Time Series Forecasting Concepts

Presented by: Derek Kane

Overview of Topics

- Importance of Time Series
- Decomposition Basics
  - Time Series Components
  - Additive vs. Multiplicative Models
- Time Series Forecasting Techniques
  - General Modeling Considerations
  - Modeling Algorithms
  - Output interpretation / Diagnostics
- Practical Application Example
- Generalized Least Squares (Maximum Likelihood)

Importance of Time Series Predictions

- The ability to observe the development of an area of interest (over time) and make predictions based upon historical observations creates a competitive advantage within the analytics based 21st century global business environment.

- Example: If an organization has the capability to better forecast the sales quantities of a product, they will be in a more favorable position to structure procurement and optimize inventory levels. This can result in an increased liquidity of the organization’s cash reserves, decrease of working capital, and improved customer satisfaction through increased order fulfillment and decreased backorders.

- Additionally, the senior management can formulate the entire capacity planning (fiscal sales budget) of the organization through better prediction of revenue (including costs & profit margin). Head Count, Capex and R&D expenditures, marketing budgets, all rely upon having a proper sales budget and time series predictions are the backbone of a successful organization.

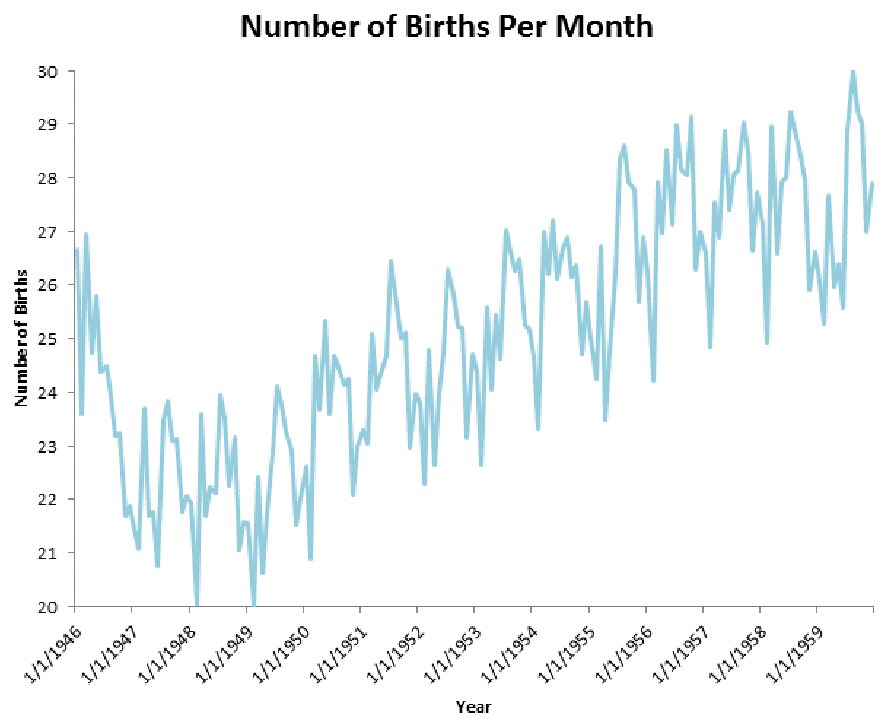
Decomposition Basics

Taking a different view…

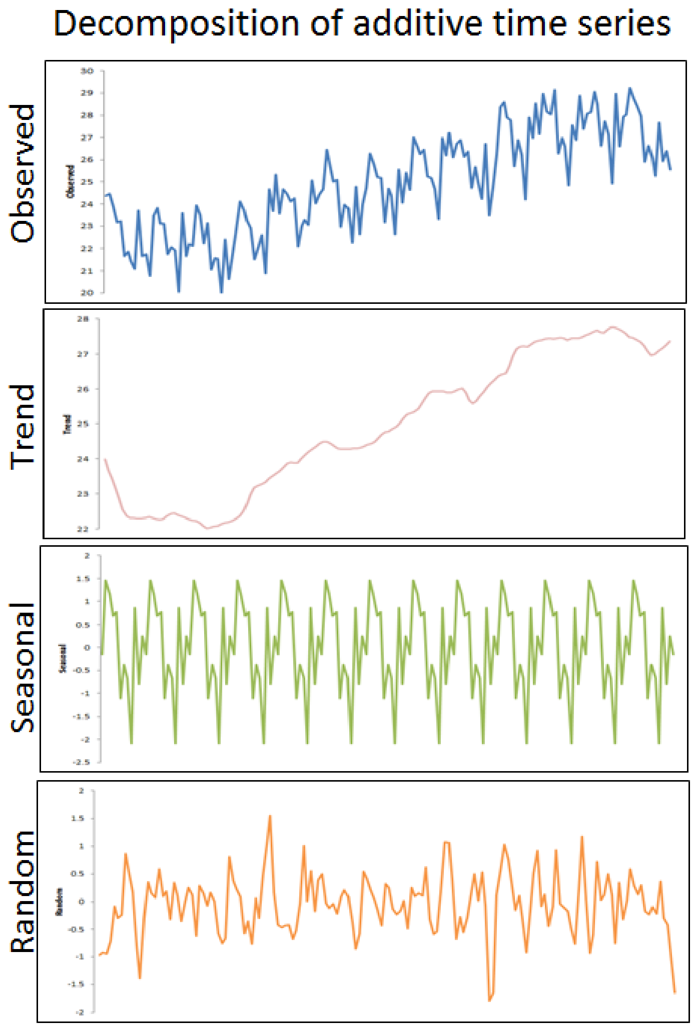
- Data Scientists and analytical professionals interpret graphs differently and we need to think about time series analysis in a different context in order to master these techniques.

- When we analyze a time series chart (see chart), we need to begin to think about decomposing the chart into its basic building blocks and then determining what type of time series we are trying to model.

- This process needs to be performed with every time series predictive task to ensure that we pair up the appropriate analytical model that best suits the data.



Decomposition Basics



- The previous example can be spilt into fundamental components:
  - Seasonality
  - Trend
  - Random fluctuations in the data (irregular components).

Decomposition - Seasonality

Seasonal Effects:

- a systematic and calendar related effect. Some examples include the sharp escalation in most Retail series which occurs around December in response to the Christmas period, or an increase in water consumption in summer due to warmer weather.
- Other seasonal effects include trading day effects (the number of working or trading days in a given month differs from year to year which will impact upon the level of activity in that month) and moving holidays (the timing of holidays such as Easter varies, so the effects of the holiday will be experienced in different periods each year).

Seasonal adjustment:

- the process of estimating and then removing from a time series influences that are systematic and calendar related.
- Observed data needs to be seasonally adjusted as seasonal effects can conceal both the true underlying movement in the series, as well as certain non-seasonal characteristics which may be of interest to analysts.

Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend.

Decomposition - Seasonality

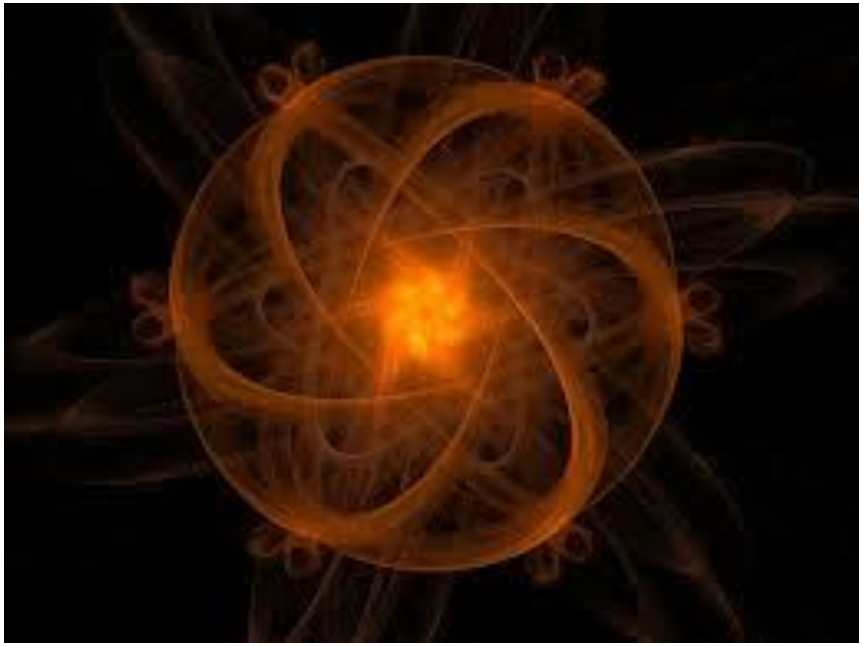
- Seasonality arises from systematic, calendar related influences such as:
  - Natural Conditions
    - Weather fluctuations that are representative of the season
  - Business and Administrative procedures
    - Start and end of the school term
    - Social and Cultural behavior Christmas
  - Social and Cultural behavior Christmas
  - Calendar Related Effects
    - Trading Day Effects the number of occurrences of each of the day of the week in a given month will differ from year to year. Ex. There were 4 weekends in March in 2000, but 5 weekends in March of 2002.
    - Moving Holiday Effects holidays which occur each year, but whose exact timing shifts. Ex. Easter, Chinese New Year, etc…

Decomposition – Trend & Irregular

- A trend can be defined as the 'long term' movement in a time series without calendar related and random (irregular) effects and is a reflection of the underlying level.

- The irregular component (sometimes also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed. It results from short term fluctuations in the series which are neither systematic nor predictable.

- In a highly irregular series, these random fluctuations can dominate movements, which will mask the trend and seasonality.



Decomposition Models

- The Seasonal, Trend, and Random components can be combined traditionally in 2 major manners to match the total observed data:

  - Additively – The implicit assumption is that the 3 components are combined in an additive fashion to match the data.

$$\text{Data} = \text{Seasonal} + \text{Trend} + \text{Random}$$

  - Multiplicatively – The implicit assumption is that the 3 components are combined in an multiplicative fashion.

$$\text{Data} = \text{Seasonal} * \text{Trend} * \text{Random}$$

Multiplicative or Additive Approach?

- The time series techniques we will discuss are approached from an additive standpoint.
- Fortunately, multiplicative models are equally easy to fit to data as additive models! The trick to fitting a multiplicative model is to take logarithms of both sides of the model.

$$\text{Log(Data)} = \text{Log(Seasonal * Trend * Random)}$$

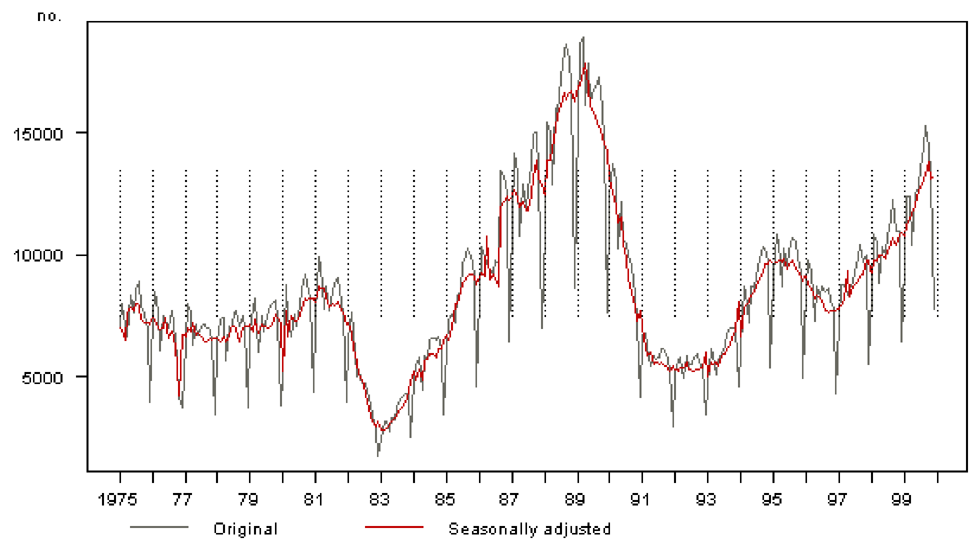
- After taking logarithms (either natural logarithms or to base 10), the three components of the time series again act additively.
- It is important to recognize when multiplicative models are appropriate. However fitting the models is no harder than fitting additive models.

Multiplicative Example

- In many time series involving quantities (e.g. money, wheat production, ...), the absolute differences in the values are of less interest and importance than the percentage changes. This is a scenario where the multiplicative approach will be used.

- In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises.

- As the underlying level of the series changes, the magnitude of the seasonal fluctuations varies as well. This shows characteristics of a multiplicative modeling approach.



Time Series Forecasting Techniques

General Approach for Analysis

1. Identify through a visual inspection whether the data has seasonality or trends.

2. Identify whether the decomposition technique required is additive or multiplicative

    Log Transform the multiplicative, if needed.

3. Test Appropriate Additive Algorithm

    Simple Moving Average Smoothing: Seasonal = No, Trend = Yes, Correlations = No

    Seasonal Adjustment: Seasonal = Yes, Trend = Yes , Correlations = No

    Simple Exponential Smoothing: Seasonal = No, Trend = Yes , Correlations = No

    Holts Exponential Smoothing: Seasonal = No, Trend = Yes , Correlations = No

    Holt-Winters Exponential Smoothing: Seasonal = Yes, Trend = Yes , Correlations = No

    ARIMA: Seasonal = Yes, Trend = Yes , Correlations = Yes

General Approach for Analysis

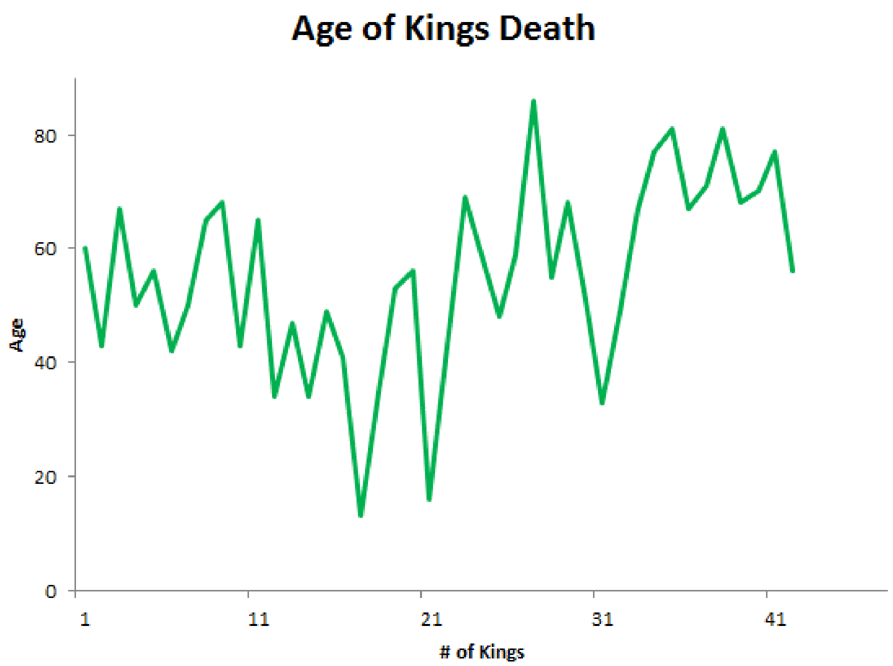
4. Perform statistical tests to verify correct model selected
    - Ljung-Box Test
    - Forecast Errors
        - Normal Distributions
        - Mean = 0
        - Constant Variance
    - Autocorrelation Function (ACF)
    - Partial-Autocorrelation Function (PACF)

5. Repeat Steps 3 & 4, if necessary

Simple MA Smoothing

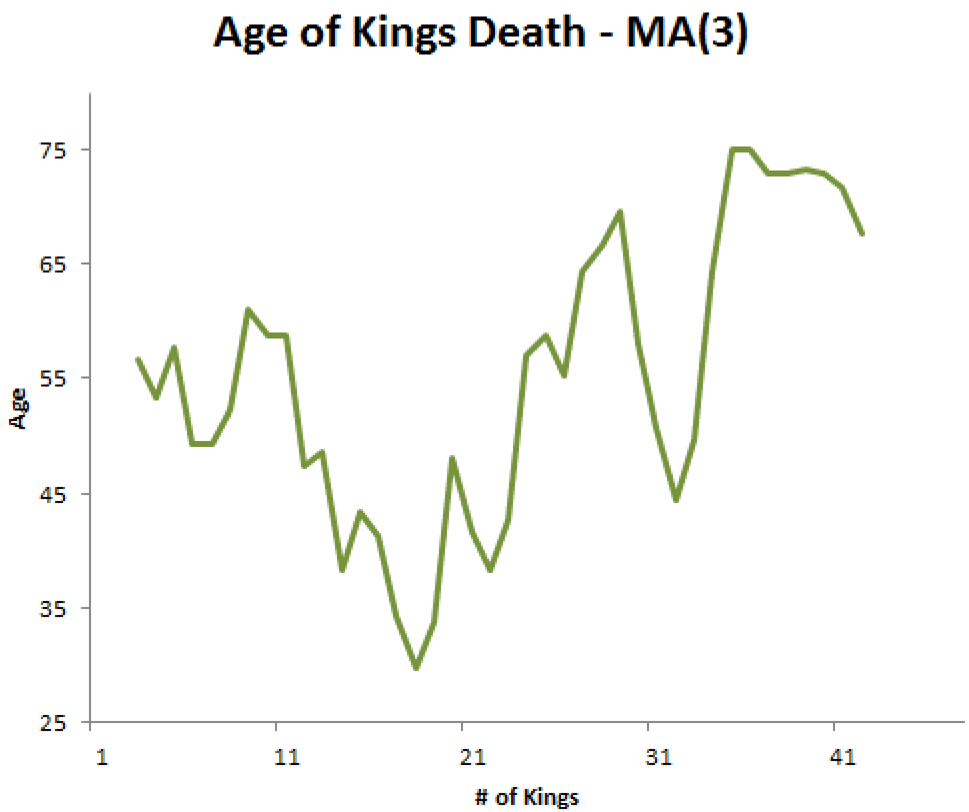
Simple MA Smoothing Time Series

- To estimate the trend component of a non-seasonal time series (consists of a trend component and an irregular component) that can be described using an additive model, it is common to use a smoothing method, such as calculating the simple moving average of the time series.



- The age of death for successive kings of England, starting from William the Conqueror. (Hipel & McLeod, 1994)
- Y-Axis: Age in Years
- X-Axis: # of Kings
- This dataset doesn't seem to show any seasonality or trending, however, the random fluctuation in data are roughly constant in size over time.
- This suggests that an "additive" model method could be used to describe the dataset.

Simple MA Smoothing Time Series

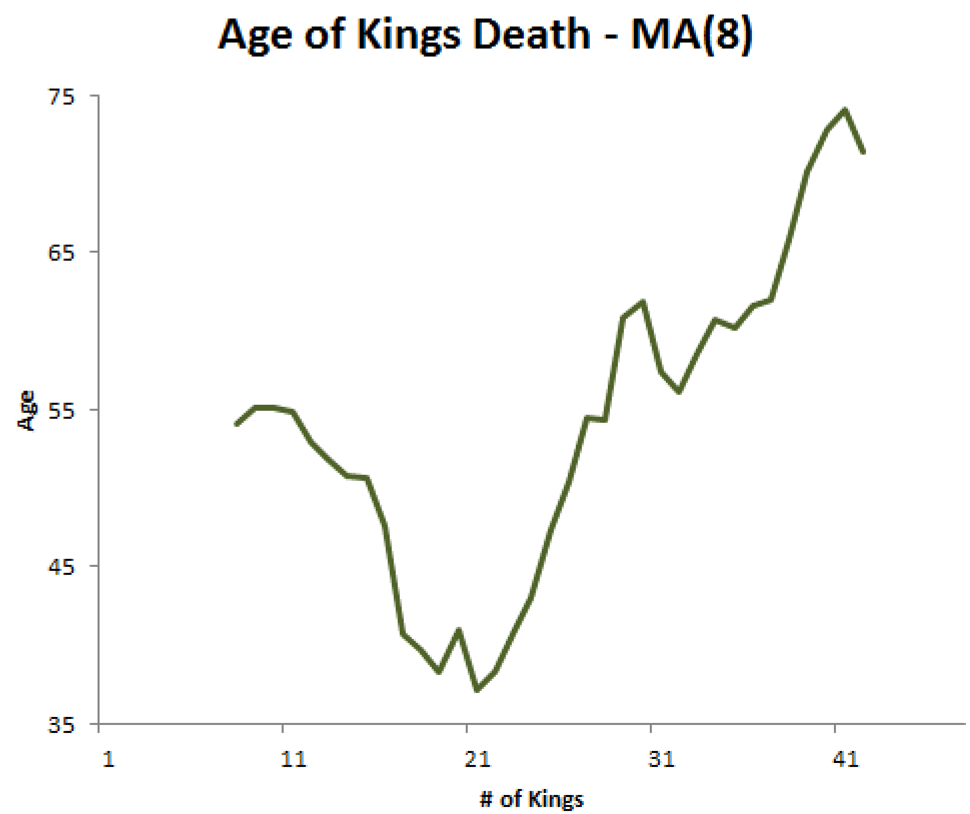


- Moving Average of order 3: MA(3)
- There still appears to be quite a lot of random fluctuations in the time series smoothed using a simple moving average of order 3. Thus, to estimate the trend component more accurately, we might want to try smoothing the data with a simple moving average of a higher order.
- This takes a little bit of trial-and-error, to find the right amount of smoothing. For example, we can try using a simple moving average of order 8:

Simple MA Smoothing Time Series

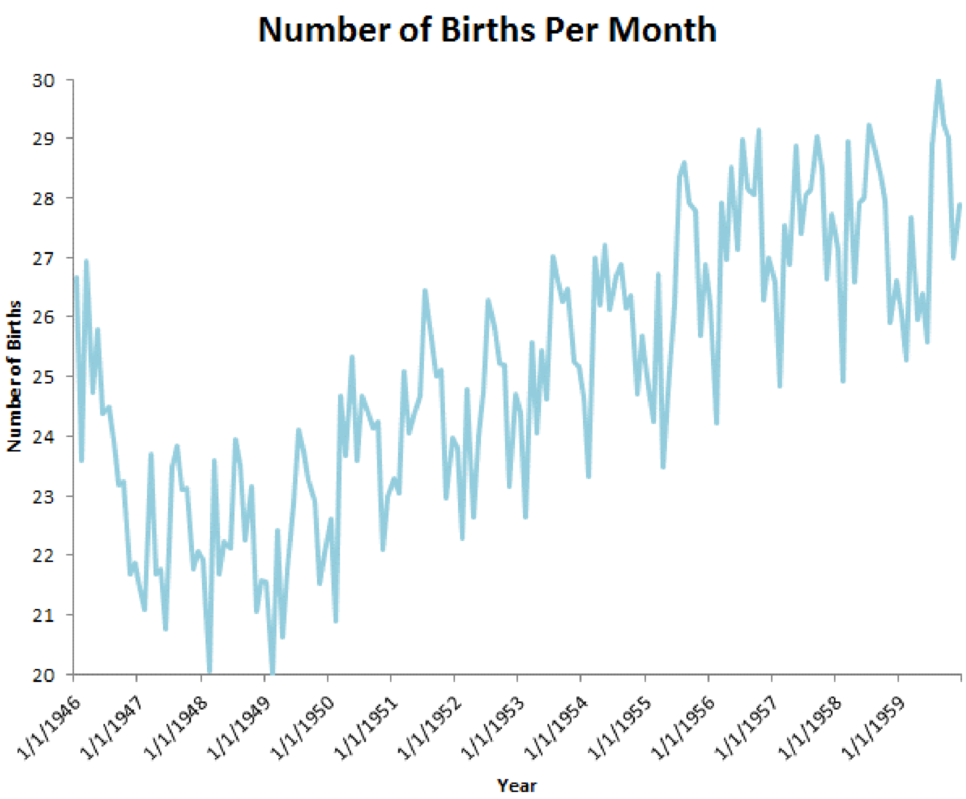
- The data smoothed with a simple moving average of order 8 MA(8) gives a clearer picture of the trend component.

