

Task 1

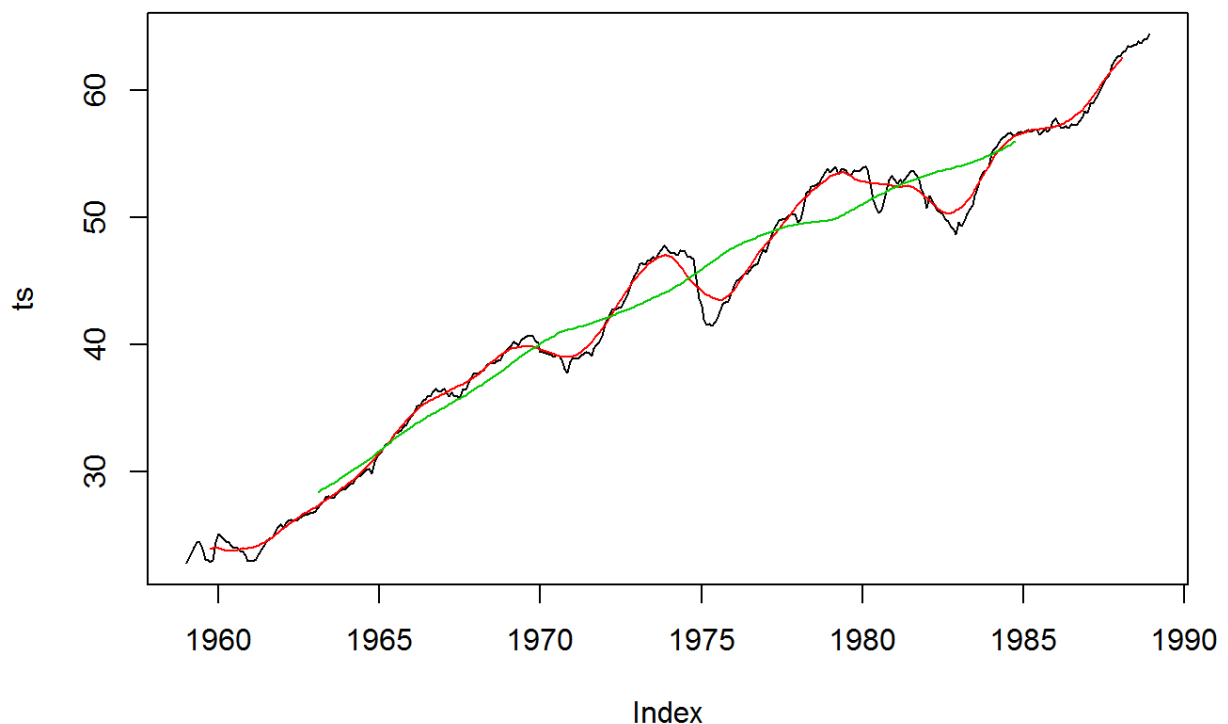
Komissarov, Maksimov, Oblygina

12 March 2017

1 этап

Reading of data

```
ts <- read.csv.zoo("training.csv",  
                  index.column = 1,  
                  FUN = as.Date, format = "%Y-%m-%d")  
  
ts.monthly <- zooreg(ts, frequency = 12, start = c(1959, 1))  
  
test <- zooreg(read.csv.zoo("testing.csv",  
                             index.column = 1,  
                             FUN = as.Date, format = "%Y-%m-%d"),  
               frequency = 12, start = c(1989, 1))
```



Как можно заметить, график не имеет ярких выбросов и имеет ярко выраженный тренд. Скользящая средняя явно возрастает. Из это можно сделать вывод о том, что ряд не является стационарным в широком смысле. Убедимся в этом, проведя тест Дики - Фуллера:

Dickey-Fuller

```
adf.test(ts, k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts  
## Dickey-Fuller = -1.4976, Lag order = 0, p-value = 0.7889  
## alternative hypothesis: stationary
```

Есть единичные корни, ряд не стационарен. Высокое p-value дает нам возможность утверждать, что ряд не стационарен.

