the same. The non-stationary ARIMA model has a one-step ahead 95% forecast interval of (0.048, 0.490) , which then stretches to a very wide (−0.850, 1.388) by end-2020 as the forecast error standard deviation increases fivefold from 0.113 to 0.571. By contrast, the stationary deviations from the segmented trend model produce an almost identical one-step-ahead 95% interval of (0.048, 0.481), but which by end-2020 has widened only to (−0.097, 0.571) as the forecast error standard deviation increases from 0.110 to just 0.170.

CET

Forecasts from the seasonal model for CET are shown in Figure 9. The pronounced, but fixed, seasonal pattern is clearly observed, as is the absence of any pronounced trend in the overall level of CET: the forecasts for January and July 2016 (after the effects of the transitory AR(2) noise have decayed away) are 4.12°C and 16.22°C, while the analogous forecasts for 2020 are 4.14°C and 16.23°C. The forecast uncertainty also remains bounded, with the forecast error standard deviation increasing from 1.33 to

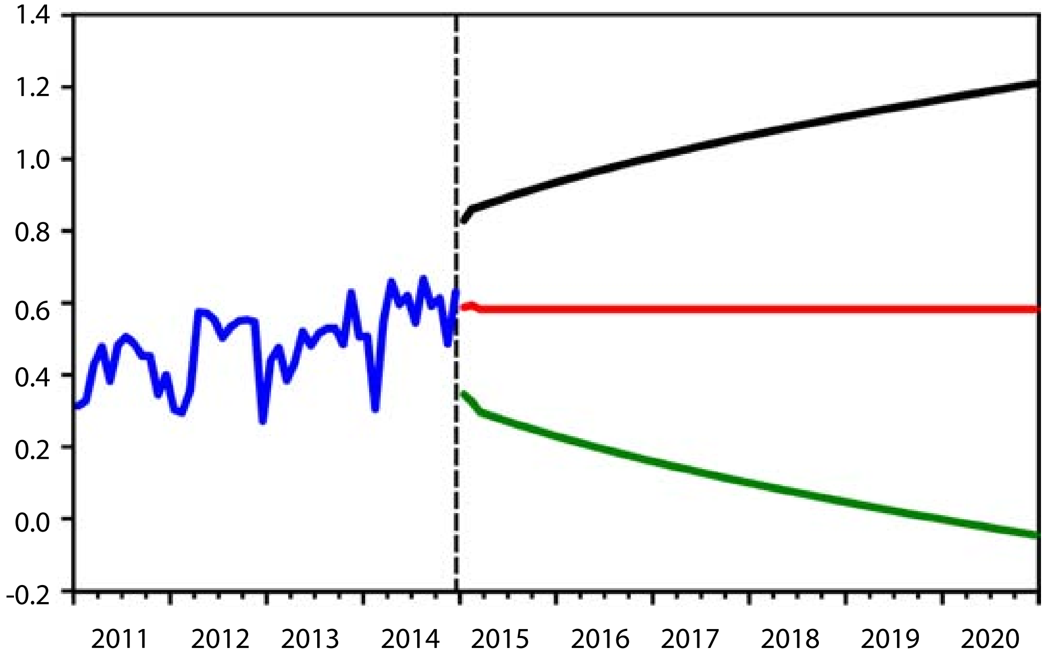

Figure 5: HADCRUT4 and forecasts from fitted ARIMA $(0, 1, 3)$ model

Monthly data, January 2011–December 2014 with forecasts out to December 2020 accompanied by 95% forecast intervals.

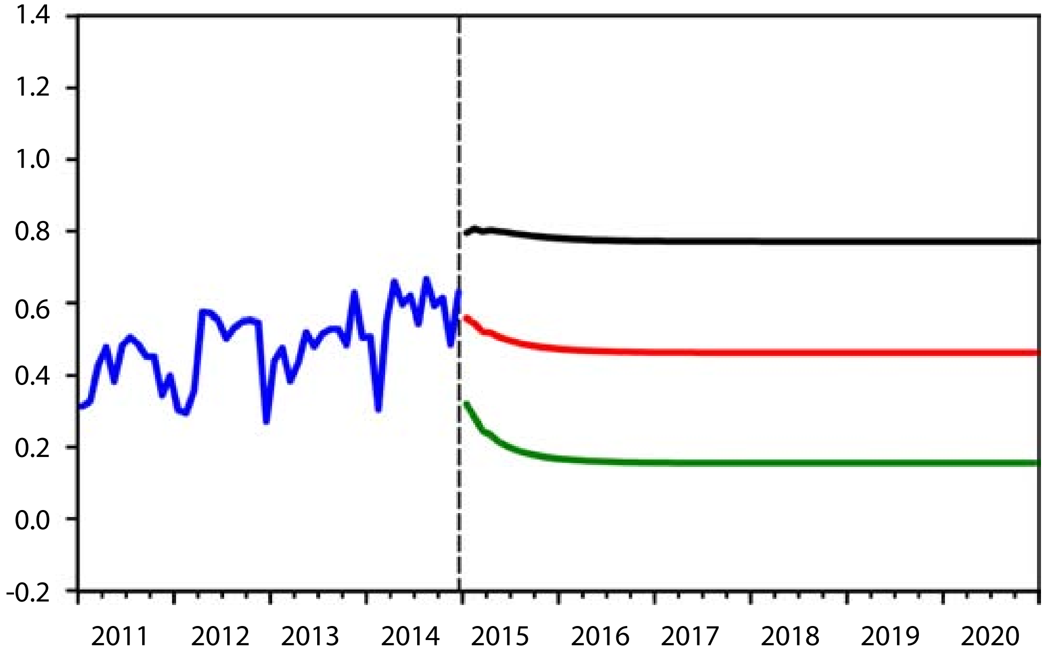

Figure 6: HADCRUT4 and forecasts from fitted segmented trend model

Monthly data, January 2011–December 2014 with forecasts out to December 2020 accompanied by 95% forecast intervals.

