ARFIMA specifications for UK inflation series

Daniele Melotti

In this paper I search for the best ARFIMA specification for UK inflation series, in a range of time between the 1st quarter of 1965 and the 3rd quarter of 2019. The research is deployed below, while the summary and the forecasts prepared with few of the researched models are located at the end of the paper.

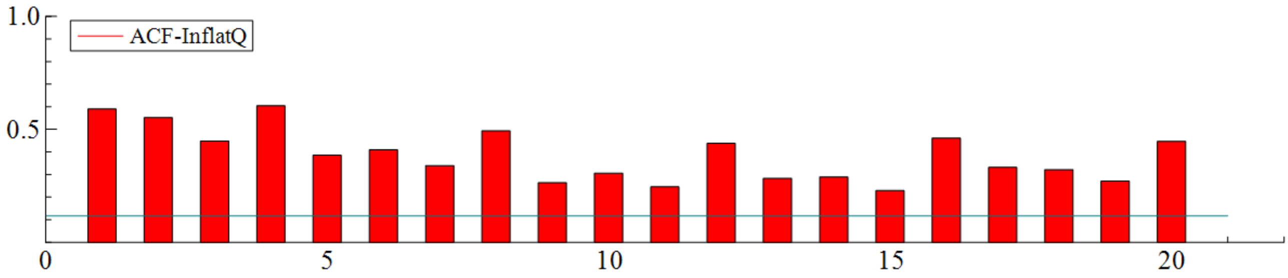
Research

The first thing to do is to produce the descriptive statistics using G@RCH in the period between 1965(1) and 2019(3). I chose ARCH test, ADF unit root test and Geweke and Porter-Hudak long memory test:

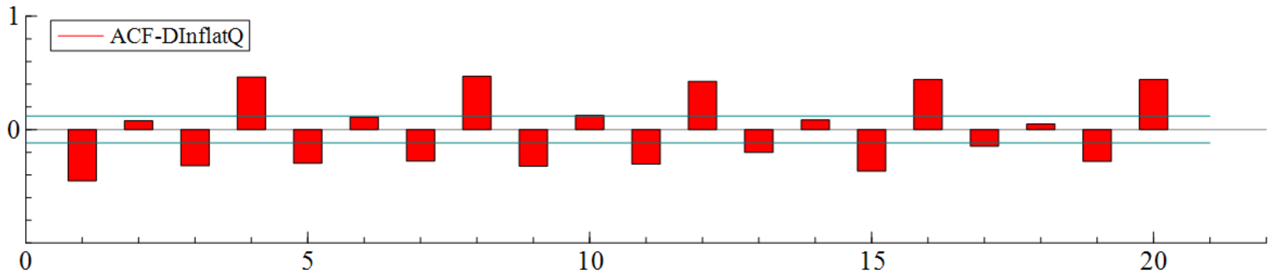
```
Series #1/1: InflatQ
---------
ARCH 1-2 test:    F(2,214)  =   52.587 [0.0000]**
ARCH 1-5 test:    F(5,208)  =   26.894 [0.0000]**
ARCH 1-10 test:   F(10,198) =   13.493 [0.0000]**
---------------
---------------------------------------------

ADF Test with 4 lags
No intercept and no time trend
H0: InflatQ is I(1)

ADF Statistics: -1.92541

Asymptotic critical values, Davidson, R. and MacKinnon, J. (1993)

        1%        5%       10%
  -2.56572  -1.94093  -1.61663

OLS Results
                  Coefficient    t-value
y_1                 -0.063909    -1.9254
dy_1                -0.394697    -5.5240
dy_2                -0.251906    -3.5529
dy_3                -0.353497    -5.0863
dy_4                 0.266034     3.9923
RSS               2493.747187
OBS                214.000000
Information Criteria (to be minimized)
Akaike          5.340449  Shibata         5.339390
Schwarz         5.419093  Hannan-Quinn    5.372228
---------------

----  Log Periodogram Regression ----
d parameter          0.581347 (0.0739716) [0.0000]
No of observations: 219; no of periodogram points: 109
```

A lot of ARCH effect is present in the InflatQ series, which according to ADF test is non-stationary at 1% and 5% significance. Geweke's test showed a very big d parameter, which is strongly significant and indicates long memory.

Plotting the ACF graph for the series:



The function decreases very slowly, but a sort of pattern is visible; every 4 lags (4th, 8th, 12th, 16th, 20th) there is a higher lag if compared to the neighbouring ones. This might signify a sort of autocorrelation in particular seasons.
And the ACF of the first differences of the InflatQ series follows:



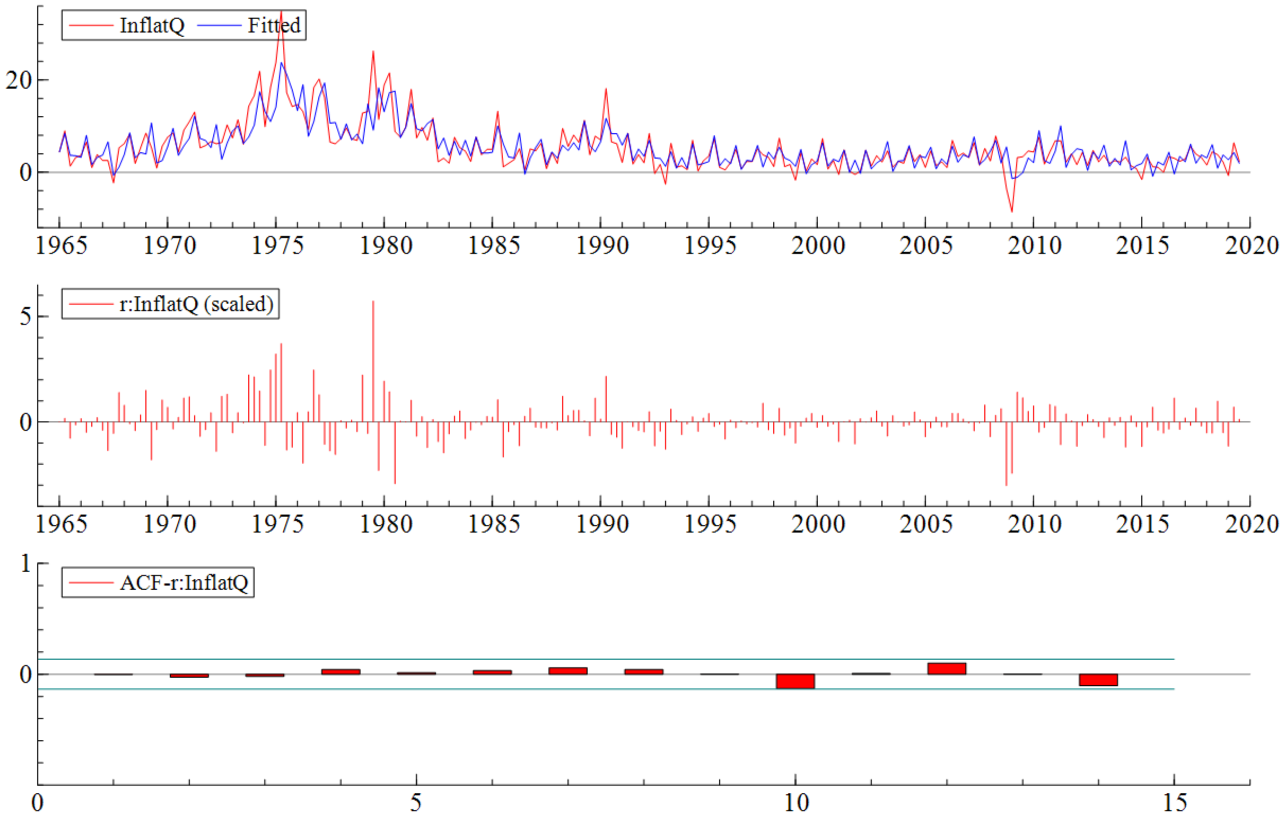
Visibly, the situation is the same as in the first graph. There are high peaks of significant autocorrelation every 4th lag. It seems like there is seasonal autocorrelation.

The first model is constructed in the same way as the first model created during the class, so it is an **ARMA(4,4)** from ARFIMA models using PcGive. It embeds the inflation series, a constant and 3 seasonals:

```
---- Maximum likelihood estimation of ARFIMA(4,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                    Coefficient  Std.Error  t-value  t-prob
d parameter           -0.458929     0.1222    -3.76   0.000
AR-1                  -0.392471     0.1693    -2.32   0.021
AR-2                 -0.0104267     0.1225  -0.0851   0.932
AR-3                   0.671042    0.09158     7.33   0.000
AR-4                   0.674103     0.1567     4.30   0.000
MA-1                    1.46958     0.2146     6.85   0.000
MA-2                    1.53986     0.3820     4.03   0.000
MA-3                   0.745838     0.3692     2.02   0.045
MA-4                   0.102181     0.1727    0.592   0.555
Constant                4.80065      1.361     3.53   0.001
Seasonal              -0.388123     0.5294   -0.733   0.464
Seasonal_1              3.62544     0.5134     7.06   0.000
Seasonal_2             -1.37505     0.5294    -2.60   0.010

log-likelihood    -551.952804
no. of observations       219  no. of parameters          14
AIC.T              1131.90561  AIC                5.16851876
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
sigma                 2.99643  sigma^2               8.97861

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
```

```
Used starting values:
      0.40000       0.19245      0.046819     -0.073668       0.21388      0.018831     -0.015270
-0.021413      0.018341        5.0443      -0.39655       3.5899      -1.3986
```

The AR(2) is not significant, as well as MA(4) and Seasonal. MA(3) is on the border. The d parameter is strongly significant and negative. The Constant is significant, as well as the last two Seasonals, indicating that there really is seasonality within the series.
A graphic analysis and a test summary should be done:



The fit of the model is quite good, but it seems that the dummy variables which are present in the database could be used as there are some outliers; perhaps this will give an even better fit to the model.
The test summary follows:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)   =    57.772 [0.0000]**
ARCH 1-1 test:    F(1,204)   =    13.302 [0.0003]**
Portmanteau(14):  Chi^2(5)   =    10.727 [0.0571]
```

The residuals are definitely non-normally distributed and there is a lot of ARCH effect. There is no autocorrelation (barely) within 14 lags.

The next model tries to fix the non-significance of AR(2) and MA(4). So, it is an **ARMA(4,3) with AR(2) fixed**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter          0.348534     0.1809     1.93   0.055
```

```
AR-1                -0.136595     0.4329   -0.316   0.753
AR-3                 0.225552     0.2751    0.820   0.413
AR-4                 0.188107    0.08721     2.16   0.032
MA-1                 0.397315     0.3706     1.07   0.285
MA-2                 0.153466     0.1673    0.918   0.360
MA-3                -0.237330     0.2574   -0.922   0.358
Constant              4.78566      2.888     1.66   0.099
Seasonal            -0.391395     0.5594   -0.700   0.485
Seasonal_1            3.63378     0.6103     5.95   0.000
Seasonal_2           -1.36508     0.5594    -2.44   0.016

log-likelihood    -554.641924
no. of observations       219  no. of parameters          12
AIC.T              1133.28385  AIC                5.17481209
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
sigma                 3.03206  sigma^2                9.1934

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.20764    -0.059906      0.21782    0.0032588     0.028682    -0.029412
5.0443     -0.39655       3.5899      -1.3986
```

This is definitely not a good model. Now there are many more non-significant parameters. Moreover, log-likelihood has become lower and AIC has increased. And looking at the test summary:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   72.477 [0.0000]**
ARCH 1-1 test:    F(1,206)  =   11.420 [0.0009]**
Portmanteau(14):  Chi^2(7)  =   5.5786 [0.5897]
```

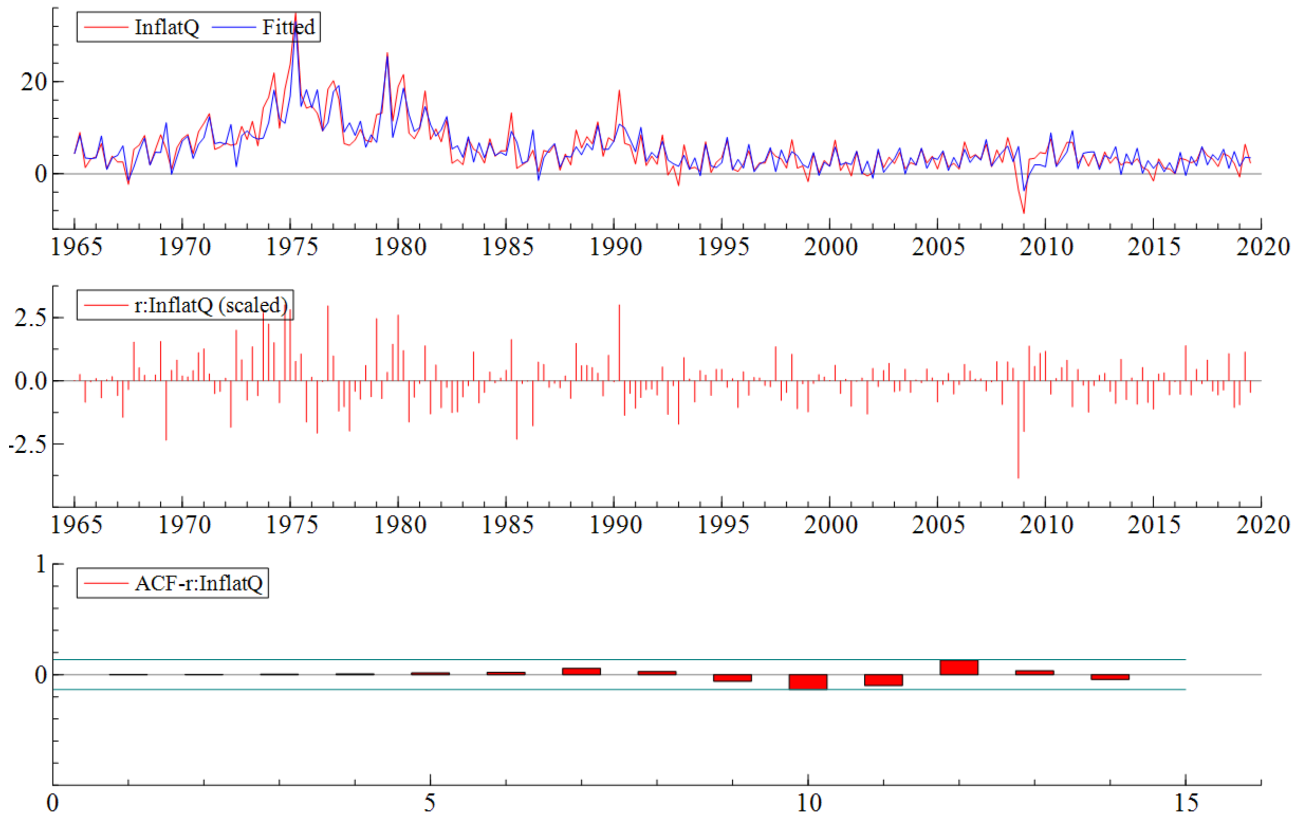
There has been no benefit apart from securing the removal of autocorrelation within 14 lags, which is not enough of a satisfactory result.

Therefore, the 3rd model sees a step backward to **ARMA(4,4), with the inclusion of two dummy variables: D75Q2 and D79Q3**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter         -0.218663     0.2805   -0.780   0.437
AR-1               -0.0922048      1.769  -0.0521   0.958
AR-2                 0.377209     0.9030    0.418   0.677
AR-3                 0.303425     0.5648    0.537   0.592
AR-4                 0.329805     0.2939     1.12   0.263
MA-1                  1.10253      1.696    0.650   0.516
MA-2                 0.476096     0.8245    0.577   0.564
MA-3                -0.133869     0.4685   -0.286   0.775
MA-4                -0.133734     0.8326   -0.161   0.873
Constant              4.75444      1.409     3.37   0.001
Seasonal            -0.384685     0.5025   -0.766   0.445
Seasonal_1            3.49670     0.5848     5.98   0.000
Seasonal_2           -1.68621     0.5035    -3.35   0.001
D75Q2                 8.46959      1.728     4.90   0.000
D79Q3                 17.9919      1.728     10.4   0.000

log-likelihood     -508.57454
no. of observations       219  no. of parameters          16
AIC.T              1049.14908  AIC                4.79063507
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
```

```
sigma                  2.45699   sigma^2              6.03682

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.066267      0.041895      0.17883      0.64354      0.086171     -0.066420
-0.14092     -0.31452       5.0443     -0.39655       3.1023      -1.8185        26.819
23.095
```

This model brings a huge increase in log-likelihood and an important decrease in the value of AIC, which is really nice. However, lots of parameters are not statistically significant, including all the AR and MA lags and the d parameter! This suggests further work is to be done in the settings of the model.
Looking at the graphic analysis:



The fit has improved especially where the dummy variables have acted (the 2nd quarter of 1975 and the 3rd quarter of 1979). However, the residuals don't look independent and identically distributed, and checking the test summary:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)   =   17.536 [0.0002]**
ARCH 1-1 test:    F(1,202)   =   7.9259 [0.0054]**
Portmanteau(14):  Chi^2(5)   =   13.028 [0.0231]*
```

Now even some statistically significant autocorrelation is present, which is one more reason why this model cannot be considered good.
The first seasonal parameter has always been not significant so far, independently from the specification. So, a test for linear restrictions can be used to check if the removal of this parameter will damage the efficiency of the models (even though it seems very unlikely as for now):

```
Test for linear restrictions (Rb=r):
R matrix
  d parameter        AR-1         AR-2         AR-3         AR-4         MA-1         MA-2
      0.00000      0.00000      0.00000      0.00000      0.00000      0.00000      0.00000
MA-3         MA-4      Constant     Seasonal   Seasonal_1   Seasonal_2        D75Q2        D79Q3
0.00000      0.00000      0.00000       1.0000      0.00000      0.00000      0.00000      0.00000
r vector
      0.00000
LinRes Chi^2(1) = 0.586028 [0.4440]
```

Looking at the p-value there is no rejection of the H_0. Namely, if this parameter is removed from the model, nothing should happen in terms of results.

So, the next model will **not include the first Seasonal**. It is an **ARMA(4,3)**, and **instead of the two dummy variables used separately, there is the addition of D75Q2+D79Q3**, which embeds both outliers together:

```
---- Maximum likelihood estimation of ARFIMA(4,d,3) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter         -0.280965     0.1708    -1.65   0.101
AR-1                 0.136479     0.2507    0.544   0.587
AR-2                 0.202565     0.1651     1.23   0.221
AR-3                 0.277550     0.1542     1.80   0.073
AR-4                 0.320748     0.2663     1.20   0.230
MA-1                 0.846373     0.3747     2.26   0.025
MA-2                 0.469322     0.4474     1.05   0.295
MA-3               0.00111957     0.3689  0.00304   0.998
Constant              4.56779      1.340     3.41   0.001
Seasonal_1            3.66557     0.4805     7.63   0.000
Seasonal_2           -1.47952     0.4823    -3.07   0.002
D75Q2+79Q3            13.5296      1.351     10.0   0.000

log-likelihood    -514.624236
no. of observations       219  no. of parameters          13
AIC.T              1055.24847  AIC                4.81848617
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
sigma                 2.52696  sigma^2               6.38553

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.34165     -0.12357      0.27887      0.19011     -0.18824     0.029455
-0.21495       4.8442       3.3362      -1.6522       24.957
```

Unfortunately, there are still many parameters that are not statistically significant, including the d parameter. The values of log-likelihood and AIC have worsened too. And looking at the test summary:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   14.553 [0.0007]**
ARCH 1-1 test:    F(1,205)  =   12.780 [0.0004]**
Portmanteau(14):  Chi^2(6)  =   11.389 [0.0771]
```

The autocorrelation has been removed. However, non-normality of the residuals and ARCH effect seem to be constantly present. ARMA specifications are not enough to remove these issues, GARCH specifications would be necessary.

The next model is an **ARMA(4,2), still including the D75Q2+D79Q3 dummy**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,2) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter         -0.280636     0.1274    -2.20   0.029
AR-1                 0.137047     0.1638    0.837   0.404
AR-2                 0.202761     0.1543     1.31   0.190
AR-3                 0.277538     0.1540     1.80   0.073
AR-4                 0.319978    0.09417     3.40   0.001
MA-1                 0.845471     0.2167     3.90   0.000
MA-2                 0.468092     0.1784     2.62   0.009
Constant              4.56787      1.340     3.41   0.001
Seasonal_1            3.66552     0.4803     7.63   0.000
Seasonal_2           -1.47955     0.4822    -3.07   0.002
D75Q2+79Q3            13.5294      1.352     10.0   0.000

log-likelihood     -514.62424
no. of observations       219  no. of parameters          12
AIC.T              1053.24848  AIC               4.80935379
mean(InflatQ)         5.49505  var(InflatQ)         29.4156
sigma                 2.52696  sigma^2              6.38553

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.23186    -0.062128   -0.0010162      0.23457    -0.087189    -0.021675
4.8442      3.3362      -1.6522       24.957
```

The d parameter is significant in this model, and all the AR and MA lags are now significant apart from AR(1), AR(2) and AR(3). The Constant and the two Seasonals are still significant. Looking at the test summary:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   14.555 [0.0007]**
ARCH 1-1 test:    F(1,206)  =   12.836 [0.0004]**
Portmanteau(14):  Chi^2(7)  =   11.387 [0.1226]
```

The statistics remain stable, apart from the p-value of Portmanteau test, which has grown even higher. This model seems to be not a bad one, but the research will continue.

The 6th model is an **ARMA(4,2) which sees the re-introduction of the two dummy variables D75Q2 and D79Q3 separately**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,2) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter         -0.296570     0.1222    -2.43   0.016
AR-1                 0.130099     0.1401    0.929   0.354
AR-2                 0.233907     0.1290     1.81   0.071
AR-3                 0.249880     0.1541     1.62   0.106
AR-4                 0.326438     0.1009     3.24   0.001
MA-1                 0.955280     0.1801     5.31   0.000
MA-2                 0.507710     0.1776     2.86   0.005
Constant              4.55531      1.340     3.40   0.001
Seasonal_1            3.77325     0.4878     7.74   0.000
Seasonal_2           -1.56937     0.4899    -3.20   0.002
D75Q2                 8.57824      1.731     4.96   0.000
```

```
D79Q3                 17.8945      1.731     10.3   0.000

log-likelihood    -508.914168
no. of observations       219  no. of parameters          13
AIC.T              1043.82834  AIC               4.76633943
mean(InflatQ)         5.49505  var(InflatQ)         29.4156
sigma                 2.46074  sigma^2              6.05522

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.23908     -0.11090     0.013195      0.22387     -0.12208     0.030367
4.8442       3.3024      -1.6184       26.819       23.095
```

The d parameter is negative and statistically significant. Only AR(1), AR(2) and AR(3) are not significant in this model. The log-likelihood has increased while the value of AIC is now lower, showing that there seems to be an advantage in using the dummy variables separately. And checking the test summary:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   16.121 [0.0003]**
ARCH 1-1 test:    F(1,205)  =   7.4561 [0.0069]**
Portmanteau(14):  Chi^2(7)  =   13.213 [0.0671]
```

The issues of ARCH effect and non-normality of the residuals remain.

The next model is again an **ARMA(4,2), this time with AR(1) and AR(3) fixed**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,2) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter          0.219287     0.1438     1.52   0.129
AR-2                0.0530653     0.1913    0.277   0.782
AR-4                 0.376356    0.07689     4.89   0.000
MA-1                 0.585831     0.1575     3.72   0.000
MA-2                 0.180195     0.1048     1.72   0.087
Constant              4.65129      1.699     2.74   0.007
Seasonal_1            3.76233     0.5227     7.20   0.000
Seasonal_2           -1.57081     0.5251    -2.99   0.003
D75Q2                 8.84747      1.773     4.99   0.000
D79Q3                 18.2275      1.773     10.3   0.000

log-likelihood    -511.258748
no. of observations       219  no. of parameters          11
AIC.T                1044.5175  AIC               4.76948629
mean(InflatQ)         5.49505  var(InflatQ)         29.4156
sigma                 2.48737  sigma^2              6.18699

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000    -0.033204      0.23775      0.11666    -0.011529       4.8442       3.3024
-1.6184       26.819       23.095
```

The d parameter is not statistically significant here, as well as AR(2) and MA(2). The test summary follows:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   15.636 [0.0004]**
ARCH 1-1 test:    F(1,207)  =   7.2255 [0.0078]**
```

```
Portmanteau(14):  Chi^2(9)  =   18.402 [0.0308]*
```

The residuals are still non-normally distributed, there is still ARCH effect and there is even statistically significant autocorrelation within 14 lags. This model is a step backwards.

Then, the next model is an **ARMA(4,2) with only AR(1) fixed**:

```
---- Maximum likelihood estimation of ARFIMA(4,d,2) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter         -0.282864     0.1212    -2.33   0.021
AR-2                 0.278273     0.1124     2.48   0.014
AR-3                 0.329583     0.1184     2.78   0.006
AR-4                 0.317099     0.1040     3.05   0.003
MA-1                  1.05848     0.1227     8.63   0.000
MA-2                 0.590207     0.1374     4.30   0.000
Constant              4.58520      1.269     3.61   0.000
Seasonal_1            3.75727     0.4724     7.95   0.000
Seasonal_2           -1.54798     0.4742    -3.26   0.001
D75Q2                 8.56108      1.712     5.00   0.000
D79Q3                 17.6523      1.712     10.3   0.000

log-likelihood    -509.358755
no. of observations       219  no. of parameters          12
AIC.T              1042.71751  AIC               4.76126717
mean(InflatQ)         5.49505  var(InflatQ)         29.4156
sigma                 2.46542  sigma^2              6.07829

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000    0.0050245    -0.021055      0.24458      0.11437    -0.039293       4.8442
3.3024      -1.6184       26.819       23.095
```

All the parameters included in this model are statistically significant. The test summary follows:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   16.896 [0.0002]**
ARCH 1-1 test:    F(1,206)  =   7.3060 [0.0074]**
Portmanteau(14):  Chi^2(8)  =   15.767 [0.0458]*
```

Unfortunately, there is a barely statistically significant autocorrelation.

The next model is quite different from the previous ones, an **ARMA(0,4)**:

```
---- Maximum likelihood estimation of ARFIMA(0,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter          0.453752    0.05511     8.23   0.000
MA-1                 0.320920    0.08961     3.58   0.000
MA-2               -0.0107952    0.08447   -0.128   0.898
MA-3                -0.190172    0.06862    -2.77   0.006
MA-4                 0.249310    0.06477     3.85   0.000
Constant              4.50164      5.109    0.881   0.379
Seasonal_1            3.73956     0.4468     8.37   0.000
Seasonal_2           -1.56575     0.4488    -3.49   0.001
D75Q2                 8.86071      1.728     5.13   0.000
D79Q3                 17.7749      1.728     10.3   0.000
```

```
log-likelihood    -511.476878
no. of observations       219  no. of parameters         11
AIC.T              1044.95376  AIC                4.77147834
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
sigma                 2.48051  sigma^2               6.15293

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.13316    -0.049728   -0.0062127      0.19680       4.8442       3.3024
-1.6184       26.819       23.095
```

The Constant and MA(2) are not statistically significant. The d parameter has turned positive and it is much greater than in the previous specifications! The test summary is displayed below:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   15.926 [0.0003]**
ARCH 1-1 test:    F(1,207)  =   5.7583 [0.0173]*
Portmanteau(14):  Chi^2(9)  =   16.523 [0.0567]
```

There is no autocorrelation, and for the first time the p-value of ARCH test is greater than 0.01.

It could be a good idea to fix MA(2). So, the 10^{th} model of this project is an **ARMA(0,4) with MA(2) fixed**:

```
---- Maximum likelihood estimation of ARFIMA(0,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter          0.449382    0.04593     9.78   0.000
MA-1                 0.328312    0.06990     4.70   0.000
MA-3                -0.186855    0.06382    -2.93   0.004
MA-4                 0.248251    0.06407     3.87   0.000
Constant              4.50719      4.861    0.927   0.355
Seasonal_1            3.73836     0.4451     8.40   0.000
Seasonal_2           -1.56417     0.4470    -3.50   0.001
D75Q2                 8.84937      1.725     5.13   0.000
D79Q3                 17.7466      1.725     10.3   0.000

log-likelihood    -511.484871
no. of observations       219  no. of parameters         10
AIC.T              1042.96974  AIC                4.76241892
mean(InflatQ)         5.49505  var(InflatQ)          29.4156
sigma                 2.48111  sigma^2                6.1559

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.12801   -0.0076234      0.19634       4.8442       3.3024      -1.6184
26.819       23.095
```

The specification is quite nice, as all the parameters are statistically significant apart from the Constant. The test summary follows:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   16.108 [0.0003]**
ARCH 1-1 test:    F(1,208)  =   5.7264 [0.0176]*
Portmanteau(14):  Chi^2(10) =   16.738 [0.0804]
```

And the outcome is very similar to the previous model’s tests.

One last thing that could be done, perhaps, is to remove the Constant from the model. A Test for linear restrictions can be done:

```
Test for linear restrictions (Rb=r):
R matrix
  d parameter        MA-1         MA-3         MA-4      Constant   Seasonal_1   Seasonal_2
      0.00000      0.00000      0.00000      0.00000       1.0000      0.00000      0.00000
D75Q2        D79Q3
0.00000      0.00000
r vector
      0.00000
LinRes Chi^2(1) = 0.859597 [0.3539]
```

Like with Seasonal, looking at the p-value there is no rejection of the H_0. Therefore, removing the Constant from the model should not make any difference in terms of results.

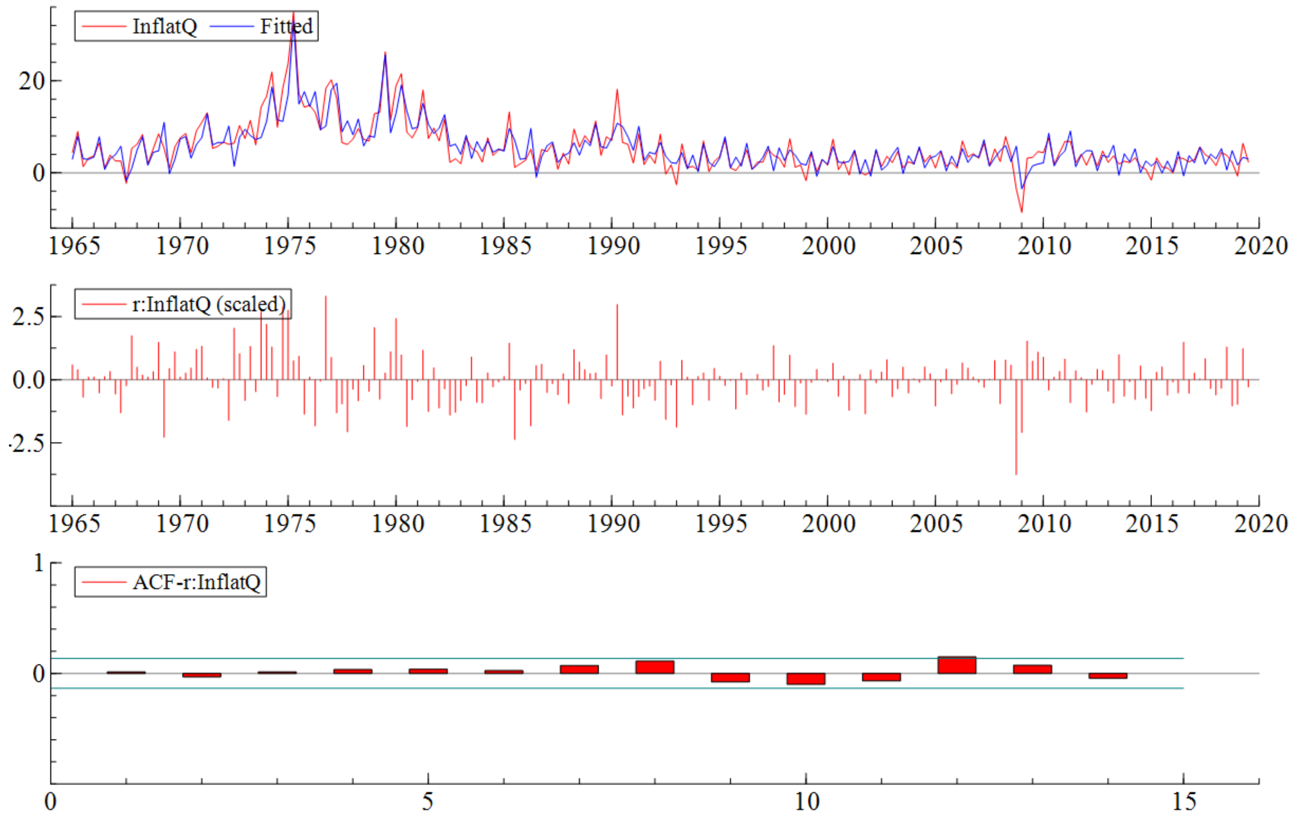
So, here is the **ARMA(0,4) with MA(2) fixed and no Constant**:

```
---- Maximum likelihood estimation of ARFIMA(0,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                  Coefficient  Std.Error  t-value  t-prob
d parameter          0.466087    0.03474     13.4   0.000
MA-1                 0.318538    0.06838     4.66   0.000
MA-3                -0.189912    0.06312    -3.01   0.003
MA-4                 0.249048    0.06420     3.88   0.000
Seasonal_1            3.74526     0.4426     8.46   0.000
Seasonal_2           -1.55407     0.4445    -3.50   0.001
D75Q2                 8.81898      1.720     5.13   0.000
D79Q3                 17.7134      1.720     10.3   0.000

log-likelihood     -511.80913
no. of observations       219  no. of parameters           9
AIC.T              1041.61826  AIC                 4.75624776
mean(InflatQ)         5.49505  var(InflatQ)           29.4156
sigma                 2.48224  sigma^2                6.16149

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.11341    -0.016944      0.48551       8.1466       3.2258       26.819
23.095
```

All the parameters are statistically significant, and the d parameter has grown a little. The tests:

```
Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   15.572 [0.0004]**
ARCH 1-1 test:    F(1,209)  =   5.5589 [0.0193]*
Portmanteau(14):  Chi^2(10) =   16.570 [0.0844]
```

Again, there is no autocorrelation, while the omnipresent non-normality of residuals and ARCH effect stay.
Also, looking at the graphical analysis of this model:



The overall fit is quite good, even if there are still spikes of a few outliers that are not perfectly covered. The ACF shows significant autocorrelation only at the 12th lag, which is quite far to be considered.

Summary

Several models have been specified, it is time to use the Progress tool and make some comparisons:

Progress to date

Model	T	p		log-likelihood	SC	HQ	AIC
Arfima(1)	219	14	EML	-551.95280	5.3852	5.2560	5.1685
Arfima(2)	219	12	EML	-554.64192	5.3605	5.2498	5.1748
Arfima(3)	219	16	EML	-508.57454	5.0382	4.8906	4.7906
Arfima(4)	219	13	EML	-514.62424	5.0197	4.8997	4.8185
Arfima(5)	219	12	EML	-514.62424	4.9951	4.8844	4.8094
Arfima(6)	219	13	EML	-508.91417	4.9675	4.8476	4.7663
Arfima(7)	219	11	EML	-511.25875	4.9397	4.8382	4.7695
Arfima(8)	219	12	EML	-509.35876	4.9470	4.8363	4.7613
Arfima(9)	219	11	EML	-511.47688	4.9417	4.8402	4.7715
Arfima(10)	219	10	EML	-511.48487	4.9172	4.8249	4.7624
Arfima(11)	219	9	EML	-511.80913	4.8955<	4.8125<	4.7562<

In terms of likelihood, the 6th model is the best one. This was the ARMA(4,2) with the exclusion of Seasonal and the inclusion of the dummy variables D75Q2 and D79Q3.
However, **in terms of information criteria such as Schwarz, Hannan-Quinn and Akaike, the best model from those elaborated is the last one, hence the ARMA(0,4) with MA(2) fixed, without Constant and Seasonal, and with the dummies D75Q2 and D79Q3**. The lower number of parameters (9) definitely plays an important role in the magnitude of the information criteria.

Below there is a table (Table 1) with the fundamental information regarding all the models:

No.	Model	Series included	Significance of parameters	Normality	ARCH effect	Autocorrelation
Arfima(1)	ARMA(4,4)	Constant, Seasonal, Seasonal_1, Seasonal_2	Only AR(2), MA(4) and Seasonal **not** significant	Non-normally distributed residuals	Present	Not present (barely)
Arfima(2)	ARMA(4,3) with AR(2) fixed	Constant, Seasonal, Seasonal_1, Seasonal_2	Only AR(4), Seasonal_1 and Seasonal_2 significant	Non-normally distributed residuals	Present	Not present
Arfima(3)	ARMA(4,4)	Constant, Seasonal, Seasonal_1, Seasonal_2, D75Q2, D79Q3	Only Constant, Seasonal_1, Seasonal_2, D75Q2 and D79Q3 significant	Non-normally distributed residuals	Present	Present
Arfima(4)	ARMA(4,3)	Constant, Seasonal_1, Seasonal_2, D75Q2+D79Q3	AR(1), AR(2), AR(4), d parameter, MA(2) and MA(3) **not** significant	Non-normally distributed residuals	Present	Not present
Arfima(5)	ARMA(4,2)	Constant, Seasonal_1, Seasonal_2, D75Q2+D79Q3	AR(1), AR(2) and AR(3) **not** significant	Non-normally distributed residuals	Present	Not present
Arfima(6)	ARMA(4,2)	Constant, Seasonal_1, Seasonal_2, D75Q2, D79Q3	Only AR(1) and AR(3) **not** significant	Non-normally distributed residuals	Present	Not present
Arfima(7)	ARMA(4,2) with AR(1) and AR(3) fixed	Constant, Seasonal_1, Seasonal_2, D75Q2, D79Q3	AR(2), MA(2) and d parameter **not** significant	Non-normally distributed residuals	Present	Present
Arfima(8)	ARMA(4,2) with AR(1) fixed	Constant, Seasonal_1, Seasonal_2, D75Q2, D79Q3	All the parameters are significant	Non-normally distributed residuals	Present	Present (barely)
Arfima(9)	ARMA(0,4)	Constant, Seasonal_1, Seasonal_2, D75Q2, D79Q3	Only Constant and MA(2) **not** significant	Non-normally distributed residuals	Present	Not present (barely)
Arfima(10)	ARMA(0,4) with MA(2) fixed	Constant, Seasonal_1, Seasonal_2, D75Q2, D79Q3	Only Constant **not** significant	Non-normally distributed residuals	Present	Not present
Arfima(11)	ARMA(0,4) with MA(2) fixed	Seasonal_1, Seasonal_2, D75Q2, D79Q3	All the parameters are significant	Non-normally distributed residuals	Present	Not present

Table 1. Specifications' summary.

The table confirms that the latest model (ARFIMA(11), underlined in green in Table 1) is the most suitable one:

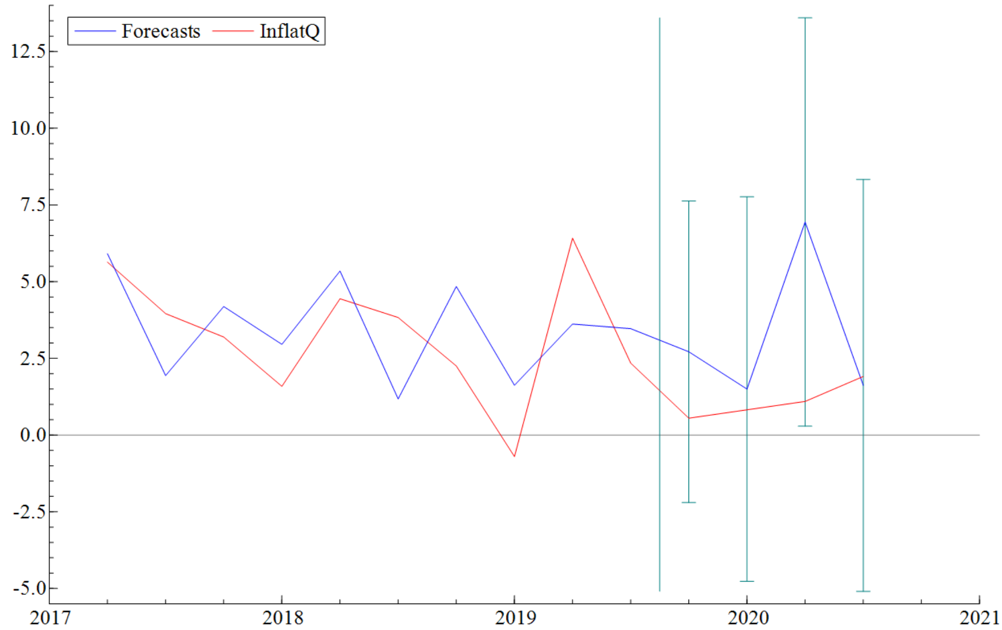
```
---- Maximum likelihood estimation of ARFIMA(0,d,4) model ----
The estimation sample is:  1965(1) - 2019(3)
The dependent variable is: InflatQ

                Coefficient  Std.Error  t-value  t-prob
d parameter        0.466087    0.03474     13.4   0.000
MA-1               0.318538    0.06838     4.66   0.000
```

```
MA-3                 -0.189912    0.06312    -3.01   0.003
MA-4                  0.249048    0.06420     3.88   0.000
Seasonal_1             3.74526     0.4426     8.46   0.000
Seasonal_2            -1.55407     0.4445    -3.50   0.001
D75Q2                  8.81898      1.720     5.13   0.000
D79Q3                  17.7134      1.720     10.3   0.000

log-likelihood     -511.80913
no. of observations       219  no. of parameters          9
AIC.T              1041.61826  AIC               4.75624776
mean(InflatQ)         5.49505  var(InflatQ)         29.4156
sigma                 2.48224  sigma^2              6.16149

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values:
      0.40000      0.11341    -0.016944      0.48551       8.1466       3.2258       26.819
23.095

Descriptive statistics for residuals:
Normality test:   Chi^2(2)  =   15.572 [0.0004]**
ARCH 1-1 test:    F(1,209)  =   5.5589 [0.0193]*
Portmanteau(14):  Chi^2(10) =   16.570 [0.0844]
```

This is the only model that has all parameters which are statistically significant and no autocorrelation within the residuals. Moreover, it is the best model in terms of information criteria too.
Perhaps, there are two more models (underlined in yellow in Table 1) which are quite nice too, namely the 6th one (highest log-likelihood) and the 10th one (very similar to the last one, with the only difference of having the Constant included in the specification, with this Constant being not significant).
A conclusion that can be dragged is that GARCH specifications will be necessary to remove the ARCH effect and the issue of non-normality of the residuals.

Forecasts

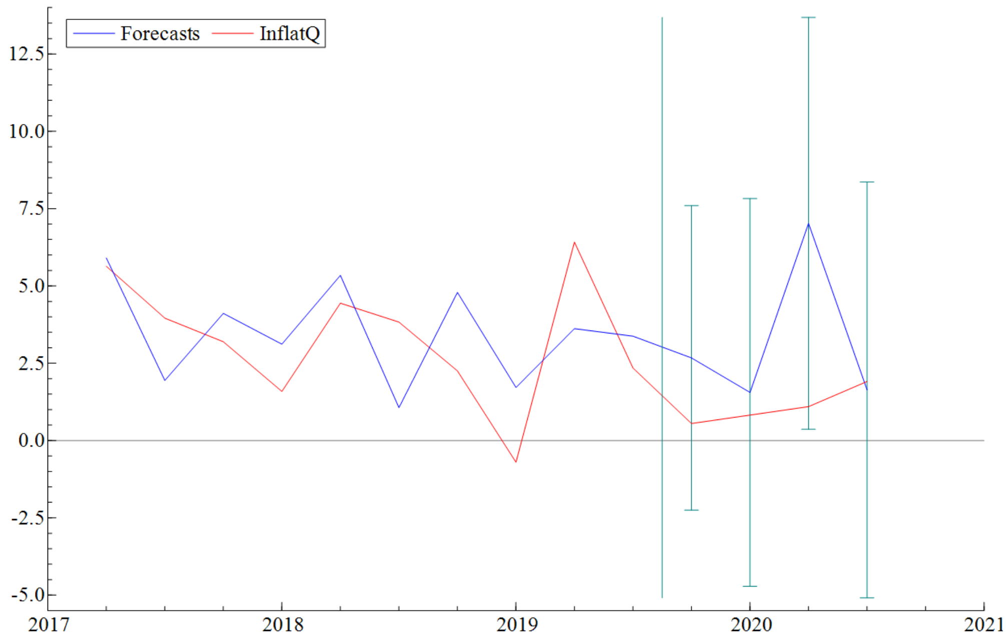
Now it's time to prepare forecasts for the models depicted as better ones in terms of significance of parameters, tests and information criteria. We need to be aware that good specifications in terms of diagnostics don't necessarily lead to good forecasts.
In order to prove this statement, the first forecasts will actually come from one of the worst models, ARFIMA(3), where one of the only positive things was the high value of log-likelihood. The number of forecasts equals 4, and they are created using errors bars and keeping 2 critical values:



The fit of the forecast is not very nice, presenting a quite big error in 2020(2). And here are the written results:

```
Forecasts from 2019(4)
Horizon     Forecast      (SE)       Actual        Error  Naive forc      (SE)
      1       2.7157     2.457      0.55002      -2.1657      2.7116     2.457
      2       1.4983     3.134      0.82361     -0.67466      1.4909     3.134
      3       6.9446     3.327       1.0955      -5.8491      6.9353     3.326
      4       1.6185     3.355       1.9100      0.29143      1.6086     3.355
   mean(Error) =       -2.0995     RMSE =        3.1402
   SD(Error)   =        2.3351     MAPE =        56.752
```

Next come the forecasts from ARFIMA(6), which was a rather good model:



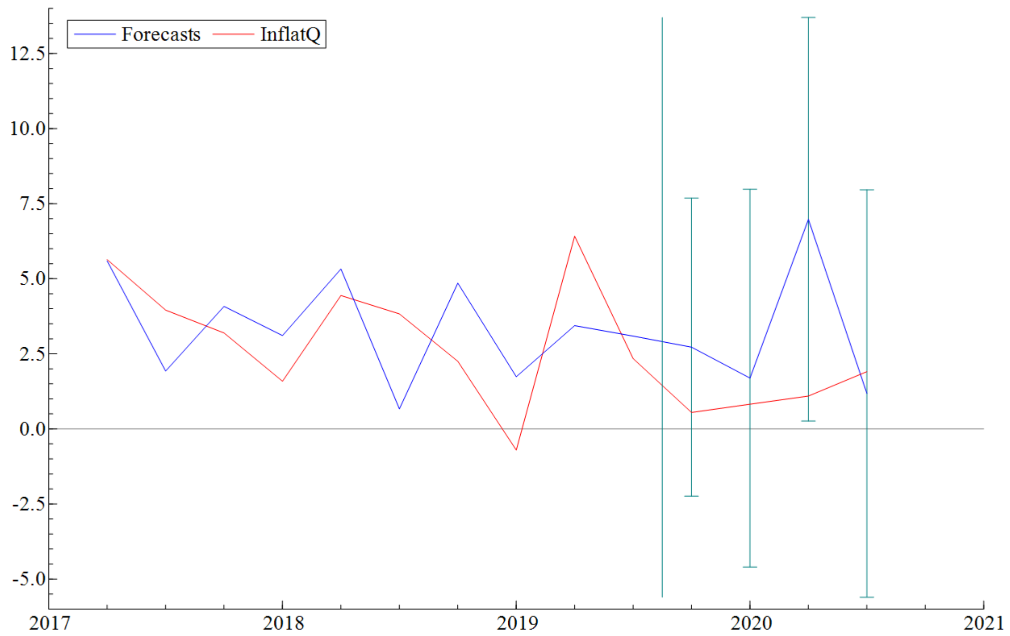
The graphical output doesn't vary much from the one of ARFIMA(3). And looking at the written output of the forecasts:

Forecasts from 2019(4)

Horizon	Forecast	(SE)	Actual	Error	Naive forc	(SE)
1	2.6717	2.461	0.55002	-2.1216	2.6651	2.461
2	1.5556	3.135	0.82361	-0.73195	1.5438	3.134
3	7.0242	3.331	1.0955	-5.9287	7.0095	3.329
4	1.6361	3.362	1.9100	0.27391	1.6202	3.360
mean(Error) =		-2.1271	RMSE =	3.1726		
SD(Error) =		2.3539	MAPE =	56.903		

It's visible that the RMSE and MAPE don't change significantly, they just increase very lightly.

In the forecasts for ARFIMA(10) the situation is very similar in terms of graphics:



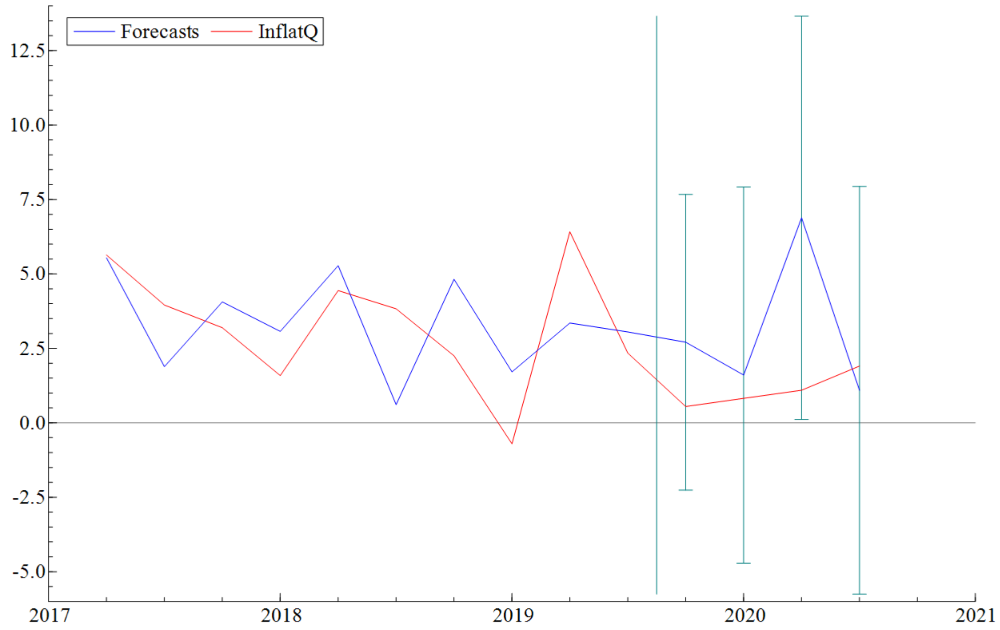
While from the written forecasts:

Forecasts from 2019(4)

Horizon	Forecast	(SE)	Actual	Error	Naive forc	(SE)
1	2.7246	2.482	0.55002	-2.1746	2.7171	2.481
2	1.6936	3.146	0.82361	-0.86999	1.6802	3.143
3	6.9811	3.359	1.0955	-5.8856	6.9642	3.355
4	1.1818	3.392	1.9100	0.72819	1.1636	3.387
mean(Error) =		-2.0505	RMSE =	3.1881		
SD(Error) =		2.4412	MAPE =	69.277		

It is visible that the values of RMSE and MAPE have increased, especially in the case of MAPE, revealing that these forecasts are worse than those from ARFIMA(3) and ARFIMA(6).

Lastly, the plot of forecasts for ARFIMA(11), which was the best model in terms of diagnostics:



The graphical difference is barely perceivable, as the only difference between this and the previous model is that the Constant was removed. And looking at the written forecasts:

Forecasts from 2019(4)

Horizon	Forecast	(SE)	Actual	Error	Naive forc	(SE)
1	2.7093	2.483	0.55002	-2.1593	2.5743	2.482
2	1.6074	3.158	0.82361	-0.78375	1.3667	3.155
3	6.8887	3.386	1.0955	-5.7932	6.5825	3.382
4	1.0925	3.423	1.9100	0.81742	0.76009	3.418

```
   mean(Error) =       -1.9797     RMSE =       3.1427
   SD(Error)   =        2.4408     MAPE =       71.843
```

Focusing on errors, the difference with the forecasts from ARFIMA(10) is very small. RMSE is lower, while MAPE is higher.

Therefore, we reach the conclusion that in terms of forecasts the best model would be ARFIMA(3), despite it's not the best one in terms of diagnostics; it is actually the worst one among those used for forecasting! This proves that a model which is not so nice in terms of diagnostics could produce better forecasts than models which are better constructed. However, the differences between the RMSE and MAPE of ARFIMA(3) and ARFIMA(6) are negligible.
Hence, I would choose ARFIMA(6) as the best overall model because it is quite good in terms of parameters' significance, log-likelihood and information criteria, and also it's forecasting ability is better than the one of ARFIMA(10) and ARFIMA(11), while it is just slightly smaller than the one of ARFIMA(3), which unfortunately isn't a good model in terms of diagnostics.