- We can see that the age of death of the English kings seems to have decreased from about 55 years old to about 38 years old during the reign of the first 20 kings, and then increased after that to about 73 years old by the end of the reign of the 40th king in the time series.



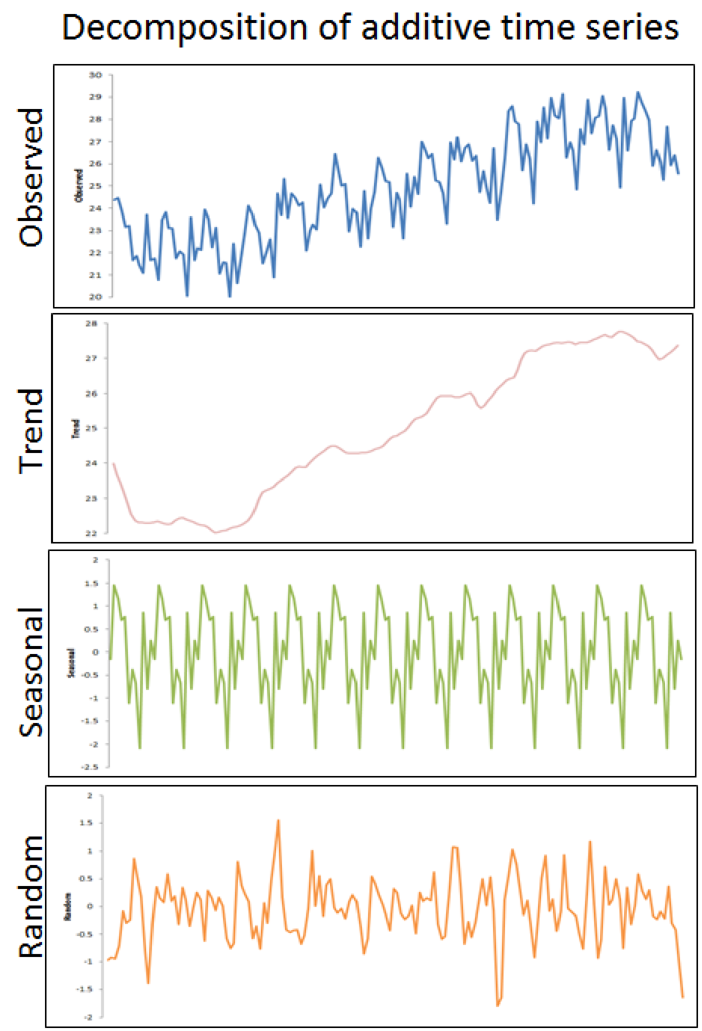
Seasonality Adjustment

Decomposing Seasonal Data & Seasonally Adjusting



- The number of births per month in New York City.
- Y-Axis: Number of Births
- X-Axis: Time in Months
- This dataset shows a seasonal variation with the number of births. There seems to be peak in every summer and a trough in every winter.
- The seasonal fluctuation are roughly constant in size over time and random fluctuations are roughly constant in size.
- This suggests that an “additive” modelling method could be used to describe the dataset.

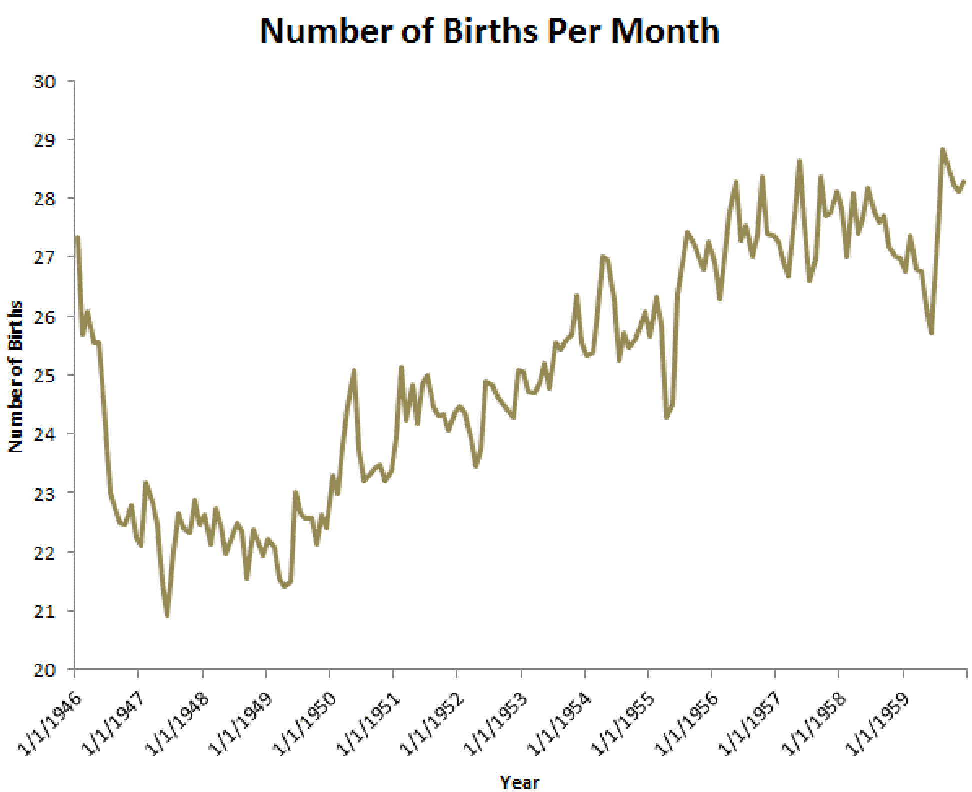
Decomposing Seasonal Data & Seasonally Adjusting



- The plot above shows the original time series (top), the estimated trend component (second from top), the estimated seasonal component (third from top), and the estimated irregular component (bottom).

- We see that the estimated trend component shows a small decrease from about 24 in 1947 to about 22 in 1948, followed by a steady increase from then on to about 27 in 1959.

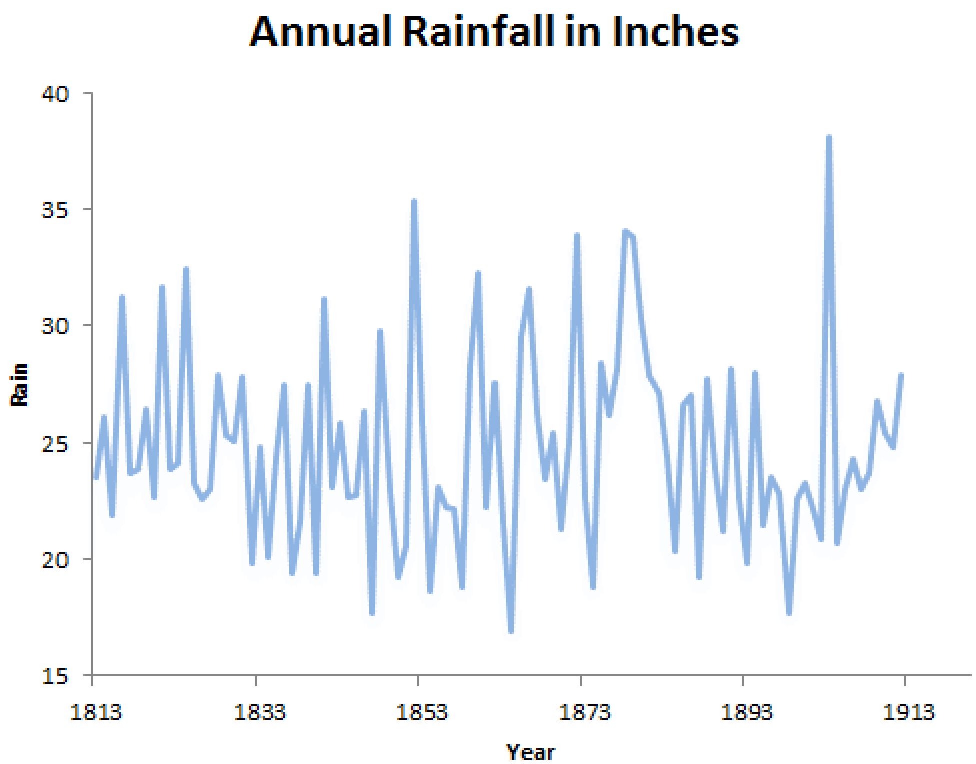
Seasonally Adjusted



- If you have a seasonal time series that can be described using an additive model, you can seasonally adjust the time series by estimating the seasonal component, and subtracting the estimated seasonal component from the original time series.
- Seasonally Adjusted: Original – Seasonal Component
- You can see that the seasonal variation has been removed from the seasonally adjusted time series. The seasonally adjusted time series now just contains the trend component and an irregular component.

Simple Exponential Smoothing

Simple Exponential Smoothing



- Contains total annual rainfall in inches for London, from 1813-1912 (original data from Hipel and McLeod, 1994).
- Y-Axis: Rainfall
- X-Axis: Months
- You can see from the plot that there is roughly constant level (the mean stays constant at about 25 inches). The random fluctuations in the time series seem to be roughly constant in size over time, so it is probably appropriate to describe the data using an additive model.
- If you have a time series that can be described using an additive model with constant level and no seasonality, you can use simple exponential smoothing to make short-term forecasts.

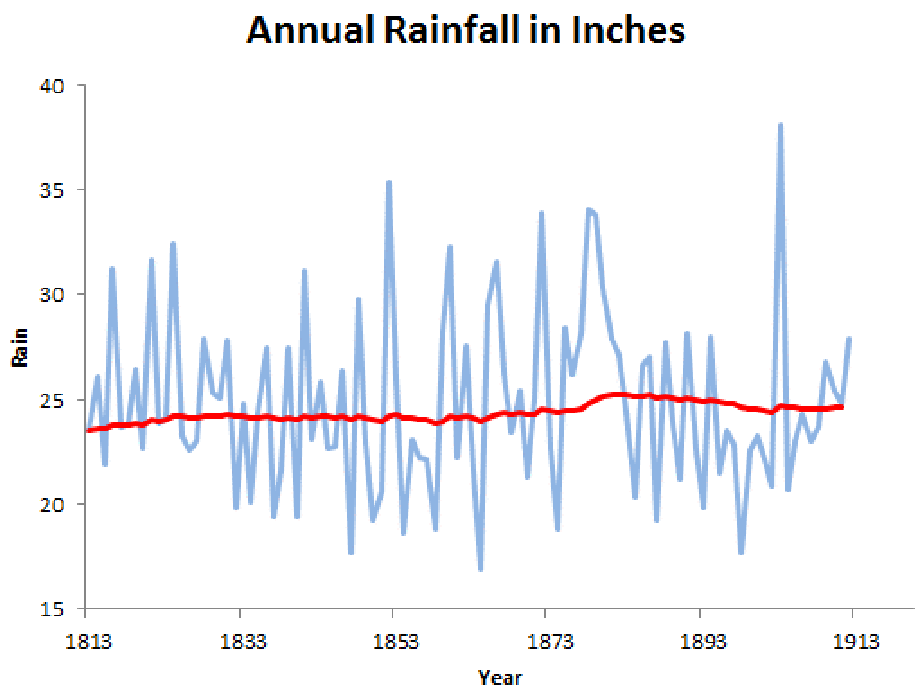
Simple Exponential Smoothing

- The simple exponential smoothing method provides a way of estimating the level at the current time point. Smoothing is controlled by the parameter alpha; for the estimate of the level at the current time point. The value of alpha; lies between 0 and 1. Values of alpha that are close to 0 mean that little weight is placed on the most recent observations when making forecasts of future values.

```
rainseriesforecasts <- HoltWinters(rainseries, beta=FALSE, gamma=FALSE)
rainseriesforecasts
Smoothing parameters:
alpha:  0.02412151
beta :  FALSE
gamma:  FALSE
Coefficients:
  [,1]
a 24.67819
```

- The output in R tells us that the estimated value of the alpha parameter is about 0.024. This is very close to zero, telling us that the forecasts are based on both recent and less recent observations (although somewhat more weight is placed on recent observations).

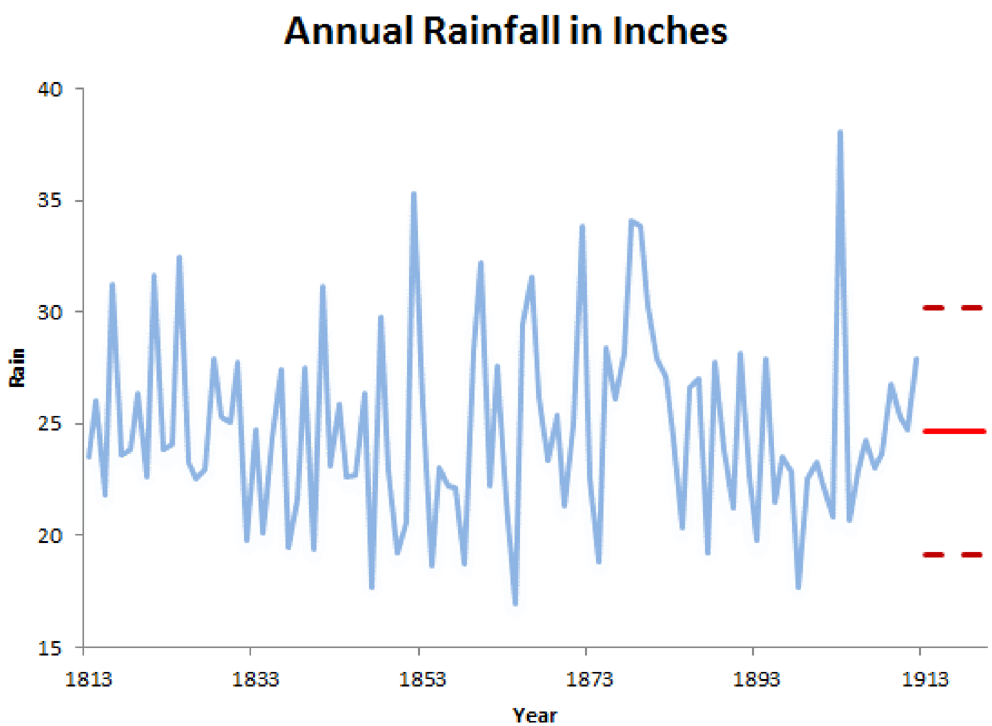
Simple Exponential Smoothing



- The plot shows the original time series in blue, and the forecasts as a red line. The time series of forecasts is much smoother than the time series of the original data here.
- As a measure of the accuracy of the forecasts, we can calculate the sum of squared errors for the in-sample forecast errors, that is, the forecast errors for the time period covered by our original time series.
- It is common in simple exponential smoothing to use the first value in the time series as the initial value for the level. For example, in the time series for rainfall in London, the first value is 23.56 (inches) for rainfall in 1813.

```
> HoltWinters(rainseries, beta=FALSE, gamma=FALSE, l.start=23.56)
```

Simple Exponential Smoothing



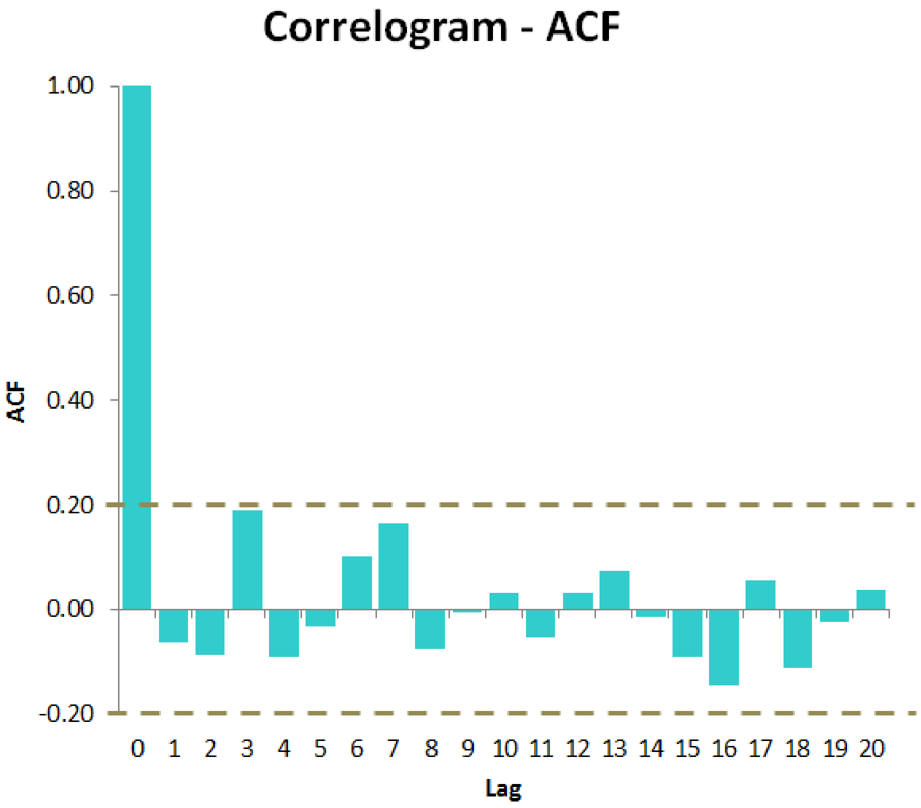
- Here the forecasts for 1913-1920 are plotted as a red line, the 80% prediction interval as the area between the maroon dotted lines.
- The 'forecast errors' are calculated as the observed values minus predicted values, for each time point. We can only calculate the forecast errors for the time period covered by our original time series, which is 1813-1912 for the rainfall data.

Statistical Tests to Improve the Model

- If the predictive model cannot be improved upon, there should be no correlations between forecast errors for successive predictions. In other words, if there are correlations between forecast errors for successive predictions, it is likely that the simple exponential smoothing forecasts could be improved upon by another forecasting technique.

- To figure out whether this is the case, we can obtain a correlogram (ACF) of the in-sample forecast errors for lags 1-20.

- You can see from the sample correlogram that the autocorrelation at lag 3 is just touching the significance bounds.



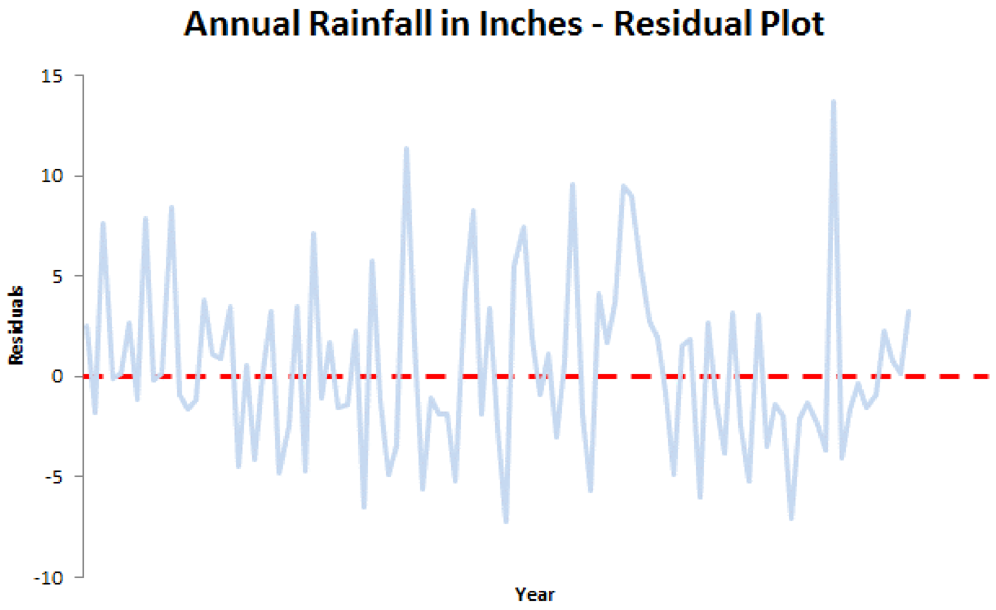
Statistical Tests to Improve the Model

- Here the Ljung-Box test statistic is 17.4, and the p-value is 0.6, so there is little evidence of non-zero autocorrelations in the in-sample forecast errors at lags 1-20.

```
Box.test(rainseriesforecasts2$residuals, lag=20, type="Ljung-Box")
  Box-Ljung test
data:  rainseriesforecasts2$residuals
X-squared = 17.4008, df = 20, p-value = 0.6268
```

- To be sure that the predictive model cannot be improved upon, it is also a good idea to check whether the forecast errors are normally distributed with mean zero and constant variance. To check whether the forecast errors have constant variance, we can make a time plot of the in-sample forecast errors:

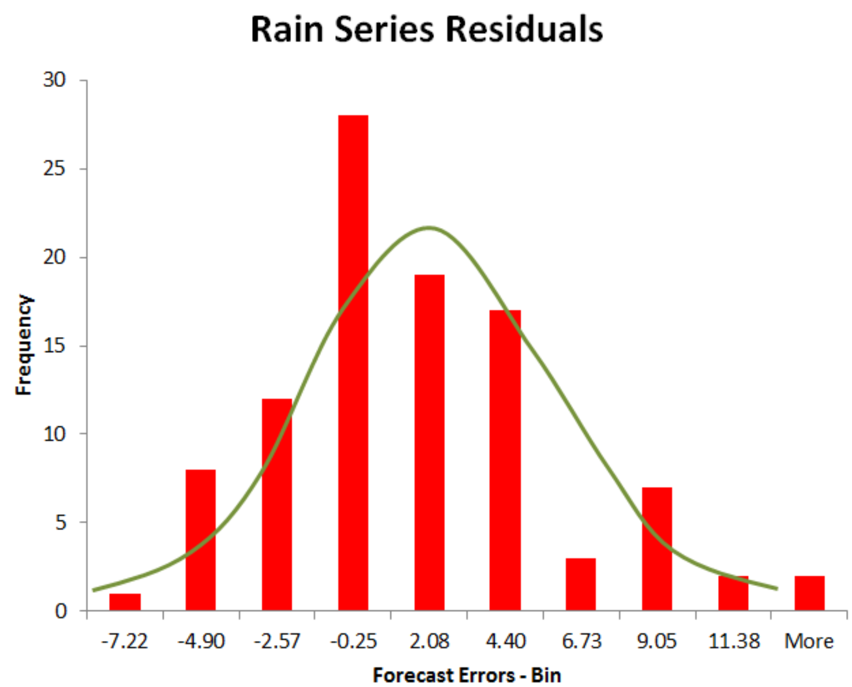
Statistical Tests to Improve the Model



- The plot shows that the in-sample forecast errors seem to have roughly constant variance over time, although the size of the fluctuations in the start of the time series may be slightly less than that at later dates.
- Constant variance can be detected by looking at the peaks and troughs to see if they are moving evenly above and below the 0 horizon.