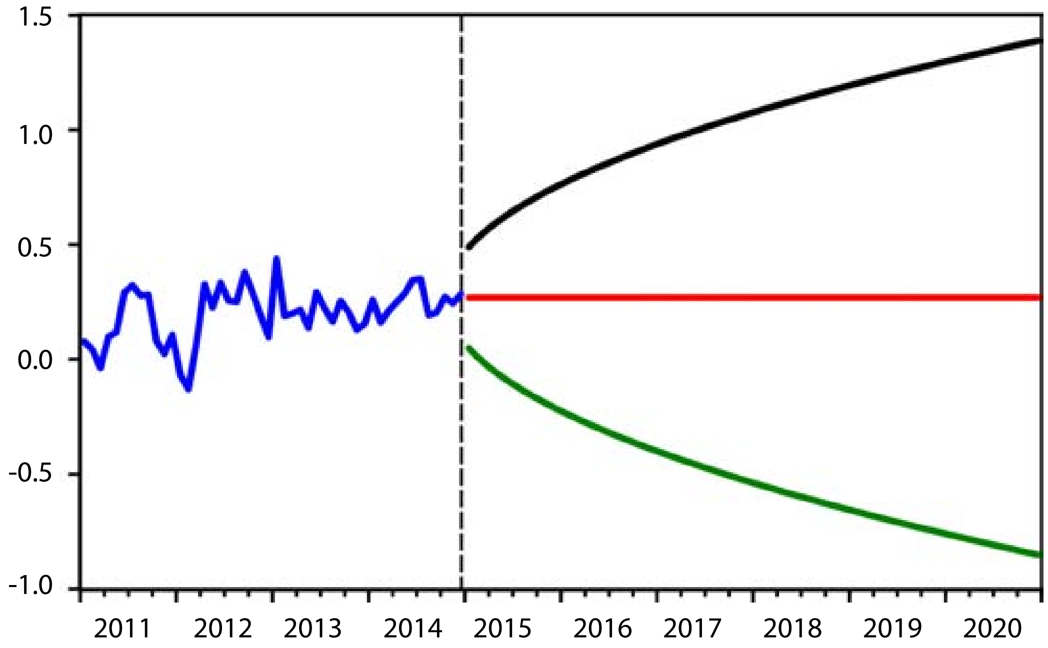

Figure 7: RSS and forecasts from fitted ARIMA $(0, 1, 1)$ model

Monthly data, January 2011–December 2014 with forecasts out to December 2020 accompanied by 95% forecast intervals.

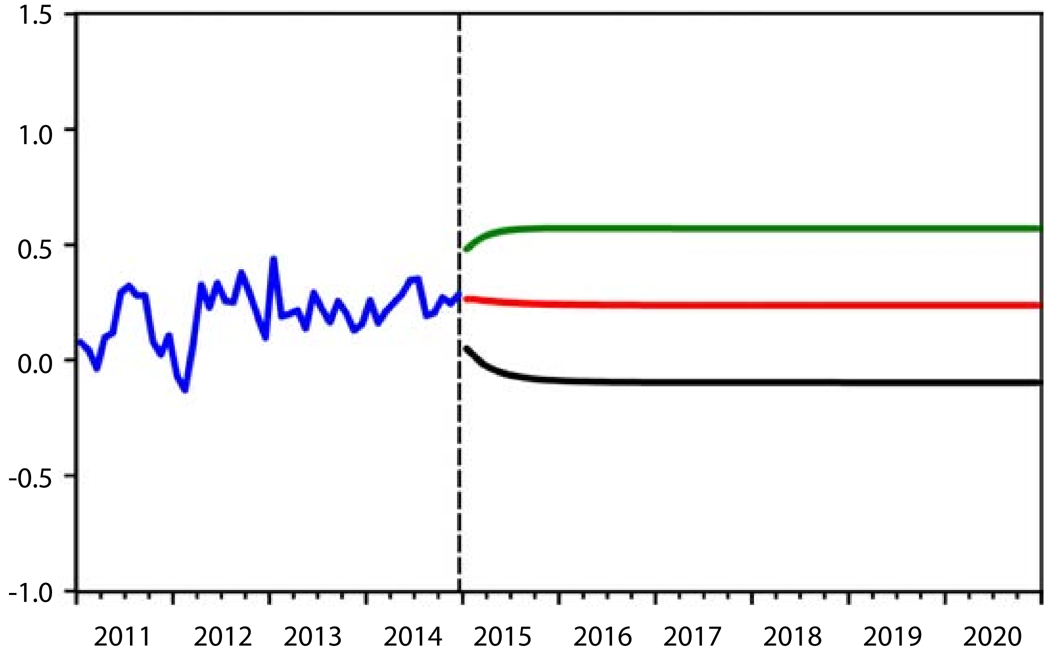

Figure 8: RSS and forecasts from fitted segmented trend model

Monthly data, January 2011–December 2014 with forecasts out to December 2020 accompanied by 95% forecast intervals.

just 1.39 over the forecast period. The structural model with fixed seasonal pattern produces almost identical forecasts and measures of forecast error uncertainty.

8 Discussion

The central aim of this report is to emphasise that, while statistical forecasting appears highly applicable to climate data, the choice of which stochastic model to fit to an observed time series largely determines the properties of forecasts of future observations and of measures of the associated forecast uncertainty, particularly as the forecast horizon increases. The importance of this result is emphasised when, as in the examples presented above, alternative well-specified models appear to fit the observed data equally well – the 'skinning the cat' phenomenon of modelling temperature time series.[11]

In terms of the series analysed throughout the paper, a clear finding presents itself for the two global temperature series. Irrespective of the model fitted, forecasts do not contain any trend, with long-horizon forecasts being flat, albeit with rather large measures of imprecision even from models in which uncertainty is bounded. This is a

