STU33010: Mid Term Assignment 2020

Michael Haskins

14/11/2020

1 Introduction

The time series dataset (file pop.csv) corresponds to several time series of population size for the UK till year 2018. In this report I will find the best Holt-Winters algorithms to predict the populations of England, Scotland, Northern Ireland and Wales (4 time series) for the years 2019 and 2020.

This report will consist of 5 sections;

1. England
2. Scotland
3. Northern Ireland
4. Wales
5. Warnings about the accuracy of these forecasts

Note: Sections 1-4 will contain the following:

- Basic analysis of this time series
- The appropriate Holt-Winters algorithm for this time series and justification for its selection.
- Forecasts for 2019 and 2020.

1.1 The time series

```
require("fma")
require("knitr")
myts = read.csv("pop.csv")

kable(myts)
```

Year	Scotland.population	GreatBritain.population	England.population	UnitedKingdom.population	England.and.Wales.population	Northern.Ireland.population	Wales.population
1971	5235600	54387600	46411700	55928000	49152000	1540400	2740300
1972	5230600	54557700	46571900	56096700	49327100	1539000	2755200
1973	5233900	54692900	46686200	56222900	49459000	1530000	2772800
1974	5240800	54708700	46682700	56235600	49467900	1526900	2785200

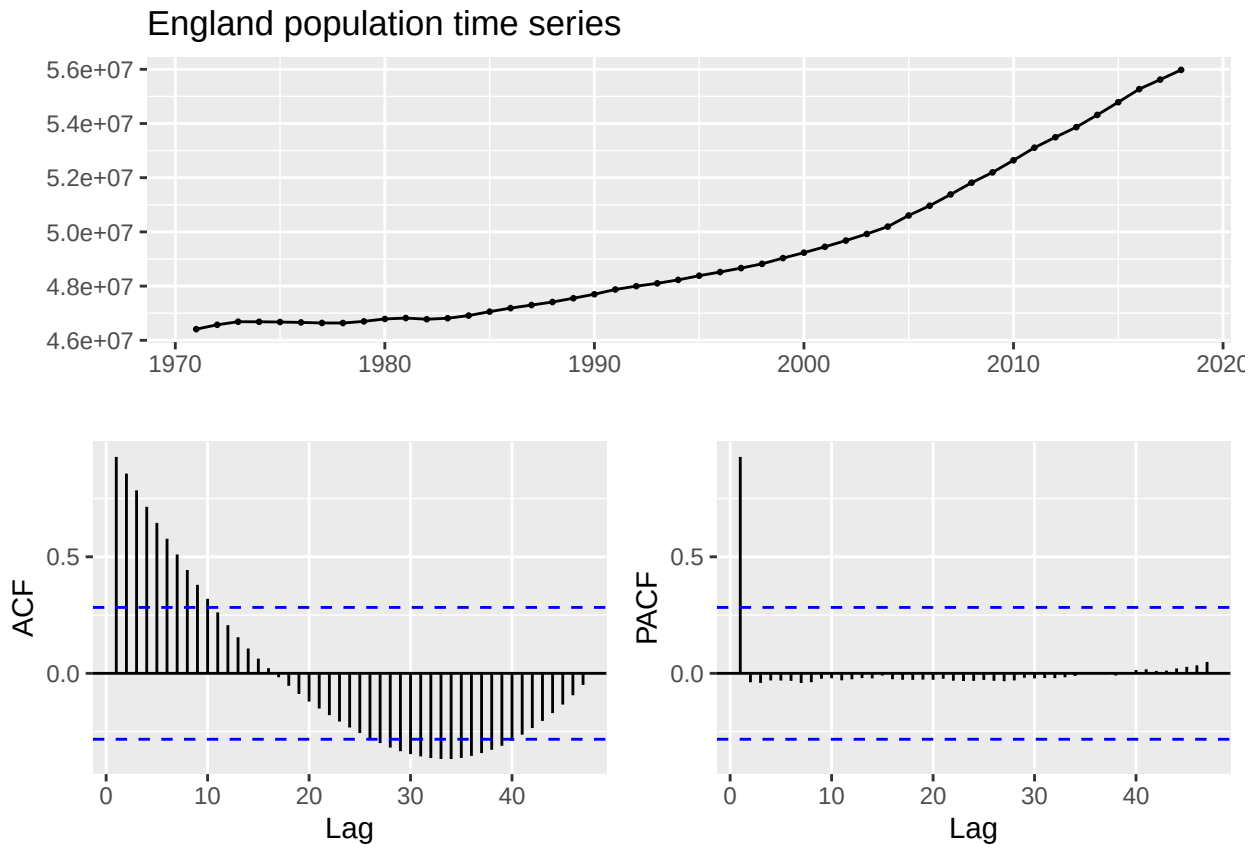
Year	Scotland.population	Great Britain.population	England.population	United Kingdom.population	England and Wales.population	Northern Ireland.population	Wales population
1975	5232400	54702200	46674400	56225700	49469800	1523500	2795400
1976	5233400	54692600	46659900	56216100	49459200	1523500	2799300
1977	5226200	54666600	46639800	56189900	49440400	1523300	2800600
1978	5212300	54654800	46638200	56178000	49442500	1523200	2804300
1979	5203600	54711800	46698100	56240100	49508200	1528300	2810100
1980	5193900	54796900	46787200	56329700	49603000	1532800	2815800
1981	5180200	54814500	46820800	56357500	49634300	1543000	2813500
1982	5164500	54746200	46777300	56290700	49581600	1544500	2804300
1983	5148100	54765100	46813700	56315700	49617000	1550600	2803300
1984	5138900	54852000	46912400	56409300	49713100	1557300	2800700
1985	5127900	54988600	47057400	56554000	49860700	1565400	2803400
1986	5111800	55110300	47187600	56683800	49998600	1573500	2810900
1987	5099000	55222000	47300400	56804000	50123000	1582000	2822600
1988	5077400	55331000	47412300	56916400	50253600	1585400	2841200
1989	5078200	55486000	47552700	57076500	50407800	1590400	2855200
1990	5081300	55641900	47699100	57237500	50560600	1595600	2861500
1991	5083300	55831400	47875000	57438700	50748000	1607300	2873000
1992	5085600	55961300	47998000	57584500	50875600	1623300	2877700
1993	5092500	56078300	48102300	57713900	50985900	1635600	2883600
1994	5102200	56218400	48228800	57862100	51116200	1643700	2887400
1995	5103700	56375700	48383500	58024800	51272000	1649100	2888500
1996	5092200	56502600	48519100	58164400	51410400	1661800	2891300
1997	5083300	56643000	48664800	58314200	51559600	1671300	2894900
1998	5077100	56797200	48820600	58474900	51720100	1677800	2899500
1999	5072000	57005400	49032900	58684400	51933500	1679000	2900600
2000	5062900	57203100	49233300	58886100	52140200	1682900	2906900
2001	5064200	57424200	49449700	59113000	52360000	1688800	2910200
2002	5066000	57668100	49679300	59365700	52602100	1697500	2922900
2003	5068500	57931700	49925500	59636700	52863200	1704900	2937700
2004	5084300	58236300	50194600	59950400	53152000	1714000	2957400
2005	5110200	58685500	50606000	60413300	53575300	1727700	2969300
2006	5133100	59084000	50965200	60827100	53950900	1743100	2985700
2007	5170000	59557400	51381100	61319100	54387400	1761700	3006300
2008	5202900	60044600	51815900	61823800	54841700	1779200	3025900
2009	5231900	60467200	52196400	62260500	55235300	1793300	3038900
2010	5262200	60954600	52642500	62759500	55692400	1804800	3050000
2011	5299900	61470800	53107200	63285100	56170900	1814300	3063800
2012	5313600	61881400	53493700	63705000	56567800	1823600	3074100
2013	5327700	62275900	53865800	64105700	56948200	1829700	3082400
2014	5347600	62756300	54316600	64596800	57408700	1840500	3092000
2015	5373000	63258400	54786300	65110000	57885400	1851600	3099100
2016	5404700	63785900	55268100	65648100	58381200	1862100	3113200
2017	5424800	64169400	55619400	66040200	58744600	1870800	3125200
2018	5438100	64553900	55977200	66435600	59115800	1881600	3138600

2 England

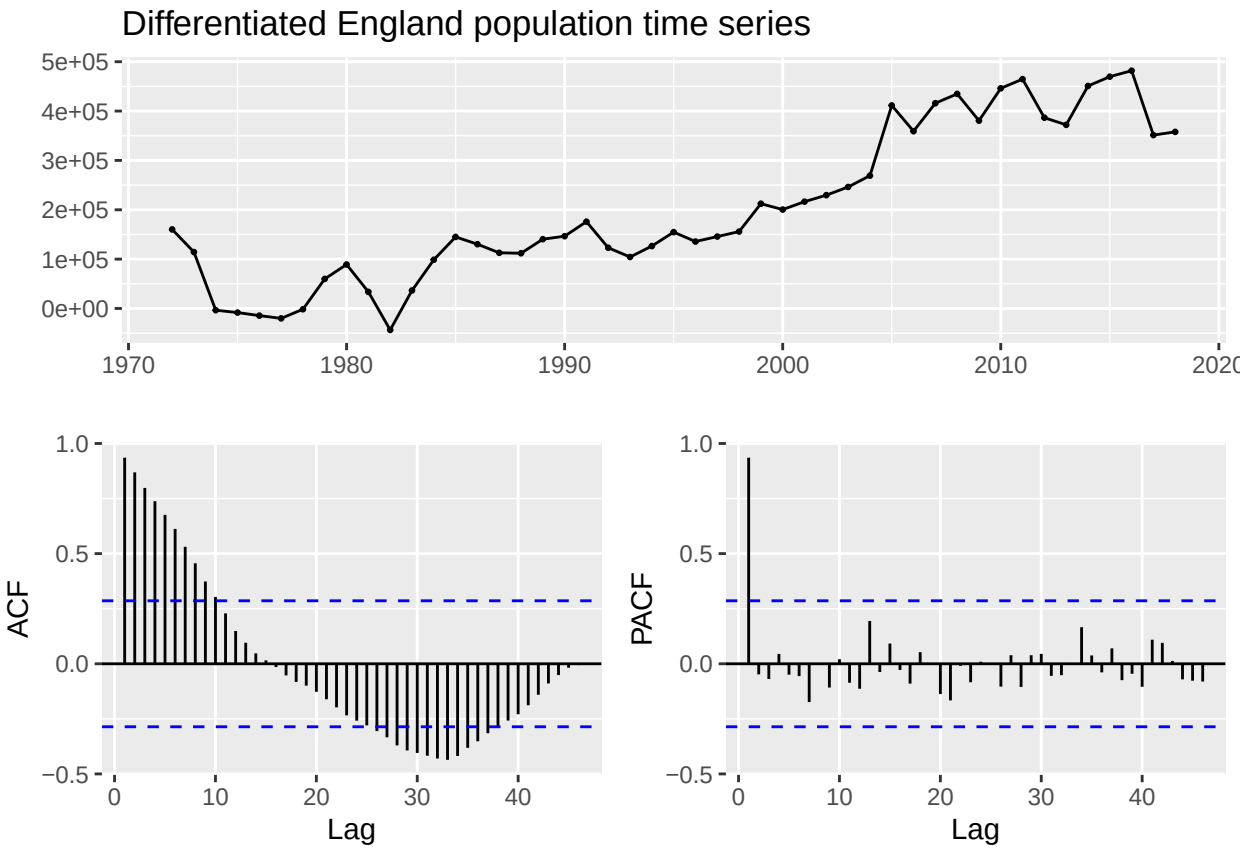
2.1 Visualization of the time series

```
mytsEngland<-ts(myts$England.population, start=1971,end=2018,frequency=1)

ggtsdisplay(mytsEngland, lag.max = 47, main = "England population time series")
```




```
##Diff to remove trend, as makes it easier to spot seasonal patterns
ggtsdisplay( diff(mytsEngland),lag.max = 47, main = "Differentiated England population time series" )
```



