Dynamic Relationships in Environmental and Energy Data: A Time Series Approach

CHITA Ionut-Cristian, COJOCARITA Cristian

Contents

1 Introduction

It is essential to comprehend how temperature, energy use, and air quality interact in order to combat global warming and enhance environmental health. In order to examine these factors this study focuses on three important time series datasets. We investigate daily temperature, air quality (PM2.5 levels), and thermal energy balance in an effort to find trends and connections that illustrate how human activity affects the environment.

1.1 Balance of Thermal Energy by Component Elements

- **Definition:** Tracks the flow of thermal energy through various stages, including primary production, imports, exports, and consumption by sectors like industry, transport, and households. It also accounts for losses and statistical differences.
- **Periodicity:** Semi-annually, from 1992 to 2022
- **Measurement Unit:** Terajoules
- **Source:** Tempo Online: **IND111A**

Contextualization:

For the purpose of creating comprehensive energy policies and plans that aim for sustainability, it is essential to comprehend the balance of thermal energy by its constituent parts. This information gives us an understanding of the efficiency and environmental effect of energy consumption by allowing us to examine how energy is generated, used, and lost across many sectors. The significance of this analysis is emphasized by a number of important research and reports.

First off, the Reveiu et al. (2015) study emphasizes how important it is to incorporate consumer behavior into techno-economic energy models. The study shows that conventional models, which solely take into account technological and economic aspects, frequently fall short of accurately capturing the real dynamics of energy usage. These models can anticipate energy consumption and spot potential for efficiency gains more precisely by include behavioral data. This method, which takes into account both technological and behavioral factors that affect energy usage and losses, is especially pertinent to comprehending the thermal energy balance [1].

Furthermore, Gherhes and Fărcas's study from 2021 on sustainable behavior among Romanian students offers a more comprehensive explanation of the significance of monitoring thermal energy balance. According to their research, behavior modifications can result in considerable energy savings and residential energy usage is a key contributor to total energy use. Having a thorough understanding of the precise thermal energy flow, including the locations of losses, may aid in the creation of focused interventions that cut waste and boost productivity. Promoting energy-efficient equipment and practices, for example, may greatly reduce thermal energy consumption and losses in homes [2].

In addition, a thorough assessment of techno-economic models is included in the paper of Timmerman et al. (2014), which classifies several models that aid in the evaluation of energy systems. They contend that these models may be made much more accurate by incorporating both current energy flow data and long-term estimates. The requirement for comprehensive thermal energy balance data is supported by this method, which may help guide current and future energy policy [3].

The balance of thermal energy by component elements serves as a tactical instrument for policymakers. We can better address the issues around energy sustainability if we have a greater understanding of the production, transmission, and consumption of energy across many industries. Developing policies that lower greenhouse gas emissions, increase energy efficiency, and encourage the use of renewable energy sources requires a comprehensive approach.

References:

1. Reveiu, A., Smeureanu, I., Dardala, M., Kanala, R. (2015). Modelling Domestic Lighting Energy Consumption in Romania by Integrating Consumers Behavior. *Procedia Computer Science*, 52, 812-818.

2. Gherhes, V., Fărcas, M. A. (2021). Sustainable Behavior among Romanian Students: A Perspective on Electricity Consumption in Households. *Sustainability*, 13(9357), 1-17.

3. Timmerman, J., Vandevelde, L., Van Eetvelde, G. (2014). Towards low carbon business park energy systems: Classification of techno-economic energy models. *Energy*, 75, 68-80.

1.2 Popesti-Leordeni AQI PM2.5 Daily

- **Definition:** The Air Quality Index (AQI) for PM2.5 represents the daily average concentration of particulate matter smaller than 2.5 micrometers in diameter. This index indicates air quality and potential health effects.

- **Periodicity:** Daily

- **Measurement Unit:** Micrograms per cubic meter (**µg/m³**)
- **Source:** AQICN Popesti-Leordeni: **https://aqicn.org/station/@109861**

Contextualization:

An important environmental health indicator is air quality, specifically the amount of fine particulate matter (PM2.5) in the atmosphere. Tracking PM2.5 levels helps determine the efficacy of air quality-improving programs as well as changes in air pollution. The significance of monitoring PM2.5 concentrations and their effects on the environment and public health is highlighted in a number of research and reports.

A thorough inventory of Romania's principal air pollutants and air quality was carried out by Năstase et al. (2018), who also noted the serious concerns that air pollution poses to the environment, the economy, food security, and public health. The report highlights that the production and use of fossil fuels is the primary source of most pollutants, including PM2.5. Romania's consumption of fossil fuels has decreased during the last ten years, mostly as a result of EU rules that promote the use of renewable energy sources. PM emissions have decreased as a result of this change, but ongoing observation is necessary to assess the long-term efficacy of these efforts.[4]

A research by Burghele et al. (2021) on interior air quality in energy-efficient buildings in Romania emphasizes the intricacies of air pollution, which further supports the need for air quality monitoring. According to the study, if energy saving measures in buildings are not adequately maintained, they may unintentionally result in poor indoor air quality. This is especially important for PM2.5 since poorly ventilated, highly insulated buildings can trap pollutants within, increasing the danger to health.[5]

In addition, because PM2.5 may enter the bloodstream and pierce deeply into the lungs, the World Health Organization (WHO) lists it as one of the most dangerous air pollutants. Long-term exposure to PM2.5 has been associated with higher death rates as well as cardiovascular and respiratory disorders. Thus, monitoring daily PM2.5 levels is essential for public health surveillance and for putting appropriate actions in place to lessen its effects.[6]

In conclusion, this information supports strategies to lower emissions and exposure to dangerous pollutants, as well as aids in the identification of pollution sources and the assessment of the efficacy of air quality legislation.

References:

1. Năstase, G., Șerban, A., Năstase, A.F., Dragomir, G., Brezeanu, A.I. (2018). Air quality, primary air pollutants and ambient concentrations inventory for Romania. Atmospheric Environment.
2. Burghele, B.D., et al. (2021). Comprehensive survey on radon mitigation and indoor air quality in energy-efficient buildings from Romania. Science of the Total Environment, 751, 141858.
3. World Health Organization (WHO). (2018). Air quality guidelines for particulate matter.

These sources offer a thorough grasp of the significance of controlling and monitoring air quality, especially PM2.5 levels, in order to protect the environment and human health.

1.3 Popesti-Leordeni Temperature Daily

- **Definition:** The daily average temperature measurement records the atmospheric temperature at a specific location, providing insights into daily climatic conditions.
- **Periodicity:** Daily
- **Measurement Unit:** Celsius

- **Source:** AQICN Popesti-Leordeni **https://aqicn.org/station/@109861**

Contextualization:

It is essential to monitor daily temperature data in order to comprehend climatic variability and recognize long-term patterns. Numerous research works have emphasized the need of temperature monitoring in evaluating the effects of climate change, particularly in areas such as Romania where notable warming has been seen in recent years.

Noteworthy trends in annual air temperature were found in a comprehensive research on climate changes in Romania by Marin et al. (2014). Other noteworthy trends in climate variables included precipitation, sunlight hours, and cloud cover. The study employed the Mann-Kendall trend test and Kendall-Theil technique, utilizing data from all accessible meteorological stations in Romania from 1961 to 2013, to identify statistically significant increases in air temperature throughout the nation. The results show a general trend toward warming, with the air temperature showing the most shifts—it rose at every location under analysis. The region is being impacted by greater global warming tendencies, which are highlighted by this steady increase in temperature.[7]

In a similar vein, Birsan et al. (2019) employed the Mann-Kendall test on gridded daily data from the ROCADA dataset covering the years 1961 to 2013 to examine variations in yearly temperature extremes over Romania. The analysis discovered rising trends in warm-related extremes and declining trends in cold-related indicators, such as the quantity of frost days. The asymmetric patterns of daylight vs nocturnal warming are highlighted by these shifts, which are especially noticeable in lowland areas. The study highlights the significance of geographic factors in climate variability by indicating that the observed warming trends are more strongly correlated with altitude than latitude.[8]

These results are further supported by Piticar and Ristoiu's (2012) study on the evolution of air temperature in Northeastern Romania. Their examination of data spanning 50 years (1961–2010) revealed notable upward trends in air temperature, especially in the summer months of June, July, and August. The robustness of these warming patterns was verified by the use of many statistical tests, such as Sen's slope estimation and the Mann-Kendall test. The height of the weather stations has an impact on temperature trends, with lower elevation stations exhibiting more dramatic warming, according to the study's hierarchical cluster analysis.[9]

All of these studies emphasize how important it is to track and examine daily temperature data in order to comprehend the dynamics of the local climate. Romania's steady warming tendencies are consistent with patterns of climate change worldwide, which offers crucial information for creating mitigation and adaptation plans.

References:

1. Marin, L., Birsan, M.V., Bojariu, R., Dumitrescu, A., Micu, D.M., Manea, A. (2014). An overview of annual climatic changes in Romania: trends in air temperature, precipitation, sunshine hours, cloud cover, relative humidity and wind speed during the 1961–2013 period. Carpathian Journal of Earth and Environmental Sciences, 9(4), 253-258.

2. Birsan, M.V., Micu, D.M., Niță, I.A., Mateescu, E., Szép, R., Keresztesi, Á. (2019). Spatio-temporal changes in annual temperature extremes over Romania (1961–2013). Romanian Journal of Physics, 64, 816

3. Piticar, A., Ristoiu, D. (2012). Analysis of air temperature evolution in northeastern Romania and evidence of warming trend. Carpathian Journal of Earth and Environmental Sciences, 7(4), 97-106

2 Application 1: ARMA/ARIMA Models in EViews

2.1 Unit Root Tests:

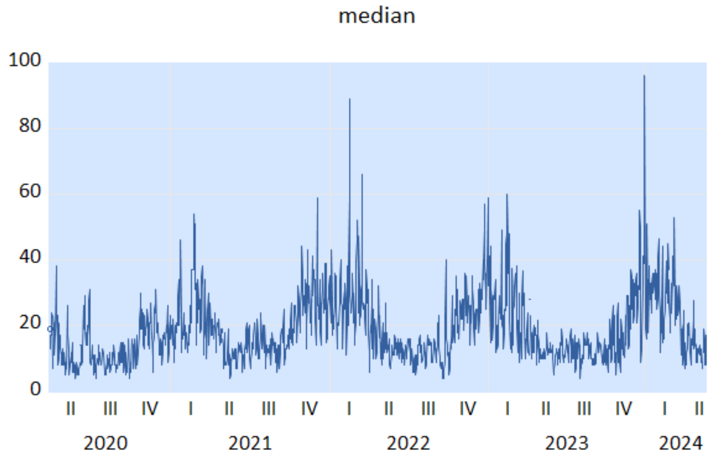


Figure 1: Air Quality Index in Bucharest (Popesti-Leordeni)

The AQI fluctuates significantly between 2020 and 2024, as seen in Figure 1, with values ranging from 0 to over 100. The spikes show times when the quality of the air was low, perhaps as a result of things like increased industrial activity or vehicle emissions.