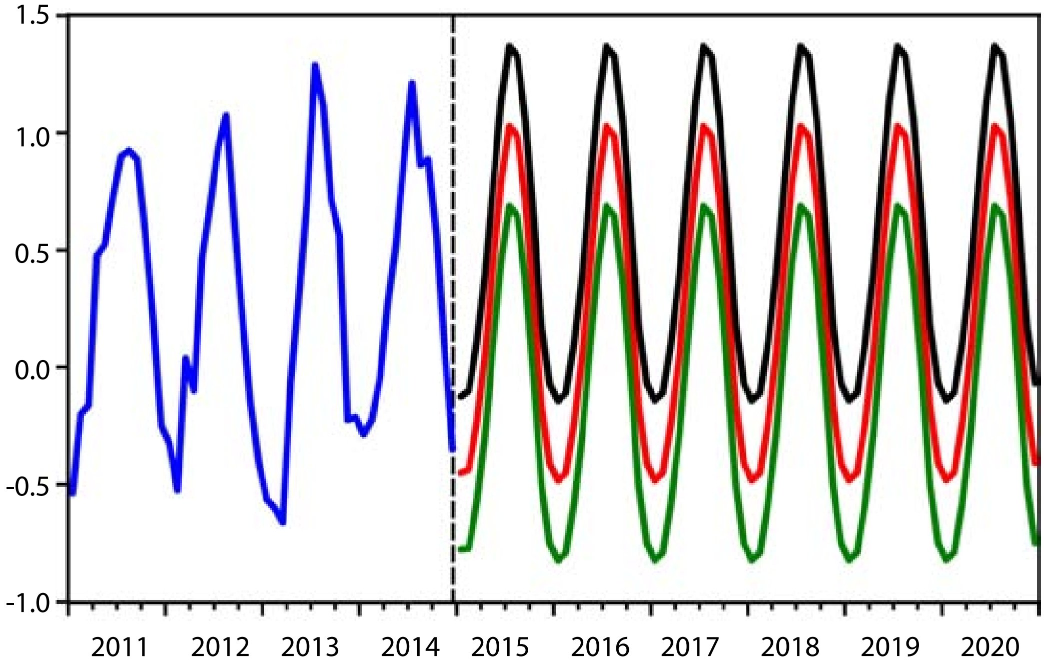

Figure 9: CET and forecasts

Monthly data, January 2011–December 2014. Forecasts per 'multiplicative ARIMA plus deterministic seasonal trends' model out to December 2020 accompanied by 95% forecast intervals.

consequence of two interacting features of the fitted models: the inability to isolate a significant drift or trend parameter and the large amount of overall noise in the observations themselves compared to the fitted 'signals'. Both of these features make forecasting global temperature series a necessarily uncertain exercise, but stochastic models are at least able to accurately measure such uncertainty.

The regional CET series does contain a modest warming signal, the extent of which has been shown to be dependent on the season: winters have tended to become warmer, spring and autumn less so, and summers have shown hardly any trend increase at all. The monthly pattern of temperatures through the year has remained stable throughout the entire 355 years of the CET record.

The models considered in the report also have the ability to be updated as new observations become available. At the time of writing, the HADCRUT4 observations for the first four months of 2015 were 0.690, 0.660, 0.680 and 0.655. Forecasts from the ARIMA $(0, 1, 3)$ model made at April 2015 are now 0.642 for May, 0.635 for June and 0.633 thereafter, up from the forecast of 0.582 made at December 2014. This uplift is a consequence of the forecasts for the first four months of 2015, these being 0.588, 0.593, 0.582 and 0.582, underestimating the actual outturns, although the latter are well inside the calculated forecast intervals.

What the analysis also demonstrates is that fitting a linear trend, say, to a preselected portion of a temperature record, a familiar ploy in the literature, cannot ever be justified.[12] At best such trends can only be descriptive exercises, but if the series is generated by a stochastic process then they are likely to be highly misleading, will have incorrect measures of uncertainty attached to them and will be completely useless for forecasting. There is simply no substitute for analysing the *entire* temperature record using a variety of well-specified models.

It may be thought that including 'predictor' variables in the stochastic models will improve both forecasts and forecast uncertainty. Long experience of forecasting nonstationary data in economics and finance tells us that this is by no means a given, even though a detailed theory of such forecasting is available.[13] Models in which 'forcing' variables have been included in this framework have been considered, with some success, when used to explain observed behaviour of temperatures.[14] Their use in forecasting, where forecasts of the forcing variables are also required, has been much less investigated, however: indeed, the difficulty in identifying stable relationships between temperatures and other forcing variables suggests that analogous problems to those found in economics and finance may well present themselves here as well.

9 Appendix: Technical details on ARIMA analysis

The lag operator

In concisely expressing several of the specific forms that equation (1) can take, we make use of the lag operator B, defined such that $Bx_t \equiv x_{t-1}$, so that $B^j x_t = x_{t-j}$ and $B^j c = c$, where c is a constant.

The ARIMA model

The simplest form of the autoregressive-integrated-moving average (ARIMA) process popularised by Box and Jenkins, and perhaps the most familiar of time series models, sets $\mu_t = \mu$, so that the level component is constant, and defines the noise to be the ARIMA (p,d,q) process:

$$\phi(B)\nabla^d \varepsilon_t = \theta(B)a_t \tag{A1}$$

Here the innovation a_t is a white noise process such that $E(a_t) = 0$, $E(a_t^2) = \sigma_a^2$ and $E(a_{t-i}a_{t-j}) = 0$ for all i and j, $i \neq j$, where the notation $E()$ denotes the theoretical mean (expectation) of the argument. Hence a_t has zero mean, constant variance σ_a^2 and zero autocovariances, so that it is uncorrelated with its past and, indeed, future values. ∇ is the first-difference operator defined as $\nabla = 1 - B$, so that, for example,

$$\nabla x_t = (1 - B)x_t = x_t - x_{t-1}$$

and

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$$

are polynomials in B of orders p and q, known as the autoregressive and moving-average operators respectively. Substituting (A1) into (1) gives

$$\phi(B)\nabla^d (x_t - \mu) = \theta(B)a_t \tag{A2}$$

so that the deviations of $x_t$from its level follow an ARIMA process. If $d > 0$ then these deviations are said to be stationary and, although x_t will be autocorrelated, it will always revert back to μ, which can then be regarded as the mean of x_t (an equivalent terminology is that the deviations have only a 'temporary' influence on x_t). If $d >$ 0 , typically 1 or possibly 2, then the deviations will be non-stationary and will not revert to a constant level and the concept of x_t having a mean is erroneous, for such deviations from any 'mean' will have permanent effects. For example, if $d = 1$ then (A2) becomes:

$$\phi(B)\nabla x_t = \phi(B)\nabla\mu + \theta(B)a_t = \theta(B)a_t$$

since $\nabla\mu = (1-B)\mu = \mu-\mu = 0$. Thus the first differences of x_t are stationary about zero while the actual values can be expressed as:

$$x_t = \frac{\theta(B)}{\phi(B)}\nabla^{-1}a_t = \psi(B)\left(1+B+B^2+\ldots+B^t\right)a_t = x_0 + \psi(B)\sum_{i=0}^{t-1}a_{t-i} \qquad \text{(A3)}$$

on putting $\psi(B) = \theta(B)/\phi(B)$, using the result that $(1-B)^{-1} = 1+B+B^2+\ldots$, and setting $a_0 = x_0$ and $a_i = 0$ for $i < 0$. x_t is thus given by its initial value x_0 plus the cumulated sum of (possibly correlated) innovations up to t. If $\phi(B) = \theta(B) = 1$ then $\nabla x_t = a_t$ and

$$x_t = x_{t-1} + a_t = x_0 + \sum_{i=0}^{t-1}a_{t-i}$$

which is the familiar representation of a (driftless) random walk. Similarly, for $d = 2$,

$$\phi(B)\nabla^2 x_t = \phi(B)\nabla^2\mu + \theta(B)a_t = \theta(B)a_t \qquad \text{(A4)}$$

since $\nabla^2\mu = \nabla\nabla\mu = 0$. Here the second-differences, $\nabla^2 x_t = (1-B)^2 x_t = x_t - 2x_{t-1} + x_{t-2}$, are stationary about zero and, by a similar argument to that used above, the first differences of x_t will be non-stationary.