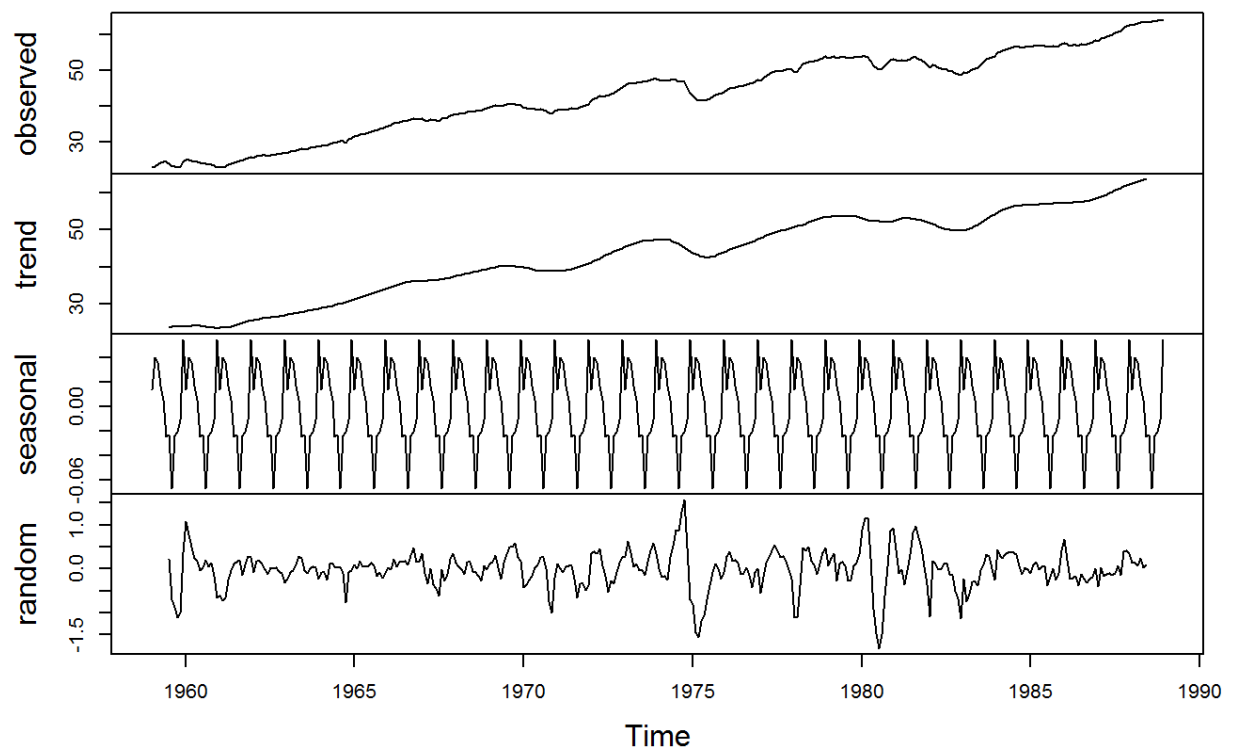
2 этап

Decomposition

Для дальнейшего анализа проведем разложение ряда на тренд, сезонность и остаток в соответствии с аддитивной и мультипликативной моделями. Выведем данные графики.

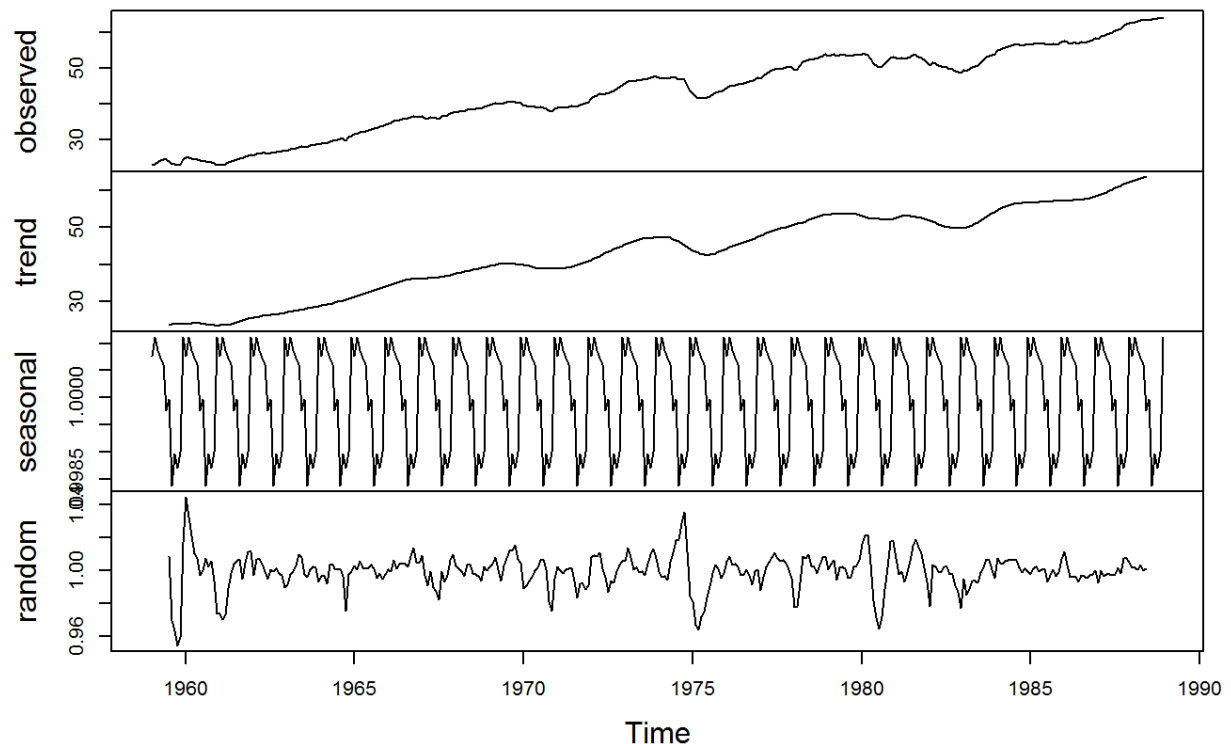
```
ts.add <- decompose(as.ts(ts.monthly), type = "additive")  
ts.mul <- decompose(as.ts(ts.monthly), type = "multiplicative")  
plot(ts.add)
```

