TS4

Dylan Hayashi

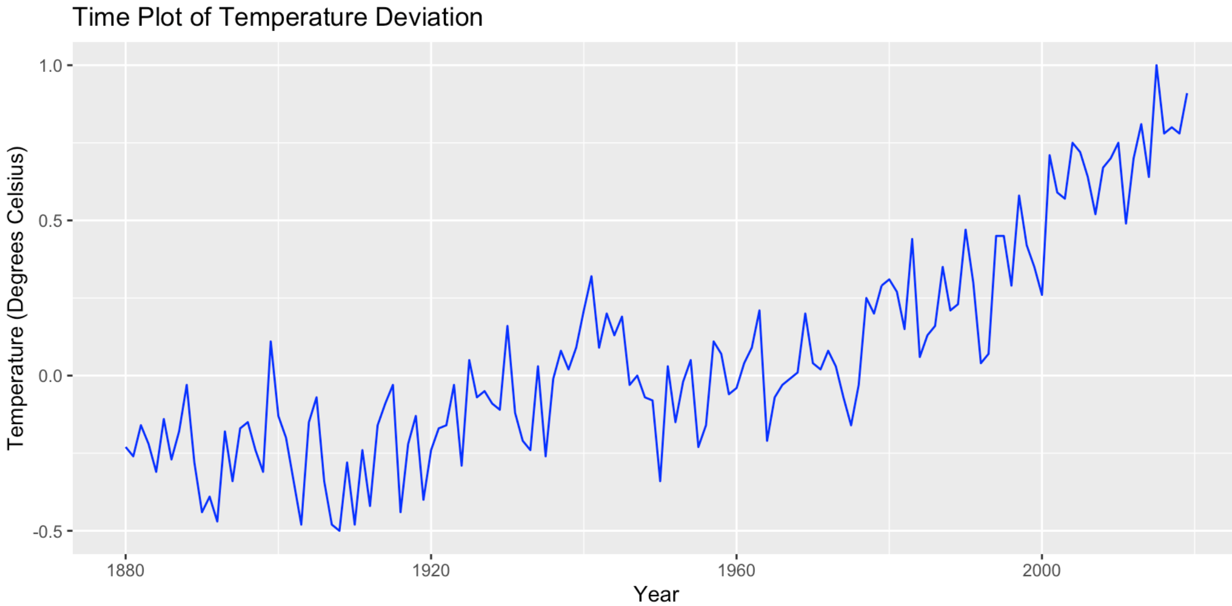
MSDS 413, Fall 2022, Section 55

Northwestern University, Time-Series Analysis & Forecasting

October 17, 2022

1. EDA

Below is a time plot of the temperature data with years on the x-axis and deviation from the base temperature on the y-axis.



Prior to conducting analysis, confirmation that this data is, in fact, time series data is required. In order to fit the definition, the data must be a time ordered sequence of observations of a stochastic variable over constant time intervals.

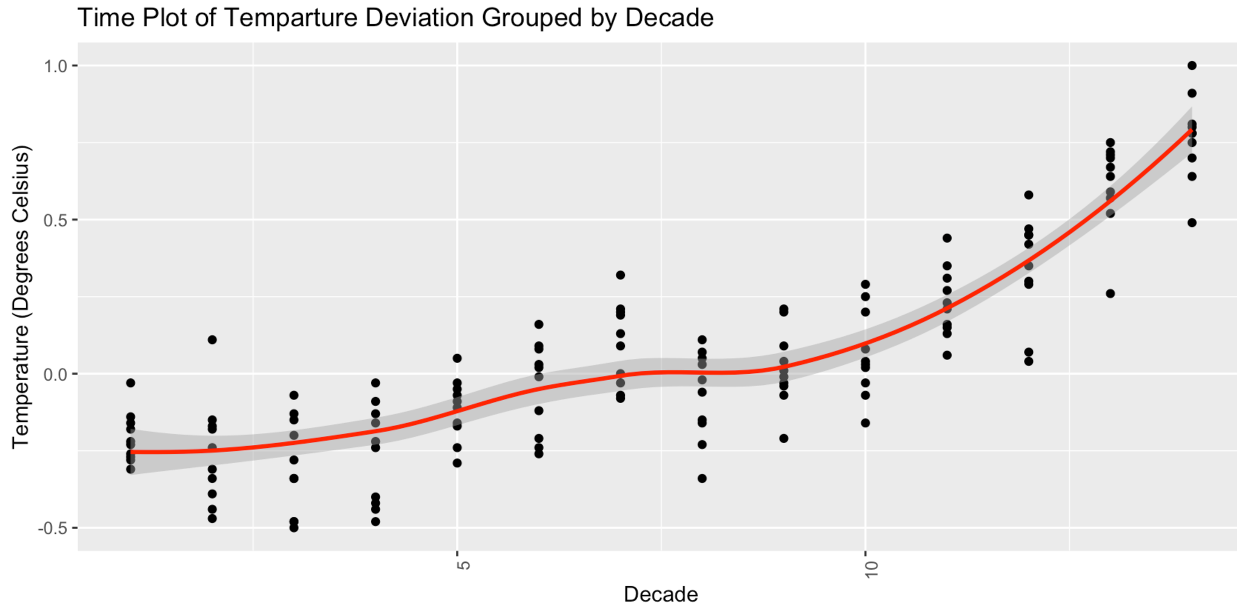
To test that this is a time ordered sequence, we can confirm that the data is indexed by unique time periods and that there is an observation of temperature for each time period. The vectors Year, unique(Year), and Value all have 140 observations, confirming that this is the case.

To test for constant intervals between time periods, we can sum the difference between each pair of successive periods, and this should equal one less than the total number of observations. The sum of these differences was 139, one less than the index length of 140, confirming that the dataset has constant time intervals.

