TIME SERIES PROJECT

Luca Paoletti

matr. 987940

December 2021

Indice

1 Introduction and Aim of the project

Series in HadCRUT4 S is the series of temperatures for the Southern Hemisphere for the years 1850 / 2021.
The series is a 2061 long series of monthly data of temperature deviations (anomaly): more information about this series is at the end of the message. For each year, we have 12 monthly observations: we should extract the average winter temperature, defined as the average of December, January and February (notice that we do not have December 1850 so we will start with Winter1851, which is the average of December1850, January1851 and February1851) and obtain a series of Yearly observations of winters (starting in 1851) (discard any other information). We should therefore have 2021-1951+1 observations at annual frequency.
Using the annual data from 1851 to 1951 I will make a model to make a one step ahead forecast of the temperatures for the period 1951-2021.
Using the Diebold - Mariano test, I will evaluate the forecasts against a naive benchmark where the forecast is the last available observation, over the sample 1951-1981. Using the Diebold - Mariano test, I will evaluate the forecasts (generated using the temperatures observed between 1851 and 1951) against a naive benchmark where the forecast is the last available observation, over the sample 1981-2021.

2 Application

2.1 Subset of the data

First of all, as I need only winter temperature (average winter temperature), I will have to subset the series with only the needed data. In this part I will use the software R.

```
#import series
series=read.delim(".../HadCRUT4_S.txt", header = FALSE)
#rename column
colnames(series)=c("Date", "Temperature")

#extract winter date (Dec, Jan, Feb)
winter_series=data.frame()
for (date in 1:nrow(series)){
  if (strsplit(series$Date[date], "/")[[1]][2] %in% c("12", "01", "02"))
    winter_series=rbind(series[date, ], winter_series)
}

#reverse df (from past to present)
winter_series=winter_series[nrow(winter_series):1, ]

#remove Jan and Feb 1850 (starting with Winter 1851)
winter_series=winter_series[-c(1:2), ]

#winter avg
n <- 3 #winter month
winter_temp=round(aggregate(winter_series$Temperature, list(rep(1:(nrow(
    winter_series) %/% n + 1), each = n,
                                  len = nrow(winter_series))), mean)[-1],
    3)
winter_avg=data.frame("Winter"=c(1851:2021),
                      "Temperature"=winter_temp)
colnames(winter_avg)=c("Winter", "AVG_Temperature")

#naive benchmark
winter_avg["Naive Benchmark"]=c(NA, winter_avg$AVG_Temperature[-nrow(winter
    _avg)])
```

In this script, once I import the series, I generate, from the original data, the winter average (from 1851 to 2021) and the Naive Benchmark. The latter, is the forecast I will use in the future to compare the precision with the Diebold Mariano test.

Once I define all the data I have to use, I will be able to plot the series and then start with the model selection.