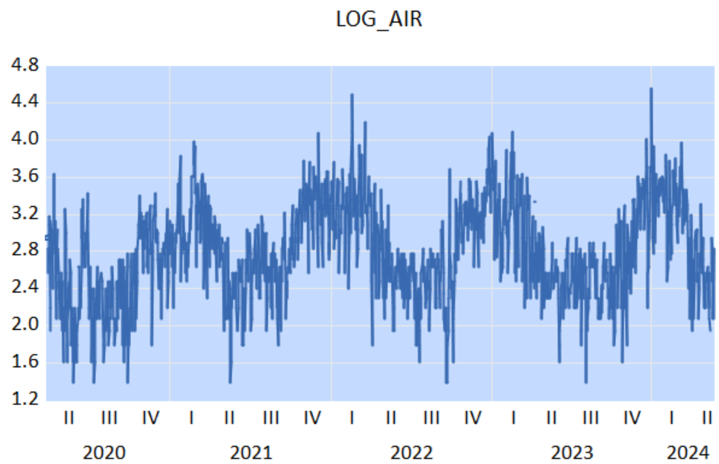


Figure 2: Log Transformation of Air Quality Index

The AQI's log transformation is displayed in Figure 2, which reduces higher values and increases lower ones to stabilize the variance. Though variations persist, this change makes underlying patterns more evident.

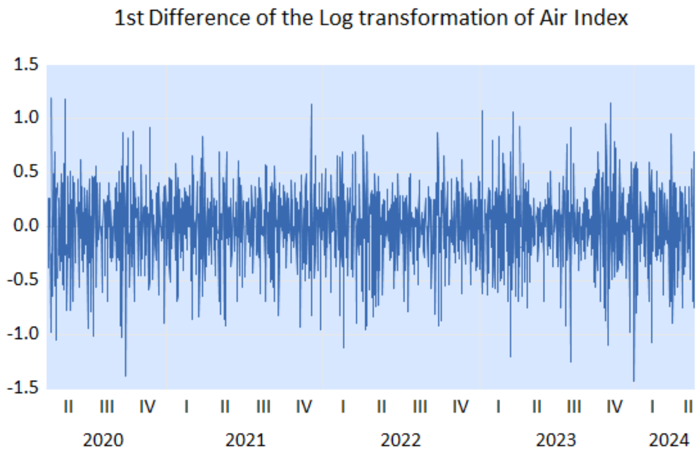
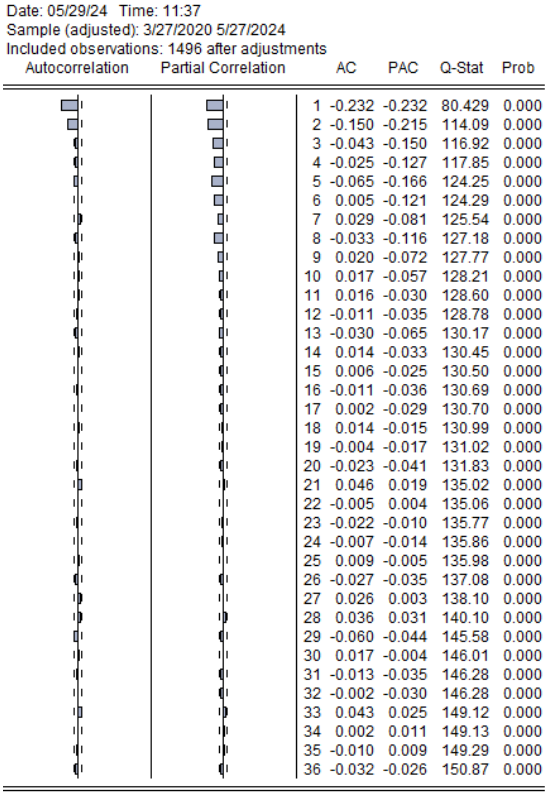


Figure 3: 1st Difference of the Log Transformation of Air Index

The initial difference of the log-transformed AQI, which eliminates trends and increases series stationaryity, is shown in Figure 3. The values show more consistent fluctuations over time and a more steady mean as they oscillate about zero.

- Original Series: High variability and significant spikes in AQI.
- Log Transformation: Stabilizes variance and clarifies patterns.
- First Difference of Log Transformation: Stabilizes mean, suggesting closer stationarity. Further tests are needed to confirm.

These transformations help prepare the AQI data for accurate time series analysis and forecasting.



Date: 05/29/24 Time: 11:37
Sample (adjusted): 3/27/2020 5/27/2024
Included observations: 1496 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.232	-0.232	80.429	0.000
		2	-0.150	-0.215	114.09	0.000
		3	-0.043	-0.150	116.92	0.000
		4	-0.025	-0.127	117.85	0.000
		5	-0.065	-0.166	124.25	0.000
		6	0.005	-0.121	124.29	0.000
		7	0.029	-0.081	125.54	0.000
		8	-0.033	-0.116	127.18	0.000
		9	0.020	-0.072	127.77	0.000
		10	0.017	-0.057	128.21	0.000
		11	0.016	-0.030	128.60	0.000
		12	-0.011	-0.035	128.78	0.000
		13	-0.030	-0.065	130.17	0.000
		14	0.014	-0.033	130.45	0.000
		15	0.006	-0.025	130.50	0.000
		16	-0.011	-0.036	130.69	0.000
		17	0.002	-0.029	130.70	0.000
		18	0.014	-0.015	130.99	0.000
		19	-0.004	-0.017	131.02	0.000
		20	-0.023	-0.041	131.83	0.000
		21	0.046	0.019	135.02	0.000
		22	-0.005	0.004	135.06	0.000
		23	-0.022	-0.010	135.77	0.000
		24	-0.007	-0.014	135.86	0.000
		25	0.009	-0.005	135.98	0.000
		26	-0.027	-0.035	137.08	0.000
		27	0.026	0.003	138.10	0.000
		28	0.036	0.031	140.10	0.000
		29	-0.060	-0.044	145.58	0.000
		30	0.017	-0.004	146.01	0.000
		31	-0.013	-0.035	146.28	0.000
		32	-0.002	-0.030	146.28	0.000
		33	0.043	0.025	149.12	0.000
		34	0.002	0.011	149.13	0.000
		35	-0.010	0.009	149.29	0.000
		36	-0.032	-0.026	150.87	0.000

Figure 4: SACF and SPACF of the transformed Time Series

Correlogram of D_LOG_AIR

- **ACF:** Significant spikes at initial lags, indicating autocorrelation.
- **PACF:** Significant spikes at initial lags, suggesting AR and/or MA components.
- **Interpretation:** The ACF and PACF show significant values, indicating the presence of AR and/or MA components in the data.

Figure 5: ProbADF calculated for AIR Index in Popesti-Leordeni

We can observe that the PP test, the ADF test, and the KPSS test all point to a stationary time series from the table that combines the p-values of the ADF, Philips-Peron, and KPSS tests. Therefore, we will choose to proceed with an integration order of d = 1. It is implied that log AIRt ARIMA(p, 1, q) for our consideration of ARIMA models.

ProbADF	ProbPP	ProbKPSS
< 0.05	0.01	0.10

We can find below a quick reminder for the hypothesis related to the ADF and the PP tests4 for a given time series:

$$\begin{cases} \mathrm{H0} : \{X_t\}_t \sim \mathbf{I}(1) \\ \mathrm{H1} : \{X_t\}_t \sim \mathbf{I}(0) \end{cases}$$

Null Hypothesis: D_LOG_AIR has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-21.82450	0.0000
Test critical values:	1% level	-3.434537	
	5% level	-2.863276	
	10% level	-2.567743	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(D_LOG_AIR)
Method: Least Squares
Date: 05/29/24 Time: 11:40
Sample (adjusted): 4/05/2020 5/27/2024
Included observations: 1488 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D_LOG_AIR(-1)	-3.050013	0.139752	-21.82450	0.0000
D(D_LOG_AIR(-1))	1.656977	0.128653	12.87948	0.0000
D(D_LOG_AIR(-2))	1.285140	0.114314	11.24220	0.0000
D(D_LOG_AIR(-3))	0.975440	0.097773	9.976533	0.0000
D(D_LOG_AIR(-4))	0.706069	0.080530	8.767726	0.0000
D(D_LOG_AIR(-5))	0.435182	0.062533	6.959232	0.0000
D(D_LOG_AIR(-6))	0.240186	0.043965	5.463140	0.0000
D(D_LOG_AIR(-7))	0.116461	0.025783	4.517025	0.0000
C	-0.000861	0.008075	-0.106672	0.9151

R-squared	0.669959	Mean dependent var	-0.001306
Adjusted R-squared	0.668174	S.D. dependent var	0.540712
S.E. of regression	0.311474	Akaike info criterion	0.511026
Sum squared resid	143.4864	Schwarz criterion	0.543114
Log likelihood	-371.2033	Hannan-Quinn criter.	0.522985
F-statistic	375.2828	Durbin-Watson stat	2.014360
Prob(F-statistic)	0.000000		

Null Hypothesis: D_LOG_AIR has a unit root
Exogenous: Constant
Bandwidth: 244 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-166.9865	0.0001
Test critical values:	1% level	-3.434517	
	5% level	-2.863267	
	10% level	-2.567738	

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

Residual variance (no correction)	0.112903
HAC corrected variance (Bartlett kernel)	0.003818

Phillips-Perron Test Equation
Dependent Variable: D(D_LOG_AIR)
Method: Least Squares
Date: 05/29/24 Time: 11:42
Sample (adjusted): 3/28/2020 5/27/2024
Included observations: 1495 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D_LOG_AIR(-1)	-1.232375	0.025204	-48.89682	0.0000
C	-0.000342	0.008696	-0.039333	0.9686

R-squared	0.615592	Mean dependent var	-0.000250
Adjusted R-squared	0.615335	S.D. dependent var	0.542129
S.E. of regression	0.336236	Akaike info criterion	0.659328
Sum squared resid	168.7902	Schwarz criterion	0.666431
Log likelihood	-490.8473	Hannan-Quinn criter.	0.661974
F-statistic	2390.899	Durbin-Watson stat	2.096688
Prob(F-statistic)	0.000000		

Null Hypothesis: D_LOG_AIR is stationary
Exogenous: Constant
Bandwidth: 165 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.060102
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.119348
HAC corrected variance (Bartlett kernel)	0.004985

KPSS Test Equation
Dependent Variable: D_LOG_AIR
Method: Least Squares
Date: 05/29/24 Time: 11:43
Sample (adjusted): 3/27/2020 5/27/2024
Included observations: 1496 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000578	0.008935	-0.064714	0.9484