Suppose now that, rather than being constant, the level follows the linear trend $\mu_t = \beta_0 + \beta_1 t$, so that (A2) becomes

$$\phi(B)\nabla^d(x_t - \beta_0 - \beta_1 t) = \theta(B)a_t \qquad \text{(A5)}$$

For $d = 0$ (A5) becomes

$$\phi(B)x_t = \alpha_0 + \alpha_1 t + \theta(B)a_t \qquad \text{(A6)}$$

where

$$\alpha_0 = \left(1 - \sum_{i=1}^{p}\phi_i\right)\beta_0 + \left(\sum_{i=1}^{p} i\phi_i\right)\beta_1$$
$$\alpha_1 = \left(1 - \sum_{i=1}^{p}\phi_i\right)\beta_1$$

so that x_t evolves as stationary deviations about a linear trend. With $d = 1$ (A5) now becomes

$$\phi(B)\nabla(x_t - \beta_0 - \beta_1 t) = \theta(B)a_t$$

Since $\nabla\beta_0 = 0$ and $\beta_1\nabla t = \beta_1$ this can be written as

$$\phi(B)\nabla x_t = \phi(B)\beta_1 + \theta(B)a_t = \alpha_1 + \theta(B)a_t \tag{A7}$$

where $\alpha_1 = (1-\phi_1-\ldots-\phi_p)\beta_1 = \phi(1)\beta_1$. In terms of levels we have $x_t = x_0+\beta_1 t+\varepsilon_t$, where $\phi(B)\nabla\varepsilon_t = \theta(B)a_t$, so that x_t itself evolves as *non-stationary* deviations about a linear trend emanating from the initial value x_0. Equivalently, analogous to (A3)

$$x_t = x_0 + \beta_1 t + \psi(B)\sum_{i=1}^{t-1} a_{t-i}$$

Thus x_t is given by the linear trend plus the cumulated sum of correlated innovations. For $d = 2$ (A5) becomes

$$\phi(B)\nabla^2(x_t - \beta_0 - \beta_1 t) = \theta(B)a_t \tag{A8}$$

Since $\nabla^2\beta_0 = \nabla\nabla\beta_0 = 0$ and

$$\beta_1\nabla^2 t = \beta_1(1-B)^2\, t = \beta_1\left(1-2B+B^2\right) t = \beta_1(t-2(t-1)+t-2) = 0$$

this can be written as

$$\nabla^2 x_t = \theta(B)a_t$$

which is again equation (A4). An equivalent form is $\nabla x_t = \beta_1 + \varepsilon_t$, where again $\phi(B)\nabla\varepsilon_t = \theta(B)a_t$, so that ∇x_t evolves as *non-stationary* deviations about an initial level given by β_1. The levels x_t will then evolve through, possibly extended, periods of increase and decrease but without following any overall trend.

Clearly for a warming trend to exist in this class of model the level function must be increasing, thus requiring a process of the form (A7) with $\alpha_1 > 0$ to generate temperatures.

Non-linear and breaking trend models

In these models the level is given by some deterministic, usually non-linear, function of t, $\mu_t = f(t)$, but they typically assume that the noise component ε_t is stationary, so that x_t evolves as stationary deviations about $f(t)$. The non-linearity is designed to capture a break or regime change and one function that has been used to capture such a shift in temperatures is the 'smooth transition', which employs the logistic function[15]

$$S_t(\gamma,\tau) = (1+\exp(-\gamma(t-\tau T)))^{-1}$$

to define the level, which we may more naturally now call the trend function, as

$$\mu_t = \alpha_1 + \beta_1 t + (\alpha_2 + \beta_2 t)S_t(\gamma,\tau) \tag{A9}$$

The logistic changes smoothly and monotonically as t increases, so the trend function smoothly transitions from the regime defined by $\alpha_1+\beta_1 t$ to the regime $\alpha_2+\beta_2 t$. The parameter τ determines the timing of the transition midpoint since, for $\gamma > 0$, $S_{-\infty}(\gamma,\tau) = 0$, $S_{\infty}(\gamma,\tau) = 1$ and $S_{\tau T}(\gamma,\tau) = 0.5$. The speed of the transition is determined by γ: if this parameter is small then $S_t(\gamma,\tau)$ will take a long time to traverse the interval from 0 to 1, the limiting case being $\gamma = 0$, when $S_t(\gamma,\tau) = 0.5$ for all t, so that

$$\mu_t = \alpha_1 + 0.5\alpha_2 + \left(\beta_1 + 0.5\beta_2\right) t$$

and there is just a single regime. For large values of γ, $S_t(\gamma,\tau)$ traverses the interval very rapidly, and as γ approaches $+\infty$ it changes instantaneously at time τT . If $\gamma < 0$ then the initial and final regimes are reversed but the interpretation of the parameters remains the same.

The smooth transition trend model has the appealing property that the midpoint of the transition can be estimated, but only two regimes are allowed for in (A9), although this may not be a problem as the transition can take some time, thus imparting 'smoothness' to the trend.

If more than two regimes are required but continuity of the trend function is still to be desired then a segmented linear trend may be considered. If there are m regimes defined by the break-points $T_1, T_2, \ldots, T_{m-1}$ then the segmented trend takes the form

$$\mu_1 = \alpha_1 + \beta_1 t + \sum_{i=2}^{m} \delta_i S_t(i) \tag{A10}$$

where

$$S_t(i) = \begin{cases} t - T_{i-1} & t > T_{i-1} \\ 0 & \text{otherwise} \end{cases}$$

Even though continuity is imposed, equation (A10) does not require a continuous first derivative, so that the slope of the trend function evolves as a sequence of discrete shifts β_1, $\beta_1+\delta_1$, $\beta_1+\delta_1+\delta_2$, etc., unlike the smooth transition. Extensions to allow for higher-order trend polynomials and, indeed, combinations of polynomials of different orders, are straightforward if algebraically more complicated to express.

