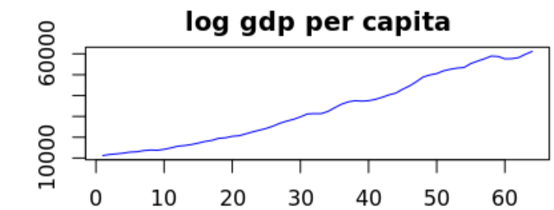
Отчёт №1 студента БЭК195 Чесноковой Елены

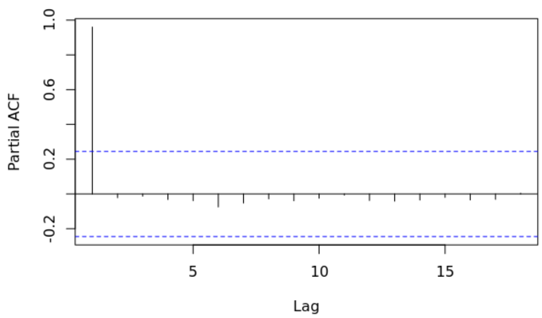
Построение прогноза для реального ВВП на душу населения в Норвегии

1. Анализ стационарности

ВВП на душу населения измеряется в годах, и в данной ситуации сезонность не влияет. Логарифмируем показатель, так как он не может быть отрицательным и может быть в виде функции экспоненты.



Проверяем показатель на стационарность. Проверим pacf.



Из графика следует, что у нас один значимый лаг. Делаем тест Дики-Фуллера.

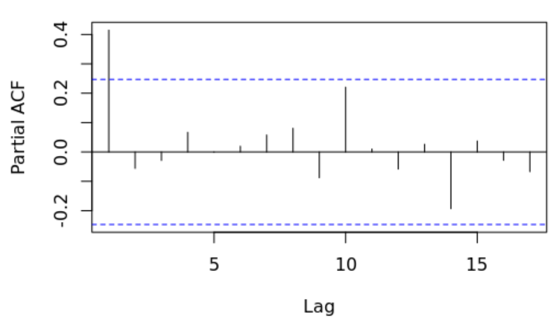
```
Value of test-statistic is: 0.806 6.3635

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

Мы получаем, что значение теста 0.806, что больше критического значения, поэтому мы не отвергаем гипотезу и нестационарности, соответственно ВВП на душу населения в Норвегии – нестационарный показатель.

Построим первую разность:

```
Value of test-statistic is: -4.3529 9.4751

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

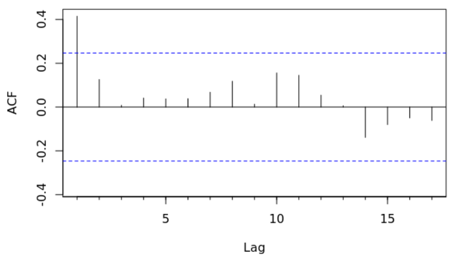
Полученный показатель больше критического значение и поэтому мы отвергаем гипотезу о нестационарности метрики. Тогда отсюда возьмём значимые лаги для построения моделей. Для этого смотрим PACF (для модели AR) и можем выделить один значимый лаг – 1.



2. Анализ различий между спецификациями моделей AR, MA и ARMA

Построение модели MA

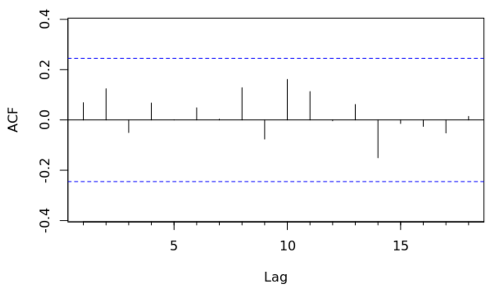
Используем ACF для взятого показателя. На графике видим, что значим лаг 1. Таким образом строим модель для него и делаем это с помощью функции ARIMA.




```
ma_gdp_1lag= Arima(data$log_gdp, c(0,1,1), include.constant =TRUE,
method = c("CSS-ML"))
coeftest(ma_gdp_1lag)
Acf(residuals(ma_gdp_1lag))
Box.test(residuals(ma_gdp_1lag), lag = 3, type = c("Ljung-Box"), fitdf
= 1)
```

```
z test of coefficients:

       Estimate Std. Error z value  Pr(>|z|)
ma1   0.3779656  0.1007032  3.7533 0.0001745 ***
drift 0.0269764  0.0026795 10.0678 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Далее анализируем остатки на автокорреляцию с помощью теста Льюнга-Бокса




```
        Box-Ljung test

data:  residuals(ma_gdp_1lag)
X-squared = 1.5377, df = 2, p-value = 0.4636
```

Остатки не автокоррелированы, из чего следует, что модель нам подходит (вероятность ошибки утверждать, что автокорреляция есть, достаточная).

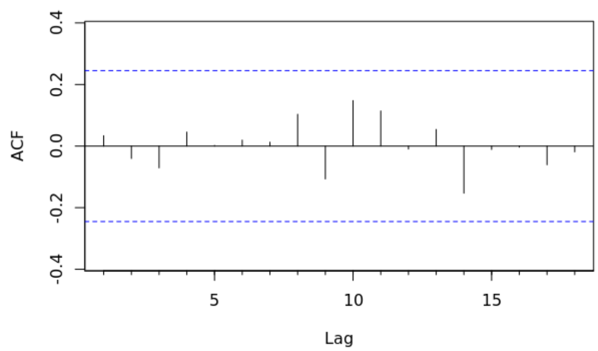
Построение модели AR

Ранее мы использовали PACF и выявили, что значим только лаг 1. Модель также построим через функцию ARIMA.

```
ar1_gdp= Arima(data$log_gdp, c(1,1,0), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(ar1_gdp)
Acf(residuals(ar1_gdp))
Box.test(residuals(ar1_gdp), lag = 3, type = c("Ljung-Box"), fitdf = 1)
```

```
z test of coefficients:

       Estimate Std. Error z value  Pr(>|z|)
ar1   0.4149632  0.1140471  3.6385 0.0002742 ***
drift 0.0270503  0.0032705  8.2710 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Проверяем остатки на автокорреляцию




```
        Box-Ljung test

data:  residuals(ar1_gdp)
X-squared = 0.53277, df = 2, p-value = 0.7661
```

Исходя из теста вероятность ошибки высокая, соответственно, из этих данных делаем вывод, что модель нас устраивает, так как в остатках нет автокорреляции (это видно на графике и ещё более точно это показывает тест).

Построение модели ARMA

Для построения этой модели используем EACF:

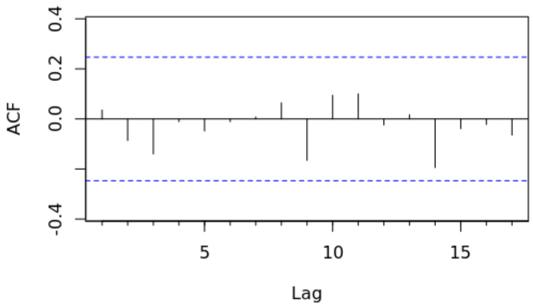
```
eacf(dif_gdp)
```

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 o x o o o o o o o o o  o  o  o
5 o o x o o o o o o o o  o  o  o
6 x o o o o o o o o o o  o  o  o
7 x o o x o o o o o o o  o  o  o
```

Для нас доступно много вариантов построения данной модели, однако мы минимизируем количество лагов, и поэтому нам подходит модель ARMA(1,1)

Строим модель с помощью функции ARIMA и проверяем остатки на автокорреляцию:

```
arma_gdp <- Arima(dif_gdp, c(1,1,1), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(arma_gdp)
Acf(residuals(arma_gdp))
Box.test(residuals(arma_gdp), lag = 3, type = c("Ljung-Box"), fitdf =
2)
```



На графике автокорреляция не наблюдается

```
        Box-Ljung test

data:  residuals(arma_gdp)
X-squared = 1.9178, df = 1, p-value = 0.1661
```

$p > 0.1$, поэтому утверждаем об отсутствии автокорреляции в остатках.

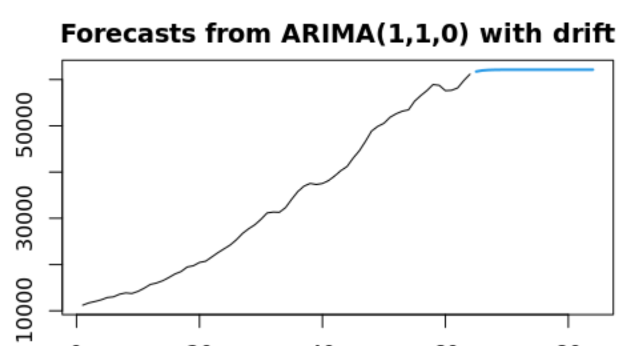
Теперь выберем наиболее подходящую для прогноза модель из выведенных. На данный момент у нас их три: MA(1), AR(1) и ARMA(1,1)
Для того, чтобы сравнить модели, используем метрики AIC И BIC и посмотрим, у кого этот параметр меньше.

```
ma_gdp_1lag$aic [1] -340.2212
ar1_gdp$aic     [1] -341.3407
arma_gdp$aic    [1] -333.7768
```

```
ma_gdp_1lag$bic  [1] -333.7918
ar1_gdp$bic      [1] -334.9113
arma_gdp$bic     [1] -325.2683
```

Отсюда видим, что самое маленькое значение метрик у модели AR(1), также у неё самое большое p-value из данных. Тогда для прогноза ВВП на душу населения будем использовать эту модель. Делаем прогноз на 20 лет, так как это примерно треть от периода, который был дан (n = 64).

3. Прогноз



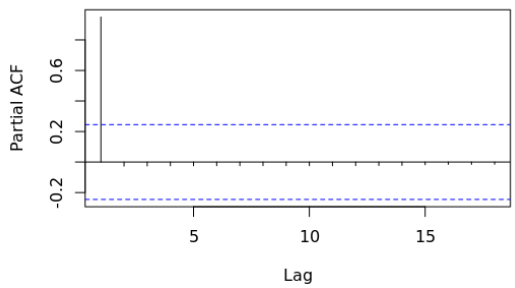
Построение прогноза для населения Норвегии

1. Анализ стационарности

Данный показатель измеряется в годах и не подвержен инфляции. В данной ситуации не буду брать логарифм.

Посмотрим pacf.

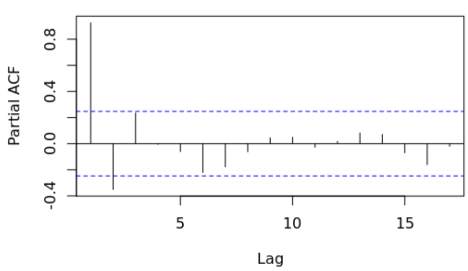
```
Value of test-statistic is: -1.5003 1.645

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```



Значимый лаг – 1

Сделаем тоже самое для первой разницы:

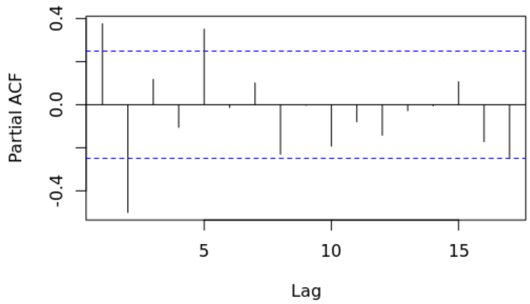
```
Value of test-statistic is: -2.1109 2.4482

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```



Значимые лаги – 1 и 2.

Сделаем для второй разницы:

```
Value of test-statistic is: -7.3698 27.1605

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

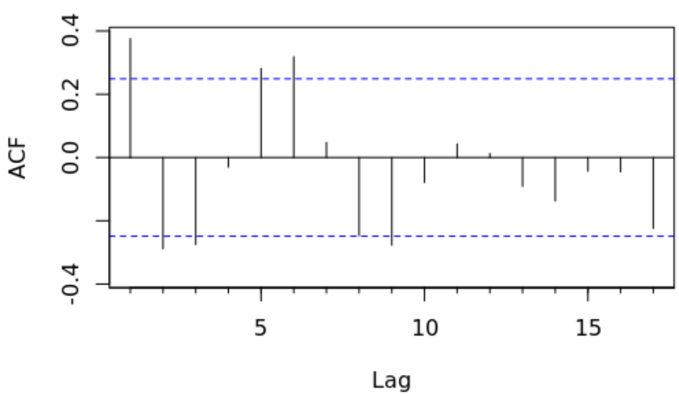


Значимые лаги – 1,2 и 5
Соответственно, гипотеза о нестационарности числа населения и первой разности не отклоняется. Показатели - нестационарные. Для второй разности ситуация обратная – отвергаем нестационарность.
Будет использовать в дальнейшем ряд второй разности числа населения..

2. Анализ различий между спецификациями моделей AR, MA и ARMA

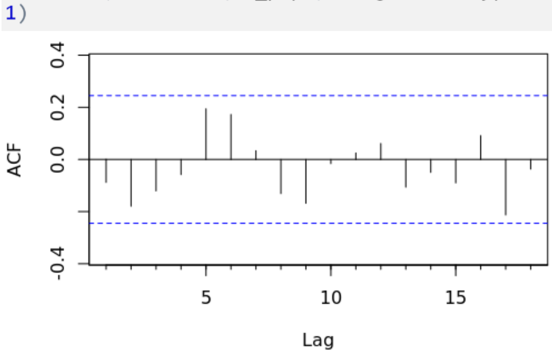
Построение модели MA

Посмотрим на ACF. Значимые лаги исходя из графика – 1,2,3,5,6,9.



Построим модель MA(1) с помощью функции ARIMA

```
ma_pop1= Arima(data$Pop, c(0,2,1), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(ma_pop1)
Acf(residuals(ma_pop1))
Box.test(residuals(ma_pop1), lag = 4 , type = c("Ljung-Box"), fitdf =
1)
```



```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ma1 0.844189   0.059631  14.157 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Проверим остатки.

```
         Box-Ljung test

data:  residuals(ma_pop1)
X-squared = 3.9527, df = 3, p-value = 0.2666
```

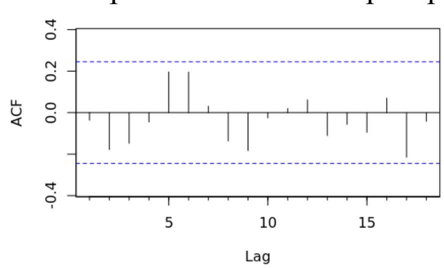
Исходя из результатов, видим, что значение достаточное, чтобы говорить об отсутствии автокорреляции в остатках.

Тогда строим MA(2):

```
ma_pop2= Arima(data$Pop, c(0,2,2), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(ma_pop2)
Acf(residuals(ma_pop2))
Box.test(residuals(ma_pop2), lag = 4 , type = c("Ljung-Box"), fitdf =
2)
```

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1  0.770258   0.162130  4.7509 2.025e-06 ***
ma2 -0.076859   0.158166 -0.4859     0.627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Посмотрим на остатки и проверим на автокорреляцию:



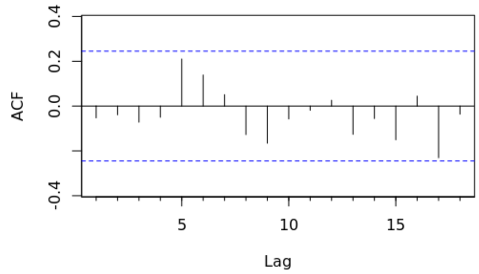

```
         Box-Ljung test

data:  residuals(ma_pop2)
X-squared = 3.8834, df = 2, p-value = 0.1435
```

Из этого следует, что автокорреляция отсутствует, а значит нам подходит данная модель.

Ещё построим MA(3):

```
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ma1  0.76432    0.12463  6.1327 8.64e-10 ***
ma2 -0.24555    0.14662 -1.6747  0.09399 .
ma3 -0.17348    0.11062 -1.5683  0.11681
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Проверим остатки на автокорреляцию:




```
        Box-Ljung test

data:  residuals(ma_pop3)
X-squared = 0.82176, df = 1, p-value = 0.3647
```

Мы видим, p-value маленькое (p >0.1), поэтому эта модель нам подходит.

Строим MA(5)

```
 z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)    
ma1  0.728736   0.203238  3.5856 0.0003363 ***
ma2 -0.276144   0.271535 -1.0170 0.3091647    
ma3 -0.193261   0.125327 -1.5420 0.1230623    
ma4  0.047629   0.345754  0.1378 0.8904351    
ma5  0.075529   0.269981  0.2798 0.7796632    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Проверяем остатки

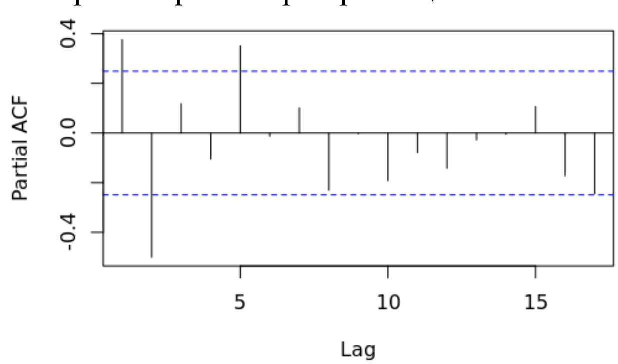
```
        Box-Ljung test

data:  residuals(ma_pop5)
X-squared = 3.6427, df = 0, p-value < 2.2e-16
```

Остатки данной модели сильно автокоррелированы.

Если строить модели по остальным значимым лагам, то получится, что p-value у них очень маленькое, что говорит о присутствии автокорреляции в остатках. Это происходит, потому что все последующие fitdf >= lag.

Построение модели AR

Смотрим на pacf второй разницы числа населения:



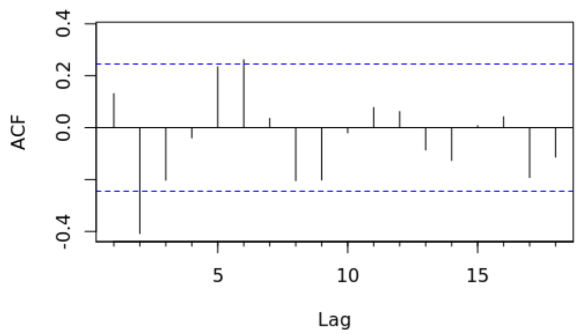
Значимые лаги – 1, 2 и 5.

Строим модель AR(1) с помощью функции ARIMA:

```
ar_pop1= Arima(data$Pop, c(1,2,0), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(ar_pop1)
Acf(residuals(ar_pop1))
Box.test(residuals(ar_pop1), lag = 4, type = c("Ljung-Box"), fitdf = 1)
```

```
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ar1  0.38007    0.11614  3.2726 0.001066 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Далее проверим остатки модели на автокорреляцию:




```
        Box-Ljung test

data:  residuals(ar_pop1)
X-squared = 15.437, df = 3, p-value = 0.001479
```

Из результатов следует наличие автокорреляции, поэтому можем сказать, что модель нам не подходит.

Далее строим AR(2):

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1  0.56552    0.11000  5.1412 2.730e-07 ***
ar2 -0.47828    0.10835 -4.4142 1.014e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

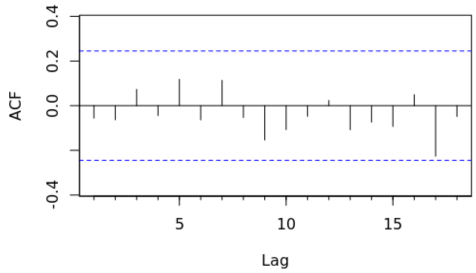
Проверим остатки:

```
        Box-Ljung test

data:  residuals(ar_pop2)
X-squared = 4.4116, df = 2, p-value = 0.1102
```

Здесь нет автокорреляция в остатках, хоть значение и не большое.

