UiO : Department of Mathematics
University of Oslo

Analysing global temperature series in state space form

Ejike Obinna Okoye

Master's Thesis, Spring 2017

This master's thesis is submitted under the master's programme *Modelling and Data Analysis*, with programme option *Statistics and Data Analysis*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Contents

Preface

Time series analysis is important in modern day statistical research. The use of regression models to obtain underlying trends and complicated periodic patterns in time series is no longer sufficient. The decomposition of a time series into unobserved components leads to a classy and more efficient way of analysing time series data. The use of state space models otherwise called dynamical linear models to treat any linear model makes the ease of interpretation of the structural time series models, together with the associated information produced by the Kalman Filter and Smoother, a desired choice for handling time series data. Periodic seasonal patterns can be treated as exogenous variables in the observation equation. Presence of outliers and breaks are easily obtained by the use of intervention effect in the state space model. Forecasting a structural temperature data with state space model can be improved upon by fine tuning the Linear trend model.

Oslo, Norway

May 29, 2017

Ejike Obinna Okoye

Acknowledgement

This thesis is the result of my studies at the University of Oslo, Institute for Mathematics and natural science. This project could not be realized without the help of a number of people. First my appreciation goes to God Almighty for seeing me through the ups and downs of my study, a big thank you to my supervisor Anders Swensen for his support and guidance to writing a successful report and lastly to my family and friends for their love and prayers.

1 Introduction

Modeling of temperature data is a very interesting topic, one which this thesis will be focused on. Over the past years and even in recent times, scientist have continuously modeled temperature series as a trend, which many believe is of more interest than particular periodicities. Modern research is focused on the use of state space models to treat time series analysis. Structural time series models are (linear Gaussian) state-space models for (univariate) time series based on a decomposition of the series into a number of components. They are specified by a set of error variances, some of which may be zero.

In sight of this thesis, State space model will be used to analyse the global temperature series. This thesis will focus on reviewing the results in the paper by T.C Mills(2010) "skinning a cat: alternative models of representing temperature trends". Here, he compares the models used by Gay-Garcia et al(2009) which is about broken trend model with stationary disturbances and the work of Kaufmann et al(2010) which compares the in-sample forecast performance of the breaking trend model with that of an error correction model including a co-integrating relationship between temperature and radiative forcing. Mills starts off by compelling the use of structural time series models, where the observed time series (y_t) is decomposed into several unobserved components which include the trend (μ_t), seasonal (S_t), cycle (C_t), an Auotregressive process (ψ_t) and an irregularity(disturbance)(v_t),

$$y_t = \mu_t + S_t + C_t + \psi_t + v_t \tag{1.0.1}$$

citing the work of Harvey & Trimbur (2003) and Trimbur (2006) who proposed structural models that allow for more flexible stochastic trends and cycle. Inclusion of additional components such as seasonal and an auto regressive process is an extension to his previous work Mills(2004). In state space models which will be introduced in section 2, the observed series y_t is referred to as the observation vector and the unobserved components are referred to as the state vector x_t.

The seasonal term is taken to be deterministic where the variance of the noise term is zero i.e $\sigma_s^2 = 0$, which then evolves as a regular cycle with annual period. Hence the season is treated as a fixed component. An AR(1) process is specified in the structural model with noise term that is not correlated with the disturbances of the observed series and the other unobserved components. The Kalman filter algorithm which is available in the package STAMP is used to estimate the parameters and components , and also provide standard errors and confidence intervals for these estimated parameters.

The trend and cycle component in the structural model is treated by generating a weighting pattern. The introduction of weights to the observations is to make predictions and smoothing of the unobserved components at different points in time. According to the paper by Koopman & Harvey (2003), filtered estimator of the state vector based on information available at time $t-1$, can be written as

$$\hat{x}_{t|t-1} = \sum_{j=1}^{t-1} w_j(\hat{x}_{t|t-1}) y_j \tag{1.0.2}$$

where the weight vectors $w_j(\hat{x}_{t|t-1})$ can be computed by a backward recursion . While the smoothed estimates of the state vector $\hat{x}_{t|T}$ is given as a weighted sum of all observations,

$$\hat{x}_{t|T} = \sum_{j=1}^{T} w_j(\hat{x}_{t|T}) y_j \tag{1.0.3}$$

where the weight matrices $w_j(\hat{x}_{t|T})$ are calculated in the two opposite directions: (a) backwards in time (j = t âĹŠ 1; ...., 1); (b) forward in time (j = t, t + 1, ..., T).

The patterns of the weights, helps the user to understand what a model actually does, and it enables comparison to be made with other methods.

In further discussion by Mills, he fits a broken trend model of the monthly global temperature series from 1850 to 2007.which gave similar estimated results to that of Gay Garcia et al (2009). The result of this linear regression trend suggested a near non-stationarity in the fitted model. Another Observation by Mills with the fitting of the breaking trend model is that the value of the estimated slope, when used to calculate for temperature increase for a forecasted year of 2100, did not give a close result to that obtained by the coupled atmospheric-ocean general circulation models.

Mills later goes on to discuss a fitted structural model to the time series data, which included a driftless random walk trend with zero variance for the slope and the slope itself, a deterministic seasonal component with zero variance for the noise term hence a constant periodic pattern from year to year, a first order cycle and an AR(1) component with estimated $\hat{\theta}$ parameter. The result obtained for the trend component was not very smooth, which was expected for a driftless random walk model and by analysing the trend weight, Mills records a sharp decline in its pattern which he attributes to the random walk nature of the trend.

The result of Mills shows that a driftless random walk trend produces a constant forecast on the global temperature series, hence eliminating a possibility of future global warming/cooling

from the forecast which is in contradiction to the projection from the coupled atmospheric-ocean general circulation models and also from the models of Gay-Garcia et al (2009) and Kaufmann et al (2010). A move to increase the order of the trend component with no restrictions on the variances in other to obtain estimated slopes which will match the projections of the general circulation models(i.e increased temperature of about 3.55^0C for the forecast period) only resulted in a model with an inferior fit as compared to the structural model with the value of the slope as zero.

In conclusion, Mills states that the fitting of the structural model to the global temperature series provides a driftless trend component which does not give any increase in forecasted temperature and it will be impossible to obtain an increase in trend temperatures that is consistent with that projected by conventional coupled atmospheric-ocean general circulation models. And if one tries to get a consistent result, it require using a model which has poor statistical fit, simpler specifications and imposing a value on the slope parameter.

In any case the idea of this thesis is to review the results obtained by Mills, also to address the issue of treating a deterministic seasonal component in a state space model and the behaviour of the residual. Also further behaviour of the structural model is discussed with regards to correlated disturbances. More information about the global temperature series is obtained by close examination of the outliers and structural breaks. Forecasts on the monthly and annual temperature data will be carried out to cross examine the results obtained by Mills(2010).

To begin with, let us take a quick look at the global temperature data plot obtained from the Hadley Centre(data is available from the Hadley Centre; HADCRUT 4).

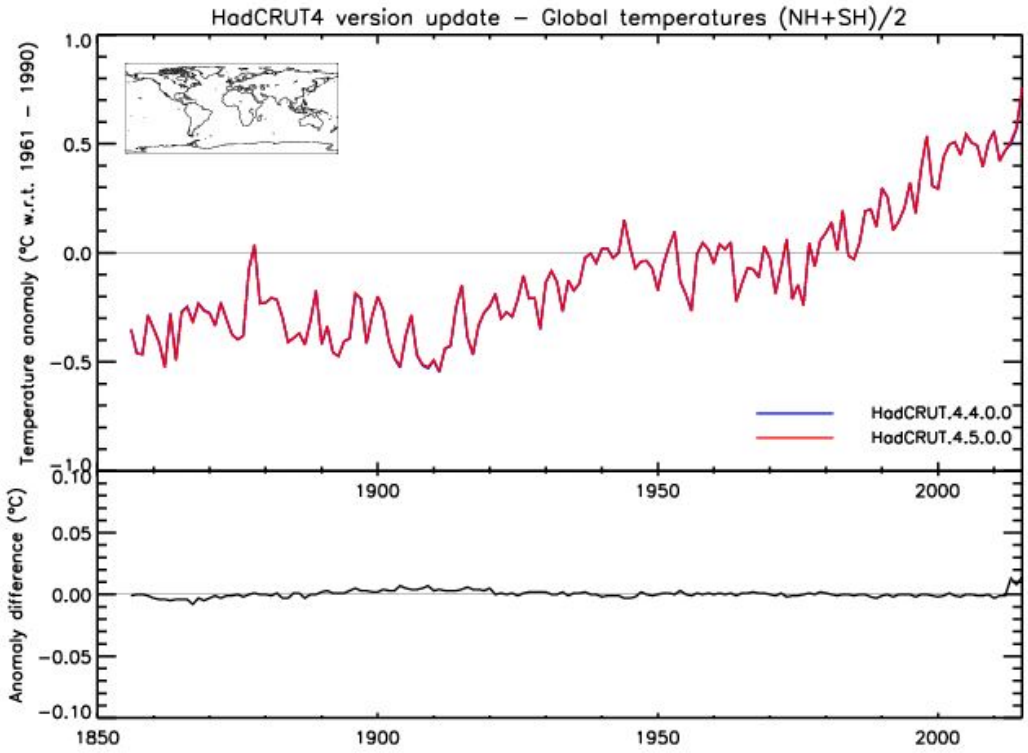

Figure 1.1: Annual temperature data from the Hadley centre

From the global temperature series record(1850-2016) shown in Figure (1.1), The temperature data obtained from the Hadley centre is taken from the HadCRUT4 dataset which is a combination of the CRUTEM4 land-surface air temperature and the HadSST3 sea-surface temperature (SST) Anomalies. The values are temperature anomalies, i.e. deviations from the corresponding 1961-1990 means. The time series we use is composed of the median global annual and monthly values.

So first off, a plot of the monthly and annual global temperature series respectively is shown below, that way we can see the underlying trend in our data and possible seasonal and cyclic pattern as well.

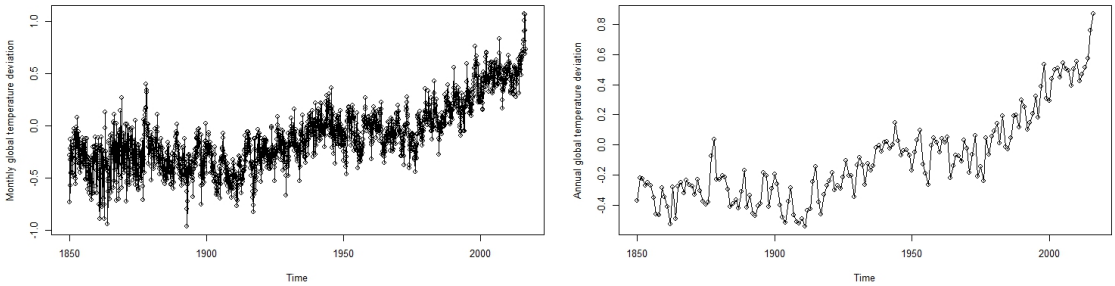

Figure 1.2: Comparison between the time series plot of the monthly global temperature data (left) and Annual global temperature data (right)

We note an upward trend in the series during the later part of the 1900's and a steady horizontal trend from period 1950 up until the latter part of the twentieth century where we see an upward trend again, which according to the book by Shumway & Stoffe (2011) has been used as an argument for the global warming hypothesis.

The issue of seasonality in a time series data is of high importance, as one would like to know if a series has constant seasonal pattern or if the seasonality varies over time. This will be discussed extensively. The deterministic annual pattern of seasonality is shown below to get an overview of the behaviour of our data. This is done by regressing seasonally centered averages over the monthly temperature series. By creating dummy variables of 1's and 0's for 11 states, and regressing these states over the temperature data, we obtain the figure below.

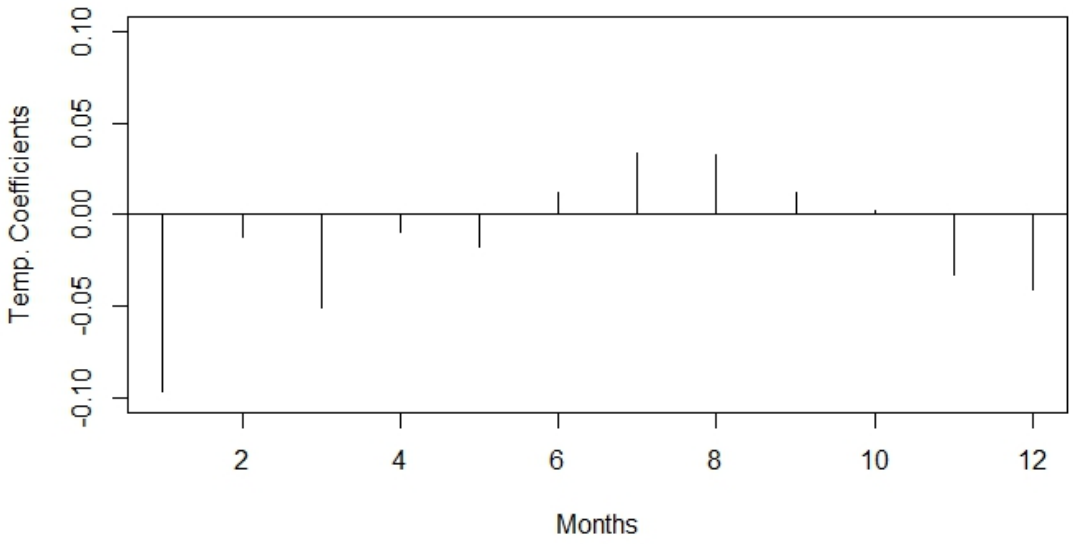

Figure 1.3: Annual seasonal pattern for monthly global temperatures

From the figure above, we can see the annual pattern of seasonality which is in line with what is expected. The first month is the intercept, hence the second month is actually the month of January and it goes on to the month of November.

In this thesis, my first task will be to replicate the result using developed procedures with the software R, discussed in the paper by T.C. Mills where he used the commercial software package STAMP. One approach to model a time series data is to treat it using a regression model in which the explanatory variables are functions of time and the coefficients are allowed to vary over time.

1.1 Fitting a broken trend by Linear regression

Now we take a look at the temperature series with fitted trend displayed below, with a break at the year 1977.

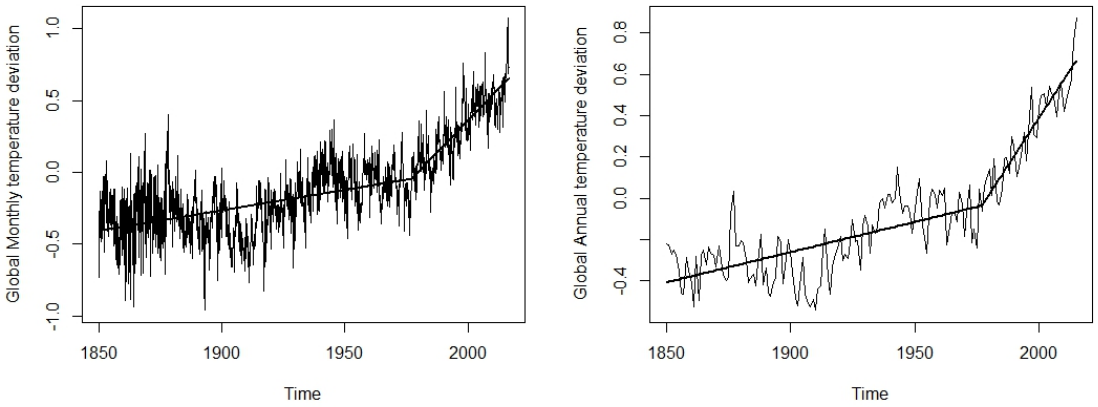

Figure 1.4: Monthly (left) and Annual(right) global temperature series 1850-2016 with fitted trend with break at 1977.

Within a regression framework, a trend would be modeled in terms of a constant and time together with a random disturbance, which is given as

$$x_t = \alpha + \beta t + w_t, \qquad t = 1, ..., T. \tag{1.1.1}$$

Where the model can be estimated using the ordinary least squares but with a disadvantage of a deterministic trend. This is explained by taking a practical case below. A simple linear regression has been used to estimate the trend by fitting the global temperature series.

$x_t = \mu_t + w_t$, t = 1850,..., 2016.

where μ_t is the fitted model which is obtained by regressing on

$t_1 = t \times I[t \leq 1977]$ and

$t_2 = t \times I[t > 1977]$

w_t is the residual. Table 1 below shows the fitted linear segmented trend x_t with break at 1977 for the monthly global temperature data.

Table 1.1: Fitted breaking trend model for monthly global temperatures 1850 - 2016

$x_t = -0.0833^{(0.007775)} + 0.003253^{(0.00011)} t_1 + 0.017539^{(0.0004584)} t_2 + residual(\tilde{x}_t)$
t_1 and t_2 are the time before and after the breaking point respectively.

The figure above(Figure 1.4) shows the monthly and annual global temperature series from 1850 to 2016, with the estimated trend μ_t, superimposed, which according to the description by T.C Mills(2010),can be seen to be characterised by a slow moving trend buried in relatively volatile seasonal and high frequency components. For the purpose of comparison, a fitted trend with break at year 1977 has been superimposed. A simple date for the break point was chosen since we are not really concerned for now about that. At a later stage of this thesis, structural breaks will be discussed for the temperature series. Also from Figure (1.4), an observed slope of 0.0029^oC before the break and 0.018^oC post-break for the monthly global temperature series, which is similar to those estimated by T.C Mills(2010) and Gay-Garcia et al(2009), using monthly and annual data respectively. It is also interesting to see that the annual temperature gives similar slopes before and after the breaks.

It is apparent that the estimated trend line obtained via simple linear regression does not quite capture the trend of the data.

It has been shown that classical regression is often insufficient for explaining all of the interesting dynamics of a time series. An example is the case of our temperature series, The autocorrelation function(ACF) of the residuals from the fitted monthly and annual global temperature data, reveals additional structure in the data that the regression did not capture, as can be seen in Figure (1.5) below.

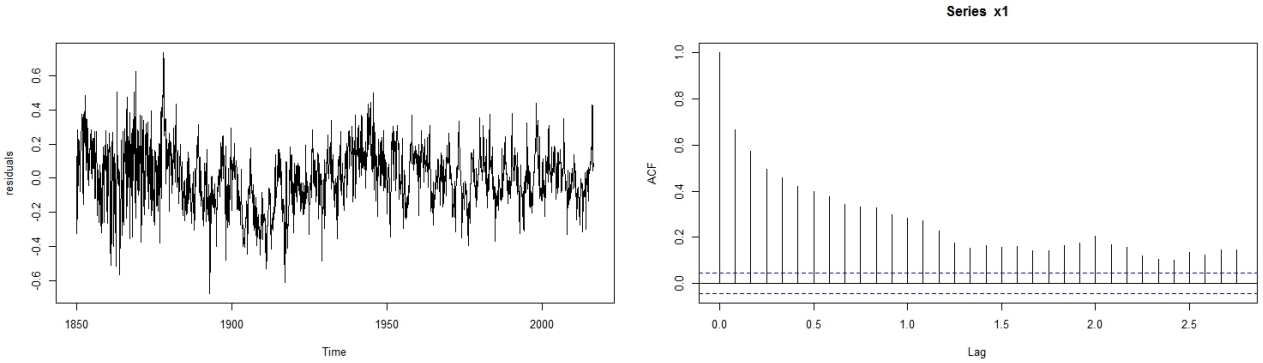

Figure 1.5: Residuals and autocorrelation function(ACF) of the residuals, of the fitted monthly global temperature series

Because this is a time series data, it is important to look at the autocorrelation function(ACF) of the residuals to see if there is any information not accounted for by the model. In the fitted monthly global temperature series, there is some remaining autocorrelation in the residuals. The autocorrelation function(ACF) of the monthly global temperature residual shows significant spikes between lag 0 and lag 1.

Also an observation from Figure 1.5, shows an increasing variance in the first half of the data than in the second half of the data. And, it appears from visual inspection that a trend is still present in the residual. For this, we will treat the residuals using the ARMA model.

Table 1.2: Fitted breaking trend model with ARMA residual

$x_t = -0.0833^{(0.007775)} + 0.003253^{(0.00011)}t_1 + 0.017539^{(0.0004584)}t_2 + \tilde{x}_t$
$\tilde{x}_t = 0.914^{(0.4488)}\tilde{x}_{t-1} + 0.140^{(0.6380)}\tilde{x}_{t-2} - 0.110^{(0.2471)}\tilde{x}_{t-3} + \tilde{w}_t - 0.443^{(0.4492)}\tilde{w}_{t-1} - 0.190\tilde{w}_{t-2} - 0.025\tilde{w}_{t-3}$
t_1 and t_2 are the time, before and after the breaking point respectively and the values in parenthesis are the standard error of the estimates.

Where Table 1.2 shows the fitted model with ARMA residual. The Sarima procedure using the astsa (applied statistical time series analysis) Package in R is used to obtain autoregressive and moving average components of the residual. Above we can see the fitted model which requires a complicated noise structure to model the high frequency components. The largest autoregressive root of the noise structure is 0.667 and the autocovariance function of the process depends only

on the time difference between s and t, (the lag h) and not on the actual time , which suggests stationarity in the residuals of the monthly global temperature series and persistence.

In Summary, the use of a regression model to analyse a time series is important but not a final approach. It raises the question for a better way of dealing with time series data. Diagnostic check on a model is very important. This investigation includes the analysis of the residuals as well as model comparisons. It can be seen from the fitted model that the check for breaks is also relevant since this is an important knowledge to know the exact periods when major events shaped the movement of the series. The residuals from a model fit will not quite have the properties of a white noise sequence. Hence this part of diagnostics can be viewed as a visual inspection of the residual autocorrelation with the main concern of detecting obvious departures from the independence assumption of a white noise series.

2 State space models

In other to account for the misleading information obtained from fitting using a linear regression model, we introduce a more general model that seems to comprise of a whole class of special cases of interest in much the same way as the linear regression performs, this is the state space model. The current (or filtered) estimate of the trend is estimated by putting the model in state space form and applying the Kalman filter. Predictions and smoothing is also carried out using related algorithms. Hence the best estimate of the trend is computed at all points in the sample using the full set of observations.

The state space model in its simple form is written as

$$x_t = \Phi x_{t-1} + w_t \tag{2.0.2}$$

Where x_t of an order one vector autoregression is referred to as the state equation, which is unobserved and determines the rule for the generation of the p x 1 vector x_t from the past p x 1 state x_{t-1}, for the time points $t = 1, ..., n$. But because x_t cannot be observed directly, but only a linear transformed version of it with added noise term which is called the observation y_t, hence the observed variables are presumed to be related to the state vector x_t through the observation equation of the system,

$$y_t = A_t x_t + v_t \tag{2.0.3}$$

where A_t is a q x p observation or measurement matrix and y_t is a $q \times 1$ vector matrix. The above equation(2.0.2) is referred to as the observation equation. The noise term w_t and v_t take the following assumption:

(i) w_t is p x 1 independent and identically distributed, zero mean normal vectors with covariance matrix Q.

(ii)The state process starts with a normal vector x_0 that has mean μ_0 and p x p covariance matrix Σ_0.

(iii)The additive observation noise v_t is white and Gaussian with qxq covariance matrix R.

(iv)For simplicity, x_0, w_t and v_t are uncorrelated. This is an assumption that helps in the explanation of early concepts.

A common way to handle the state equation is to model it as a trend plus random walk, where $\Phi = 1$.

The Kalman filter and smoother for a simple case of the state space model using monthly temperature data described in equation (2.0.2) and (2.0.3), is examined below. The initial state is normal, say, $x_0 \sim N(\mu_0, \Sigma_0)$, where μ_0 and Σ_0 are the initial mean and covariance respectively. We assume for simplicity, $\{w_t\}$ and $\{v_t\}$ are uncorrelated.

A plot of the smoothed and filtered series with the following assumed starting values values for the parameters , A = 1, Φ = 1, Q = 1 and R = 1 is shown below. This assumption is used for the purpose of understanding the concept.

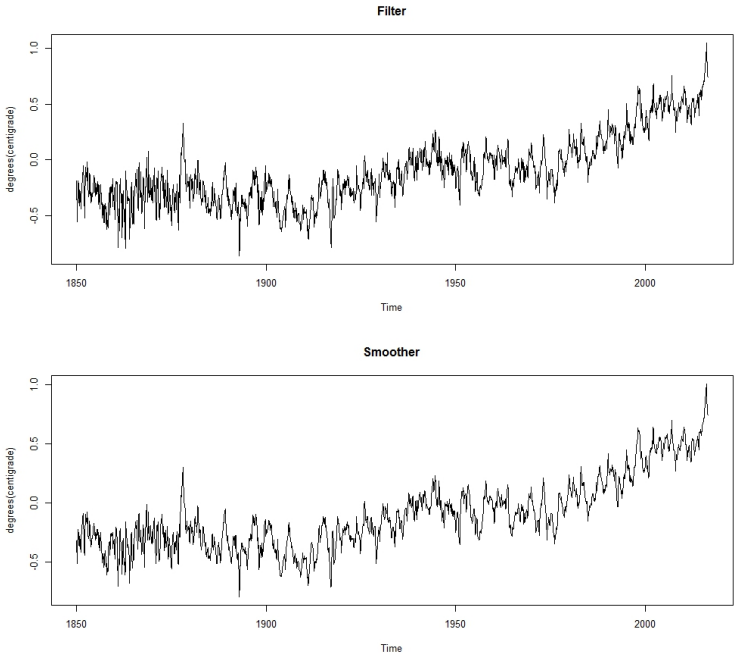

Figure 2.6: Kalman Filter(up) and Kalman Smoother(down), showing the trend of the monthly global temperatures series 1850 - 2016

The real advantages of the state-space formulation, however, does not really come through in the assumptions leading to Figure 2.6. Basic assumptions where used to achieve the results for the filter and smoothing. The fine tuning that can be achieved for various forms of the matrix A_t and the transition matrix Φ allow for a better filtering and smoothing of the trend. This will lead us to the discussion about structural models.

We also note that fixed inputs may enter into the states or observations, In this case, an r x 1 vector of inputs u_t is introduced into the model(1) and (2), thus the model can be written as

$$y_t = A_t x_t + \Gamma u_t + v_t \tag{2.0.4}$$

$$x_t = \Phi x_{t-1} + \Upsilon u_t + w_t \tag{2.0.5}$$

Where Γ and Υ are q x r and p x r vectors respectively.

The question of interest for the state space model relates to first predicting, filtering and smoothing of known parameters and afterwards a more interesting aspect will be estimating the unknown parameters $\Phi, Q, R, \Upsilon, \Gamma$ that define the particular model and, estimating the values of the underlying unobserved process x_t given the data $Y_s = y_1, ..., y_s$, to time s. The use of the Kalman filter and smoother gives the desired solutions to the problem mentioned above.

To begin with, a simple explanation as to how the Kalman filter and smoother works is given but for a more comprehensive study on this, I recommend taking a look at the book by Shumway & Stoffe(2011).

The Kalman filter $x_t^s = E(x_t \mid y_s),\ s = 1, ..., t$ specifies how to update the filter from x_{t-1}^{t-1} to x_t^t once a new observation y_t is obtained. The Kalman smoother on the other hand estimates x_t based on the the entire data sample $y_1, ..., y_n$, where $t \leq n$, $x_t^s = E(x_t \mid y_s),\ s = 1, ..., n$. Smoothing implies that each estimated value is a function of all the observation, whereas the filtered estimator depends on the present and past. Also P_t^s which is mean squared error also called the conditional error covariance and is expressed as

$$P_t^s = P_{t_1,t_2}^s = E\{(x_{t_1} - x_{t_1}^s)(x_{t_2} - x_{t_2}^s)^{'} | Y_s\} \qquad t_1 = t_2 = t$$

The Kalman filter property for the state space model specified in (2.0.3) and (2.0.4) with initial conditions $x_0^0 = \mu_0$ and $P_0^0 = \Sigma_0$, for $t = 1, ..., n$, can be written below as

$$\begin{aligned} x_t^{t-1} &= \Phi x_{t-1}^{t-1} + \Upsilon u_t \\ P_t^{t-1} &= \Phi P_{t-1}^{t-1} \Phi^{'} + Q \\ x_t^t &= x_t^{t-1} + P_t^{t-1} A_t^{'} (A_t P_t^{t-1} A_t^{'} + R)^{-1} (y_t - A_t x_t^{t-1} - \Gamma u_t) \end{aligned} \tag{2.0.6}$$

where $K_t = P_t^{t-1} A_t^{'} (A_t P_t^{t-1} A_t^{'} + R)^{-1}$ is called the Kalman gain. x_t^{t-1} and x_t^t are the predictor and filter state respectively. Important By products of the filter are the innovations(prediction

errors)

$$\epsilon_t = y_t - E(y_t \mid Y_{t-1}) = y_t - A_t x_t^{t-1} - \Gamma u_t \tag{2.0.7}$$

Where ϵ_t is referred to as the innovation(prediction error), and its variance covariance matrices is

$$\begin{aligned}\Sigma_t = var(\epsilon_t) &= var[y_t - A_t x_t^{t-1} - \Gamma u_t] \\ &= var[A_t(x_t - x_t^{t-1}) + v_t] = A_t P_t^{t-1} A_t^{'} + R\end{aligned} \tag{2.0.8}$$

for $t = 1, ..., n$. The proof of this property can be seen in the appendix.

Remark:

(i) The recursions x_t^{t-1} and P_t^{t-1} can be used for prediction when $t > n$, with x_n^n, P_n^n as starting values.

(ii) Time dependent parameters $\Phi_t, \Upsilon_t, \Gamma_t, Q_t, R_t$, can be allowed.

(iii) Simultaneous density of $x_0, x_1, ..., x_n,\ y_1, y_2, ..., y_n$ is

$f(x_0) = \prod_{t=1}^{n} f_w(x_t - \phi x_{t-1}) f_v(y_t - A_t x_t)$.

The Kalman smoother x_t^n property is given as for $t = n, n-1, ..., 1$

$$x_{t-1}^n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \tag{2.0.9}$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1}) \tag{2.0.10}$$

$$J_{t-1} = P_{t-1}^{t-1}\Phi^{'} + [P_{t-1}^{t-1}]^{-1} \tag{2.0.11}$$

The derivation for the smoother is given in the appendix, but the summary of the property of the Kalman smoother and the prediction error is given above. where is the predicting error.

How to choose the initial values

$$\tilde{x}_t = \phi \tilde{x}_{t-1} + \tilde{w}_t, \qquad t = 1, ..., n$$

- if x_t is stationary, the distribution of x_0 can be taken to be the stationary distribution.
- Condition on the initial observation by solving

  $y_1 = A\tilde{x}_1$

$\tilde{x}_1 = \phi x_0,$, if possible.

Structural time series models is one in which the trend, seasonal, cycle and error terms, plus other relevant components, are modelled explicitly.

2.1 Maximum likelihood estimation in state space model

Parameter estimation that specify the state space model (3) and (4) is very important in analysing various components of the time series model. The idea is that, $\Theta = \{\mu_0, \Sigma_0, \Phi, Q, R, \Upsilon, \Gamma\}$ used to represent the vector of unknown parameters containing the elements of the initial mean μ_0,the covariance Σ_0, the transition matrix Φ, the state and observation covariance matrices Q and R and the inputs Υ and Γ is estimated using the maximum likelihood estimation. The maximum likelihood, for a time series model where the observations $y_1, ..., y_n$ are not independent, is defined using a conditional probability density function to write the joint density function

$$L(y; \Theta) = \prod_{t=1}^{n} p(y_t | Y_{t-1}) \tag{2.1.1}$$

where $p(y_t|Y_{t-1})$ denotes the distribution of y_t conditional on the information set at time $t-1$, that is $Y_{t-1} = \{y_{t-1}, y_{t-2}, ..., y_1\}$. The maximum likelihood is used under the assumption that the initial state is normal $x_0 \sim N(\mu_0, \Sigma_0)$ and the errors $v_1, v_2, ...v_n$ and $w_1, w_2, ..., w_n$ are jointly normal and uncorrelated vector variables. $w_t \sim N(0, Q)$ and $v_t \sim N(0, \sigma_v^2)$

The likelihood is computed using the innovations

$\epsilon_t = y_t - A_t x_t^{t-1} - \Gamma u_t,$

which are independent normal where $E(\varepsilon_t) = 0$ and the covariance matrix

$\Sigma_t = A_t P_t^{t-1} A_t^{'} + R$

The log likelihood $-\log L(y; \Theta)$ is proportional to

$$\frac{1}{2}\sum_{t=1}^{n} \log \mid \Sigma_t(\Theta) \mid + \frac{1}{2}\sum_{t=1}^{n} \epsilon_t(\Theta)^{'} \Sigma_t(\Theta)^{-1} \epsilon_t(\Theta) \tag{2.1.2}$$

where the dependence of the innovations on the parameter Θ has been emphasized. Using the Kalman filter for given Θ, this can be calculated.

Procedure

- select initial and starting values for the parameters (Θ^0)

- For Θ^0, compute the likelihood $L(y;\Theta^0)$ using the Kalman filter(kf)
- Apply a numerical optimization algorithm to $L(y;\Theta^0)$
- Repeat this process for n steps until the value of Θ corresponding to the maximum Likelihood is found.

2.2 Structural models with Annual Temperature Series

An introduction to structural models can be better understood by reading the book by Harvey(1989), ... The idea of using structural models is very essential in the sense that it gives one the opportunity to better analyse time series data combining various components, i.e Trend, season, cycles, auto regressiveness, e.t.c. Structural models can be extended into state space by redefining the observation equation to include all the components one needs to model.

2.2.1 Trend

The question of defining a trend is one which has troubled the minds of statisticians and economists for many years. A trend is conceived of as that part of a series which changes relatively slowly over time. In other words, the properties of smoothing plays a big role in the definition. A trend, viewed in terms of prediction,is that part of the series which when extrapolated gives the clearest indication of the future long term movements in the series. A simple case of the structural model is one in which the underlying level of the series changes over time. This level is defined by a random walk which includes a random disturbance term(white noise). This can be written as

$$\begin{aligned} y_t &= \mu_t + v_t \qquad v_t \sim NID(0,\sigma_v^2), \qquad t = 1,...,T \\ \mu_t &= \mu_{t-1} + \omega_t \qquad \omega_t \sim NID(0,\sigma_\omega^2) \end{aligned} \tag{2.2.1}$$

where μ_t is the local level and v_t is a white noise disturbance term. NID denotes normally and independently distributed, and two disturbances are mutually uncorrelated. This leads us the Local linear trend model which replaces the deterministic trend in (1.0.1) by a stochastic trend.

It can also be modeled as a random walk with slope β_t.

$$\begin{aligned} y_t &= \mu_t + v_t & v_t &\sim NID(0, \sigma_v^2) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_t & \omega_t &\sim NID(0, \sigma_\omega^2) \\ \beta_t &= \beta_{t-1} + \eta_t & \eta_t &\sim NID(0, \sigma_\eta^2) \end{aligned} \tag{2.2.2}$$

where β_t is the slope term, η_t is it's noise term. The level and slope disturbances are mutually uncorrelated and uncorrelated with v_t. When $\sigma_\eta^2 = 0$, we have a case of a trend as a random walk plus drift. To express this model in state space form, equation (2.2.2) is written as

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + v_t$$

And the state equation written as

$$x_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \omega_t \\ \eta_t \end{pmatrix}$$

Where $R = var(v_t) = r_{11} = \sigma_v^2$ and

$$Q = \begin{pmatrix} \sigma_\omega^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}$$

Where A_t is a 1 x 2 matrix, Φ_t is 2 x 2 matrix and the state noise terms are uncorrelated with mean zero and a 2 x 2 covariance matrix Q.

The focus will be on estimating the variance of the white noise disturbance term in the observation equations $R = var(v_t)$ and state equation Q respectively.