2.2 Analysis of time series

It can be determined that there is a trend due to the upward slope of the time series and that there is no seasonal pattern visible due to the straight line of the time series and also due to the fact there is no pattern visible in the PACF graph.

2.3 Appropriate Holt-Winters algorithm

Since there is a trend and no seasonal pattern the Holt-Winters algorithms that are appropriate to use are the SES(Which can handle a small trend) or DES.

It is important to choose the Holt-Winters algorithm that has the lowest Sum of Square Error (SSE), as this will yield the most accurate forecasts.

```
##Below Code will yield the SSE of SES Holt-Winters Algorihtm
SESSSE = HoltWinters(mytsEngland, beta = FALSE, gamma = FALSE)$SSE

##Below Code will yield the SSE of DES Holt-Winters Algorihtm
DESSSE = HoltWinters(mytsEngland, gamma = FALSE)$SSE

if( SESSSE > DESSSE){
  print("DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES")
} else {
  print("SES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than DES")
}
```

```
[1] "DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES"
```

2.4 Forecasts for 2019 and 2020

As was determined above DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES, meaning the forecasts generated using DES will be more accurate than those of SES.

```
##Prediction for 2019 using DES
predict(HoltWinters(mytsEngland, gamma = FALSE), n.ahead=1)[1]
```

```
[1] 56335000
```

```
##Prediction for 2020 using DES
predict(HoltWinters(mytsEngland, gamma = FALSE), n.ahead=2)[2]
```

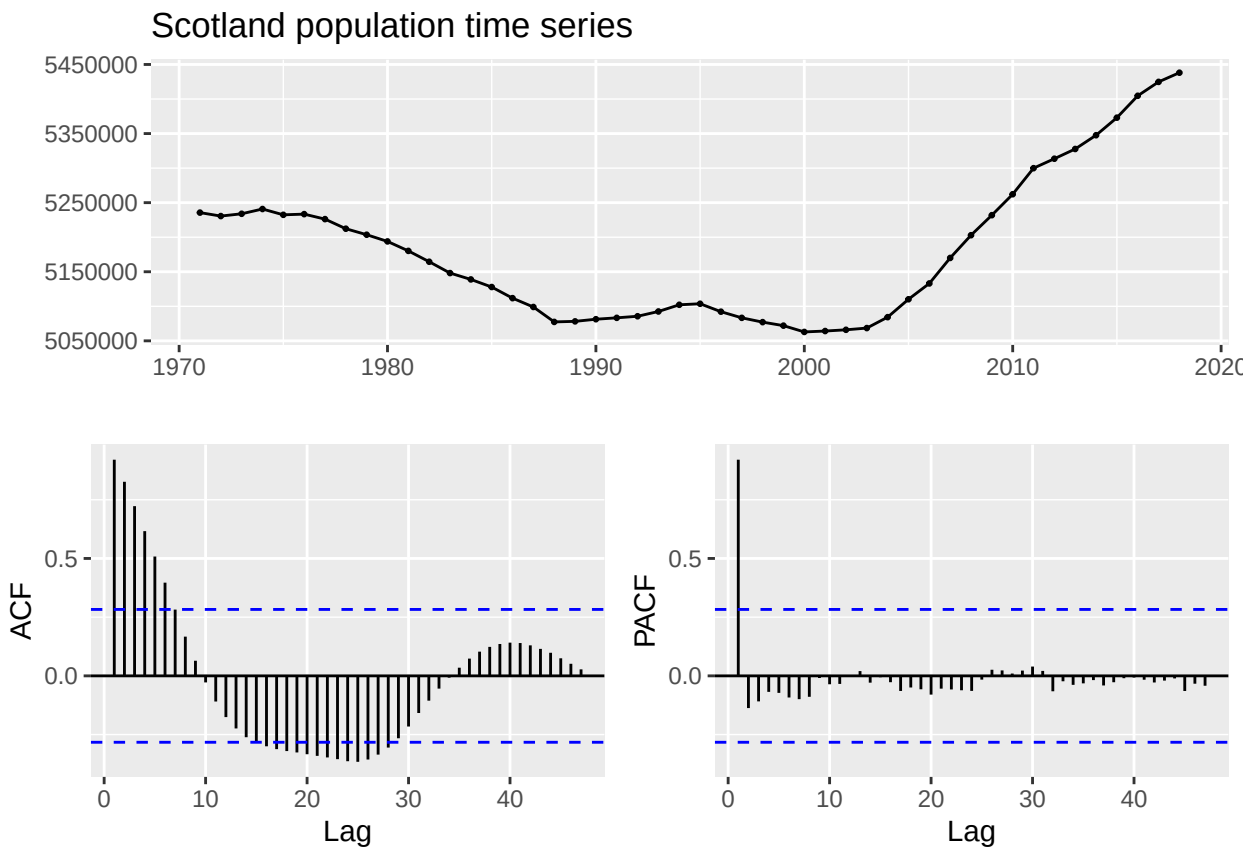
```
[1] 56692800
```

3 Scotland

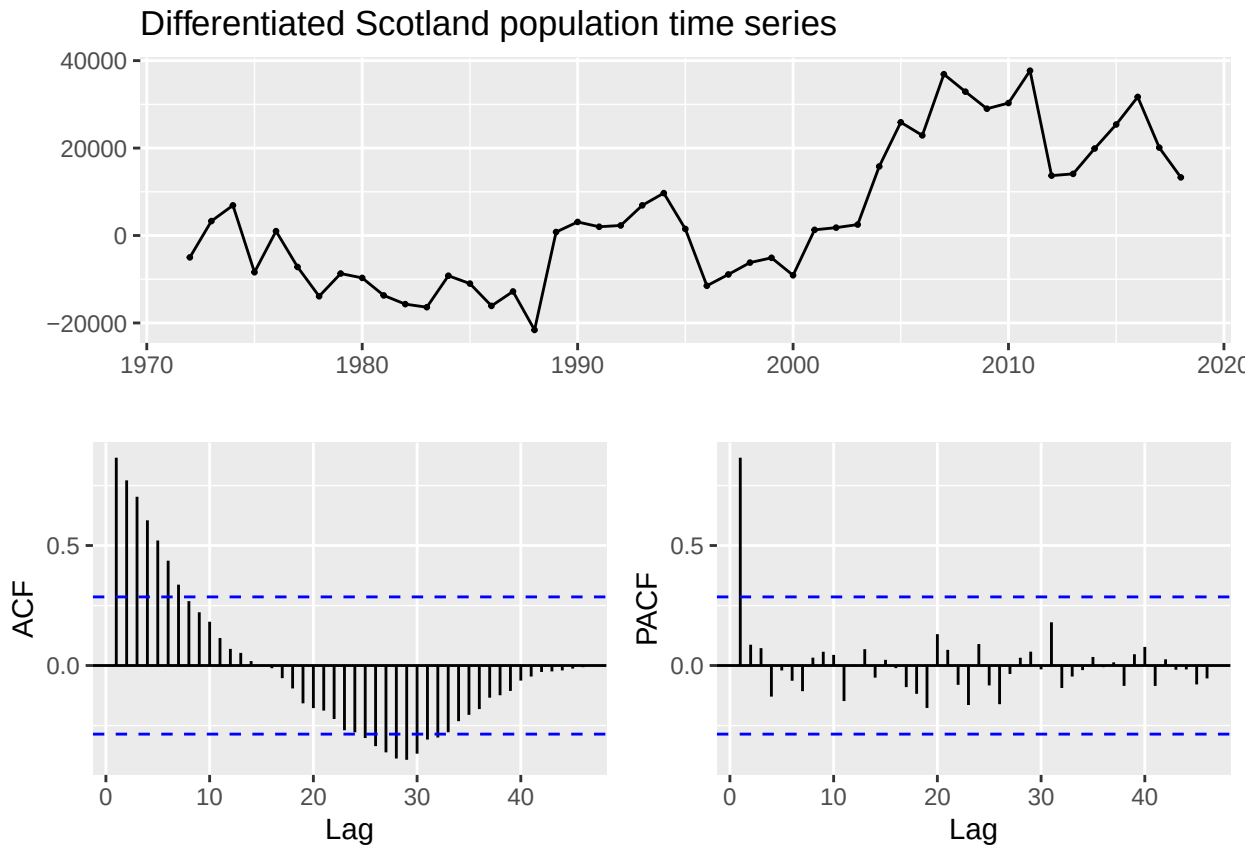
3.1 Visualization of the time series

```
mytsScotland<-ts(myts$Scotland.population, start=1971,end=2018,frequency=1)

ggtsdisplay(mytsScotland,lag.max = 47, main = "Scotland population time series")
```




```
##Diff to remove trend, as makes it easier to spot seasonal patterns
ggtsdisplay( diff(mytsScotland),lag.max = 47, main = "Differentiated Scotland population time series" )
```



3.2 Analysis of time series

There is a trend in this time series. We can see that the population was trending generally downward until 2003 and ever since its been on a strong upward trend. This is reflected in the ACF plot where we can see it is initially sloping downward and then when the lag is around 16 (Reflecting the trend turnaround in circa 2003) the trend reverses. Since the Holt-Winters algorithms place decreasing emphasis on older results more emphasis will be placed on the data that reflects the upward trend as this data is more recent so is more relevant.

It can also be determined that there is no seasonality in the time series as even when the trend is removed, by differentiating, it is still impossible to see anything resembling a seasonal pattern in the PACF graph.

3.3 Appropriate Holt-Winters Algorithm

Since there is a trend and no seasonal pattern the Holt-Winters algorithms that are appropriate to use are the SES(Which can handle a small trend) or DES.

It is important to choose the Holt-Winters algorithm that has the lowest Sum of Square Error (SSE), as this will yield the most accurate forecasts.

```
##Below Code will yield the SSE of SES Holt-Winters Algorihtm
SESSSE = HoltWinters(mytsScotland, beta = FALSE, gamma = FALSE)$SSE

##Below Code will yield the SSE of DES Holt-Winters Algorihtm
DESSSE = HoltWinters(mytsScotland, gamma = FALSE)$SSE
```

```
if( SESSSE > DESSSE){
  print("DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES")
} else {
  print("SES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than DES")
}
```

```
[1] "DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES"
```

3.4 Forecasts for 2019 and 2020

As was determined above DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES.

```
##Prediction for 2019 using DES
predict(HoltWinters(mytsScotland, gamma = FALSE), n.ahead=1)[1]
```

```
[1] 5453202
```

```
##Prediction for 2020 using DES
predict(HoltWinters(mytsScotland, gamma = FALSE), n.ahead=2)[2]
```

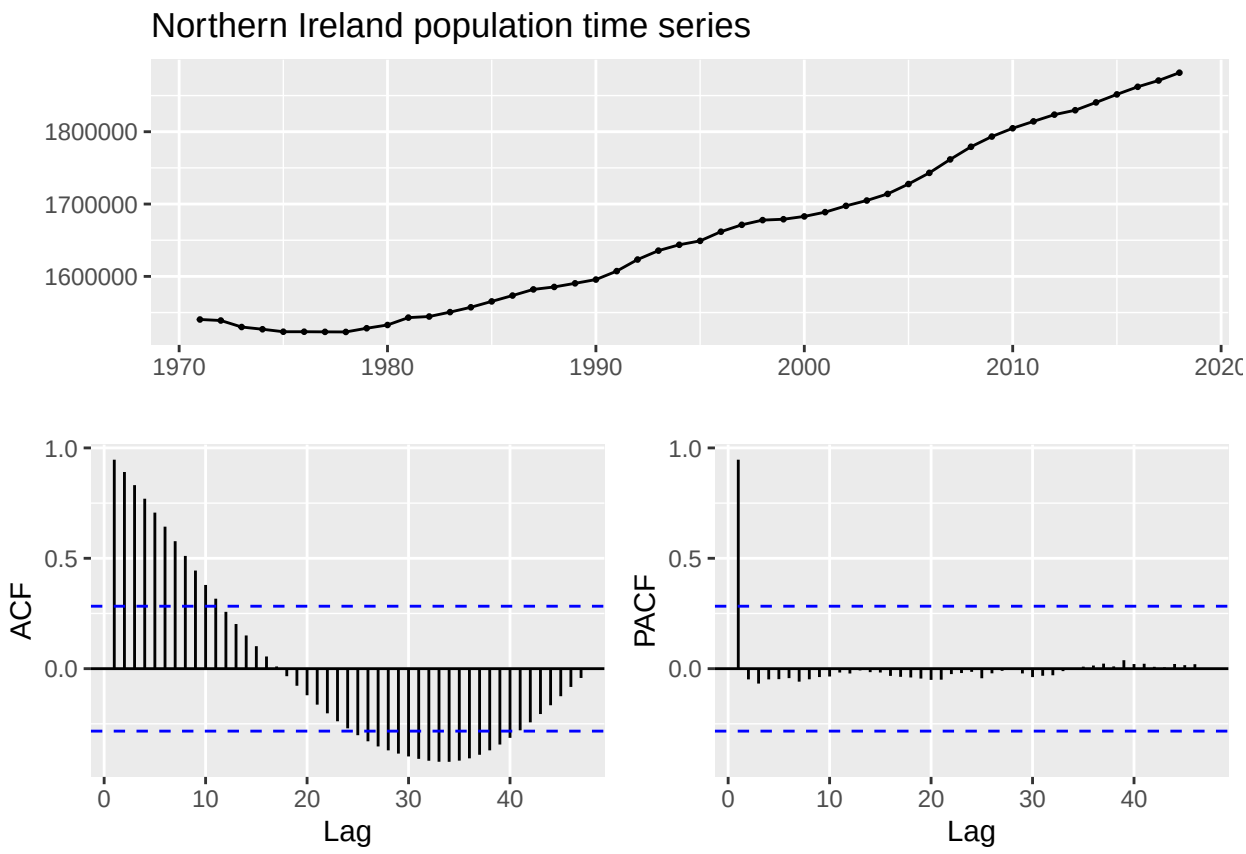
```
[1] 5468304
```

4 Northern Ireland

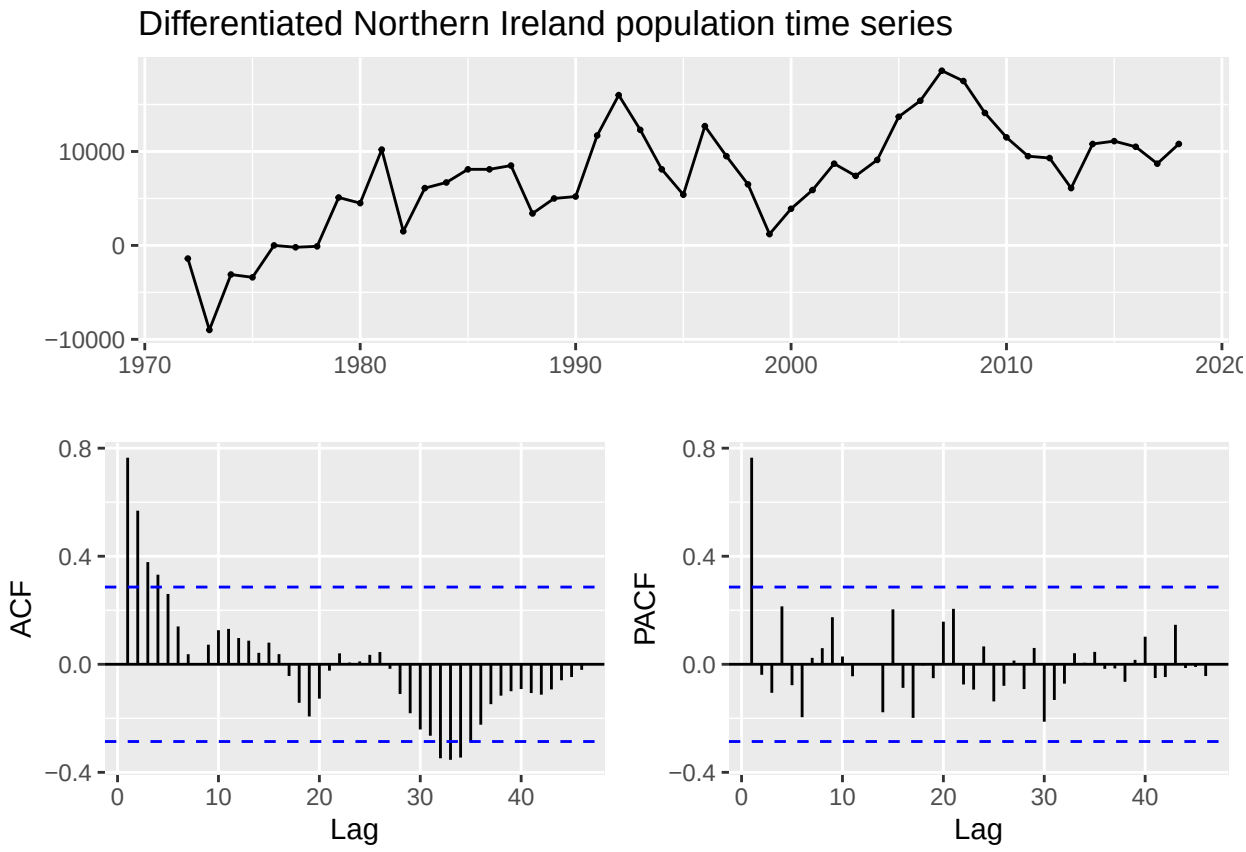
4.1 Visualization of the time series

```
mytsNorthernIreland<-ts(myts$Northern.Ireland.population, start=1971,end=2018,frequency=1)

ggtsdisplay(mytsNorthernIreland, main = "Northern Ireland population time series", lag.max = 47)
```




```
##Diff to remove trend, as makes it easier to spot seasonal patterns
ggtsdisplay( diff(mytsNorthernIreland), main = "Differentiated Northern Ireland population time series"
```



4.2 Analysis of time series

It can be determined that there is a consistent upward trend in the time series due to the slope of the plotted line of the time series being generally positive.

It can also be determined that there is no seasonality in the time series as even when the trend is removed, by differentiating, it is still impossible to see anything resembling a seasonal pattern in the PACF graph.

4.3 Appropriate Holt-Winters Algorithm

Since there is a trend and no seasonal pattern in the time series the Holt-Winters algorithms that are appropriate to use are the SES(Which can handle a small trend) or DES.

It is important to choose the Holt-Winters algorithm that has the lowest Sum of Square Error (SSE), as this will yield the most accurate forecasts.

```
##Below Code will yield the SSE of SES Holt-Winters Algorihtm
SESSSE = HoltWinters(mytsNorthernIreland, beta = FALSE, gamma = FALSE)$SSE

##Below Code will yield the SSE of DES Holt-Winters Algorihtm
DESSSE = HoltWinters(mytsNorthernIreland, gamma = FALSE)$SSE

if( SESSSE > DESSSE){
  print("DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES.")
} else {
```

```
  print("SES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than DES.")
}
```

```
[1] "DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES."
```

4.4 Forecasts for 2019 and 2020

As was determined above DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES.

```
##Prediction for 2019 using DES
predict(HoltWinters(mytsNorthernIreland, gamma = FALSE), n.ahead=1)[1]
```

```
[1] 1892030
```

```
##Prediction for 2020 using DES
predict(HoltWinters(mytsNorthernIreland, gamma = FALSE), n.ahead=2)[2]
```

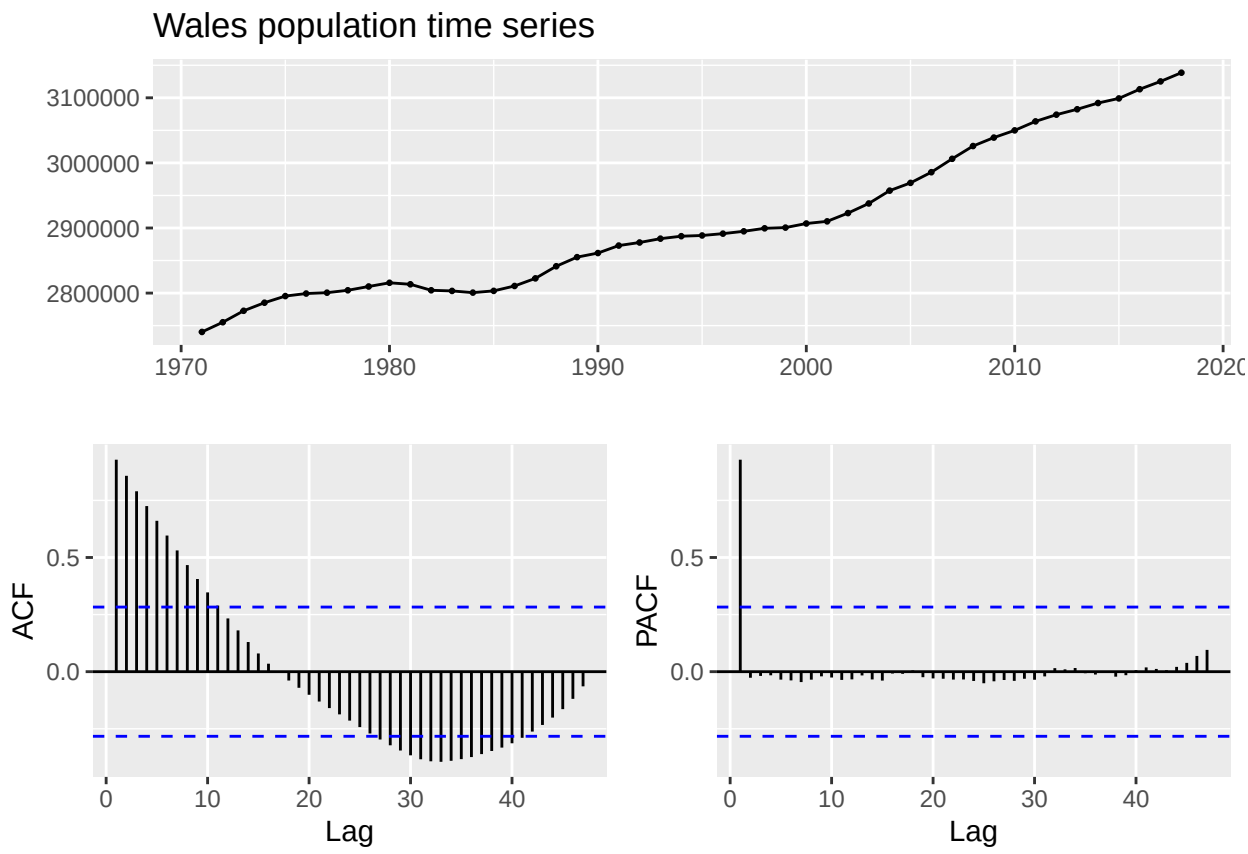
```
[1] 1902460
```

5 Wales

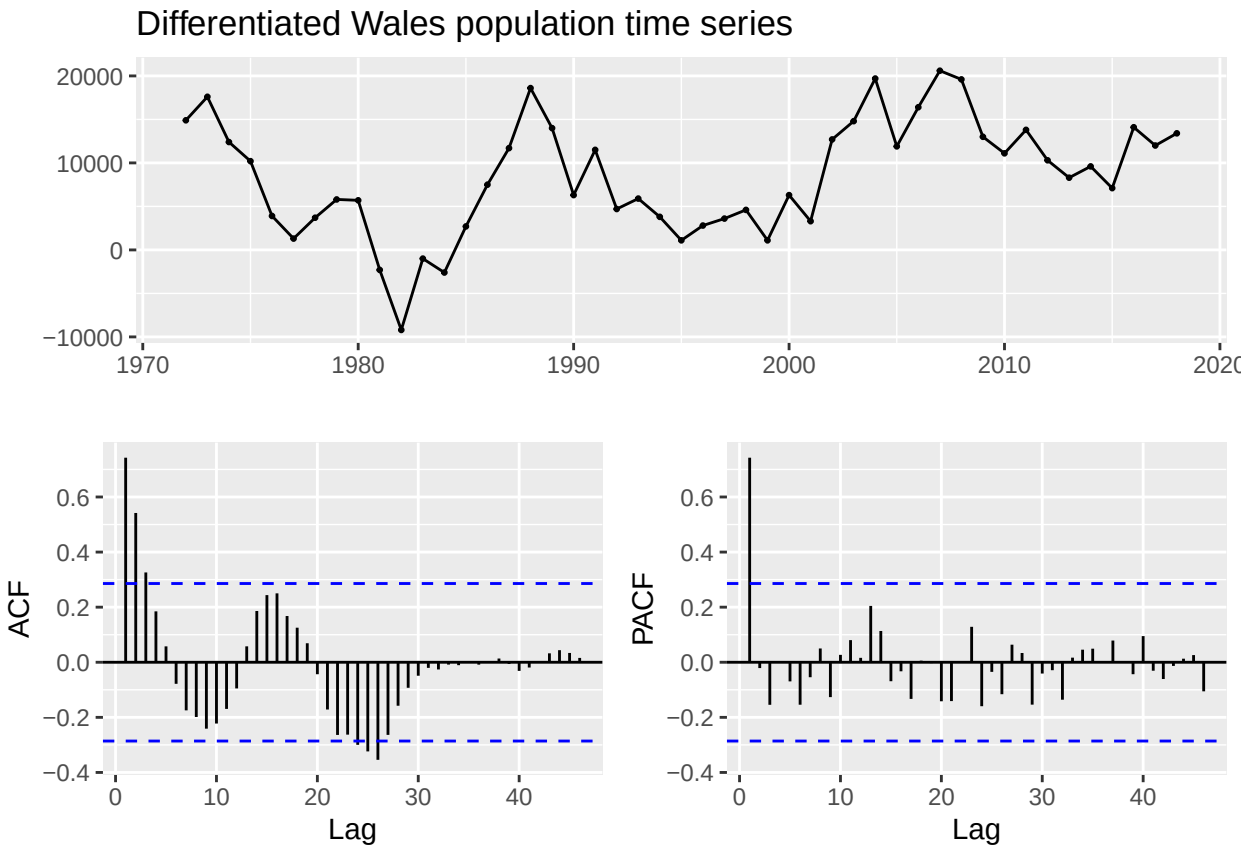
5.1 Visualization of the time series

```
mytsWales<-ts(myts$Wales.population, start=1971,end=2018,frequency=1)

ggtsdisplay(mytsWales, lag.max = 47, main = "Wales population time series")
```




```
##Diff to remove trend, as makes it easier to spot seasonal patterns
ggtsdisplay( diff(mytsWales),lag.max = 47, main = "Differentiated Wales population time series" )
```



5.2 Analysis of time series

It can be determined that there is a trend in the time series due to the upward slope of the plotted line of the time series and the ACF graph having a generally downward sloping line.