Наконец строим AR(5) и проверяем остатки.

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1  0.67321    0.11780  5.7150 1.097e-08 ***
ar2 -0.66896    0.13945 -4.7970 1.610e-06 ***
ar3  0.39700    0.15455  2.5688  0.010205 *
ar4 -0.30650    0.13652 -2.2451  0.024761 *
ar5  0.33884    0.11347  2.9862  0.002824 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
        Box-Ljung test

data:  residuals(ar_pop5)
X-squared = 1.9812, df = 0, p-value < 2.2e-16
```

Остатки модели сильно автокоррелированы.

Построение ARMA

Строим EACF:

```
eacf(dif2_pop)
```

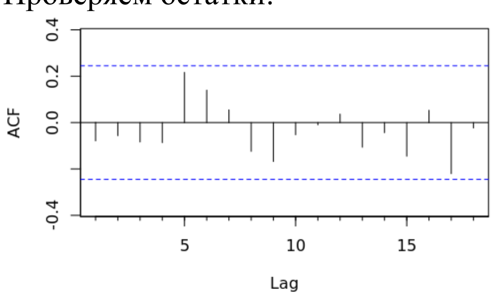
```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x o x x o o x o o  o  o  o
1 x x x o o x o o o o o  o  o  o
2 o o x o o o o o o o o  o  o  o
3 x o x o o o o o o o o  o  o  o
4 x o x o o o o o o o o  o  o  o
5 o x x o o o o o o o o  o  o  o
6 o o x o o o o o o o o  o  o  o
7 x o x o o o o o o o o  o  o  o
```

Построим модель ARMA(2,1):

```
arma_pop <- Arima(data$Pop, c(2,2,1), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(arma_pop)
Acf(residuals(arma_pop))
Box.test(residuals(arma_pop), lag = 4, type = c("Ljung-Box"), fitdf =
2)
```

```
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -0.0096604  0.1458810 -0.0662   0.9472
ar2 -0.1988426  0.1370955 -1.4504   0.1469
ma1  0.8022441  0.0948606  8.4571   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Проверяем остатки:



```
        Box-Ljung test

data:  residuals(arma_pop)
X-squared = 1.6418, df = 2, p-value = 0.44
```

Соответственно, в остатках нет автокорреляции, и мы можем использовать данную модель ($p > 0.1$).
Исходя из рассмотренных моделях на данный момент у нас есть ARMA(2,1), MA(1), MA(2), MA(3)и AR(2).

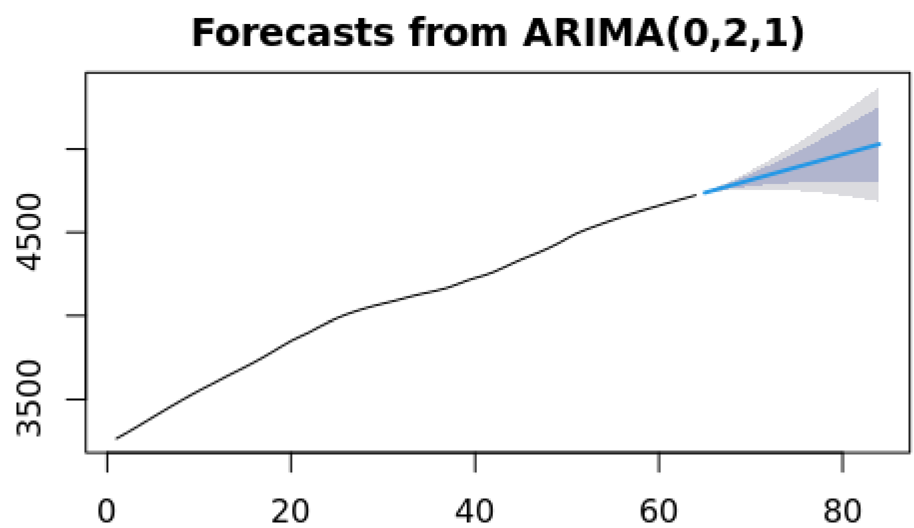
Сравниваем модели по AIC и BIC:

```
arma_pop$aic     [1] 251.3187
ma_pop1$aic      [1] 249.4979
ma_pop2$aic      [1] 251.254
ma_pop3$aic      [1] 250.9178
ar_pop2$aic      [1] 255.2798
```