Of course, continuity of the trend function can be dispensed with, in which case equation (A10) may be replaced by

$$\mu_t = \alpha_1 + \beta_1 t + \sum_{i=2}^{m} (\alpha_i + \beta_i t) D_t(i) \tag{A11}$$

where

$$D_t(i) = \begin{cases} 1 & T_{i-1} \le t < T_i \\ 0 & \text{otherwise} \end{cases}$$

Under the discontinuous segmented trend (A11), shocks to x_t given by non-zero values of ε_t are usually transitory, in that x_t will revert back to the trend line that it is currently on, but occasionally permanent, when the shock shifts μ_t immediately onto a new trend line.

Structural models

A structural model allows the level to be non-stationary, in its most general form taking the specification[16]

$$\mu_t = \mu_{t-1} + \beta_{t-1} + v_t \tag{A12}$$
$$\beta_t = \beta_{t-1} + w_t$$

Thus the level follows a random walk with a slope (or drift) that also follows a random walk. The errors v_t and w_t are independent zero-mean white noises with variances σ_v^2 and σ_w^2 respectively: if $\sigma_w^2 = 0$ then the slope is constant, β say, whereas if $\sigma_v^2 = 0$ changes to the level are entirely due to shifts in the slope. If both variances are zero then the level becomes $\mu_t = \mu_{t-1} + \beta = \mu_0 + t\beta$, a linear trend. Substituting (A12) into (A1) yields

$$\nabla^2 x_t = w_t + \nabla v_t + \nabla^2 \varepsilon_t$$

the right-hand side of which can be written as the moving average $a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$, albeit with some complicated restrictions imposed on θ_1 and θ_2: for example, for the smooth trend model which has $\sigma_v^2 = 0$, $\theta_1 = -4\theta_2/(1-\theta_2)$ and $-1 \le \theta_2 \le 0$. The general structural model can thus be thought of as a restricted ARIMA $(0, 2, 2)$ process.

Seasonal extensions

The temperature series investigated so far are both 'global' and hence contain no seasonal fluctuations. To deal with a regional temperature series, whose evolution will necessarily include a seasonal fluctuation, the level and noise components of equation (1) need to be extended. Equation (A11) may be extended to:

$$\mu_t = \alpha_1 + \beta_1 t + \sum_{i=2}^{s} (\alpha_i' + \beta_i' t) d_i(t) = \sum_{i=1}^{s} (\alpha_i + \beta_i t) d_i(t) \tag{A13}$$

Here s is the seasonal period: for temperatures recorded at quarterly intervals, $s = 4$, while for monthly data, $s = 12$. The $d_i(t)$ are seasonal 'dummy' variables defined as

$$d_i(t) = \begin{cases} 1 & \text{if } t = \text{interval } i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots, s$$

In equation (A13) $\alpha_i' = \alpha_i - \alpha_1$ and $\beta_i = \beta_i' - \beta_1$, $i = 2, \ldots, s$, so that the first expression gives the seasonal 'factors' as deviations from a reference value, taken for convenience as the first observation of the year, e.g. the first quarter (typically 'winter') if $s = 4$, January if $s = 12$. The level model thus allows for deterministic (linear) seasonal trends and extensions to higher-order polynomials and breaking and segmented seasonal trends are clearly possible if thought desirable.

A stochastic seasonal pattern can be introduced by extending the noise model (A1) to become

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D\varepsilon_t = \theta(B)\Theta(B^s)a_t \tag{A14}$$

Here $\nabla_s = 1 - B^s$ is the sth-difference operator and

$$\Phi(B^s) = 1 - \Phi_1 B^s - \ldots - \Phi_P B^{Ps}$$
$$\Theta(B^s) = 1 - \Theta_1 B^s - \ldots - \Theta_Q B^{Qs}$$

are polynomials of order P and Q respectively in B^s. Equation (A14) is the multiplicative *seasonal* ARIMA model, often denoted ARIMA $(p, d, q) \times (P, D, Q)_s$.[17]

The structural model can be extended to include a seasonal component such that $x_t = \mu_t + \psi_t + \varepsilon_t$, where the level component is again given by (A12) and where $\psi_t = \psi_{t-s} + \omega_t$, ω_t being a white noise with variance σ_ω^2.[18] The seasonal component thus follows a seasonal random walk and the seasonal pattern is allowed to change over time.

Forecasting ARIMA processes

To develop the properties of forecasts, suppose that we observe the set of observations $(x_{1-d}, x_{2-d}, \ldots, x_T)$ from a general ARIMA(p, d, q) process:

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t \tag{A15}$$

How do we forecast a future value x_{T+h}? If we let

$$\alpha(B) = \phi(B)\nabla^d = \left(1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_{p+d}B^{p+d}\right)$$

(A15) becomes, for time $T + h$,

$$\alpha(B)x_{T+h} = \theta_0 + \theta\,(B)\,a_{T+h}$$

or, when written out fully,

$$\begin{aligned}x_{T+h} = \alpha_1 x_{T+h-1} + \alpha_2 x_{T+h-2} + \ldots + \alpha_{p+d}x_{T+h-p-d} \\ + \theta_0 + a_{T+h} - \theta_1 a_{T+h-1} - \ldots - \theta_q a_{T+h-q}\end{aligned}$$

Clearly, observations from $T+1$ onwards will be unavailable, but a minimum mean square error (MMSE) forecast of x_{T+h} made at time T, which we shall denote $f_{T,h}$, is given by the conditional expectation

$$\begin{aligned} f_{T,h} = E\Big(&\alpha_1 x_{T+h-1} + \alpha_2 x_{T+h-2} + \ldots + \alpha_{p+d} x_{T+h-p-d} \\ &+ \theta_0 - a_{T+h} - \theta_1 + a_{T+h-1} - \ldots - \theta_q a_{T+h-q} | x_T, x_{T-1}, \ldots\Big) \end{aligned} \quad \text{(A16)}$$

Now

$$E\left(x_{T+j} | x_T, x_{T-1}, \ldots\right) = \begin{cases} x_{T+j}, & j \leq 0 \\ f_{T,j}, & j > 0 \end{cases}$$

and

$$E\left(a_{T+j} | x_T, x_{T-1}, \ldots\right) = \begin{cases} a_{T+j}, & j \leq 0 \\ 0, & j > 0 \end{cases}$$

so that, to evaluate $f_{T,h}$, all we need to do is:

1. replace past expectations ($j < 0$) by known values, x_{T+j} and a_{T+j}, and
2. replace future expectations ($j > 0$) by forecast values $f_{T,j}$ and 0.