Using the Kalman filter property, the Kfilter0 and Ksmooth0 script in the astsa package in R, uses the cholesky decomposition of Q and R. They are denoted cQ and cR. Practically, the scripts only require that Q or R may be reconstructed as $t(cQ)\% * \%(cQ)$ or $t(cR)\% * \%(cR)$ respectively. Hence for the sake of clarity, the state covariance matrix can written as

$$cQ = \begin{pmatrix} q11 & 0 \\ 0 & q22 \end{pmatrix}$$

Parameters to be estimated are the standard deviation of the error terms; $\sigma_v = \sqrt{r_{11}}$, $\sigma_\omega =$

$\sqrt{q_{11}}$ and $\sigma_\eta = \sqrt{q_{22}}$. The initial assumptions: the initial mean $\mu_0 = -0.35$, the covariance Σ_0 =diag$(0.01, 2)$, the initial measurement and state covariance values was started at $\sigma_v = 0.01$ and $\sigma_\omega = 0.01$, $\sigma_\eta = 0.01$ respectively.

After about 10 iterations of a Newton - Raphson, the observation uncertainty together with it's standard error is $\sigma_v = 0.08058_{0.00902}$, compared with the state uncertainties together with their standard error $\hat{\sigma}_\omega = 0.04628_{0.01369}$ and $\hat{\sigma}_\eta = 0.00119_{0.00203}$. It can be seen here that the slope is insignificant to the model. The Figure below shows the smoothed trend estimate, this uses the Kfilter0 and Ksmooth0 scripts of the astsa package in R, developed by Shumway & Stoffe (2011)

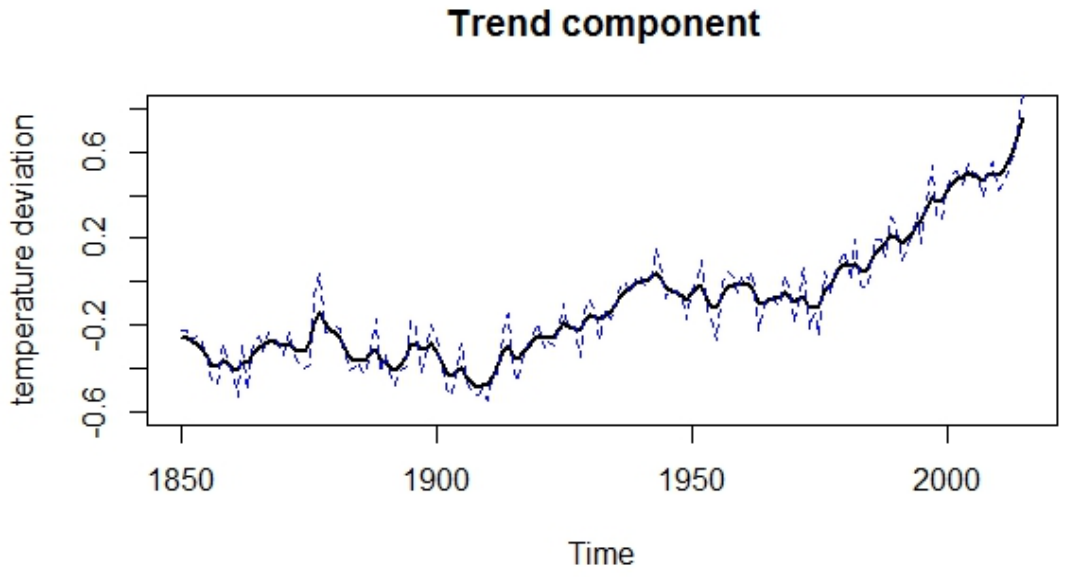

Figure 2.7: Smoothed trend estimate

Where the black line is the smoothed trend and the blue dotted line is the initial observation.

Next will be to take a look at a more inclusive structural model, that is considering the time series to be the sum of the trend component, a cycle component and a white noise.

2.2.2 Checking for periodicity in the residuals

The notion that a time series exhibits repetitive behaviour over time is important in the field of statistics. The modeling of the global temperature data as a trend plus irregular (noise term)

is obviously not sufficient for analysis. Since we know that for temperature data, it is likely for periodicity to occur, thus has led to the investigation of the residuals obtained from fitting the model. This investigation is carried out by the use of the spectral analysis. A fundamental objective of spectral analysis is to identify the dominant frequencies in a series. Therefore the idea that a time series is composed of periodic components, appearing in proportion to their underlying variances gives rise to the study of spectral density. If the residual is an autoregressive moving average process(ARMA), its spectral density can be obtained using the fact that it is a linear process, i.e., $res_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

The spectral density of a general ARMA(p,q) process, $\phi(D)x_t = \theta(D)w_t$ is given by

$$f_x(\omega) = \sigma_w^2 \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2} \tag{2.2.3}$$

where $\phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k$ and $\theta(z) = 1 - \sum_{k=1}^{q} \theta_k z^k$.

Since the main idea here is to check for periodicity in the residuals, a parametric spectral estimator is obtained by fitting an autoregressive AR(p)to the residual obtained from two cases($y_t - Ax_t^n$ and $y_t - Ax_t^{t-1}$) (a) fitting the smoothed trend to the observation (b) fitting the filtered value to the observation, where the order p is determined by one of the model selection criteria, such as, AIC, AICc and BIC by choosing the value of p at the lowest AIC and BIC.

If $\hat{\phi_1}, \hat{\phi_2}, ..., \hat{\phi_p}$ and $\hat{\sigma}_w^2$ are the estimates from an AR(p) fit to the res_t, a parametric spectral estimate of $f_{res}(w)$ is obtained by substituting these estimates into equation(7), that is,

$$\hat{f}_{res}(\omega) = \frac{\hat{\sigma}_w^2}{|\hat{\phi}(e^{-2\pi i\omega})|^2}, \tag{2.2.4}$$

where

$$\hat{\phi}(z) = 1 - \hat{\phi_1} z - \hat{\phi_2} z^2 - ... - \hat{\phi_p} z^p \tag{2.2.5}$$

ω is the frequency with condition that $0 \leq \omega \leq \frac{1}{2}$. The use of the arma.spec script in R is used to obtain the plot of the spectrum $\hat{f}_{res}(\omega)$ against its frequency(ω).

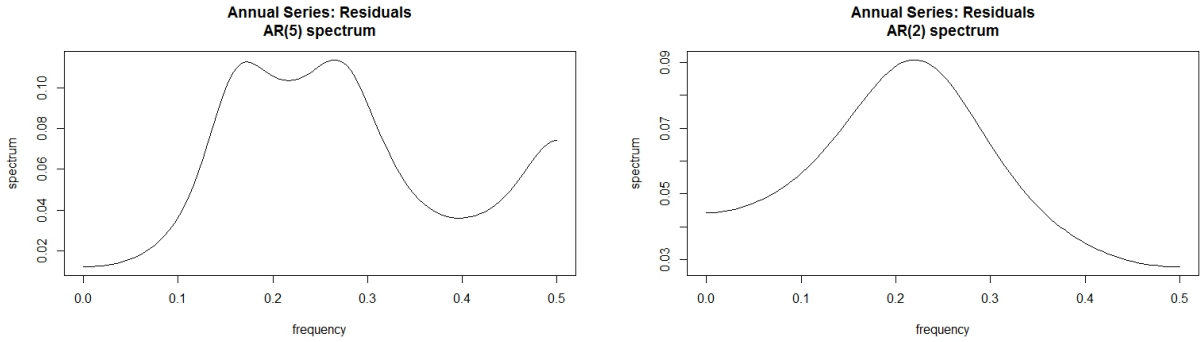

Figure 2.8: Autoregressive spectral estimators for the residuals obtained by fitting annual smoothed trend(left) and for the residuals obtained by fitting annual filtered value (right).

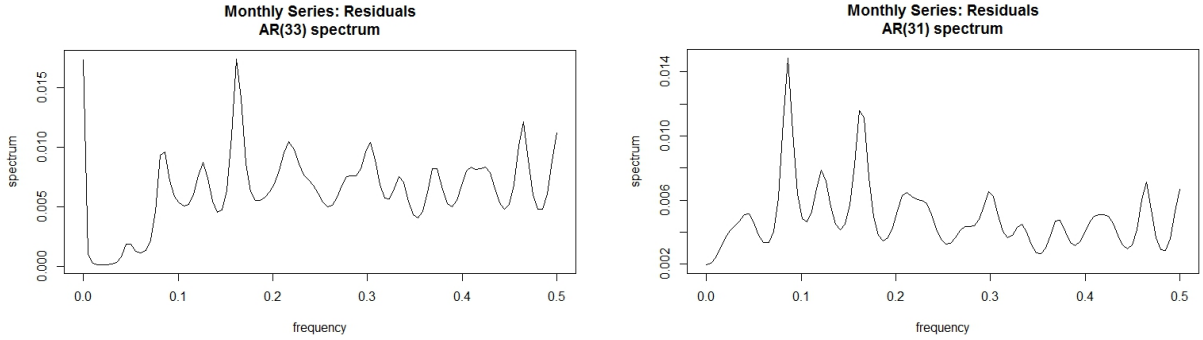

Figure 2.9: Autoregressive spectral estimators for the residuals obtained by fitting monthly smoothed trend(left) and for the residuals obtained by fitting monthly filtered value (right).

The figures above show that there exists an underlying periodic component in the residuals obtained by fitting the smoothed and filtered data to the observation. Comparing the results obtained for the parametric spectral estimates of the residual using filtered and smoothed data for both the annual and monthly case, shows a higher order of p for the smoothed data and a lower order of p for the filtered data. That been said, we can see that the first peak looking at the monthly series corresponding to the fitted filtered data is at a frequency $\omega = 1/50 = 0.02$, corresponding to a period of 50 months which is similar to the result obtained for the Southern Oscillation index(SOI) series by Shumway & Stoffe (2011) corresponding to the El Nino period. And the second peak is at a frequency of $\omega = 1/12.3 = 0.081$, with a period of approximately 12 months.

Therefore the need to extend the structural model to contain a Seasonal and cycle component

is important.

2.2.3 Trend + cycle Component

Treatment of time series as a trend plus cycle is a good deal for analysing temperature data. The model presented in Harvey and Koopman (1997), which was generalized by Harvey and Trimbur(2003) in such a way that it can produce smoother extracted cycles considering higher order stochastic cycles. A simple case will be to consider a first order stochastic cycle.

$$y_t = \mu_t + C_t + v_t, \tag{2.2.6}$$

Where the unobserved components are the trend μ_t, cycle C_t and the noise term v_t. The trend is given as in equation (2.2.1), since the slope term is insignificant to the model. A cycle can be expressed either as a sine or cosine wave. In its simplest form, C_t is written as a mixture of the sine and cosine waves.

$$C_t = \bar{C}\cos\lambda_c t + \bar{C}^*\sin\lambda_c t \tag{2.2.7}$$

Where λ_c is the frequency of the cycle measured in radians and its period is $2\pi/\lambda_c$. Just like the seasonal, the cycle can be allowed to change stochastically over time.

$$\begin{aligned} C_t &= C_{t-1}\cos\lambda_c + C^*_{t-1}\sin\lambda_c + \kappa_t \\ C^*_t &= -C_{t-1}\sin\lambda_c + C^*_{t-1}\cos\lambda_c + \kappa^*_t \end{aligned} \tag{2.2.8}$$

Where κ_t and κ^*_t are the two white noise disturbances which are assumed to be uncorrelated with each other and with the noise term in the observation. The property of the cyclic noise term is expressed in Carvalho et al(2007) as

$$E(\kappa_t\kappa_t) = E(\kappa^*_t\kappa^*_t) = \sigma_\kappa, \qquad\qquad E(\kappa_t\kappa^*_t) = 0$$

where σ_κ is the covariance matrix.

The cycle model can be stretched further by introducing the damping factor ρ, to give

$$\begin{bmatrix} C_t \\ C^*_t \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} C_{t-1} \\ C^*_{t-1} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix} \tag{2.2.9}$$

$$\begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \sim WN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\kappa^2 & 0 \\ 0 & \sigma_\kappa^2 \end{bmatrix} \right)$$

where $\rho \in [0, 1]$ is a damping factor required to ensure the stationarity of C_t. If it is equal to one, the cycle is non stationary . Cycles of this form fit naturally into the structural time series model framework.

Hence, the observation equation can be written as

$$y_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ C_t \\ C_t^* \end{pmatrix} + v_t$$

And the state equation written as

$$x_t = \begin{pmatrix} \mu_t \\ C_t \\ C_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho\cos(\lambda) & \rho\sin(\lambda) \\ 0 & -\rho\sin(\lambda) & \rho\cos(\lambda) \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ C_{t-1} \\ C_{t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_t \\ \kappa_t \\ \kappa_t^* \end{pmatrix}$$

Where $R = var(v_t) = r_{11} = \sigma_v^2$ and

$$Q = \begin{pmatrix} \sigma_\omega^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{pmatrix}$$

Where A_t is a 1 x 3 matrix, Φ_t is 3 x 3 matrix and the state noise terms are uncorrelated with mean zero and a 3 x 3 covariance matrix Q.

The parameters which are to be estimated are the three variance parameters and two cyclical parameters. $\sigma_\omega^2, \sigma_v^2, \sigma_\kappa^2, \rho, \lambda$. Taking starting values for the parameters and writing the scripts using the Kalman filter and Kalman smoother, the smoothed component is displayed below.

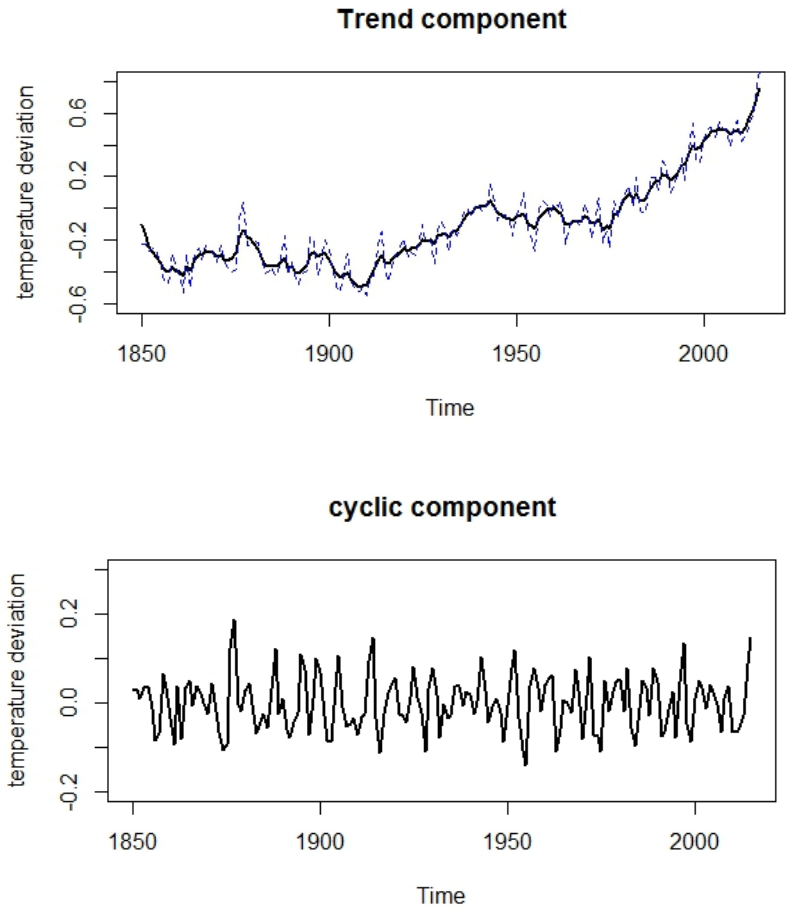

Figure 2.10: Estimated trend component, T_t^n (Top) and the estimated cycle component, C_t^n (bottom).

The final estimated parameters after 25 iterations for the first order cycle are:$\sigma_\omega = 0.0593, \sigma_v = 0.0414, \sigma_\kappa = 0.0301, \rho = 0.6136, \lambda = 1.1966$ and also the period is given as $\frac{2*\pi}{\lambda} = 5.2508$. The assumptions(specification) were chosen on the basis of parsimony, i.e that the noise terms are uncorrelated. Estimation of the parameters is done using the kfilter0 and ksmoother0 script in the Astsa package.

The trend component is a Local level model (simple driftless random walk model). There is a first order cycle having a period of about 5 years with a standard error of approximately one year. The variance of the stochastic cycle is computed as 0.0042, which clearly shows how the cycle component evolve over time(vary over time). It seems fairly regular atleast after 1900. Between 1850 and 1900 the cycle component looks volatile.

2.2.4 Cyclical Trend Component

The concept of this model is different from the trend plus cycle model. Cyclical components can be combined with other components such as the trend which is defined by (2.2.1). A cyclical trend component model defines a cyclic pattern embedded in the trend and not in the observation series hence the cyclical variations are also included in the long-term trend. This is put into the state space form by amending the measurement equation and the first row of the transition equation.

The unobserved state equation and observed equation can be written as:

$$\begin{aligned} \mu_t &= \mu_{t-1} + C_{t-1} + \omega_t, \\ C_t &= \rho(C_{t-1}\cos\lambda_c + C^*_{t-1}\sin\lambda_c) + \kappa_t \\ C^*_t &= \rho(-C_{t-1}\sin\lambda_c + C^*_{t-1}\cos\lambda_c) + \kappa^*_t \end{aligned} \tag{2.2.10}$$

$$y_t = \mu_t + v_t \tag{2.2.11}$$

The state equation in state space form is written as

$$x_t = \begin{pmatrix} \mu_t \\ C_t \\ C^*_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & \rho\cos(\lambda) & \rho\sin(\lambda) \\ 0 & -\rho\sin(\lambda) & \rho\cos(\lambda) \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ C_{t-1} \\ C^*_{t-1} \end{pmatrix} + \begin{pmatrix} \omega_t \\ \kappa_t \\ \kappa^*_t \end{pmatrix} \quad \omega_t \sim N(0,\sigma^2_\omega),\ \kappa_t \sim N(0,\sigma^2_\kappa),\ \kappa^*_t \sim N(0,\sigma^2_\kappa)$$

And the measurement equation is

$$y_t = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ C_t \\ C^*_t \end{pmatrix} + v_t$$

$$y_t = \mu_t + v_t, \qquad v_t \sim N(0,\sigma^2_v)$$

Where A_t is a 1 x 3 matrix, Φ_t is 3 x 3 matrix and the disturbance terms uncorrelated with mean zero and a covariance 3 x 3 matrix Q_t.

