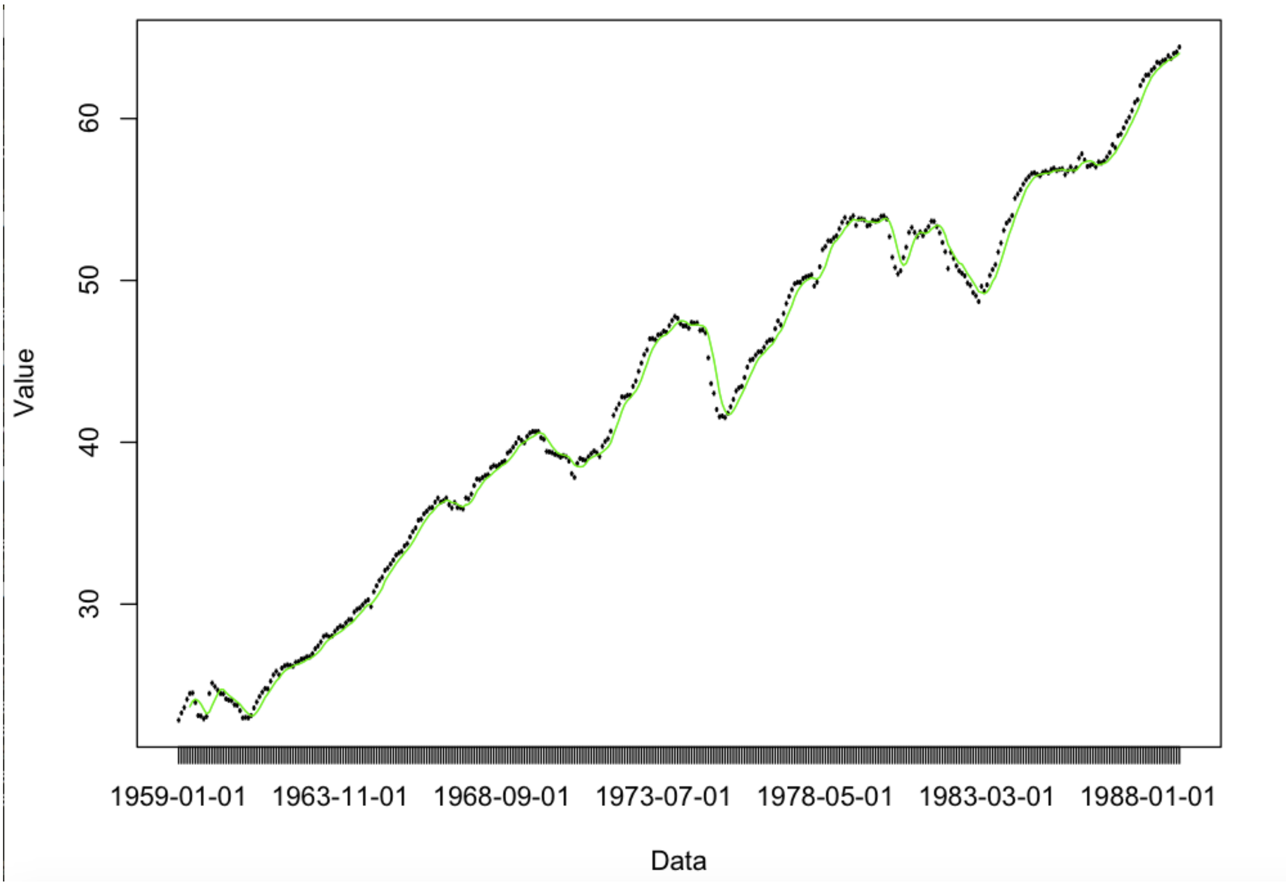
Пункт 1.

Считываем данные и проводим визуальную оценку нашего ряда. Для этого построим на графике скользящие статистики: SMA, WMA, EMA. При расчёте скользящего среднего значение функции вычисляется каждый раз заново, при этом учитывается конечное значимое множество предыдущих значений. Усредняющие функции демонстрируют ярко выраженные тренд, на основе чего мы можем сделать вывод, что ряд является не стационарным, так как математическое ожидание растет со временем.



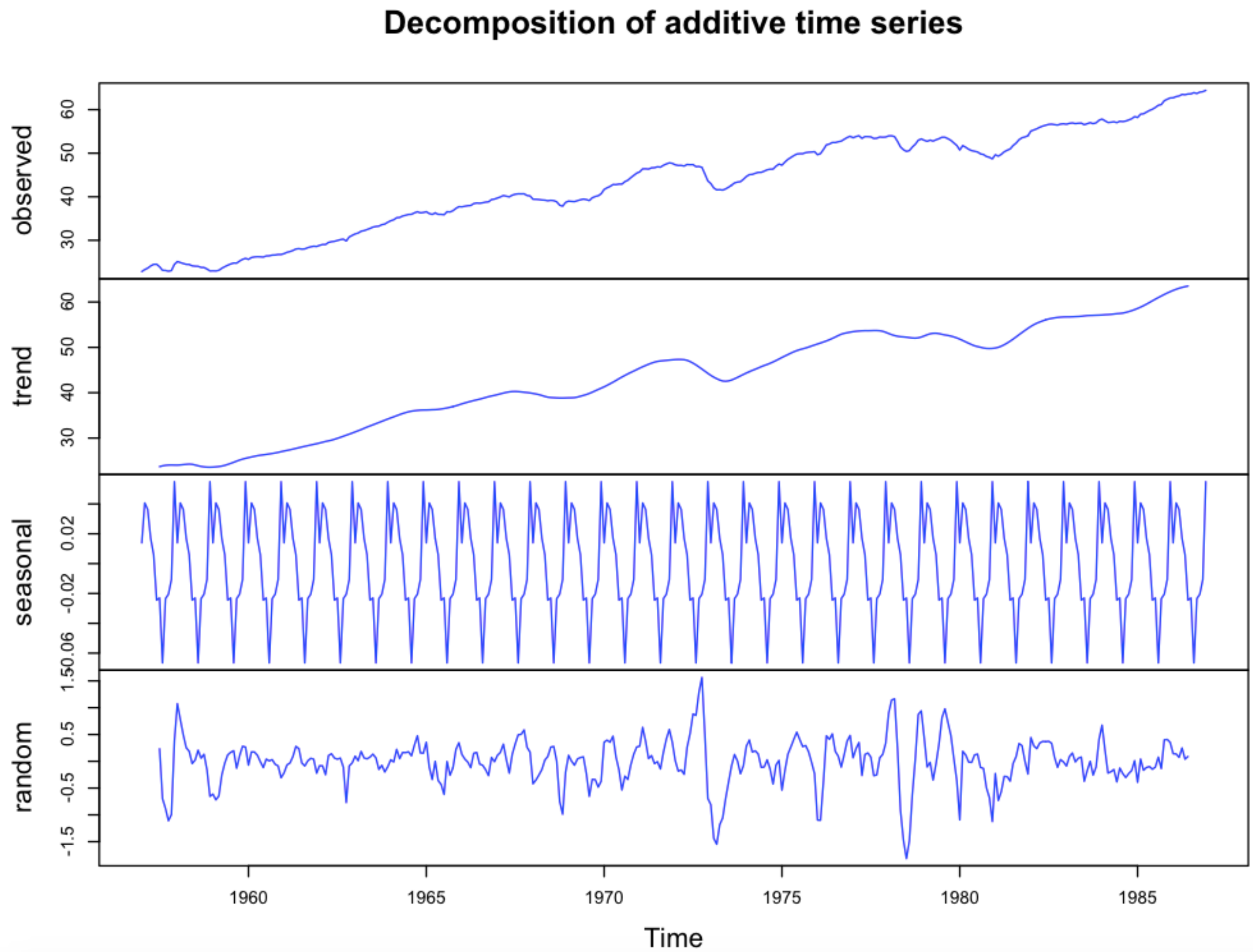
Для более подробного анализа ряда на стационарность проведем тест Дики - Фуллера.

```
> adf.test(pppdata.02, alternative=c('stationary'))

	Augmented Dickey-Fuller Test

data:  pppdata.02
Dickey-Fuller = -3.2505, Lag order = 7, p-value = 0.07962
alternative hypothesis: stationary
```

В силу того, что p-value велико(рначения p-value выше уровня значимости (5%)), отвергается гипотеза о стационарности. Соответственно мы делаем вывод, что наш ряд не стационарный.

Пункт 2.

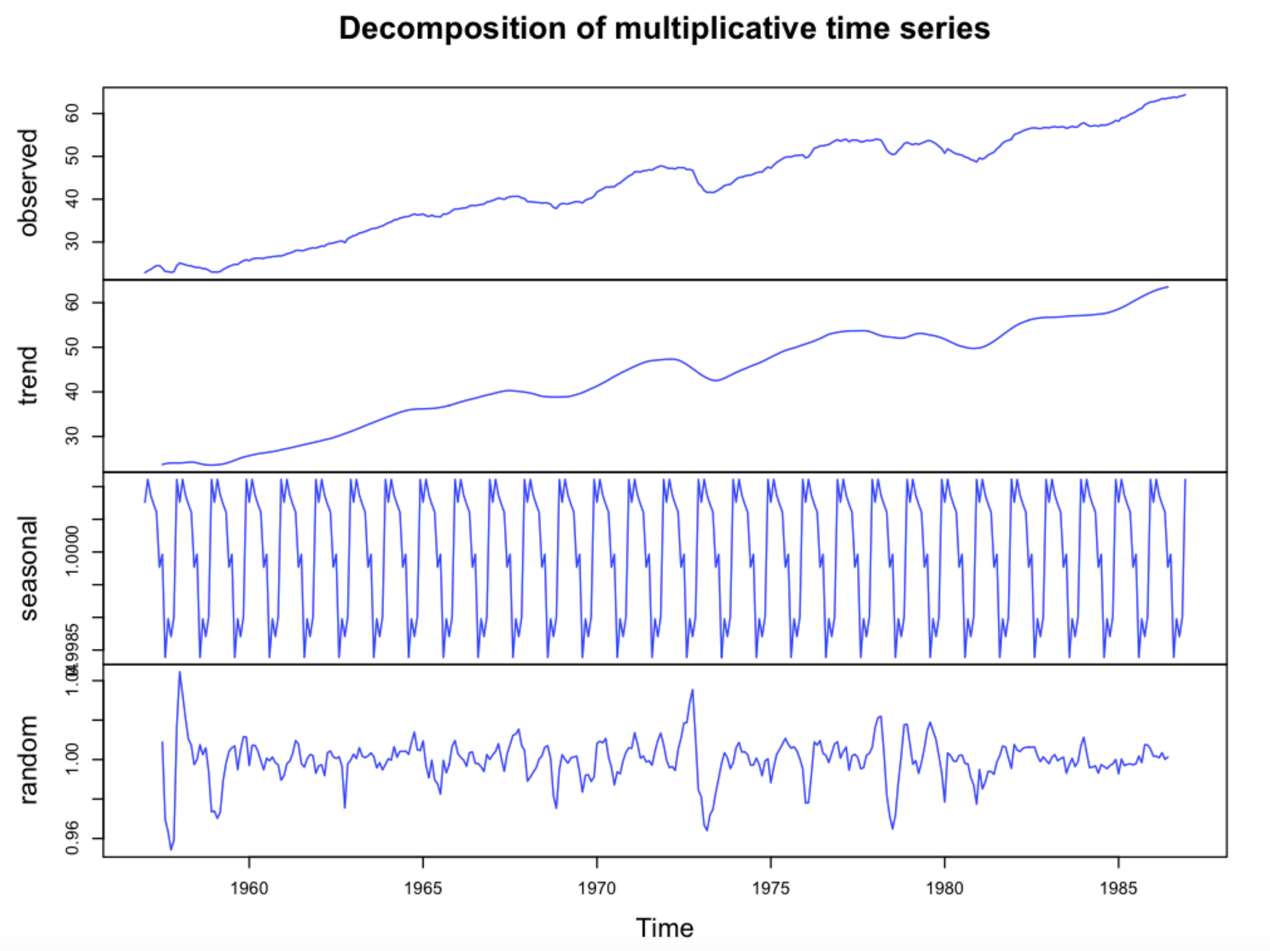
Визуализируем тренд, сезональность, остаток временного ряда в соответствии с аддитивной, мультипликативной моделями.



Для аддитивный модели: Основываясь на выводах первого пункта график тренда показывает нам, что ряд не стационарен. Для проверки на стационарность сезонной компоненты применим тест Дики - Фуллера.

```
> adf.test(as.ts(decompose_add_pppdata.02$seasonal),
alternative=c('stationary'))

	Augmented Dickey-Fuller Test

data:  as.ts(decompose_add_pppdata.02$seasonal)
Dickey-Fuller = -63.636, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Предупреждение:
В adf.test(as.ts(decompose_add_pppdata.02$seasonal), alternative =
c("stationary")) :
  p-value smaller than printed p-value
```

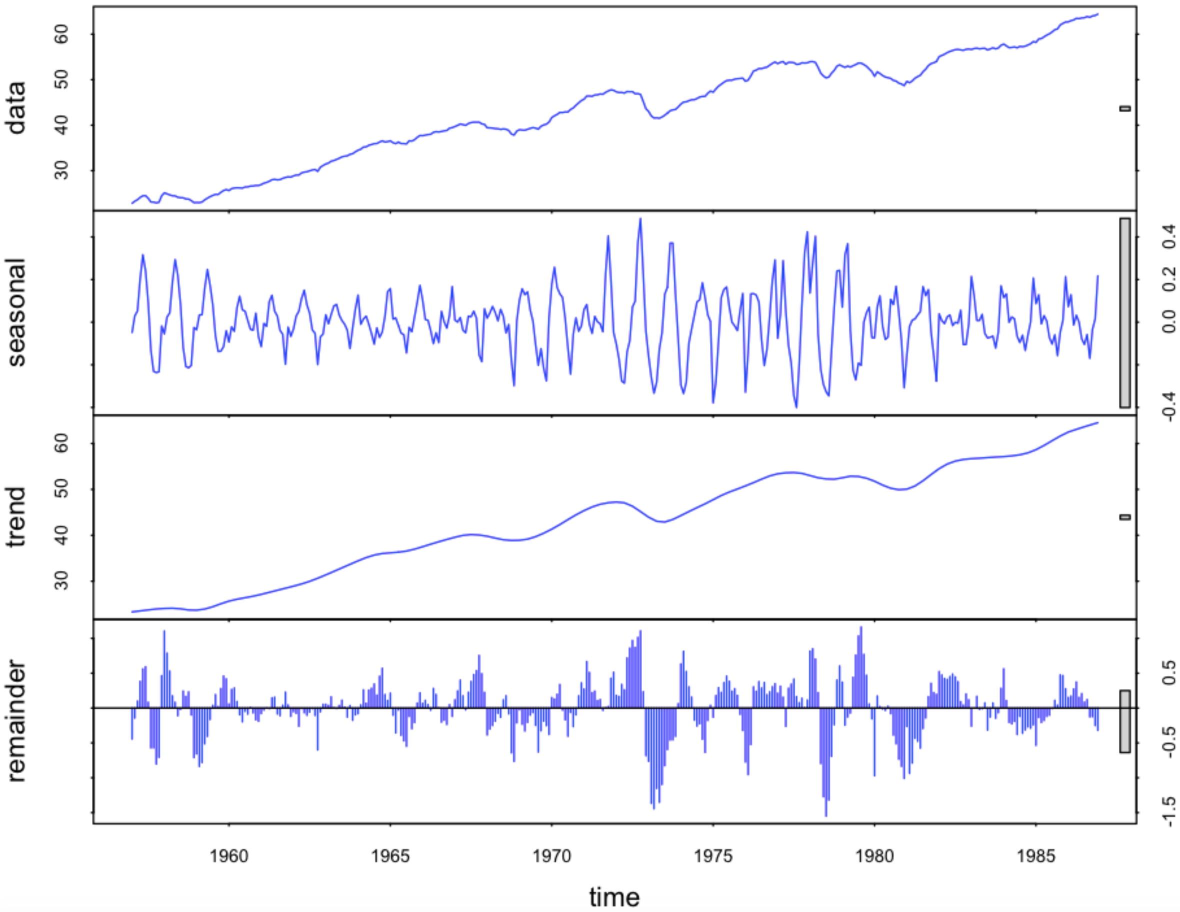
Поскольку *p*-value мало, то с уровнем значимости 0.01 данный ряд можно считать стационарным (отвергнута гипотеза о наличии единичного корня).



Для мультипликационной модели: Основываясь на выводах первого пункты график тренда показывает нам, что ряд не стационарен. Для проверки на стационарность сезонной компоненты применим тест Дики - Фуллера.

```
> adf.test(as.ts(decompose_mult_pppdata.02$seasonal),
alternative=c('stationary'))

	Augmented Dickey-Fuller Test

data:  as.ts(decompose_mult_pppdata.02$seasonal)
Dickey-Fuller = -37.814, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Предупреждение:
В adf.test(as.ts(decompose_mult_pppdata.02$seasonal), alternative =
c("stationary")) :
  p-value smaller than printed p-value
```

Поскольку *p*-value мало, то с уровнем значимости 0.01 данный ряд можно считать стационарным (отвергнута гипотеза о наличии единичного корня).



Исследуем остатки на стационарность:

```
> adf.test(random_stl_pppdata.02, alternative=c('stationary'))

	Augmented Dickey-Fuller Test

data:  random_stl_pppdata.02
Dickey-Fuller = -6.807, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Предупреждение:
В adf.test(random_stl_pppdata.02, alternative = c("stationary")) :
  p-value smaller than printed p-value
```

Поскольку *p*-value мало, то с уровнем значимости 0.01 данный ряд можно считать стационарным (отвергнута гипотеза о наличии единичного корня).

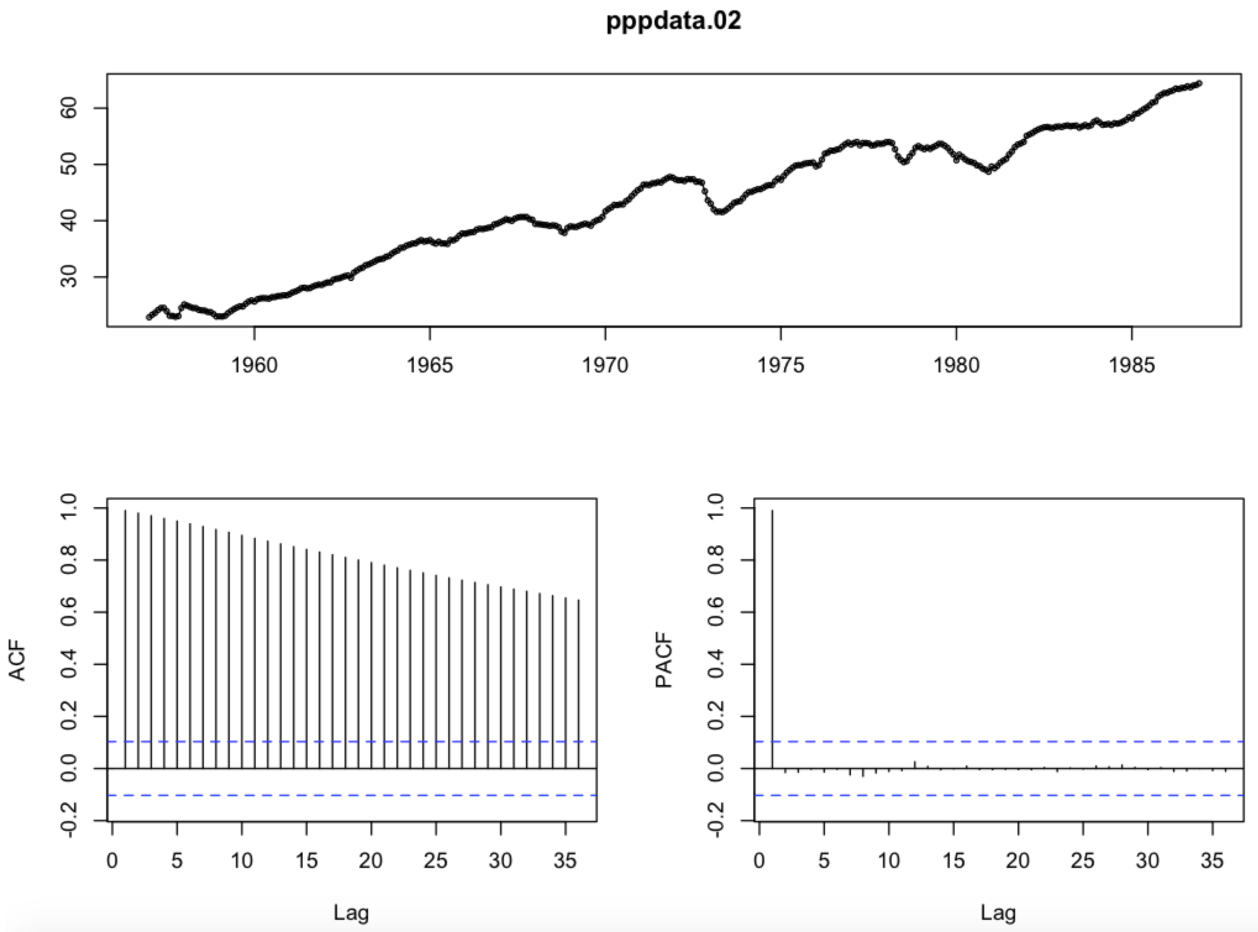
Пункт 3.

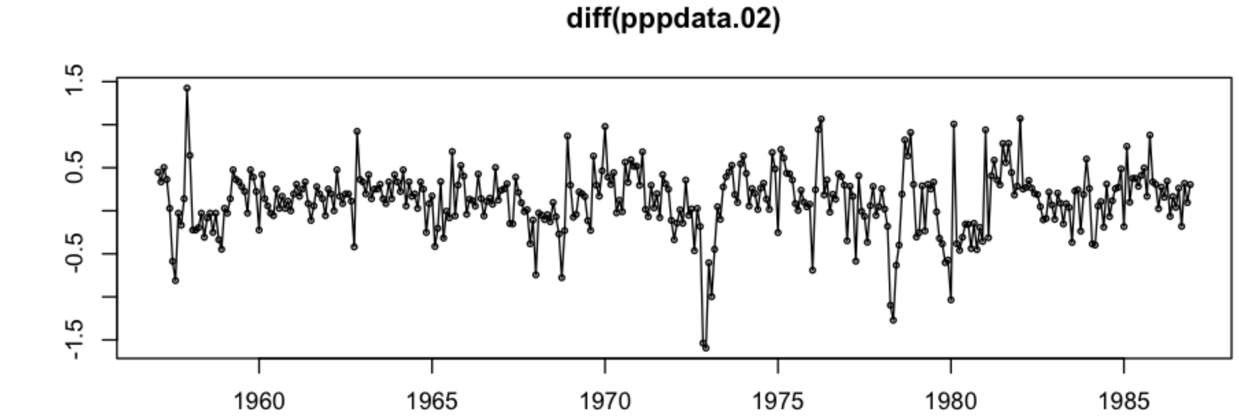
Для проверки на интегрированность k- ого порядка: В силу того, что исходный ряд не стационарный, берем ряд разностей исходного ряда и проверяем его на стационарность

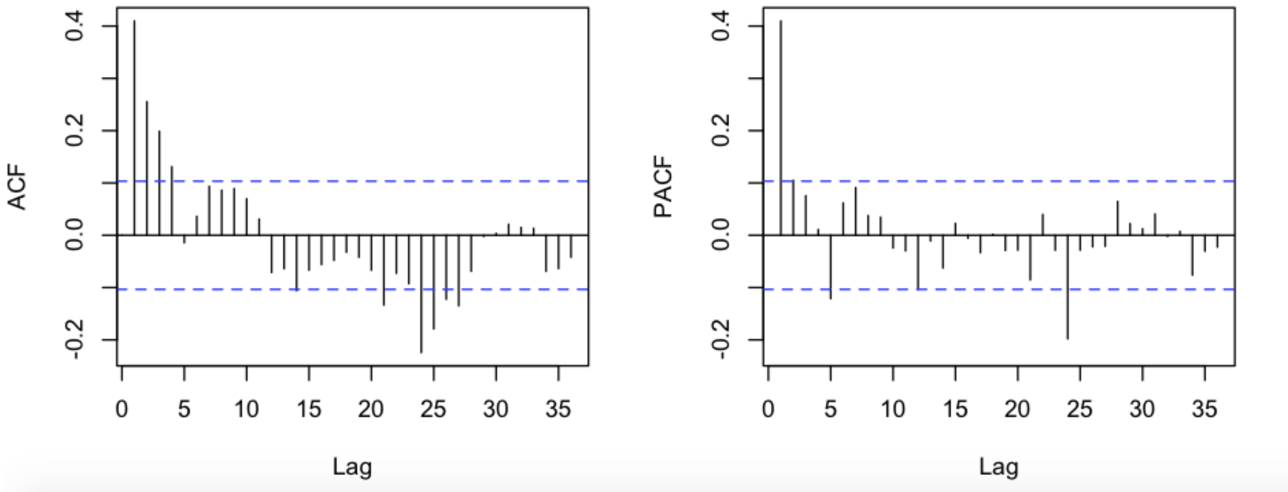
```
> adf.test(diff(pppdata.02), alternative=c('stationary'))

	Augmented Dickey-Fuller Test

data:  diff(pppdata.02)
Dickey-Fuller = -5.0661, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Предупреждение:
В adf.test(diff(pppdata.02), alternative = c("stationary")) :
  p-value smaller than printed p-value
```

Тест Дики-Фуллера говорит нам о том, что ряд разностей первого порядка стационарный. Соответственно наш ряд является интегрированным первого порядка.

Построим ACF и PACF

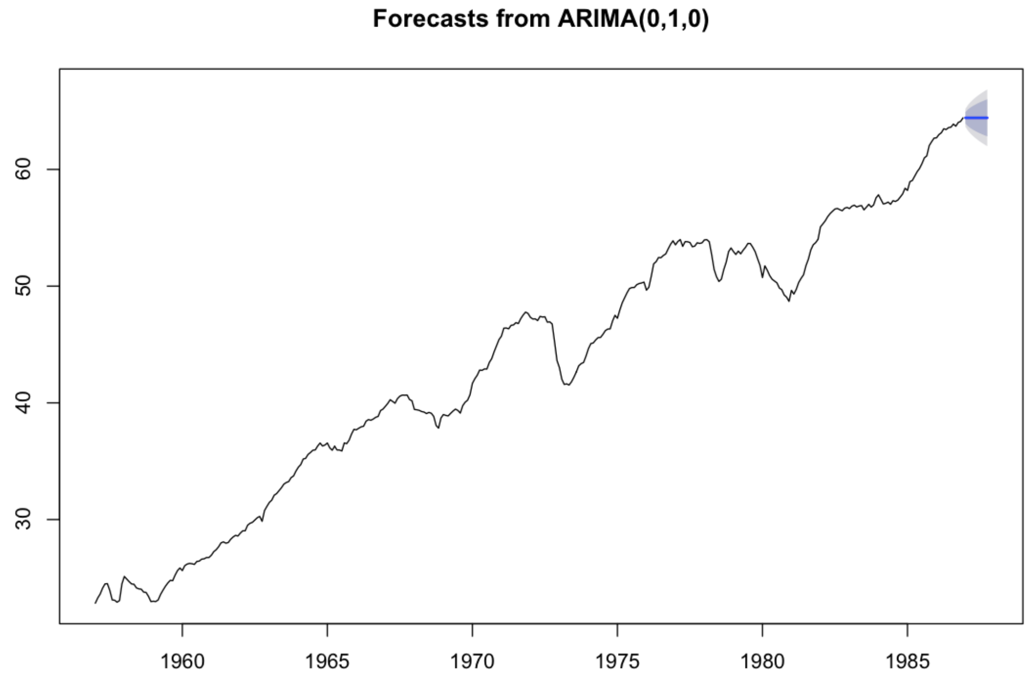




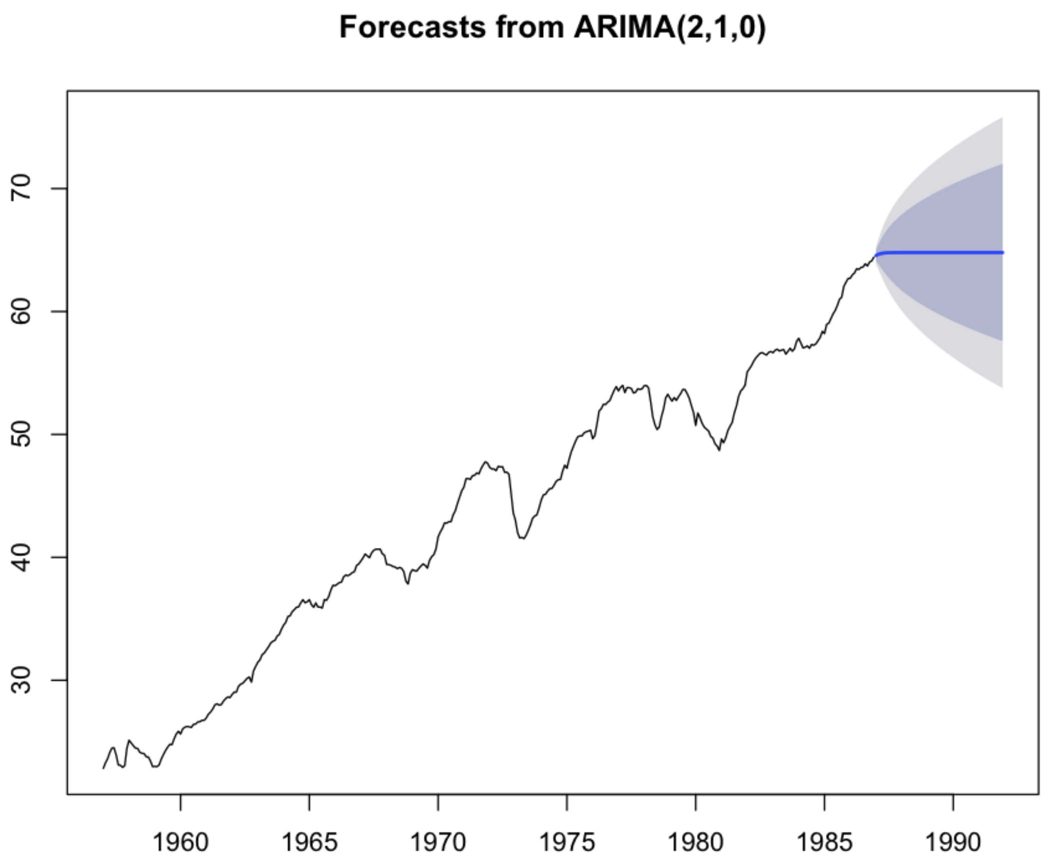


Анализируем ACF и PACF для разностей первого порядка. Построим различные модели ARIMA и проанализируем их, для этого определим параметры p, d, q. В коррелограмме ACF находим минимальный лаг, который значимо отличается