Global CO_2 Emissions in 1997

Meral Basit, Alex Hubbard, Mohamed Bakr

Contents

1 Background

1.1 Carbon Emissions

Carbon emissions refer to the release of carbon, particularly carbon dioxide (CO_2), into the atmosphere. This process primarily occurs through the burning of fossil fuels such as coal, oil, and natural gas, as well as through deforestation and various industrial processes. CO_2 is a greenhouse gas, meaning it traps heat in the Earth's atmosphere and contributes to the greenhouse effect, which leads to global warming and climate change.

In our report we are trying to understand the trend of the atmospheric CO2 by asking the following research question:

> *Is there a significant upward trend in atmospheric CO2 levels over time?*

1.2 Null Hypothesis

There is no significant upward trend in atmospheric CO_2 levels over time. $H_0 : \beta_1 \leq 0$ Where: β_1 is the trend coefficient over time in a linear regression model of the form $CO2_t = \beta_0 + \beta_1 \cdot t + \epsilon_t$. $CO2_t$ is the atmospheric CO_2 level at time t.

2 Measurement and Data

2.1 Measuring Atmospheric Carbon

In this study, we will use the Mauna Loa Atmospheric CO_2 Concentration time series dataset that is a available in `R`. The time series of 468 observations of the monthly Atmospheric concentrations of CO_2 from 1959 to 1997 expressed in parts per million (ppm). This means that, for example, a value of 320 means there are 320 CO2 molecules for every 1 million air particles (after the water vapor is removed from the sample). The data was reported in the preliminary 1997 SIO (Keeling and Whorf, 1997).

The data was collected at the Mouna Loa Observatory which is located on the island of Hawaii at an elevation of 11,135 feet above sea level which makes this location "well situated to measure air masses that are representative of very large areas" (ESRL global monitoring laboratory - mauna loa observatory, n.d.), (Global monitoring laboratory - carbon cycle greenhouse gases, n.d.). The values for February, March and April of 1964 were missing and have been obtained by interpolating linearly between the values for January and May of 1964.

2.2 Historical Trends in Atmospheric Carbon

The Keeling Curve is a graph of the accumulation of carbon dioxide in the Earth's atmosphere based on continuous measurements taken at the Mauna Loa Observatory on the island of Hawaii from 1958 to the present day. The curve is named for the scientist Charles David Keeling, who started the monitoring program and supervised it until his death in 2005 (Keeling Curve, 2024)

As we see in Figure 1, which is the time series in the top left, there is a pretty strong seasonality as well as a linear upward trend. Additionally, the dataset is not mean stationary but it may be variance stationary.

The `ACF` decays slowly but does not dampen below the significance level even after lag 24. This suggests a strong autocorrelation in the CO_2 values while the `PACF` drops shortly after lag 1 but