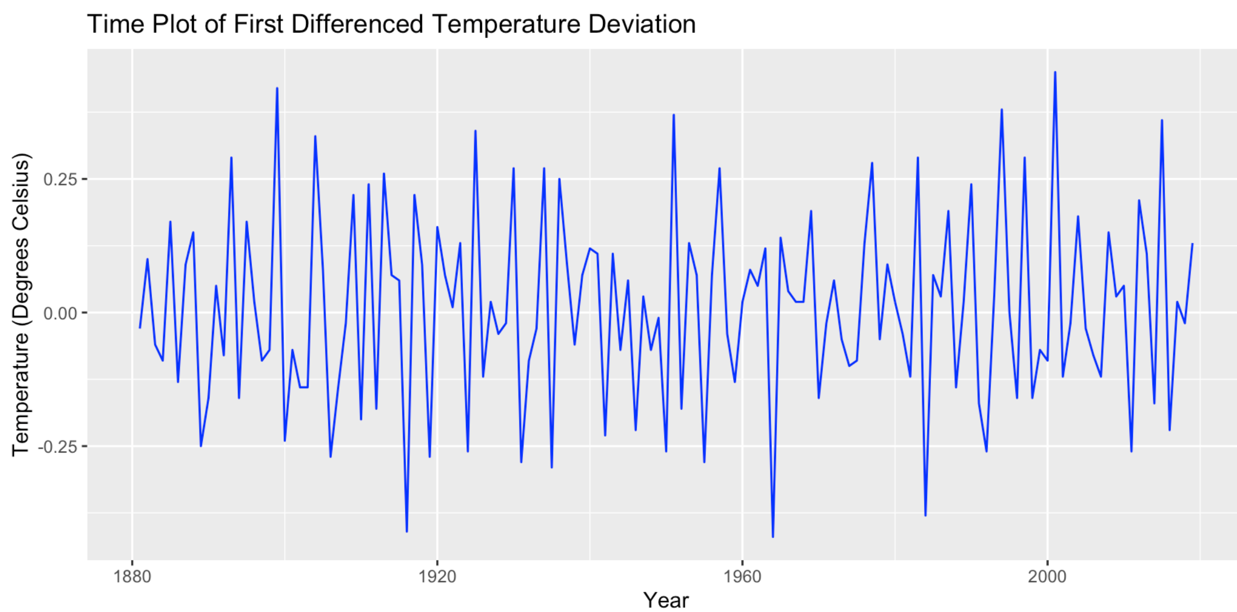
The data summary on the right, which is of the Value variable, shows that the variable has variance. Having variance is indicative of stochasticity. This confirms that the data match the definition of time-series in its entirety.

nobs	141.000000
NAs	0.000000
Minimum	-0.500000
Maximum	1.000000
1. Quartile	-0.170000
3. Quartile	0.250000
Mean	0.068865
Median	0.000000
Sum	9.710000
SE Mean	0.029423
LCL Mean	0.010694
UCL Mean	0.127036
Variance	0.122066
Stdev	0.349379
Skewness	0.716895
Kurtosis	-0.181891

The time plot above shows an overall level change, which is indicative of a linear trend. This is not confirmed by performing an SLR of Value on Year, which after adjusting year for starting in 1880, produced a t-test 95% confidence interval of (-3.057, 2.624) and a p-value of 0.51. The interval and p-value are sufficient to show failure to reject the null hypothesis that B_1 is not 0. Despite this, I am still going to perform first differencing, as I believe there is a linear trend based on the plot. I am also inclined to believe that the range of Value (-0.5, 1) may make it hard to statistically distinguish the coefficient B_1 from 0. A linear trend is also suggested by an upward sloping line of best fit when the data points are organized by decade, as shown below.



Applying first differencing produced data that takes the following form as a time plot:

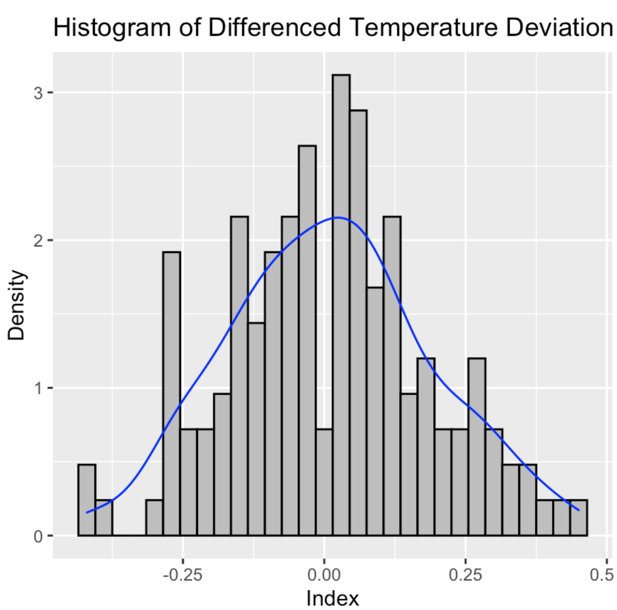


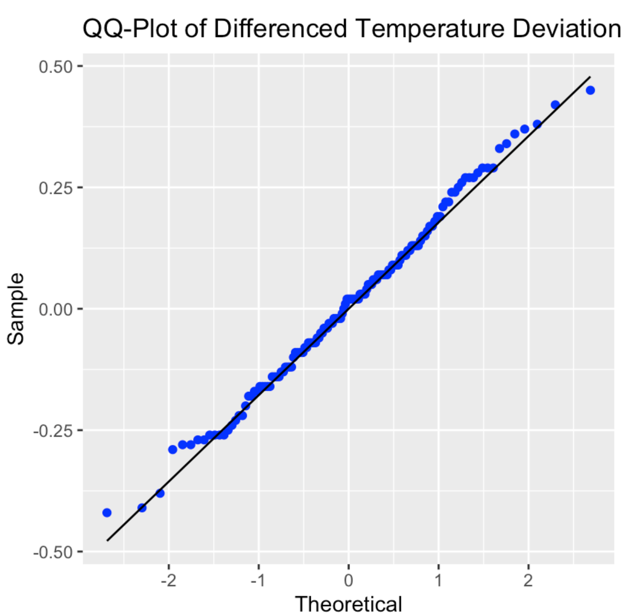
The new mean of the differenced temperature variable is 0.0082 (shown in the data summary on the next page) with a t.test 95% confidence interval of (-0.022, 0.038) and p-value of 0.59. This p-value and confidence interval indicate that we fail to reject the null hypothesis that the mean is 0, and we can conclude that the mean is statistically indistinguishable from 0.

The first differenced data are stationary. As previously stated, the mean of the data is zero. An ADF test substantiates linear-trend and drift stationarity, as it returns a p-value of 0.01, allowing us to reject the null hypothesis of non-stationarity.

The first differenced data are also normally distributed. The histogram on the left below shows a unimodal distribution with decent symmetry. The QQ plot shows that almost all of the data lies on the normal line. While both skew and excess kurtosis are quite close to 0 (0.097 and -0.339), neither of their 95% confidence intervals include zero: (0.095, 0.114) and (-0.339, 0.318). However, the distribution is normally distributed according to Shapiro-Wilk, Lilliefors, and Anderson-Darling tests, each of which had p-values above 0.05 (0.745, 0.912, and 0.826), indicating we do not reject their null hypotheses that the distribution is normal. Given the results of these tests, the form of the data taken in the plots, and the proximity of skew and excess kurtosis to 0, I will consider this data normally distributed.

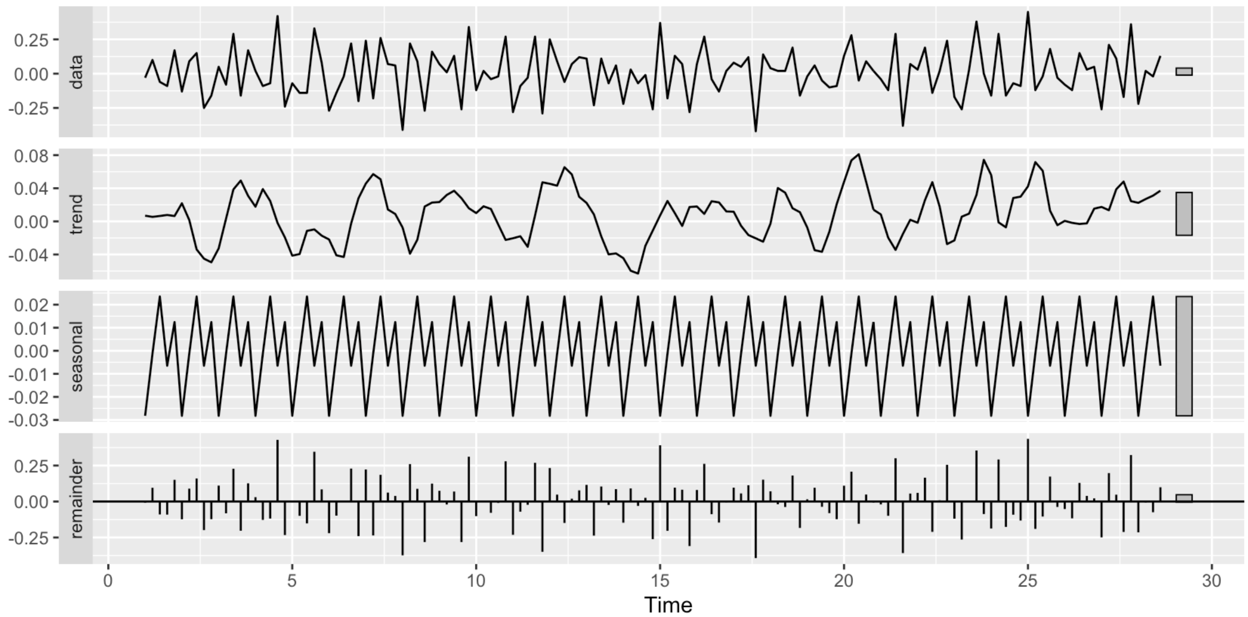
nobs	139.000000
NAs	0.000000
Minimum	-0.420000
Maximum	0.450000
1. Quartile	-0.120000
3. Quartile	0.120000
Mean	0.008201
Median	0.020000
Sum	1.140000
SE Mean	0.015185
LCL Mean	-0.021824
UCL Mean	0.038227
Variance	0.032051
Stdev	0.179028
Skewness	0.097016
Kurtosis	-0.339051





The data do feature auto-correlation. This is demonstrated by a Box-Ljung test of the null hypothesis that the observations are independent, which returns a p-value of 0.001, sufficiently low to reject the null hypothesis and assume auto-correlation.

Below is a decomposition of the data:



The trend line appears to be a moderately smoothed version of the data, following similar peaks and valleys but with less frequent changes. It does not appear to be a smooth, upward or downward line, as one would hope for. This degree of variance in the trend line indicates that the decomposition did not explain the variance in the data well enough to produce a clean trend line. The seasonal section indicates that there is a well defined seasonal component to the data, which features four steps: increase to global seasonal maximum, decrease to a local minimum, increase to a local maximum, and then a decrease to the global seasonal minimum. The residuals panel shows a sinusoidal pattern of sorts, indicating that they hold some information of variance that the decomposition was not able to capture.

2. Seasonal Autoregressive Moving Average (SARMA) Models

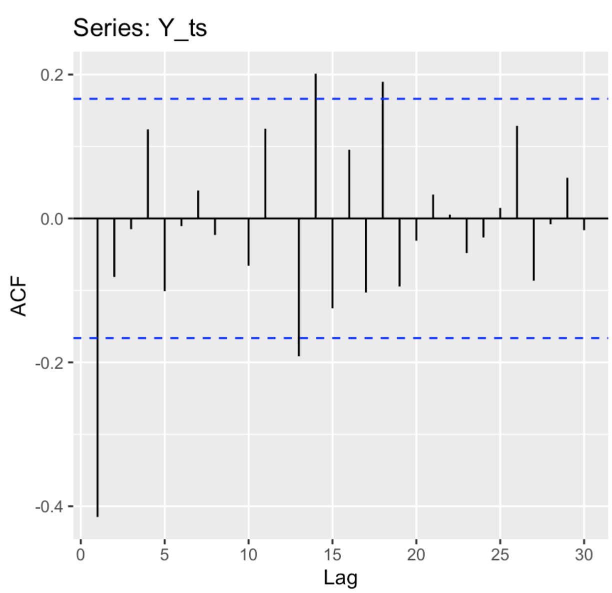
2.1 Trend Removal

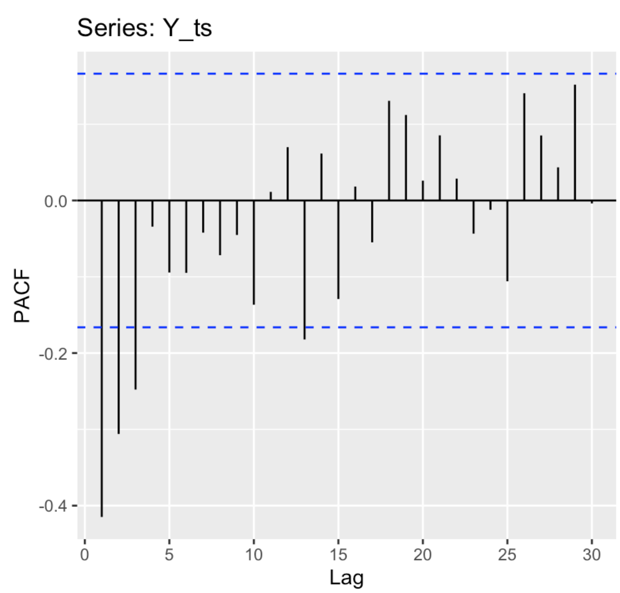
Performing an SLR of the first differenced data on year produced a B_1 coefficient of 0.007, with a t-test 95% confidence interval of (-2.264, 1.934) and p-value of 0.5. Thus, the null hypothesis that B_1 is equal to 0 is accepted. This demonstrates that the linear trend has been removed, as the time dependent coefficient is zero.

2.2 ARIMA Model

Below are the ACF and PACF plots of the data. The ACF shows that the first lag is significant, indicating that an MA(1) component is appropriate. The PACF shows that the first three lags are significant, indicating that an AR(3) component is appropriate. I have already established that first differencing is required, so the appropriate model is one of the form ARMA(3,1,1). The general form of this model would be:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \theta_1 z_{t-1}$$



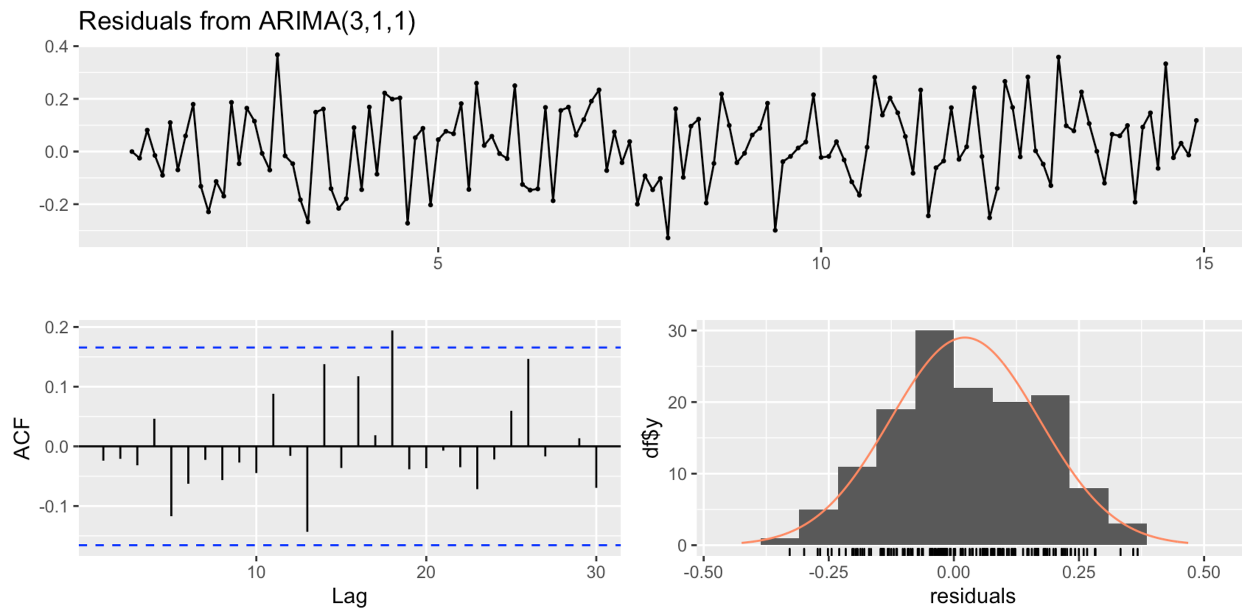


The specific form of this model is shown to the right. This model has five degrees of freedom. Overall, the model's residuals indicate it fits well; they are normally distributed, stationary, and are non-auto-correlated. However, they do have detectable business cycles.

```
Series: y
ARIMA(3,1,1)

Coefficients:
          ar1      ar2      ar3      ma1
      -0.0274  -0.1001  -0.0253  -0.6029
s.e.   0.4102   0.2548   0.2004   0.4059

sigma^2 = 0.02328:  log likelihood = 65.84
AIC=-121.68   AICc=-121.23   BIC=-107.01
```

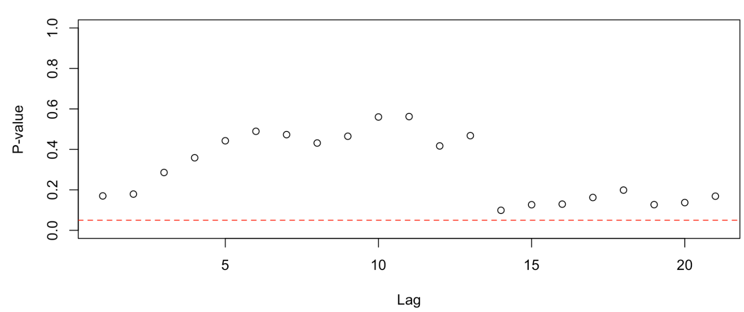


Normality of the residuals can be seen in the histogram above in the residuals panel. The distribution is unimodal and has good symmetry. Skew and excess kurtosis are both close to zero (-0.033 and -0.601), but neither of their 95% confidence intervals contain zero (-0.035, -0.015) and (-0.601, -0.578). However, Shapiro-Wilk, Lilliefors, and Anderson-Darling tests all indicate normality, as they each have a p-value high enough to accept their null hypotheses of normality (0.62, 0.55, and 0.61).

nobs	140.000000
NAs	0.000000
Minimum	-0.328062
Maximum	0.367156
1. Quartile	-0.083385
3. Quartile	0.140476
Mean	0.022028
Median	0.017237
Sum	3.083891
SE Mean	0.012570
LCL Mean	-0.002825
UCL Mean	0.046880
Variance	0.022119
Stdev	0.148726
Skewness	-0.033057
Kurtosis	-0.601101

Stationarity of the residuals is demonstrated by multiple methods. As shown on the right, the mean of the residuals is 0.022, which is quite close to zero. A t-test of the mean produced a confidence interval of (-0.003, 0.047) and a p-value of 0.082, allowing us to accept the null hypothesis that the mean is zero. A McLeod test produced the plot on the bottom right of this page, which shows that none of the lags have non-constant variance. Thus, these residuals have met the criteria for stationarity. An ADF test of the residuals corroborates stationarity. The p-value produced by the test was 0.01, allowing us to assert linear-trend and drift stationarity. However, a KPSS produced a test statistic of 0.034, which is above the critical value required to assert stationarity. Given the residuals' mean and variance, I consider these residuals stationary.

A Box-Ljung test of the residuals returned a p-value of 0.9426, allowing us to accept the null hypothesis that the residuals are independent, and thus do not feature auto-correlation.



The residuals do have detectable business cycles of lengths 7.578 and 2.690. Given that we are not assessing the relationship between economic cycles and temperature deviation, I perceive these business cycles to be patterns in the data that the model was not able to account for. Despite this, I still believe that the residuals overall indicate a good fit for the reasons listed above.

2.3 SARIMA Model

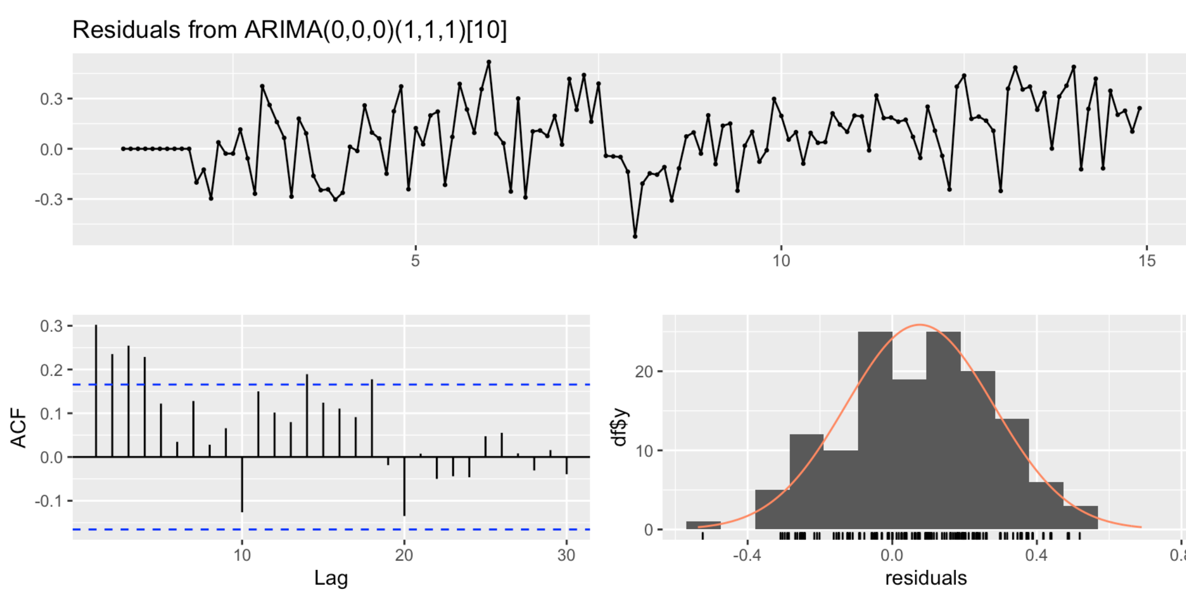
For this section, I use a SARIMA(0,0,0)(1,1,1) model. The precise form of this model is shown to the right. This model has three degrees of freedom.

```
Series: y
ARIMA(0,0,0)(1,1,1)[10]

Coefficients:
         sar1    sma1
      -0.8661  0.7182
s.e.   0.1252  0.1758

sigma^2 = 0.05179:  log likelihood = 8.15
AIC=-10.31   AICc=-10.12   BIC=-1.7
```

I experimented with several combinations of (Q,d,P), which included (1,1,2), (2,1,1), and (2,1,2). The residuals of these models appeared similar in their residuals panel, and the models each had similar results of tests such as those for normality and stationarity. However, AR(1) and MA(1) were the only variables that were determined to be statistically significant by the parameter test function, the second AR and MA variables were not. Therefore, I proceed with a (1,1,1) seasonal component. This model's residual panel is shown below.



