FORECASTING INFLATION

Course of Time Series

UNIVERSITA' DEGLI STUDI DI MILANO

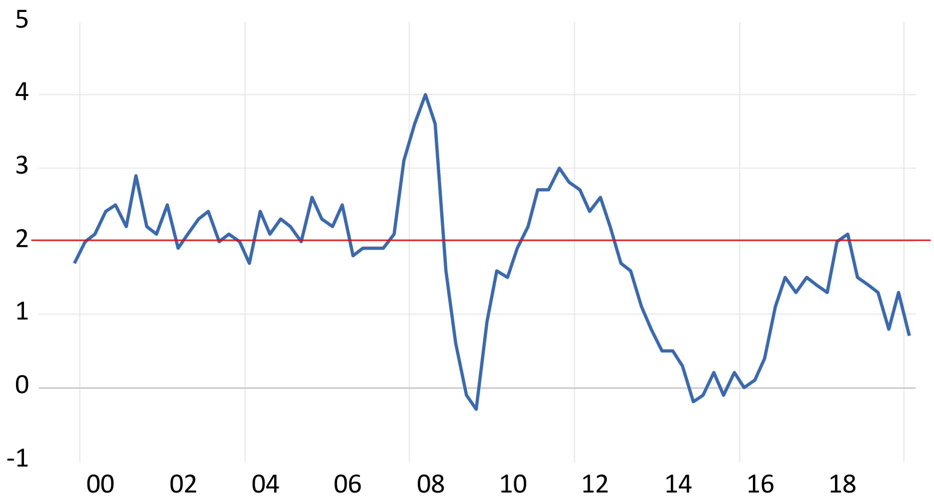
Sofia Gervasoni

Index

Introduction

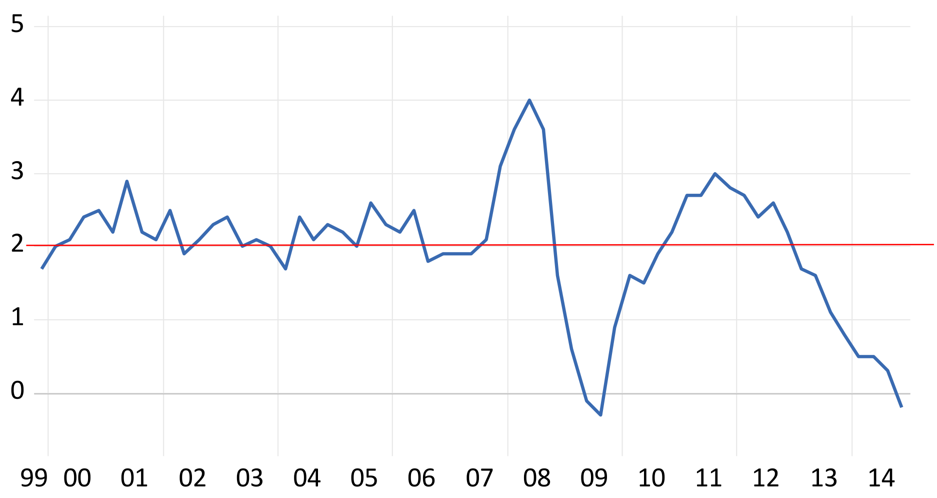
I started my analysis looking at the graph of the of the realisations of the quarterly inflation between 1999:Q4 and 2020:Q1 in the euro area. From the graph of this series, it is possible to notice that the series seems to satisfy the property of mean reverting, since even after shocks (as it happened for example between 2008 and 2009), the values revert to the mean and continue fluctuating around it. In this case the mean is around 2, and this is due to the fact that ECB, to maintain price stability, define annual HICP inflation rate of 2% over the medium term. So, from an economic point of view, there is always mean reverting in inflation since the Central Banks tend to keep it under control and stabilize it. It is also possible to notice that values between 1999 and 2007 present lower variance than the ones between 2007 and 2020 (even if always between certain boundaries).

REALISED



Since the aim of this analysis is to forecast inflation over the 2015Q1 to 2020Q1 and compare the results with other forecasts, I set my sample of observations between 1999Q4 and 2014Q4. On this sample I will estimate the model that I will use to forecast inflation for the period of interest (2015Q1 - 2020Q1).

REALISED



Test for Unit Root

First, I run an Augmented Dickey-Fuller (ADF) test, to test if there is a unit root. As said, in this case it seems that the values fluctuate around a mean equal to 2 (different from zero), and this suggest using the ADF test with the intercept (our Case 2 of this test). In this test H_0: "there is a unit root".

$$REALISED_t = \alpha + \rho\ REALISED_{t-1} + \epsilon_t$$

Where α is the constant (intercept) and ρ is the correlation of $REALISED_t$ with its past. In particular, I am testing $H_0 : \{\rho = 1\}$ against $H_A : \{|\rho|<1\}$.