Three examples will illustrate the procedure. Consider first the AR(4) model

$$\left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4\right) x_t = \theta_0 + a_t$$

so that $\alpha(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4\right)$. Here

$$x_{T+h} = \phi_1 x_{T+h-1} + \phi_2 x_{T+h-2} + \phi_3 x_{T+h-3} + \phi_4 x_{T+h-4} + \theta_0 + a_{T+h}$$

and hence, for $h = 1$, we have

$$f_{T,1} = \phi_1 x_T + \phi_2 x_{T-1} + \phi_3 x_{T-2} + \phi_4 x_{T-3} + \theta_0$$

For $h = 2, 3$ and 4, we have, respectively,

$$\begin{aligned} f_{T,2} &= \phi_1 f_{T,1} + \phi_2 x_T + \phi_3 x_{T-1} + \phi_4 x_{T-2} + \theta_0 \\ f_{T,3} &= \phi_1 f_{T,2} + \phi_2 f_{T,1} + \phi_3 x_T + \phi_4 x_{T-1} + \theta_0 \\ f_{T,4} &= \phi_1 f_{T,3} + \phi_2 f_{T,2} + \phi_3 f_{T,1} + \phi_4 x_T + \theta_0 \end{aligned}$$

and, for $h > 4$

$$f_{T,h} = \phi_1 f_{T,h-1} + \phi_2 f_{T,h-2} + \phi_3 f_{T,h-3} + \phi_4 f_{T,h-4} + \theta_0$$

As the forecast horizon $h \to \infty$, it may be shown that, since x_t is stationary

$$f_{T,h} \to \frac{\theta_0}{1-\phi_1-\phi_2-\phi_3-\phi_4} = E\left(x_t\right) = \mu$$

so that for long horizons the best forecast of a future observation is eventually the mean of the process, although the trajectory of the forecasts towards this limit will depend on the values taken by the autoregressive parameters.

Next consider the ARIMA $(0,1,3)$ model $\nabla x_t = \left(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3\right) a_t$. Here $\alpha(B)=(1-B)$ and so

$$x_{T+h} = x_{T+h-1} + a_{T+h} - \theta_1 a_{T+h-1} - \theta_2 a_{T+h-2} - \theta_3 a_{T+h-3}$$

For $h=1$ we have

$$f_{T,1} = x_T - \theta_1 a_T - \theta_2 a_{T-1} - \theta_3 a_{T-2}$$

and for $h=2,3$

$$\begin{aligned} f_{T,2} &= f_{T,1} - \theta_2 a_T - \theta_3 a_{T-1} \\ f_{T,3} &= f_{T,2} - \theta_3 a_T \end{aligned}$$

and for $h>3$

$$f_{T,h} = f_{T,h-1} = f_{T,3}$$

Thus, after two initial 'jumps', for all horizons $h > 2$, the forecasts from origin T will follow a straight line parallel to the time axis and passing through $f_{T,3}$. Clearly, if $\theta_1 = \theta_2 = \theta_3 = 0$, x_t follows a random walk and we have the well-known that result that $f_{T,h} = x_T$: the optimal forecast of all future values of a random walk is the current value.

Finally then, consider the ARIMA $(0,2,2)$ model $\nabla^2 x_t = \left(1-\theta_1 B-\theta_2 B^2\right) a_t$, with $\alpha(B) = (1-B)^2 = \left(1-2B+B^2\right)$:

$$x_{T+h} = 2x_{T+h-1} - x_{T+h-2} + a_{T+h} - \theta_1 a_{T+h-1} - \theta_2 a_{T+h-2}$$

For $h=1$ we have

$$f_{T,1} = 2x_T - x_{T-1} - \theta_1 a_T - \theta_2 a_{T-1}$$

for $h = 2$

$$f_{T,2} = 2f_{T,1} - x_T - \theta_1 a_T$$

for $h = 3$

$$f_{T,3} = 2f_{T,2} - f_{T,1}$$

and thus for $h \geq 3$

$$f_{T,h} = 2f_{T,h-1} - f_{T,h-2}$$

Hence, for all horizons, the forecasts from origin T will follow a straight line passing through the forecasts $f_{T,1}$ and $f_{T,2}$ and these will determine the slope of the line.

Forecast errors

The h-step ahead forecast error for origin T is

$$e_{T,h} = x_{T+h} - f_{T,h} = a_{T+h} + \psi_1 a_{T+h-1} + \ldots + \psi_{h-1} a_{T+1}$$

where $\psi_1, \ldots, \psi_{h-1}$ are the first $h-1$ ψ-weights in $\psi(B) = \alpha^{-1}(B)\theta(B)$. The forecast error is therefore a linear combination of the unobservable future shocks entering the system after time T and, in particular, the one-step ahead forecast error will be

$$e_{T,1} = x_{T+1} - f_{T,1} = a_{T+1}$$

Thus, for a MMSE forecast, the one-step ahead forecast errors must be uncorrelated. However, it may be shown that h-step ahead forecasts made at different origins will not be uncorrelated, and neither will be forecasts for different lead times made at the same origin. The variance of the forecast error $e_{T,h}$ is then

$$V\left(e_{T,h}\right) = \sigma_1^2 \left(1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2\right) \tag{A17}$$

To obtain the ψ-weights for the AR(4) model, we have to equate coefficients of powers of B in the expression $\alpha(B)\phi(B) = 1$, leading to

$$\begin{aligned}
\psi_1 &= \phi_1 \\
\psi_2 &= \psi_1\phi_1 + \phi_2 = \phi_1^2 + \phi_2 \\
\psi_3 &= \psi_2\phi_1 + \psi_1\phi_2 + \phi_3 = \phi_1^3 + 2\phi_1\phi_2 + \phi_3 \\
\psi_4 &= \phi_3\phi_1 + \psi_2\phi_2 + \psi_1\phi_3 + \phi_4
\end{aligned}$$

and, for $h > 4$,

$$\psi_h = \psi_{h-1}\phi_1 + \psi_{h-2}\phi_2 + \psi_{h-3}\phi_3 + \psi_{h-4}\phi_4$$

Since we are assuming stationarity, these 'ψ-weights' may be shown to get progressively smaller so that, consequently, $V\left(e_{T,h}\right)$ converges to a finite value, which is the variance of the process about the ultimate forecast μ. Thus, as the forecast horizon h becomes longer, forecasts from any stationary process converge to the mean of x_t, with the forecast error variance eventually being bounded by the actual variance of the series.