R-squared	0.000000	Mean dependent var	-0.000578
Adjusted R-squared	0.000000	S.D. dependent var	0.345583
S.E. of regression	0.345583	Akaike info criterion	0.713500
Sum squared resid	178.5442	Schwarz criterion	0.717050
Log likelihood	-532.6981	Hannan-Quinn criter.	0.714823
Durbin-Watson stat	2.459288		

Figure 6: ProbADF,ProbPP,ProbKPSS computed for Log Air Index Difference of Transformation

Augmented Dickey-Fuller Unit Root Test on MEDIAN

Null Hypothesis: MEDIAN has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.144023	0.0008
Test critical values:	1% level	-3.434543	
	5% level	-2.863279	
	10% level	-2.567744	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MEDIAN)
Method: Least Squares
Date: 05/29/24 Time: 00:11
Sample (adjusted): 4/07/2020 5/27/2024
Included observations: 1486 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MEDIAN(-1)	-0.088962	0.021467	-4.144023	0.0000
D(MEDIAN(-1))	-0.354216	0.031725	-11.16527	0.0000
D(MEDIAN(-2))	-0.359675	0.032479	-11.07394	0.0000
D(MEDIAN(-3))	-0.305408	0.033122	-9.220770	0.0000
D(MEDIAN(-4))	-0.297662	0.033407	-8.910061	0.0000
D(MEDIAN(-5))	-0.263374	0.033147	-7.945745	0.0000
D(MEDIAN(-6))	-0.246425	0.032585	-7.562489	0.0000
D(MEDIAN(-7))	-0.164388	0.031634	-5.196620	0.0000
D(MEDIAN(-8))	-0.164029	0.030221	-5.427725	0.0000
D(MEDIAN(-9))	-0.143670	0.028169	-5.100313	0.0000
D(MEDIAN(-10))	-0.068574	0.025953	-2.642239	0.0083
C	1.635444	0.431680	3.788553	0.0002

R-squared	0.220866	Mean dependent var	-0.008748
Adjusted R-squared	0.215051	S.D. dependent var	7.186659
S.E. of regression	6.367189	Akaike info criterion	6.548236
Sum squared resid	59757.58	Schwarz criterion	6.591066
Log likelihood	-4853.339	Hannan-Quinn criter.	6.564199
F-statistic	37.98572	Durbin-Watson stat	1.996202
Prob(F-statistic)	0.000000		

1. **Stationarity:** The ADF test results indicate that the AIR index (median) is stationary after differencing. This is evidenced by the ADF test statistic being lower than the critical values and the p-value being below 0.05.

2. **Significant Lags:** The test equation shows that several lagged differences of the median AQI are significant, with p-values less than 0.05, indicating that past values significantly affect the current differenced value.

3. **Model Fit:** The R-squared value of 0.220866 suggests that approximately 22.1% of the variance in the differenced series is explained by the model. While this indicates some explanatory power, there is still substantial unexplained variance.

2.2 Box-Jenkins methodology and model validity

Using the auto.arima function, the best model is indeed confirming our intuition. Based on AIC, BIC, and LL criterions the best model is ARIMA(1, 1, 1) with drift. Since the program may not be correct every time, we will keep using ARIMA(1, 1, 0) with drift, based on our intuitions, and compare the two models. First, looking at the criteria, ARIMA(1, 1, 1) is indeed slightly better.

From the estimation output, we obtain the following model for ARIMA(1, 1, 1):

```
LOG_AIR = 2.79014008875 + 0.500059151561*D(LOG_AIR) +
[AR(1)=0.769568358748,MA(1)=0.999999220557,UNCOND,ESTSMPL="3/27/2020 5/27/2024"]
```

For ARIMA(1, 1, 0):

```
LOG_AIR = 2.78935651999 + 0.500546810534*D(LOG_AIR) + [AR(1)=0.899799843248,UNCOND]
```

We must now verify the two models' validity because, as the ACF functions in Figure 8 and 9 show, the autocorrelation is negligible. As a result, the standard errors are accurately computed, suggesting that we may easily verify if the coefficients are true.

Date: 05/29/24 Time: 12:18
Sample (adjusted): 3/27/2020 5/27/2024
Q-statistic probabilities adjusted for 2 ARMA terms and 1 dynamic regressor

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.102	-0.102	15.522	
		2	-0.044	-0.055	18.487	
		3	0.039	0.029	20.776	0.000
		4	0.052	0.058	24.876	0.000
		5	0.012	0.027	25.084	0.000
		6	0.074	0.083	33.227	0.000
		7	0.098	0.116	47.705	0.000
		8	0.042	0.072	50.374	0.000
		9	0.089	0.110	62.264	0.000
		10	0.086	0.107	73.424	0.000
		11	0.081	0.107	83.375	0.000
		12	0.061	0.085	88.981	0.000
		13	0.036	0.042	90.903	0.000
		14	0.080	0.073	100.52	0.000
		15	0.068	0.063	107.57	0.000
		16	0.054	0.044	111.95	0.000
		17	0.065	0.045	118.31	0.000
		18	0.075	0.050	126.87	0.000
		19	0.056	0.036	131.62	0.000
		20	0.042	0.014	134.33	0.000
		21	0.099	0.062	149.31	0.000
		22	0.056	0.033	154.01	0.000
		23	0.034	0.007	155.79	0.000
		24	0.052	0.008	159.95	0.000
		25	0.059	0.010	165.32	0.000
		26	0.033	-0.012	166.94	0.000
		27	0.075	0.024	175.62	0.000
		28	0.086	0.040	187.02	0.000
		29	-0.003	-0.038	187.04	0.000
		30	0.068	0.014	194.21	0.000
		31	0.037	-0.019	196.26	0.000
		32	0.054	0.002	200.72	0.000
		33	0.086	0.045	212.13	0.000
		34	0.055	0.023	216.85	0.000
		35	0.038	0.009	219.02	0.000
		36	0.020	-0.021	219.62	0.000

*Probabilities may not be valid for this equation specification.

Dependent Variable: LOG_AIR
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/29/24 Time: 12:12
Sample: 3/27/2020 5/27/2024
Included observations: 1496
Failure to improve objective (non-zero gradients) after 16 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.790140	0.037523	74.35878	0.0000
D(LOG_AIR)	0.500059	0.000832	601.2534	0.0000
AR(1)	0.769568	0.016064	47.90503	0.0000
MA(1)	0.999999	1.851419	0.540126	0.5892
SIGMASQ	0.026417	0.001540	17.15696	0.0000

R-squared	0.898109	Mean dependent var	2.791271
Adjusted R-squared	0.897836	S.D. dependent var	0.509357
S.E. of regression	0.162806	Akaike info criterion	-0.782926
Sum squared resid	39.52028	Schwarz criterion	-0.765177
Log likelihood	590.6287	Hannan-Quinn criter.	-0.776313
F-statistic	3285.583	Durbin-Watson stat	2.200044
Prob(F-statistic)	0.000000		

Inverted AR Roots	.77
Inverted MA Roots	-1.00

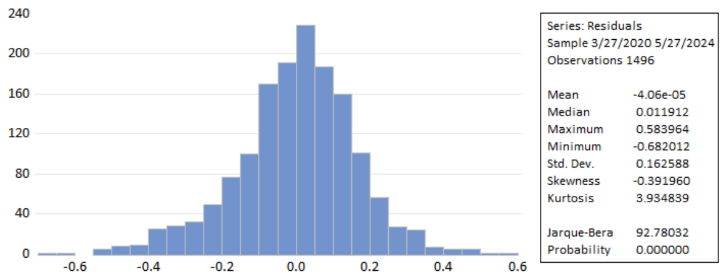


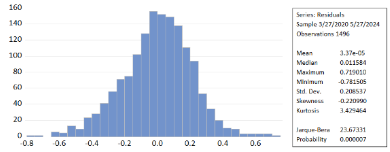
Figure 7: Residuals from ARIMA(1,1,1)

Dependent Variable: LOG_AIR
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/29/24 Time: 14:00
Sample: 3/27/2020 5/27/2024
Included observations: 1496
Convergence achieved after 4 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.789357	0.054406	51.26947	0.0000
D(LOG_AIR)	0.500547	0.008858	56.50653	0.0000
AR(1)	0.899800	0.011184	80.45460	0.0000
SIGMASQ	0.043459	0.001459	29.77932	0.0000

R-squared	0.832381	Mean dependent var	2.791271
Adjusted R-squared	0.832044	S.D. dependent var	0.509357
S.E. of regression	0.208747	Akaike info criterion	-0.291612
Sum squared resid	65.01415	Schwarz criterion	-0.277412
Log likelihood	222.1255	Hannan-Quinn criter.	-0.286321
F-statistic	2469.719	Durbin-Watson stat	1.426827
Prob(F-statistic)	0.000000		

Inverted AR Roots	.90



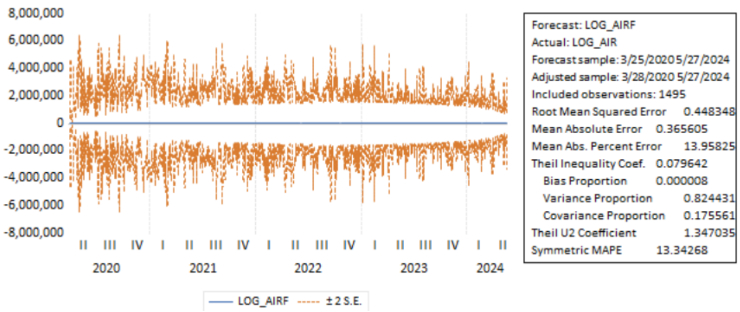
Date: 05/29/24 Time: 14:01
Sample (adjusted): 3/27/2020 5/27/2024
Q-statistic probabilities adjusted for 1 ARMA term and 1 dynamic regressor