Null Hypothesis: REALISED has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.521973	0.1155
Test critical values:	1% level	-3.546099	
	5% level	-2.911730	
	10% level	-2.593551	

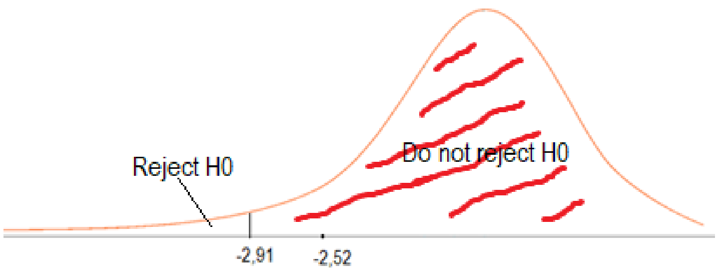
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(REALISED)
Method: Least Squares
Date: 11/27/21 Time: 13:30
Sample (adjusted): 2000Q2 2014Q4
Included observations: 59 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
REALISED(-1)	-0.200315	0.079428	-2.521973	0.0145
D(REALISED(-1))	0.363288	0.132197	2.748080	0.0081
C	0.376716	0.173121	2.176027	0.0338

R-squared	0.154714	Mean dependent var	-0.037288
Adjusted R-squared	0.124525	S.D. dependent var	0.503402
S.E. of regression	0.471017	Akaike info criterion	1.381663
Sum squared resid	12.42398	Schwarz criterion	1.487301
Log likelihood	-37.75907	Hannan-Quinn criter.	1.422900
F-statistic	5.124885	Durbin-Watson stat	2.111484
Prob(F-statistic)	0.009039		

Looking at the results of the ADF test, I concluded that there could be a unit root. The null hypothesis in this case could not be rejected, since the p-value (0.1155) is high (higher than 0.01, 0.05 and 0.1) and the estimated t-statistic (-2.52) is higher than the critical value at 5% (-2.91) for this test. In this case the critical values are adjusted for the sample size, for a 5% level we could also take -2.86 as a critical value (that is the asymptotic critical value).



*this distribution is not the real one, it is just to give you an idea of what I am saying above

These results are in contrast with the economic theory (and intuition), because in a unit root model there is no tendency to revert the mean, but as we have seen the inflation always tend to do it (i.e., HICP inflation rate of 2%). Thinking that the inflation has a unit root and an unbounded variance is not realistic, since as seen in the first graph the range in which it flunctuated is in a certain way controlled. Taking a model with unit root in this case means taking a forecast as the last observation and adjust it from there, but this does not inclued a reversion to the mean.

So, I run the ADF test in first difference and I found out the following resultsIn this case the model is written as

$$\boldsymbol{\Delta} REALISED_t = \alpha + (\rho\text{-}1)\ REALISED_{t-1} + \epsilon_t$$

So, in this case we are testing H_0: {ρ-1=0}, but the test is equivalent to the previous one (same limit distribution).

Null Hypothesis: D(REALISED) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.866730	0.0000
Test critical values:	1% level	-3.546099	
	5% level	-2.911730	
	10% level	-2.593551	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(REALISED,2)
Method: Least Squares
Date: 11/27/21 Time: 13:34
Sample (adjusted): 2000Q2 2014Q4
Included observations: 59 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(REALISED(-1))	-0.756780	0.128995	-5.866730	0.0000
C	-0.031517	0.064213	-0.490819	0.6254

R-squared	0.376494	Mean dependent var	-0.013559
Adjusted R-squared	0.365555	S.D. dependent var	0.618524
S.E. of regression	0.492667	Akaike info criterion	1.455343
Sum squared resid	13.83507	Schwarz criterion	1.525768
Log likelihood	-40.93262	Hannan-Quinn criter.	1.482834
F-statistic	34.41852	Durbin-Watson stat	2.014110
Prob(F-statistic)	0.000000		

Running the ADF test in first difference the null hypotesis of having a unit root could be rejected (pval=0)[1]. That means that D(realised) [2] is intergrated of order 0 (I(0)), that means that the process $REALISED_t$ is integrated of order 1 (I(1)).

This would suggest modelling inflation in first differences.

[1] In this case we are testing H_0:{ρ-1=0}

[2] First difference of REALISED ($\boldsymbol{\Delta} REALISED_t = REALISED_t - REALISED_{t-1}$)

Due to the discordance between what the unit test pointed out and the economic intuition, I will estimate two different models: the first in ***levels*** (assuming stationariety) and the second one in ***first differences*** (solving the unit root problem). I will use both to do forecast and check which one works better.

Model in levels

Model selection and estimation

Sample: 1999Q4 2014Q4
Included observations: 61

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.783	0.783	39.237	0.000
		2	0.512	-0.260	56.306	0.000
		3	0.217	-0.235	59.413	0.000
		4	-0.079	-0.236	59.828	0.000
		5	-0.216	0.163	63.045	0.000
		6	-0.280	-0.057	68.529	0.000
		7	-0.286	-0.080	74.364	0.000
		8	-0.279	-0.187	80.000	0.000
		9	-0.237	0.064	84.137	0.000
		10	-0.180	-0.003	86.570	0.000
		11	-0.130	-0.048	87.860	0.000
		12	-0.073	-0.076	88.275	0.000

Looking at the correlogram of the series it is possible to observe that the global autocorrelation falls to zero after the second lag, whereas the partial autocorrelation falls to zero after the first lag and this suggest to try an ARMA(1,2)[3].

To choose the right model, I tried to estimate[4] different ARMA models (with different number of parameters) and then I used the Schwarz Information Criteria (BIC)[5] to select the model that I will use for my forecasts.

As expected the right model is an ARMA(p,q) model with p=1 and q=3 (this is not so far from what I expected).

As it is possible to see in the following table there are some values for Schwarz Criterium that are lower than the one for the ARMA(1,3), but in these models were excluded because of “convergence not achieved after 500 iterations” or non-invertible MA components.