Decomposition of additive time series



```
plot(ts.mul)
```

Decomposition of multiplicative time series



Как можно заметить тренд в обеих моделях имеет чёткое соответствие стационарности с имеющимся у нас временным рядом. То есть он также не стационарен. А сезонность и остаток являются стационарными в широком смысле, так как все статистические характеристики не меняются с течением времени. Для того, чтобы удостовериться в этом, проведем тест Дики-Фуллера.

```
adf.test(na.remove(ts.add$trend), k = 0)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: na.remove(ts.add$trend)
## Dickey-Fuller = -0.6874, Lag order = 0, p-value = 0.9711
## alternative hypothesis: stationary
```

```
adf.test(na.remove(ts.add$seasonal), k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: na.remove(ts.add$seasonal)  
## Dickey-Fuller = -10.169, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(na.remove(ts.add$random), k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: na.remove(ts.add$random)  
## Dickey-Fuller = -7.3949, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(na.remove(ts.mul$trend), k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: na.remove(ts.mul$trend)  
## Dickey-Fuller = -0.6874, Lag order = 0, p-value = 0.9711  
## alternative hypothesis: stationary
```

```
adf.test(na.remove(ts.mul$seasonal), k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: na.remove(ts.mul$seasonal)  
## Dickey-Fuller = -9.192, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(na.remove(ts.mul$random), k = 0)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: na.remove(ts.mul$random)  
## Dickey-Fuller = -7.5842, Lag order = 0, p-value = 0.01  
## alternative hypothesis: stationary
```