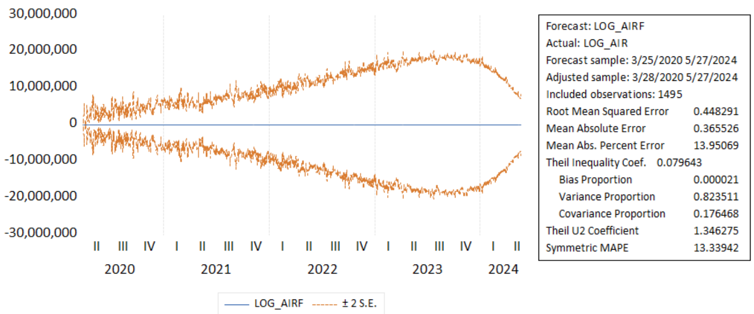
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.287	0.287	123.09	
		2	-0.313	-0.430	269.73	0.000
		3	-0.127	0.166	293.95	0.000
		4	-0.066	-0.287	300.55	0.000
		5	-0.061	0.104	306.18	0.000
		6	0.016	-0.127	306.57	0.000
		7	0.054	0.101	310.99	0.000
		8	0.025	-0.083	311.91	0.000
		9	0.047	0.145	315.23	0.000
		10	0.073	-0.025	323.36	0.000
		11	0.055	0.143	327.94	0.000
		12	0.010	-0.050	328.09	0.000
		13	-0.006	0.123	328.15	0.000
		14	0.032	-0.002	329.67	0.000
		15	0.038	0.099	331.86	0.000
		16	0.020	-0.005	332.45	0.000
		17	0.033	0.105	334.14	0.000
		18	0.044	0.003	337.08	0.000
		19	0.017	0.068	337.54	0.000
		20	0.025	0.031	338.51	0.000
		21	0.067	0.096	345.24	0.000
		22	0.036	-0.006	347.24	0.000
		23	-0.009	0.061	347.36	0.000
		24	0.007	-0.002	347.44	0.000
		25	0.014	0.032	347.72	0.000
		26	0.013	0.007	348.00	0.000
		27	0.063	0.086	354.12	0.000
		28	0.046	-0.045	357.32	0.000
		29	-0.018	0.043	357.84	0.000
		30	0.001	-0.025	357.84	0.000
		31	0.017	0.015	358.30	0.000
		32	0.040	0.036	360.77	0.000
		33	0.078	0.050	370.08	0.000
		34	0.048	0.012	373.61	0.000
		35	-0.009	0.003	373.75	0.000
		36	-0.032	-0.026	375.33	0.000

*Probabilities may not be valid for this equation specification.

Figure 8: Residuals from ARIMA(1,1,0)

Forecast Output Analysis:





⊟ Evaluation

Forecast Evaluation
Date: 05/29/24 Time: 00:35
Sample: 3/25/2020 5/27/2024
Included observations: 1497
Evaluation sample: 3/25/2020 5/27/2024
Training sample: 3/27/2020 12/31/2021
Number of forecasts: 7

Combination tests
Null hypothesis: Forecast i includes all information contained in others

Forecast	F-stat	F-prob
STDEV	29.29636	0.0000
STDEVF	588.1284	0.0000

Diebold-Mariano test (HLN adjusted)
Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prob	< prob
Abs Error	3.545496	0.0004	0.9998	0.0002
Sq Error	-3.268817	0.0011	0.0006	0.9994

Evaluation statistics

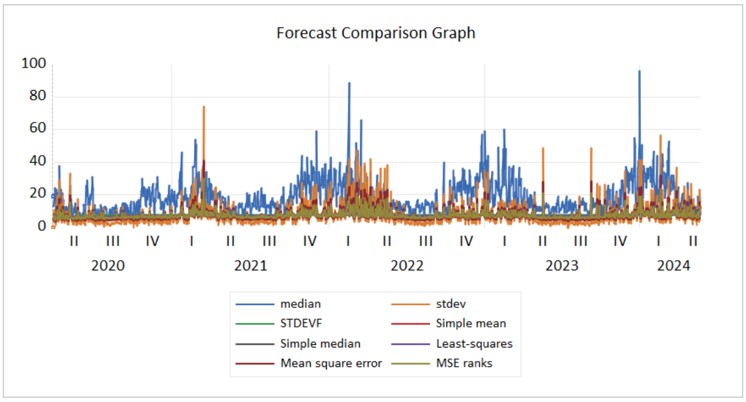
Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
STDEV	13.97050	11.71705	62.61927	95.38901	0.444926	2.012063
STDEVF	14.79767	11.17637	50.59520	72.41844	0.518276	1.758328
Simple mean	13.93996	11.17632	54.39474	78.48816	0.474771	1.778965
Simple median	13.93996	11.17632	54.39474	78.48816	0.474771	1.778965
Least-squares	14.46230	11.11401	51.02059	72.88495	0.505216	1.740454
Mean square error	13.94325	11.17432	54.34581	78.39979	0.475072	1.777944
MSE ranks	14.13214	11.10908	52.43950	75.12847	0.488597	1.746487

*Trimmed mean could not be calculated due to insufficient data

Forecast Evaluation Table

- The combination tests indicate that each forecast model (STDEV, STDEVF) contains unique information not captured by the other, as both F-statistics are significant (p ¡ 0.05).
- The significant p-values (¡ 0.05) for both absolute and squared errors indicate that there is a statistically significant difference in the accuracy of the forecasts.
- Best Performing Model: The Simple Mean model appears to perform best overall based on RMSE and MAE.
- STDEV Models: While the STDEV and STDEVF models capture more extreme fluctuations, they tend to have higher errors and variance.
- Overall Fit: The forecasts provide a reasonable approximation of the actual values, but the choice of model may depend on the specific error metrics prioritized (e.g., RMSE vs. MAPE).

⊟ Forecast Comparison



3 Application 2: SARIMA Models

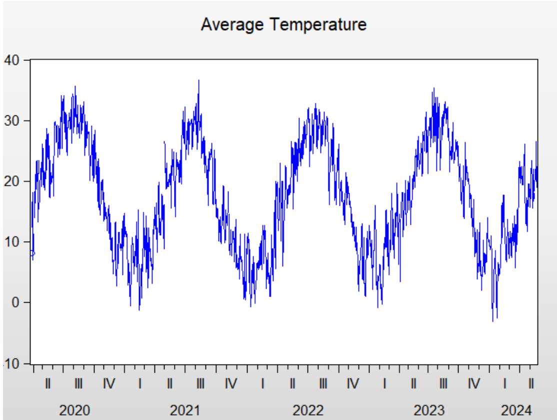
3.1 Unit Root Tests:

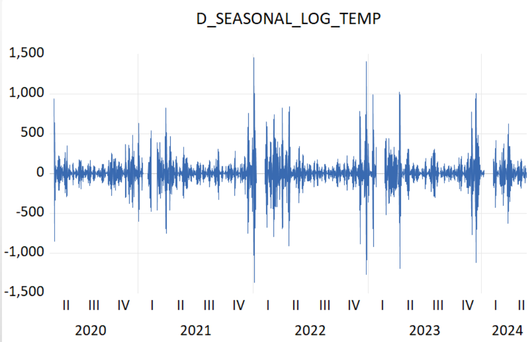
In this section, we will forecast the median daily temperatures in Bucharest, Popesti-Leordeni, for the period from 2020 to 2024. The first graph, representing the log-transformed median temperatures (LOG_TEMP), exhibits a clear seasonal pattern with regular peaks and troughs. This initial observation suggests that the time series contains a significant seasonal component and may not be stationary, indicating the need for differencing.

Examining the decomposed components, we observe distinct seasonality in the data. The seasonal component shows that temperatures tend to be higher during the summer months (June to September) and lower during the winter months. This seasonal variation is typical of temperate climates, where summer brings higher temperatures due to increased solar radiation, while winter results in lower temperatures.

The trend component reveals a general downward trend from 2020 to early 2022, followed by a slight upward trend from mid-2022 to 2024. This trend indicates that there are underlying long-term factors affecting the median temperatures, in addition to the seasonal effects.

To summarize, the presence of both non-stationarity and a seasonal component is evident in the temperature data. Consequently, we hypothesize that a first-order difference (d = 1) and a seasonal difference (D = 1) are necessary to achieve stationarity and adequately model the seasonal effects. This approach will allow us to use SARIMA models effectively for forecasting future temperatures in Bucharest, Popesti-Leordeni.





3.2 Box-Jenkins Methodology for Seasonal Data:

- ADF Test: The differenced series D(LOG_TEMP) is stationary, as the test statistic is significantly lower than the critical values, and the p-value is less than 0.05.
- PP Test: The differenced series D(LOG_TEMP) is stationary, as the test statistic is significantly lower than the critical values, and the p-value is less than 0.05.
- KPSS Test: The differenced series D(LOG_TEMP) is stationary, as the test statistic is much lower than the critical values, indicating that we fail to reject the null hypothesis of stationarity.

Null Hypothesis: D(LOG_TEMP) has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-22.84113	0.0000
Test critical values:	1% level	-3.434718	
	5% level	-2.863356	
	10% level	-2.567786	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOG_TEMP,2)
Method: Least Squares
Date: 05/29/24 Time: 15:49
Sample (adjusted): 4/01/2020 5/27/2024
Included observations: 1429 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOG_TEMP(-1))	-1.886752	0.082603	-22.84113	0.0000
D(LOG_TEMP(-1),2)	0.695344	0.069898	9.947984	0.0000
D(LOG_TEMP(-2),2)	0.410616	0.056439	7.275433	0.0000
D(LOG_TEMP(-3),2)	0.223822	0.040677	5.502455	0.0000
D(LOG_TEMP(-4),2)	0.040673	0.025212	1.613260	0.1069
C	-0.006069	0.007641	-0.794308	0.4271

R-squared	0.568706	Mean dependent var	-0.004232
Adjusted R-squared	0.567191	S.D. dependent var	0.438518
S.E. of regression	0.288493	Akaike info criterion	0.355901
Sum squared resid	118.4340	Schwarz criterion	0.378006
Log likelihood	-248.2909	Hannan-Quinn criter.	0.364156
F-statistic	375.2754	Durbin-Watson stat	1.896886
Prob(F-statistic)	0.000000		

Null Hypothesis: D(LOG_TEMP) has a unit root
Exogenous: Constant
Bandwidth: 59 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-60.32862	0.0001
Test critical values:	1% level	-3.434606	
	5% level	-2.863307	
	10% level	-2.567759	

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

Residual variance (no correction)	0.100373
HAC corrected variance (Bartlett kernel)	0.021874

Phillips-Perron Test Equation
Dependent Variable: D(LOG_TEMP,2)
Method: Least Squares
Date: 05/29/24 Time: 15:50
Sample (adjusted): 3/28/2020 5/27/2024
Included observations: 1465 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOG_TEMP(-1))	-1.076186	0.025251	-42.61877	0.0000
C	-0.002665	0.008286	-0.321624	0.7478

R-squared	0.553876	Mean dependent var	-0.011690
Adjusted R-squared	0.553571	S.D. dependent var	0.474491
S.E. of regression	0.317033	Akaike info criterion	0.541742
Sum squared resid	147.0460	Schwarz criterion	0.548964
Log likelihood	-394.8262	Hannan-Quinn criter.	0.544436
F-statistic	1816.360	Durbin-Watson stat	1.922275
Prob(F-statistic)	0.000000		

Null Hypothesis: D(LOG_TEMP) is stationary
Exogenous: Constant
Bandwidth: 60 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.085034
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.116107
HAC corrected variance (Bartlett kernel)	0.027898