Schwarz Criterium	***c***	***MA(1)***	***MA(2)***	***MA(3)***	***MA(4)***	***MA(5)***
c	2.608808	1.825981	1.744018	1.206893	1.195287	~~1.143108~~
AR(1)	1.532299	1.511063	1.38477	***1.163117***	1.227834	1.292808
AR(2)	1.487301	1.49167	1.608789	1.247297	~~0.771644~~	~~0.770682~~
AR(3)	1.523958	1.578686	1.574985	~~0.86181~~	1.338586	1.196262
AR(4)	1.566320	1.539646	1.494364	1.652201	1.41009	1.326726
AR(5)	1.562413	1.604862	1.561464	1.476744	1.500029	1.367135

The following output contains the estimations for the model that I selected, that is an ARMA(1,3):

[3] In AR(p) models the global autocorrelation goes to zero exponentially, whereas the partial autocorrelation goes to zero after p lags (and vice versa for MA(q), where the global autocorrelation goes to zero after q lags).
[4] I used the Conditional Least Square estimates (CLS)
[5] We look for the lowest BIC (Schwarz IC)

Dependent Variable: REALISED
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/24/21 Time: 21:47
Sample (adjusted): 2000Q1 2014Q4
Included observations: 60 after adjustments
Failure to improve likelihood (non-zero gradients) after 23 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1999Q2 1999Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.852899	0.316728	5.850135	0.0000
AR(1)	0.496974	0.138240	3.595015	0.0007
MA(1)	0.651281	0.097172	6.702360	0.0000
MA(2)	0.801238	0.049891	16.05972	0.0000
MA(3)	0.768830	0.084703	9.076738	0.0000

R-squared	0.823466	Mean dependent var	1.986667
Adjusted R-squared	0.810627	S.D. dependent var	0.875943
S.E. of regression	0.381184	Akaike info criterion	0.988588
Sum squared resid	7.991585	Schwarz criterion	1.163117
Log likelihood	-24.65765	Hannan-Quinn criter.	1.056856
F-statistic	64.13859	Durbin-Watson stat	1.986720
Prob(F-statistic)	0.000000		

Inverted AR Roots	.50		
Inverted MA Roots	.08+.97i	.08-.97i	-.82

As we can observe from this output all the p-values related to each parameter are 0, and that means that the null hypothesis that the estimated parameters are statistically not significant could be rejected. This is a confirmation that the parameters that we have in the model are all significant (and relevant). From the estimated coefficients it is possible to notice that the dependence of the series on its past vanish quicker than the dependence on the past shocks, the estimated coefficients for the MA part of the model are high (nearly close to 1) and vanish very slow.

Model validation

The step after the model selection is the model validation, that could be done using a Portmanteau test. This test asses the null hypothesis that the residuals are independently distributed.

Sample (adjusted): 2000Q1 2014Q4
Q-statistic probabilities adjusted for 4 ARMA terms

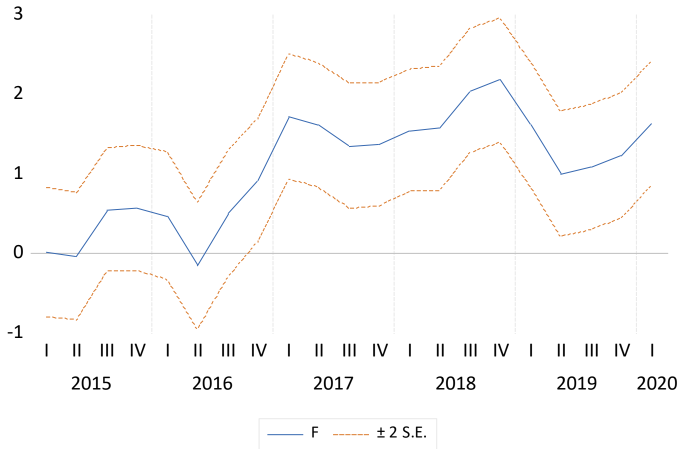
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.026	-0.026	0.0436	
		2	0.039	0.038	0.1399	
		3	0.034	0.036	0.2157	
		4	-0.107	-0.107	0.9769	
		5	-0.035	-0.044	1.0596	0.303
		6	0.046	0.053	1.2075	0.547
		7	0.009	0.023	1.2135	0.750
		8	-0.094	-0.108	1.8391	0.765
		9	-0.070	-0.091	2.1941	0.822
		10	0.021	0.037	2.2280	0.898

The assumption that the residuals are independently distributed is not rejected by the data and this could also be seen by looking at the correlograms of the residuals: the bars are all within the confidence interval.

Thus, the Portmanteau test on the residuals confirms that the ARMA(1,3) is acceptable, if the series is stationary (and I assumed it at the beginning of this analysis).

Forecasts

Since the model is valid, it could be used to do forecasts. So, I set as forecast sample the period 2015Q1-2020Q1 and since I wanted the forecast to update every period, I selected the "static forecast". I obtained the following forecast that I called *F*.