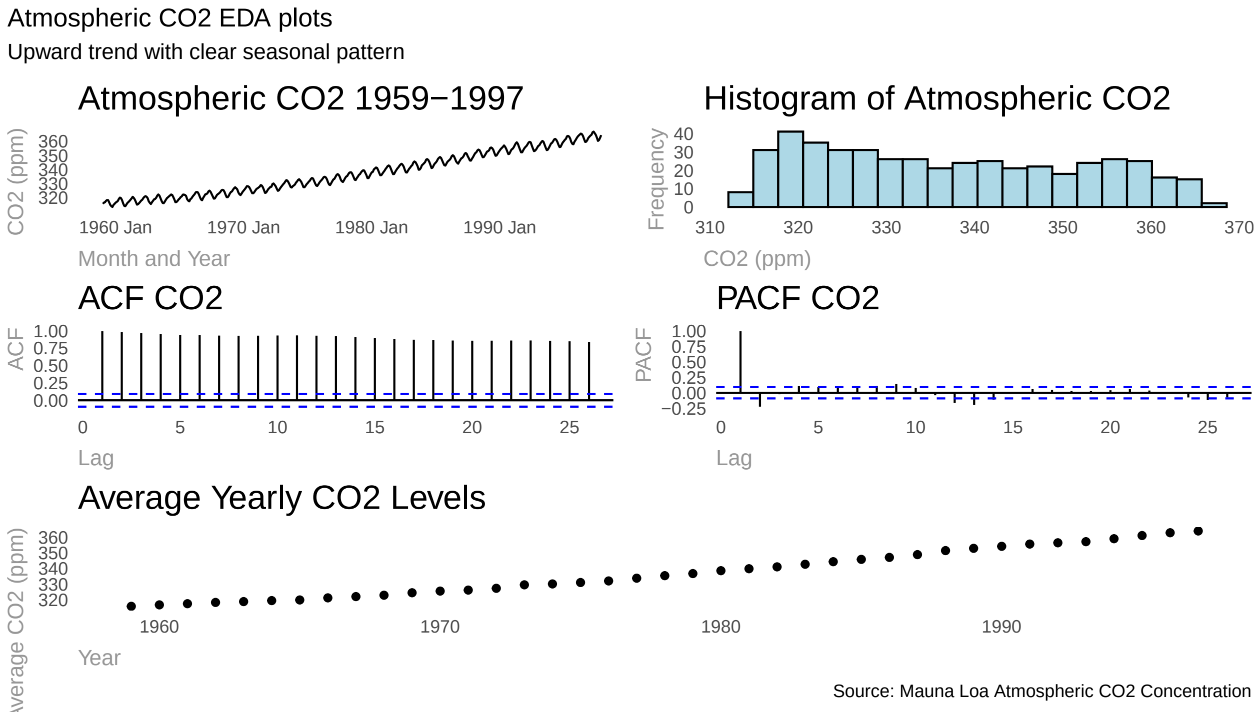

Figure 1: Atmospheric CO2 EDA standard plots

still has an oscillating pattern with a few lags above the significance level. This suggests that this series may have a unit root. Both `ACF` and `PACF` show seasonality patterns.

The histogram in Figure 1 top right shows that there is a wide range of values with a slight right skew. The yearly average plot on the bottom illustrates the linear trend in the data series more explicitly.

3 Models and Forecasts

In this section, we will analyze and compare two different models to gain a better understanding of the complex dynamics of the time series process. We will assess a linear model and an ARIMA model to identify the most appropriate time series model for our analysis.

3.1 Linear vs Quadratic Models

We started by fitting a linear model of the form: $CO2_t = \beta_0 + \beta_1 t + \epsilon_t$. Based on the fit results, the estimated coefficient $\beta_1 = 0.109$ indicates that the CO_2 levels increase by ≈ 0.109 units per month. The p-value of the time index is < 0.05 which suggests that the coefficient is statistically significant. We reject the null hypothesis that the coefficient $\beta_1 = 0$. That also provides evidence that the CO_2 levels have an upward linear trend.

Both R^2 and the adjusted R^2 value is 0.969, which means that the linear model can explain 96.9% of the CO_2 levels variance, suggesting that the model effectively captures the main patterns in the data.

Next, we fit a quadratic model of the form: $CO2_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$. Based on the fit results of the quadratic time trend model, the estimated coefficient $\beta_1 = 0.0674$ indicates that the CO_2 levels increase by ≈ 0.0674 units per month. The p-value of the time index is < 0.05 which suggests that the coefficient is statistically significant. We reject the null hypothesis that the coefficient $\beta_1 = 0$. that also provides evidence that the CO_2 levels have an upward linear trend.

The estimated quadratic term coefficient $\beta_2 = 0.0000886$. The positive coefficient suggests that the rate of increase in CO_2 levels is accelerating. The p-value of the time index is < 0.05 which suggests that the coefficient is statistically significant. We reject the null hypothesis that the coefficient $\beta_2 = 0$.

Both R^2 and the adjusted R^2 value is 0.979, which means that the linear model can explain 97.9% of the CO_2 levels variance, suggesting that the model effectively captures the main patterns in the data.

After analyzing the two models, it is obvious that the significant coefficients show a clear upward trend in CO2 levels over time, with a rapidly increasing rate. However, it seems that the quadratic model is slightly outperforming the linear model when comparing the R^2 results and examining the residual plots in Figure 2.

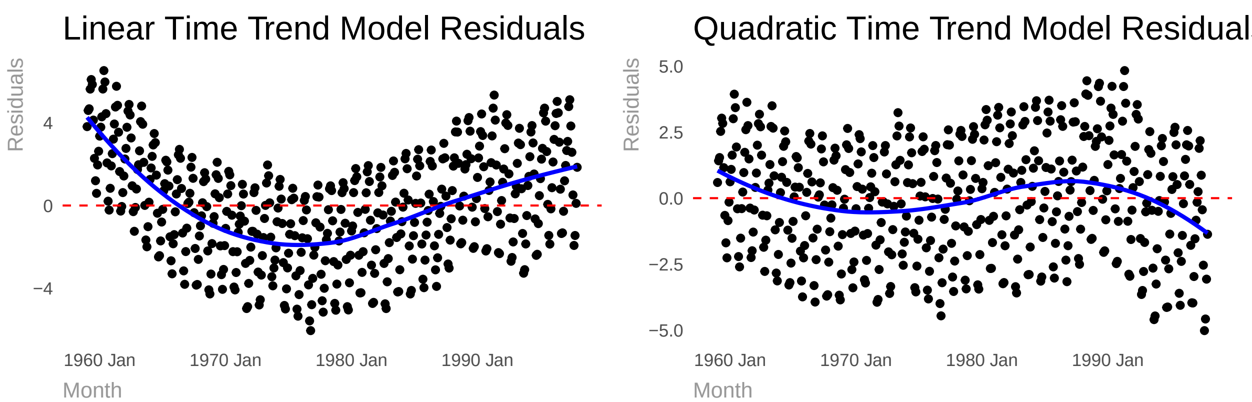

Figure 2: Residuals of the Linear and Quadratic Models

A logarithmic transformation can stabilize variance and make growth trends more linear. However, in this case, the log-linear model with $R^2 = 0.972$ does not greatly improve the fit compared to the quadratic model with $R^2 = 0.979$.