Statistical Tests to Improve the Model

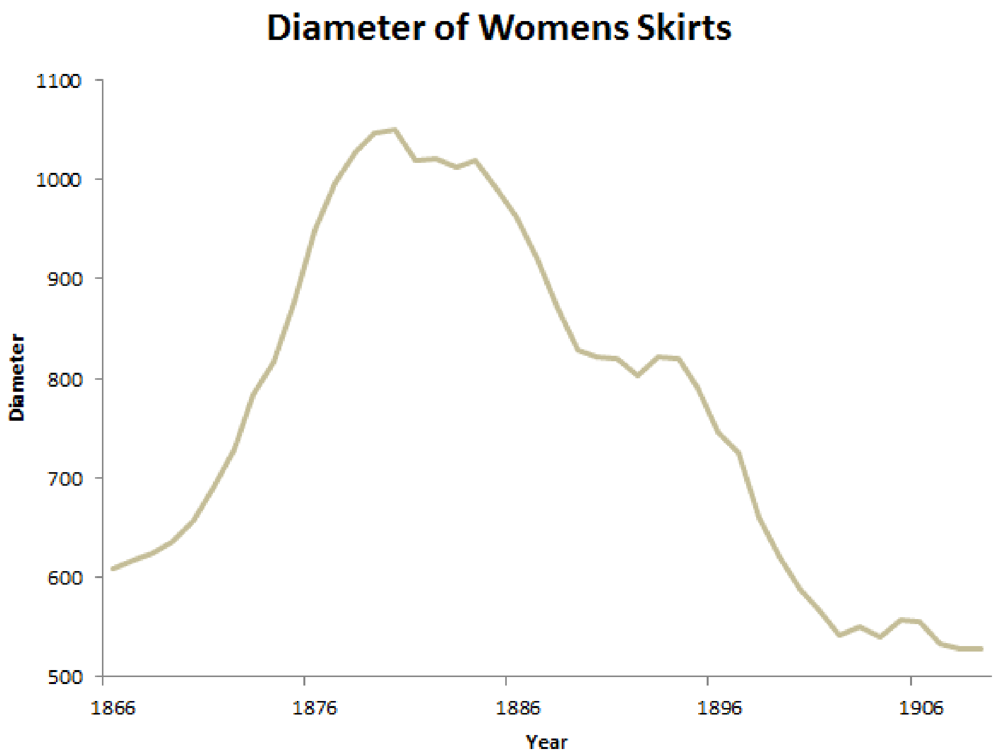
- To check whether the forecast errors are normally distributed with mean zero, we can plot a histogram of the forecast errors, with an overlaid normal curve that has mean zero and the same standard deviation as the distribution of forecast errors.
- The plot shows that the distribution of forecast errors is roughly centered on zero, and is more or less normally distributed, although it seems to be slightly skewed to the right compared to a normal curve. However, the right skew is relatively small, and so it is plausible that the forecast errors are normally distributed with mean zero.



Conclusion: The Ljung-Box test showed that there is little evidence of non-zero autocorrelations in the in-sample forecast errors, and the distribution of forecast errors seems to be normally distributed with mean zero. This suggests that the simple exponential smoothing method provides an adequate predictive model for London rainfall, which probably cannot be improved upon.

Holts Exponential Smoothing

Holts-Exponential Smoothing



- An example of a time series that can probably be described using an additive model with a trend and no seasonality is the time series of the annual diameter of women's skirts at the hem, from 1866 to 1911. (original data from Hipel and McLeod, 1994).
- Y-Axis: Diameter of Skirts
- X-Axis: Time in Months

- We can see from the plot that there was an increase in hem diameter from about 600 in 1866 to about 1050 in 1880, and that afterwards the hem diameter decreased to about 520 in 1911.

Holts-Exponential Smoothing

- If you have a time series that can be described using an additive model with increasing or decreasing trend and no seasonality, you can use Holt's exponential smoothing to make short-term forecasts.

- Holt's exponential smoothing estimates the level and slope at the current time point. Smoothing is controlled by two parameters, alpha, for the estimate of the level at the current time point, and beta for the estimate of the slope b of the trend component at the current time point. As with simple exponential smoothing, the parameters alpha and beta have values between 0 and 1, and values that are close to 0 mean that little weight is placed on the most recent observations when making forecasts of future values.

- The simple exponential smoothing method provides a way of estimating the level at the current time point. Smoothing is controlled by the parameter alpha; for the estimate of the level at the current time point. The value of alpha; lies between 0 and 1. Values of alpha that are close to 0 mean that little weight is placed on the most recent observations when making forecasts of future values.

Holts-Exponential Smoothing

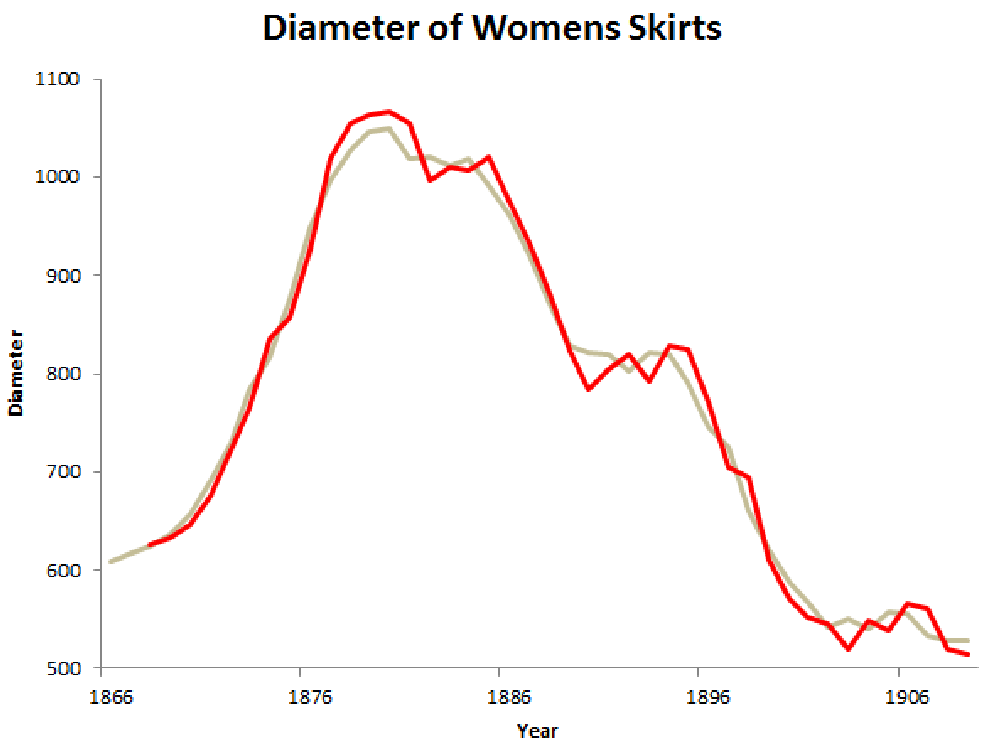
- To make forecasts, we can fit a predictive model using the HoltWinters() function in R. To use HoltWinters() for Holt's exponential smoothing, we need to set the parameter gamma=FALSE (the gamma parameter is used for Holt-Winters exponential smoothing, as described below).

- For example, to use Holt's exponential smoothing to fit a predictive model for skirt hem diameter, we type in R:

```
> skirtsseriesforecasts <- HoltWinters(skirtsseries, gamma=FALSE)
> skirtsseriesforecasts
  Smoothing parameters:
  alpha:  0.8383481
  beta :  1
  gamma:  FALSE
  Coefficients:
    [,1]
  a 529.308585
  b   5.690464
```

- The estimated value of alpha is 0.84, and of beta is 1.00. These are both high, telling us that both the estimate of the current value of the level, and of the slope b of the trend component, are based mostly upon very recent observations in the time series. This makes good intuitive sense, since the level and the slope of the time series both change quite a lot over time. The value of the sum-of-squared-errors for the in-sample forecast errors is 16954.

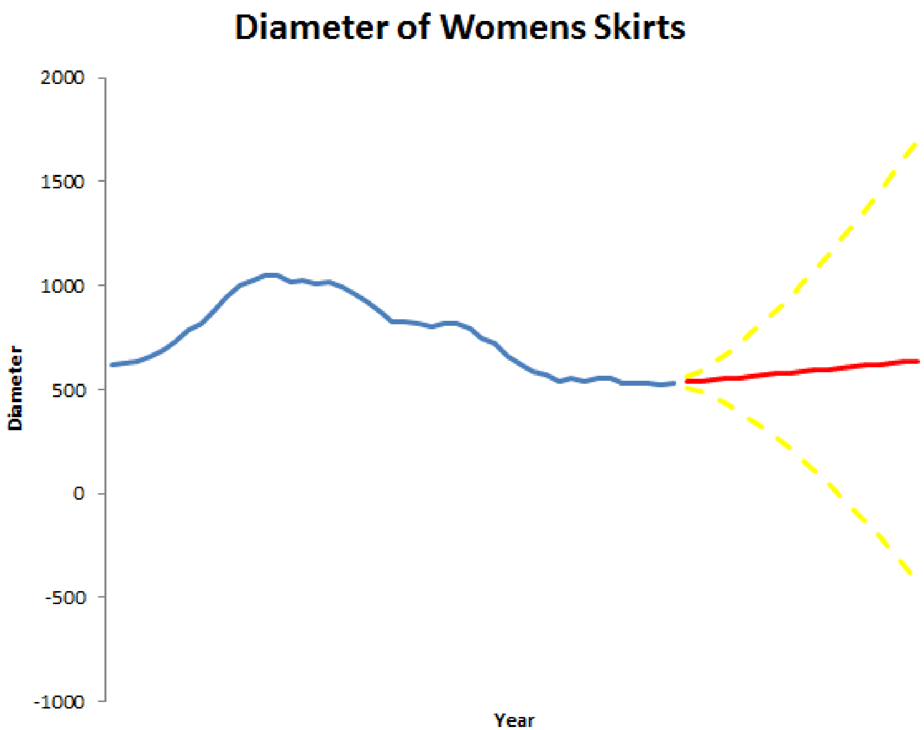
Holts-Exponential Smoothing



- We can plot the original time series as a brown line, with the forecasted values as a red line on top of that.
- We can see from the picture that the in-sample forecasts agree pretty well with the observed values, although they tend to lag behind the observed values a little bit.
- It is common to set the initial value of the level to the first value in the time series (608 for the skirts data), and the initial value of the slope to the second value minus the first value (9 for the skirts data).

```
HoltWinters(skirtsseries, gamma=FALSE, l.start=608, b.start=9)
```

Holts-Exponential Smoothing



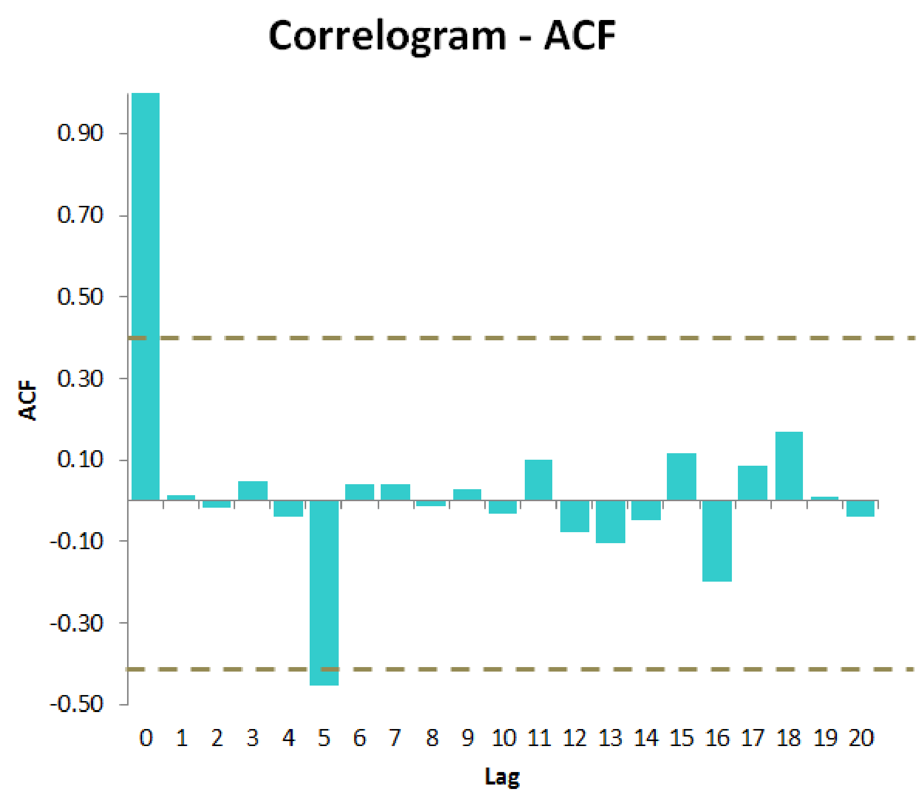
- The forecasts are shown as a yellow line, with the 80% prediction intervals.
- As for simple exponential smoothing, we can make forecasts for future times not covered by the original time series by using the forecast.HoltWinters() function in the “forecast” R package.
- For example, our time series data for skirt hems was for 1866 to 1911, so we can make predictions for 1912 to 1930 (19 more data points)

Statistical Tests to Improve the Model

- As for simple exponential smoothing, we can check whether the predictive model could be improved upon by checking whether the in-sample forecast errors show non-zero autocorrelations at lags 1-20.

- Here the correlogram shows that the sample autocorrelation for the in-sample forecast errors at lag 5 exceeds the significance bounds.

- However, we would expect one in 20 of the autocorrelations for the first twenty lags to exceed the 95% significance bounds by chance alone.



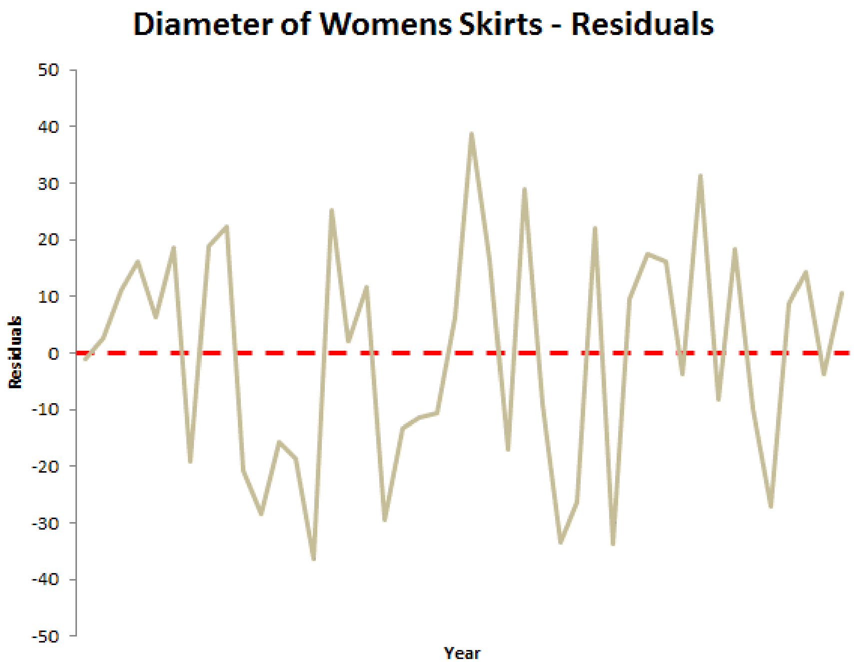
Statistical Tests to Improve the Model

- As for simple exponential smoothing, we can check whether the predictive model could be improved upon by checking whether the in-sample forecast errors show non-zero autocorrelations at lags 1-20. For example, for the skirt hem data, we can make a correlogram, and carry out the Ljung-Box test, by typing:

```
  Box-Ljung test
data:  skirtsseriesforecasts2$residuals
X-squared = 19.7312, df = 20, p-value = 0.4749
```

- Despite having the Lag = 5 exceed the significance bounds, when we carry out the Ljung-Box test, the p-value is 0.47, indicating that there is little evidence of non-zero autocorrelations in the in-sample forecast errors at lags 1-20.

Statistical Tests to Improve the Model

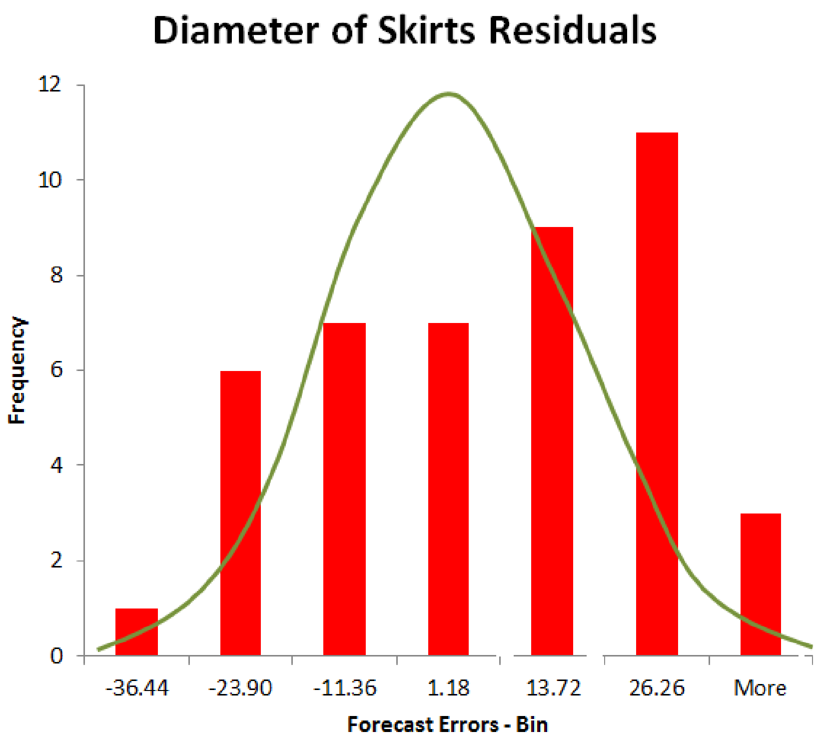


- As for simple exponential smoothing, we should also check that the forecast errors have constant variance over time, and are normally distributed with mean zero.
- The forecast errors appear to exhibit constant variance over time.

Statistical Tests to Improve the Model

- The time plot of forecast errors shows that the forecast errors have roughly constant variance over time.

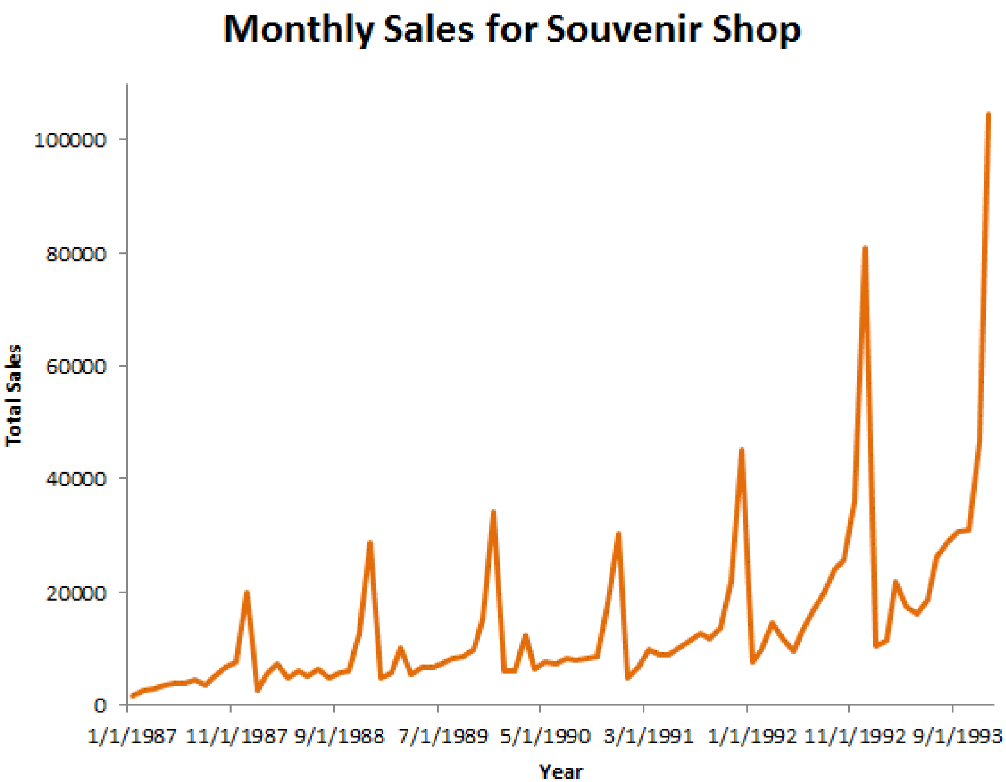
- The histogram of forecast errors show that it is plausible that the forecast errors are normally distributed with mean zero and constant variance.



- **Conclusion:** The Ljung-Box test shows that there is little evidence of autocorrelations in the forecast errors, while the time plot and histogram of forecast errors show that it is plausible that the forecast errors are normally distributed with mean zero and constant variance. Therefore, we can conclude that Holt's exponential smoothing provides an adequate predictive model for skirt hem diameters, which probably cannot be improved upon.

Holt-Winters Exponential Smoothing

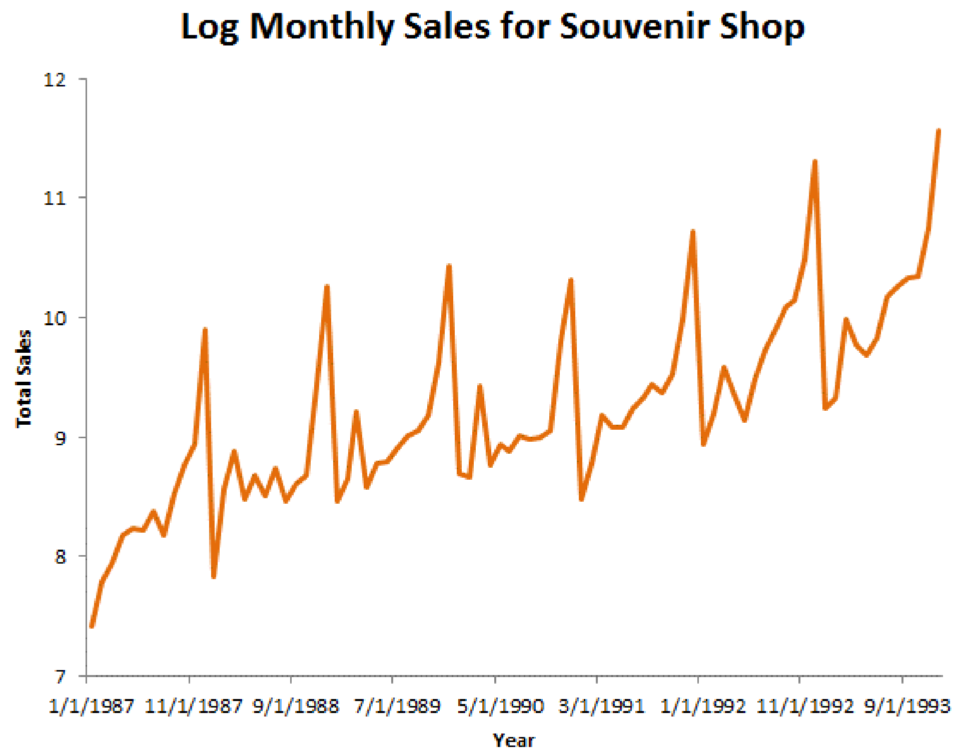
Holt-Winters Exponential Smoothing



- Monthly sales for the souvenir shop at a beach resort town in Queensland, Australia
- Y-Axis: # Souvenirs
- X-Axis: Time in Months
- In this case, it appears that an additive model is not appropriate for describing this time series, since the size of the seasonal fluctuations and random fluctuations seem to increase with the level of the time series.
- Thus, we may need to adjust this multiplicative time series in order to get a transformed time series that can be described using an additive model. For example, we can transform the time series by calculating the natural log of the original data:

Holt-Winters Exponential Smoothing

- For example, we can transform the time series by calculating the natural log of the original data:
- Here we can see that the size of the seasonal fluctuations and random fluctuations in the log-transformed time series seem to be roughly constant over time, and do not depend on the level of the time series. Thus, the log-transformed time series can probably be described using an additive model.
- This chart can be described as having an increasing (decreasing) trend and seasonality.



Holt-Winters Exponential Smoothing

- If you have a time series that can be described using an additive model with increasing or decreasing trend and seasonality, you can use Holt-Winters exponential smoothing to make short-term forecasts.

- Holt-Winters exponential smoothing estimates the level, slope and seasonal component at the current time point. Smoothing is controlled by three parameters: alpha, beta, and gamma, for the estimates of the level, slope b of the trend component, and the seasonal component, respectively, at the current time point. The parameters alpha, beta and gamma all have values between 0 and 1, and values that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values.

- The simple exponential smoothing method provides a way of estimating the level at the current time point. Smoothing is controlled by the parameter alpha; for the estimate of the level at the current time point. The value of alpha; lies between 0 and 1. Values of alpha that are close to 0 mean that little weight is placed on the most recent observations when making forecasts of future values.

Holt-Winters Exponential Smoothing

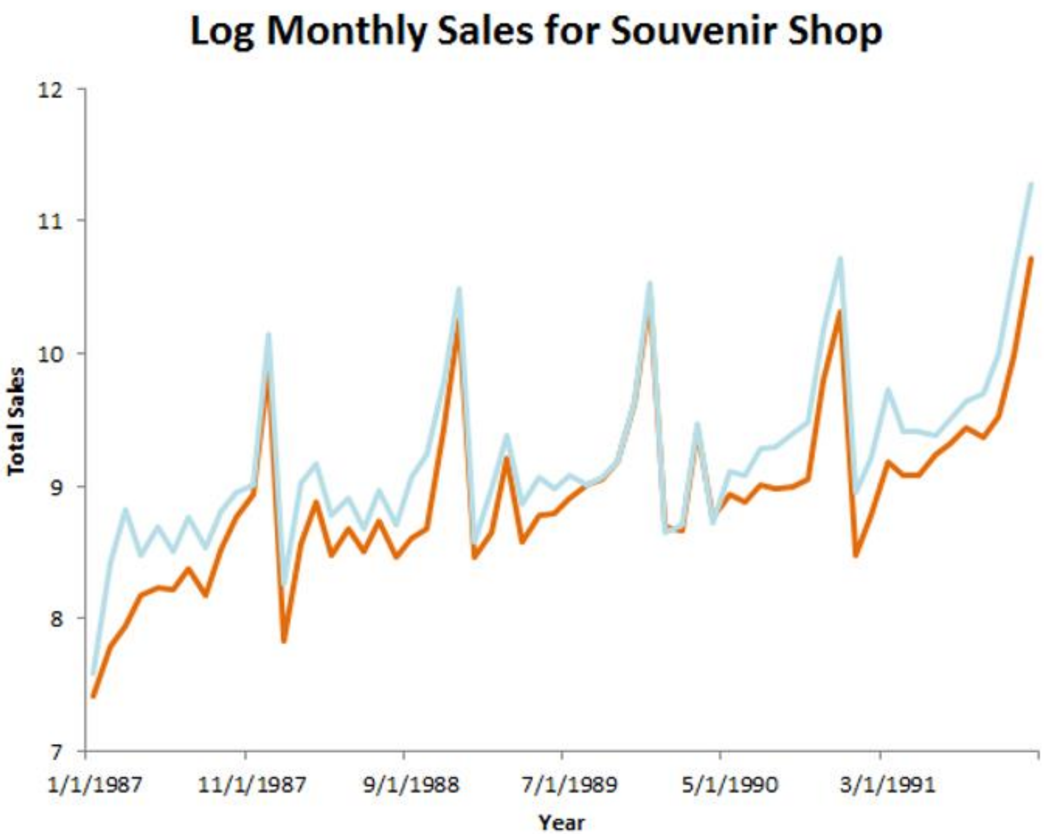
- To make forecasts, we can fit a predictive model using the HoltWinters() function.

```
logsouvenirtimeseries <- log(souvenirtimeseries)
souvenirtimeseriesforecasts <- HoltWinters(logsouvenirtimeseries)
souvenirtimeseriesforecasts
Holt-Winters exponential smoothing with trend and additive seasonal component.
Smoothing parameters:
alpha:  0.413418
beta :  0
gamma:  0.9561275
```

- The estimated values of alpha, beta and gamma are 0.41, 0.00, and 0.96, respectively. The value of alpha (0.41) is relatively low, indicating that the estimate of the level at the current time point is based upon both recent observations and some observations in the more distant past. The value of beta is 0.00, indicating that the estimate of the slope b of the trend component is not updated over the time series, and instead is set equal to its initial value.

- This makes good intuitive sense, as the level changes quite a bit over the time series, but the slope b of the trend component remains roughly the same. In contrast, the value of gamma (0.96) is high, indicating that the estimate of the seasonal component at the current time point is just based upon very recent observations.

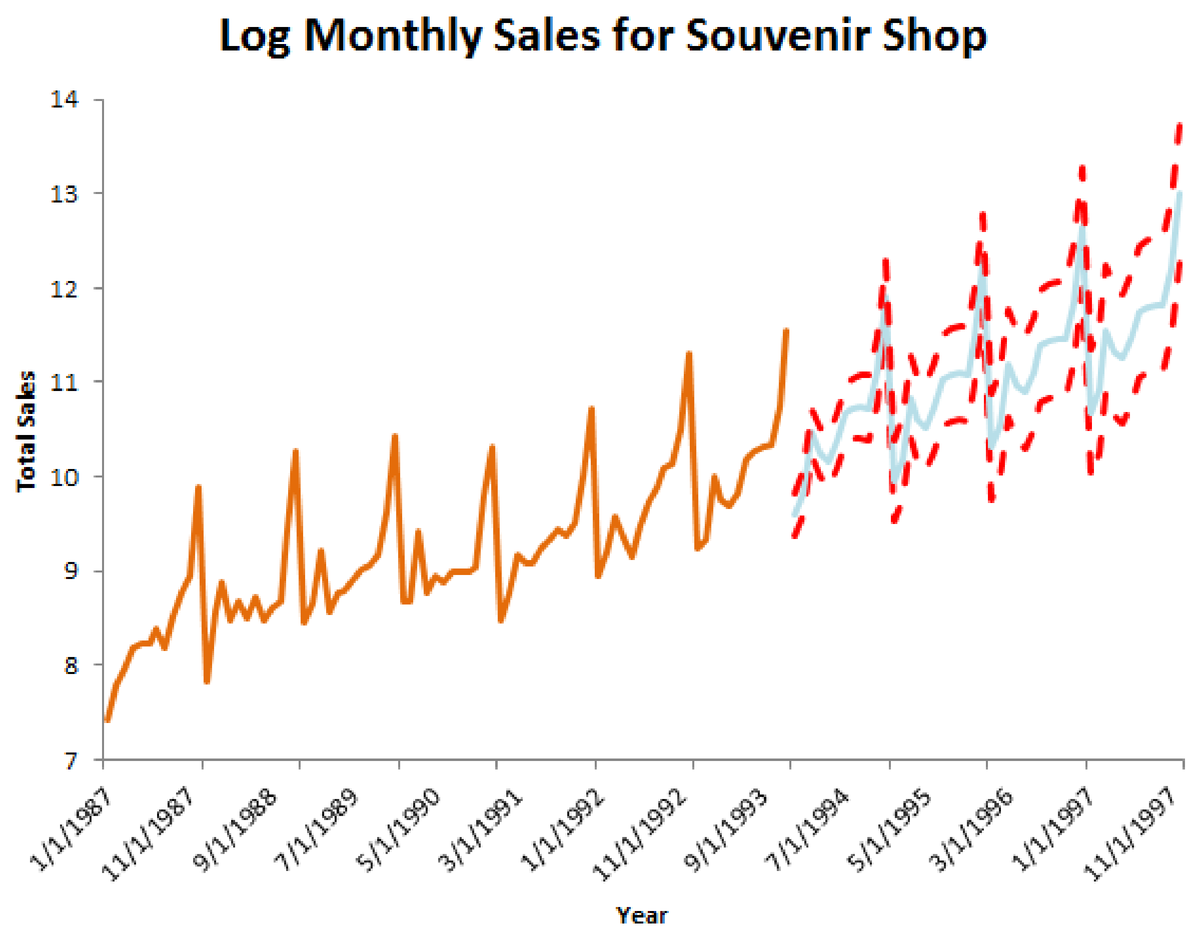
Holt-Winters Exponential Smoothing



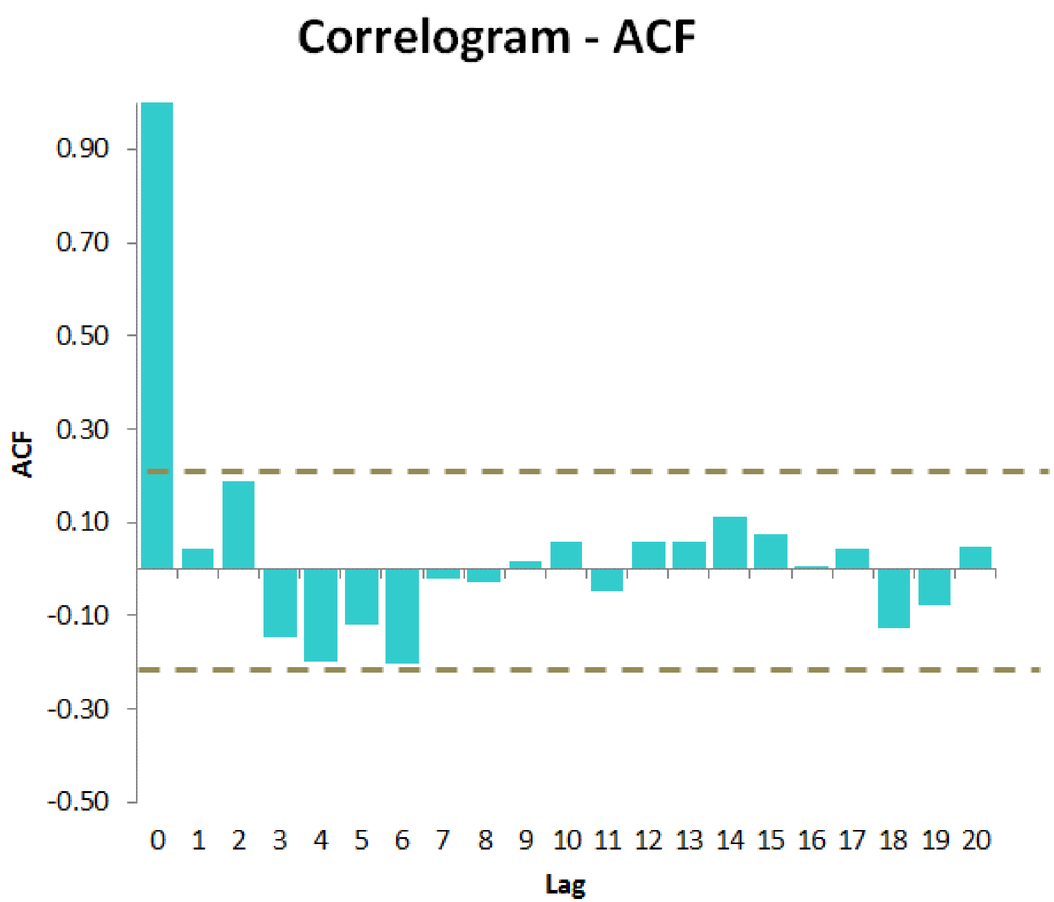
- As for simple exponential smoothing and Holt's exponential smoothing, we can plot the original time series as a brown line, with the forecasted values as a blue line on top of that:
- We see from the plot that the Holt-Winters exponential method is very successful in predicting the seasonal peaks, which occur roughly in November every year.

Holt-Winters Exponential Smoothing

- The forecasts are shown as a blue line, and the red dotted lines show 80% prediction intervals.



Statistical Tests to Improve the Model



- The correlogram shows that the autocorrelations for the in-sample forecast errors do not exceed the significance bounds for lags 1-20..

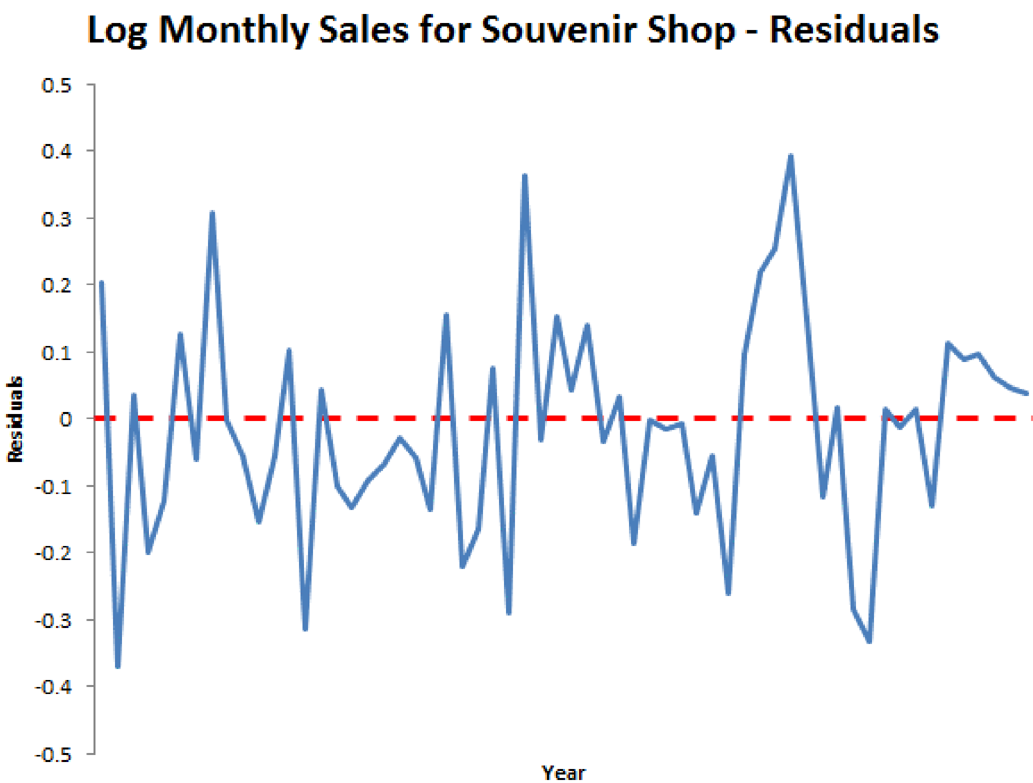
Statistical Tests to Improve the Model

- We can investigate whether the predictive model can be improved upon by checking whether the in-sample forecast errors show non-zero autocorrelations at lags 1-20, by making a correlogram and carrying out the Ljung-Box test:

```
acf(souvenirtimeseriesforecasts2$residuals, lag.max=20)
Box.test(souvenirtimeseriesforecasts2$residuals, lag=20, type="Ljung-Box")
Box-Ljung test
data:  souvenirtimeseriesforecasts2$residuals
X-squared = 17.5304, df = 20, p-value = 0.6183
```

- The p-value for Ljung-Box test is 0.61, indicating that there is little evidence of non-zero autocorrelations at lags 1-20

Statistical Tests to Improve the Model



- We can check whether the forecast errors have constant variance over time.
- The forecast errors appear to exhibit constant variance over time.

Statistical Tests to Improve the Model

- From the histogram of forecast errors, it seems plausible that the forecast errors are normally distributed with mean zero.

- Conclusion: There is little evidence of autocorrelation at lags 1-20 for the forecast errors, and the forecast errors appear to be normally distributed with mean zero and constant variance over time. This suggests that Holt-Winters exponential smoothing provides an adequate predictive model of the log of sales at the souvenir shop, which probably cannot be improved upon. Furthermore, the assumptions upon which the prediction intervals were based are probably valid.

Auto-Regressive Integrated Moving Average Models

Introduction to ARIMA Models

- Exponential smoothing methods are useful for making forecasts, and make no assumptions about the correlations between successive values of the time series.

- However, if you want to make prediction intervals for forecasts made using exponential smoothing methods, the prediction intervals require that the forecast errors are uncorrelated and are normally distributed with mean zero and constant variance.

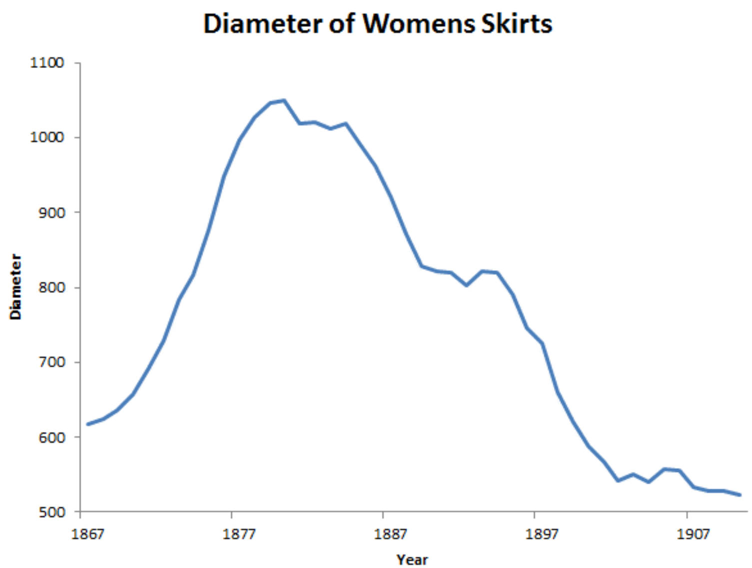
- While exponential smoothing methods do not make any assumptions about correlations between successive values of the time series, in some cases you can make a better predictive model by taking correlations in the data into account.

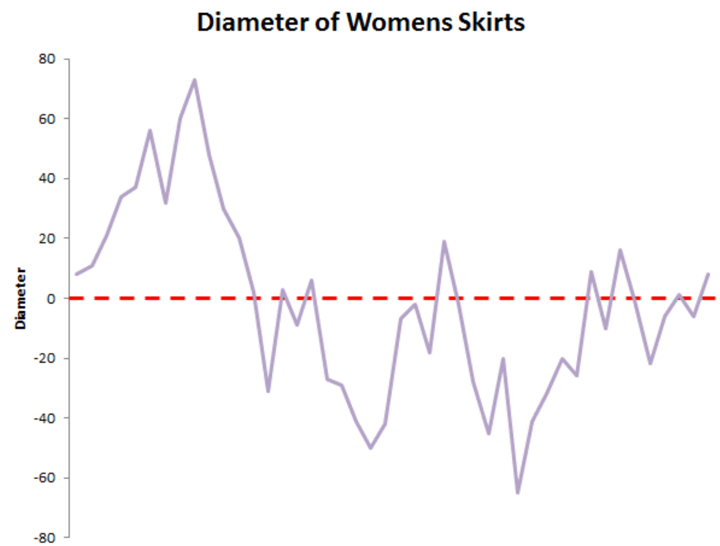
- Autoregressive Integrated Moving Average (ARIMA) models include an explicit statistical model for the irregular component of a time series, that allows for non-zero autocorrelations in the irregular component.