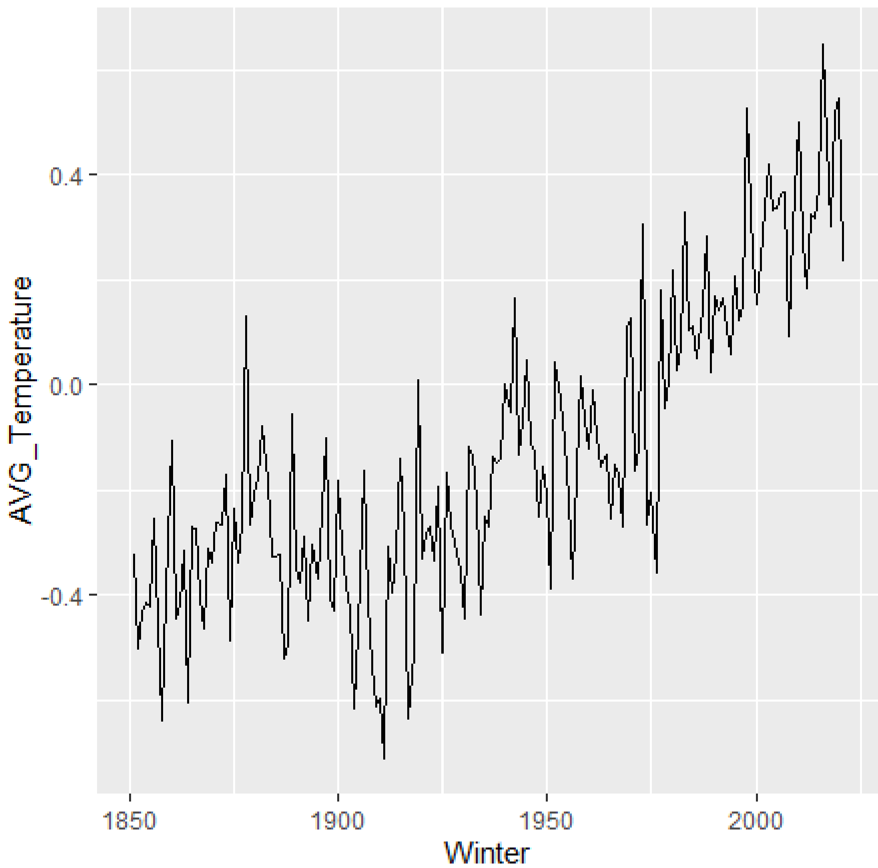


Figura 1: Average winter data (from 1851 to 2021)

2.2 Unit roots test

From Figura 2 is possible to see how the series change the trend from 1980.
This representation suggests the possible presence of a unit root (or more) after 1980.

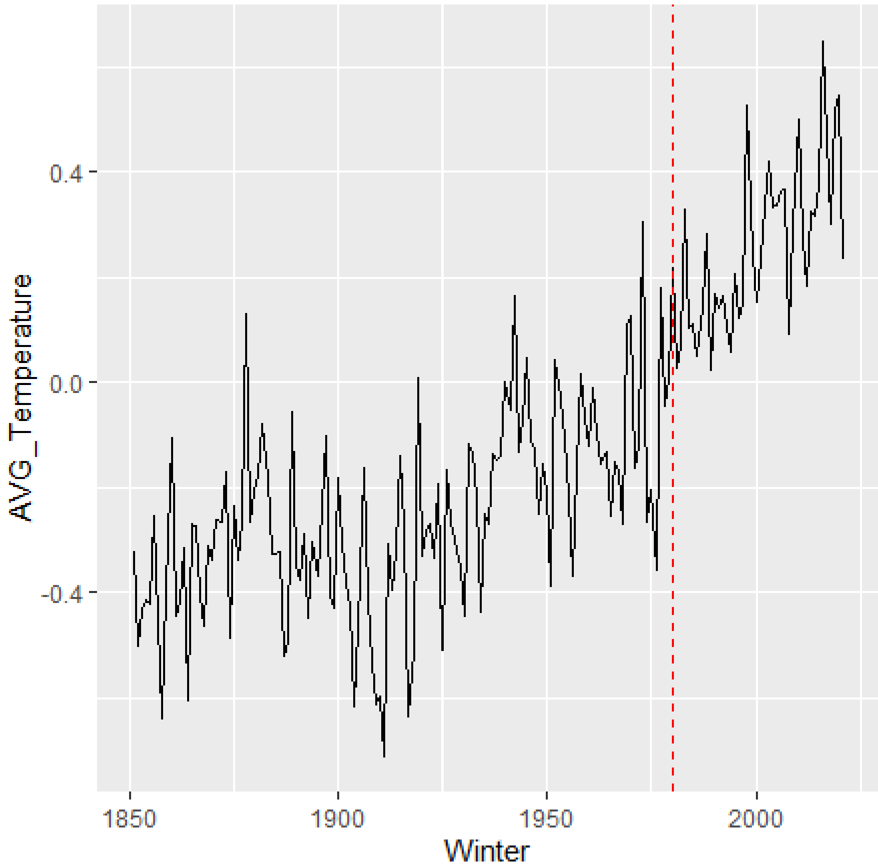


Figura 2: Dashed line underlines the change in trend

In fact, in this period, the global warming starts to affect the world and, as we can see, and the average temperature starts to increase sharply.
To check the presence of the unit root I will use the Augmented Dickey Fuller test. This test has as Null Hypothesis H_0 : {there is Unit Root}, consequently, if the p-value is statistically significant we can reject or not reject the null hypothesis. In our case, I run the test twice:

- ADF for the sample 1851 - 1980 (searching for NO unit root)
- ADF for the sample 1980 - 2021 (searching for unit root)

In the first case the result are written in the following table:

Null Hypothesis: AVG_TEMPERATURE has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.088025	0.0300
Test critical values:	1% level	-3.482453	
	5% level	-2.884291	
	10% level	-2.578981	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(AVG_TEMPERATURE)
Method: Least Squares
Date: 12/10/21 Time: 15:29
Sample (adjusted): 1854 1980
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AVG_TEMPERATURE(-1)	-0.315387	0.102132	-3.088025	0.0025
D(AVG_TEMPERATURE(-1	-0.345935	0.103434	-3.344504	0.0011
D(AVG_TEMPERATURE(-2	-0.293942	0.087857	-3.345694	0.0011
C	-0.073142	0.030092	-2.430602	0.0165

R-squared	0.342355	Mean dependent var	0.005118
Adjusted R-squared	0.326315	S.D. dependent var	0.199505
S.E. of regression	0.163750	Akaike info criterion	-0.749960
Sum squared resid	3.298138	Schwarz criterion	-0.660379
Log likelihood	51.62246	Hannan-Quinn criter.	-0.713564
F-statistic	21.34365	Durbin-Watson stat	2.062442
Prob(F-statistic)	0.000000		

Figura 3: ADF sample 1851 - 1980

It can be observed that, due to the p-value (0.0300) we can easily reject the null hypothesis and consequently confirm the initial suggestion of absence of unit root.

On the other hand, we expect the presence of unit root (at least one) in the second sample (1980-2021).

Null Hypothesis: AVG_TEMPERATURE has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.632115	0.4577
Test critical values:	1% level	-3.596616	
	5% level	-2.933158	
	10% level	-2.604867	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(AVG_TEMPERATURE)
Method: Least Squares
Date: 12/10/21 Time: 15:42
Sample: 1980 2021
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AVG_TEMPERATURE(-1)	-0.252585	0.154760	-1.632115	0.1109
D(AVG_TEMPERATURE(-1	-0.396899	0.161007	-2.465106	0.0183
D(AVG_TEMPERATURE(-2	-0.554417	0.139227	-3.982096	0.0003
C	0.079182	0.041718	1.898057	0.0653

R-squared	0.508926	Mean dependent var	0.005310
Adjusted R-squared	0.470157	S.D. dependent var	0.166764
S.E. of regression	0.121388	Akaike info criterion	-1.289258
Sum squared resid	0.559931	Schwarz criterion	-1.123766
Log likelihood	31.07442	Hannan-Quinn criter.	-1.228599
F-statistic	13.12714	Durbin-Watson stat	2.052716
Prob(F-statistic)	0.000005		

Figura 4: ADF sample 1980 - 2021

In this second sample, we can underline the fact that H_0 is not rejected (p-value: 0.4577) and we can confirm our initial idea. To check the number of unit roots of this sample, we have to repeat the test on the first difference and see if the Null Hypothesis is still not rejected (more than 1 unit root) or if the test will reject H_0: $I(1) \rightarrow$ integrated of order 1.

Null Hypothesis: D(AVG_TEMPERATURE) has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic - based on SIC, maxlag=9)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-10.68804	0.0000
Test critical values:	1% level		-3.596616	
	5% level		-2.933158	
	10% level		-2.604867	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(AVG_TEMPERATURE,2)
Method: Least Squares
Date: 12/10/21 Time: 15:50
Sample: 1980 2021
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(AVG_TEMPERATURE(-1))	-2.220387	0.207745	-10.68804	0.0000
D(AVG_TEMPERATURE(-1)	0.660482	0.125730	5.253195	0.0000
C	0.018459	0.019269	0.957942	0.3440