3.2 Polynomial Model

As you can see in Figure 3, adding a seasonal dummy variable to a polynomial model helps to capture some of the seasonal variation that we see in the historical CO_2 levels and improves the model's fit with $R^2 = 0.998$.

3.3 ARIMA Models

Based on the EDA and `ACF` and `PACF` plots in Figure 1, it was challenging to determine the parameters p and q using figures alone. Therefore, we used `AICc` to estimate the parameters.

To estimate the i parameter, we used the `KPSS` test to check for stationarity. Initially, the p-value was 0.01, leading us to reject the null hypothesis, indicating non-stationarity. After differencing

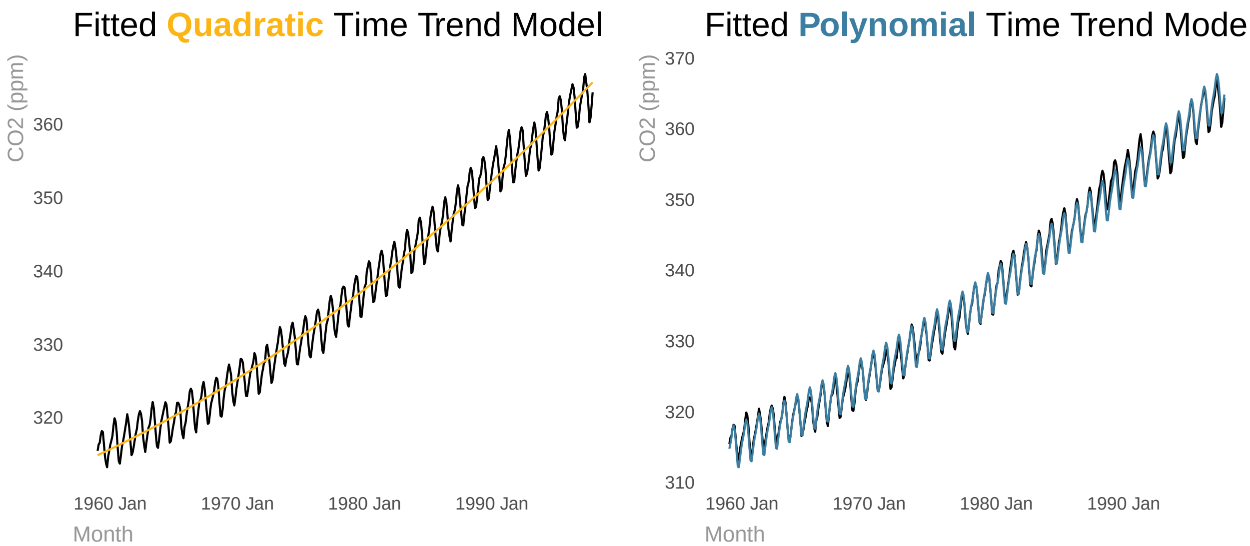

Figure 3: Quadratic Model vs Polynomial with seasonal dummies

Table 1: ARIMA Models Comparison

.model	sigma2	log_lik	AIC	AICc	BIC
arima_fit_log.search	0.00	2330.6	-4649	-4649	-4625
arima_fit_log	0.00	2305.8	-4602	-4601	-4581
arima_fit.search	0.08	-71.6	157	157	185
arima_fit	0.08	-72.3	159	159	187
baseline	0.10	-108.1	226	226	246

once, the p-value went up to 0.1, and we failed to reject the null hypothesis, indicating that the data is stationary after one differencing.

We estimated an ARIMA(2,1,0)(2,1,0)[12] model as a baseline plus four additional models. One model used the data as is, and one had a log transformation. We included models with stepwise and non-stepwise searches for each dataset. By examining Table 1, we find that the model with the lowest AICc value is ARIMA(3,1,0)(2,1,0)[12] with a log transformation (non-stepwise search). This aligns with our earlier observation that the log transformation yields better results.

The residuals line and `ACF` plots Figure 4 from the ARIMA(3,1,0)(2,1,0)[12] model show that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise which is confirmed by `Ljung Box` test. With a p-value = 0.358, we fail to reject the null hypothesis H_0 : data is independently distributed.

3.4 Forecasts

Using the polynomial model with seasonal dummies, we predict that CO2 levels will reach 413 ppm on average by 2020.