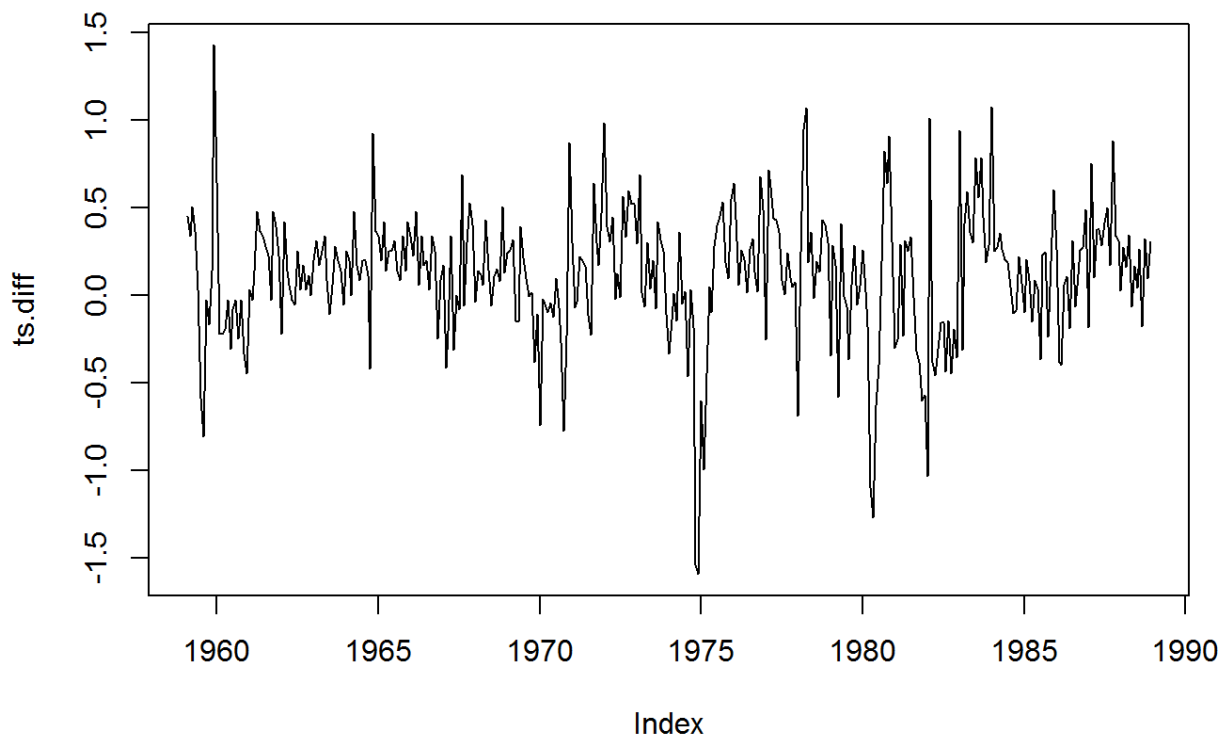
Наше предположение верно. В обеих моделях значение параметра p-value у тренда большое, следовательно, ряд не является стационарным. А для остатка и сезональности он мал, что подтверждает их стационарность.

3 этап

Order of integration

Для определения порядка интегрированности временного ряда достаточно проанализировать последовательно его разности.

```
ts.diff <- diff(ts, differences = 1)
plot(ts.diff)
```



```
adf.test(ts.diff, k = 0)
```

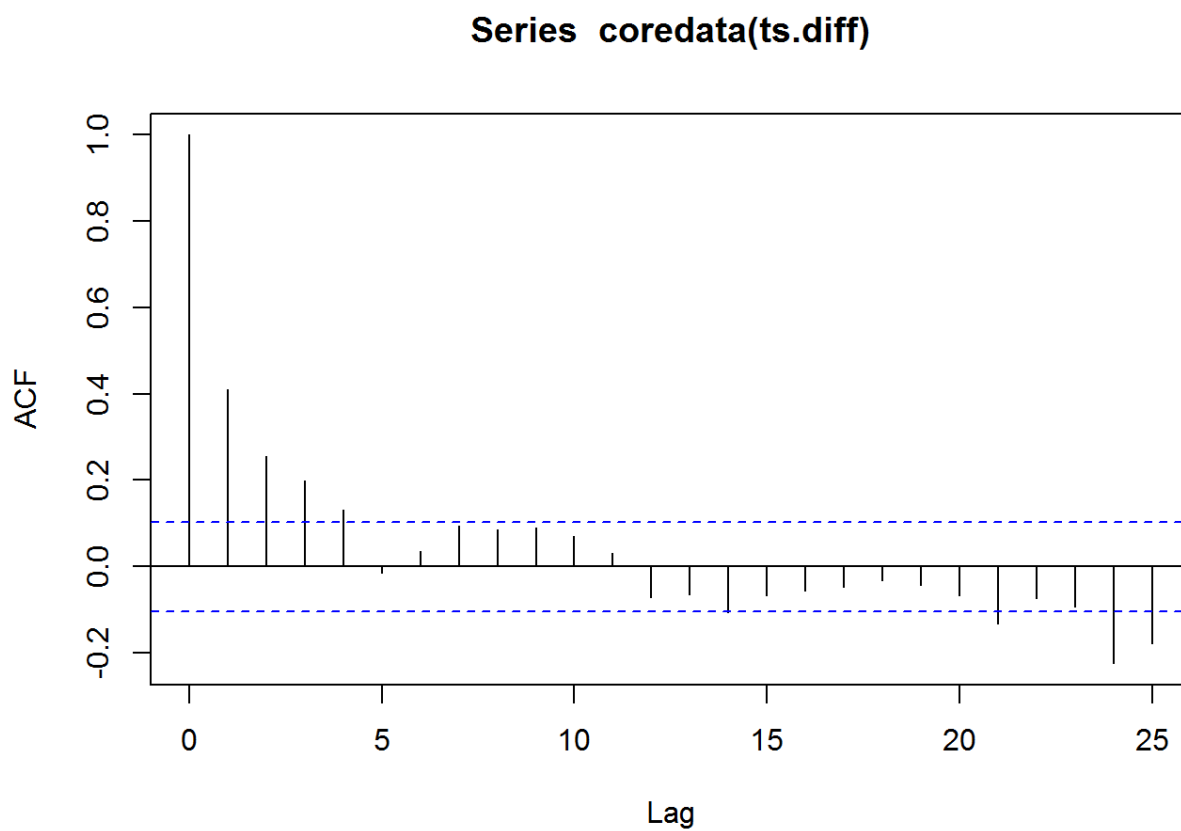
```
##
## Augmented Dickey-Fuller Test
##
## data: ts.diff
## Dickey-Fuller = -12.2, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary
```