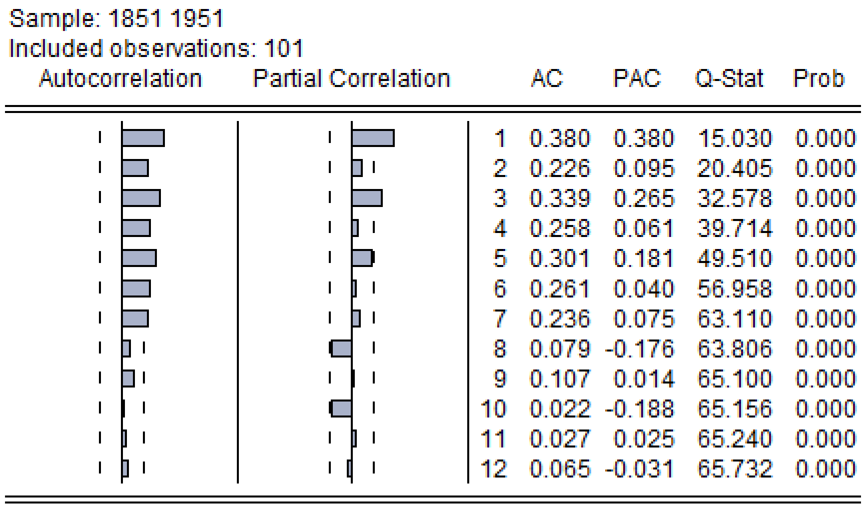
R-squared	0.791551	Mean dependent var	-0.008738
Adjusted R-squared	0.780861	S.D. dependent var	0.264781
S.E. of regression	0.123950	Akaike info criterion	-1.269125
Sum squared resid	0.599182	Schwarz criterion	-1.145006
Log likelihood	29.65163	Hannan-Quinn criter.	-1.223631
F-statistic	74.04793	Durbin-Watson stat	2.112357
Prob(F-statistic)	0.000000		

Figura 5: ADF sample 1980 - 2021 (on first difference)

With this last output (Figura 5) we can conclude the subsection stating that, this series has no unit root ($I(0)$) from 1851 to 1980, but in the last years (1980 - 2021) the process varies from a process integrated of order 0 ($I(0)$) to a process integrated of order 1 ($I(1)$).
This assumption confirms our initial idea about global warming; in fact the change in the series 'starts' when the problem of global warming starts to rise sharply.

2.3 Model Selection

Once I have found out the unit root problem, I can now consider the model selection (ARMA model selection).
I am interested in the period from 1851 to 1951 because from this model, I will then make forecast on the future periods.
The first thing to do, as to find the order of the ARMA model, is to consider the Auto-correlation function (ACF: ρ_j) and the Partial Auto-correlation function (PACF: α_j).



Sample: 1851 1951
Included observations: 101

	AC	PAC	Q-Stat	Prob
1	0.380	0.380	15.030	0.000
2	0.226	0.095	20.405	0.000
3	0.339	0.265	32.578	0.000
4	0.258	0.061	39.714	0.000
5	0.301	0.181	49.510	0.000
6	0.261	0.040	56.958	0.000
7	0.236	0.075	63.110	0.000
8	0.079	-0.176	63.806	0.000
9	0.107	0.014	65.100	0.000
10	0.022	-0.188	65.156	0.000
11	0.027	0.025	65.240	0.000
12	0.065	-0.031	65.732	0.000

Figura 6: ACF and PACF (1851 - 1951)

Observing Figura 6, I have to mention a few concepts about ACF and PACF before:

- in a model $AR(p)$: ρ goes to 0 after p lags while α goes to zero slowly
- in a model $MA(q)$: α goes to 0 after q lags while ρ goes to zero slowly

Looking at these correlograms, once I know that under the dashed lines the value is statistically equal to 0, I can try to select a few models:

- AR(1)
- ARMA(1, 1)
- AR(3).

Once we have selected these 3 models, I will have to study the Information Criteria (IC) in order to define the model with the lowest IC value. In particular with the IC criteria I am going to weigh the additional parameters that will decrease the bias but, at the same time, will increase the variance (parsimonious modelling).

There are two different IC values, depending on the penalty (weight of the parameters)

$$AIC = -2L(\hat{\beta}) + 2(p + q)$$

$$BIC = -2L(\hat{\beta}) + (\ln T)(p + q)$$

In this particular case the values are summarized in the following table:

	AIC	BIC
AR(1)	-0.833677	-0.781573
AR(3)	-0.872826	-0.767317
ARMA(1, 1)	-0.899993	-0.821838

So far I need the model with the lowest IC, the best model is the ARMA(1, 1). In this case both the AIC and BIC confirm the fact, but, if I have to choose one IC, I will follow the information given by the BIC. This decision is due to the fact that AIC, in large samples, may select larger than correct p and q, while BIC makes consistent estimation of p and q.

2.4 Portmanteau test on residuals

After the model selection, I have to test the incorrelation of the residuals. In fact, if the data are really ARMA(1, 1), the residuals $\epsilon_t(\hat{\beta})$ should be very close to the truth ϵ_t. For the residuals I will introduce the abbreviation

$$\hat{\epsilon}_t = \epsilon_t(\hat{\beta})$$

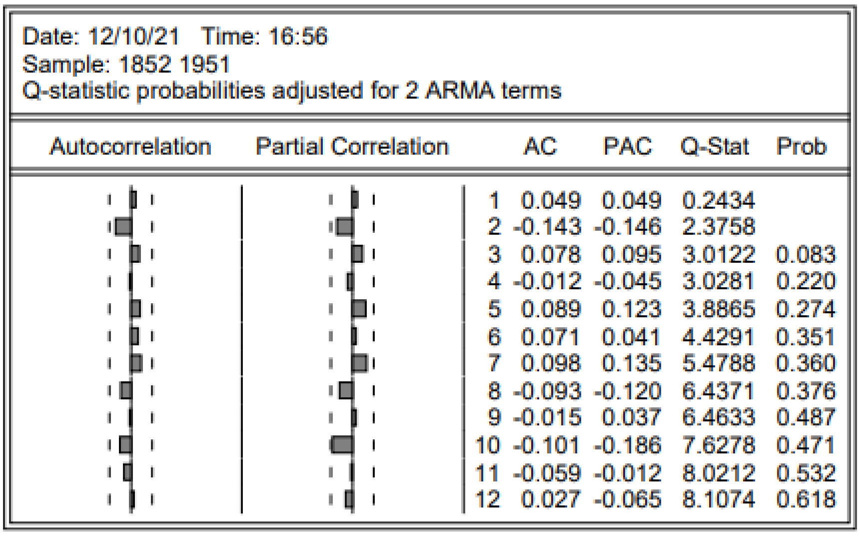
and consider the sample auto-correlation of the residuals

$$r_j = \frac{\frac{1}{T}\sum_{t=j+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-j}}{\frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t^2}$$

then the Portmanteau statistic for the sample auto-correlation has limit distribution

$$T\sum_{j=1}^{k} r_j^2 \xrightarrow{d} \chi^2_{k-(p+q)}$$

If the selected model is correct, I expect the value of the p-value to be always statistically high in order to not reject the Null Hypothesis of H_0 : {no auto-correlation of the residuals}.


Date: 12/10/21 Time: 16:56
Sample: 1852 1951
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.049	0.049	0.2434	
		2	-0.143	-0.146	2.3758	
		3	0.078	0.095	3.0122	0.083
		4	-0.012	-0.045	3.0281	0.220
		5	0.089	0.123	3.8865	0.274
		6	0.071	0.041	4.4291	0.351
		7	0.098	0.135	5.4788	0.360
		8	-0.093	-0.120	6.4371	0.376
		9	-0.015	0.037	6.4633	0.487
		10	-0.101	-0.186	7.6278	0.471
		11	-0.059	-0.012	8.0212	0.532
		12	0.027	-0.065	8.1074	0.618

Figura 7: Portmanteau Test for residuals of an ARMA(1, 1)

As we can see from Figura 7, the bar of the bar chart are always under the dashed line and, consequently, statistically equal to 0. Moreover, the p-values of the Q-statistic are always over 0.05 (5%) which means we never reject the null hypothesis at the percentage level mentioned before.

2.5 Diebold and Mariano test

I am interested in forecasting $Y_{1952-2021}$ using information available at time 1851-1951, and I have two possible forecasts, the ARMA(1, 1) one and the Naive Benchmark forecast. We can then compare these forecasts looking at the errors made with each one.
Since we have already found the unit roots after 1980, we can divide this comparison in two different groups:

- DM test for foreacast in the period 1952-1981
- DM test for foreacast in the period 1982-2021

Denote these forecast errors of the first period as

$$e^{arma}_{1952-1981} = Y_{1952-1981} - \hat{Y}^{arma}_{1952-1981}$$

$$u^{naive}_{1952-1981} = Y_{1952-1981} - \hat{Y}^{naive}_{1952-1981}$$

and the same for the second time period

$$e^{arma}_{1982-2021} = Y_{1982-2021} - \hat{Y}^{arma}_{1982-2021}$$

$$u^{naive}_{1982-2021} = Y_{1982-2021} - \hat{Y}^{naive}_{1982-2021}.$$

A criterion to choose a forecast is that it should have small Mean Square Errors (MSE). Therefore, I want to compare $(e^{arma}_{1952-1981})^2$ and $(u^{naive}_{1952-1981})^2$ (and for the second time period).

$$d_{1952-1981} = (u^{naive}_{1952-1981})^2 - (e^{arma}_{1952-1981})^2$$

Two forecasts have equal predictive ability if $E[d_{1952-1981}] = 0$.
If $E[d_{1952-1981}] > 0$, then $u^{naive}_{1982-2021}$ is better; if $E[d_{1952-1981}] < 0$, then $e^{arma}_{1952-1981}$ is better.

In order to statistically check this result, we can run a Diebold and Mariano test. Under the Null Hypothesis $H_0 : \{E[d_t] = 0\}$, the t-statistic is

$$\sqrt{T}\frac{\bar{d} - E[d_t]}{\sqrt{L\hat{R}V(d_t)}} \xrightarrow{d} N(0, 1)$$

In practice, I will run a regression of d_t on the constant (only) and using the Bartlett Kernel estimate of the Long Run Variance (LRV). In this way, the p-value will tell me if the expected value of d_t is statistically different from zero ($H_0 : \{E[d_t] = 0\}$), that means not to reject the null hypothesis and see which estimator is better; if the coefficient of C is positive, $e^{arma} < u^{naive}$, ARMA(1, 1) has a higher forecasting power; otherwise, if $e^{arma} > u^{naive}$, the coefficient of the constant will be negative and consequently the Naive Forecast has higher predictive power. Of course, all of these suggestions are valid if and only if I reject the Null Hypothesis of $E[d_t] = 0$, in the other hand, if the p-value is higher than 0.05, the two estimators have a similar predictive power and they are both efficient.

The following are the result of the DM test for the two different periods, before and after 1981.

Dependent Variable: ENAIVE^2-EARMA11^2
Method: Least Squares
Date: 12/04/21 Time: 09:59
Sample: 1952 1981
Included observations: 30
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.017303	0.010649	1.624914	0.1150

R-squared	-0.000000	Mean dependent var	0.017303
Adjusted R-squared	-0.000000	S.D. dependent var	0.067145
S.E. of regression	0.067145	Akaike info criterion	-2.531146
Sum squared resid	0.130747	Schwarz criterion	-2.484440
Log likelihood	38.96720	Durbin-Watson stat	2.326820

Figura 8: Diebold and Mariano Test (1952 - 1981)

As we can see from Figura 8, the p-value is higher than 0.05 (5%) which means I don't reject the null hypothesis of $E[d_t] = 0$. The two estimators, from 1952 to 1981 have the similar predictive power but, if I have to choose one, I will choose the ARMA(1, 1) model; in fact the coefficient on the constant is positive, as a consequence of the fact that the residuals of the model are lower than the residuals of the Naive Benchmark.

Let's now consider the other time period, 1982-2021

Dependent Variable: ENAIVE^2-EARMA11^2
Method: Least Squares
Date: 12/04/21 Time: 10:00
Sample: 1982 2021
Included observations: 40
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.025817	0.007386	-3.495356	0.0012

R-squared	-0.000000	Mean dependent var	-0.025817
Adjusted R-squared	-0.000000	S.D. dependent var	0.056330
S.E. of regression	0.056330	Akaike info criterion	-2.890509
Sum squared resid	0.123748	Schwarz criterion	-2.848287
Log likelihood	58.81018	Durbin-Watson stat	2.098098

Figura 9: Diebold and Mariano Test (1982 - 2021)

The results of the overview above (Figura 9) are different. This time the p-value is near 0 and lower than 0.05 (5%), consequently I reject the null hypothesis and, looking at the coefficient on the constant I can define the Naive Benchmark as better than the ARMA model for the prediction between 1982 and 2021.
This is what I expected since I saw the plot of the Series (Figura 1). With the ARMA

model I've trained a model on a series with no unit roots, stationary, from 1852 to 1951. How can the model know that after 1980 there would be change? The consequence, in fact, is underlined in the Diebold and Mariano tests: till 1980, the series is similar to the past and both estimators are good, but from 1980 the trend change and the ARMA forecasts can't follow it. The power of the Naive Benchmark is that it will always follow the observation of the year before, without loosing the trend of the series, if there is one.