ARIMA(p,d,q)

Differencing a Time Series

- ARIMA models are defined for stationary time series. Therefore, if you start off with a non-stationary time series, you will first need to 'difference' the time series until you obtain a stationary time series. If you have to difference the time series d times to obtain a stationary series, then you have an ARIMA(p,d,q) model, where d is the order of differencing used.
- For example, the time series of the annual diameter of women's skirts at the hem, from 1866 to 1911 is not stationary in mean, as the level changes a lot over time:

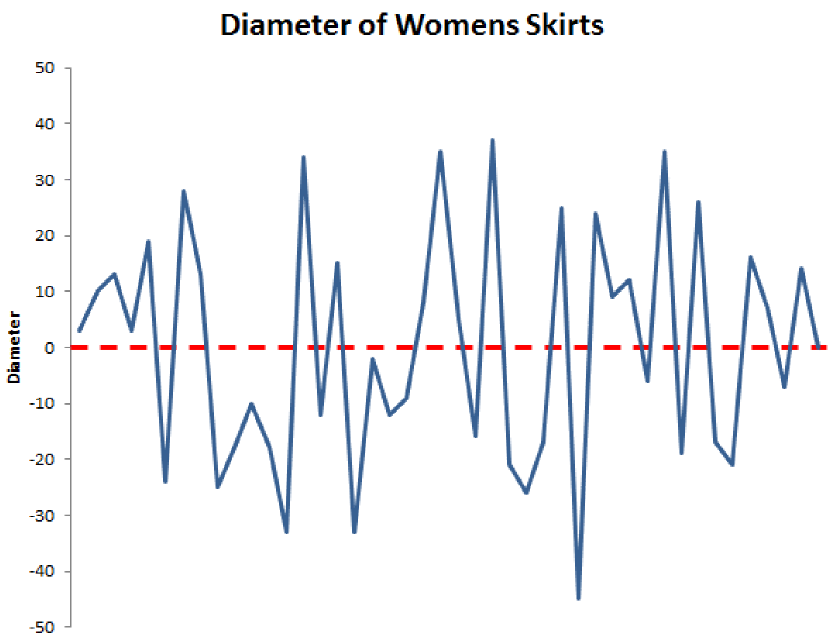




- Original Dataset

- Difference = 1
- The resulting time series of first differences (above) does not appear to be stationary in mean.

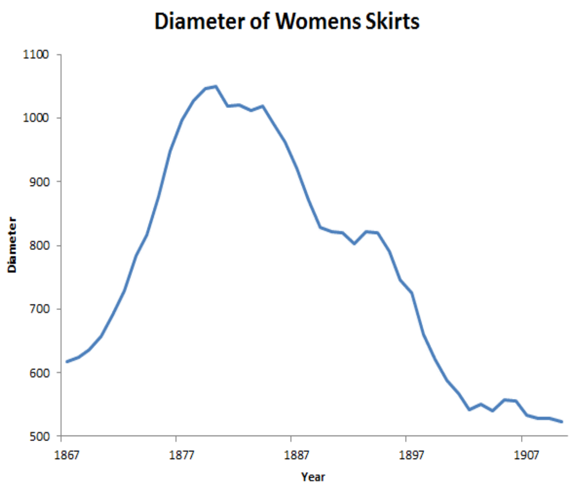
Differencing a Time Series

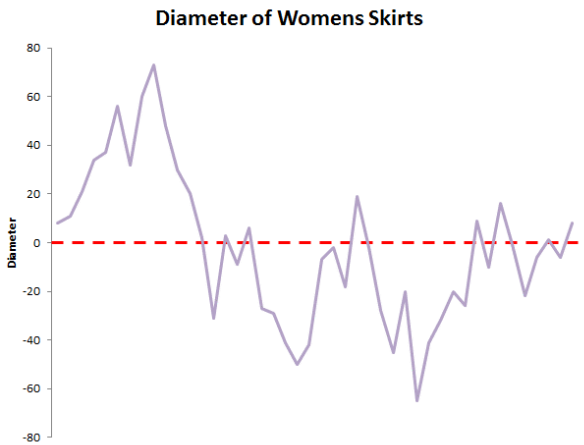


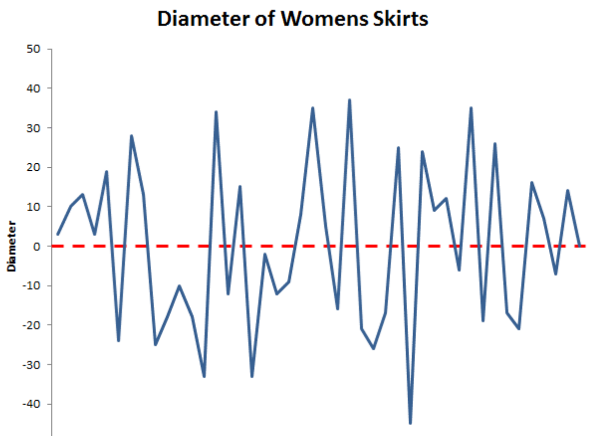
- Difference = 2

- The time series of second differences does appear to be stationary in mean and variance, as the level of the series stays roughly constant over time, and the variance of the series appears roughly constant over time.
- If you need to difference your original time series data d times in order to obtain a stationary time series, this means that you can use an ARIMA(p,d,q) model for your time series, where d is the order of differencing used.
- For example, for the time series of the diameter of women's skirts, we had to difference the time series twice, and so the order of differencing (d) is 2. This means that you can use an ARIMA(p,2,q) model for your time series.
- The next step is to figure out the values of p and q for the ARIMA model.

Differencing a Time Series



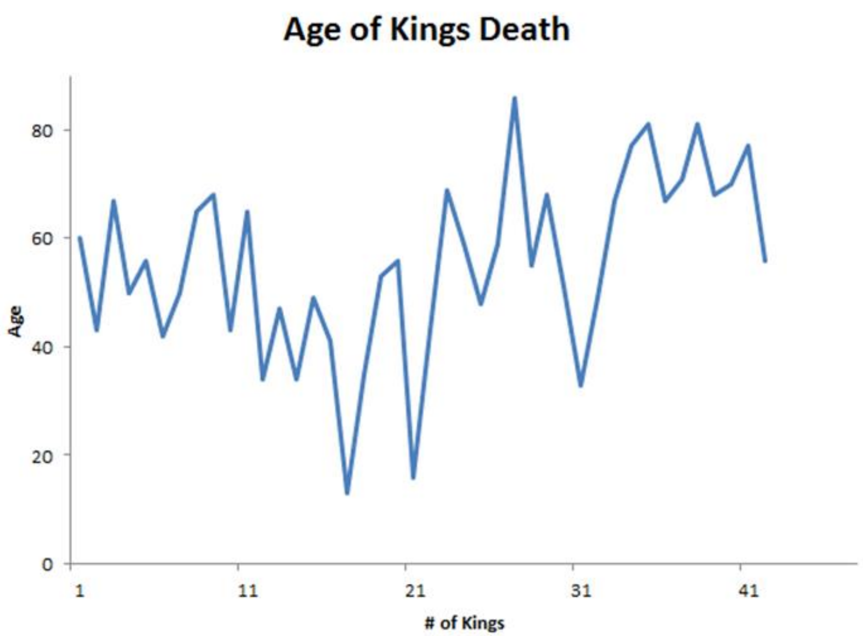


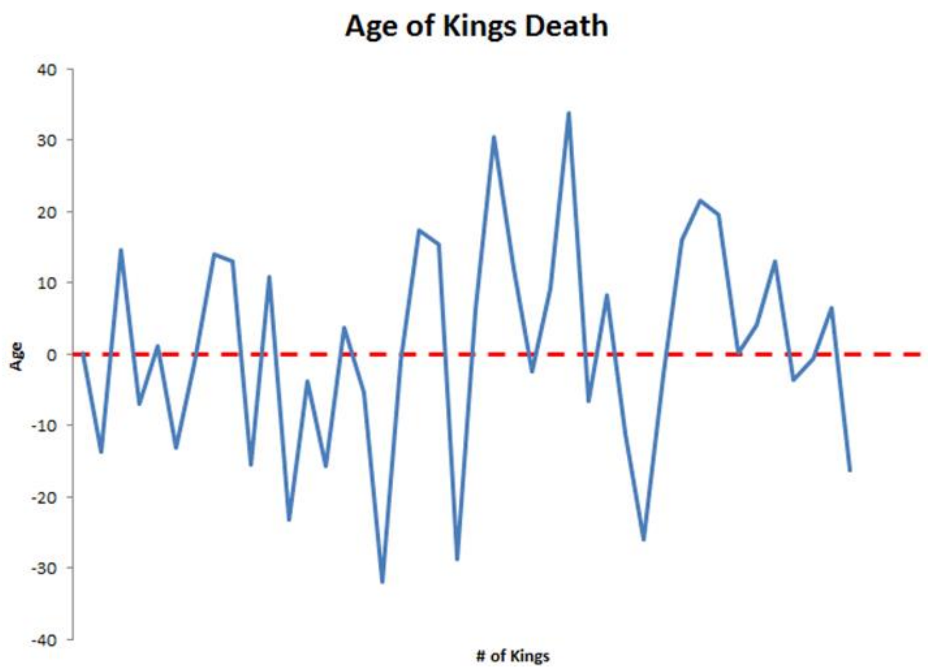


- Original = 1
- Difference = 1
- Difference = 2

Additional Differencing Example

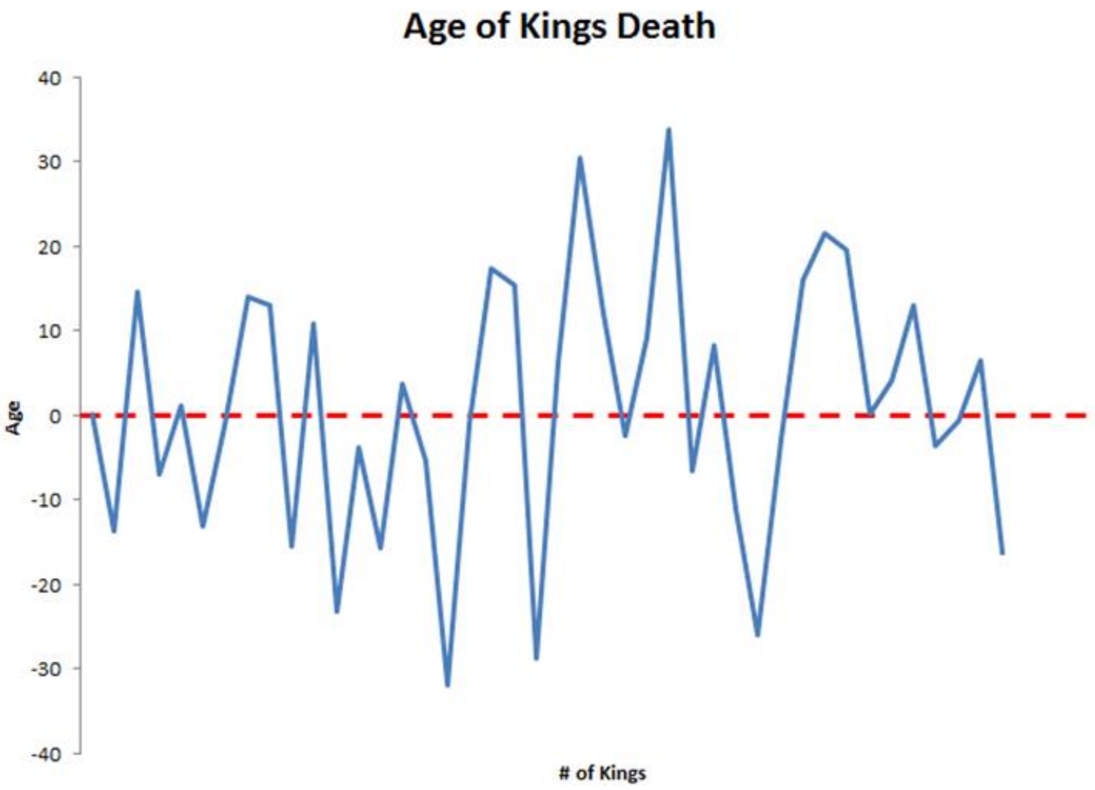
- Another example is the time series of the age of death of the successive kings of England:





- Original Dataset
- Difference = 1

Additional Differencing Example



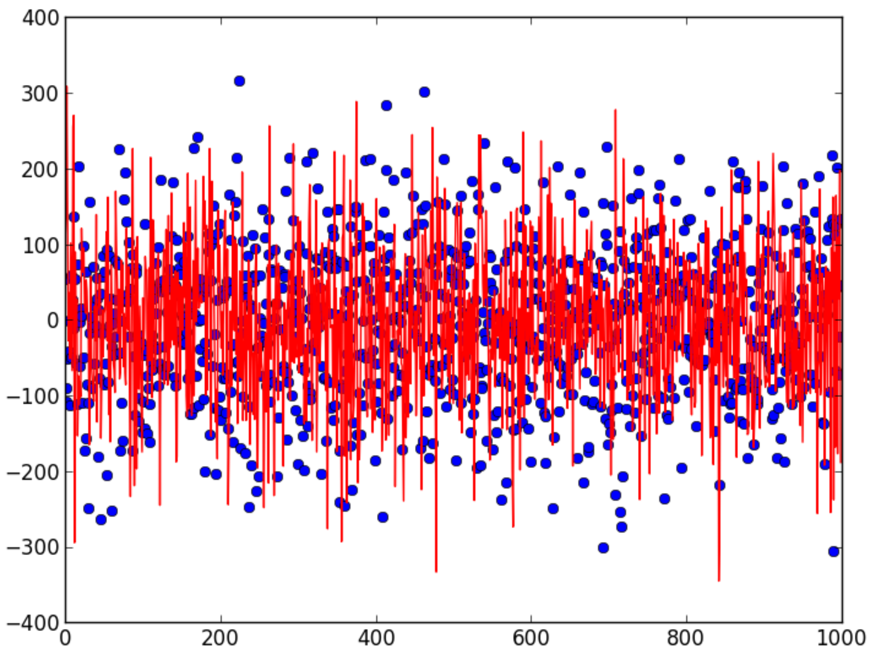
- Difference = 1

- The time series of first differences appears to be stationary in mean and variance, and so an ARIMA(p,1,q) model is probably appropriate for the time.
- By taking the time series of first differences, we have removed the trend component of the time series of the ages at death of the kings, and are left with an irregular component.
- We can now examine whether there are correlations between successive terms of this irregular component; if so, this could help us to make a predictive model for the ages at death of the kings.

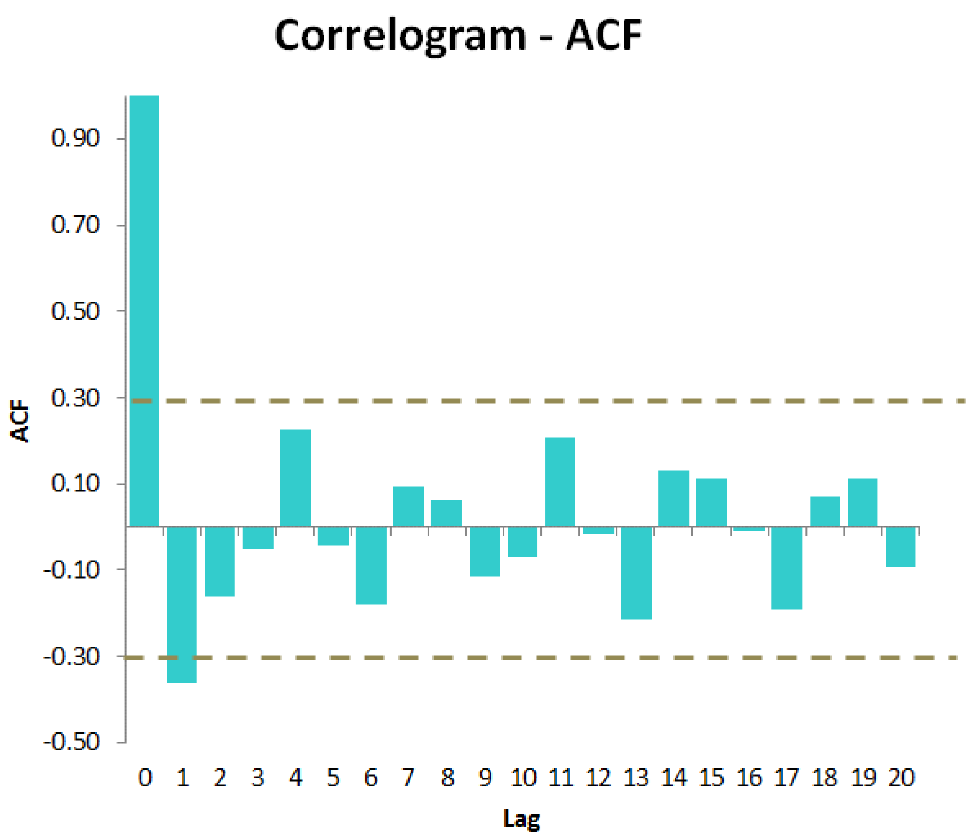
Selecting a Candidate ARIMA Models

- If your time series is stationary, or if you have transformed it to a stationary time series by differencing d times, the next step is to select the appropriate ARIMA model, which means finding the values of most appropriate values of p and q for an ARIMA(p,d,q) model.

- To do this, you usually need to examine the correlogram (ACF) and partial correlogram (PACF) of the stationary time series.



Statistical Tests to Improve the Model



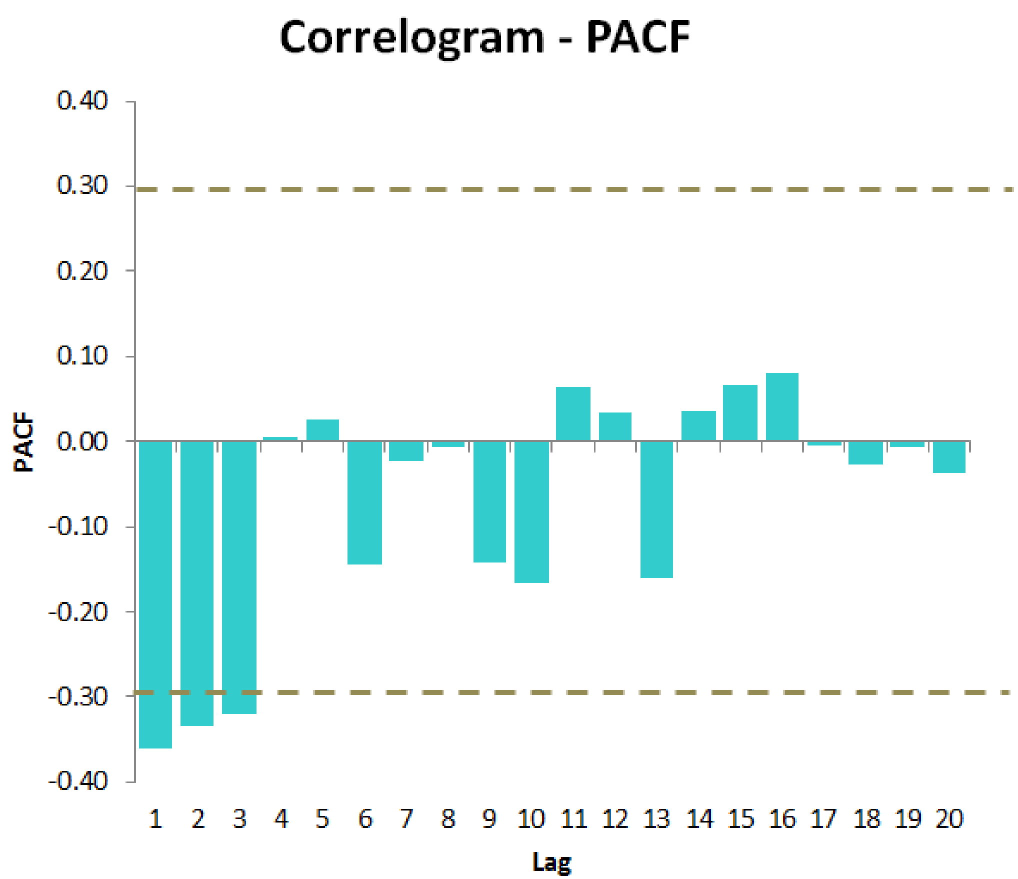
- Example of the Ages at Death of the Kings of England

- For example, to plot the correlogram for lags 1-20 of the once differenced time series of the ages at death of the kings of England, and to get the values of the autocorrelations, we type:

- We see from the correlogram that the autocorrelation at lag 1 (-0.360) exceeds the significance bounds, but all other autocorrelations between lags 1-20 do not exceed the significance bounds.

Statistical Tests to Improve the Model

- Example of the Ages at Death of the Kings of England
- Plot the partial correlogram (PACF) for lags 1-20 for the once differenced time series of the ages at death of the English kings, and get the values of the partial autocorrelations:
- The partial correlogram shows that the partial autocorrelations at lags 1, 2 and 3 exceed the significance bounds, are negative, and are slowly decreasing in magnitude with increasing lag (lag 1: -0.360, lag 2: -0.335, lag 3:-0.321). The partial autocorrelations tail off to zero after lag 3.



Selecting a Candidate ARMA Model

ARIMA(p,d,q) & ARMA(p,q)

- Since the ACF is zero after lag 1, and the PACF tails off to zero after lag 3, this means that the following ARMA (autoregressive moving average) models are possible for the time series of first differences:
  - an ARMA(3,0) model, that is, an autoregressive model of order p=3, since the partial autocorrelogram is zero after lag 3, and the autocorrelogram tails off to zero (although perhaps too abruptly for this model to be appropriate)
  - an ARMA(0,1) model, that is, a moving average model of order q=1, since the autocorrelogram is zero after lag 1 and the partial autocorrelogram tails off to zero
  - an ARMA(p,q) model, that is, a mixed model with p and q greater than 0, since the autocorrelogram and partial correlogram tail off to zero (although the correlogram probably tails off to zero too abruptly for this model to be appropriate)

- We use the principle of parsimony to decide which model is best: that is, we assume that the model with the fewest parameters is best. The ARMA(3,0) model has 3 parameters, the ARMA(0,1) model has 1 parameter, and the ARMA(p,q) model has at least 2 parameters. Therefore, the ARMA(0,1) model is taken as the best model.

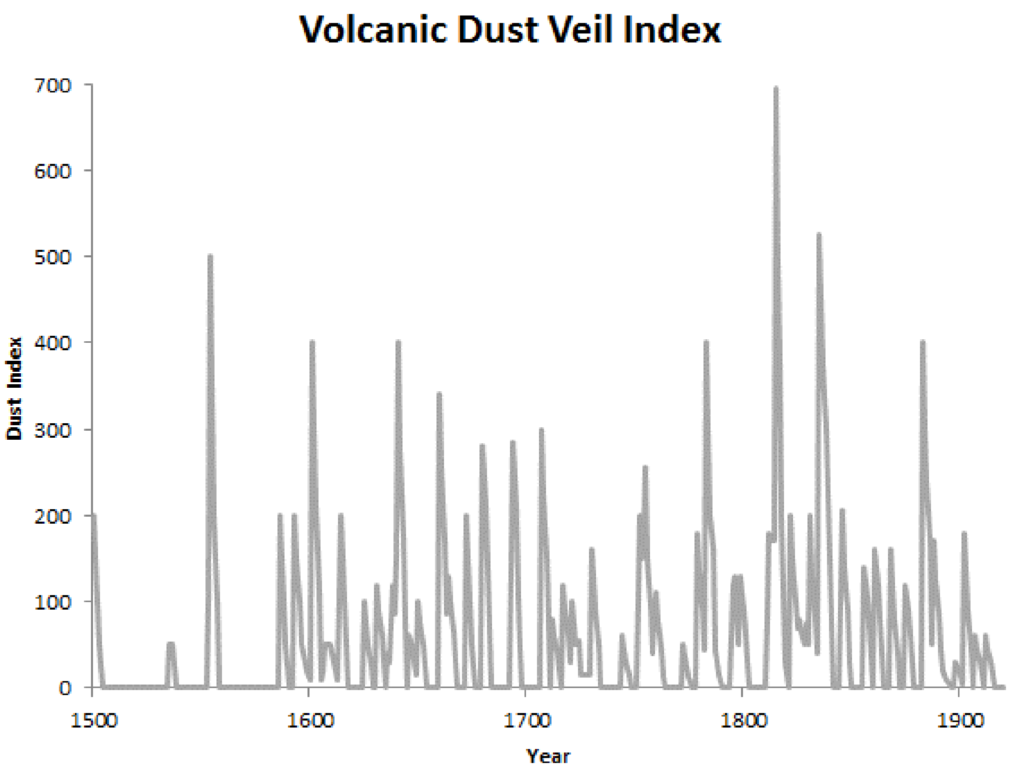
.

Selecting a Candidate ARIMA Model

- Since an ARMA(0,1) model (with p=0, q=1) is taken to be the best candidate model for the time series of first differences of the ages at death of English kings, then the original time series of the ages of death can be modelled using an ARIMA(0,1,1) model (with p=0, d=1, q=1, where d is the order of differencing required).

- A MA (moving average) model is usually used to model a time series that shows short-term dependencies between successive observations. Intuitively, it makes good sense that a MA model can be used to describe the irregular component in the time series of ages at death of English kings, as we might expect the age at death of a particular English king to have some effect on the ages at death of the next king or two, but not much effect on the ages at death of kings that reign much longer after that.

ARIMA - Example 2



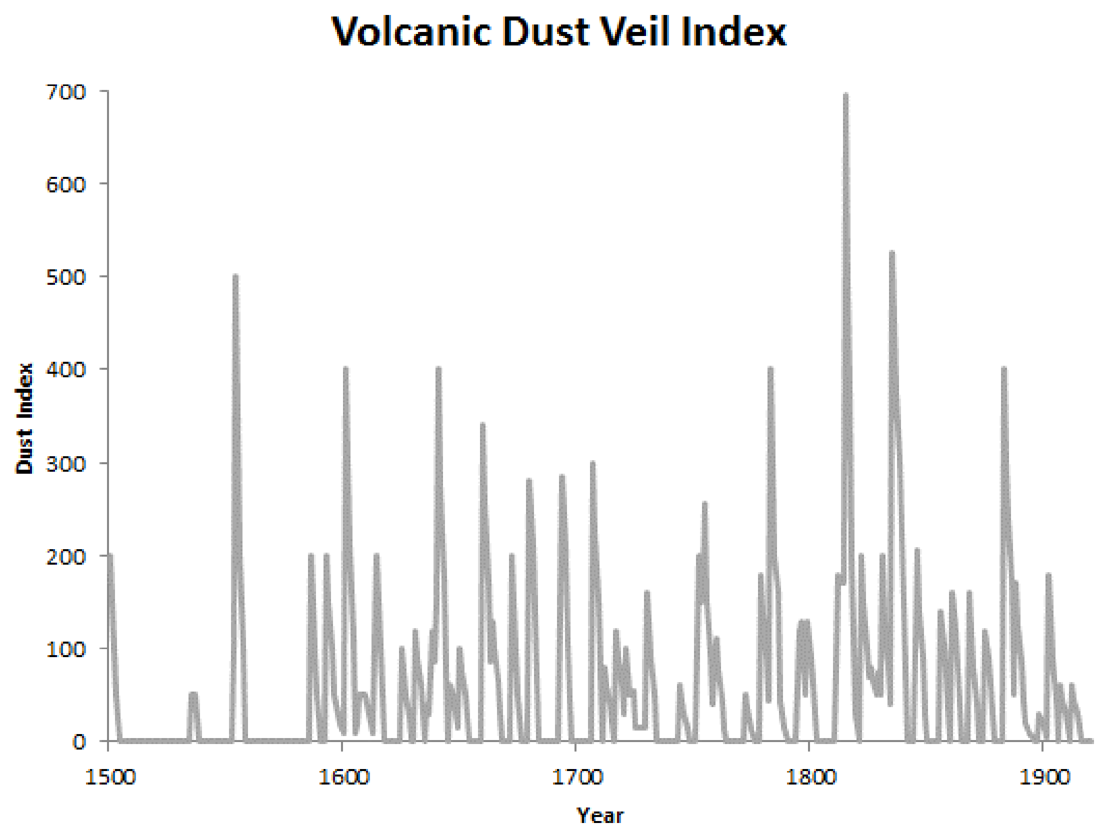
- The volcanic dust veil index data in the northern hemisphere, from 1500-1969 (original data from Hipel and Mcleod, 1994).
- This is a measure of the impact of volcanic eruptions' release of dust and aerosols into the environment.
- From the time plot, it appears that the random fluctuations in the time series are roughly constant in size over time, so an additive model is probably appropriate for describing this time series.

ARIMA - Example 2

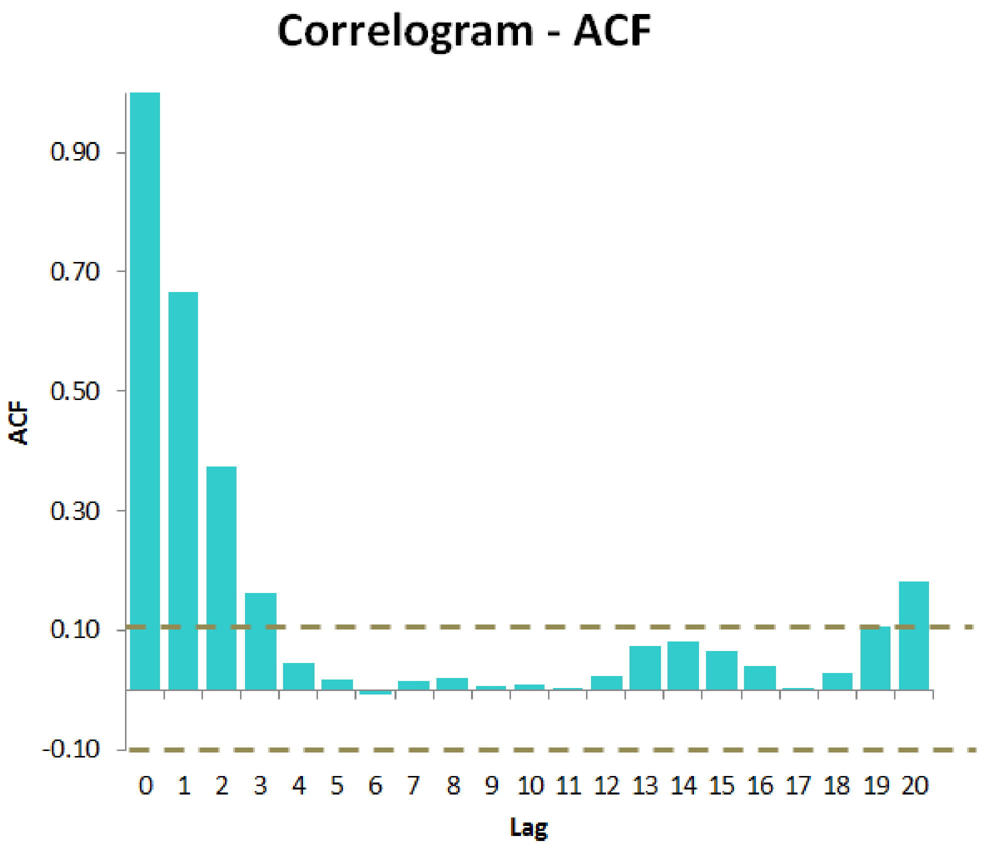
- Furthermore, the time series appears to be stationary in mean and variance, as its level and variance appear to be roughly constant over time.

- Therefore, we do not need to difference this series in order to fit an ARIMA model, but can fit an ARIMA model to the original series (the order of differencing required, d, is zero here).

- We can now plot a correlogram and partial correlogram for lags 1-20 to investigate what ARIMA model to use:



Statistical Tests to Improve the Model



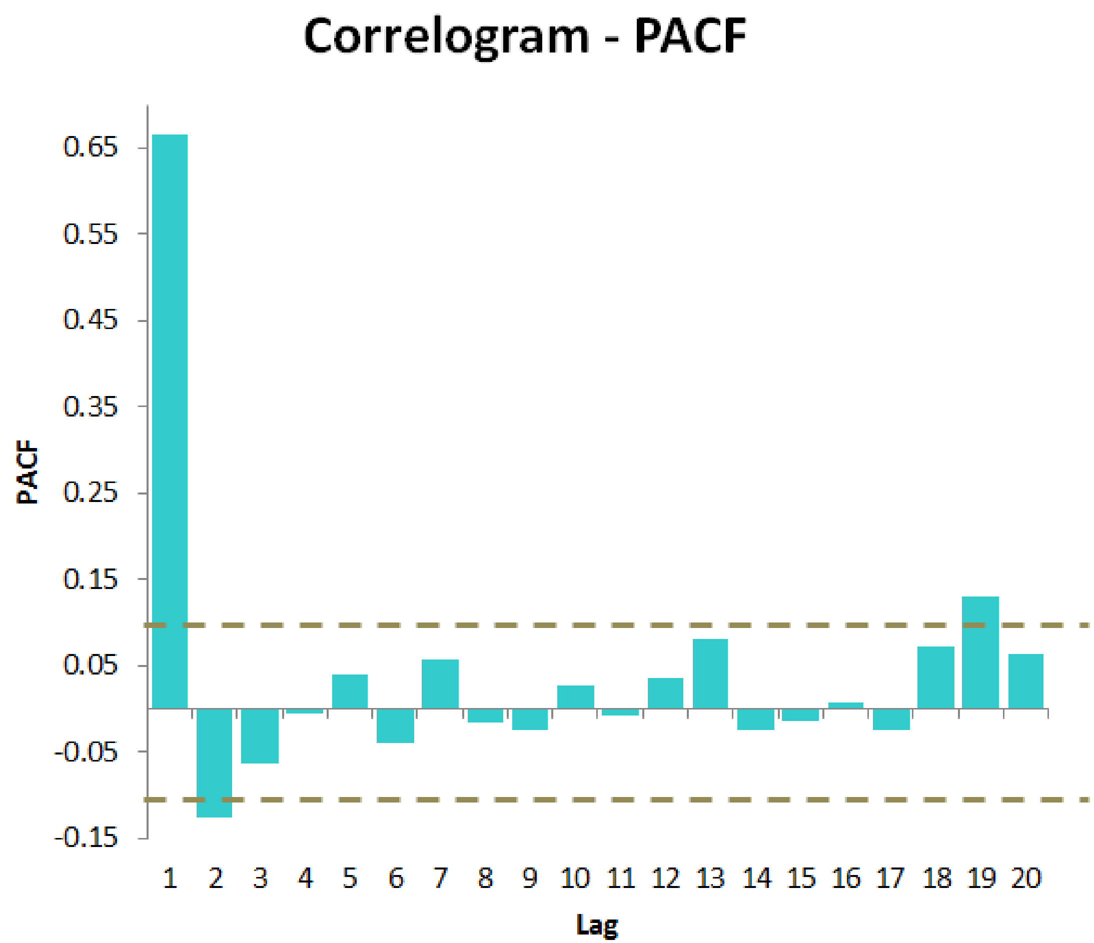
- We see from the correlogram that the autocorrelations for lags 1, 2 and 3 exceed the significance bounds, and that the autocorrelations tail off to zero after lag 3. The autocorrelations for lags 1, 2, 3 are positive, and decrease in magnitude with increasing lag (lag 1: 0.666, lag 2: 0.374, lag 3: 0.162).

- The autocorrelation for lags 19 and 20 exceed the significance bounds too, but it is likely that this is due to chance, since they just exceed the significance bounds (especially for lag 19), the autocorrelations for lags 4-18 do not exceed the significance bounds, and we would expect 1 in 20 lags to exceed the 95% significance bounds by chance alone.

Statistical Tests to Improve the Model

- From the partial autocorrelogram, we see that the partial autocorrelation at lag 1 is positive and exceeds the significance bounds (0.666), while the partial autocorrelation at lag 2 is negative and also exceeds the significance bounds (-0.126).

- The partial autocorrelations tail off to zero after lag 2.



Selecting a Candidate ARMA Model

ARIMA(p,d,q) & ARMA(p,q)

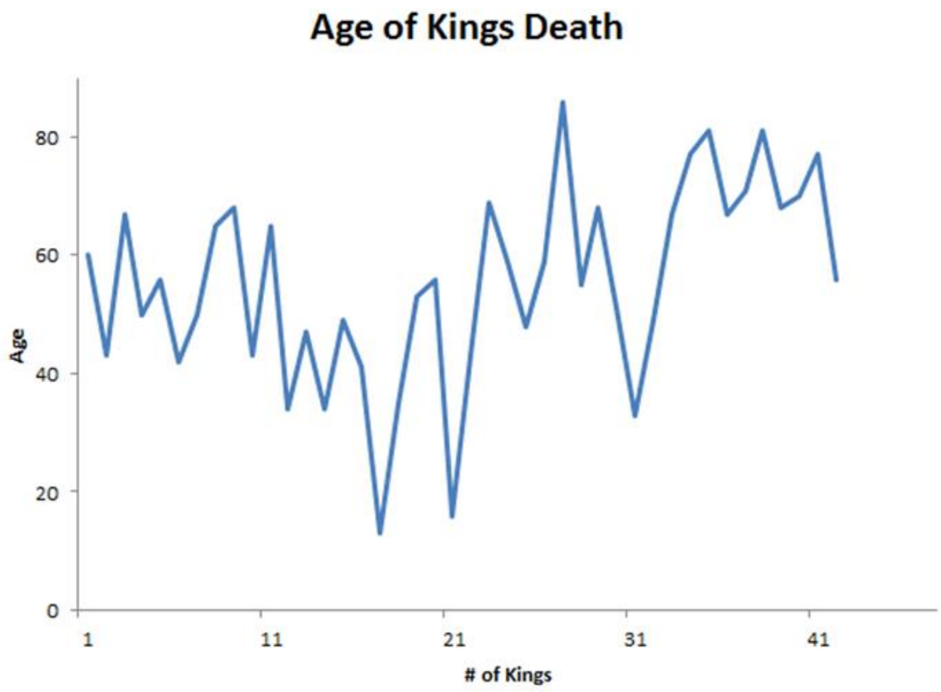
- Since the correlogram tails off to zero after lag 3, and the partial correlogram is zero after lag 2, the following ARMA models are possible for the time series:
  - an ARMA(2,0) model, since the partial autocorrelogram is zero after lag 2, and the correlogram tails off to zero after lag 3, and the partial correlogram is zero after lag 2
  - an ARMA(0,3) model, since the autocorrelogram is zero after lag 3, and the partial correlogram tails off to zero (although perhaps too abruptly for this model to be appropriate)
  - an ARMA(p,q) mixed model, since the correlogram and partial correlogram tail off to zero (although the partial correlogram perhaps tails off too abruptly for this model to be appropriate)

- The ARMA(2,0) model has 2 parameters, the ARMA(0,3) model has 3 parameters, and the ARMA(p,q) model has at least 2 parameters. Therefore, using the principle of parsimony, the ARMA(2,0) model and ARMA(p,q) model are equally good candidate models.

Selecting a Candidate ARIMA Model

- If an ARMA(2,0) model (with p=2, q=0) is used to model the time series of volcanic dust veil index, it would mean that an ARIMA(2,0,0) model can be used (with p=2, d=0, q=0, where d is the order of differencing required). Similarly, if an ARMA(p,q) mixed model is used, where p and q are both greater than zero, than an ARIMA(p,0,q) model can be used.
- An AR (autoregressive) model is usually used to model a time series which shows longer term dependencies between successive observations.
- Intuitively, it makes sense that an AR model could be used to describe the time series of volcanic dust veil index, as we would expect volcanic dust and aerosol levels in one year to affect those in much later years, since the dust and aerosols are unlikely to disappear quickly.

Forecasting Using ARIMA

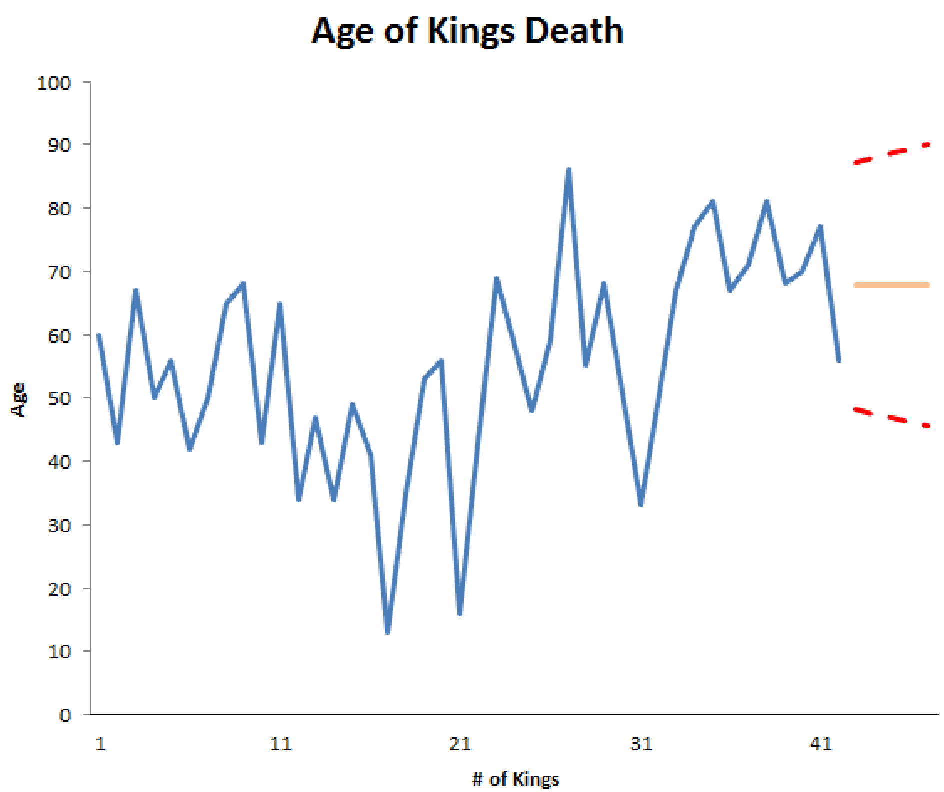


- Example of the Ages at Death of the Kings of England
- Once you have selected the best candidate ARIMA(p,d,q) model for your time series data, you can estimate the parameters of that ARIMA model, and use that as a predictive model for making forecasts for future values of your time series.
- For example, we discussed above that an ARIMA(0,1,1) model seems a plausible model for the ages at deaths of the kings of England.

Forecasting Using ARIMA

- We can plot the observed ages of death for the first 42 kings, as well as the ages that would be predicted for these 42 kings and for the next 5 kings using our ARIMA(0,1,1) model, by typing:

- As in the case of exponential smoothing models, it is a good idea to investigate whether the forecast errors of an ARIMA model are normally distributed with mean zero and constant variance, and whether the are correlations between successive forecast errors.