Forecast: F	
Actual: REALISED	
Forecast sample: 2015Q1 2020Q1	
Included observations: 21	
Root Mean Squared Error	0.372579
Mean Absolute Error	0.300045
Mean Abs. Percent Error	NA
Theil Inequality Coef.	0.153126
Bias Proportion	0.129655
Variance Proportion	0.001042
Covariance Proportion	0.869302
Theil U2 Coefficient	NA
Symmetric MAPE	67.69707

Model in first difference

Model selection and estimation

Due to the presence of the unit root pointed out from the ADF at the beginning, I also estimate a model in first different and to understand which model I could estimate I looked at the correlogram of the series in first difference.

Sample (adjusted): 2000Q1 2014Q4
Included observations: 60 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.239	0.239	3.6130	0.057
		2	0.103	0.049	4.2951	0.117
		3	0.000	-0.037	4.2951	0.231
		4	-0.439	-0.464	17.083	0.002
		5	-0.192	0.007	19.570	0.002
		6	-0.092	0.043	20.157	0.003
		7	-0.042	0.023	20.281	0.005
		8	-0.035	-0.296	20.368	0.009
		9	0.021	-0.009	20.400	0.016
		10	-0.020	-0.012	20.430	0.025
		11	0.028	0.094	20.490	0.039
		12	0.081	-0.096	20.994	0.050
		13	-0.160	-0.319	23.028	0.041
		14	0.102	0.218	23.866	0.048
		15	-0.029	0.032	23.935	0.066
		16	-0.100	-0.135	24.779	0.074
		17	0.200	-0.024	28.233	0.042
		18	-0.116	-0.115	29.423	0.043
		19	0.029	0.179	29.501	0.059
		20	0.058	-0.028	29.819	0.073

As I did before, I look at the correlograms to understand which model I could estimate. In this case it is hard to deduce a model looking at the correlogram, so I estimated different ARMA models (with different parameters) and I compared their Schwarz information criterium (that you can find in the following table).

Schwarz Criterium	***c***	***MA(1)***	***MA(2)***	***MA(3)***
c	1.507063	1.521185	1.418870	1.321124
AR(1)	1.525768	1.593017	1.472034	***1.317625***
AR(2)	1.611867	1.681867	1.429976	1.389552
AR(3)	1.695244	1.595243	1.509870	1.4416

Also in this case estimating the ARMA models[6] with different number parameters and comparing the resulting values of the Schwarz information criterion, I obtained that the suggested model is an ARMA(1,3).

Dependent Variable: D(REALISED)
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/27/21 Time: 15:20
Sample (adjusted): 2000Q2 2014Q4
Included observations: 59 after adjustments
Failure to improve likelihood (non-zero gradients) after 28 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1999Q3 2000Q1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.070992	0.128798	-0.551191	0.5838
AR(1)	-0.474362	0.131883	-3.596848	0.0007
MA(1)	0.830993	0.064242	12.93539	0.0000
MA(2)	0.822598	0.062089	13.24875	0.0000
MA(3)	0.927452	0.047112	19.68600	0.0000

R-squared	0.378723	Mean dependent var	-0.037288
Adjusted R-squared	0.332702	S.D. dependent var	0.503402
S.E. of regression	0.411220	Akaike info criterion	1.141562
Sum squared resid	9.131512	Schwarz criterion	1.317625
Log likelihood	-28.67609	Hannan-Quinn criter.	1.210290
F-statistic	8.229429	Durbin-Watson stat	2.042209
Prob(F-statistic)	0.000029		

Inverted AR Roots	-.47		
Inverted MA Roots	.07-.98i	.07+.98i	-.97

As we can observe from this output all the p-values related to each parameter are close to 0 (except for the constant), and that means that the null hypothesis that the estimated parameters are statistically not significant could be rejected. This is a confirmation that the parameters that we have in the model are all statistically significant.

In this case the estimated coefficient related to AR(1) is negative, whereas the persistence of the MA part of the model is higher and seems to increase over time (from 0.83 to 0.92).

[6] I used the Conditional Least Square estimates (CLS)

As said the estimation of the constant does not result statistically significant, since the p-value related is higher than 0.05. Despite it, I decided to keep the constant in my model.

Model validation

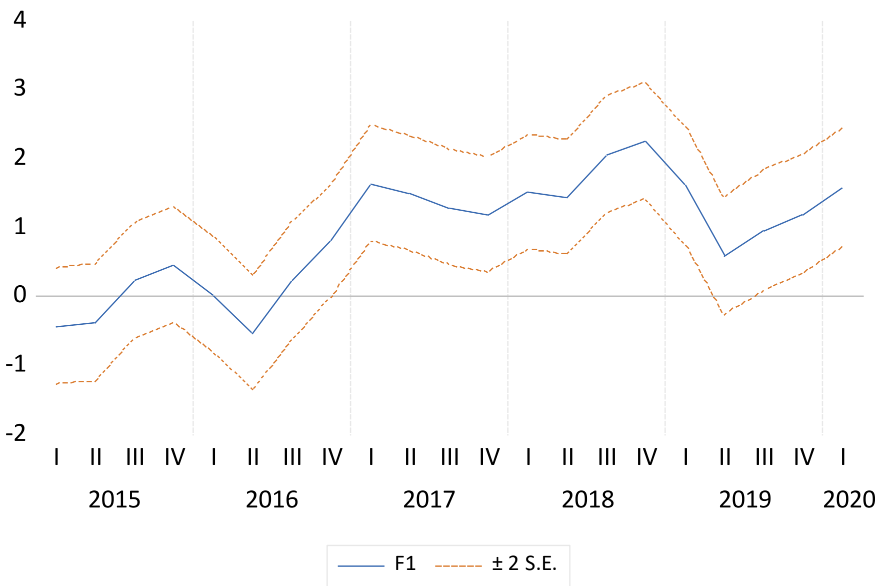
Date: 11/27/21 Time: 15:25
Sample (adjusted): 2000Q2 2014Q4
Q-statistic probabilities adjusted for 4 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.037	-0.037	0.0848	
		2	-0.145	-0.146	1.4082	
		3	-0.119	-0.134	2.3200	
		4	-0.107	-0.147	3.0690	
		5	-0.174	-0.245	5.0952	0.024
		6	0.154	0.064	6.7144	0.035
		7	0.031	-0.065	6.7790	0.079
		8	-0.042	-0.091	6.9033	0.141
		9	-0.108	-0.165	7.7409	0.171
		10	0.082	0.017	8.2346	0.221
		11	0.069	0.063	8.5913	0.283
		12	0.020	-0.020	8.6231	0.375
		13	-0.138	-0.179	10.120	0.341
		14	0.090	0.072	10.765	0.376
		15	0.000	0.046	10.765	0.463
		16	-0.127	-0.155	12.123	0.436