For the ARIMA $(0,1,3)$ model an analogous procedure obtains $\psi_1 = 1 - \theta_1, \psi_2 = 1 - \theta_1 - \theta_2$ and $\psi_j = 1 - \theta_1 - \theta_2 - \theta_3$ for $j > 2$. Thus we have

$$V\left(e_{T,h}\right) = \sigma_a^2\left(1 + \left(1-\theta_1\right)^2 + \left(1-\theta_1-\theta_2\right)^2 + (h-2)\left(1-\theta_1-\theta_2-\theta_3\right)^2\right)$$

which increases linearly with h and so cannot converge to a finite value. Consequently, the longer the forecast horizon h the greater the forecast error variance and the more imprecise forecasts necessarily become. Similarly, the ARIMA $(0,2,2)$ model has ψ-weights given by $\psi_j = 1 + \theta_2 + j(1 - \theta_1 - \theta_2)$, $j = 1, 2, \ldots$, and an h-step ahead forecast error variance of

$$\begin{aligned} V\left(e_{T,h}\right) = \sigma_a^2\Big(&1 + (h-1)\left(1+\theta_2\right)^2 + \tfrac{1}{6}h(h-1)(2h-1)\left(1-\theta_1-\theta_2\right)^2 \\ &+ h(h-1)\left(1+\theta_2\right)\left(1-\theta_1-\theta_2\right)\Big) \end{aligned}$$

which again increases with h but, since cubes of h are involved, potentially more rapidly than the ARIMA process with $d = 1$.

These examples thus show how the degree of differencing (equivalently, the order of integration) determines not only how successive forecasts are related to each other, but also the behaviour of the associated error variances.

Model selection

There are two generally-used methods of selecting an appropriate ARIMA process for modelling a time series. The first is the traditional three-stage approach of Box and Jenkins, that of identification, estimation and diagnostic checking. The initial identification stage requires the examination of the sample autocorrelation and partial autocorrelation functions for various differences of the series and then selecting a small group of models, possibly just the one, whose theoretical autocorrelation and partial autocorrelation functions most closely resemble those from the sample. These models are then estimated, typically by least squares, and then subjected to diagnostic checking to assess their adequacy, in the sense of whether the residuals satisfactorily ‘mimic’ the white noise assumption made for the innovations. The three stages may be iterated until a satisfactory model is obtained.

The second method is to fit a range of models and select that which minimises an information criterion, a statistic that trades off goodness of fit against model complexity, in the sense that models containing more estimated parameters are penalised more heavily.

In practice these two methods are often combined and this is the approach taken here, with identification leading to a, hopefully small, set of potential models. These are then estimated and the preferred model selected on the basis of both information criteria and adequacy of fit. Extensions to models other than the ARIMA class are straightforward.

10 Bibliography

Bai, J. (1997). 'Estimating multiple breaks one at a time'. *Econometric Theory* 13, 315–352.

Bai, J. and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes'. *Econometrica* 66, 47–78.

Bai, J. and Perron, P. (2003). 'Computation and analysis of multiple structural change models'. *Journal of Applied Econometrics* 18, 1–22.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Clements, M.P. and Hendry, D.F. (1999). *Forecasting Non-Stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Gay-Garcia, C., Estrada, F. and Sànchez A. (2009). 'Global and hemispheric temperatures revisited'. *Climatic Change* 94, 333–349.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A.C. and Todd, P.H.J. (1983). 'Forecasting economic time series with structural and Box-Jenkins models'. *Journal of Business and Economic Statistics* 1, 299–315.

Harvey, D.I. and Mills, T.C. (2001). 'Modelling global temperature trends using cointegration and smooth transitions'. *Statistical Modelling* 1, 143–159.

Harvey, D.I. and Mills, T.C. (2002). 'Unit roots and double smooth transitions'. *Journal of Applied Statistics* 29, 675–683.

Karl, T.R., Arguez, A., Huang, B., Lawrimore, J.H., McMahon, J.R., Menne, M.J., Peterson, T.C., Vose, R.S. and Zhang, H.-M. (2015). 'Possible artifacts of data biases in the recent global surface warming hiatus'. *Science* 4 June 2015.

Kaufmann, R.K., Kauppi, H. and Stock, J.H. (2010). 'Does temperature contain a stochastic trend? Evaluating conflicting statistical results'. *Climatic Change* 101, 395–495.

Lewandowsky, S., Oreskes, N., Risbey, J.S. and Newell, B.R. (2015). 'Seepage: Climate change denial and its effect on the scientific community'. *Global Environmental Change* 33, 1–13.

McKitrick, R.R. and Vogelsang, T. (2014). 'HAC-robust trend comparisons among climate series with possible level shifts'. *Environmetrics* DOI: 10.1002/env.2294

Mills, T.C. (2006). 'Modelling current trends in Northern Hemisphere temperatures'. *International Journal of Climatology* 26, 867–884.

Mills, T.C. (2007). 'Time series modelling of two millennia of northern Hemisphere temperatures: long memory or level shifts?'. *Journal of the Royal Statistical Society, Series A* 170, 83–94.

Mills, T.C. (2009a). 'Modelling current temperature trends'. *Journal of Data Science* 7, 89–97.

Mills, T.C. (2009b). 'How robust is the long-run relationship between temperatures and radiative forcing?'. *Climatic Change* 94, 351–361.

Mills, T.C. (2010a). 'Skinning a cat: stochastic models for assessing temperature trends'. *Climatic Change* 10, 415–426.

Mills, T.C. (2010b). 'Is global warming real? Analysis of structural time series models of global and hemispheric temperatures'. *Journal of Cosmology* 8, 1947–1954.

Mills, T.C. (2011). *The Foundations of Modern Time Series Analysis*. Basingstoke: Palgrave Macmillan.

Mills, T.C. (2012a). 'Box-Jenkins modelling of global temperatures'. *Journal of Statistics: Advances in Theory and Applications* 7, 49–65.

Mills, T.C. (2012b). 'Non-parametric modelling of temperature records'. *Journal of Applied Statistics* 39, 361–383.

Mills, T.C. (2013a). *A Very British Affair. Six Britons and the Development of Time Series Analysis*. Basingstoke: Palgrave Macmillan.

Mills, T.C. (2013b). 'Breaks and unit roots in global and hemispheric temperatures: an updated analysis'. *Climatic Change* 118, 745–755.

Mills, T.C. (2015a). (Editor) *Time Series Econometrics*. London: Routledge.

Mills, T.C. (2015b). *Time Series Econometrics: A Concise Introduction*. Basingstoke: Palgrave Macmillan.

Mills, T.C. and Mills, A.G. (1992). 'Modelling the seasonal pattern in U.K. macroeconomic time series'. *Journal of the Royal Statistical Society, Series A* 155, 61–75.

Mills, T.C., Tsay, R.S. and Young, P.C. (2011). 'Introduction to the Special Issue commemorating the 50th anniversary of the Kalman Filter and the 40th anniversary of Box & Jenkins'. *Journal of Forecasting* 30, 1–5.

Pierce, D.A. (1978). 'Seasonal adjustment when both deterministic and stochastic seasonality are present'. In A. Zellner (editor), *Seasonal Analysis of Economic Time Series*, 365-397. Washington, DC: US Department of Commerce, Bureau of the Census.

Stern D.I. and Kaufmann R.K. (2000). 'Detecting a global warming signal in hemispheric series: a structural time series analysis'. *Climatic Change* 47, 411–438.

Notes

1. See Mills (2011, 2013a) for the historical development of time series analysis and Mills (2015a) for a collection of the key early papers in the subject.

2. Box and Jenkins (1970): for an appreciation of the impact of this book on time series analysis, see Mills, Tsay and Young (2011).

3. Sir Arthur Schuster investigated periodicities in earthquake frequencies and sunspot activity, while Udny Yule and Gilbert Walker, after whom the conventional method of estimating autoregressions was named, used sunspot and air pressure data to illustrate their techniques: see Mills (2011, chapters 3 and 6).

4. See, for example, Gay-Garcia et al (2009), Kaufmann et al (2010), Mills (2006, 2007, 2009a, 2010a, 2010b, 2012a, 2012b, 2013b) and Stern and Kaufmann (2000).

5. The HADCRUT4 and CET series are both maintained by the UK Met Office's Hadley Centre for Climate Change (the former jointly with the Climatic Research Unit at the University of East Anglia) and are available from the Met Office website. The RSS series is available from the National Space Science & Technology Centre website at the University of Alabama at Huntsville.

6. Because the emphasis in this report is on forecasting in discrete time, continuous time frameworks, which are often used to construct theoretical models of the climate, are not considered.

7. Statisticians define the variance as the mean of the squared deviations about the mean; its square root is known as the standard deviation.

8. While the models fitted to the various temperature series are illustrative of the model classes available, they have been selected so that they are free of obvious misspecifications. Moreover they have all been fitted by commercially available software: apart from the structural models, which were estimated using the STAMP module of OXMETRICS 7, all other models were estimated using ECONOMETRIC VIEWS (EVIEWS) 8. Thus all the models may be readily replicated and no doubt improved upon. For an introductory treatment of ARIMA model identification see Mills (2015b).

9. The break-points were determined 'exogenously', in other words by visual examination of a plot of the series. This was done for two related reasons. First, methods for determining breaks endogenously remain in a relatively early stage of development (see Bai, 1997; Bai and Perron, 1998, 2003, McKitrick and Vogelsang 2014) and their properties in dynamic regression models have not been completely established. Second, these methods require observations to be 'trimmed' from the beginning and end of the sample to ensure that the tests have reasonable properties: any trimming at the end of the sample will make it almost impossible to find a break that occurs near the end of the sample, as may well have happened in this series, this being the well documented 'pause' or 'hiatus' in temperatures. Consequently, other researchers may wish to explore alternative break points: certainly bringing the last break point forwards from December 2001 will begin to produce a significant positive trend for the fifth regime.

10. This model can be slightly improved by incorporating an autoregressive conditionally heteroskedastic (ARCH), rather than white noise, innovation into the noise component. Little

change to the coefficient estimates is found, however, and so this additional complication is avoided in the presentation here.

11. As coined by Mills (2010a).

12. Two recent and particularly egregious examples are Karl et al (2015) and Lewandowsky et al. (2015).

13. See Clements and Hendry (1999).

14. See, for example, Mills (2009b).

15. See Harvey and Mills (2001, 2002).

16. See, for example, Harvey (1989).

17. This model was originally introduced by Box and Jenkins (1970). Methods of identifying models with the general structure of (A13) and (A14) and of testing for the presence of deterministic and stochastic seasonality are discussed in Pierce (1978) and Mills and Mills (1992).

18. See Harvey and Todd (1983).

GWPF TECHNICAL PAPERS

The Global Warming Policy Foundation is an all-party and non-party think tank and a registered educational charity which, while openminded on the contested science of global warming, is deeply concerned about the costs and other implications of many of the policies currently being advocated.

Our main focus is to analyse global warming policies and their economic and other implications. Our aim is to provide the most robust and reliable economic analysis and advice. Above all we seek to inform the media, politicians and the public, in a newsworthy way, on the subject in general and on the misinformation to which they are all too frequently being subjected at the present time.

The key to the success of the GWPF is the trust and credibility that we have earned in the eyes of a growing number of policy makers, journalists and the interested public. The GWPF is funded overwhelmingly by voluntary donations from a number of private individuals and charitable trusts. In order to make clear its complete independence, it does not accept gifts from either energy companies or anyone with a significant interest in an energy company.

Views expressed in the publications of the Global Warming Policy Foundation are those of the authors, not those of the GWPF, its trustees, its Academic Advisory Council members or its directors.

Published by the Global Warming Policy Foundation

For further information about GWPF or a print copy of this report, please contact:

The Global Warming Policy Foundation
55 Tufton Street, London, SW1P 3QL
T 0207 3406038 M 07553 361717
www.thegwpf.org

Registered in England, No 6962749
Registered with the Charity Commission, No 1131448