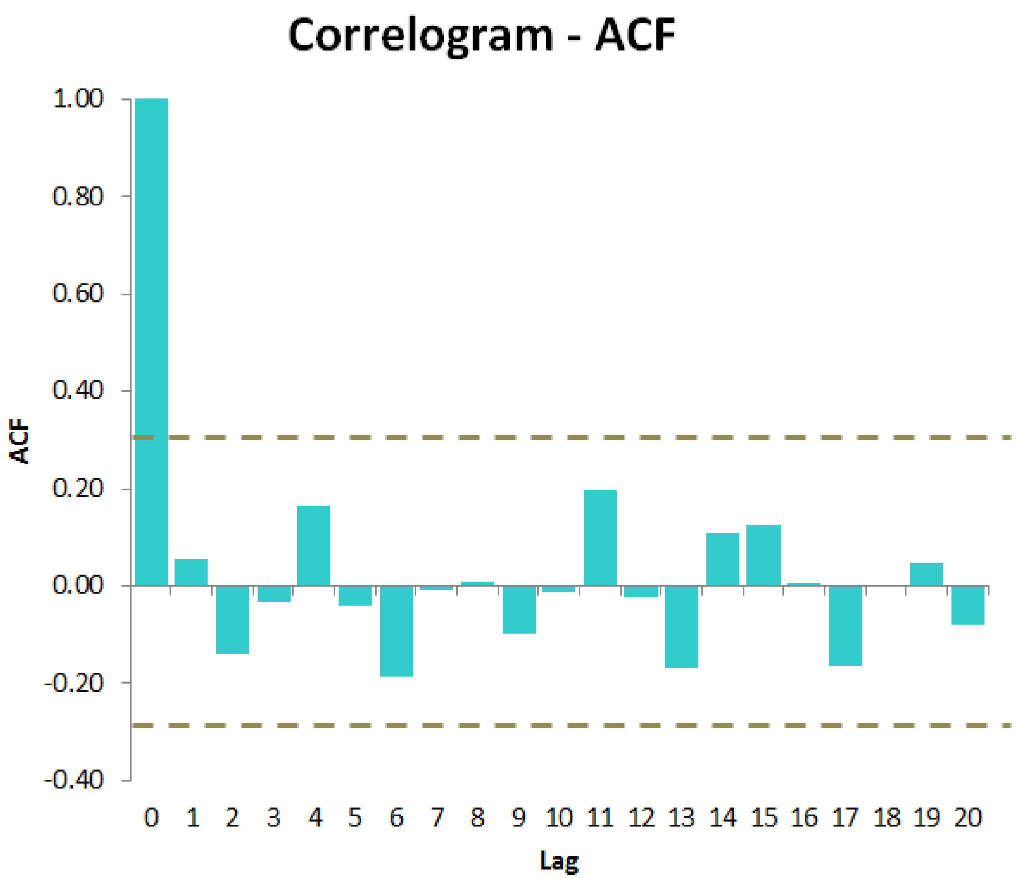
Statistical Tests to Improve the Model

- For example, we can make a correlogram of the forecast errors for our ARIMA(0,1,1) model for the ages at death of kings, and perform the Ljung-Box test for lags 1-20, by typing:

```
acf(kingstimeseriesforecasts$residuals, lag.max=20)
Box.test(kingstimeseriesforecasts$residuals, lag=20, type="Ljung-Box")
Box-Ljung test
data:  kingstimeseriesforecasts$residuals
X-squared = 13.5844, df = 20, p-value = 0.851
```

- The p-value for Ljung-Box test is 0.851, indicating that there is little evidence of non-zero autocorrelations at lags 1-20

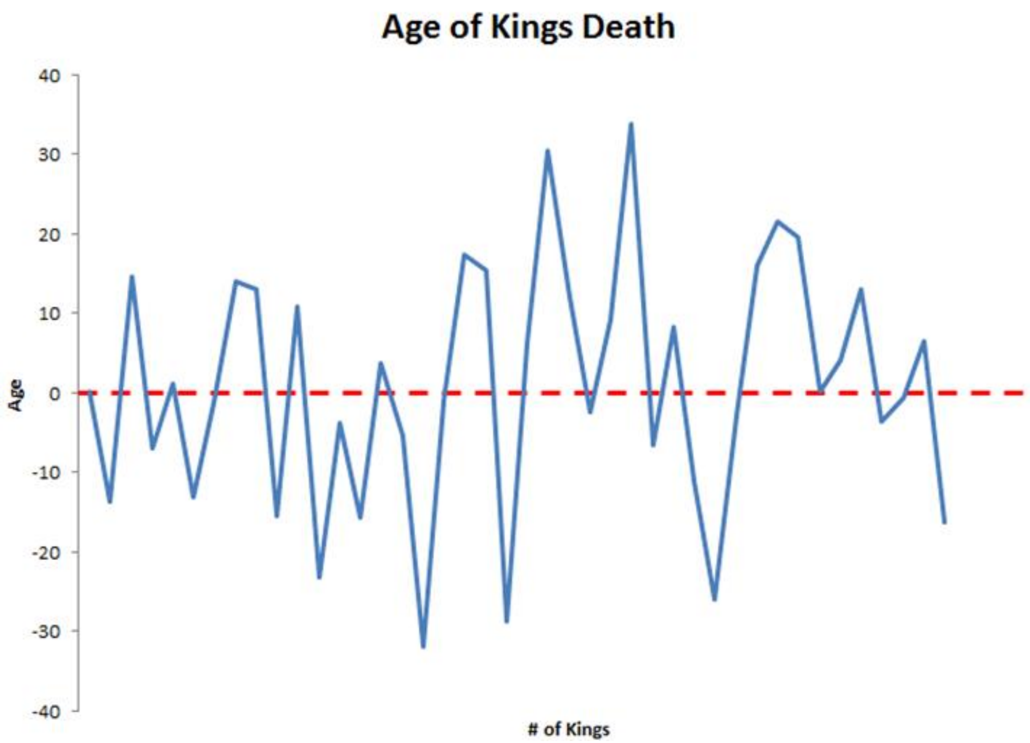
Statistical Tests to Improve the Model



- The correlogram shows that none of the sample autocorrelations for lags 1-20 exceed the significance bounds.

- Also, the p-value for the Ljung-Box test is 0.9, we can conclude that there is very little evidence for non-zero autocorrelations in the forecast errors at lags 1-20.

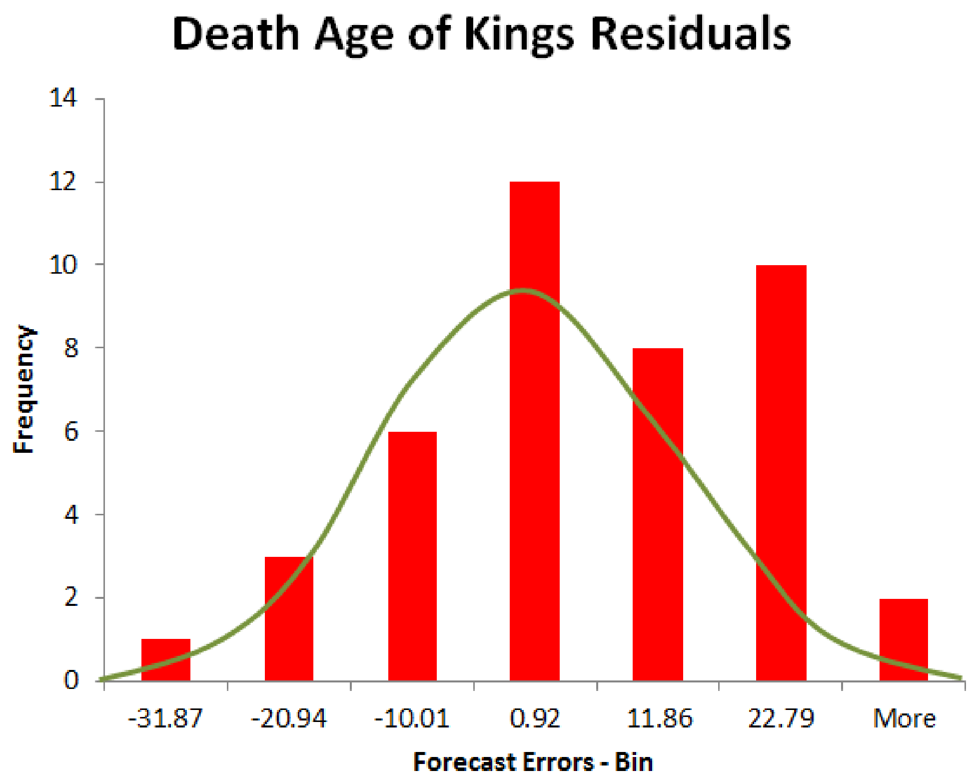
Statistical Tests to Improve the Model



- We can check whether the forecast errors have constant variance over time, and are normally distributed with mean zero, by making a time plot of the forecast errors and a histogram (with overlaid normal curve):

Statistical Tests to Improve the Model

- The time plot of the in-sample forecast errors shows that the variance of the forecast errors seems to be roughly constant over time (though perhaps there is slightly higher variance for the second half of the time series).
- The histogram of the time series shows that the forecast errors are roughly normally distributed and the mean seems to be close to zero. Therefore, it is plausible that the forecast errors are normally distributed with mean zero and constant variance.

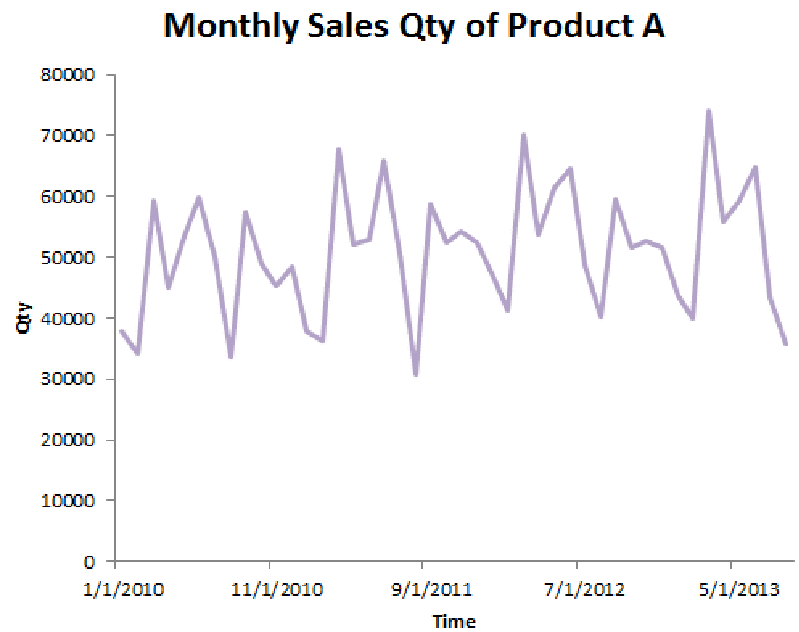


- **Conclusion:** Since successive forecast errors do not seem to be correlated, and the forecast errors seem to be normally distributed with mean zero and constant variance, the ARIMA(0,1,1) does seem to provide an adequate predictive model for the ages at death of English kings.

Practical Application Example
Product Sales Quantity

Understanding the Data

Time	Product A
1/1/2010	37884
2/1/2010	34287
3/1/2010	59362
4/1/2010	45002
5/1/2010	53332
6/1/2010	59909
7/1/2010	49942
8/1/2010	33522
9/1/2010	57556
10/1/2010	48895
11/1/2010	45268
12/1/2010	48460
etc...	etc...

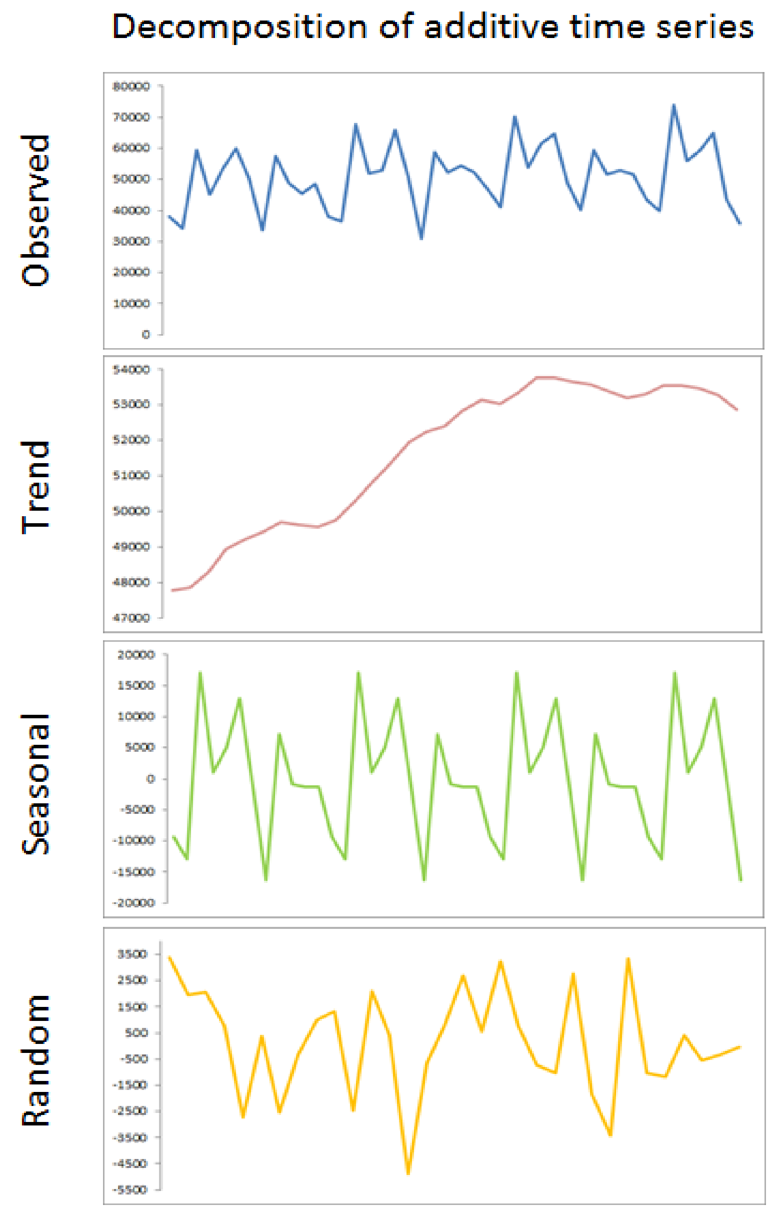


- What does the time series graph show us in terms of the modeling technique to use?
  - Is there a trend? If so, is there any direction?
  - Is there seasonality?

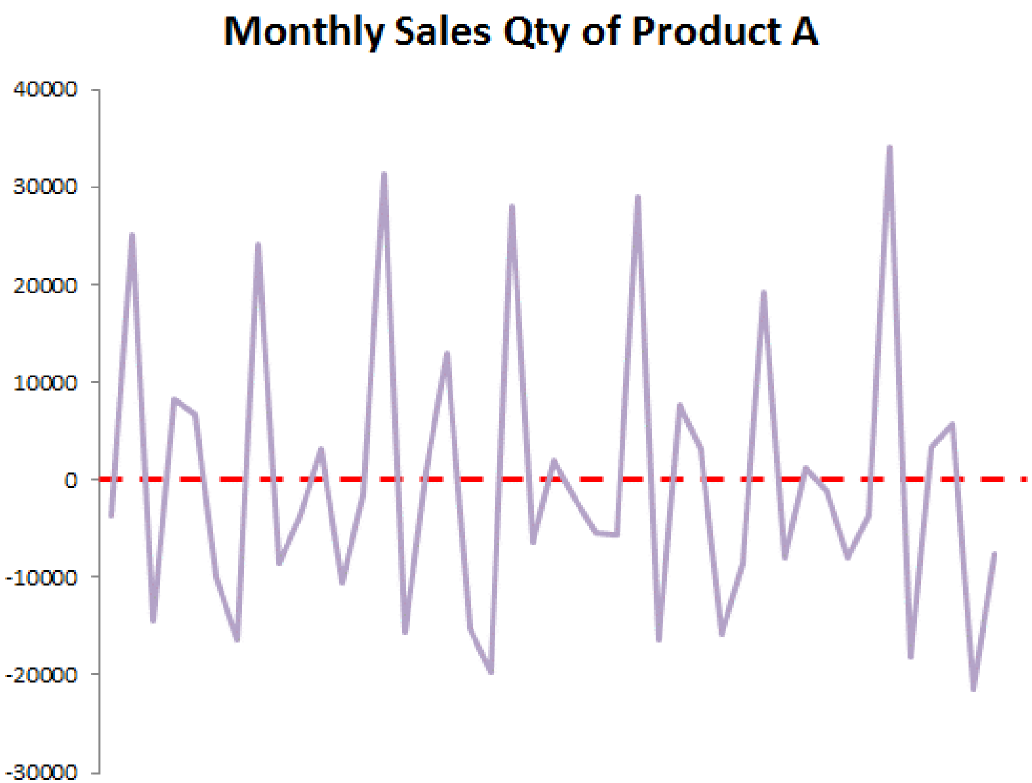
- Which potential model(s) would you use for this application?
  - Holt-Winter's exponential smoothing
  - ARIMA(p,d,q)

Decompose the Time Series

- In order to further isolate the algorithm to use for analysis, lets decompose the time series to see the individual components.
- There appears to be a slight positive trend line (exaggerated in the graphic due to the axis parameters)
- There is a regular seasonal component to the data.
- Finally, the random component of the data reaches peaks of 3500 units and troughs of -4500 units. The gap of this spread is wide enough that we should use an ARIMA(p,d,q) model in order to capture this noise in our final analysis.



Differencing the Time Series



- Difference = 1

- The time series of first difference does appear to be stationary in mean and variance.
- Since we had to difference the time series once, and so the order of differencing (d) is 1. This means that you can use an ARIMA(p,1,q) model for your time series.
- The next step is to figure out the values of p and q for the ARIMA model.

Statistical Tests to Improve the Model

- There is a technique in R called auto.arima() which will identify the best candidate ARIMA(p,d,q) based upon some evaluation parameters (default is Akaike Information Criterion or AIC).

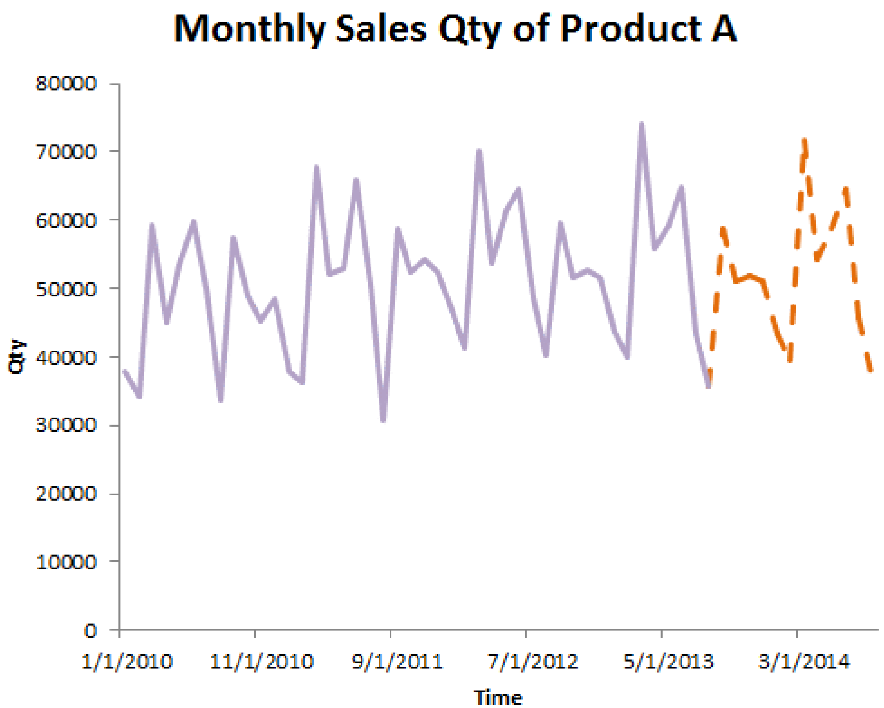
```
Series: ts2
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.7412  -0.3908
s.e.   0.1429   0.2064

sigma^2 estimated as 13331175:  log likelihood=-299.67
AIC=605.33   AICc=606.22   BIC=609.64
```

- When we process this technique the resulting model is an ARIMA(0,1,1). Please note that we can evaluate the ACF, PACF, Box-Ljung Test, etc… using the techniques outlined in the presentation but will skip this evaluation for brevity.

Forecasting the Results



Forecast Period	Product A Estimated Qty
9/1/2013	58715
10/1/2013	51094
11/1/2013	51835
12/1/2013	51061
1/1/2014	43475
2/1/2014	39498
3/1/2014	71626
4/1/2014	54353
5/1/2014	58725
6/1/2014	64579
7/1/2014	45697
8/1/2014	36199

- We specified that the forecast extends for t=12 or 12 months. The final forecast takes into consideration seasonal variations, evolving macro/micro trends, and leverages the random noise when building the qty estimates.
- The final estimated sales projections can be leveraged to determine procurement requirements, sales budgets/goals, etc…

Time Series – Generalized Least Square Model

Regression Models on Time Series Data

- There are many instances where we are interested in not only predicting the development of a time series dataset but we also want to understand how individual independent variables (X) impact the dependent variable (Y) spanning time.

- Ex. Sales development over time taking GDP, unemployment rate, and other variables into account.

- We would naïvely initially approach this task using the standard OLS Linear regression model with a form of:

$$Y = \text{Intercept} + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon$$

- Note: A critical assumption when building the OLS is that the errors (ε) are homoscedastic and serially uncorrelated. It is common when working with Time Series data that the errors are unlikely to be independent and are serially correlated.

Regression Models on Time Series Data

- OLS regression will not be sufficient when modeling time series data!!!!

- In order to address the serial correlation issue, we can employ a Generalized Least Squares model fit by Maximum Likelihood. This approach focuses on resolving some of the underlying concerns with the error term, (ε).

- There are several standard models for auto correlated regression errors, with the most common being the first order auto regressive process, AR(1):

$$\varepsilon_t = \Phi\varepsilon_{t-1} + \nu_t$$

- The ν_t represents random shocks and are assumed to be Gaussian white noise and Φ represents the correlation component.

- The AR(2) model:

$$\varepsilon_t = \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + \nu_t$$

Regression Models on Time Series Data

- In contrast, in the first order moving-average process, MA(1), the current error depends upon the random shock from the current and previous periods (rather than upon the previous regression error).

$$\text{MA(1):}\ \varepsilon_t = \Psi \nu_{t-1} + \nu_t$$

$$\text{MA(2):}\ \varepsilon_t = \Psi_1 \nu_{t-1} + \Psi_2 \nu_{t-2} + \nu_t$$

- Finally, AR and MA terms can be combined in ARMA(p,q) processes.