The assumption that the residuals are independently distributed is not rejected by the data and this could also be seen by looking at the correlograms of the residuals: the bars are all within the confidence interval. Moreover, in this case the model is stationary (no unit root), so the Portmantua test holds.

Forecasts

Since the model is valid, it could be used to do forecasts. So, I set as forecast sample the period 2015Q1-2020Q1 and since I wanted the forecast to update every period, I selected the “static forecast”. I obtained the following forecast that I called *F1*.



Forecast: F1	
Actual: REALISED	
Forecast sample: 2015Q1 2020Q1	
Included observations: 21	
Root Mean Squared Error	0.413486
Mean Absolute Error	0.333993
Mean Abs. Percent Error	NA
Theil Inequality Coef.	0.173813
Bias Proportion	0.007041
Variance Proportion	0.092643
Covariance Proportion	0.900316
Theil U2 Coefficient	NA
Symmetric MAPE	67.95786

Dibod and Mariano test to compare forecasts

The DM test is a test to compare the predictive accuracy of two forecasts. Under the null hypothesis the two forecasts have the same accuracy (H_0: $E(d_t) = 0$), whereas under the alternative hypothesis is that the two forecasts have different levels of accuracy (H_1: $E(d_t) \neq 0$). Here, d_t is the loss differential between the two forecasts ($d_t = e_{1t}^2 - e_{2t}^2$). Moreover, e_{1t} and e_{2t} are the forecast errors obtained respectively from forecast 1 ($\widehat{Y_1}$) and forecast 2 ($\widehat{Y_2}$). More in detail, $e_{1t} = Y_1 - \widehat{Y_1}$ and $e_{2t} = Y_2 - \widehat{Y_2}$ (where Y_1 and Y_2 represent the realised values). If $E(d_t) > 0$ the second forecast is better (the first one has a higher MSE than the second forecast), whereas if $E(d_t) < 0$ the first forecast is better than the second one (the second one has a higher MSE than the first forecast).

Comparison between my forecast (F) - (F1) and the forecasts from the professional forecasters (F_a)

I set as forecast sample the period 2015Q1-2020Q1, so that I selected the years on which I am interested in for the forecast. Then, I calculated the forecast errors both for my forecast (*e=realised-F*) and for the forecast provided by the professional forecasters (*e_a= realised - F_a*). Then I compared the forecast, testing H_0: E(d)=0 (where *$d = e_a^2 - e^2$) with the DM test*[7].

Dependent Variable: EA^2-E^2
Method: Least Squares
Date: 11/25/21 Time: 11:40
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.328804	0.152976	2.149379	0.0440

R-squared	0.000000	Mean dependent var	0.328804
Adjusted R-squared	0.000000	S.D. dependent var	0.530896
S.E. of regression	0.530896	Akaike info criterion	1.617947
Sum squared resid	5.637013	Schwarz criterion	1.667686
Log likelihood	-15.98844	Hannan-Quinn criter.	1.628742
Durbin-Watson stat	0.776608		

As it is possible to see from the output of the DM test, the two forecast could have different predictive ability, since the p-value suggest rejecting the null hypothesis at a 95% confidence interval (even if the null hypothesis could not be rejected at a 99% confidence interval). Looking at the estimated coefficient of C, the professional forecast benchmark has a bigger mean square error than the one of my forecasts.