Using our best model the ARIMA(3,1,0)(2,1,0)[12], we predicted that the CO_2 levels will reach 420 ppm by 2025 Mar with [308, 560]95% confidence interval while it predicts it will reach 500 ppm by 2050 May with [227, 964]95% confidence interval.

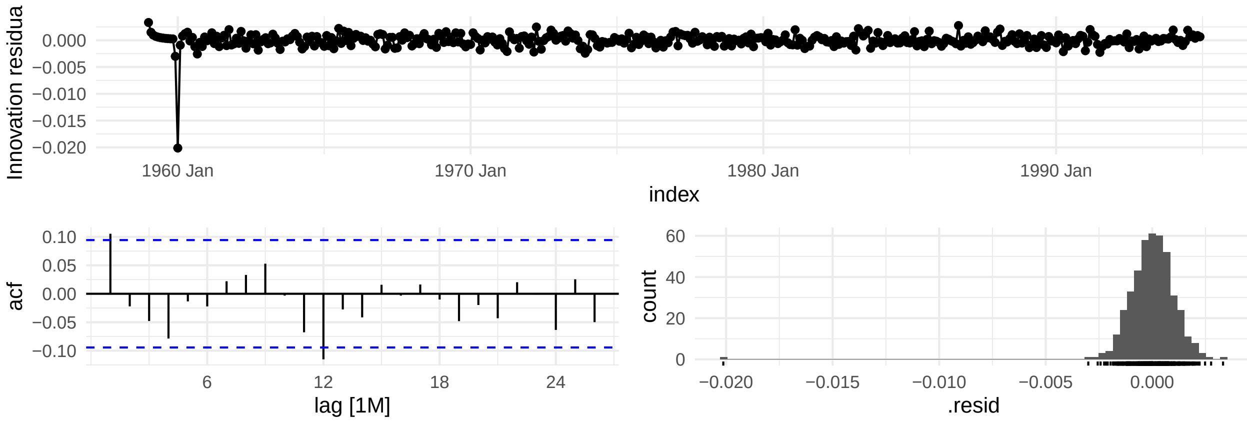

Figure 4: ARIMA(3,1,0)(2,1,0)[12] Model Residuals

By 2100, our model predicts that CO2 levels will be at about 850 ppm on average with confidence intervals varying between [91, 3741]95% and [89, 3835]95%. These predictions are may be highly inaccurate, since we are forecasting very far into the future. Additionally, our model does not account for any unforeseen events that might impact the rate of increase. For example, increases in efficiency and decarbonization of the grid may decrease the slope of the graph. The prediction intervals provide a measure of this uncertainty: the further the projected point, the wider the intervals.

4 Conclusions

The polynomial model fits the CO_2 time series data well, explaining nearly 99.8% of the variance in CO_2 levels. The significant coefficients indicate a strong upward trend in atmospheric CO_2 levels over time with an accelerating rate of increase. We reject the null hypothesis $H_0 : \beta_1 \leq 0$. This model provides a more nuanced understanding of the trend and variability in CO_2 levels compared to the simpler linear models estimated here.

The selected ARIMA model (e.g., ARIMA(3, 1, 0)(2, 1, 0)[12]) effectively captures the trend and seasonality in the CO_2 time series data. The model diagnostics indicate a good fit, and the model suggests significant increases in atmospheric CO_2 levels if current trends continue. However, while the point estimates provide specific values, the prediction intervals highlight the uncertainty in the long-term forecasts.

Finally, ARIMA model with RMSE 0.43 performed better than the Polynomial model with RMSE 0.72.

A Appendix: Model comparison

Table 2: Estimated Time Series Models

	Output Variable: expected Atmospheric CO_2 Levels			
		log(value)	value	
	Linear Model	Log Linear Model	Quadratic Model	Polynomial Model
	(1)	(2)	(3)	(4)
time(index)	0.11*** (0.001)	0.0003*** (0.0000)	0.07*** (0.003)	0.07*** (0.001)
I(time(index)ˆ2)			0.0001*** (0.0000)	0.0001*** (0.0000)
month2				0.66*** (0.16)
month3				1.41*** (0.16)
month4				2.54*** (0.16)
month5				3.02*** (0.16)
month6				2.35*** (0.16)
month7				0.83*** (0.16)
month8				−1.23*** (0.16)
month9				−3.06*** (0.16)
month10				−3.24*** (0.16)
month11				−2.05*** (0.16)
month12				−0.94*** (0.16)
Constant	312.00*** (0.24)	5.74*** (0.001)	315.00*** (0.30)	315.00*** (0.15)
Observations	468	468	468	468
R^2	0.97	0.97	0.98	1.00
Adjusted R^2	0.97	0.97	0.98	1.00
Residual Std. