UK driver deaths modeling

Saikat Banerjee

June 11, 2018

Synopsis

This data analysis makes use of the built-in UKDriverDeaths dataset in R. This analysis is focused on exploratory visualization, fitting the time-series data using two basic models and evaluate their relative performance.

Loading of the data in R and verification of it as a time-series

```
# Dataset for driver deaths in UK
driver <- UKDriverDeaths
# Verifying identity
is.ts(UKDriverDeaths)
```

```
## [1] TRUE
```

```
# The start and end of the UKDriverDeaths data
start(UKDriverDeaths); end(UKDriverDeaths)
```

```
## [1] 1969    1
```

```
## [1] 1984   12
```

How many data points are there each year?

```
frequency((UKDriverDeaths))
```

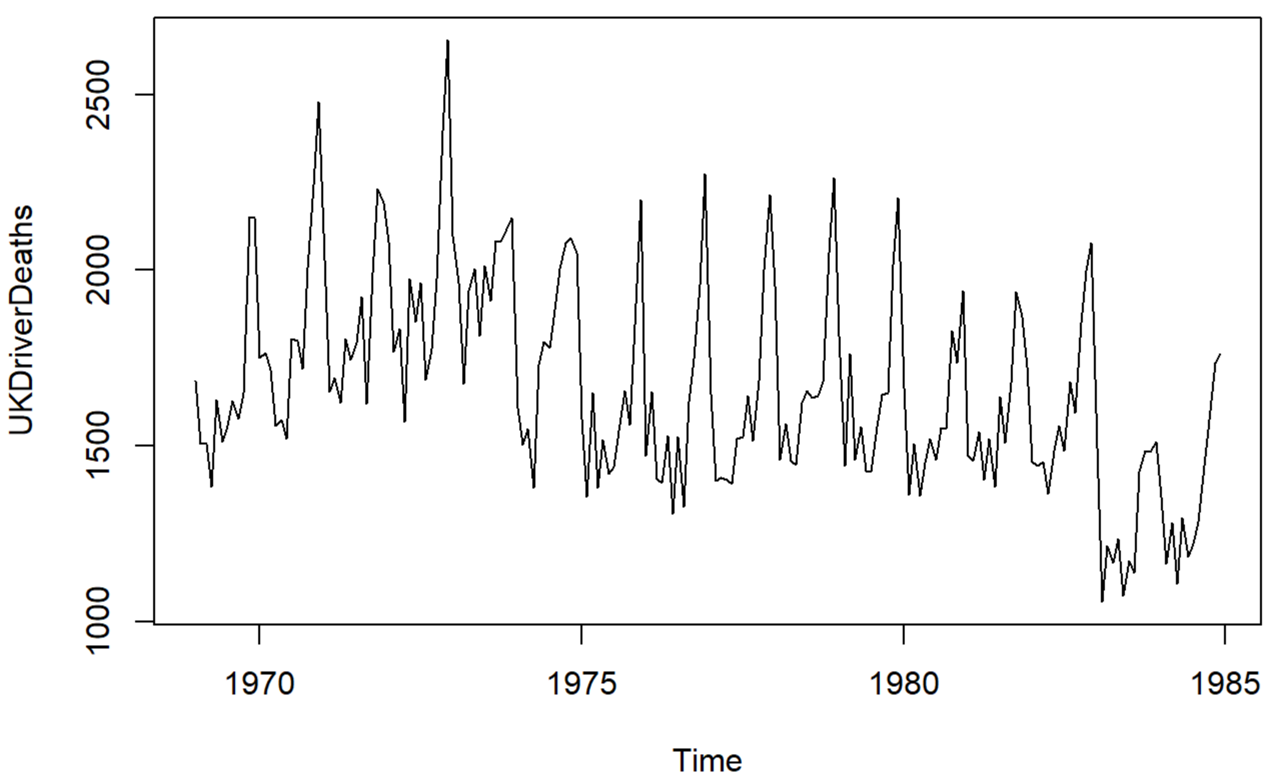
```
## [1] 12
```

```
cycle(UKDriverDeaths)
```