[7] I adjusted for the dependence in the errors, using the HAC covariance method

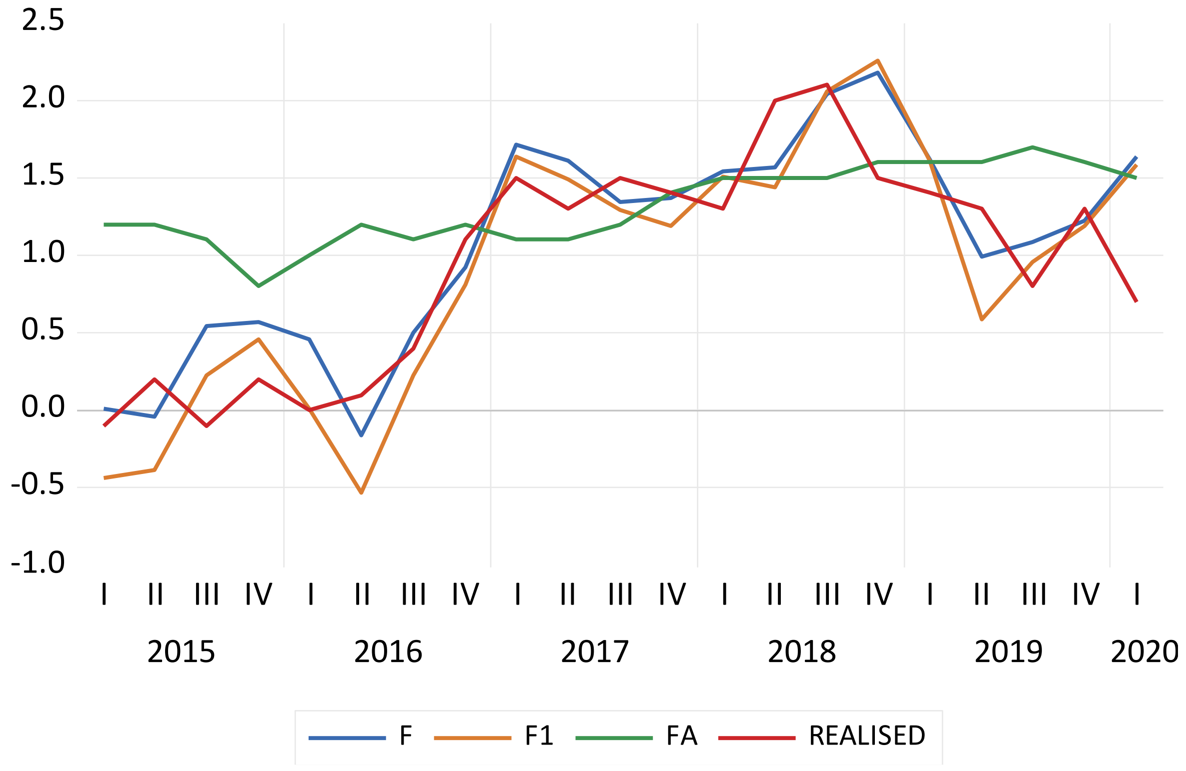
Comparing the forecasts for the model in *first differences* (F1) and the forecasts of the professional forecasters (Fa), It is possible to conclude that the two forecast could have more or less the same predictive ability, since the p-value suggest to not rejecting the null hypothesis at a 95% confidence interval (pval= 0.075 > 0.05). Even in this case, the estimated coefficient of C is positive (even if smaller than before), so the professional forecast benchmark has a bigger mean square error than the one of my forecasts.

Dependent Variable: EA^2-E1^2
Method: Least Squares
Date: 11/27/21 Time: 15:56
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.296649	0.157920	1.878475	0.0750

R-squared	0.000000	Mean dependent var	0.296649
Adjusted R-squared	0.000000	S.D. dependent var	0.552168
S.E. of regression	0.552168	Akaike info criterion	1.696520
Sum squared resid	6.097792	Schwarz criterion	1.746259
Log likelihood	-16.81346	Hannan-Quinn criter.	1.707314
Durbin-Watson stat	1.130497		

Comparing the values obtained from my forecasts, the values obtained from the professional forecasters and the effective values, it is possible to observe that my forecasts (blue and orange line) are much closer to the real values (red line) than the professional forecasts (green line).



Comparison between my forecasts (F)-(F1) and a naive forecast made taking as forecast the last observed variable (F_b)

The naïve forecast takes as forecast the last observed variable (*f_b=realised(-1) in the sample 2015q1-2020q1*). After calculating the forecast error for the forecast *F_b* (*e_b= realised - F_b*), I run again the DM test. In this case I am testing H_0: E(d)=0 (where *$d= e^2 - e_b^2$).*

Dependent Variable: E^2-EB^2
Method: Least Squares
Date: 11/29/21 Time: 18:48
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000243	0.042069	0.005787	0.9954

R-squared	0.000000	Mean dependent var	0.000243
Adjusted R-squared	0.000000	S.D. dependent var	0.204746
S.E. of regression	0.204746	Akaike info criterion	-0.287647
Sum squared resid	0.838417	Schwarz criterion	-0.237908
Log likelihood	4.020291	Hannan-Quinn criter.	-0.276852
Durbin-Watson stat	1.700072		

In this case I could not reject the null hypothesis for every confidence interval (p-value lower than 0.01 and the test statistic is beyond the critical value). This means that the two models could have the same predictive accuracy. Also, the coefficient is really close to zero, even if the naïve forecast seems doing slightly better than the model that I estimated.

Also, for the model estimated in *first differences* the DM test seems not rejecting the null hypothesis that the naïve benchmark and the model that I have estimated have the same predictive ability. My forecast again is doing more or less as the naïve benchmark (even if the naïve benchmark is doing slightly better), since as it is possible to observe the naïve benchmark has a mean square error close to the ones produced by my models (even if for the naïve benchmark is slightly lower).