Согласно тесту Дики-Фуллера первая разность исходного ряда является стационарной, значит, порядок интегрированности равен 1.

Correlation coefficients

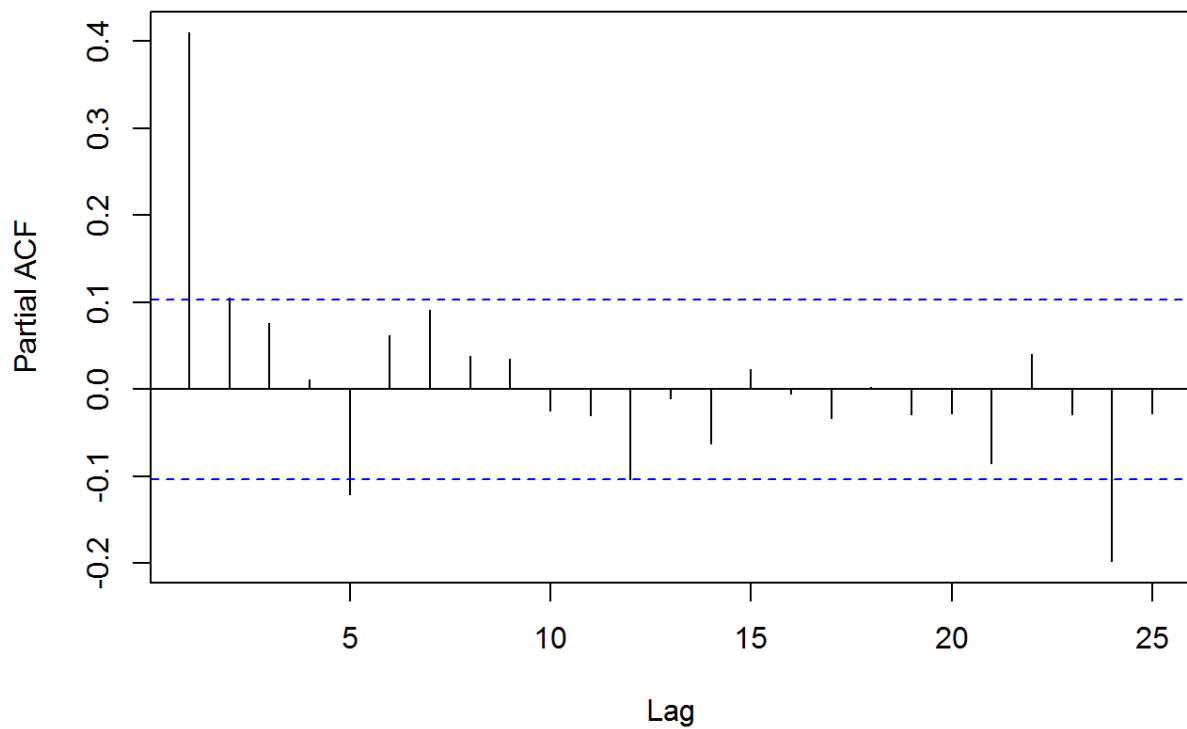
Построим функции автокорреляции и частичной автокорреляции.

```
acf(coredata(ts.diff))
```



```
pacf(coredata(ts.diff))
```

Series coredata(ts.diff)

