Effect of Temperatures on Ozone Pollutant Levels in the United States

Jacob Croskey

ECON 6760: Time Series Analysis

May 6, 2024

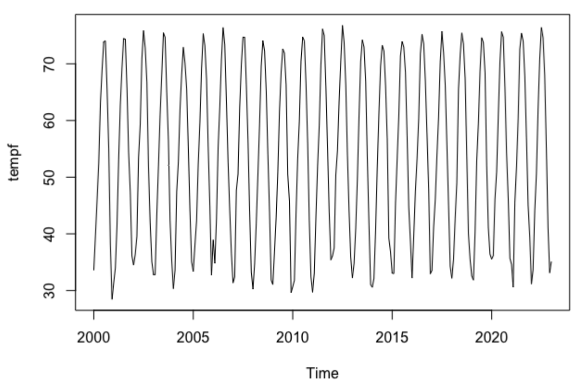
Motivation and Data Explanation:

This paper will analyze the relationship between average temperature levels and its effect on ozone pollutant levels in the United States from 2000-2023. My dependent variable of interest is ground-level ozone (O3) levels which form from the interaction between nitrogen oxides and volatile organic compounds. This interaction is catalyzed by sunlight and higher temperatures, which is why I chose average temperature levels in the United States as my explanatory variable. I want to investigate the obvious question: do changes in temperatures in the United States affect ground-level ozone levels in an impactful way?

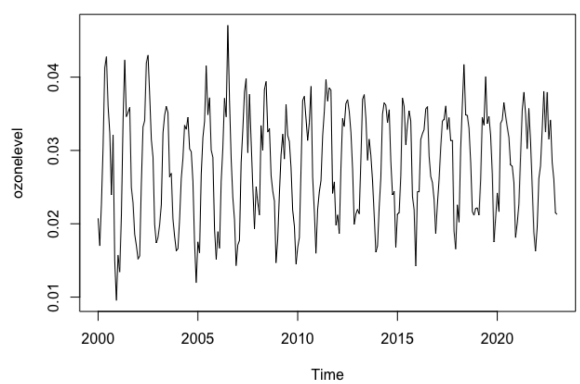
I gathered time series data for national temperatures from the National Centers for Environmental Information (NCEI) database which is a branch of the National Oceanic and Atmospheric Administration (NOAA). The monthly data I acquired ranged from January 2000 to January 2023 and is an average temperature across the entire United States, measured in Fahrenheit.

For the ozone level data, I used data from the Environmental Protection Agency (EPA). This dataset held data on four pollutants (Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide) with multiple daily observations across every state from the same time frame as the temperature data. The ozone levels were measured in parts per million (ppm). To make this a monthly time series, I only selected observations where the date was the first day of the month. Following this step, I grouped by date and took the mean to focus on national levels versus the state levels. See both time series plots below:

Temperatures



Ozone Levels



The motivation being investigating the relationship between temperature and ozone pollution levels is to better understand the dynamics of air quality and public health. I was particularly interested in the time period of 2000-2023 because it ranges from a couple of years before I was born to just over a year ago. I hope to see how these effects have changed over my lifetime. From the plots above, there is obvious seasonality to both temperatures and ozone levels. By examining how changes in temperature correlate with variations in ozone levels, researchers can better predict and manage air quality issues, especially during warmer months when ozone can reach harmful levels.

ARDL Model:

The format of my ARDL model is as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + \sum_{i=1}^{11} \delta_i D_i + \varepsilon_t$$

where $\mathbf{y_t}$ is ozone level, $\mathbf{x_t}$ is temperature, $\boldsymbol{\varepsilon_t}$ is the error term, $\boldsymbol{\beta_0}$ is the constant term which is normalized for January ozone levels, and $\sum_{i=1}^{11} \delta_i D_i$ is the sum of the dummy variables made for each month where the coefficients indicate how each month's ozone level differs from January (the reference month).

Note: *p<0.1; **p<0.05; ***p<0.01

<u>Variable</u>	<u>Estimated Coefficient</u>
Constant (β_0)	0.0102** (0.004)
Lagged Ozone (y_{t-1})	0.242*** (0.06)
Temperature (x_t)	0.0002** (0.0001)
Lagged Temperature (x_{t-1})	-0.00005 (0.0001)
February Dummy	-0.001 (0.001)
March Dummy	0.004*** (0.001)
April Dummy	0.007*** (0.002)
May Dummy	0.007** (0.003)
June Dummy	0.005 (0.004)
July Dummy	0.003 (0.005)
August Dummy	0.003 (0.005)
September Dummy	0.001 (0.005)
October Dummy	-0.001 (0.004)
November Dummy	-0.003 (0.002)
December Dummy	-0.004*** (0.001)