от нуля. Поэтому в качестве параметра q можем взять коэффициент q = 4, аналогично для p = 2, d =1, так как ряд интегрируем 1-го порядка. Переберем различные варианты и выберем оптимальную модель для прогнозирования:

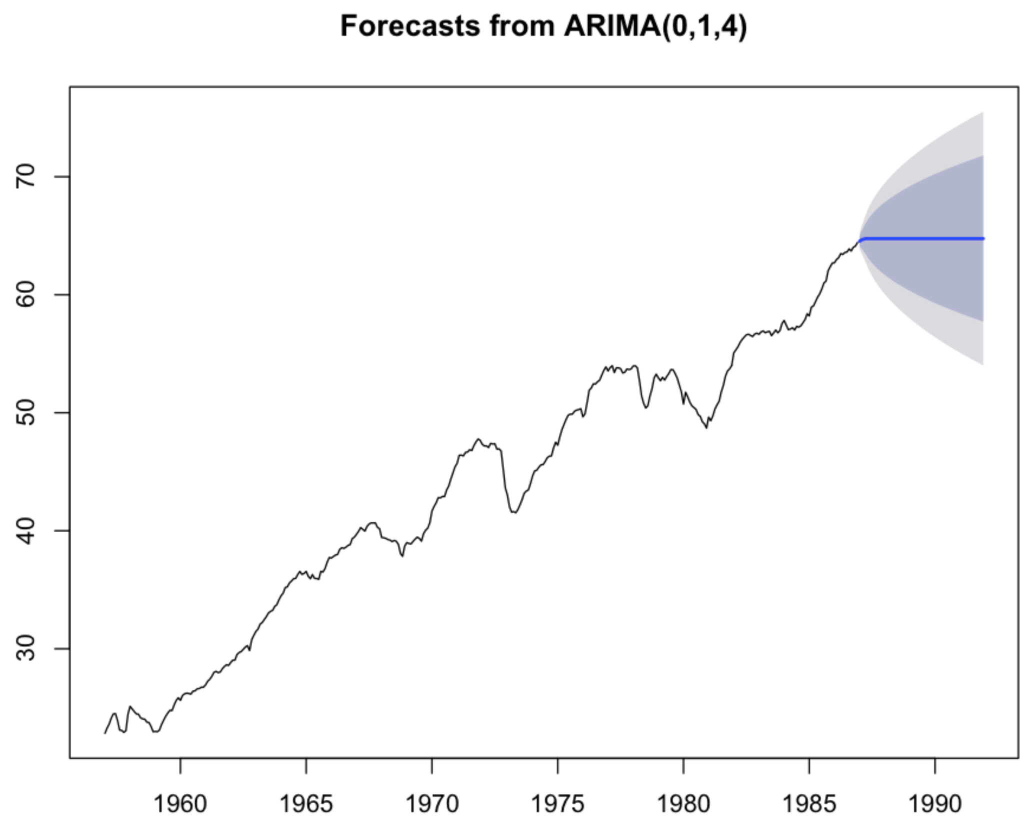
1. ARIMA(0, 1, 0) Критерий Акаики для данной модели = 346.9324