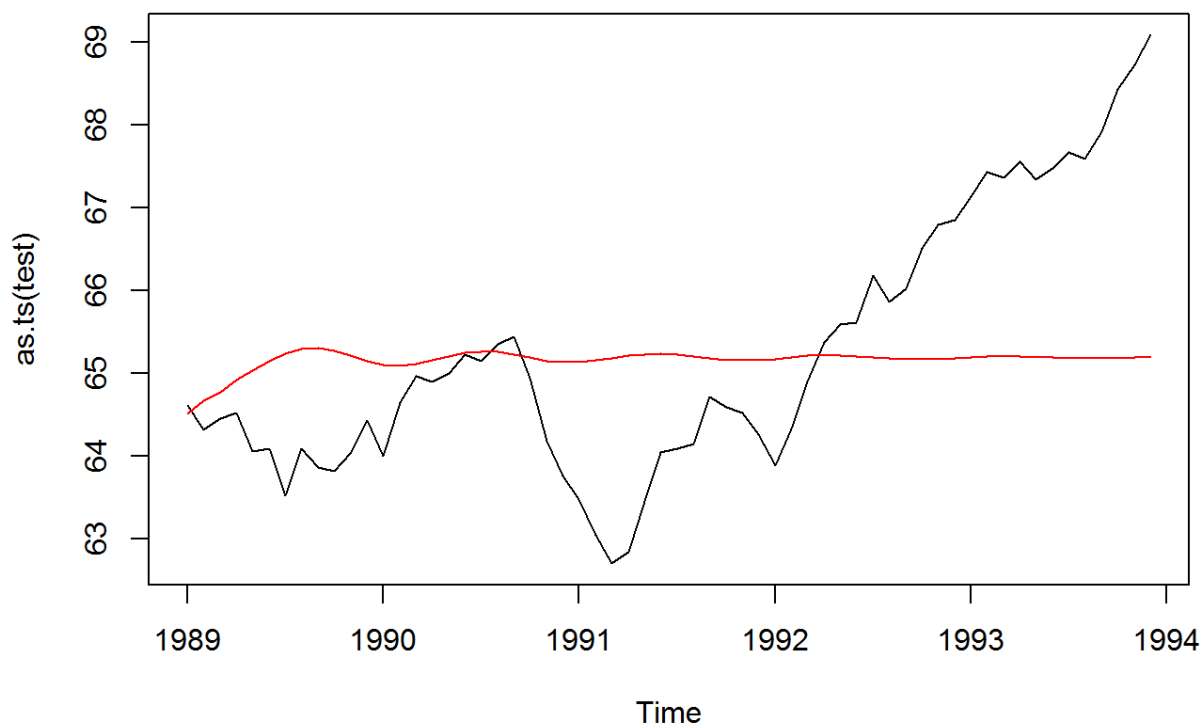


По наличию пиков можно предположить оптимальность моделей ARIMA(4, 1, 6) и ARIMA(3, 1, 5)

ARIMA 1

Применим модель ARIMA(4, 1, 6). Сделаем прогноз на 60 месяцев.

```
ts.arima_1 = Arima(ts.monthly, order = c(4, 1, 6))  
for.arima_1 = forecast(ts.arima_1, h = 60)  
plot(as.ts(test)); lines(for.arima_1$mean, col = 2)
```



Посчитаем r2-score.

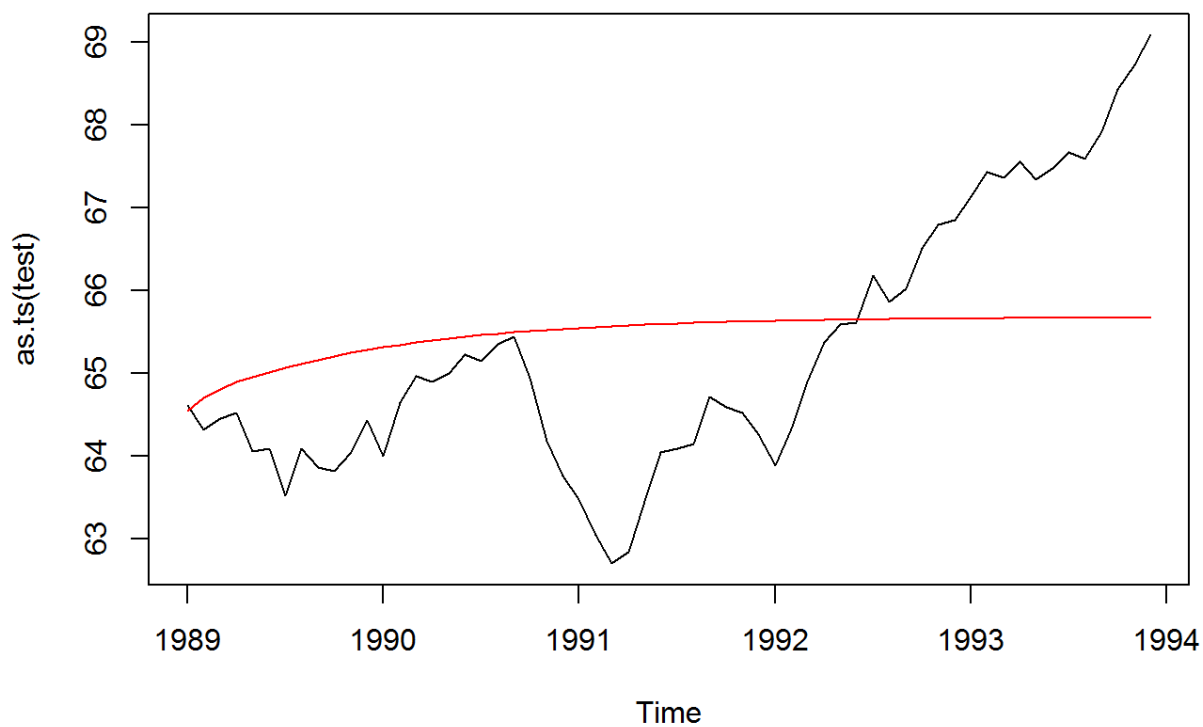
```
R2_Score(for.arima_1$mean, test)
```

```
## [1] 0.01096331
```

ARIMA 2

Применим модель ARIMA(4, 1, 6). Сделаем прогноз на 60 месяцев.

```
ts.arima_2 = Arima(ts.monthly, order = c(3, 1, 5))  
for.arima_2 = forecast(ts.arima_2, h = 60)  
plot(as.ts(test)); lines(for.arima_2$mean, col = 2)
```

Посчитаем r2-score.

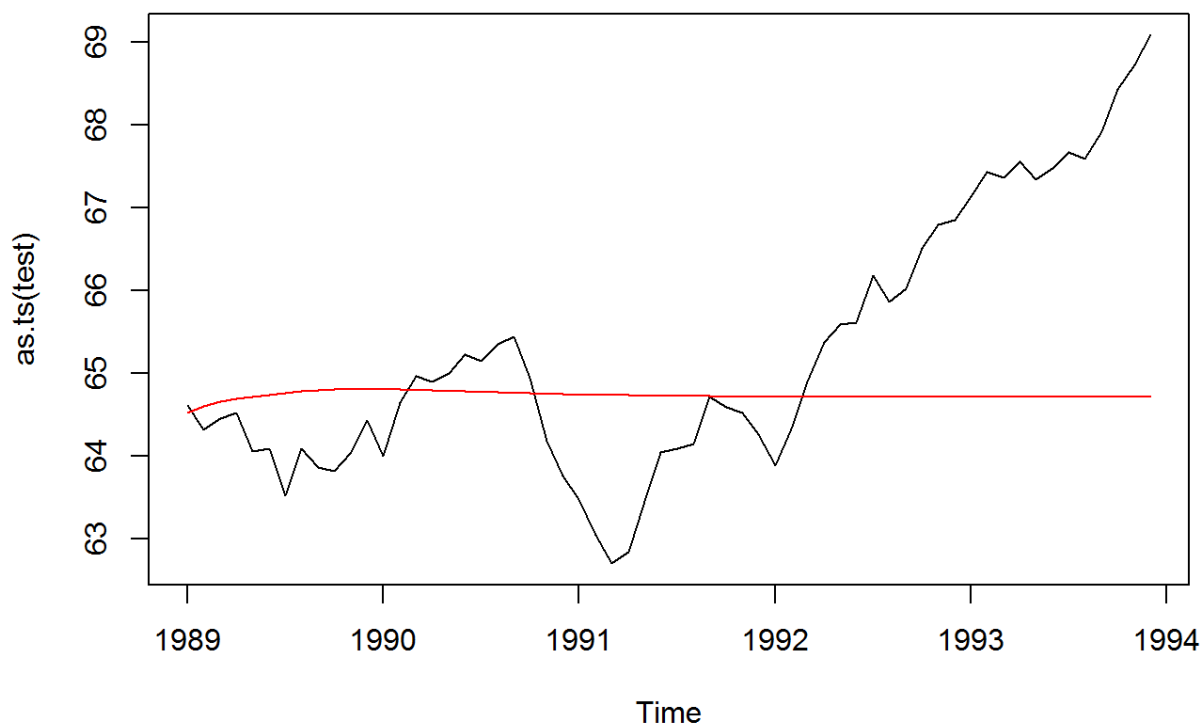
```
R2_Score(for.arima_2$mean, test)
```

```
## [1] 0.1008945
```

Neural Network

Применим нейронную сеть с 5 вложенными переменными. Сделаем прогноз на 60 месяцев.

```
ts.nn <- nnetar(as.ts(ts.monthly), p = 5)
for.nn = forecast(ts.nn, h = 60)
plot(as.ts(test)); lines(for.nn$mean, col = 2)
```



Посчитаем r2-score.

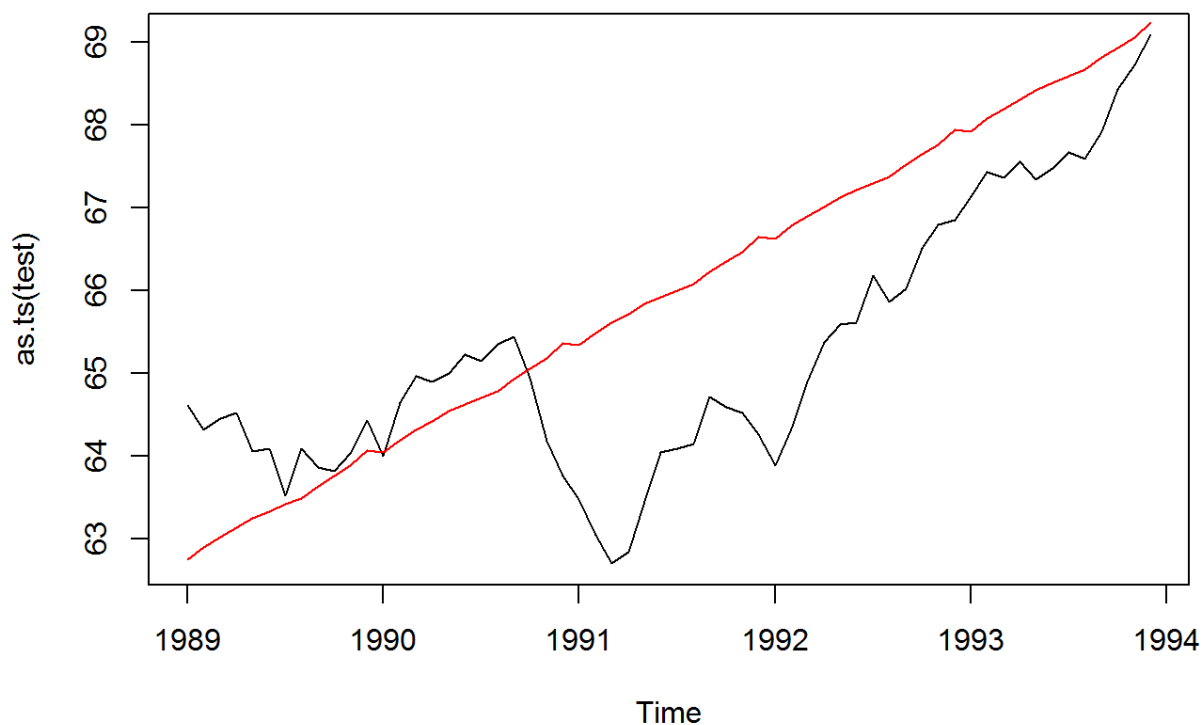
```
R2_Score(for.nn$mean, test)
```

```
## [1] -0.1235844
```

Linear Model

Применим простую линейную модель, предварительно декомпозирующую ряд на тренд и сезонность. Сделаем прогноз на 60 месяцев.

```
ts.lm <- tslm(as.ts(ts.monthly) ~ trend + season)
for.lm = forecast(ts.lm, h = 60)
plot(as.ts(test)); lines(for.lm$mean, col = 2)
```



Посчитаем r2-score.

```
R2_Score(for.lm$mean, test)
```

```
## [1] 0.1968743
```

Conclusion

ARIMA и нейронная сеть предсказывают результат на конец прогнозируемого срока хуже, чем линейная модель. Нейронная сеть и ARIMA предсказывают результаты начала срока лучше линейной модели. По результатам r2-score только линейная модель применима для тестовых данных.