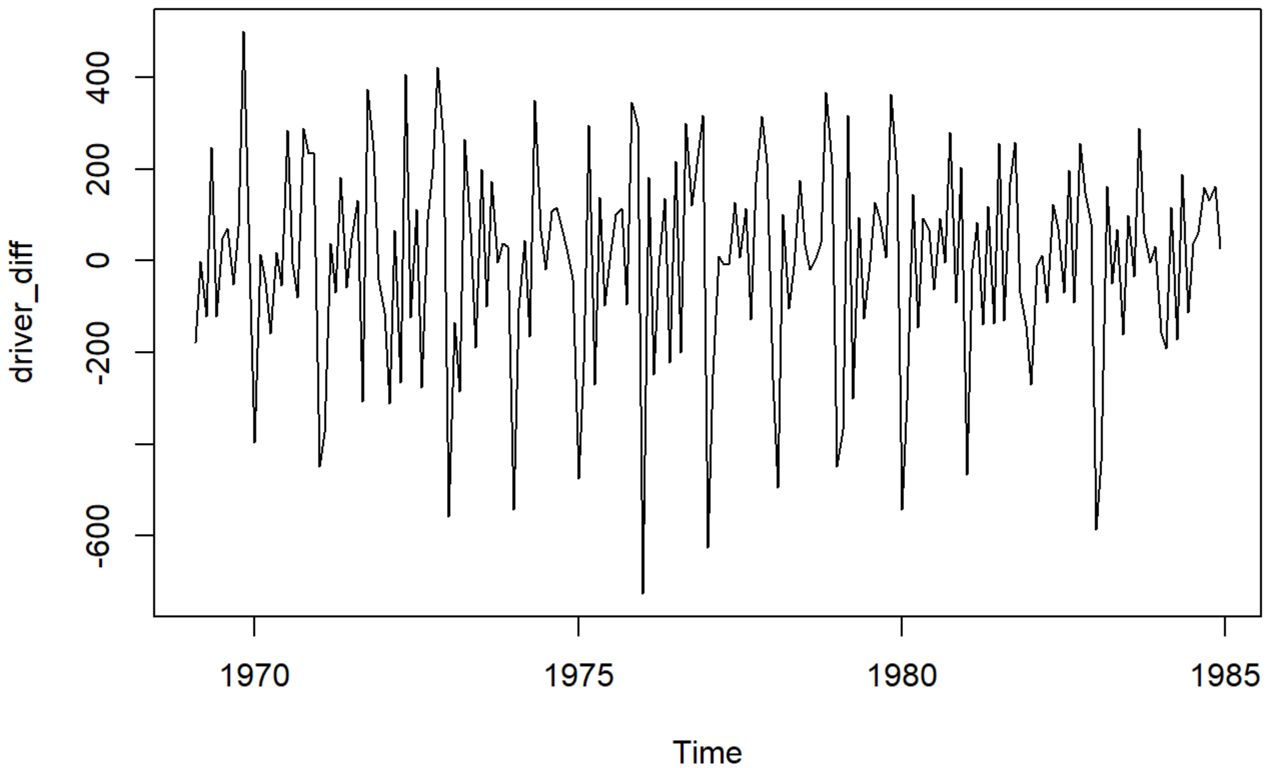
```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1969   1   2   3   4   5   6   7   8   9  10  11  12
## 1970   1   2   3   4   5   6   7   8   9  10  11  12
## 1971   1   2   3   4   5   6   7   8   9  10  11  12
## 1972   1   2   3   4   5   6   7   8   9  10  11  12
## 1973   1   2   3   4   5   6   7   8   9  10  11  12
## 1974   1   2   3   4   5   6   7   8   9  10  11  12
## 1975   1   2   3   4   5   6   7   8   9  10  11  12
## 1976   1   2   3   4   5   6   7   8   9  10  11  12
## 1977   1   2   3   4   5   6   7   8   9  10  11  12
## 1978   1   2   3   4   5   6   7   8   9  10  11  12
## 1979   1   2   3   4   5   6   7   8   9  10  11  12
## 1980   1   2   3   4   5   6   7   8   9  10  11  12
## 1981   1   2   3   4   5   6   7   8   9  10  11  12
## 1982   1   2   3   4   5   6   7   8   9  10  11  12
## 1983   1   2   3   4   5   6   7   8   9  10  11  12
## 1984   1   2   3   4   5   6   7   8   9  10  11  12
```

Exploratory visulaization and basic stats on the dataset

```
# Visualizing the deaths of the UK drivers
plot(UKDriverDeaths)
```