The concept of the cyclical trend is that the cycle is incorporated into the trend. Hence the

smoothed trend of the series is expected to have cyclical effects. We can compare this model to the previous case where the cycle was not a part of the trend in a later section. Below is a plot of cyclical trend.

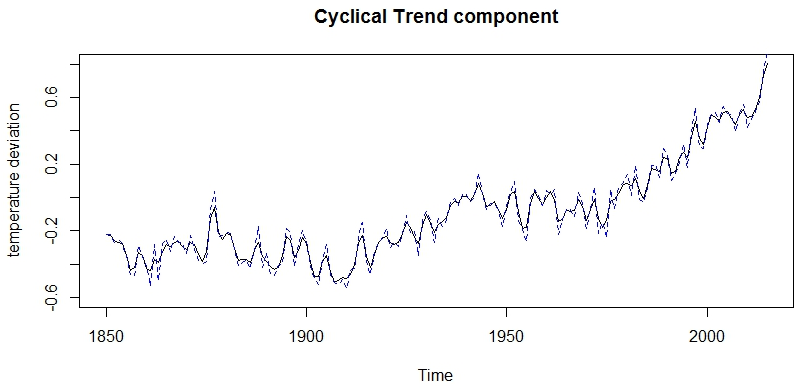

Figure 2.11: Cyclical trend of the annual temperature data

Using the same initial values $\Sigma_0 = 0.001, \mu_0 = -0.3$ as in the case of the trend plus cycle, and starting values; $\sigma_\omega = 0.1, \sigma_v = 0.1, \sigma_\kappa = 0.1, \rho = 0.4, \lambda = 0.3$, The final estimated parameters after 45 iterations are: $\sigma_\omega^2 = 0.00055, \sigma_\epsilon^2 = 0.06096, \sigma_\kappa^2 = 0.06857, \rho = 0.6755, \lambda = 1.3727$ and also the period is given as $\frac{2*\pi}{\lambda} = 4.5772$. The assumptions(specification) were chosen on the basis of parsimony, i.e that the noise terms are uncorrelated. Estimation of the parameters is done using the kfilter0 and ksmoother0 script.

Next we take a look at an overall comparison between the various model already discussed. i.e Trend, Trend plus cycle model and the Cyclical trend model.

2.2.5 Comparison between the various structural components

An appropriate way will be to compare the structural models which includes a look at the likelihood and AIC values, residuals and goodness of fit of the various models.

Parameters	Trend	Trend + Cycle	Cyclical trend
$\hat{\rho}$		0.6136	0.7117
		(0.2222)	(0.0765)
$\hat{\lambda_c}$		1.1966	1.4725
		(0.2505)	(0.1262)
$\hat{\sigma}_w$	0.04628	0.0593	0.00055
	(0.01369)	(0.0101)	(0.1508)
$\hat{\sigma}_v$	0.08058	0.0325	0.06096
	(0.00902)	(0.0414)	(0.0082)
$\hat{\sigma}_\kappa$		0.0651	0.06857
		(0.0301)	(0.0103)
$\frac{2*\pi}{\lambda}$		5.2508	4.5772
Likelihood	-285.3353	-283.5029	-282.7258
AIC	576.6706	577.0058	575.4516

Table 2.3: Table showing the various models and the estimated value of there parameters

The AIC is given as $-2l + 2p$, where l is the maximized likelihood value and p is the number of estimated parameters.

The table above shows the parameter estimates obtained by decomposing the observed Annual global temperature series into a set of unobserved components(trend, cycle, cyclical trend), together with their standard errors in parenthesis .

A look at the above table shows that the standard deviation of the noise term in the trend plus cycle component is higher than that of the cyclical trend component. A logical conclusion from this is that when the cycle is embedded in the trend, since the disturbances are uncorrelated, we get a trend that is dominated by a cyclic pattern. The AIC between the three models shows that an appropriate choice is the cyclical trend model with the lowest AIC value of 575.4516 but we know that the model with lowest AIC value does not always represent the best fit and from the model trends in the above plots we can see that an appropriate model choice will be the trend plus cycle which gives a smoother trend and investigation of the residuals shows that the model is a good fit.

2.3 Structural models with Monthly Temperature Series

The point of decomposing the structural model to the monthly temperature series is to find some sort of disparity with the results obtained by decomposing the structural model to the annual temperature series. First we take a look at the trend components for the monthly temperature data. Of course there will be similarities with the results obtained from the smoothed trend of the annual temperature data and that of the trend of the monthly temperature data.

2.3.1 Trend

So here we have a Local linear trend model with a noise term. This model is referred to as a random walk model.

The state space form of the trend is the same as in subsection 2.2.1. With the variance of its disturbance insignificant from zero and thus a deterministic(constant) slope, with the same initial value used for the annual temperature data, the result obtained after about 9 iterations of a Newton - Raphson, the observation uncertainty was $\hat{\sigma}_v = 0.09383$, compared with the state uncertainty $\hat{\sigma}_\omega = 0.05729$. Figure(2.12) below shows the smoothed trend estimate, this uses the Kfilter0 and Ksmooth0 scripts of the astsa package in R, developed by Shumway & Stoffe (2011)

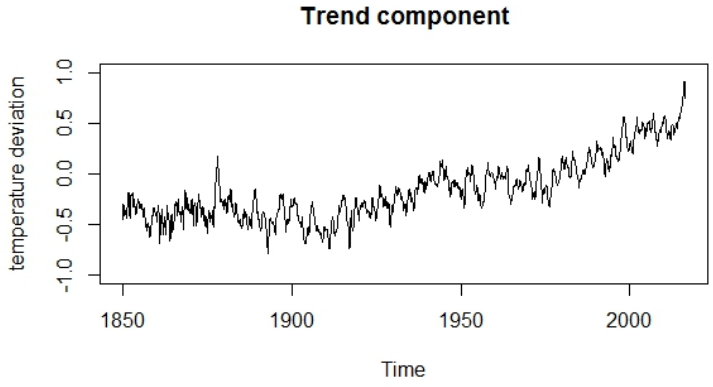

Figure 2.12: Smoothed trend estimate

Comparing the smoothed trend of the monthly global temperature to that of smoothed trend of the annual global temperature, we can deduce that the Kalman filter and smoother algorithm does not perform very well for the Linear trend model with very large data.

Next a look at a more inclusive structural model, that is considering the time series to be the

sum of the trend component, a seasonal component, and a white noise.

2.3.2 Trend + Seasonal Component

As with a trend, a seasonal component is defined in terms of the predictions it yields. The estimated seasonal is that part of the series which, when extrapolated, repeats itself over any one-year time period and averages out to zero over such a time period(Harvey(1989). In a structural model, the seasonal part satisfies the conditions below.

The observed series is expressed as

$$Y_t = \mu_t + S_t + v_t, \tag{2.3.1}$$

where μ_t is the trend and S_t is the seasonal component. Let the trend be a local linear trend model as given in (2.2.2), with $\phi = 1$, that is ,

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_t \\ \beta &= \beta_{t-1} + \eta_t \end{aligned} \tag{2.3.2}$$

and let the seasonal component be modeled as

$$S_{t+1} = -\sum_{j=1}^{s-1} S_{t+1-j} + \gamma_t, \qquad \gamma_t \sim N(0, \sigma_\gamma^2) \tag{2.3.3}$$

for $t = 1, .., n$, for monthly data $s = 12$.

Durbin and Koopman (2001) showed that there are other alternatives under which the seasonal component can be modeled.

For constant seasonal pattern over time, the seasonal values for month 1 to s can be modelled by constants $S_1^*, ..., S_s^*$ where $\sum_{i=1}^{s} S_i^* = 0$. Which follows that $\sum_{j=0}^{s-1} S_{t+1-j} = 0$ and thus,

$$S_{t+1} = -\sum_{j=1}^{s-1} S_{t+1-j} \qquad t = s-1, s, ... \tag{2.3.4}$$

An alternative is express the the seasonal in a trigonometric form, which can be done for two versions of the seasonal, i.e one version for a constant seasonal and the other version for a time varying seasonal which is made stochastic by replacing the constant seasonal by random walks

Durbin and Koopman (2001)

$$S_t = \sum_{j=1}^{[s/2]} S_{j,t}$$

Where each $S_{j,t}$, $j = 1, ..., 6$ is generated by

$$\begin{bmatrix} S_{j,t} \\ S^*_{j,t} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} S_{j,t-1} \\ S^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{j,t} \\ \gamma^*_{j,t} \end{bmatrix} \lambda_j = \frac{2\pi j}{s}, \quad j = 1, ..., [s/2] \tag{2.3.5}$$

Where λ_j is the frequency (in radians) and $\gamma_{j,t} \sim \text{NID}\ (0, \sigma^2_\gamma)$ and $\gamma^*_{j,t} \sim \text{NID}\ (0, \sigma^2_\gamma)$ are mutually uncorrelated noise having a common variance. The component $S^*_{j,t}$ simply allows the seasonal to be modelled as a stochastic combination of sine and cosine waves. This is similar to the cycle component.

The paper by T.C Mills(2010) focused on using the trigonometric form which is referred to as the quasi random walk model, where the stochastic terms were assumed to be zero i.e $\sigma^2_\gamma = 0$, thus leading to the conclusion that the seasonal component has a deterministic pattern (being constant from year to year). Equation (2.3.5) is referred to as the main time domain model for the seasonal component in structural time series analysis.

In our discussion, we will try to see if modeling the seasonal component will give a deterministic pattern (i.e., if the variance of the noise term is close to zero).

Equation (2.3.3) can also be written as

$$S_t = -S_{t-1} - S_{t-2} - S_{t-3} - S_{t-4} - S_{t-5} - S_{t-6} - S_{t-7} - S_{t-8} - S_{t-9} - S_{t-10} - S_{t-11} + \gamma_t \tag{2.3.6}$$

Which corresponds to the assumption that the seasonal component is expected to sum to zero over a complete period. We can express equations(2.3.1),(2.3.2) and (2.3.6) in state space form. The observation equation can be written as

$$y_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ S_t \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ S_{t-4} \\ S_{t-5} \\ S_{t-6} \\ S_{t-7} \\ S_{t-8} \\ S_{t-9} \\ S_{t-10} \end{pmatrix} + v_t \tag{2.3.7}$$

And the state equation written as

$$x_t = \begin{pmatrix} \mu_t \\ \beta_t \\ S_t \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ S_{t-4} \\ S_{t-5} \\ S_{t-6} \\ S_{t-7} \\ S_{t-8} \\ S_{t-9} \\ S_{t-10} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ S_{t-4} \\ S_{t-5} \\ S_{t-6} \\ S_{t-7} \\ S_{t-8} \\ S_{t-9} \\ S_{t-10} \\ S_{t-11} \end{pmatrix} + \begin{pmatrix} \omega_t \\ \eta_t \\ \gamma_t \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Where $R = r_{11} = var(v_t)$ and

$$Q = \begin{pmatrix} q_{11} & 0 & 0 & 0 & . & . & . & 0 \\ 0 & q_{22} & 0 & 0 & . & . & . & 0 \\ 0 & 0 & 0 & q_{33} & . & . & . & 0 \\ . & . & . & & . & . & . & . \\ . & . & . & & . & . & . & . \\ . & . & . & & . & . & . & . \\ . & . & . & & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & & 0 & 0 \end{pmatrix}$$

The parameters to be estimated just like we did with the trend component, are σ_v the standard deviation of the white noise term in the observation equation, σ_ω, σ_β and σ_γ , the state standard deviation corresponding to the level, slope and the seasonal component respectively.

The assumption for the initial mean and the diagonal covariance matrix with $\Sigma_{0_{ii}}$, for $i = 1, ..., 12$ is the same as in the previous trend component. The initial observation and state covariance values were taken as $\sigma_v = 0.01$, and $\sigma_\omega = 0.01$, $\sigma_\eta = 0.01$, $\sigma_\gamma = 0.01$, respectively. Just like in the trend case, we use the Kfilter0 and Ksmooth0 scripts to obtain a smoothed trend and seasonal estimate which is shown below.

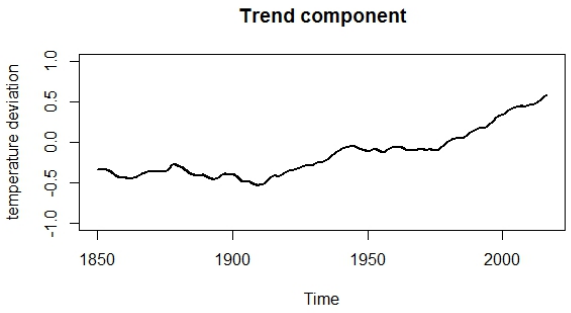

Figure 2.13: Estimated trend component, T_t^n for monthly temperature data.

The figure above shows the smoothed trend. After 16 iterations of the Newton - Raphson iteration, the observation uncertainty reduces to $\sigma_v = 0.09271$ and the state uncertainties are $\sigma_\omega = 0.05286$ and $\sigma_\gamma = 1.7048$ x 10^{-7} for the trend and seasonal respectively. where the standard deviation for the slope is insignificant from zero, thus a deterministic(constant) slope is obtained. Thus it can be seen that the variance of the noise term in the seasonal component $\hat{\sigma}_\gamma^2 = (1.7048$ x $10^{-7})^2$ is zero i.e the seasonal component has a zero innovation variance so that it has a

deterministic pattern, thus near constant from year to year. The likelihood value is given as (K = -3106.0408)

The question of modelling the fixed seasonal pattern in the state space can be dealt with by considering the seasons as explanatory variables or input(exogenous variables) in the state or observation equation,

$$\begin{aligned} \mu_t &= \mu_{t-1} + \Upsilon u_t + \omega_t \\ y_t &= \mu_t + \Gamma u_t + v_t \end{aligned} \tag{2.3.8}$$

where u_t is the input which will be specified as a seasonal variable i.e the regression effect of the time series model is placed in Υu_t.

$$u_t = \begin{bmatrix} S_1 & S_2 & S_3 & . & . & . & . & . & S_{11} & constant \end{bmatrix}^T$$

A centered seasonal average is employed which is used to obtain the annual pattern plus a constant term(intercept). The Γ or Υ is the amplitude that specifies the measure of change over a specific period, this variable will be estimated.

$$S_1 = \begin{pmatrix} \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

$$S_2 = \begin{pmatrix} \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

$$S_3 = \begin{pmatrix} \frac{-1}{12} & \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

$$S_4 = \begin{pmatrix} \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

$$S_5 = \begin{pmatrix} \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

$$S_6 = \begin{pmatrix} \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} \end{pmatrix}$$

....

$$S_{11} = \begin{pmatrix} \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{-1}{12} & \frac{11}{12} & \frac{-1}{12} \end{pmatrix}$$

This method gives a different approach to handling the seasonal effect by assuming it as an exogenous variable in the state space model. The result obtained from this shows an estimated amplitude of ($\Upsilon = 0.2$) which shows the height of the period.

A look at the autocorrelation function of the residuals for both cases where we have seasonal modeled in the unobserved component and when the seasonal is treated as an exogenous variable in the observation equation, this will show that there is very little correlation between the residuals when we introduce the seasonal as an exogenous variable.

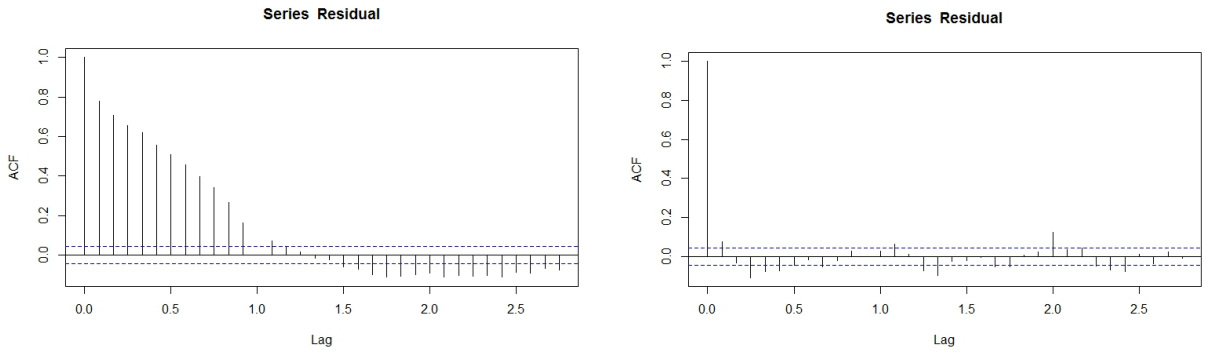

Figure 2.14: Autocorrelation function for irregular residual without inputs(exogenous variable) (Left), Autocorrelation function for irregular residual with inputs(exogenous variable)(Right)

In conclusion to the introduction of the seasonal component as an input in the state space is that this helps as an intervention variable to reduce the dependence on correlation in the residuals. In the next chapter we will see how the use of intervention variables will be used to handle outliers and structural breaks.

2.3.3 Trend + cycle Component

The state space form of the trend plus cycle model is the same as in subsection 2.2.3. With the initial mean and covariance set at $\mu_0 = -0.35$ and $\Sigma_0 = 0.01$, starting values of $\rho = 0.7, \lambda = 0.6, \sigma_w = 0.1, \sigma_\kappa = 0.1$, and $\sigma_v = 0.1$. $\beta = 0$ which implies that the trend model is a local level model.

After 30 iterations of a Newton - Raphson, the result is given as $\rho = 0.8045, \lambda_c = 0.0179, \sigma_\omega = 0.0116, \sigma_v = 0.0811, \sigma_\kappa = 0.0746$. The Figure below shows the smoothed trend estimate and cycle component, this uses the Kfilter0 and Ksmooth0 scripts of the astsa package in R, developed by Shumway & Stoffe (2011)

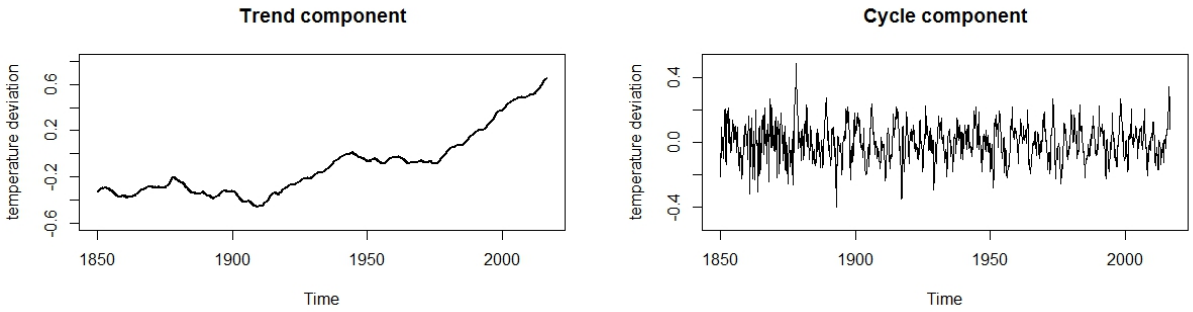

Figure 2.15: Estimated trend component, T_t^n (Left) and the estimated cycle component, S_t^n (Right) for monthly temperature data.

A summary to this is that a cyclic period of $2\pi/\hat{\lambda}_c = 351.0159$, which is about 29 years for the first order cycle. The trend component is by no means smooth but the warming trends for the last 30 years of the twentieth century are clearly apparent.

2.3.4 Cyclical Trend Component

The model is the same as in section 2.2.4. Initial values remain the same as in the case of trend plus cycle. We use the Kfilter0 and Ksmooth0 script to obtain the filters and smoothers for the estimated parameters. The figure below shows the cyclical trend for the monthly temperature data.

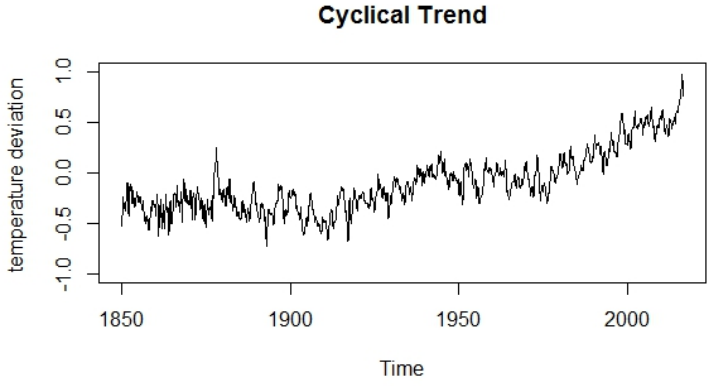

Figure 2.16: Cyclical trend component

The result obtained after 33 iterations where the uncertainty of the trend disturbance is given as $\sigma_w = 0.00033$ which is insignificant from zero, thus we can see a dominant cyclic behaviour

and a deterministic trend. The values for the parameter estimation is shown in the table 2.4

2.3.5 Trend + Cycle + Autoregressive AR(1)

The structural model can be enhanced further by introducing an Autoregressive component.

The autoregressive component can be specified as an AR(1) process

$$\psi_t = \phi_1 \psi_{t-1} + \xi_t, \qquad \xi_t \sim N(0, \sigma_\xi^2)$$

Thus the structural components are:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \omega_t, \\ C_t &= \rho(C_{t-1}\cos\lambda_c + C_{t-1}^* \sin\lambda_c) + \kappa_t \\ C_t^* &= \rho(-C_{t-1}\sin\lambda_c + C_{t-1}^* \cos\lambda_c) + \kappa_t^* \\ \psi_t &= \phi_1 \psi_{t-1} + \xi_t \end{aligned} \tag{2.3.9}$$

And the observed equation is written as

$$y_t = \mu_t + C_t + \psi_t + v_t \tag{2.3.10}$$

The state space form of this model is written as

$$y_t = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ C_t \\ C_t^* \\ \psi_t \end{pmatrix} + v_t$$

$$v_t \sim N(0, \sigma_v^2)$$

And the state equation is written as

$$x_t = \begin{pmatrix} \mu_t \\ C_t \\ C_t^* \\ \psi_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \rho\cos(\lambda) & \rho\sin(\lambda) & 0 \\ 0 & -\rho\sin(\lambda) & \rho\cos(\lambda) & 0 \\ 0 & 0 & 0 & \theta_1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ C_{t-1} \\ C_{t-1}^* \\ \psi_{t-1} \end{pmatrix} + \begin{pmatrix} \omega_t \\ \kappa_t \\ \kappa_t^* \\ \xi_t \end{pmatrix}$$

$\omega_t \sim N(0, \sigma_\omega^2), \quad \kappa_t \sim N(0, \sigma_\kappa^2), \ \kappa_t^* \sim N(0, \sigma_\kappa^2)$

Where $R = var(v_t) = r_{11} = \sigma_v^2$ and

$$Q = \begin{pmatrix} \sigma_\omega^2 & 0 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 & 0 \\ 0 & 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & 0 & \sigma_\xi^2 \end{pmatrix}$$

Where A_t is a 1 x 4 matrix, Φ_t is 4 x 4 matrix and the state noise terms are uncorrelated with mean zero and a 4 x 4 covariance matrix Q.

The parameters which are to be estimated are the four standard deviation parameters and two cyclical parameters. $\sigma_\omega, \sigma_v, \sigma_\kappa, \sigma_\xi, \rho, \lambda$. Taking starting values for the parameters and writing the scripts using the Kalman filter and Kalman smoother, the smoothed component is displayed below, together with the cycle, AR(1) process.

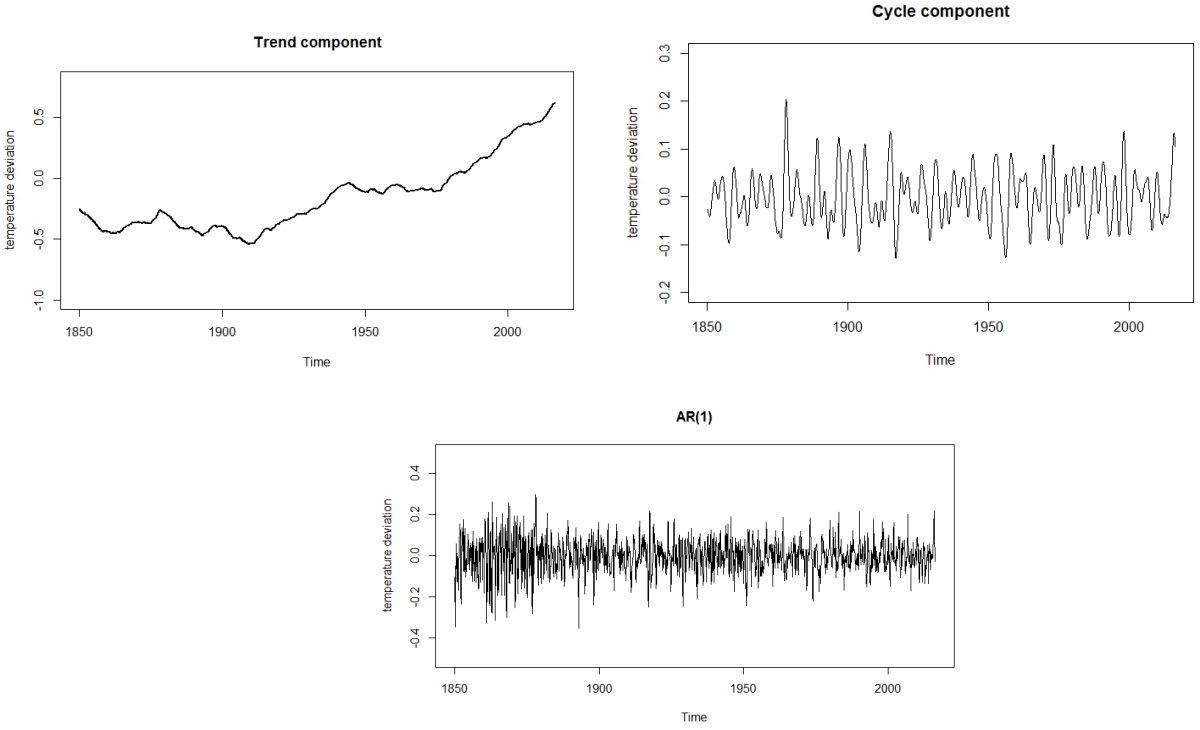

Figure 2.17: trend + cycle + AR(1)+ irregular component

The values of the estimated parameters are given in table 2.4. The autoregressive process

as seen above is a non stationary process with $\phi_1 = 1$. Harvey (1989) introduced the autoregressive component into the structural model by generalising the irregular term to a stationary autoregressive process which results to the trend and irregular component being subject to the autoregressive effect. The equation below shows the structural model by Harvey (1989).

$$\begin{aligned} y_t &= \mu_t + \psi_t \\ \mu_t &= \mu_{t-1} + \omega_t, \\ \psi_t &= \phi_1\psi_{t-1} + \phi_2\psi_{t-2} + - - - - - - - + \phi_p\psi_{t-p} + v_t \end{aligned} \tag{2.3.11}$$

where v_t is the irregular term, ψ_t is the AR process of order p.

2.3.6 Trend + Season + cycle Component

Another exciting model to discuss is a structural model that decomposes the data into a trend, season and cycle component.

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_t + \omega_t, \\ S_t &= -S_{t-1} - S_{t-2} - S_{t-3} - S_{t-4} - S_{t-5} - S_{t-6} - S_{t-7} - S_{t-8} - S_{t-9} - S_{t-10} - S_{t-11} + \gamma_t \\ C_t &= \rho(C_{t-1}\cos\lambda_c + C^*_{t-1}\sin\lambda_c) + \kappa_t \\ C^*_t &= \rho(-C_{t-1}\sin\lambda_c + C^*_{t-1}\cos\lambda_c) + \kappa^*_t \end{aligned} \tag{2.3.12}$$

$$y_t = \mu_t + S_t + C_t + v_t \tag{2.3.13}$$

The state space form of this model is written as

$$y_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \\ S_t \\ . \\ . \\ S_{t-10} \\ C_t \\ C_t^* \end{pmatrix} + v_t$$

$$v_t \sim N(0, \sigma_v^2)$$

And the state equation is written as

$$x_t = \begin{pmatrix} \mu_t \\ \beta_t \\ S_t \\ . \\ . \\ S_{t-10} \\ C_t \\ C_t^* \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & seasonal & 0 \\ 0 & 0 & 0 & Cycle \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ S_{t-1} \\ . \\ . \\ S_{t-11} \\ C_{t-1} \\ C_{t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_t \\ \eta_t \\ \gamma_t \\ . \\ . \\ . \\ \kappa_t \\ \kappa_t^* \end{pmatrix}$$

$\omega_t \sim N(0, \sigma_\omega^2),\ \gamma_t \sim N(0, \sigma_\gamma^2),\ \kappa_t \sim N(0, \sigma_\kappa^2),\ \kappa_t^* \sim N(0, \sigma_\kappa^2)$

The variance term is described as

$R = var(v_t) = r_{11} = \sigma_v^2$

and

$$Q = \begin{pmatrix} \sigma_\omega^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\gamma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\kappa^2 \end{pmatrix}$$

Where A_t is a 1 x 14 matrix, Φ_t is 14 x 14 matrix and the state noise terms are uncorrelated with mean zero and a covariance 14 x 14 matrix Q.

Next will be to estimate the parameters in the state space model which are: $\sigma_\omega, \sigma_\eta, \sigma_\gamma, \sigma_\kappa, \rho, \lambda_c, \sigma_v$, with initial values μ_0, Σ_0 and starting values for the parameters, the Kalman filter and smoother script Kfilter0 a Ksmooth0 are used for prediction, filtering and smoothing based on the Kalman filter and Smoother property.

The result of the estimated parameters obtained after 26 iterations of a Newton - Raphson, is shown in table 2.4., with the variance of the slope equal to zero , thus a deterministic(constant) slope is obtained. The plots of the trend , seasonal and cycle component is shown below

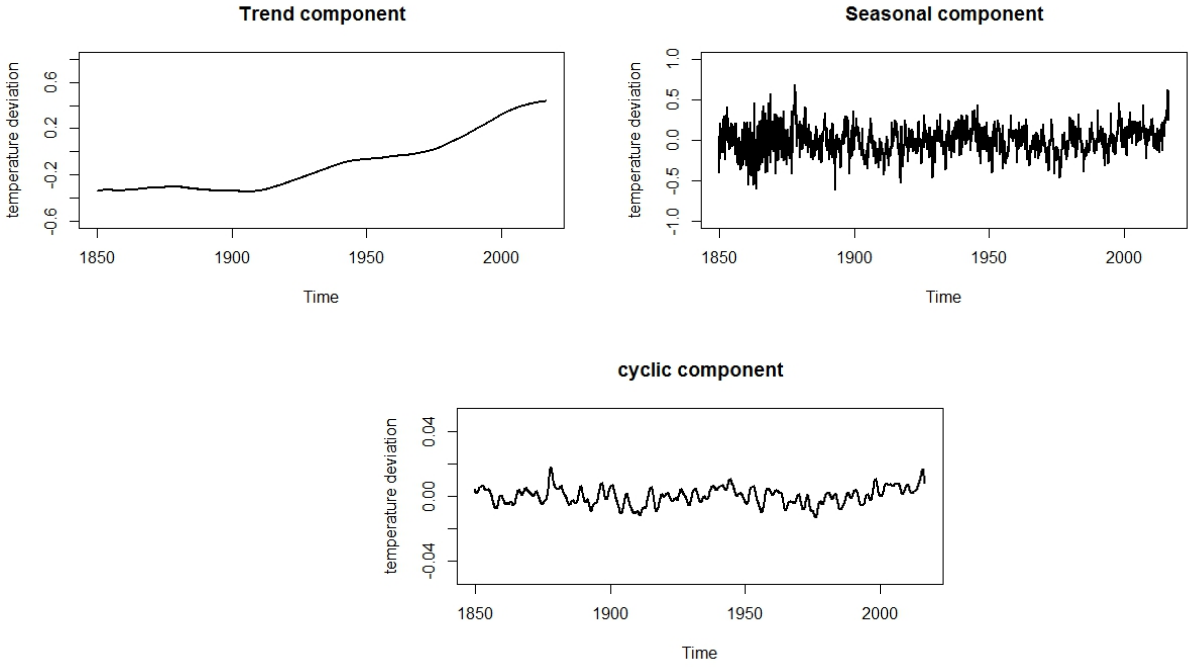

Figure 2.18: Estimated trend component, T_t^n (Top left) and the estimated seasonal component, S_t^n (Top Right) and the cyclic component(bottom) for monthly temperature data.

The result obtained shows a deterministic seasonal pattern with zero innovation variance. A look at the autocorrelation function of the residual shows very high correlation, an introduction of intervention variables i.e the use of the seasonal as an exogenous variable solves this problem.

2.3.7 Comparison between the various structural components

An appropriate way will be to compare the structural models which includes a look at the likelihood and AIC values, residuals and goodness of fit of the various models.

Parameters	Trend	T + S	T + C	Cyclical trend	T + S + C	T + C + AR(1)
$\hat{\theta}_1$						0.5846*
						(0.0859)
$\hat{\rho}$			0.8045*	0.7365*	0.8256*	0.9474*
			(0.0302)	(0.0345)	(0.0524)	(0.0236)
$\hat{\lambda_c}$			0.0179	0.7333*	0.2041*	0.1303*
			(0.0836)	(0.1008)	(0.1277)	(0.023)
$\hat{\sigma}_\omega$	0.05729*	0.05286*	0.0116*	0.00033	0.03646*	0.0121*
	(0.00332)	(0.003)	(0.0023)	(0.0151)	(0.00389)	(0.0022)
$\hat{\sigma}_\gamma$		0.00111			0.0003	
		(0.0015)			(0.0011)	
$\hat{\sigma}_v$	0.09383*	0.09271*	0.0811*	0.09286*	0.04329*	0.0705*
	(0.00259)	(0.0024)	(0.0036)	(0.0025)	(0.0261)	(0.0073)
$\hat{\sigma}_\kappa$			0.0746*	0.0406*	0.09426*	0.0385*
			(0.0047)	(0.0052)	(0.0068)	(0.0168)
σ_ξ						0.0857*
						(0.0074)
Likelihood	-3177.202	-3160.703	-3150.06	-3156.003	-3131.3201	-3130.285
AIC	6358.404	6327.406	6310.12	6322.006	6274.64	6274.57

Table 2.4: Comparing the estimated parameters of the component of the structural model [1]

The AIC is given as $-2l + 2p$, where l is the maximized likelihood value and p is the number of estimated parameters.

The table above shows the parameter estimates obtained by decomposing the observed monthly global temperature series into a set of unobserved components(trend,seasonal, cycle, cyclical trend, AR(1) process), together with their standard errors in parenthesis . All the parameters of the model are estimated with a good degree of precision.

There are several insignificant parameters in the above table but the significant parameters which are marked with $*$ are bigger than the 95% confidence level hence, a 5% significance level.

It is important to state that, obtaining significant maximum likelihood estimate, comes from choosing appropriate initial values and starting values which can be very difficult when we have

[1] *Estimated value that represents the 5% significance level

several parameters to be estimated. Thus the effect of increased number of parameters is penalized by the AIC.

Since the models might be considered to be nested, it is expected that the likelihood value should increase as the number of parameters increases but that could be affected by the complexity of the models and appropriate initial and starting values.

Model selection by the use of the lowest AIC value will tend to favour the structural model that includes the trend, cycle and AR(1) process. Looking across all five models, one consistent thing that can be clearly spotted is the smoothness of the trend, thus one can deduce that the more structure the models get the smoother the trend becomes, and the higher the number of significant parameters estimated.

The cycle, although clearly stochastic, is fairly regular. High volatility can be seen in the cycle in the first 50 years but thereafter displays a homogeneous behaviour.

2.4 Models with correlated errors

From the state space model equation given in section 2.3, the state space model is written with a slight adjustment to the state equation. The idea behind this modification is that the state noise process starts at $t = 0$. This is done due to alignment of the covariance between w_t and v_t.

$$x_{t+1} = \Phi x_t + \Upsilon u_{t+1} + \Theta \omega_t \qquad t = 0, 1, 2, ..., n \tag{2.4.1}$$

$$y_t = A_t x_t + \Gamma u_t + v_t \qquad t = 0, 1, 2, ..., n \tag{2.4.2}$$

In the state equation, $x_0 \sim N_p(\mu_0, \Sigma_0)$, Φ is p x p, Υ is p x r , Θ is p x m and $\omega_t \sim$ iid $N_m(0, Q)$. The inclusion of the matrix Θ is to allow us avoid the use of singular state noise process Shumway & Stoffer (2011). In the observation equation, A_t is q x p , Γ is q x r and $v_t \sim$ iid $N_q(0, R)$. In this model, we allow for correlation between the state noise ω_t and the observation noise v_t at time t. i.e.

$$cov(\omega_s, v_t) = \begin{cases} S, & \text{if } s = t, \\ 0, & \text{if } s \neq t \end{cases} \tag{2.4.3}$$

Where S is an m x q matrix. Thus, from equation (2.4.1) and (2.4.2), and with the condition of correlation between the noise term, together with the innovations $\epsilon_t = y_t - A_t x_t^{t-1} - \Gamma u_t$ and

the innovation variance $\Sigma_t = A_t P_t^{t-1} A_t^{'} + R$, we obtain the one step ahead prediction x_{t+1}^t, and the filter values x_t^t together with the prediction errors P_{t+1}^t, P_t^t respectively.

The proof to the Kalman filter and smoother with correlated errors is shown here. A two step equation may be obviously combined into a one step update which computes x_{t+1}^t from x_t^t or x_{t+1}^{t-1} from x_t^{t-1}.

Hence the one step ahead prediction x_{t+1}^t is obtained by carrying out the following computation.

$$\begin{aligned} x_{t+1} &= \Phi x_t + \Upsilon u_{t+1} + \Theta \omega_t \qquad t = 0, 1, 2, ..., n \\ x_{t+1}^{t-1} &= E(x_{t+1}|Y_{t-1}) = E(\Phi x_t + \Upsilon u_{t+1} + \Theta \omega_t | Y_{t-1}) \\ &= \Phi x_t^{t-1} + \Upsilon u_{t+1} \end{aligned} \tag{2.4.4}$$

and thus

$$\begin{aligned} P_{t+1}^{t-1} &= E[(x_{t+1} - x_{t+1}^{t-1})(x_{t+1} - x_{t+1}^{t-1})^{'} | Y_{t-1}] \\ &= E[(\Phi x_t + \Upsilon u_{t+1} + \Theta \omega_t - \Phi x_t^{t-1} - \Upsilon u_{t+1})(\Phi x_t + \Upsilon u_{t+1} + \Theta \omega_t - \Phi x_t^{t-1} - \Upsilon u_{t+1})^{'} | Y_{t-1}] \\ &= [\Phi(x_t - x_t^{t-1}) + \Theta \omega_t][\Phi(x_t - x_t^{t-1}) + \Theta \omega_t]^{'} \\ &= \Phi P_t^{t-1} \Phi^{'} + \Theta Q \Theta^{'} \end{aligned} \tag{2.4.5}$$

Next we obtain the conditional covariance between x_{t+1} and ϵ_t given Y_{t-1} is

$$\begin{aligned} cov(x_{t+1}, \epsilon_t | Y_{t-1}) &= cov(x_{t+1}, y_t - A_t x_t^{t-1} - \Gamma u_t | Y_{t-1}) \\ &= cov(x_{t+1} - x_{t+1}^{t-1}, y_t - A_t x_t^{t-1} - \Gamma u_t | Y_{t-1}) \\ &= cov(\Phi(x_t - x_t^{t-1}) + \Theta \omega_t, A_t(x_t - x_t^{t-1}) + v_t) \\ &= \Phi P_t^{t-1} A_t^{'} + \Theta cov(\omega_t, v_t) \\ &= \Phi P_t^{t-1} A_t^{'} + \Theta S \end{aligned} \tag{2.4.6}$$

Using these results, we have that the joint conditional distribution of x_{t+1} and ϵ_t given Y_{t-1}

is normal

$$= \begin{pmatrix} x_{t+1} \\ \epsilon_t \end{pmatrix} \mid Y_{t-1} \sim N\left(\begin{pmatrix} x_{t+1}^{t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} P_{t+1}^{t-1} & \Phi P_t^{t-1} A_t^{'} + \Theta S \\ A_t P_t^{t-1} \Phi^{'} + \Theta S & A_t P_t^{t-1} A_t^{'} + R \end{pmatrix} \right)$$

And thus , using the property of a multivariate normal distribution, we can write

$$\begin{aligned} x_{t+1}^t &= E[x_{t+1}|y_1, y_2, ..., y_{t-1}, y_t] = E[x_{t+1}|Y_{t-1}, \epsilon_t] \\ &= x_{t+1}^{t-1} + (\Phi P_t^{t-1} A_t^{'} + \Theta S)\Sigma_t^{-1}\epsilon_t \\ &= \Phi x_t^{t-1} + \Upsilon u_{t+1} + K_t \epsilon_t \end{aligned} \tag{2.4.7}$$

where $K_t = (\Phi P_t^{t-1} A_t^{'} + \Theta S)(A_t P_t^{t-1} A_t^{'} + R)$ is the Kalman gain.

The corresponding covariance vector P_{t+1}^t is given as

$$\begin{aligned} P_{t+1}^t &= cov(x_{t+1}|Y_{t-1}, \epsilon_t) \\ &= P_{t+1}^{t-1} - (\Phi P_t^{t-1} A_t^{'} + \Theta S)\Sigma_t^{-1}(\Phi^{'} P_t^{t-1} A_t^{'} + \Theta S) \\ &= \Phi P_t^{t-1} \Phi^{'} + \Theta Q \Theta^{'} - K_t \Sigma_t K_t \end{aligned} \tag{2.4.8}$$

The filter equations is the same as the previous filter property given in subsection 2.1

$$\begin{aligned} x_t^t &= E[x_t|y_1, ..., y_{t-1}, y_t] = E[x_t|Y_{t-1}, \epsilon_t] \\ &= x_t^{t-1} + P_t^{t-1} A_t^{'}[A_t P_t^{t-1} A_t^{'} + R]^{-1}\epsilon_t \\ P_t^t &= P_t^{t-1} - P_t^{t-1} A_{t+1}^{'} \Sigma_t^{-1} A_t P_t^{t-1} \end{aligned} \tag{2.4.9}$$

It can be seen from the proof above that the covariance $cov(\omega_t, v_t)$ is only used in the one step ahead prediction i.e the predictor x_{t+1}^t is generated from the past predictor x_t^{t-1} when the noise terms are correlated, And this leads to the filter update.

The structure of the covariance matrix is given as

$$\Sigma = \begin{bmatrix} Q & cov(\omega_t, v_t) \\ cov(\omega_t, v_t) & R \end{bmatrix}$$

$$cov(\omega_t, v_t) = \beta\ \sigma_\omega\ \sigma_v$$

Where β is the correlation coefficient between the noise term of the state and the observation and σ_w and σ_v are the standard deviation of the noise terms. The correlation coefficient provides additional information by telling us the degree to which both variables move together. The correlation will always have a measurement value between -1 and 1, and adds a strength value on how the noise term move together. If the correlation is 1, they move perfectly together, and if the correlation is -1, the noise term move perfectly in opposite directions. If the correlation is 0, then the two noise term move in random directions from each other.

2.4.1 Trend with correlated error

Applying the idea of correlated error to a simple case of random walk trend component for equations (2.4.1) and (2.4.2), ignoring any inputs, the only values to be estimated are the standard deviation of the state equation Q, and that of the observation equation R and the covariance between the two noise terms S.

This said, we look at two simple scenarios;

- when $\omega_t = v_t$, hence cov$(\omega_t, v_t) = var(v_t) = R$
- when $\omega_t \neq v_t$, hence cov$(\omega_t, v_t) = \sigma_{\omega v}$

Next we take a trend model for the first case where the noise term is equal to that of the observation disturbance $\omega_t = v_t$. For this case, A_t is an array of 1's with dimension of nrow =1, ncol=1, for each value of the length of the given data. Inputs = 0, initial mean(μ_0) = -0.2, Σ_0 = 0.001 and Phi(Φ)= 1. The parameter to be estimated is the standard deviation of the observation noise. the correlation coefficient ($\beta = 1$). starting values for $\sigma_v = 0.1$. The estimation process uses the "L-BFGS-B" method when using the optim function in R.

After the optimization process, the estimated value for the observation uncertainty σ_v = 0.1151, and standard error = 0.00633. The Kalman smoother property fails in this condition owing to the choice of singularity placed on the correlated noise term. This can be corrected by defining $v = \Theta\omega$, where Θ is not singular, hence we can obtain an identifiable model assuming perfectly correlated disturbances.

For the case 2 where the noise terms are not equal $\omega_t \neq v_t$. matching the non zero auto-covariances of $\triangle y_t$ yields a (nonlinear) system of two equations in three unknowns$(\sigma^2_\omega, \sigma^2_v, \sigma_{\omega v})$

which has infinite solutions. Imposing $\sigma_{\omega v} = 0$ gives a unique solution. Tommaso Proietti(1991)

A more interesting aspect of the use of correlated error term will be for a structural time series model.

2.4.2 Trend plus Cycle with correlated error

So a basic univariate representation of the output series, y_t , deals with the decomposition into a Local level model component, μ_t and a stochastic cycle C_t

$$y_t = \mu_t + C_t + v_t, \qquad t = 1, 2, ..., T$$

$$\begin{aligned}
\mu_t &= \mu_{t-1} + \omega_t \\
C_t &= \rho(C_{t-1}\cos\lambda_c + C^*_{t-1}\sin\lambda_c) + \kappa_t \\
C^*_t &= \rho(-C_{t-1}\sin\lambda_c + C^*_{t-1}\cos\lambda_c) + \kappa^*_t
\end{aligned}$$

where κ_t and κ^*_t are independent N(0, σ^2_κ) variables.

The property of this model is similar to the property of the previous model but with some adjustment to the issue of correlation between the observation noise and the component of the state equation noise terms at time t, which is in this case defined as

$$S = E\left[v_t * \begin{pmatrix} w_t \\ \kappa_t \\ \kappa^*_t \end{pmatrix}\right], \tag{2.4.10}$$

where S is the covariance between the noise term in the trend and cycle equation and the observation equation. S is an m x q matrix which in this case will be a 1 x 3 matrix since q is the number of noise term in the state equation. The correlation between the noise terms can be briefly summarized as

$$S = \begin{cases} (0, \sigma_{v\kappa}, \sigma_{v\kappa}) & \text{No correlation between the trend and the observation noise terms} \\ (\sigma_{v\omega}, 0, 0), & \text{No correlation between the cycle and the observation noise terms} \\ (\sigma_{v\omega}, \sigma_{v\kappa}, \sigma_{v\kappa}) & \text{Correlation between the trend, Cycle and the observation noise terms} \end{cases}$$

where $\sigma_{v\kappa} = \beta\sigma_v\sigma_\kappa$, $\sigma_{v\omega} = \beta\sigma_v\sigma_\omega$

The variance structure of the state equation is given below as

$$\begin{pmatrix} \omega_t \\ \kappa_t \\ \kappa_t^* \end{pmatrix} \sim \mathbf{NID}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\omega^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{pmatrix}\right]$$

where

$$Q = \begin{pmatrix} \sigma_\omega^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & \sigma_\kappa^2 \end{pmatrix}$$

Q is the variance structure of the state noise and $R = \sigma_v^2$ is the variance of the observation noise. The result obtained by considering only the correlation between the state disturbances and the observation disturbance did not show any useful information as regards to the modified Kalman filter property. This is simply due to the issue of the uniqueness of the solution. Parameter estimation on this basis tends to be obscured or returning results without accurate or non computable standard errors.

Alternatively, will be to assume correlation between the noise terms of the state components only which is indeed very practical as is done in the articles of Manuel & John(2016) and Shigeru & Han (2015). Thus the variance structure can be written as the upper half of the covariance matrix.

$$\begin{pmatrix} \omega_t \\ \kappa_t \\ \kappa_t^* \end{pmatrix} \sim \mathbf{NID}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\omega^2 & \sigma_{\omega\kappa} & \sigma_{\omega\kappa} \\ & \sigma_\kappa^2 & \sigma_{\omega\kappa} \\ & & \sigma_\kappa^2 \end{pmatrix}\right]$$

Where $\sigma_{\omega\kappa} = \beta\sigma_\omega\sigma_\kappa$

The trend and cycle disturbances are allowed to be correlated, with β being the correlation coefficient. Although this might be more complex to model, it definitely will have useful information as to the dependence of the model on correlation.

3 Structural breaks and Outliers

An outlier is an observation which is inconsistent with a model which is seen to be appropriate for a large number of the observations. It can be captured by the use of a dummy explanatory variable which is given the value one, at the time of the outlier and zero elsewhere. Meanwhile a structural break is said to take place when the level of the time series changes in either upward or downward direction usually due to some specific event. It is modeled using a step process whereby the input(explanatory variable), is given the value zero before the time of the event and one after the time of the event. From the article by Koopman(1993), additive outlier and structural change interventions are modelled by using Γu_t and Υu_t respectively from the state and observation equation given below, Whereby explanatory variables can be brought into the model so as to capture exogenous effects and various types of interventions.

$$\begin{aligned} x_t &= \Phi x_{t-1} + \Upsilon u_t + w_t \\ y_t &= x_t + \Gamma u_t + \epsilon_t \end{aligned} \tag{3.0.11}$$

Thus the observation y_t can be written as a regression model with a stochastic local linear trend component μ_t as

$$\begin{aligned} y_t &= \mu_t + \Gamma u_t + \epsilon_t, \qquad t = 1, ..., T. \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_t \\ \beta_t &= \beta_{t-1} + \varsigma_t \end{aligned} \tag{3.0.12}$$

with the trend having a slope β_t, where ω_t and ς_t are mutually uncorrelated white noise processes with variances σ^2_ω and σ^2_ς. u_t is a k x 1 vector of observed explanatory variables and Γ is a corresponding vector of parameters which may be known or unknown(and so will have to be estimated). A regression model using these explanatory variables(inputs), shows significant serial correlation in the residuals and this yields a good fit.

u_t may include deterministic and/or stochastic inputs.

$$S_{0t} = \begin{cases} 1 & \text{if } t = T_0, \\ 0 & elsewhere \end{cases} \tag{3.0.13}$$

$$S_{1t} = \begin{cases} 0 & \text{if } t < T_0, \\ 1 & \text{if } t \geq T_0 \end{cases} \tag{3.0.14}$$

$$S_{2t} = \begin{cases} 1, 2, 3, ...n & \text{if } t \geq T_0, \qquad n = number\ of\ observation\ after\ T_0 \\ 0 & \text{if } t < T_0 \end{cases} \tag{3.0.15}$$

This is a time dependent process where S_{1t}, S_{2t} and S_{0t} are used as the deterministic inputs. T_0 is the time of break and S_{1t}, S_{2t} are the step intervention variable and staircase intervention variable respectively and S_{0t} is the outlier intervention variable. They can be made stochastic by introducing a random noise term. Just as was stated earlier, the inputs can be extended to the observation or to the state equation. For structural breaks, the focus is on the level and slope. Hence it is referred to as level breaks and slope breaks. The inputs are thus applied to the stochastic level and slope equations of the model in (3.0.12). For outliers, the input S_{0t} is introduced into the observation equation as a linear regression. A level break can be thought of as a large disturbance to the trend component resulting in a level shift. In an econometric model, exogenous variables(interventions) are usually regarded as stochastic because they are more flexible.

3.1 Detection of Outliers and Structural breaks

Diagnostic or testing of a time series model for irregularities is normally carried out using the residuals at the final estimates, which are the innovations (ϵ_t). In unobserved components model, other residuals such as the residuals for the level, slope, cycle etc, are available. Auxiliary residuals as described by Harvey & Koopman (1992) is usually preferred to test for an outlier or a structural break at any point in time. The auxiliary residuals are the estimators of the disturbances. The disadvantage of the use of the auxiliary residual as noted by Harvey & Koopman () (1992), is that they are serially correlated, even in a correctly specified model with known parameters. Using the state space model given by the Kalman filter property,

$$y_t = \mu_t + \varepsilon_t, \qquad t = 1, ..., T \tag{3.1.1}$$

$var(\varepsilon_t) = \sigma_\varepsilon^2 I$, $var(y) = \sigma^2 V$, and $E(\mu_t \varepsilon_t) = 0$. If y is a multivariate normal, then the covariance matrix of (y, ε_t) is given as

$$\tilde{\varepsilon} = E(\varepsilon \mid y) = (\sigma_\varepsilon^2/\sigma^2)V^{-1}y \tag{3.1.2}$$

and the unconditional covariance matrix is given as

$$Var(\tilde{\varepsilon}) = (\sigma_\varepsilon^4/\sigma^2)V^{-1}$$

The covariance matrix of y is obtained by

$$Var(y) = E[(\mu + \varepsilon)(\mu + \varepsilon)^{'}] = V$$

where $\tilde{\varepsilon}$ is referred to as the auxiliary residual which is equivalent to the smoothing error obtained by subtracting the smoothed trend from the observation i.e ($\tilde{\varepsilon} = y - \tilde{\mu_s}$), using the Kalman smoother property. $\tilde{\varepsilon}$ is the expectation of the residual ε conditioned on the whole observation and not at time t.

The process of obtaining the covariance of the observation can be very time consuming but this has been implemented in the STAMP package by Harvey & Koopman.

A structural break which results from a shift in the level of the series that is of greater magnitude than might reasonably be expected given the model specification, can be best detected by an outlying value of $\tilde{\omega_t}$.

Now we look at the annual temperature series from 1850 - 2015. A driftless random walk plus noise model(Local level model) with $\hat{\sigma_\varepsilon} = 0.0776$ and $\hat{\sigma_\omega} = 0.0532$ fits the data, where the variance of the slope is insignificantly different from zero and and the slope has a constant value close to zero. The auxiliary residuals for the irregular is plotted in figures (3.19) alongside the observation series. Large irregular auxiliary residuals in 1963 and 1973 indicate outlying value and possible break respectively.

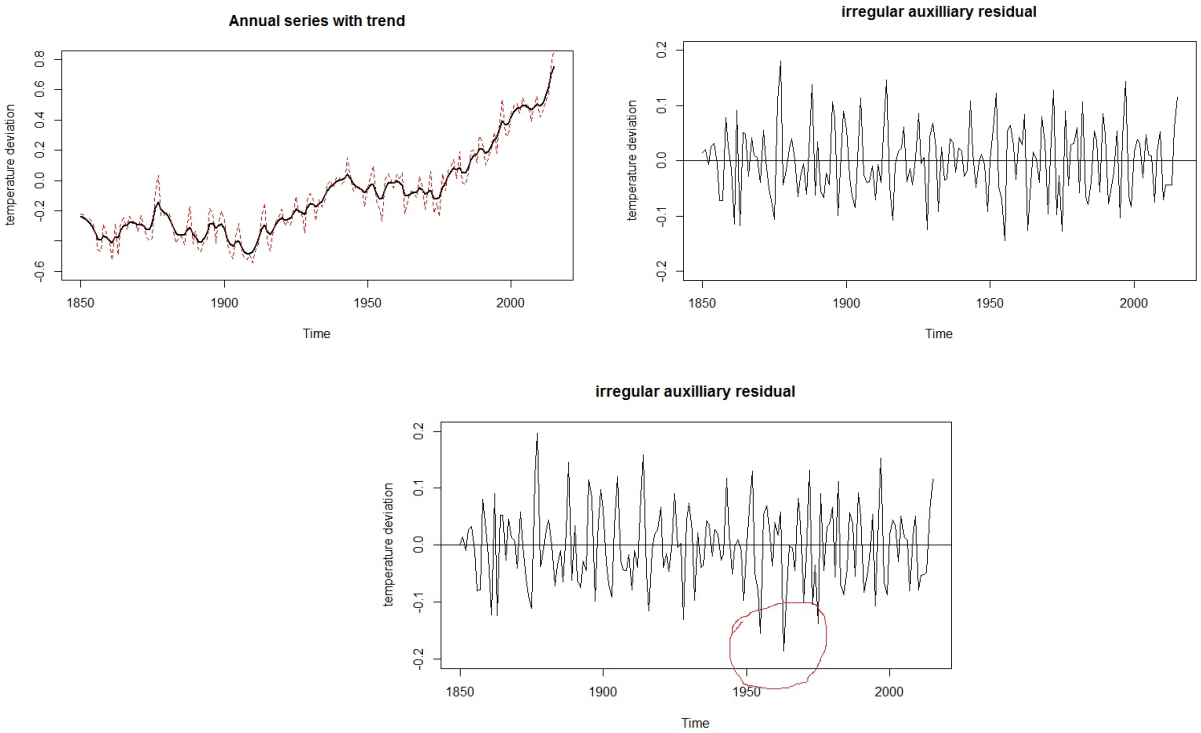

Figure 3.19: (a) Annual temperature series with stochastic trend (b) Irregular auxiliary residual (c) Irregular auxiliary residual with outlier intervention(bottom)

Refitting the model with intervention effect (3.0.14) for outliers in 1963 and a level shift in 1973. The effect of an outlier intervention in figure 3.19(c) for the irregular auxiliary residual shows a strong presence of an outlier in 1963. This can be attributed to the period of extreme cold temperature drop in the northern hemisphere, particularly the regions of Europe and North America. The resulting effect of this local level intervention is seen in the trend plot of figure (3.20), where a large level shift is seen from observed series. This conclusion seems reasonable. A recommendation will be to look closely at the standardized irregular and level auxiliary residuals respectively. As highlighted by Harvey & Koopman (1992), the standardized residuals may indicate outliers and structural change but will not normally give a clear indication as to the source of the problem. Hence certain tests are required as formal procedure to detect unusual large residuals i.e test for excess kurtosis and a test for skewness.

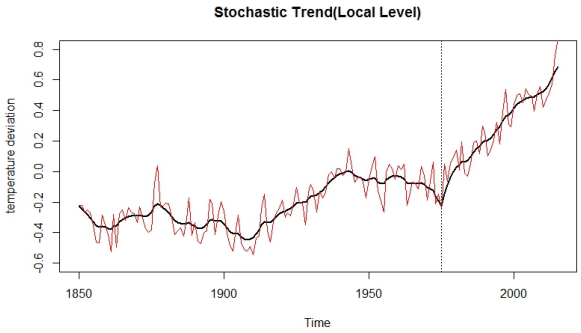

Figure 3.20: Annual temperature series with stochastic trend(local level) having fitted interventions.

4 Forecasting Structural time series

When we talk about forecasting, the goal is to predict future values of a time series y_{n+m}, $m = 1, 2, ...$, which is based on the data collected to the present $y = \{y_n, y_{n-1}, - - - - -, y_1\}$.

In economic research, the focus has been on forecasting techniques such as the use of regression methods i.e ordinary least squares(OLS).

In the state space model, for $t > n$, the predictor or forecast giving by the Kalman filter property x_t^{t-1} is updated, assuming we have the complete history of the process. State space models for forecasting is until date the most preferred method because unlike the other techniques, if the observed data is affected by measurement error or displays a seasonal pattern, In this case, state-space models allow the separation of the true underlying signal from the measurement noise. As only the signal is relevant for prediction, this can considerably improve the quality of the forecast.

Hence, given $\{x_1, - - - - -, x_n\}$, we want to forecast the value of the time series at the next time point x_{n+1}. This process is called the one step ahead prediction and this is obtained from the Kalman filter property.

For $t > n$

$$x_{n+1}^n = \Phi x_n^n + \Upsilon u_{n+1} \tag{4.0.3}$$

and the mean square one step ahead prediction error is given as

$$p_{n+1}^n = \Phi p_n^n \Phi^{'} + Q \tag{4.0.4}$$

with initial conditions x_n^n & p_n^n, which is the filtered values at $t = n$.

Durbin & Koopman (2001), demonstrates that the forecasts for the observation value y_{n+j}, for $j = 2, - - - - - -, J$ is generated by treating $y_{n+1}, - - - - -, y_{n+J}$ as missing values, which results in the similar one step ahead state prediction given above but under the condition that the innovation and its covariance are zero, i.e $\epsilon_{n+j} = 0$ and $\Sigma_{n+j} = \text{var}(\epsilon_t) = 0$ for $j = 1, - - - - -, J - 1$. These are the conditions used by Durbin & Koopman (2001) to handle missing observations in a time series.

Long run monthly forecasts using Local level model produces constant predicted values for the observation data set. The reason is, for a forecasted series, the level is constant, thus repeating the value of past predicted state, which in turn removes the possibility of future global warming

from the forecast. Thus agreeing with the conclusion of Mills(2010).

$$\begin{aligned} \hat{y}_t &= \hat{\mu}_t \\ \hat{\mu}_t &= \hat{\mu}_{t-1} + \omega_t \end{aligned} \tag{4.0.5}$$

where $\omega_t = 0$

An improvement to this is considering the Local linear trend model. Although the estimated slope is constant and has very small value close to zero , the model can still be seen as imposing a near linear forecast on the monthly global temperature series, hence a future global warming trend might be deduced from this. as will be shown in figure 4.21(b) .
For $t > n$,

$$\begin{aligned} \hat{y}_t &= A_t \hat{\mu}_t \\ \hat{\mu}_t &= \hat{\mu}_{t-1} + \hat{\beta}_{t-1} + \omega_t \\ \hat{\beta}_t &= \hat{\beta}_{t-1} + \eta_t \end{aligned} \tag{4.0.6}$$

where $\omega_t = 0$, $\eta_t = 0$

Using a local linear trend model, we compare the forecast for the annual and monthly temperature series with n ahead steps = 32 years(i.e forecast between 2016 - 2050)

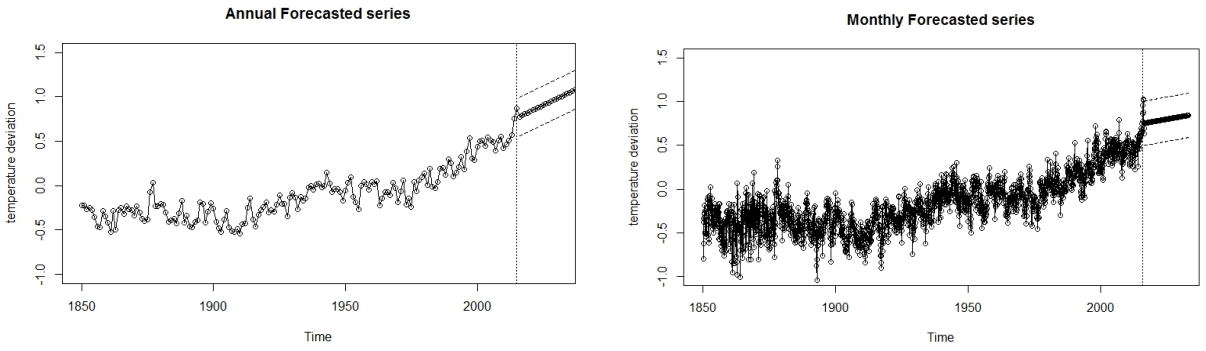

Figure 4.21: (a) Annual Forecast temperature series (left) (b) Monthly Forecast temperature series(right). The forecasts are shown as a continuation of the data. The dashed lines indicate the upper and lower 95 % prediction intervals

Since the estimated slope is none zero and constant, we get a random walk model with drift which does not impose a constant forecast on the annual global temperature series with significant

estimated slope $\hat{\beta} = 0.00601$, root mean squared error of 0.0038 and likelihood $L = -285.8525$ of the model compared to the random walk with drift trend model with an insignificant estimated slope of $\hat{\beta} = 0.00061$, the root mean square error of 0.0012 and likelihood $L = -3133.864$, which also imposes a near linear forecast on the monthly global temperature series. The forecasts can be extended to the structural models for both the annual and monthly global temperatures. The results obtained from this can be compared to that of Mills(2010) and the Hadley centre's HadCM2 and HadCM3 models.

4.1 Monthly local level trend , cycle plus AR model

As we seen before there is a major difference that exists in forecasting between the annual global temperature series and the monthly global temperature series. A significant difference present is contained in the trend component where the slope estimate shows an increasing temperature for the forecast period for the annual temperature series and a near constant temperature forecast for monthly temperature series. Either way, for forecasts using a structural model, the main attention is on the trend since an increase in the temperature is only indicated here. A structural time series model with the trend, cycle and an autoregressive component as the unobserved component is fitted to the observed annual global temperature data. The focus is on predicting future data based on the Kalman filter property of the state space model. Specifications of the trend, cycle and autoregressive components remains the same as given in previous section, hence a plot of the forecast for the year 2050 is given below.

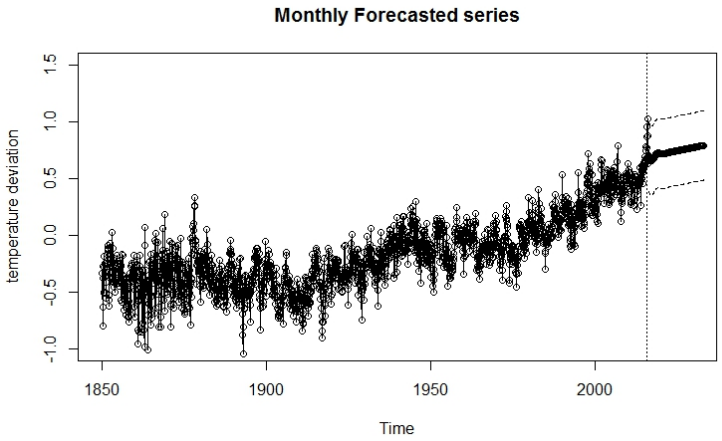

Figure 4.22: Monthly Forecast temperature series. The forecasts are shown as a continuation of the data. The dashed lines indicate the upper and lower 95 % prediction intervals

The Long run forecast as seen for the monthly temperature data, still shows that the random walk trend with drift imposes a linear forecast on the monthly temperature series. Describing an increase in the temperature forecast can be captured by the use of estimated slopes in the trend component, since the estimated slope will only show an increasing temperature forecast which although is what so many researchers hope for, but is not necessarily the case. Hence the use of this class of structural model for predicting future temperature increase for monthly temperature series seems unrealistic and the same conclusion was also arrived at by Mills(2010).

For very large data, the Kalman filter and smoother tends to give unsatisfactory results , which is due to high possibility of correlation in the data and also the presence of Influential points which are more likely to occur in clusters in large samples and therefore more difficult to identify.

5 Conclusion

This work spans the use of breaking trend regression model to fit a global temperature data with the purposes of estimating the slopes which show how much increase in temperature has been recorded from the initial time of data collection until present. Investigations and diagnostic is carried out by checking for correlations in the residuals and also checking for periodicity in the residuals. The results of this investigation shows significant autocorrelation in the residuals indicating the presence of underlying noise terms which is not accounted for. By treating the residual as an autoregressive moving average (ARMA) process whereby we can obtain its spectral density, the result from the parametric spectral estimate shows underlying periodic patterns for both annual and monthly data, thus, leads to a discussion on the need to treat global temperature data as a structural time series model. Structural time series models are set up as a regression model where explanatory variables are functions of time. They offer a generalized and useful approach to the estimation of unobserved components in an observed time series. We select appropriate models by considering the goodness of fit of the model by comparing the AIC values. Parameters are estimated and accompanied with some measures of precision.

An important aspect of fitting structural models, is the underlying changes of the unobserved trend component. Unobserved components are made stochastic by introducing a random noise process which are Gaussian distributed with zero mean and variance σ^2. The noise terms(error terms) are assumed to be independent and uncorrelated of one another and also with the error term in the observation . The current/or filtered estimate of the trend is obtained by putting the model in state space form and applying the Kalman filter algorithm. Predictions and Smoothing of the estimated trend is also done by applying the Kalman filter and smoother algorithm respectively for $t > n$ and $t = n$, where $t = 1, ..., n$. This assures computing the best estimate of the trend at all points in the sample using the full observations. Of importance is also the unobserved components such as seasonality, cyclical pattern and autoregressive process. Although seasonality is usually treated as deterministic, allowing the seasonal pattern to evolve over time by the use of stochastic seasonality is another way to treat the seasonal component although not very effective since the noise uncertainty of the seasonal component goes to zero hence making it deterministic. Another effective way of treating the seasonal component was to introduce it (centered seasonal average) as an exogenous variable or intervention effect into the state space model. The effect is seen in the autocorrelation function of the residual, where the dependence on serial correlation is reduced. Treating the error terms by assuming correlation between them

is another way to estimate unobserved components using state space model. Although Shumway & Stoffer(2011) used this concept for fitting ARMAX models and regression(multivariate series) models with autocorrelated errors, this can also be applied to fitting a univariate series. The shortcoming encountered when we allowed for correlation of the state and observation error is that for structural models, correlation between unobserved components was not taken into account. Tests for outliers and structural breaks are very important for a structural time series framework. visible unusual large values were detected in the irregular noise term, which represents the presence of outliers in the stochastic trend(local level) model considered. An extension of this would be to do the same for models with more unobserved components. We saw that introducing intervention effects to the model helps in inducing the presence of an outlier or structural break.

A number of of interesting features of the annual and monthly global temperature series has been uncovered by fitting structural models, of most interest is the trend component, since this features in debate over global warming. The use of structural models where the Local linear trend component is characterized by a random walk for predicting future temperature increase only shows positive result on the annual global temperature series but retains a near linear forecast on the monthly global temperature series which for large n.ahead steps will only produce constant forecasts. The reason for this is seen as the estimated slope for the annual global temperature series has significantly non zero value as compared to the estimated slope for the monthly global temperature series which is insignificantly close to zero.

6 Recommendation

For the case of state space models with correlated errors, further work can be done here by taking into consideration correlation between error terms in the state equation when considering structural models of the form trend , seasonal, cycle and an autoregressive process. This type of result are important since they show how well the Kalman filter property for correlated errors carries out filtering of the unobserved component as compared to the case when the errors are not correlated.

Bootstrapping is also of interest for state space models. Since the Maximum likelihood estimators of the parameters of the state space models are asymptotically normal, comparing the difference of the standard errors obtained by bootstrap method and that of the asymptotic methods might help give insight into the data i.e presence of outliers in the data. The bootstrap, however, is used to check the accuracy of the estimates by examining the bootstrap interval.

7 Appendix

The Kalman Filter

$$x_t^{t-1} = E(x_t|Y_{t-1}) = E(\Phi x_{t-1}^{t-1} + \Upsilon u_t + \omega_t|Y_{t-1}) = \Phi x_{t-1}^{t-1} + \Upsilon u_t \tag{7.0.1}$$

and

$$\begin{aligned} P_t^{t-1} =& E\{(x_t - x_t^{t-1})(x_t - x_t^{t-1})^{'}\} \\ =& E\{[\Phi(x_{t-1} - x_{t-1}^{t-1}) + \omega_t][\Phi(x_{t-1} - x_{t-1}^{t-1}) + \omega_t]^{'}\} \\ =& \Phi P_{t-1}^{t-1}\Phi^{'} + Q \end{aligned}$$

The by products of the filter are the innovation(prediction error) ϵ_t and its variance-covariance matrices Σ_t

$$\begin{aligned} \epsilon_t =& y_t - E(y_t|Y_{t-1}) = y_t - A_t x_t^{t-1} - \Gamma u_t \\ \Sigma_t = var(\epsilon_t) =& var[y_t - A_t x_t^{t-1} - \Gamma u_t] \\ =& var[A_t(x_t - x_t^{t-1}) + v_t] = A_t P_t^{t-1} A_t^{'} + R, \qquad t = 1, ..., n \end{aligned}$$

Where x_t^{t-1} and P_t^{t-1} are the one step ahead state prediction/forecast and covariance error respectively. To obtain the filter state, we note that $E(\epsilon_t y_s^{'}) = 0$ for $s < t$, which implies that the innovations are independent of the past observations.

The conditional covariance between x_t and ϵ_t given Y_{t-1} is

$$\begin{aligned} cov(x_t, \epsilon_t|Y_{t-1}) =& cov(x_t, y_t - A_t x_t^{t-1} - \Gamma u_t|Y_{t-1}) \\ =& cov(x_t - x_t^{t-1}, y_t - A_t x_t^{t-1} - \Gamma u_t|Y_{t-1}) \\ =& cov[x_t - x_t^{t-1}, A_t(x_t - x_t^{t-1}) + v_t] \\ =& P_t^{t-1} A_t^{'} \end{aligned}$$

Thus we can write down the joint distribution of x_t and ϵ_t given Y_{t-1} is normal.

$$\begin{pmatrix} x_t \\ \epsilon_t \end{pmatrix} \mid Y_{t-1} \sim N\left(\begin{bmatrix} x_t^{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} P_t^{t-1} & P_t^{t-1}A_t^{'} \\ A_t P_t^{t-1} & \Sigma_t \end{bmatrix} \right)$$

We arrive the expression for the Kalman filter for the state equation, by applying the property of the a normal distribution

$$E(x|z) = \mu_x + \Sigma_{xz}\Sigma_{zz}^{-1}(z - \mu_z)$$

$$\begin{aligned} x_t^t = E(x_t|y_1, ..., y_{t-1}, y_t) =& E(x_t|Y_{t-1}, \epsilon_t) \\ =& x_t^{t-1} + K_t\epsilon_t \end{aligned}$$

where $K_t = P_t^{t-1}A_t^{'}\Sigma_t^{-1} = P_t^{t-1}A_t^{'}(A_t P_t^{t-1}A_t^{'} + R)^{-1}$

Using the property of the normal distribution, the covariance error is given as

$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$

$$P_t^t = cov(x_t|Y_{t-1}, \epsilon_t) = P_t^{t-1} - P_t^{t-1}A_t^{'}\Sigma_t^{-1}A_t P_t^{t-1}$$

Note that this filter property is for time invariant parameters.

The Kalman Smoother

$Y_{t-1} = \{y_1, ..., y_{t-1}\}$ and $\eta_t = \{v_t, ..., v_n, \omega_{t+1}, ..., \omega_n\}, \qquad t = n, n-1, ..., 1$

where Y_0 is empty, let

$q_{t-1} = E\{x_{t-1}|Y_{t-1}, x_t - x_t^{t-1}, \eta_t\}$

Since $Y_{t-1}, x_t - x_t^{t-1}, \eta_t$ are mutually independent, and x_{t-1} and η_t are independent, by applying the property of the a normal distribution, we get

$q_{t-1} = x_{t-1}^{t-1} + J_{t-1}(x_t - x_t^{t-1})$

where,

$J_{t-1} = cov(x_{t-1}, x_t - x_{t-1})[P_t^{t-1}]^{-1} = P_{t-1}^{t-1}\Phi^{'}[P_t^{t-1}]^{-1}$

Since $Y_{t-1}, x_t - x_t^{t-1}, \eta_t$ generate $Y_n = y_1, ..., y_n$,

$$x_{t-1}^n = E\{x_{t-1}|Y_n\} = E\{q_{t-1}|Y_n\} = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \tag{7.0.2}$$

The error covariance P_{t-1}^n is obtained by a straight forward calculation. Moving both sides of equation (7.0.2),and adding x_{t-1} to both sides,

$$x_{t-1} - x_{t-1}^n = x_{t-1} - x_{t-1}^{t-1} - J_{t-1}(x_t^n - \Phi x_{t-1}^{t-1})$$
$$(x_{t-1} - x_{t-1}^n) + J_{t-1}x_t^n = (x_{t-1} - x_{t-1}^{t-1}) + J_{t-1}\Phi x_{t-1}^{t-1})$$

Next we multiply each side by the transpose of itself and take the expectation,

$$P_{t-1}^n + J_{t-1}E(x_t^n x_t^{n'})J_{t-1}' = P_{t-1}^{t-1} + J_{t-1}\Phi E(x_{t-1}^{t-1}x_{t-1}^{t-1'})\Phi' J_{t-1}' \tag{7.0.3}$$

Noting that from the expression of P_t^{t-1} the cross product $E[x_t x_t^{t-1}] = 0$

$E[x_t^n x_t^{n'}] = E[x_t x_t'] - P_t^n = \Phi E(x_{t-1}x_{t-1}')\Phi' + Q - P_t^n$

and

$E(x_{t-1}^{t-1}x_{t-1}^{t-1'}) = E(x_{t-1}x_{t-1}') - P_{t-1}^{t-1}$ Hence substituting into (7.0.3), we get

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \tag{7.0.4}$$

8 Source Code

```
> Kfilter0
function (num, y, A, mu0, Sigma0, Phi, cQ, cR)
{
    Q = t(cQ) %*% cQ
    R = t(cR) %*% cR
    Phi = as.matrix(Phi)
    pdim = nrow(Phi)
    y = as.matrix(y)
    qdim = ncol(y)
    xp = array(NA, dim = c(pdim, 1, num))
    Pp = array(NA, dim = c(pdim, pdim, num))
    xf = array(NA, dim = c(pdim, 1, num))
    Pf = array(NA, dim = c(pdim, pdim, num))
    innov = array(NA, dim = c(qdim, 1, num))
    sig = array(NA, dim = c(qdim, qdim, num))
    x00 = as.matrix(mu0, nrow = pdim, ncol = 1)
    P00 = as.matrix(Sigma0, nrow = pdim, ncol = pdim)
    xp[, , 1] = Phi %*% x00
    Pp[, , 1] = Phi %*% P00 %*% t(Phi) + Q
    sigtemp = A %*% Pp[, , 1] %*% t(A) + R
    sig[, , 1] = (t(sigtemp) + sigtemp)/2
    siginv = solve(sig[, , 1])
    K = Pp[, , 1] %*% t(A) %*% siginv
    innov[, , 1] = y[1, ] - A %*% xp[, , 1]
    xf[, , 1] = xp[, , 1] + K %*% innov[, , 1]
    Pf[, , 1] = Pp[, , 1] - K %*% A %*% Pp[, , 1]
    sigmat = as.matrix(sig[, , 1], nrow = qdim, ncol = qdim)
    like = log(det(sigmat)) + t(innov[, , 1]) %*% siginv %*%
        innov[, , 1]
    for (i in 2:num) {
        if (num < 2)
```

```
            break
        xp[, , i] = Phi %*% xf[, , i - 1]
        Pp[, , i] = Phi %*% Pf[, , i - 1] %*% t(Phi) + Q
        sigtemp = A %*% Pp[, , i] %*% t(A) + R
        sig[, , i] = (t(sigtemp) + sigtemp)/2
        siginv = solve(sig[, , i])
        K = Pp[, , i] %*% t(A) %*% siginv
        innov[, , i] = y[i, ] - A %*% xp[, , i]
        xf[, , i] = xp[, , i] + K %*% innov[, , i]
        Pf[, , i] = Pp[, , i] - K %*% A %*% Pp[, , i]
        sigmat = as.matrix(sig[, , i], nrow = qdim, ncol = qdim)
        like = like + log(det(sigmat)) + t(innov[, , i]) %*%
            siginv %*% innov[, , i]
    }
    like = 0.5 * like
    list(xp = xp, Pp = Pp, xf = xf, Pf = Pf, like = like, innov = innov,
        sig = sig, Kn = K)
}
<environment: namespace:astsa>

> Ksmooth0
function (num, y, A, mu0, Sigma0, Phi, cQ, cR)
{
    kf = astsa::Kfilter0(num, y, A, mu0, Sigma0, Phi, cQ, cR)
    pdim = nrow(as.matrix(Phi))
    xs = array(NA, dim = c(pdim, 1, num))
    Ps = array(NA, dim = c(pdim, pdim, num))
    J = array(NA, dim = c(pdim, pdim, num))
    xs[, , num] = kf$xf[, , num]
    Ps[, , num] = kf$Pf[, , num]
    for (k in num:2) {
        J[, , k - 1] = (kf$Pf[, , k - 1] %*% t(Phi)) %*% solve(kf$Pp[,
            , k])
```

```
        xs[, , k - 1] = kf$xf[, , k - 1] + J[, , k - 1] %*% (xs[,
            , k] - kf$xp[, , k])
        Ps[, , k - 1] = kf$Pf[, , k - 1] + J[, , k - 1] %*% (Ps[,
            , k] - kf$Pp[, , k]) %*% t(J[, , k - 1])
    }
    x00 = mu0
    P00 = Sigma0
    J0 = as.matrix((P00 %*% t(Phi)) %*% solve(kf$Pp[, , 1]),
        nrow = pdim, ncol = pdim)
    x0n = as.matrix(x00 + J0 %*% (xs[, , 1] - kf$xp[, , 1]),
        nrow = pdim, ncol = 1)
    P0n = P00 + J0 %*% (Ps[, , k] - kf$Pp[, , k]) %*% t(J0)
    list(xs = xs, Ps = Ps, x0n = x0n, P0n = P0n, J0 = J0, J = J,
        xp = kf$xp, Pp = kf$Pp, xf = kf$xf, Pf = kf$Pf, like = kf$like,
        Kn = kf$K)
}
<environment: namespace:astsa>

> Kfilter2
function (num, y, A, mu0, Sigma0, Phi, Ups, Gam, Theta, cQ, cR,
    S, input)
{
    Q = t(cQ) %*% cQ
    R = t(cR) %*% cR
    Phi = as.matrix(Phi)
    pdim = nrow(Phi)
    y = as.matrix(y)
    qdim = ncol(y)
    rdim = ncol(as.matrix(input))
    if (max(abs(Ups)) == 0)
        Ups = matrix(0, pdim, rdim)
    if (max(abs(Gam)) == 0)
        Gam = matrix(0, qdim, rdim)
```

```
ut = matrix(input, num, rdim)
xp = array(NA, dim = c(pdim, 1, num))
Pp = array(NA, dim = c(pdim, pdim, num))
xf = array(NA, dim = c(pdim, 1, num))
Pf = array(NA, dim = c(pdim, pdim, num))
Gain = array(NA, dim = c(pdim, qdim, num))
innov = array(NA, dim = c(qdim, 1, num))
sig = array(NA, dim = c(qdim, qdim, num))
like = 0
xp[, , 1] = Phi %*% mu0 + Ups %*% as.matrix(ut[1, ], rdim)
Pp[, , 1] = Phi %*% Sigma0 %*% t(Phi) + Theta %*% Q %*% t(Theta)
for (i in 1:num) {
    B = matrix(A[, , i], nrow = qdim, ncol = pdim)
    innov[, , i] = y[i, ] - B %*% xp[, , i] - Gam %*% as.matrix(ut[i,
        ], rdim)
    sigma = B %*% Pp[, , i] %*% t(B) + R
    sigma = (t(sigma) + sigma)/2
    sig[, , i] = sigma
    siginv = solve(sigma)
    Gain[, , i] = (Phi %*% Pp[, , i] %*% t(B) + Theta %*%
        S) %*% siginv
    K = as.matrix(Gain[, , i], nrow = qdim, ncol = pdim)
    xf[, , i] = xp[, , i] + Pp[, , i] %*% t(B) %*% siginv %*%
        innov[, , i]
    Pf[, , i] = Pp[, , i] - Pp[, , i] %*% t(B) %*% siginv %*%
        B %*% Pp[, , i]
    sigma = matrix(sigma, nrow = qdim, ncol = qdim)
    like = like + log(det(sigma)) + t(innov[, , i]) %*% siginv %*%
        innov[, , i]
    if (i == num)
        break
    xp[, , i + 1] = Phi %*% xp[, , i] + Ups %*% as.matrix(ut[i +
        1, ], rdim) + K %*% innov[, , i]
```

```
        Pp[, , i + 1] = Phi %*% Pp[, , i] %*% t(Phi) + Theta %*%
            Q %*% t(Theta) - K %*% sig[, , i] %*% t(K)
    }
    like = 0.5 * like
    list(xp = xp, Pp = Pp, xf = xf, Pf = Pf, K = Gain, like = like,
        innov = innov, sig = sig)
}
<environment: namespace:astsa>

> Ksmooth2
function (num, y, A, mu0, Sigma0, Phi, Ups, Gam, Theta, cQ, cR,
    S, input)
{
    kf = astsa::Kfilter2(num, y, A, mu0, Sigma0, Phi, Ups, Gam,
        Theta, cQ, cR, S, input)
    pdim = nrow(as.matrix(Phi))
    xs = array(NA, dim = c(pdim, 1, num))
    Ps = array(NA, dim = c(pdim, pdim, num))
    J = array(NA, dim = c(pdim, pdim, num))
    xs[, , num] = kf$xf[, , num]
    Ps[, , num] = kf$Pf[, , num]
    for (k in num:2) {
        J[, , k - 1] = (kf$Pf[, , k - 1] %*% t(Phi)) %*% solve(kf$Pp[,
            , k])
        xs[, , k - 1] = kf$xf[, , k - 1] + J[, , k - 1] %*% (xs[,
            , k] - kf$xp[, , k])
        Ps[, , k - 1] = kf$Pf[, , k - 1] + J[, , k - 1] %*% (Ps[,
            , k] - kf$Pp[, , k]) %*% t(J[, , k - 1])
    }
    list(xs = xs, Ps = Ps, J = J, xp = kf$xp, Pp = kf$Pp, xf = kf$xf,
        Pf = kf$Pf, like = kf$like, Kn = kf$K)
}
<environment: namespace:astsa>
```

BIBLIOGRAPHY

References

Textbooks:

Robert H. Shumway, David S. Stoffer (2011)third edition, Time series analysis and its applications: With R examples.Springer Science + Business Media, LLC 233, New York.

Durbin .J and S.J. Koopman (2001) Time Series Analysis by State Space Methods, Oxford University Press.

Harvey C. Andrew (1989) Forecasting, structural time series models and the kalman filter. Cambridge university press USA.

Related works

Terence C. Mills(2010) 'Skinning a cat': alternative models of representing temperature trends. Climatic Change 101:415-426.

Mills TC(2004), Time series modelling of trends in Northern Hemisphere average temperature series. Energy Environ 15: 743-753

Kaufmann RK, Kauppi H, Stock JH (2006b), The relationship between radiative foring and temperature: what do statistical analyses of the instrumental temperature record measure? Climate change 77:279-289

Gay-Garcia C, Estrada F, Sanchez A (2009) Global and hemispheric temperatures revisited. Climatic change 94:333-349.

Carvalho Vasco, Harvey Andrew, Trimbur Thomas (2007), A Note on Common Cycles, Common Trends,and Convergence. Journal of Business and Economic Statistics

Tommaso Proietti (2002), State space decomposition under the hypothesis of non zero correlation between trend and cycle, with an application to the euro-zone.Dept. of Economics, European University Institute

Tommaso Proietti (1991), Forecasting with Structural Time Series Models. Blackwell Publishers, Oxford.

Harvey .A, S.J Koopman and J. Penzer(1997), MESSY TIME SERIES: A UNIFIED APPROACH. Advances in Econometrics, Volume 13 (1998), pages 103-143.

Harvey .A.C, S.J Koopman (1992), Diagnostic checking of unobserved-components time series models. Journal of business & Economics statistics, vol. 10, No. 4(oct.,1992), pp.377-389

Harvey .A.C, T.M Trimbur and H.K. Van Dijk (2004), Bayes Estimates of the Cyclical Component in twentieth century US gross domestic product. Econometric Institute Report EI 2004-45.

J. Casals and S. Sotoca (2001), The Exact Likelihood for a State Space Model with Stochastic Inputs. Computers and Mathematics with Applications 42 (2001) 199-209.

M.H. Ambrosio and D.J. Dias (2011), Measuring the Cyclical Component of a Time Series: Proposal of a New Methodology. Brazilian Review of Econometrics vol. 30, no.1, pp. 91âĂŞ122.

Lili M., Wang H.A and J. Chen (2010), Analysis of Kalman Filter with Correlated Noises under Different Dependence. Journal of Information & Computational Science 7: 5 (2010) 1147-1154.

Koopman .S.J (1993), Disturbance Smoother for State Space Models. Biometrika, Vol. 80, No. 1 (Mar., 1993), pp. 117-126. Oxford University Press on behalf of Biometrika Trust. http://www.jstor.org/stable/2336762

Koopman S.J. and Harvey A. (2003), Computing observation weights for signal extraction and filtering. Journal of Economic Dynamics & Control 27 (2003) 1317 âĂŞ 1333. 2002

James D. Hamilton (1994), STATE-SPACE MODELS. Handbook of Econometrics, Volume IV.

Robert Kohn and Craig F. Ansley (1989), A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models. Biometrika, Vol. 76, No. 1 (Mar., 1989), pp. 65-79. Oxford University Press on behalf of Biometrika Trust.

Erik BÃÿlviken, Nils Christophersen, and Geir Storvik (1998). Linear dynamical models, Kalman filtering and statistics. Lecture notes to IN-ST 259. University of Oslo, October 1998.

Fernando Tusell (2011), Kalman Filtering in R. Journal of Statistical Software, March 2011, Volume 39, Issue 2.

Giovanni Petris, and Sonia Petrone (2011), State Space Models in R. Journal of Statistical Software, May 2011, Volume 41, Issue 4.

Manuel Gonzalez-Astudillo and John M. Roberts (2016), When Can Trend-Cycle Decompositions Be Trusted? Finance and Economics Discussion Series 2016-099. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2016.099.

Shigeru Iwata and Han Li (2015), What are the Differences in Trend Cycle Decompositions by Beveridge and Nelson and by Unobserved Component Models? Vol. 34 , Iss. 1-2,2015 CopyrightÂĩ Taylor & Francis Group, LLC

Web Page Sources

HARDCRUT 4: http://www.metoffice.gov.uk/hadobs/hadcrut4/

National aeronautics and space administration (NASA) (2016), url-http://data.National aeronautics and space administration (NASA) (2016), url-http://data.giss.nasa.gov/gistemp/station/data/

Calculating Covariance For Stocks | Investopedia http://www.investopedia.com/articles/financial-theory/11/calculating-covariance