<u>Model Specifications:</u>

For this ARDL model, I wanted to analyze the effect of temperature on ozone pollution levels from 2000 to 2023. To start, I checked for stationarity using the Augmented Dickey-Fuller test in which both variables returned the alternative hypothesis of stationarity. After verifying this, I used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) methods to test for the optimal number of lags between 1 and 6. After both methods returned that 1 lag was the optimal number, I lagged both temperature and ozone level by one and incorporated them into my ARDL model. Finally, in order to control for the seasonal variation in both temperature and ozone levels as seen above, I created 11 dummy variables for months February-December. To avoid running into a perfect multicollinearity issue, I decided to omit one month (January) which means that every other dummy shows its effect relative to the omitted month. After fully specifying my model, I estimated the model in R; the coefficients of which can be seen in the table on the right.

<u>Model Interpretation:</u>

- Intercept (January Baseline): The intercept, representing January, has an estimate of 0.01023. This value is statistically significant ($p = 0.01578$), indicating that there is a baseline level of ozone in January around which other months' ozone levels vary.

- Lagged Ozone (lag_Ozone): The coefficient of 0.2417 with a p-value < 0.0001 (highly significant) indicates a strong positive relationship between the previous month's ozone levels and the current month's levels. This suggests that ozone levels have a significant carryover effect from one month to the next.
- Current Temperature (Temp): Shows a positive coefficient of 0.0002172, significant at the 5% level ($p = 0.0262$). This suggests that as temperatures increase, ozone levels also tend to rise, reflecting the immediate impact of temperature on ozone formation.
- Lagged Temperature (lag_Temp): The negative coefficient of -0.0000469, although not statistically significant ($p = 0.6332$), suggests that the previous month's temperature does not significantly influence the current month's ozone levels.

Seasonal Effects (Monthly Dummies):

- February: Shows a slight decrease in ozone levels compared to January, but this is not significant ($p = 0.5297$).
- March to May: Each of these months shows a positive and significant increase in ozone levels compared to January, with April showing the strongest effect ($p = 0.000983$).
- June to August: The coefficients are positive but not statistically significant, suggesting these months do not differ significantly from January in terms of ozone levels.
- September to November: These months also show non-significant differences from January, with slight decreases for October and November.
- December: Shows a significant decrease in ozone levels compared to January ($p = 0.0050$), indicating lower ozone levels during this month.

Overall Interpretation:

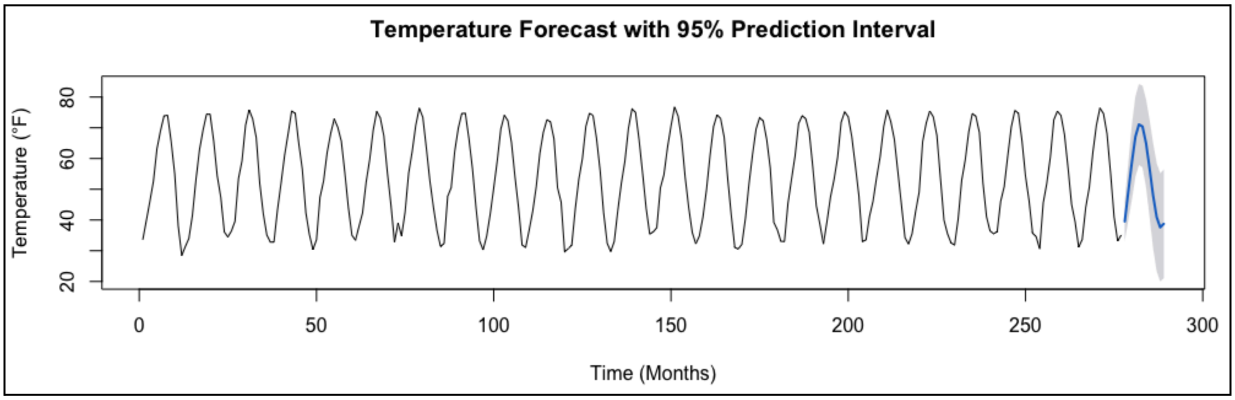
This model effectively captures the dynamics of ozone levels, illustrating the importance of both current and historical temperatures and ozone levels. The significant impact of temperatures supports the understanding that higher temperatures can enhance ozone formation through increased photochemical reactions. The lagged effect of ozone levels confirms the persistence in atmospheric conditions affecting ozone. Seasonal variations highlighted by the monthly dummies demonstrate that ozone levels fluctuate throughout the year, likely due to changes in weather conditions and human activities such as heating and driving patterns, which vary seasonally.