Исходя из полученных результатов, лучше всего подойдёт модель MA(1), так как для неё AIC и BIC наименьшие. В своём прогнозе я буду использовать эту модель как лучшую.

```
arma_pop$bic     [1] 259.8273
ma_pop1$bic      [1] 253.7521
ma_pop2$bic      [1] 257.6354
ma_pop3$bic      [1] 259.4263
ar_pop2$bic      [1] 261.6612
```

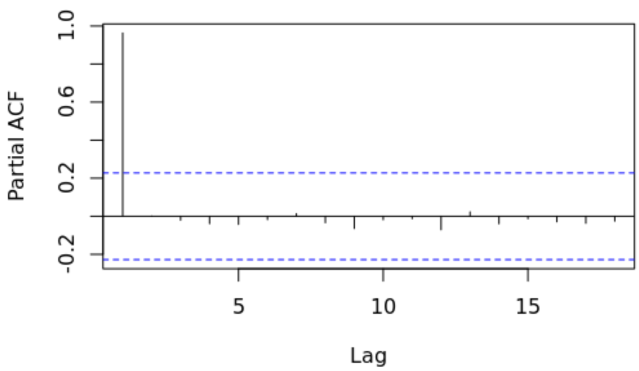
3. Прогноз



Построение прогноза средней ЗП в Республике Карелия

1. Анализ стационарности

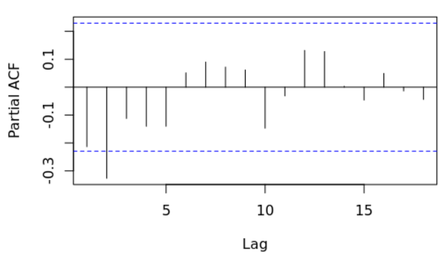
Будем использовать логарифм, так как зарплата не может быть отрицательной. Посмотрим PACF и сделаем DF тест. Заработная плата уже скорректирована на инфляцию и сезонность.




```
Value of test-statistic is: -0.4463 20.335

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

Не отвергаем гипотезу о нестационарности.

Теперь создаём первую разницу для этого показателя и также смотрим PACF и делаем тест на стационарность.




```
Value of test-statistic is: -9.0422 40.8916

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

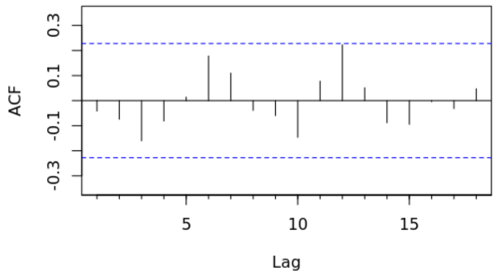
Т-статистика показывает маленькое значение, поэтому отвергаем гипотезу и нестационарности показателя. Для построения моделей возьмём первую разница заработной платы. На графике видим, что у нас один значимый лаг – 2.

2. Анализ различий между спецификациями моделей AR, MA и ARMA

Построение AR модели

Значимый лаг – 2. Строим модель AR(2) с помощью функции ARIMA, после чего проверяем остатки на автокорреляцию.

```
z test of coefficients:

        Estimate Std. Error z value  Pr(>|z|)
ar1   -0.2828972  0.1109442 -2.5499  0.010775 *
ar2   -0.3242050  0.1099050 -2.9499  0.003179 **
drift  0.0180585  0.0015886 11.3675 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

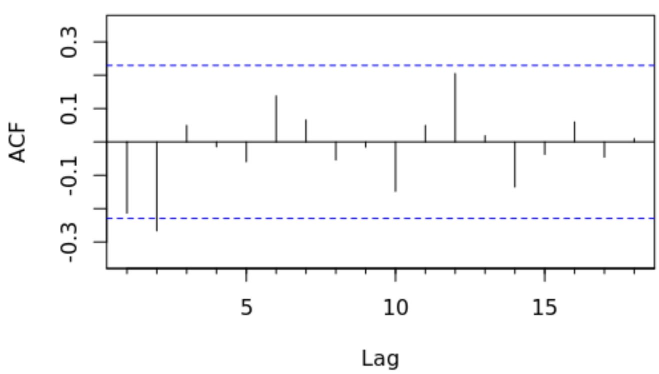