Dependent Variable: E1^2-EB^2
Method: Least Squares
Date: 11/29/21 Time: 18:49
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.032399	0.040532	0.799341	0.4335

R-squared	0.000000	Mean dependent var	0.032399
Adjusted R-squared	0.000000	S.D. dependent var	0.223628
S.E. of regression	0.223628	Akaike info criterion	-0.111220
Sum squared resid	1.000187	Schwarz criterion	-0.061481
Log likelihood	2.167812	Hannan-Quinn criter.	-0.100426
Durbin-Watson stat	1.987602		

Comparison between my forecast (F) and the 2% constant, that is the ECB target (F_c)

The ECB inflation target is equal to 2% in the medium term, so as we have seen at the beginning the values of inflation tend to fluctuate around this value. So, the forecast with ECB target for inflation will take value 2 for all the period between 2015Q1 and 2020Q1. After calculating the forecast error for the forecast F_c (e_c=*realised* – F_c), I run again the DM test. In this case I am testing H_0: E(d)=0 (where $d = e^2 - e_c^2$).

Dependent Variable: E^2-EC^2
Method: Least Squares
Date: 11/25/21 Time: 12:43
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.425471	0.549281	-2.595158	0.0173

R-squared	0.000000	Mean dependent var	-1.425471
Adjusted R-squared	0.000000	S.D. dependent var	1.581936
S.E. of regression	1.581936	Akaike info criterion	3.801624
Sum squared resid	50.05043	Schwarz criterion	3.851363
Log likelihood	-38.91705	Hannan-Quinn criter.	3.812418
Durbin-Watson stat	0.210845		

In this case, my forecast is doing better than the ECB target benchmark (I could reject the null hypothesis at a 95% confidence interval, since 0.0173<0.05). As it possible to see from the estimation on the coefficient C, the ECB target benchmark has a significantly bigger mean square error (since the sign of the estimated coefficient is negative).

Also, for the model estimated with the first differences the DM test seems rejecting the null hypothesis that the ECB target benchmark and the model that I have estimated have the same predictive ability (0.0212<0.05). My forecast (F1) again is doing better than the ECB target benchmark, since as it is possible to observe the ECB target benchmark has a significantly bigger mean square error.

Dependent Variable: E1^2-EC^2
Method: Least Squares
Date: 11/27/21 Time: 16:11
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.393315	0.557384	-2.499739	0.0212

R-squared	0.000000	Mean dependent var	-1.393315
Adjusted R-squared	0.000000	S.D. dependent var	1.610785
S.E. of regression	1.610785	Akaike info criterion	3.837768
Sum squared resid	51.89254	Schwarz criterion	3.887507
Log likelihood	-39.29656	Hannan-Quinn criter.	3.848562
Durbin-Watson stat	0.289686		

Comparison between the forecasts produced by the model in levels (F) and the one in first difference (F1)

As it is possible to observe from the output of the DM test the null *hypothesis [H_0: $E(d)=0$ (where $d= e^2- e_1^2$)]* could not be rejected, this means that the two models could have the same predictive ability. The negative value for the coefficient suggests that the model estimated in levels is doing slightly better than the one estimated in first differences.

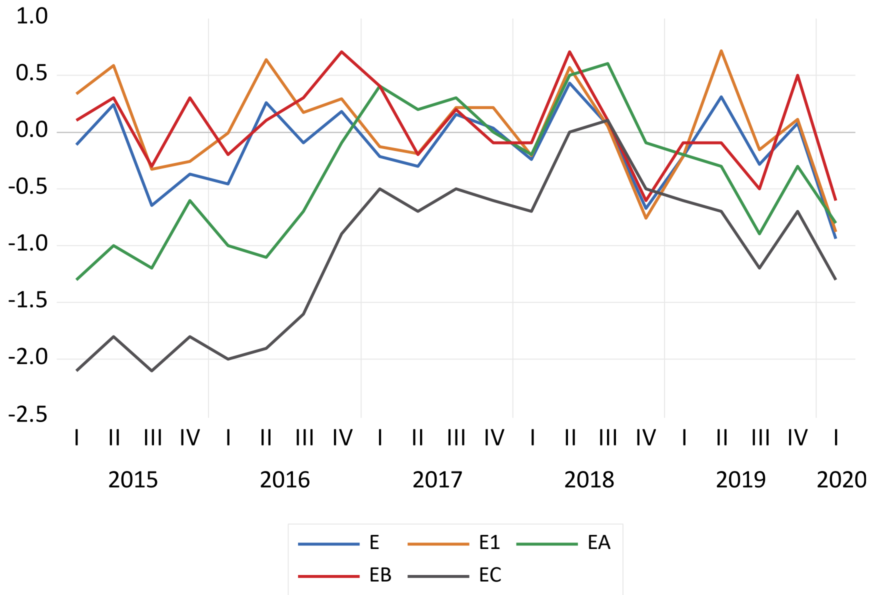
Dependent Variable: E^2-E1^2
Method: Least Squares
Date: 11/27/21   Time: 16:15
Sample: 2015Q1 2020Q1
Included observations: 21
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.032156	0.031124	-1.033141	0.3139

R-squared	0.000000	Mean dependent var	-0.032156
Adjusted R-squared	0.000000	S.D. dependent var	0.165425
S.E. of regression	0.165425	Akaike info criterion	-0.714152
Sum squared resid	0.547307	Schwarz criterion	-0.664413
Log likelihood	8.498596	Hannan-Quinn criter.	-0.703357
Durbin-Watson stat	2.454956		