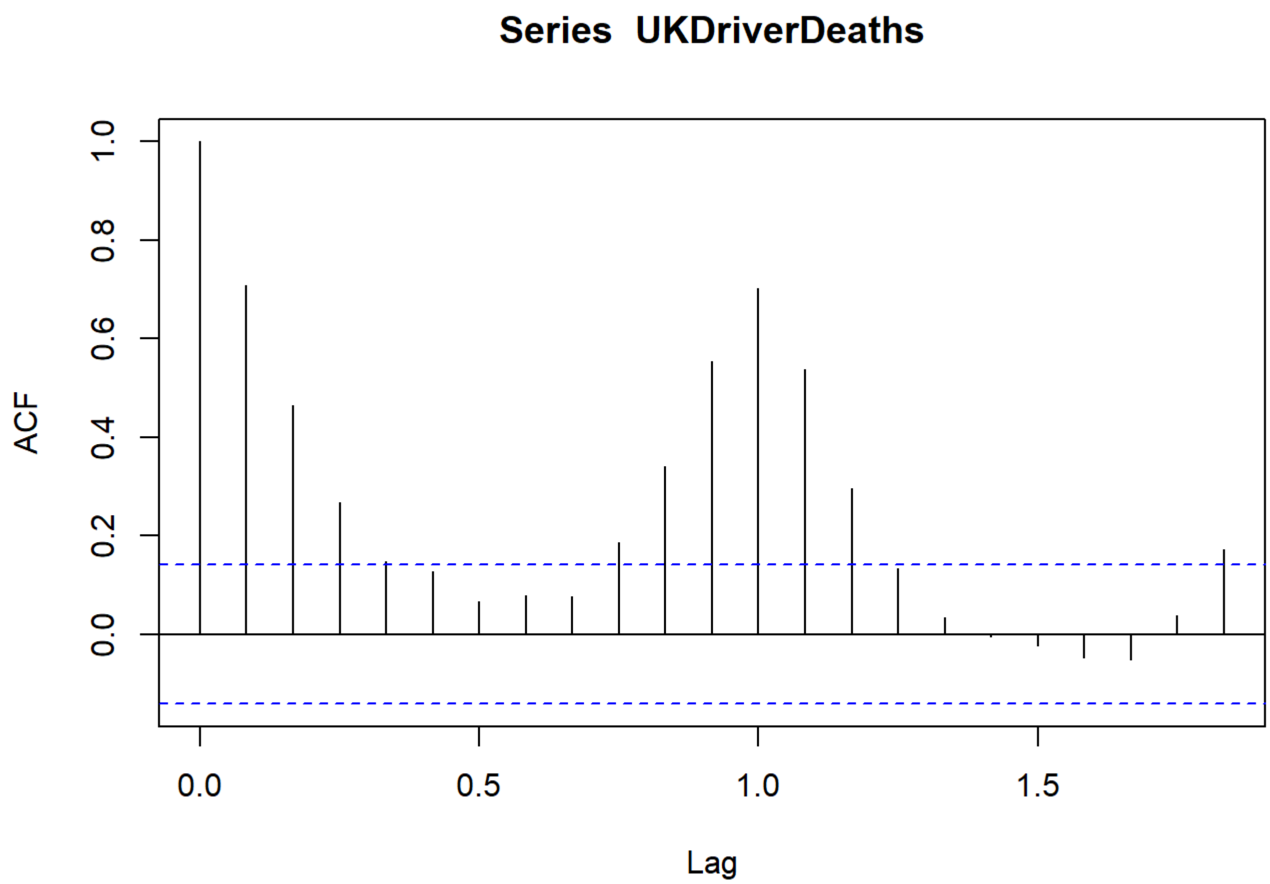


```
# mean and sd of the ts
mu <- mean(UKDriverDeaths)
sigma <- sd(UKDriverDeaths)
# Visualizing the first difference of the data,
# showing the trend with time is removed
driver_diff <- diff(UKDriverDeaths); ts.plot(driver_diff)
```



ACF function applied to the dataset

```
# calculate ACF function for the data
acf_driver <- acf(UKDriverDeaths)
```