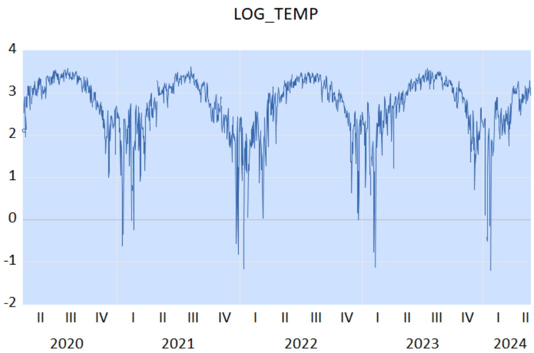
KPSS Test Equation
Dependent Variable: D(LOG_TEMP)
Method: Least Squares
Date: 05/29/24 Time: 15:51
Sample (adjusted): 3/27/2020 5/27/2024
Included observations: 1475 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003495	0.008875	0.393774	0.6938

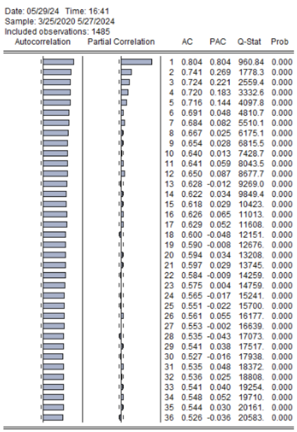
R-squared	0.000000	Mean dependent var	0.003495
Adjusted R-squared	0.000000	S.D. dependent var	0.340860
S.E. of regression	0.340860	Akaike info criterion	0.685986
Sum squared resid	171.2571	Schwarz criterion	0.689577
Log likelihood	-504.9145	Hannan-Quinn criter.	0.687325
Durbin-Watson stat	1.937631		

a)Reject H0 with ADF test; b)Reject H0 with PP test; c)Fail to reject KPSS test

After performing the first finite difference, we notice that the time series is now stationary using the p-values of the ADF and Philips-Peron test and the result of the KPSS test. On the other hand, using the seasonal graph, we notice two things that confirm the price increase in summer:



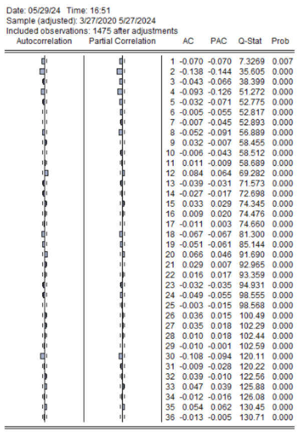
The largest price increase in absolute terms takes place at the end of each year, meaning in winter. The biggest price decrease in absolute value takes place in the beginning of the year, at the end of winter



Date: 05/29/24 Time: 16:41
Sample: 3/25/2020 5/27/2024
Included observations: 1485

	AC	PAC	Q-Stat	Prob
1	0.804	0.804	960.84	0.000
2	0.741	0.269	1778.3	0.000
3	0.724	0.221	2559.4	0.000
4	0.720	0.183	3332.6	0.000
5	0.716	0.144	4097.8	0.000
6	0.691	0.048	4810.7	0.000
7	0.684	0.082	5510.1	0.000
8	0.667	0.025	6175.1	0.000
9	0.654	0.028	6815.5	0.000
10	0.640	0.013	7428.7	0.000
11	0.641	0.059	8043.5	0.000
12	0.650	0.087	8677.7	0.000
13	0.628	-0.012	9269.0	0.000
14	0.622	0.034	9849.4	0.000
15	0.618	0.029	10423.	0.000
16	0.626	0.065	11013.	0.000
17	0.629	0.052	11608.	0.000
18	0.600	-0.048	12151.	0.000
19	0.590	-0.008	12676.	0.000
20	0.594	0.034	13208.	0.000
21	0.597	0.029	13745.	0.000
22	0.584	-0.009	14259.	0.000
23	0.575	0.004	14759.	0.000
24	0.565	-0.017	15241.	0.000
25	0.551	-0.022	15700.	0.000
26	0.561	0.055	16177.	0.000
27	0.553	-0.002	16639.	0.000
28	0.535	-0.043	17073.	0.000
29	0.541	0.038	17517.	0.000
30	0.527	-0.016	17938.	0.000
31	0.535	0.048	18372.	0.000
32	0.536	0.025	18808.	0.000
33	0.541	0.040	19254.	0.000
34	0.548	0.052	19710.	0.000
35	0.544	0.030	20161.	0.000
36	0.526	-0.036	20583.	0.000

(a) Simple correlogram

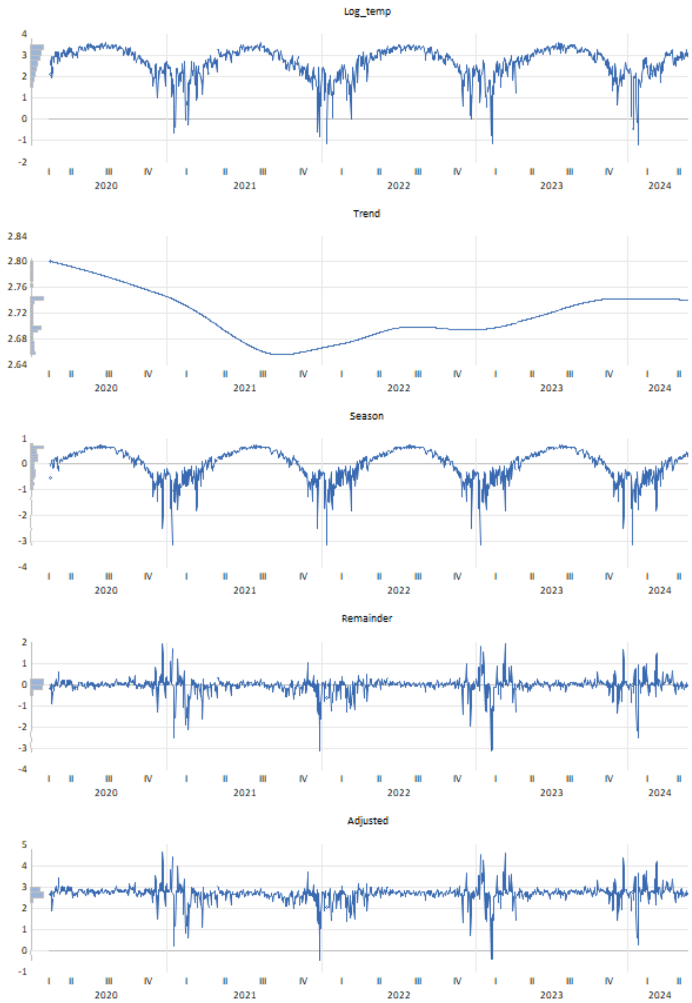


Date: 05/29/24 Time: 16:51
Sample (adjusted): 3/27/2020 5/27/2024
Included observations: 1475 after adjustments

	AC	PAC	Q-Stat	Prob
1	-0.070	-0.070	7.3269	0.007
2	-0.138	-0.144	35.605	0.000
3	-0.043	-0.066	38.399	0.000
4	-0.093	-0.126	51.272	0.000
5	-0.032	-0.071	52.775	0.000
6	-0.005	-0.055	52.817	0.000
7	-0.007	-0.045	52.893	0.000
8	-0.052	-0.091	56.889	0.000
9	0.032	-0.007	58.455	0.000
10	-0.006	-0.043	58.512	0.000
11	0.011	-0.009	58.689	0.000
12	0.084	0.064	69.282	0.000
13	-0.039	-0.031	71.573	0.000
14	-0.027	-0.017	72.698	0.000
15	0.033	0.029	74.345	0.000
16	0.009	0.020	74.476	0.000
17	-0.011	0.003	74.660	0.000
18	-0.067	-0.067	81.300	0.000
19	-0.051	-0.061	85.144	0.000
20	0.066	0.046	91.690	0.000
21	0.029	0.007	92.965	0.000
22	0.016	0.017	93.359	0.000
23	-0.032	-0.035	94.931	0.000
24	-0.049	-0.055	98.555	0.000
25	-0.003	-0.015	98.568	0.000
26	0.036	0.015	100.49	0.000
27	0.035	0.018	102.29	0.000
28	0.010	0.018	102.44	0.000
29	-0.010	-0.001	102.59	0.000
30	-0.108	-0.094	120.11	0.000
31	-0.009	-0.028	120.22	0.000
32	0.039	-0.010	122.56	0.000
33	0.047	0.039	125.88	0.000
34	-0.012	-0.016	126.08	0.000
35	0.054	0.062	130.45	0.000
36	-0.013	-0.005	130.71	0.000

(b) Correlogram of the First order difference

We can see that in the correlogram of first order, there is a slight seasonal pattern of from 6 to 6 months, meaning from winter to summer.



- **Seasonality:** The seasonal component shows a clear, consistent pattern with peaks and troughs corresponding to different seasons, indicating strong seasonality in the temperature data.
- **Trend:** The trend component reveals a declining trend in median temperatures from 2020 to early 2022, followed by a slight increase from mid-2022 onwards.
- **Residuals:** The remainder component suggests that the residuals are random and centered around zero, indicating a good fit of the trend and seasonal components.
- **Adjusted Series:** The adjusted series confirms the removal of seasonal effects, allowing for clearer analysis of the underlying trend and irregular components.

Dependence Variables SARIMA

Dependent Variable: LOG_TEMP
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/29/24 Time: 17:21
Sample: 4/02/2020 5/27/2024
Included observations: 1420
Failure to improve objective (non-zero gradients) after 61 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.656221	0.071854	36.96690	0.0000
D(LOG_TEMP)	0.559959	0.003767	148.6568	0.0000
D(LOG_TEMP,7)	-0.000952	4.50E-05	-21.13909	0.0000
AR(1)	0.919541	0.009830	93.54049	0.0000
SAR(1)	0.996711	0.010963	90.91777	0.0000
MA(1)	-0.999985	2.125291	-0.470517	0.6381
SMA(1)	1.000000	29.59130	0.033794	0.9730
SIGMASQ	0.019842	0.013785	1.439424	0.1503

R-squared	0.952806	Mean dependent var	2.733752
Adjusted R-squared	0.952572	S.D. dependent var	0.648634
S.E. of regression	0.141259	Akaike info criterion	-1.027049
Sum squared resid	28.17538	Schwarz criterion	-0.997424
Log likelihood	737.2044	Hannan-Quinn criter.	-1.015982
F-statistic	4072.422	Durbin-Watson stat	1.770584
Prob(F-statistic)	0.000000		

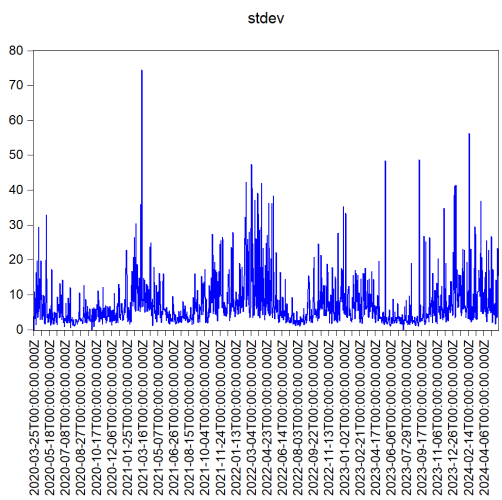
Inverted AR Roots	1.00	.92
Inverted MA Roots	1.00	-1.00

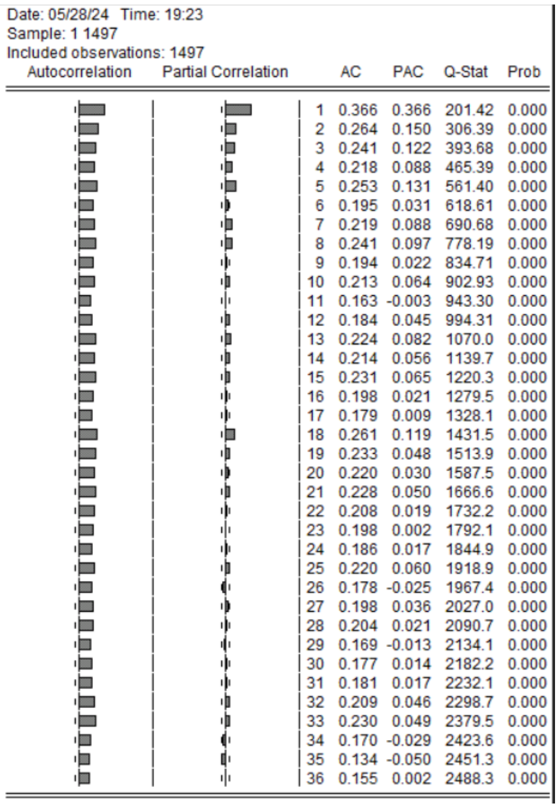
- The SARIMA model (with parameters specified) fits the data well, explaining over 95% of the variance in the log-transformed median daily temperatures.
- Significant terms include the constant, seasonal differencing, and AR(1) and SAR(1) terms, indicating these components are crucial for capturing the underlying patterns in the temperature data.
- The non-significance of the MA(1) and SMA(1) terms suggests they may not contribute meaningfully to the model and could potentially be removed in further model refinement.
- **D(LOG_TEMP,7): 0.059954, Std. Error = 0.000954, t-Statistic = 62.85569, Prob = 0.0000**
- The first seasonal differencing term is highly significant.
- **AR(1): 0.999937, Std. Error = 0.008430, t-Statistic = 118.5994, Prob = 0.0000**
- The first-order autoregressive term is highly significant, indicating strong autocorrelation.
- **SAR(1): 0.919920, Std. Error = 0.086986, t-Statistic = 10.57482, Prob = 0.0000**
- The first-order seasonal autoregressive term is significant.
- **MA(1): -0.033364, Std. Error = 0.045048, t-Statistic = -0.740834, Prob = 0.4587**

- The first-order moving average term is not significant.
- **Durbin-Watson stat: 1.770584**
- This statistic is close to 2, suggesting no significant autocorrelation in the residuals.

4 Application 3: Multivariate Time Series Analysis

4.1 Correlogram:

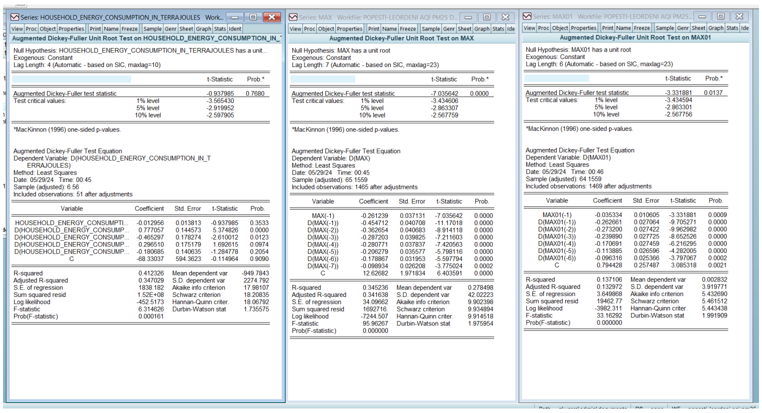




Date: 05/28/24 Time: 19:23
Sample: 1 1497
Included observations: 1497

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.366	0.366	201.42	0.000
		2	0.264	0.150	306.39	0.000
		3	0.241	0.122	393.68	0.000
		4	0.218	0.088	465.39	0.000
		5	0.253	0.131	561.40	0.000
		6	0.195	0.031	618.61	0.000
		7	0.219	0.088	690.68	0.000
		8	0.241	0.097	778.19	0.000
		9	0.194	0.022	834.71	0.000
		10	0.213	0.064	902.93	0.000
		11	0.163	-0.003	943.30	0.000
		12	0.184	0.045	994.31	0.000
		13	0.224	0.082	1070.0	0.000
		14	0.214	0.056	1139.7	0.000
		15	0.231	0.065	1220.3	0.000
		16	0.198	0.021	1279.5	0.000
		17	0.179	0.009	1328.1	0.000
		18	0.261	0.119	1431.5	0.000
		19	0.233	0.048	1513.9	0.000
		20	0.220	0.030	1587.5	0.000
		21	0.228	0.050	1666.6	0.000
		22	0.208	0.019	1732.2	0.000
		23	0.198	0.002	1792.1	0.000
		24	0.186	0.017	1844.9	0.000
		25	0.220	0.060	1918.9	0.000
		26	0.178	-0.025	1967.4	0.000
		27	0.198	0.036	2027.0	0.000
		28	0.204	0.021	2090.7	0.000
		29	0.169	-0.013	2134.1	0.000
		30	0.177	0.014	2182.2	0.000
		31	0.181	0.017	2232.1	0.000
		32	0.209	0.046	2298.7	0.000
		33	0.230	0.049	2379.5	0.000
		34	0.170	-0.029	2423.6	0.000
		35	0.134	-0.050	2451.3	0.000
		36	0.155	0.002	2488.3	0.000

4.2 Unit Root Tests:



Series: HOUSEHOLD_ENERGY_CONSUMPTION_IN_TERRAJOULES Work...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on HOUSEHOLD_ENERGY_CONSUMPTION_IN_

Null Hypothesis: HOUSEHOLD_ENERGY_CONSUMPTION_IN_TERRAJOULES has a unit...
Exogenous: Constant
Lag Length: 4 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.937985	0.7680
Test critical values:	1% level	-3.565430	
	5% level	-2.919952	
	10% level	-2.597905	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(HOUSEHOLD_ENERGY_CONSUMPTION_IN_TERRAJOULES)
Method: Least Squares
Date: 05/29/24 Time: 00:45
Sample (adjusted): 6 56
Included observations: 51 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HOUSEHOLD_ENERGY_CONSUMPTI...	-0.012956	0.013813	-0.937985	0.3533
D(HOUSEHOLD_ENERGY_CONSUMP...	0.777057	0.144573	5.374826	0.0000
D(HOUSEHOLD_ENERGY_CONSUMP...	-0.465297	0.178274	-2.610012	0.0123
D(HOUSEHOLD_ENERGY_CONSUMP...	0.296510	0.175179	1.692615	0.0974
D(HOUSEHOLD_ENERGY_CONSUMP...	-0.180685	0.140635	-1.284778	0.2054
C	-68.33037	594.3623	-0.114964	0.9090

R-squared	0.412326	Mean dependent var	-949.7843
Adjusted R-squared	0.347029	S.D. dependent var	2274.792
S.E. of regression	1838.182	Akaike info criterion	17.98107
Sum squared resid	1.52E+08	Schwarz criterion	18.20835
Log likelihood	-452.5173	Hannan-Quinn criter.	18.06792
F-statistic	6.314626	Durbin-Watson stat	1.735575
Prob(F-statistic)	0.000161		

Series: MAX Workfile: POPESTI-LEORDENI AQI PM25 D...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ide

Augmented Dickey-Fuller Unit Root Test on MAX

Null Hypothesis: MAX has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.035642	0.0000
Test critical values:	1% level	-3.434606	
	5% level	-2.863307	
	10% level	-2.567759	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MAX)
Method: Least Squares
Date: 05/29/24 Time: 00:45
Sample (adjusted): 65 1559
Included observations: 1465 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MAX(-1)	-0.261239	0.037131	-7.035642	0.0000
D(MAX(-1))	-0.454712	0.040708	-11.17018	0.0000
D(MAX(-2))	-0.362654	0.040683	-8.914118	0.0000
D(MAX(-3))	-0.287203	0.039825	-7.211603	0.0000
D(MAX(-4))	-0.280771	0.037837	-7.420563	0.0000
D(MAX(-5))	-0.206279	0.035577	-5.798116	0.0000
D(MAX(-6))	-0.178867	0.031953	-5.597794	0.0000
D(MAX(-7))	-0.098934	0.026208	-3.775024	0.0002
C	12.62682	1.971834	6.403591	0.0000

R-squared	0.345236	Mean dependent var	0.278498
Adjusted R-squared	0.341638	S.D. dependent var	42.02223
S.E. of regression	34.09662	Akaike info criterion	9.902398
Sum squared resid	1692716.	Schwarz criterion	9.934894
Log likelihood	-7244.507	Hannan-Quinn criter.	9.914518
F-statistic	95.96267	Durbin-Watson stat	1.975954
Prob(F-statistic)	0.000000		

Series: MAX01 Workfile: POPESTI-LEORDENI AQI PM25...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ide

Augmented Dickey-Fuller Unit Root Test on MAX01

Null Hypothesis: MAX01 has a unit root
Exogenous: Constant
Lag Length: 6 (Automatic - based on SIC, maxlag=23)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.331881	0.0137
Test critical values:	1% level	-3.434594	
	5% level	-2.863301	
	10% level	-2.567756	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MAX01)
Method: Least Squares
Date: 05/29/24 Time: 00:46
Sample (adjusted): 64 1559
Included observations: 1469 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MAX01(-1)	-0.035334	0.010605	-3.331881	0.0009
D(MAX01(-1))	-0.262661	0.027064	-9.705271	0.0000
D(MAX01(-2))	-0.273200	0.027422	-9.962982	0.0000
D(MAX01(-3))	-0.239890	0.027725	-8.652526	0.0000
D(MAX01(-4))	-0.170691	0.027459	-6.216295	0.0000
D(MAX01(-5))	-0.113885	0.026596	-4.282005	0.0000
D(MAX01(-6))	-0.096316	0.025366	-3.797067	0.0002
C	0.794428	0.257487	3.085318	0.0021

R-squared	0.137106	Mean dependent var	0.002832
Adjusted R-squared	0.132972	S.D. dependent var	3.919771
S.E. of regression	3.649868	Akaike info criterion	5.432690
Sum squared resid	19462.77	Schwarz criterion	5.461512
Log likelihood	-3982.311	Hannan-Quinn criter.	5.443438
F-statistic	33.16292	Durbin-Watson stat	1.991909
Prob(F-statistic)	0.000000		

Interpretation:

Household Energy Consumption (HOUSEHOLD_ENERGY_CONSUMPTION_IN_TERRAJOULES):

- Test Statistic: -1 (greater than -2.5 at the 10% level)
- p-value: 0.75 (greater than 0.05)
- **Conclusion:** The series is non-stationary because we do not reject the null hypothesis of a unit root.

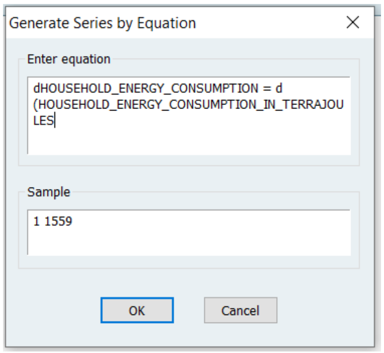
Air Quality Index (MAX):

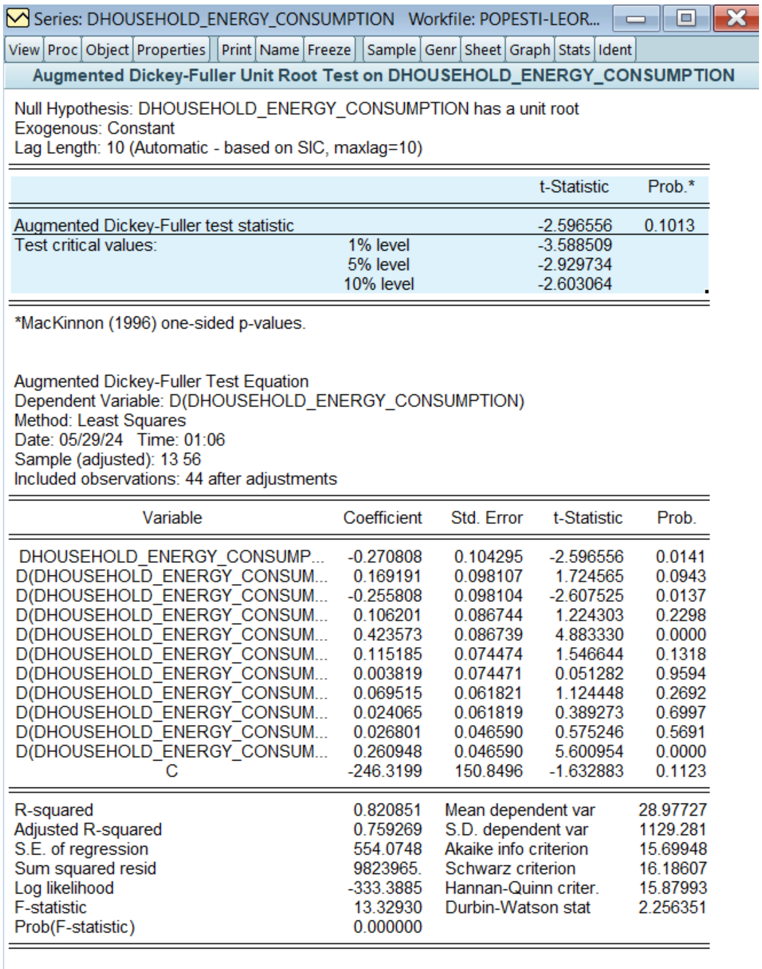
- Test Statistic: -7 (less than -3.5 at the 1% level)
- p-value: 0 (less than 0.01)
- **Conclusion:** The series is stationary because we reject the null hypothesis of a unit root.

Temperature (MAX01):

- Test Statistic: -3 (less than -2.8 at the 5% level but greater than -3.5 at the 1% level)
- p-value: 0.01 (less than 0.05)
- **Conclusion:** The series is stationary at the 5% significance level, so we reject the null hypothesis of a unit root.

Differencing Non-Stationary Series:





Series: DHOUSEHOLD_ENERGY_CONSUMPTION Workfile: POPESTI-LEOR...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on DHOUSEHOLD_ENERGY_CONSUMPTION

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

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.596556	0.1013
Test critical values:	1% level	-3.588509	
	5% level	-2.929734	
	10% level	-2.603064	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DHOUSEHOLD_ENERGY_CONSUMPTION)
Method: Least Squares
Date: 05/29/24 Time: 01:06
Sample (adjusted): 13 56
Included observations: 44 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DHOUSEHOLD_ENERGY_CONSUMP...	-0.270808	0.104295	-2.596556	0.0141
D(DHOUSEHOLD_ENERGY_CONSUM...	0.169191	0.098107	1.724565	0.0943
D(DHOUSEHOLD_ENERGY_CONSUM...	-0.255808	0.098104	-2.607525	0.0137
D(DHOUSEHOLD_ENERGY_CONSUM...	0.106201	0.086744	1.224303	0.2298
D(DHOUSEHOLD_ENERGY_CONSUM...	0.423573	0.086739	4.883330	0.0000
D(DHOUSEHOLD_ENERGY_CONSUM...	0.115185	0.074474	1.546644	0.1318
D(DHOUSEHOLD_ENERGY_CONSUM...	0.003819	0.074471	0.051282	0.9594
D(DHOUSEHOLD_ENERGY_CONSUM...	0.069515	0.061821	1.124448	0.2692
D(DHOUSEHOLD_ENERGY_CONSUM...	0.024065	0.061819	0.389273	0.6997
D(DHOUSEHOLD_ENERGY_CONSUM...	0.026801	0.046590	0.575246	0.5691
D(DHOUSEHOLD_ENERGY_CONSUM...	0.260948	0.046590	5.600954	0.0000
C	-246.3199	150.8496	-1.632883	0.1123

R-squared	0.820851	Mean dependent var	28.97727
Adjusted R-squared	0.759269	S.D. dependent var	1129.281
S.E. of regression	554.0748	Akaike info criterion	15.69948
Sum squared resid	9823965.	Schwarz criterion	16.18607
Log likelihood	-333.3885	Hannan-Quinn criter.	15.87993
F-statistic	13.32930	Durbin-Watson stat	2.256351
Prob(F-statistic)	0.000000		

Interpretation:

The differenced energy consumption series (dHOUSEHOLD_ENERGY_CONSUMPTION) has the following results:

- The ADF test statistic (-2.596556) is greater than the critical value at the 5% level (-2.929734) and very close to the 10% level (-2.603064).
- The p-value (0.1013) is greater than 0.05.
- **Conclusion:** The series is not clearly stationary at the 5% significance level but is close to being stationary at the 10% level. Given that it is almost at the threshold for the 10% significance level, we consider differencing the series again to ensure stationarity or proceeding cautiously by assuming it is nearly stationary.

Series: D2HOUSEHOLD_ENERGY_CONSUMPTION Workfile: POPESTI-...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on D2HOUSEHOLD_ENERGY_CONSUMPTION

Null Hypothesis: D2HOUSEHOLD_ENERGY_CONSUMPTION has a unit root
Exogenous: Constant
Lag Length: 9 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.673521	0.0867
Test critical values:	1% level	-3.588509	
	5% level	-2.929734	
	10% level	-2.603064	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(D2HOUSEHOLD_ENERGY_CONSUMPTION)
Method: Least Squares
Date: 05/29/24 Time: 01:08
Sample (adjusted): 13 56
Included observations: 44 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2HOUSEHOLD_ENERGY_CONSUM...	-0.983308	0.367795	-2.673521	0.0116
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.009897	0.335724	-0.029479	0.9767
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.428088	0.300477	-1.424695	0.1636
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.418744	0.261082	-1.603881	0.1183
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.091994	0.215188	-0.427505	0.6718
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.089356	0.181528	-0.492246	0.6258
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.198068	0.140074	-1.414025	0.1667
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.195765	0.110703	-1.768385	0.0862
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.238900	0.070043	-3.410763	0.0017
D(D2HOUSEHOLD_ENERGY_CONSU...	-0.236526	0.049441	-4.783967	0.0000
C	76.21905	92.73710	0.821883	0.4170

R-squared	0.903551	Mean dependent var	107.8636
Adjusted R-squared	0.874324	S.D. dependent var	1693.469
S.E. of regression	600.3479	Akaike info criterion	15.84521
Sum squared resid	11893783	Schwarz criterion	16.29126
Log likelihood	-337.5947	Hannan-Quinn criter.	16.01063
F-statistic	30.91502	Durbin-Watson stat	2.035719
Prob(F-statistic)	0.000000		

Interpretation:

The second differenced energy consumption series (d2HOUSEHOLD_ENERGY_CONSUMPTION) has the following results:

- The ADF test statistic (-2.673521) is still greater than the critical value at the 5% level (-2.929734) but less than the critical value at the 10% level (-2.603064).
- The p-value (0.0867) is greater than 0.05 but less than 0.10.
- **Conclusion:** The series is now stationary at the 10% significance level, as the ADF test statistic is less than the 10% critical value. Given that we are often stricter, this is a marginal case, but for practical purposes, we can proceed with the analysis considering it as stationary.

4.3 Cointegration Analysis:

Johansen Cointegration Test:

Date: 05/29/24 Time: 01:12
Sample (adjusted): 62 1559
Included observations: 1477 after adjustments
Trend assumption: Linear deterministic trend
Series: MAX MAX01
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.109878	186.9971	15.49471	0.0000
At most 1 *	0.010158	15.07980	3.841465	0.0001

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.109878	171.9173	14.26460	0.0000
At most 1 *	0.010158	15.07980	3.841465	0.0001

Max-eigenvalue test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

MAX	MAX01
-0.044658	-0.082236
0.005743	-0.100639

Unrestricted Adjustment Coefficients (alpha):

D(MAX)	11.67969	-0.471095
D(MAX01)	0.249688	0.369674

1 Cointegrating Equation(s): Log likelihood -11325.97

Normalized cointegrating coefficients (standard error in parentheses)

MAX	MAX01
1.000000	1.841482
	(0.18355)

Adjustment coefficients (standard error in parentheses)

D(MAX)	-0.521586
	(0.03914)
D(MAX01)	-0.011150
	(0.00436)

Interpretation of Johansen Cointegration Test Results:

Trace Test and Max-Eigenvalue Test:
Both tests show there are 2 cointegrating equations. This means MAX (air quality) and MAX01 (temperature) have a stable, long-term relationship. Cointegration Coefficient:
The relationship is:

```
MAX + 1.841482 x MAX01=0
```

This means for every increase in temperature, air quality increases by about 1.84 units. **Adjustment Coefficients:**

- Air quality (MAX) adjusts quickly to changes, correcting about 52% of deviations each period.
- Temperature (MAX01) adjusts slowly, correcting only about 1% of deviations each period.