Overall, this model's residuals indicate that it is not a good fit of the data. While the residuals are normally distributed and have no detectable business cycles, they do not feature stationarity and do feature auto-correlation.

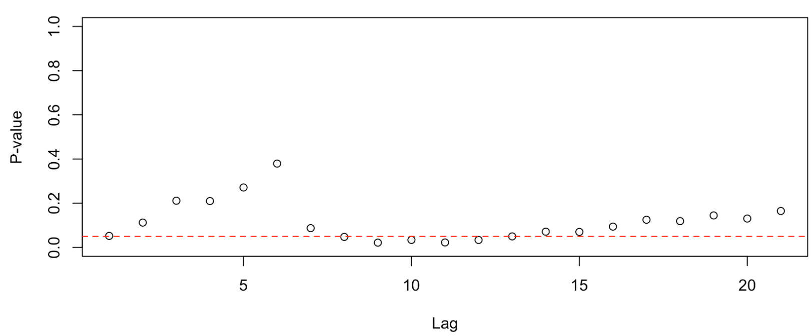
The residuals panel above shows the residuals' distribution; they appear normally distributed given their unimodal and symmetrical shape. Shapiro-Wilk, Lilliefors, and

Anderson-Darling tests all indicate normality. Each of these tests returned a p-value above 0.05, meaning that their null hypotheses of normality are accepted. Their p-values were 0.24, 0.62, and 0.31 respectively. The 95% confidence intervals for the residuals' skew (-0.159, -0.14) and excess kurtosis (-0.383, -0.35) do not contain zero, however, these values are close enough that I would not reject normality purely based on them. Given the histogram and results of the three tests, I consider these residuals normal.

As stated, the residuals are not stationary. The summary table on the right shows that their mean is 0.0076, quite close to zero. However, a t-test of the mean returned a 95% confidence interval of (0.042, 0.11) with a p-value of 0.00002. Thus, the mean is not zero. A McLeod-Li test of the residuals produced the plot below, which indicates that multiple lags show non-constant variance. Given that the residuals' mean is not zero and there is not constant variance, this would indicate that they are not stationary. This is at odds with the results of an ADF test, which produced a p-value of 0.01, indicating that the residuals have linear-trend and drift stationarity. However, a KPSS test produced a test statistic of 0.093, indicating the acceptance of the null hypothesis of non-stationarity. Purely based on the mean and variance, I assert that the residuals are not stationary.

nobs	140.000000
NAs	0.000000
Minimum	-0.524236
Maximum	0.518549
1. Quartile	-0.043148
3. Quartile	0.205169
Mean	0.076112
Median	0.091591
Sum	10.655740
SE Mean	0.017291
LCL Mean	0.041926
UCL Mean	0.110299
Variance	0.041855
Stdev	0.204585
Skewness	-0.145610
Kurtosis	-0.382370

As previously mentioned, the residuals do feature auto-correlation. This is demonstrated by performing a Box-Ljung test on them, which produced a p-value of 0.0000008, allowing us to reject the hypotheses of independence of observations and assert auto-correlation.



2.4 Model Comparison

By measures of in-sample goodness-of-fit, my ARIMA(3,1,1) model is preferable to my SARIMA(0,0,0)(1,1,1)$_{10}$ model. As described in section 2.2, the residuals for my ARIMA model indicated that the model fit the data quite well. The residuals were normally distributed, stationarity, and did not feature auto-correlation. Conversely, the residuals of the SARIMA model were not normally distributed and did feature auto-correlation.

2.5 Out-of-Sample Testing

The results of my two models' backtests are shown below. The errors and bias of the ARIMA model were equal to or less than the errors and bias of the SARIMA model. Given this, I would argue that the ARIMA model performed better in the out-of-sample backtest than the SARIMA model.

	ARIMA(3,1,1)				$SARIMA(0,0,0)(1,1,1)_{10}$			
Measurement	t+1	t+2	t+3	t+4	t+1	t+2	t+3	t+4
RMSE	0.066	0.071	0.093	0.122	0.205	0.205	0.192	0.249
MAE	0.050	0.050	0.069	0.122	0.197	0.195	0.179	0.249
Bias	0.031	0.050	0.062	0.122	0.197	0.195	0.179	0.249

3. ARMA×SARMA Models

3.1 Model Creation

I began this section by running a model that combines the two models that I created in section 2, which resulted in a SARIMA(3,1,1)$(1,1,1)_{10}$ model. After applying the parameter test function, I found that only AR_1 and SAR_1 variables were statistically significant. I compared the two models by running backtests and found that the simplified SARIMA(0,1,1)$(0,1,1)_{10}$ model performed better. After experimenting with additional various combinations of orders for both AR and MA in the main and seasonal components, as well as the inclusion of differencing in the seasonal component, I recurrently found that the SARIMA(0,1,1)$(0,1,1)_{10}$ performed best. Thus, it is the model that I use in this section. The precise form of this model is shown below to the right. This model has four degrees of freedom.

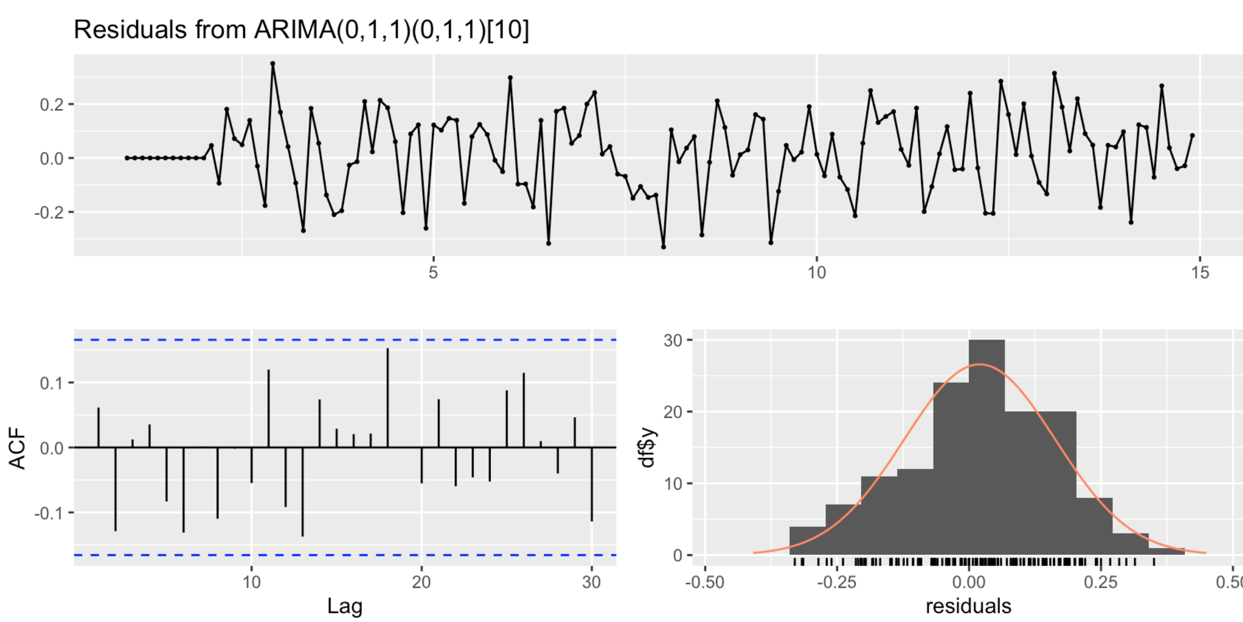
```
Series: y
ARIMA(0,1,1)(0,1,1)[10]

Coefficients:
          ma1     sma1
      -0.6984  -1.0000
s.e.   0.0750   0.0998

sigma^2 = 0.02285:  log likelihood = 48.18
AIC=-90.36   AICc=-90.17   BIC=-81.78
```

Overall, the residuals of this model indicate that it fits the data quite well. The residuals are normally distributed, stationary, and non-auto-correlated. They do have a detectable business cycle, but the results as a whole are quite good. The residuals panel is shown at the bottom of the page.

The histogram in the panel shows that their distribution is unimodal and has good symmetry. Skew (-0.251) and excess kurtosis (-0.329) are quite close to zero, however, neither of their 95% confidence intervals contain zero. They are (-0.258, -0.242) and (-0.328, -0.303) respectively. These would indicate non-normality, but this is at odds with the results of Shapiro-Wilk, Lilliefors, and Anderson-Darling tests, which all indicate normality. Their p-values were 0.28, 0.07, and 0.23, all sufficiently high to accept their null hypotheses of normality. Given these test results and the histogram in the residuals panel, I assert that these residuals are normally distributed.

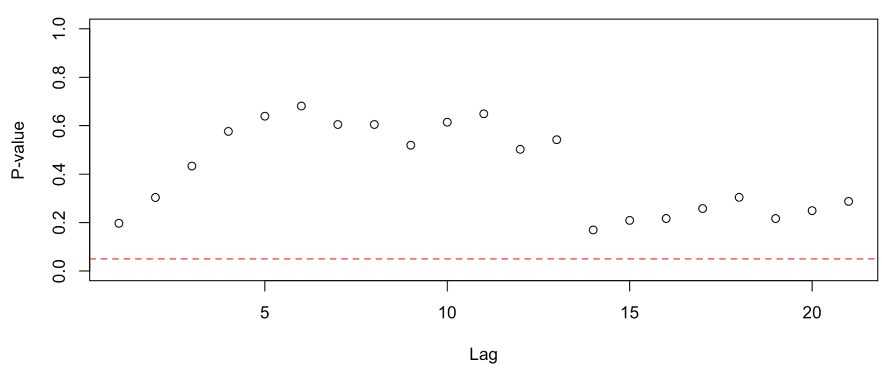


nobs	140.000000
NAs	0.000000
Minimum	-0.330201
Maximum	0.350513
1. Quartile	-0.066822
3. Quartile	0.122933
Mean	0.020204
Median	0.022124
Sum	2.828495
SE Mean	0.012091
LCL Mean	-0.003702
UCL Mean	0.044109
Variance	0.020466
Stdev	0.143061
Skewness	-0.251173
Kurtosis	-0.329137

Stationarity of the residuals can be demonstrated by their mean and variance. As the table on the right shows, their mean is 0.02, quite close to zero. A t-test of this mean produced a 95% confidence interval of (-0.002, 0.048) and a p-value of 0.07, both of which allow us to state that the residuals mean is zero. A McLeod-Lit test produced the plot shown below. It indicates that all lags have constant variance. Thus, the residuals are mean zero with constant variance, fitting the definition of stationary. Linear-trend and drift stationarity are supported by an ADF test of the residuals, which produced a p-value of 0.01, allowing us to reject the null hypothesis of non-stationarity. A KPSS test produced a test-statistic of 0.049, which indicates rejection of the null hypothesis that the residuals feature random-walk stationarity. Despite this, the residuals are mean zero with constant variance, and therefore, I assert have stationarity.

A Box-Ljung test of the residuals produced a p-value of 0.41, allowing us to accept the null hypothesis that they are independent. This indicates that the residuals do not feature auto-correlation.

A business cycle of duration 3.26 was detected in the residuals. As in my analysis of my ARIMA model, I conclude that this indicates there is a pattern in the residuals that the model did not capture. Given that we are not using economic data nor performing economic analysis, I do not conclude any economic meaning from this.