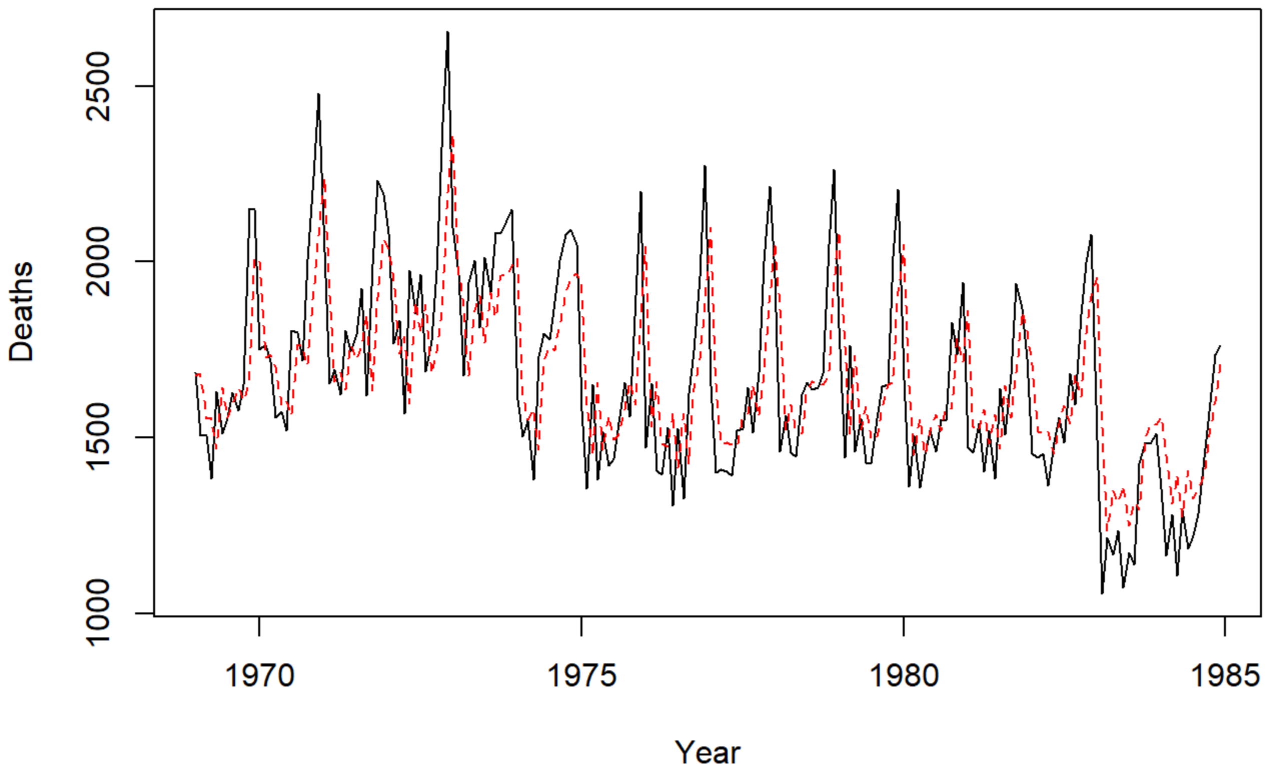


Autogresive modeling, fitting and prediction

```
# apply basic AR model to the data
driver_fit_AR <- arima(driver, order = c(1,0,0))
print(driver_fit_AR)
```

```
##
## Call:
## arima(x = driver, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.7060  1671.2584
## s.e.  0.0505    49.3558
##
## sigma^2 estimated as 41447:  log likelihood = -1293.47,  aic = 2592.94
```

```
ts.plot(driver, gpars = list(ylab = "Deaths", xlab = "Year"))
driver_fitted_AR <- driver - residuals(driver_fit_AR)
points(driver_fitted_AR, lty = 2, col = "red", type = "l")
```




```
# Forecasting events in the future year (1985)
```