Final considerations

In this case the naïve benchmark is doing better than all the models estimated and the other forecasts. The predictive accuracy of the naïve forecast and the ARMA(1,3) model that I estimated (in levels) are more all less the same, since their mean square errors are very close. As it is possible to observe from the following graph, that compares the forecasts errors for the different forecasts, my forecast (F) obtained using the ARMA(1,3) model in levels (blue line) and the forecast obtained with the naïve method (F_b, red line) are the best, since the values of their forecast errors (e and e_b respectively) are more or less always close to 0. Also the model ARMA(1,3) estimated in first differences (e_1, orange line) is doing well, even if slightly worse than the model estimated in levels and the naïve benchmark. The forecast errors of the professional forecasts (e_a, green line) are higher than the ones produced from to my forecasts and the naïve benchmark, and this is possible to see also from the mean square error (*this is because we cheated a little bit!*). The shape of the lines that represent the forecast errors using the ECB target benchmark (black line) and the professional forecast is the same, but the professional forecast tends to have lower mean square error.

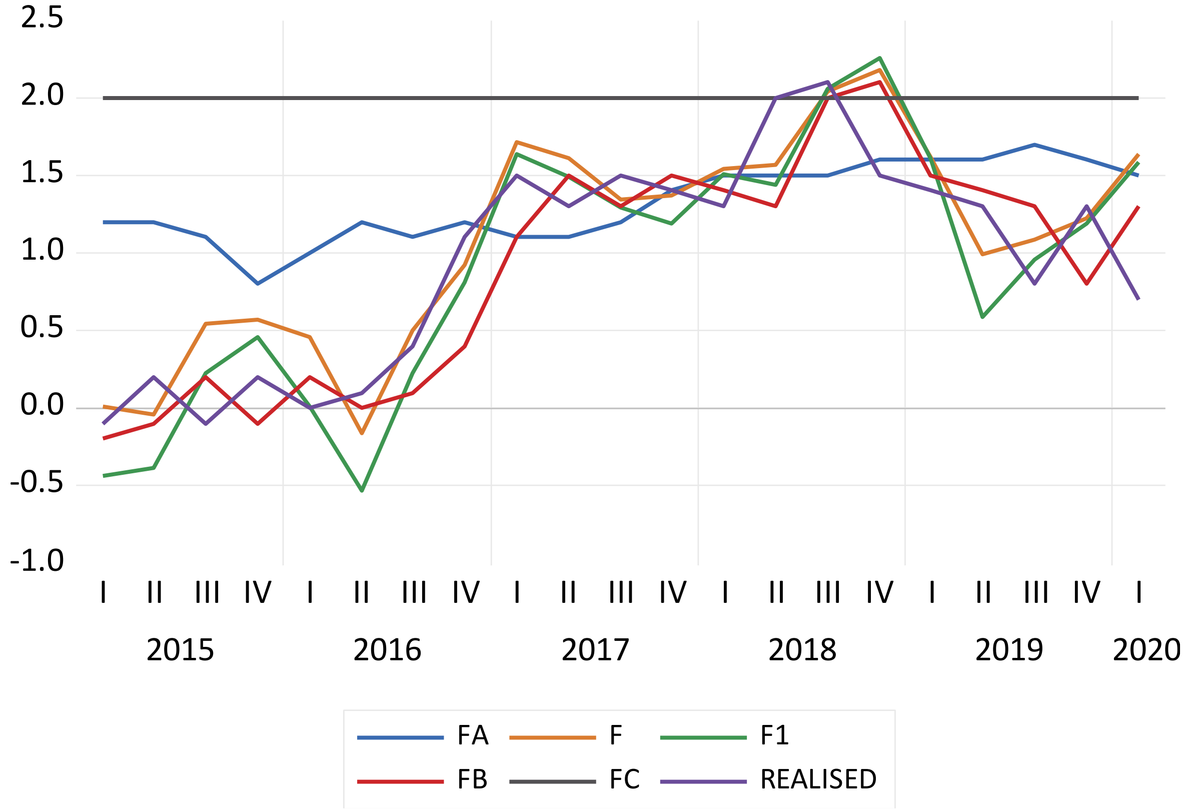


	E^2	E_1^2	Ea^2	Eb^2	Ec^2
mean	0,138815	0,17097	0,467619	0,138571	1,564286

Comparing the mean square errors, we come to the same conclusion that we obtained from the graph. The forecasts obtained with the ARMA(1,3) estimated in levels and the ones obtained with the naïve forecasts are the best one, since the MSEs connected to them are the lowest (0,138815 and 0,138571 respectively). The MSE of the forecasts done with the ARMA(1,3) estimated in first differences is close to the one related to the forecast errors done with the model estimated in levels, even if a little bit higher (0,17097). The fourth lower MSE is the one connected with the professional forecasts (0,467619), followed by the ECB target benchmark (1,564286) which is the higher.

It is possible to conclude that even if the unit root test (ADF test) pointed out the possible presence of a unit root, assuming stationarity and estimating a model in levels in this case led to better forecasts (even if not so far from the ones obtained with the model estimated in first differences).

The following graph represent the values assumed by the different forecast compared to the real values (purple line). From this graph it is possible to observe that the forecasts obtained from my models (levels and first difference) and the naïve benchmark are the closer to the real values (F, F_1 and F_b). The professional forecasts (Fa) and the ECB target benchmark (Fc) are the most far from the realised values.



2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
I II III IV I II III IV I II III IV I II III IV I II III IV I
2015
2016
2017
2018
2019
2020
FA
F
F1
FB
FC
REALISED