3.2 ARIMAxSARIMA and ARIMA Model Comparison via Model Diagnostics

The model diagnostics of my ARIMA and ARIMAxSARIMA models were quite similar. The residuals of both models were normally distributed, stationary, non-auto-correlated, and had detectable business cycles. Comparison of the models' summary statistics and p-values / test statistics do not indicate a clear winner either. For instance, the ARIMA model has lower skew, but the ARIMAxSARIMA model has lower excess kurtosis. Based on model diagnostics alone, it is not clear that one model is definitively better than the other.

3.3 ARIMAxSARIMA and ARIMA Model Comparison via Backtesting

Backtesting provided a much clearer picture of how the two models perform relative to one another. Compared to the ARIMA model, the ARIMAxSARIMA model has equivalent or lower values for RMSE, MAE, and Bias across all time periods. Therefore, I argue that the combination model is better.

	ARIMA(3,1,1)				$SARIMA(0,1,1)(0,1,1)_{10}$			
Measurement	t+1	t+2	t+3	t+4	t+1	t+2	t+3	t+4
RMSE	0.066	0.071	0.093	0.122	0.054	0.054	0.051	0.077
MAE	0.050	0.050	0.069	0.122	0.049	0.050	0.048	0.077
Bias	0.031	0.050	0.062	0.122	0.014	0.002	0.017	0.077

3.4 ARIMAxSARIMA and SARIMA Model Comparison via Model Diagnostics

The model diagnostics of the SARIMA and ARIMAxSARIMA model indicate that the combination model is a better fit of the data. The residuals of the combination model have more ideal characteristics; they are normally distributed, stationarity, and do not feature auto-correlation. Conversely, the residuals of the SARIMA model were not stationarity and did feature auto-correlation.

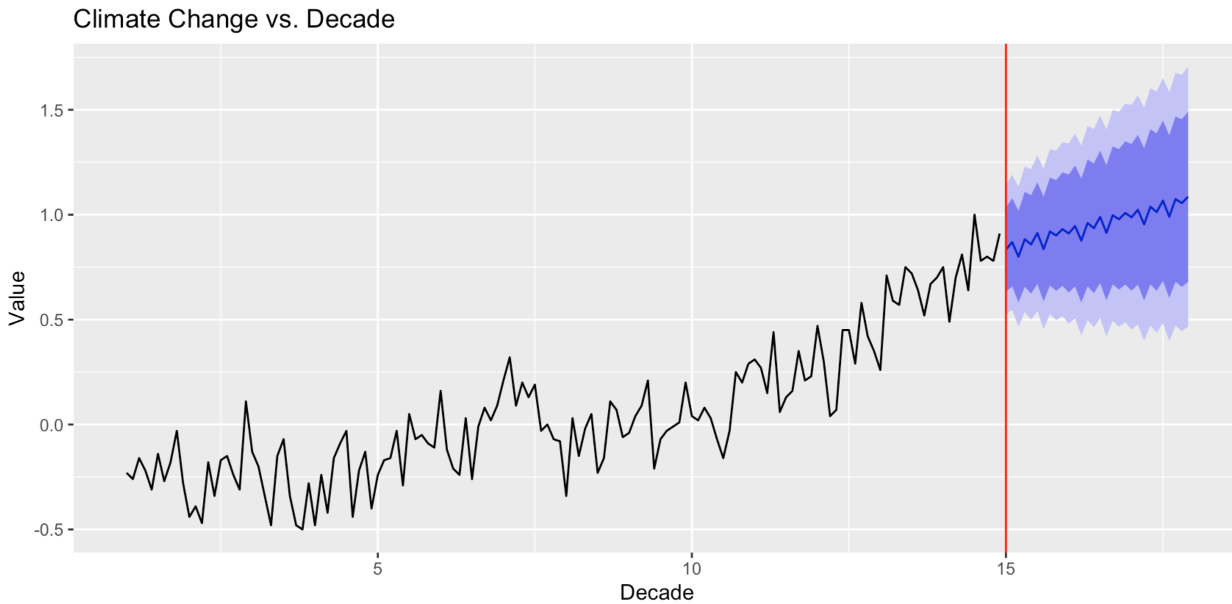
3.5 ARIMAxSARIMA and SARIMA Model Comparison via Backtesting

Comparison of the SARIMA and ARIMAxSARIMA model also indicates that the combination model is superior. Measurements of error and bias across all time periods are lower for the combination model than for the SARIMA model.

	$SARIMA(0,0,0)(1,1,1)_{10}$				$SARIMA(0,1,1)(0,1,1)_{10}$			
Measurement	t+1	t+2	t+3	t+4	t+1	t+2	t+3	t+4
RMSE	0.205	0.205	0.192	0.249	0.054	0.054	0.051	0.077
MAE	0.197	0.195	0.179	0.249	0.049	0.050	0.048	0.077
Bias	0.197	0.195	0.179	0.249	0.014	0.002	0.017	0.077

4. Report

Below is a plot that shows my predictions for the global temperature deviation in degrees Celsius from base year 1880 between 2020 and 2050. These predictions were created by a model that identifies patterns in past data and makes probabilistic forecasts based on them. Probabilistic means that the predicted outcomes have associated probabilities. The dark blue line indicates what I have determined to be the expected outcome. The shading above and below the line represents other trajectories, with darker shading indicating greater likelihood. The lightest shaded regions represent the 5% most extreme outcomes.



The most notable feature of this prediction is that the temperature is predicted to increase. This is shown by the expected outcome line moving upwards. There are several important possible outcomes that this model also makes predictions about. According to the graph, there is a greater than 50% chance that we will cross the 1 degree Celsius line by 2050, and there is a 2.5% chance that we will cross the 1.5 degree Celsius line. This 1.5 degree threshold is the same outlined in the Paris Agreement, and the year 2050 is the general target year for net-zero initiatives.