- The ARMA(1,1) model:

$$\varepsilon_t = \Phi \varepsilon_{t-1} + \nu_t + \Psi \nu_{t-1}$$

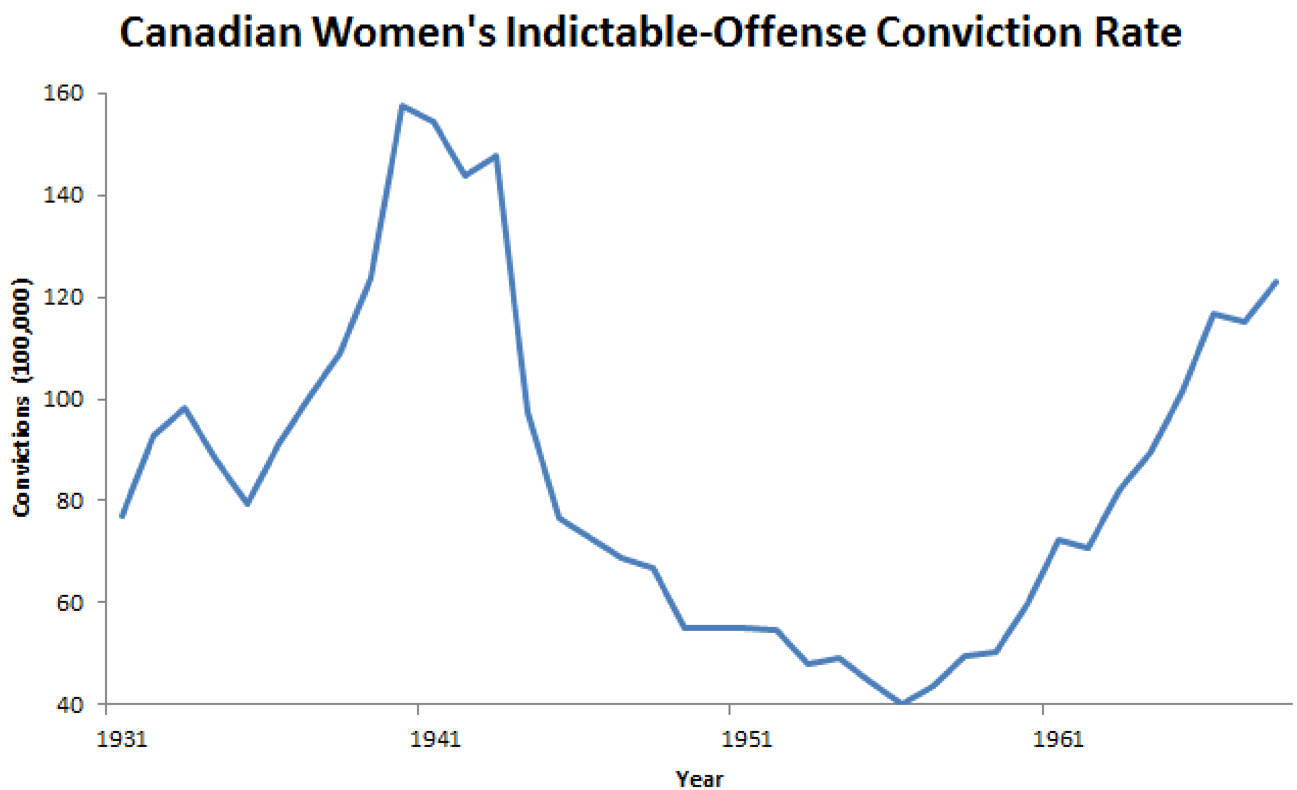
Regression Models on Time Series Data

- Now that we have a basic understanding of the GLS approach, lets work through an example.
- Here is the Hartnagel dataset from 1931 – 1968 of different macro-economic attributes of women in Canada.
- We will estimate the coefficient, fconvict, on tfr, partic, degrees, and mconvict. The rationale for including mconvict is to control for omitted variables that affect the crime rate in general.

Variable	Description
year	1931 - 1968
tfr	the total fertility rate, births per 1000 women
partic	women's labor force participation rate, per 1000
degrees	women's post-secondary degree rate, per 10,000
fconvict	women's indictable-offense conviction rate, per 10,000
ftheft	women's theft conviction rate, per 100,000
mconvict	men's indictable-offense conviction rate, per 100,000
mtheft	men's theft conviction rate, per 100,000

year	tfr	partic	degrees	fconvict	ftheft	mconvict	mtheft
1935	2755	238	13.2	79.4	20.4	765.7	247.1
1936	2696	240	13.2	91.0	22.1	816.5	254.9
1937	2646	241	12.2	100.4	22.4	821.8	272.4
1938	2701	242	12.6	108.9	21.8	956.8	285.8
1939	2654	244	12.3	123.6	21.1	1035.7	292.2
1940	2766	245	12.0	157.3	21.4	951.6	256.0
1941	2832	246	11.7	154.3	25.3	850.9	205.8
etc...	etc...	etc...	etc...	etc...	etc...	etc...	etc...

Understanding the Data



- The dataset does not appear to have any seasonality components, however, we could decompose the time series and seasonally adjust the dataset to ensure that we are only focusing on the trend and random components.

Multiple Linear Regression (OLS)

Multiple Linear Regression Output

lm(formula = fconvict ~ tfr + partic + degrees + mconvict)

Residuals:

Min	1Q	Median	3Q	Max
-42.964	-9.204	-3.566	6.149	48.385

Coefficients	**Estimate**	**Std. Error**	**t- value**	**Pr(>|t|)**	
(Intercept)	127.64	59.95704	2.129	0.0408	*
tfr	-0.04657	0.008033	-5.797	1.75E-06	***
partic	0.253416	0.115132	2.201	0.0348	*
degrees	-0.21205	0.211454	-1.003	0.3232	
mconvict	0.059105	0.045145	1.309	0.1995	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

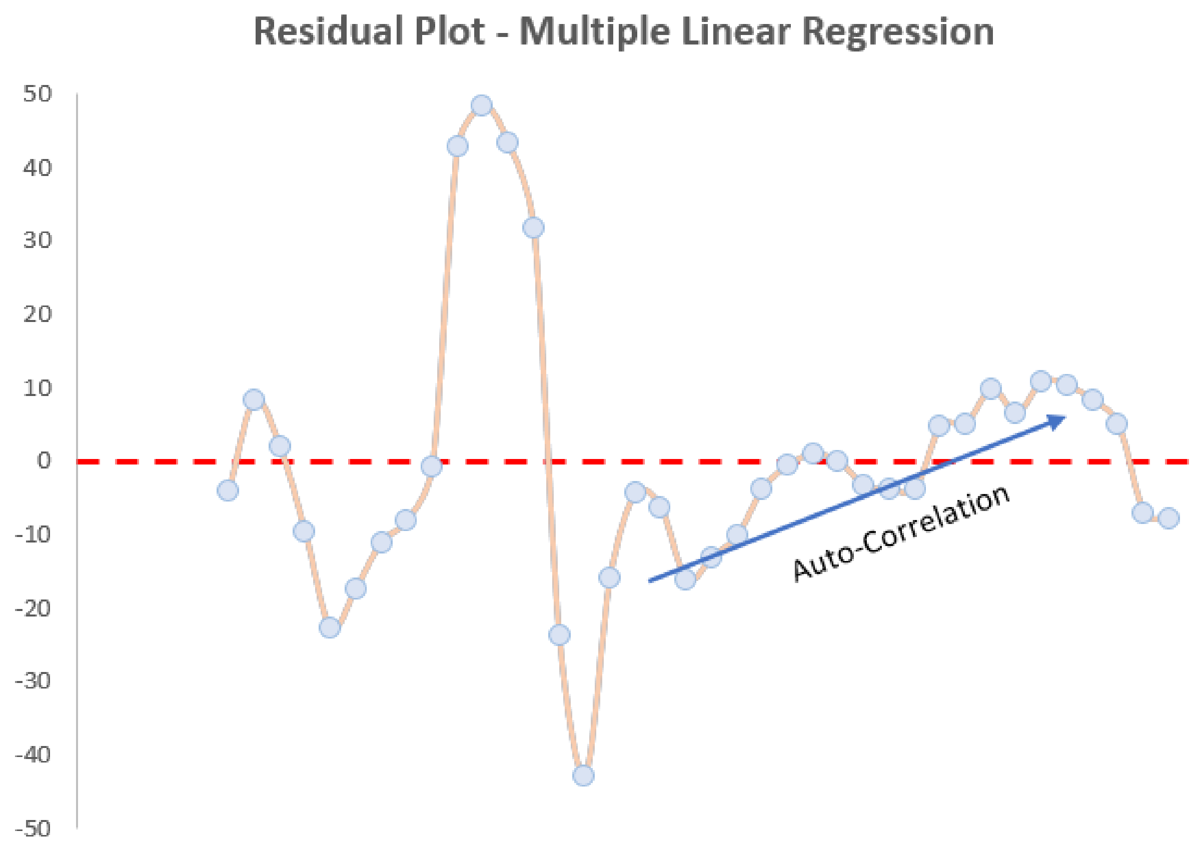
Residual standard error: 19.19 on 33 degrees of freedom
Multiple R-squared: 0.6948, Adjusted R-squared: 0.6578
F-statistic: 18.78 on 4 and 33 DF, p-value: 3.905e-08

Women's crime rate appears to decline with fertility and increase with labor-force participation; the other 2 coefficients are statistically insignificant.

The $R^2 = 0.6948$ which is fairly strong and indicates that the model is capable of predicting a large portion of observations.

Multiple Linear Regression (OLS)

- However, the residual plot suggests substantial autocorrelation.
- The residuals are not uniform and randomly distributed across the 0 line. This can be observed by the large peak and trough.
- Additionally, there seems to be a direction of the residuals which are moving upward indicating that serial correlation is a problem and that this OLS violates statistical assumptions.

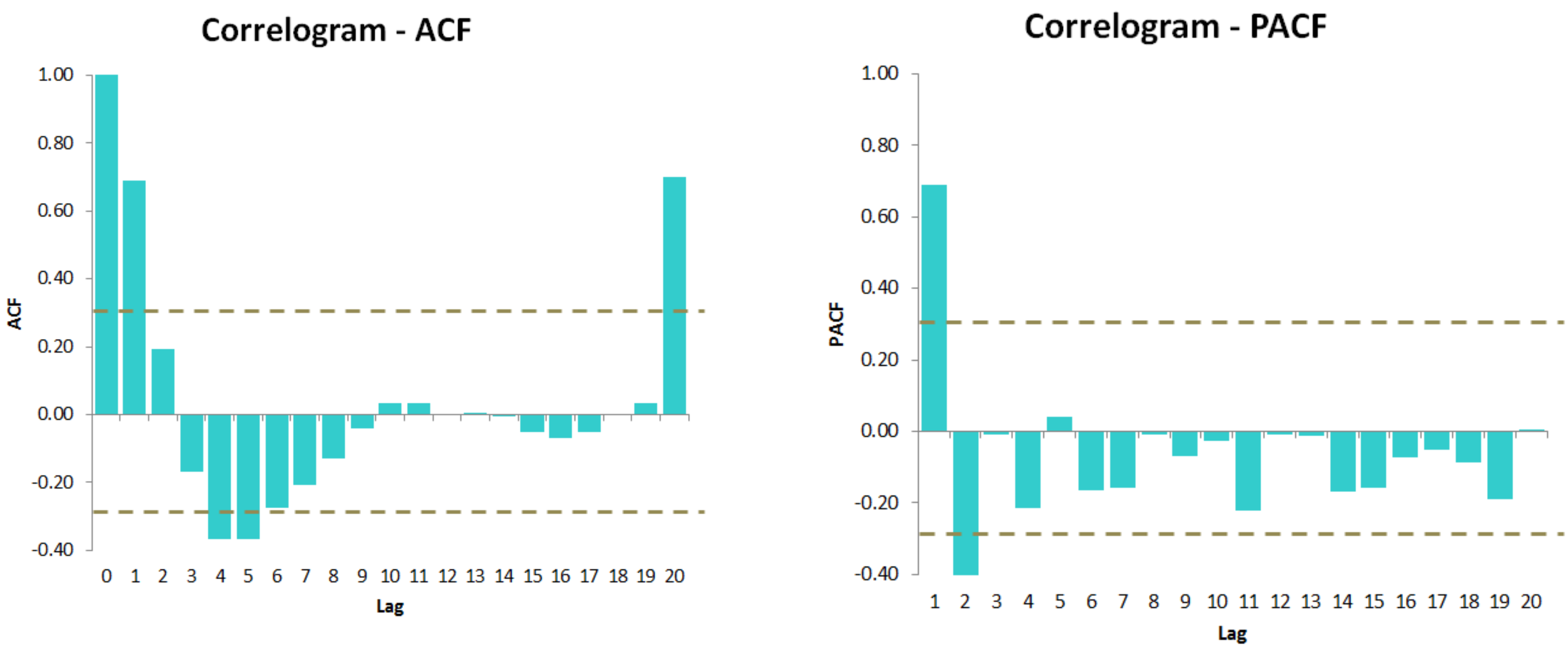


Determining the model approach

Model	ACF	PACF
White Noise	All Zeros	All Zeros
AR(p)	Exponential Decay	p significant lags before dropping to zero
MA(q)	q significant lags before dropping to zero	Exponential Decay
ARMA(p,q)	Decay after q^{TH} lag	Decay after p^{TH} lag

- Like the examples before, we need to create the ACF and PACF correlograms and examine them in detail.
- However, we need to look carefully at the patterns of each one to determine which model approach, AR(p), MA(q), or ARMA(p,q), makes the most sense for the GLM approach.
- The reference cart can be used as a general guideline for this evaluation.

Determining the model approach



- The ACF diagram shows an exponential decay to lag=2 (sinusoidal pattern) and the PACF shows significant lags before dropping to 0 at lag=2.

- This indicates that the AR(p) approach should be used and the PACF showed a drop to 0 at lag = 2. Therefore, we will use the AR(2) model for our analysis. The PACF also suggests that the $\Phi_1 > 0$ and $\Phi_2 < 0$

Durbin Watson Test

- We follow up by computing the Durbin Watson statistics for the OLS regression. This function produces bootstrapped p-values for the Durbin Watson statistics.
- Three of the first five Durbin-Watson statistics are significant, including the first.
- This indicates that autocorrelation is present in the dataset and will be improved upon through non OLS modeling methods.

Durbin-Watson Test (lag=5)

lag	Autocorrelation	D-W Statistic	p-value
1	0.688345	0.6168636	0.000
2	0.192267	1.5993563	0.154
3	-0.168570	2.3187448	0.316
4	-0.365278	2.6990538	0.006
5	-0.367324	2.6521103	0.016

GLS Model Fit by Maximum Likelihood

Generalized Least Squares Fit by Maximum Likelihood

fconvict ~ tfr + partic + degrees + mconvict

Standardized Residuals:

Min	Q1	Med	Q3	Max
-2.49915	-0.3717	-0.14945	0.33724	2.90947

Data:

AIC	BIC	logLik
305.41	318.52	-144.7

Coefficients	**Estimate**	**Std. Error**	**t- value**	**Pr(>|t|)**
(Intercept)	83.34	59.471	1.4014	0.1704
tfr	-0.04	0.009	-4.3086	1.00E-04
partic	0.288	0.112	2.5677	0.015
degrees	-0.21	0.207	-1.0158	0.3171
mconvict	0.076	0.035	2.1619	0.038

Residual standard error: 17.702

Degrees of freedom: 38 total; 33 residual

Correlation Structure: ARMA(2,0)

Parameter estimate(s):

Phi (Φ_1)	Phi (Φ_2)
1.06835	-0.551

Correlation:	(Intr)	tfr	partic	degres
tfr	-0.773			
partic	-0.57	0.176		
degrees	0.093	0.033	-0.476	
mconvict	-0.689	0.365	0.047	0.082

Observations:

- The Φ_1 = 1.06835 and Φ_2 = -0.551 which was what we expected. (Φ_1 >0 and Φ_2 <0)
- The mconvict variable was not statistically significant in the OLS model but is significant in the GLS model.
- The Aikake Information Criteria (AIC) is 305.41. Lower values indicate stronger models.

GLS Model Fit by Maximum Likelihood

ANOVA				
	df	AIC	BIC	logLik
GLS with AR(0)	6	339.00	348.83	-163.50
GLS with AR(1)	7	312.42	323.89	-149.21
GLS with AR(2)	8	305.41	318.52	-144.71
GLS with AR(3)	9	307.40	322.13	-144.70

- To verify that the AR(2) model was the preferred model, we re-ran the models for AR(0), AR(1), and AR(3) and performed ANOVA.
- The AIC is the lowest for the AR(2) model, 305.41, but followed very closely by the AR(3) model, 307.40. Following the principle of parsimony the AR(2) model is less complex and therefore should be used as our final model.

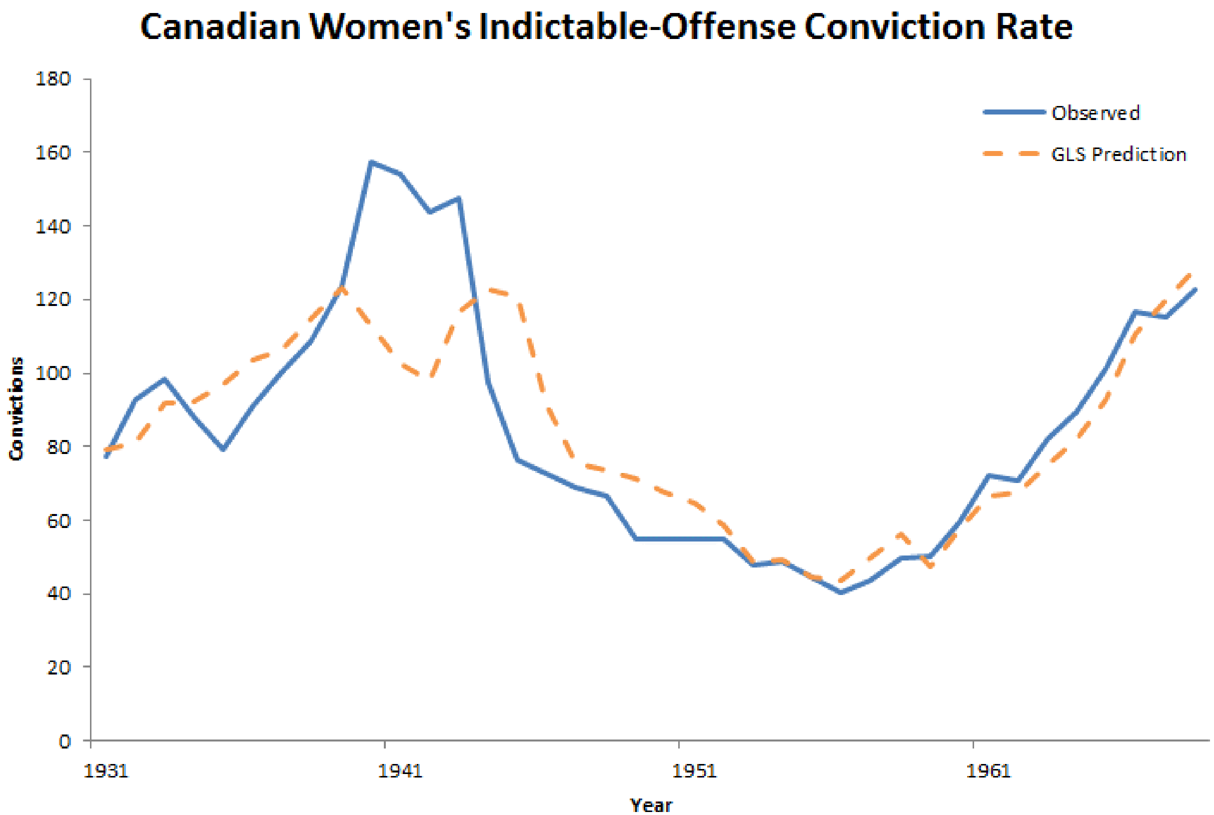
Final GLS Model

Generalized Least Squares Fit by Maximum Likelihood							
fconvict ~ tfr + partic + mconvict + 0							
Standardized Residuals:					**Data:**		
Min	Q1	Med	Q3	Max	AIC	BIC	logLik
-1.96814	-0.40519	-0.21775	0.2036	3.0624	304.59	314.42	-146.3
Coefficients	**Estimate**	**Std. Error**	**t- value**	**Pr(>|t|)**	**Correlation Structure: ARMA(2,0)**		
tfr	-0.02366	0.006897	-3.4309	0.0016	*Parameter estimate(s):*		
partic	0.304232	0.083568	3.6406	0.0009	Phi (Φ_1)	Phi (Φ_2)	
mconvict	0.100061	0.026751	3.7405	0.0007	1.15437	-0.53	
Residual standard error: 20.028					**Correlation:**	*tfr*	*partic*
Degrees of freedom: 38 total; 35 residual					*partic*	-0.528	
					mconvict	-0.478	-0.468

Additional Changes:

- Removed the Intercept and the Degree variables.
- The Aikake Information Criteria (AIC) is 304.59 compared to 305.41.

Final GLS Model

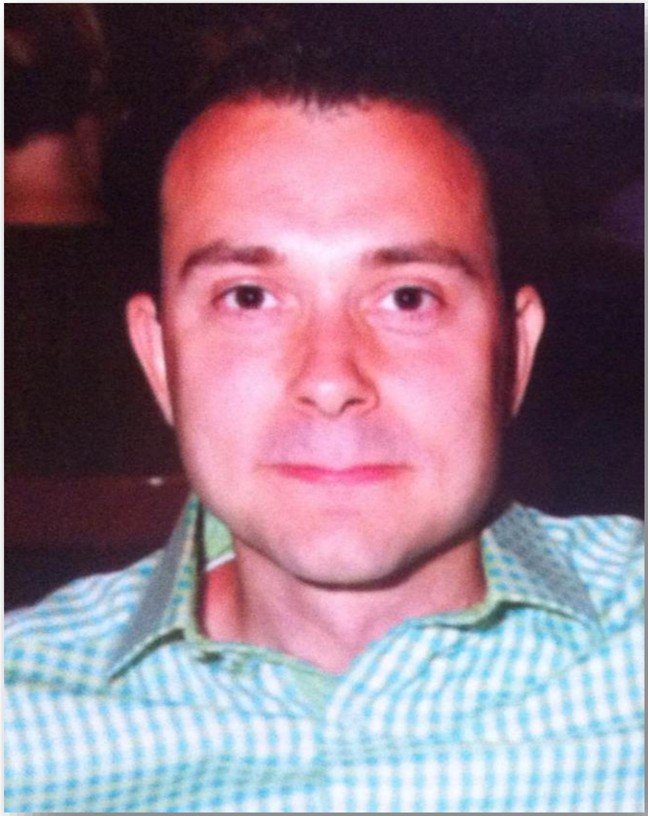


$$\text{Conviction Rate} = -0.024 * \text{Fertility} + 0.304 * \text{Labor} + 0.10 \text{ Male Convict} + \varepsilon$$

$$\varepsilon_t = 1.154 * \varepsilon_{t-1} - 0.53 * \varepsilon_{t-2} + \nu_t$$

About Me

- Reside in Wayne, Illinois
- Active Semi-Professional Classical Musician (Bassoon).
- Married my wife on 10/10/10 and been together for 10 years.
- Pet Yorkshire Terrier / Toy Poodle named Brunzie.
- Pet Maine Coons' named Maximus Power and Nemesis Gul du Cat.
- Enjoy Cooking, Hiking, Cycling, Kayaking, and Astronomy.
- Self proclaimed Data Nerd and Technology Lover.



Fine