```
        Box-Ljung test

data:  residuals(ar_w)
X-squared = 3.1272, df = 2, p-value = 0.2094
```

P-value в тесте дало значение больше 0.1, поэтому мы говорим здесь об отсутствии автокорреляции в остатках.

Построение MA модели

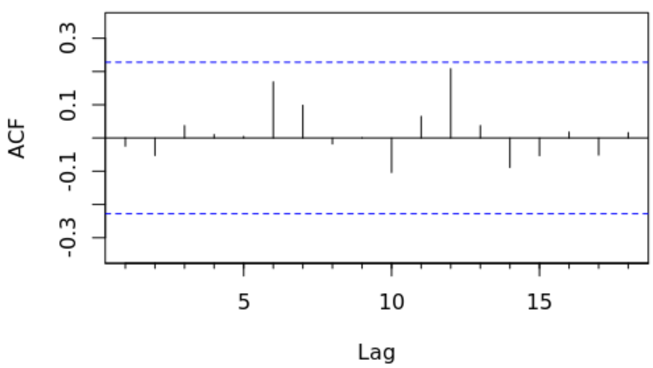
Смотрим на ACF для того, чтобы обозначить лаги, по которым строить модель. Значимый лаг – 2.



Строим модель MA(2) функцией ARIMA и проверяем остатки на автокорреляцию:

```
ma_w= Arima(wag$wage, c(0,1,2), include.constant =TRUE, method =
c("CSS-ML"))
coeftest(ma_w)
Acf(residuals(ma_w))
Box.test(residuals(ma_w), lag = 4 , type = c("Ljung-Box"), fitdf = 1)
```

```
z test of coefficients:

        Estimate Std. Error z value Pr(>|z|)
ma1   -0.3234205  0.1173486 -2.7561  0.00585 **
ma2   -0.2563996  0.1124128 -2.2809  0.02256 *
drift  0.0182707  0.0010891 16.7760  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
        Box-Ljung test

data:  residuals(ma_w)
X-squared = 0.38133, df = 3, p-value = 0.9441
```

Тест показал, что в остатках нет автокорреляции.

Построение ARMA

```
eacf(d1)
```

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o x o o o o o o o o o  o  o  o
1 x x o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x o x x o o o o o o o  o  o  o
5 x x x o o o o o o o o  o  o  o
6 x o o x o o o o o o o  o  o  o
7 x o o x o o o o o o o  o  o  o
```

Модель MA(2) я уже построила, поэтому сейчас я построю модель ARMA(1,2)

```
z test of coefficients:

        Estimate Std. Error z value Pr(>|z|)
ar1   -0.2062911  0.3422934 -0.6027  0.54673
ma1   -0.1374007  0.3149351 -0.4363  0.66263
ma2   -0.3420614  0.1594711 -2.1450  0.03195 *
drift  0.0182356  0.0011112 16.4108  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Проверим остатки:

```
        Box-Ljung test

data:  residuals(arma_w)
X-squared = 0.10734, df = 2, p-value = 0.9477
```

Из этого следует, что остатки не автокоррелированы.

Итого мы получаем три модели без автокорреляции в остатках:

MA(2), AR(2), ARMA(1,2)

Сравниваем модели:

```
arma_w$aic    [1] -344.9048
ma_w$aic      [1] -346.5574
ar_w$aic      [1] -344.6216
```

```
arma_w$bic    [1] -333.4525
ma_w$bic      [1] -337.3956
ar_w$bic      [1] -335.4598
```

Мы получили, что наименьшие значение данных метрик имеет MA(1), также p-value больше, чем у модели AR(2). Для прогноза я буду использовать модель MA(1).

3. Прогноз