It can also be determined that there is no seasonality in the time series as even when the trend is removed, by differentiating, it is still impossible to see anything resembling a seasonal pattern in the PACF graph.

5.3 Appropriate Holt-Winters Algorithm

Since there is a trend and no seasonal pattern in the time series the Holt-Winters algorithms that are appropriate to use are the SES(Which can handle a small trend) or DES.

It is important to choose the Holt-Winters algorithm that has the lowest Sum of Square Error (SSE), as this will yield the most accurate forecasts.

```
##Below Code will yield the SSE of SES Holt-Winters Algorihtm
SESSSE = HoltWinters(mytsWales, beta = FALSE, gamma = FALSE)$SSE

##Below Code will yield the SSE of DES Holt-Winters Algorihtm
DESSSE = HoltWinters(mytsWales, gamma = FALSE)$SSE

if( SESSSE > DESSSE){
  print("DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES.")
} else {
```

```
  print("SES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than DES.")
}
```

```
[1] "DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES."
```

5.4 Forecasts for 2019 and 2020

As was determined above DES is the best Holt-Winters Algorithm for this time series, as it has lower SSE than SES.

```
##Prediction for 2019 using DES
predict(HoltWinters(mytsWales, gamma = FALSE), n.ahead=1)[1]
```

```
[1] 3151837
```

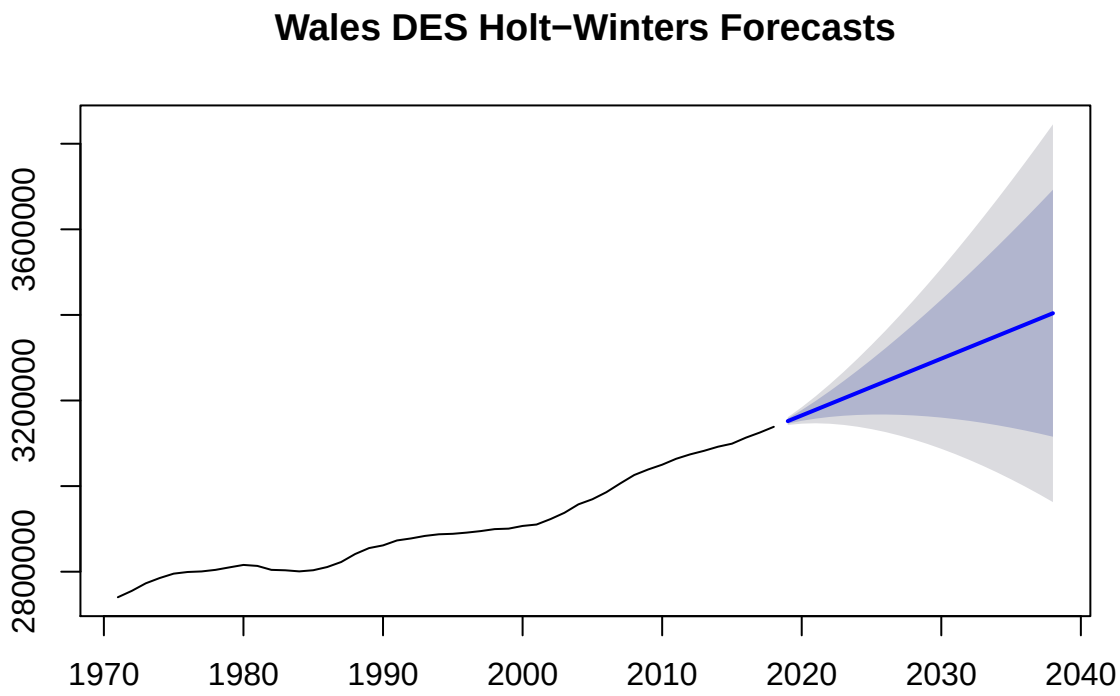
```
##Prediction for 2020 using DES
predict(HoltWinters(mytsWales, gamma = FALSE), n.ahead=2)[2]
```

```
[1] 3165107
```

6 Warnings about the accuracy of the forecasts in this report

It is important to remember that forecasts are rarely 100% accurate and their accuracy decreases the further into the future they are. The forecasts for 2019 and 2020 should be quite accurate as they are only one year and two years into the future. Below is a plot of the population forecasts for Wales. Note how wide the 95% and 80% confidence intervals are the further into the future the forecast.

```
## Light grey area is the 95% CI
## Blue area is the 80% CI
## The blue line is the plotted line of all the forecasts
  plot(forecast(HoltWinters(mytsWales, gamma = FALSE), h = 20), main = "Wales DES Holt-Winters Forecast
```



6.1 External factors that can affect the forecasts

Holt-Winters Algorithms only take into account previous years data and do not account for changes in any of the below external factors when calculating forecasts, changes in these external factors can greatly affect the accuracy of the forecasts.

- Government Policies can affect population, such as emigration/immigration polices and child benefit policies.
- Recessions/ Economic hardship typically result in population decreases.
- Unexpected Events
  - War & Famine: Both result in decreases in populations.
  - Diseases & Virus: Covid-19 has resulted in thousands of deaths across the UK, thus reducing the population, and has made forecasts less accurate as this is a new occurrence and the past data that has been used to create these forecasts was not subject to the affects of Covid-19 on the populations across the UK.