AR Forecast Model for Temperature (x_t):
To forecast for future temperature values 12 periods out, I used an ARIMA(3,0,0) model. The ARIMA(3,0,0) model indicates a third-order autoregressive structure without differencing or moving average components. I ended up with this specification by using R to determine the optimal number of lags and producing an ARIMA that would find a balance between model fit and complexity. The point forecasts began at 39.54°F which seamlessly picks up from the last observation (January 2023) of 35.13°F and reached up to 71.08°F, demonstrating expected seasonal variations or inherent patterns in the temperature data. Accompanying these point forecasts, the 95% prediction intervals were calculated to provide a measure of the uncertainty associated with each forecasted value. These intervals widened as the forecast horizon extended, which is expected as more distant predictions yield increasing uncertainty. Specifically, the intervals ranged from approximately 33.28°F to 45.80°F in the initial forecast period, gradually expanding to between 21.08°F and 56.39°F by the twelfth period. The statistical performance of the model was robust, with a residual standard error (sigma^2) of 10.2, indicating the average squared deviations from the observed temperatures were moderate, suggesting a reasonable fit to the historical data.

Temperature Forecast Results:

Month	Point Forecast (°F)	Lower Bound	Upper Bound
February 2023	39.54	33.28	45.80
March 2023	49.21	40.08	58.33
April 2023	58.70	46.76	70.63
May 2023	67.08	54.13	80.04
June 2023	71.08	57.95	84.21
July 2023	70.42	57.15	83.68
August 2023	65.03	50.89	79.17
September 2023	56.81	41.17	72.44
October 2023	47.93	31.03	64.83
November 2023	40.94	23.44	58.43
December 2023	37.60	20.05	55.15
January 2024	38.73	21.08	56.39

Plotted Forecast with 95% Confidence Interval:

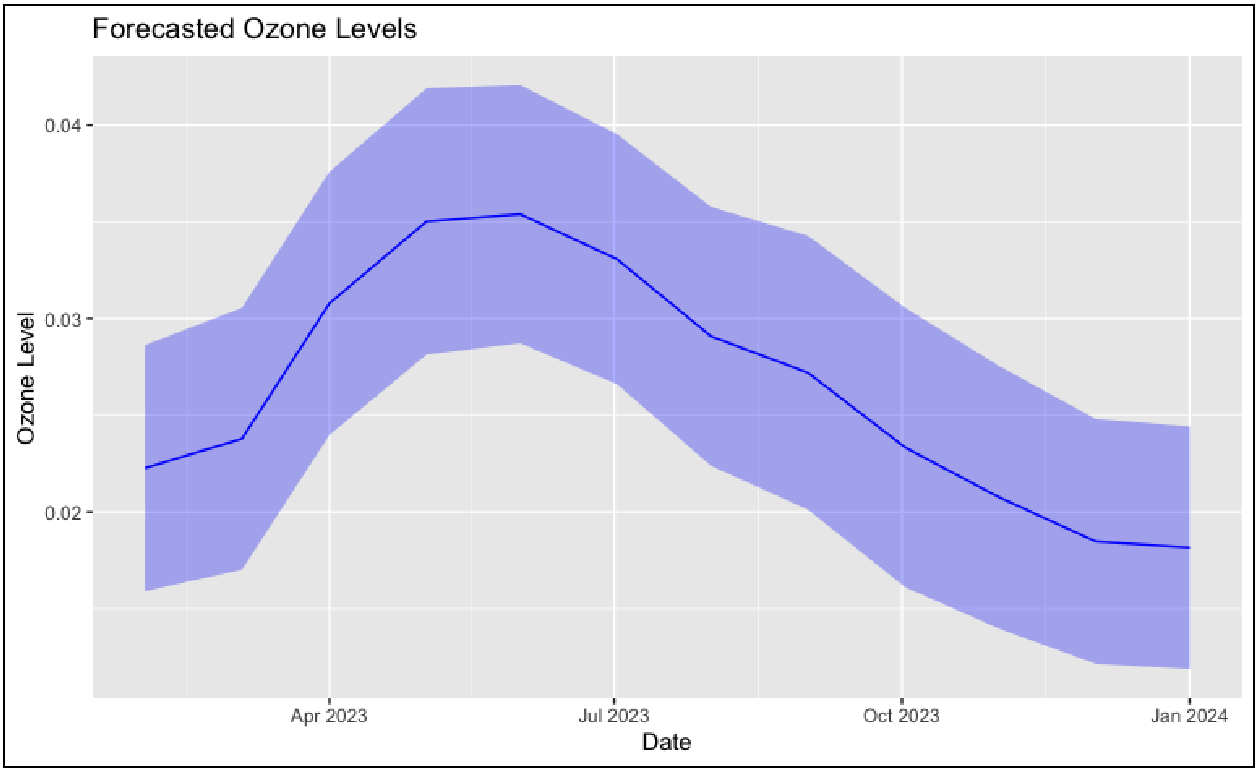


AR Forecast Model for Ozone Level (y_t):
Using the causal model and AR forecasts for temperature, I was able to create a forecast for ozone pollution levels 12 months out from the last observation in the sample. Along with including temperature, its lag, and the lag for ozone level to assist in the prediction, I categorized each forecasted date into its respective month in order to capture seasonal influences to each part of the year. The resulting forecasts, as detailed in the provided table below, indicate a varied range of expected ozone levels, starting from an initial forecast of 0.0223ppm and adjusting across the year as influenced by temperature and seasonal factors. Similar to the temperature forecast above, this forecast also continues smoothly from the last observed ozone pollution level of 0.0213ppm (January 2023). The 95% prediction intervals for these forecasts also broaden slightly with time, a reflection of the increasing uncertainty in predictions as the forecast horizon extends.

Ozone Level Forecast Results:

Month	Point Forecast (ppm)	Lower Bound	Upper Bound
February 2023	0.0223	0.0159	0.0286
March 2023	0.0238	0.0170	0.0306
April 2023	0.0308	0.0240	0.0376
May 2023	0.0350	0.0281	0.0419
June 2023	0.0354	0.0287	0.0421
July 2023	0.0331	0.0266	0.0395
August 2023	0.02908	0.0224	0.0358
September 2023	0.0272	0.0201	0.0343
October 2023	0.0233	0.0161	0.0306
November 2023	0.0208	0.0140	0.0276
December 2023	0.0185	0.0122	0.0248
January 2024	0.0181	0.0119	0.0244

Plotted Forecast with 95% Confidence Interval:



Conclusion

In this analysis, the goal was to develop a robust understanding of the factors influencing ozone levels through the use of time series modeling, particularly focusing on the impacts of temperature and seasonality. While the models employed provided valuable insights and forecasted outcomes, several limitations must be acknowledged. These limitations highlight the potential for bias in our interpretations and inferences regarding causality.

- Assumption of Exogeneity: A critical assumption for deriving causal inference from regression models is the exogeneity of the independent variables. The assumption is that x_t (temperature) is not influenced by y_t (ozone levels). However, this might not strictly hold due to potential omitted variables that could affect both temperature and ozone levels, such as economic activity, urban heat island effects, or regulatory changes affecting environmental conditions. If such variables are omitted, they could induce endogeneity, leading to biased estimates and incorrect causal inferences.
- Potential for Reverse Causality and Simultaneity: The relationship between temperature and ozone might not be unidirectional. For instance, extensive urbanization can lead to higher temperatures while also contributing to higher pollution levels, including ozone. This simultaneity in the relationship could complicate the causal interpretation that higher temperatures lead to higher ozone levels, as both could be simultaneously influencing each other.
- Seasonality and Model Specification: The selection of an ARIMA(3,0,0) model may not be the best model to use to account for seasonal variation in the data. Although ARIMA models are robust for non-seasonal predictions, the lack of a seasonal component could lead to underfitting seasonal patterns, potentially skewing forecasts and causal interpretations. Though my forecast did catch seasonal variation, exploring alternative models such as SARIMAs could introduce new insights and explain this more accurately.

The analysis conducted provides a structured approach to understanding the dynamics between ozone levels, temperature, and seasonality, employing statistical models to predict future conditions. However, the inherent limitations related to potential endogeneity issues and the potential seasonal variation issues in the initial model underscore the need for cautious interpretation of the results.

Despite these limitations, it is still important to recognize the implications and causal relationship of the model. The main takeaways are as follows:

- Direct Link Established: For every 1-degree increase in temperature, ozone levels rise by approximately 0.0002172 ppm, confirming a direct and statistically significant relationship between temperature and ozone levels.

- Public Health Impact: The relationship highlights the potential for increased ozone pollution during warmer periods, necessitating enhanced public health advisories and preventive measures during heatwaves.
- Future Research Direction: The study underscores the importance of incorporating seasonal variations and potentially omitted variables like urbanization effects in future models to better understand and predict ozone dynamics.

Investigating the relationship between temperature and ozone pollution levels is crucial for understanding the dynamics of air quality and public health. Temperature can significantly influence ozone formation, as warmer temperatures accelerate the chemical reactions that produce ozone in the atmosphere. By examining how changes in temperature correlate with variations in ozone levels, researchers can better predict and manage air quality issues, especially during warmer months when ozone can reach harmful levels. This analysis is particularly important in urban areas, where high temperatures and ozone concentrations can combine to create serious health risks for residents.