Summary:

- **Stable Relationship:** Temperature and air quality move together over time.
- **Influence:** Temperature changes have a significant effect on air quality.
- **Adjustment Speed:** Air quality adjusts quickly; temperature adjusts slowly.

VAR:

Vector Autoregression Estimates
Date: 05/29/24 Time: 01:13
Sample (adjusted): 59 1559
Included observations: 1489 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	MAX	MAX01
MAX(-1)	0.284332	-0.009361
	(0.02580)	(0.00293)
	[11.0208]	[-3.19904]
MAX(-2)	0.118071	-0.003417
	(0.02584)	(0.00293)
	[4.56948]	[-1.16599]
MAX01(-1)	0.501162	0.779259
	(0.22522)	(0.02554)
	[2.22522]	[30.5064]
MAX01(-2)	-1.440058	0.129462
	(0.22501)	(0.02552)
	[-6.39997]	[5.07286]
C	49.77409	2.650986
	(3.24866)	(0.36846)
	[15.3214]	[7.19478]
R-squared	0.239399	0.838607
Adj. R-squared	0.237349	0.838171
Sum sq. resids	1725552.	22197.25
S.E. equation	34.09943	3.867520
F-statistic	116.7723	1927.730
Log likelihood	-7365.394	-4124.337
Akaike AIC	9.899790	5.546457
Schwarz SC	9.917607	5.564273
Mean dependent	47.77502	22.48750
S.D. dependent	39.04668	9.614022
Determinant resid covariance (dof adj.)		17382.57
Determinant resid covariance		17266.02
Log likelihood		-11489.31
Akaike information criterion		15.44568
Schwarz criterion		15.48132
Number of coefficients		10

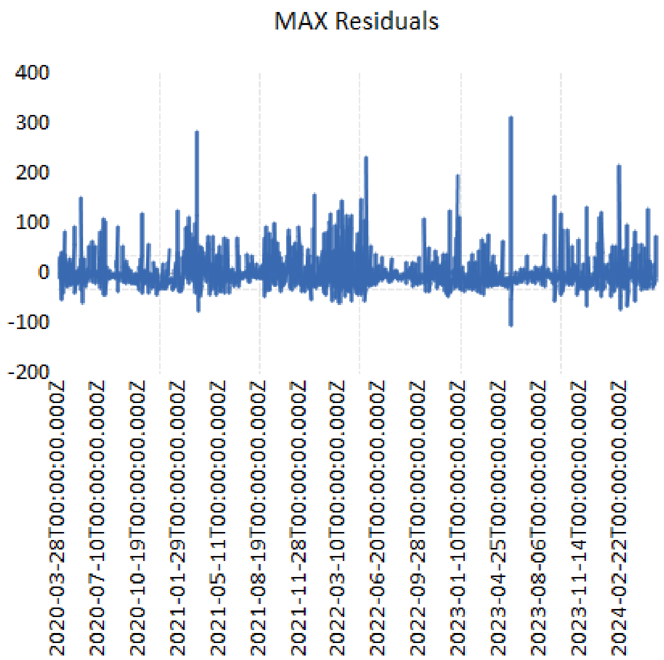
Estimate VECM:

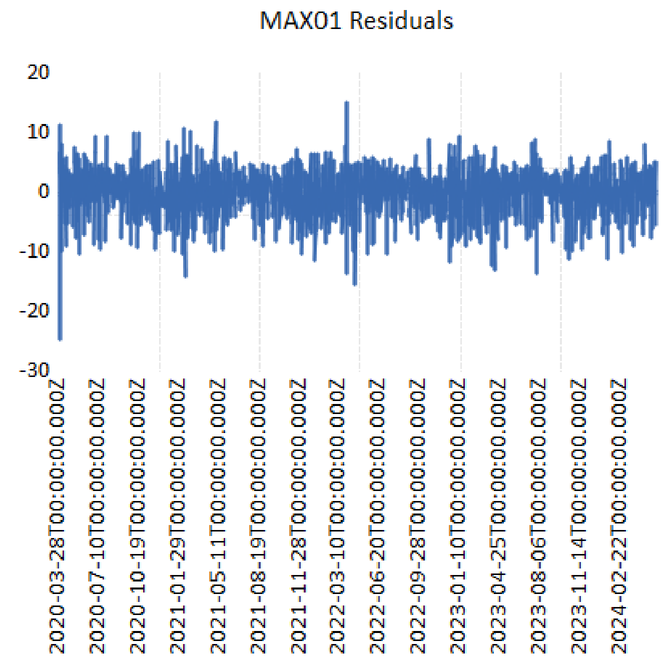
VAR Lag Order Selection Criteria
Endogenous variables: MAX MAX01
Exogenous variables: C
Date: 05/29/24 Time: 01:17
Sample: 1 1559
Included observations: 1465

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-12757.34	NA	125887.0	17.41889	17.42612	17.42159
1	-11298.32	2912.064	17270.66	15.43252	15.45418	15.44060
2	-11257.13	82.09689	16415.72	15.38175	15.41786	15.39522
3	-11224.64	64.67849	15789.40	15.34285	15.39340	15.36170
4	-11204.86	39.30414	15453.03	15.32131	15.38631*	15.34555
5	-11190.58	28.34885	15237.65	15.30728	15.38671	15.33691
6	-11182.39	16.22892	15150.80	15.30156	15.39544	15.33658
7	-11172.21	20.15522	15023.48	15.29312	15.40144	15.33352*
8	-11164.66	14.91964*	14950.91*	15.28828*	15.41104	15.33407

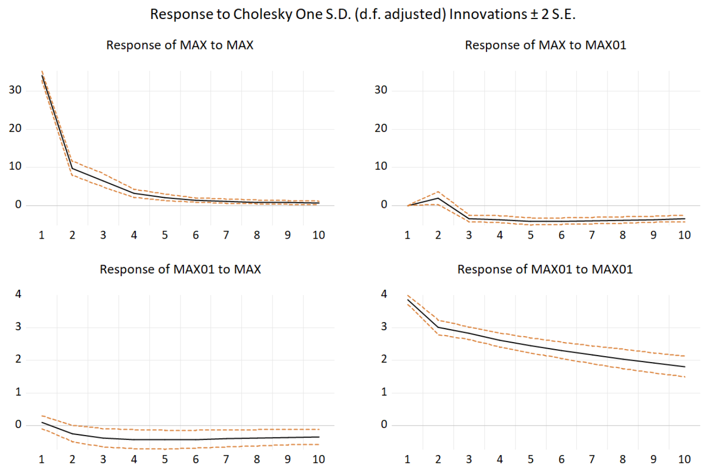
* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residuals

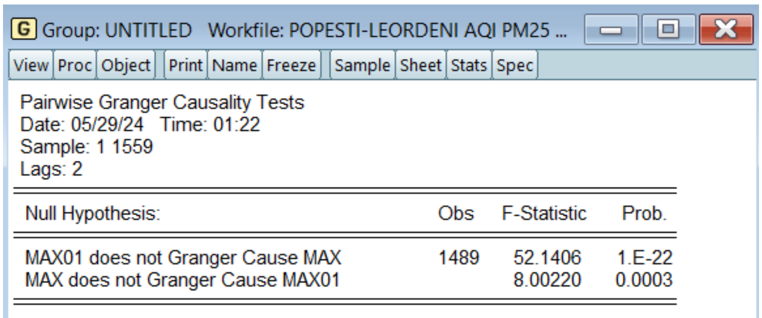




Impulse Response Function:



Granger Causality Test:



Group: UNTITLED Workfile: POPESTI-LEORDENI AQI PM25 ...

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Pairwise Granger Causality Tests
Date: 05/29/24 Time: 01:22
Sample: 1 1559
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
MAX01 does not Granger Cause MAX	1489	52.1406	1.E-22
MAX does not Granger Cause MAX01		8.00220	0.0003

Interpretation:

Null Hypothesis 1:

- MAX01 does not Granger Cause MAX The F-statistic is 52.1406, which is very high. The p-value is effectively 0 (1.E-22), which is far below the typical significance level of 0.05. **Conclusion:** We reject the null hypothesis that MAX01 (temperature) does not Granger cause MAX (air quality index). This means that past values of MAX01 provide significant information for predicting future values of MAX.

Null Hypothesis 2:

- MAX does not Granger Cause MAX01
- The F-statistic is 8.00220, which is significant.
- The p-value is 0.0003, which is also below the significance level of 0.05. **Conclusion:** We reject the null hypothesis that MAX (air quality index) does not Granger cause MAX01 (temperature). This means that past values of MAX provide significant information for predicting future values of MAX01.

Summary of Granger Causality Results:

- Bidirectional Granger Causality: The results show that there is a bidirectional Granger causality between temperature (MAX01) and air quality (MAX). Both variables Granger-cause each other, indicating that past values of one variable help in predicting the future values of the other.
- This bidirectional causality suggests a dynamic interaction between temperature and air quality. Understanding this interaction can be crucial for modeling and forecasting purposes, as it indicates that changes in temperature and air quality are interlinked and can influence each other over time.

Conclusion:

The complex relationships between temperature and air quality were examined in this project. We established numerous important discoveries using a variety of statistical techniques, such as Granger causality tests, unit root tests, and cointegration analyses:

- **Stationarity and Seasonality:** The temperature series was found to be stationary at the 5% significance level, while the air quality index (AQI) also exhibited stationarity. Seasonal effects were significant, indicating the necessity of incorporating both seasonal and trend components in our models.
- **Cointegration:** There exists a stable, long-term relationship between temperature and air quality, as demonstrated by the Johansen cointegration test. This indicates that over time, these two variables move together, reinforcing the need for integrated environmental policies.
- **Granger Causality:** The bidirectional Granger causality between temperature and air quality suggests that each variable can predict the future values of the other. This underscores the complex interplay between climatic factors and pollution levels, which should be considered in forecasting models and policy formulations.
- **Impact of Temperature on Air Quality:** The results indicated that temperature changes significantly impact air quality. Lower temperatures were associated with increased levels of pollutants, emphasizing the importance of monitoring and mitigating temperature variations to improve air quality. We think that the main reason for that are the various ineffictient methods of heating during the cold season.
- **Policy Implications:** These findings support the development of comprehensive strategies to manage both temperature and air quality. Policymakers are encouraged to consider the interdependence of these variables in their efforts to reduce emissions and enhance public health.

Overall, the project's insights are crucial for guiding future research and policy-making aimed at improving environmental quality and addressing the challenges posed by climate change and urban pollution.