3 Conclusion

The purpose of the project was to make predictions on the future temperature and see how two different types of estimation work on it. In this particular case, because of the model selection, I choose an ARMA model with $p = 1$ and $q = 1$ (ARMA(1, 1)), to be compared with a Naive Benchmark, where the forecast is the last available observation. Figura 2 clearly marks the difference of the series before and after 1980; such change is due to the development of the Global Warming. In a statistic way this change is due to the presence of the unit root, $I(1)$, after this critical year. Despite of that, the project continues with the forecast evaluation (Diebold and Mariano Test).
The latter has been split into two time periods:

- 1952-1981 where the ARMA model forecasts work better
- 1982-2021 where the Naive forecasts work better

From the following plot we can understand why and how the two forecasts are different.

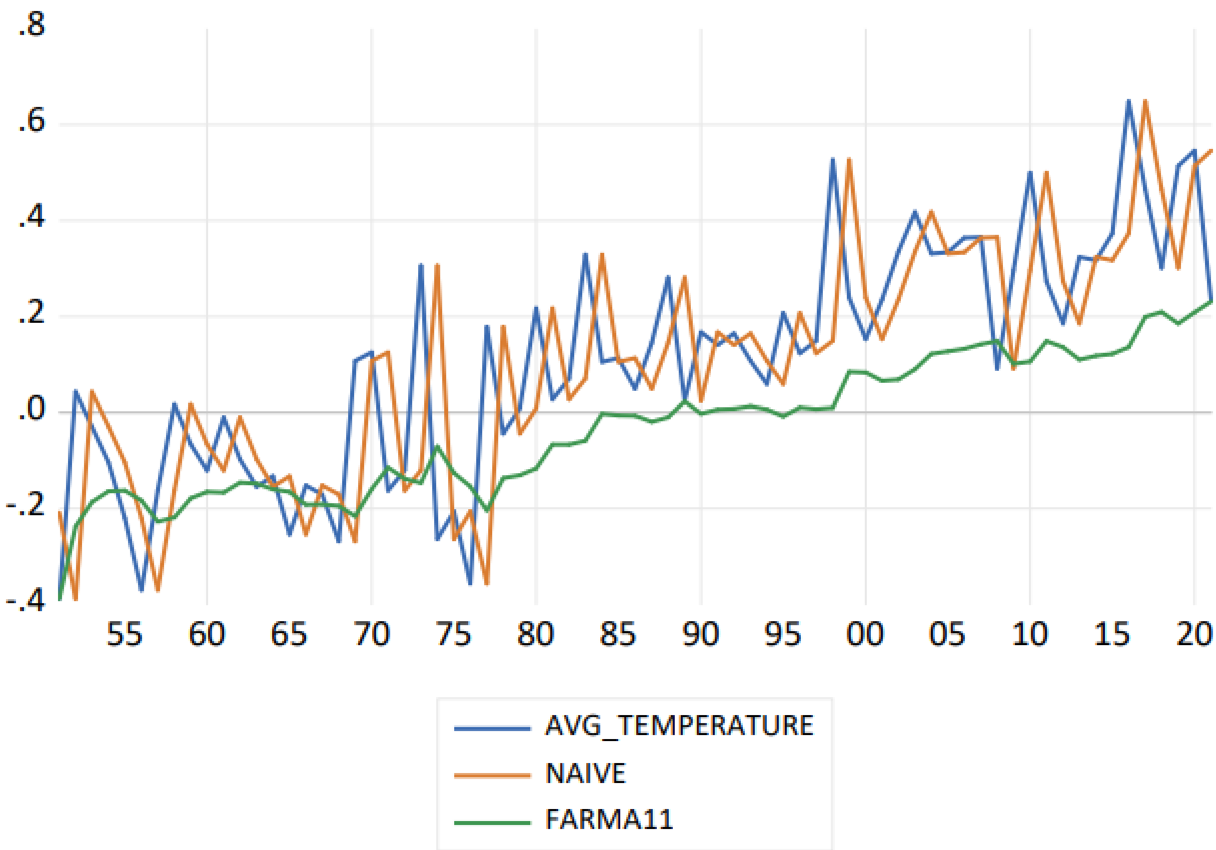


Figura 10: Forecasts 1852 - 2021

Figura 10 shows well how the naive forecast follows the series while the ARMA's one doesn't. The residuals are consequently different between the two different time periods (before and after 1980) and so is the Diebold and Mariano Test (Figura 8 and Figura 9). The last thing to say is that I chose 1980 as explanatory year for the change of temperature, but this, of course, is just a personal suggestion. In fact Global Warming is a constant problem and consequently selecting only one year could cause loss of information on the predictive power. Despite that, all the initial assumption on the change of temperature has been confirmed by this empirical project.