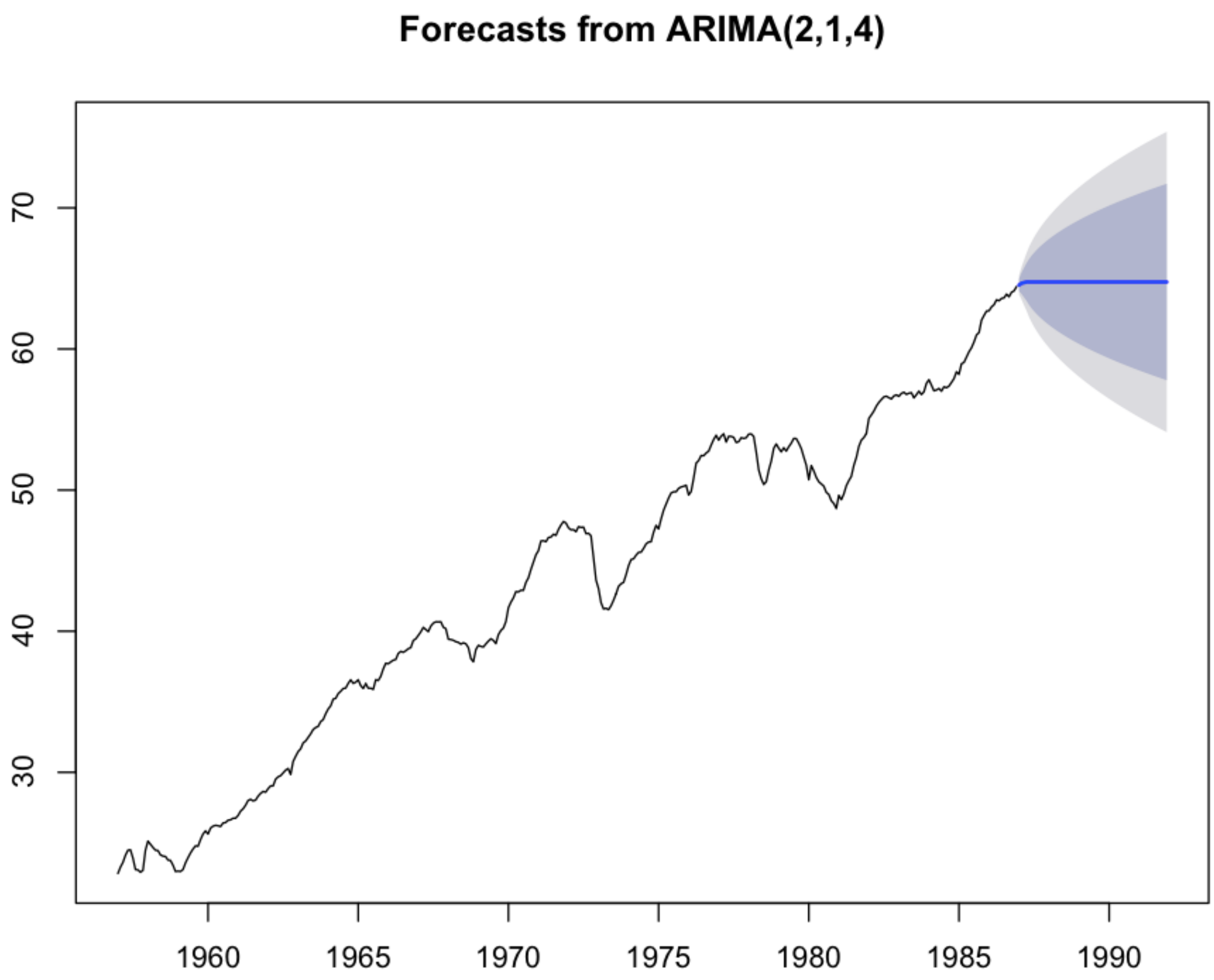
2. ARIMA(, 1, 0) Критерий Акаики для данной модели = 258.5099



3. ARIMA(0, 1, 4) Критерий Акаики для данной модели = 253.3399



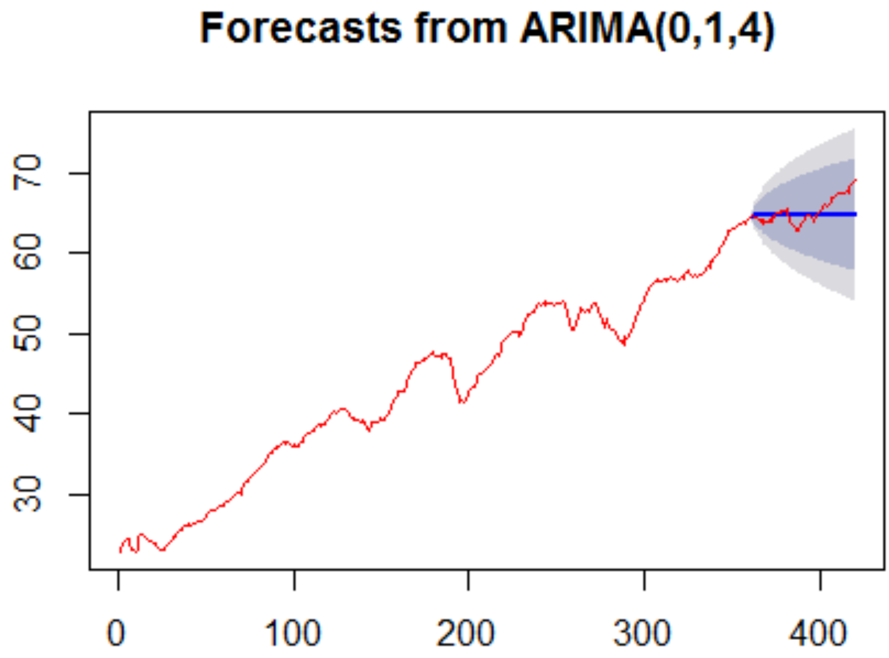
4. ARIMA(2,1,4) Критерий Акаики для данной модели = 257.3009



С помощью функции автокорреляции и функции частичной автокорреляциии были построены несколько моделей и посчитаны параметры с помощью функции R2_Score. Для определения какая из них наилучшим образом делает прогноз, воспользуемся критерием Акаики.

Наименьшему критерию соответствуем модель ARIMA(0,1,4), в данном случае AIC = 253.3399.

Построим график прогноза и testing.csv:



Вывод: Были подобраны модели для прогнозирования с помощью ACF и PACF, и отобрана наилучшая с помощью критерия Акаики, однако по итоговому графику нельзя сказать, что прогноз близок к реальным.

Итог:
Запорожец: код
Соболькова: анализ
Агаджанян: отчет + графики