Moving Average modeling, fitting and prediction

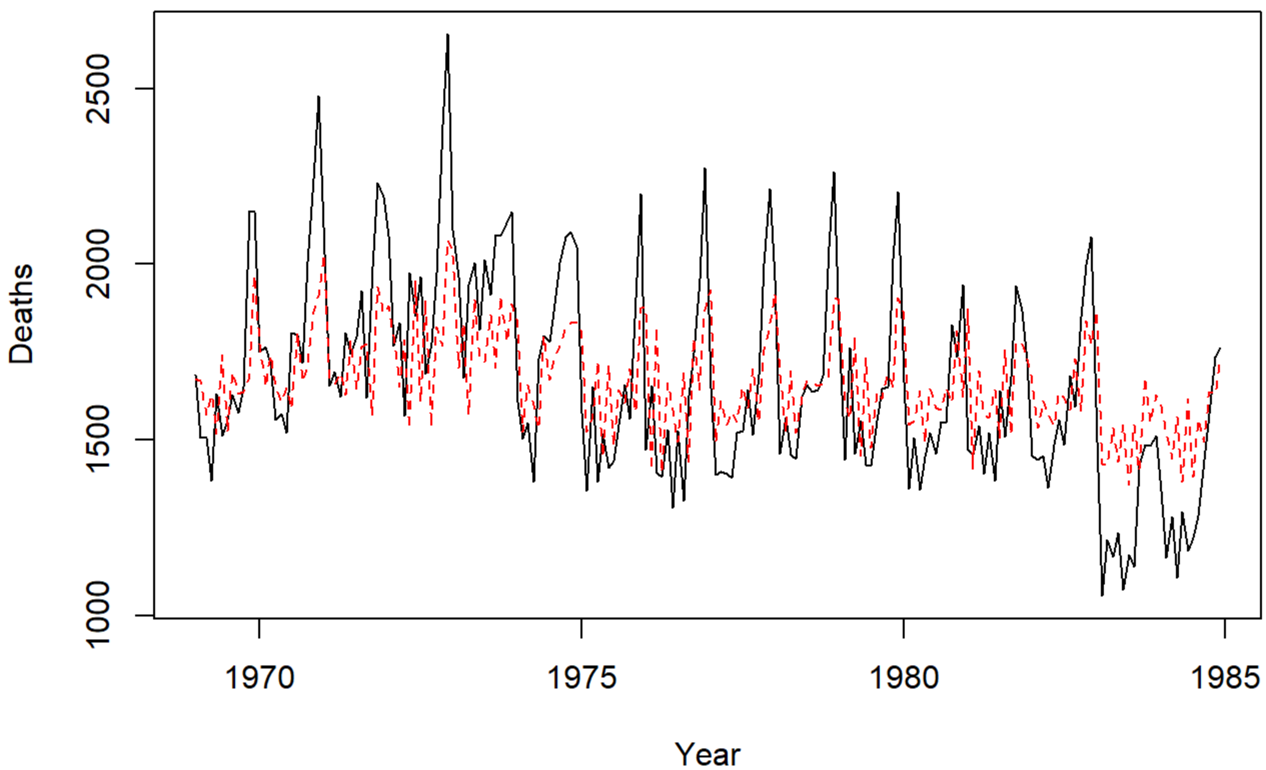
```
predict(driver_fit_AR, n.ahead = 12)
```

```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1985 1736.025 1716.981 1703.537 1694.046 1687.346 1682.615 1679.276
##           Aug      Sep      Oct      Nov      Dec
## 1985 1676.919 1675.254 1674.079 1673.250 1672.664
##
## $se
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1985 203.5861 249.2070 269.0710 278.4423 282.9970 285.2399 286.3511
##           Aug      Sep      Oct      Nov      Dec
## 1985 286.9034 287.1782 287.3151 287.3833 287.4172
```

```
# apply basic MA model to the data
driver_fit_MA <- arima(driver, order = c(0,0,1))
print(driver_fit_MA)
```

```
##
## Call:
## arima(x = driver, order = c(0, 0, 1))
##
## Coefficients:
##          ma1  intercept
##       0.6352  1670.8228
## s.e.  0.0545    26.4398
##
## sigma^2 estimated as 50399:  log likelihood = -1312.16,  aic = 2630.31
```

```
ts.plot(driver, gpars = list(ylab = "Deaths", xlab = "Year"))
driver_fitted_MA <- driver - residuals(driver_fit_MA)
points(driver_fitted_MA, lty = 2, col = "red", type = "l")
```




```
# Forecasting events in the future year (1985)
predict(driver_fit_MA, n.ahead = 12)
```

```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1985 1686.474 1670.823 1670.823 1670.823 1670.823 1670.823 1670.823
##           Aug      Sep      Oct      Nov      Dec
## 1985 1670.823 1670.823 1670.823 1670.823 1670.823
##
## $se
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1985 224.4983 265.9603 265.9603 265.9603 265.9603 265.9603 265.9603
##           Aug      Sep      Oct      Nov      Dec
## 1985 265.9603 265.9603 265.9603 265.9603 265.9603
```

Which model is better?

```
AIC(driver_fit_AR); BIC(driver_fit_AR)
```

```
## [1] 2592.941
```

```
## [1] 2602.713
```

```
AIC(driver_fit_MA); BIC(driver_fit_MA)
```

```
## [1] 2630.314
```

```
## [1] 2640.087
```

AR is a better model than MA for this dataset based on the AIC and BIC values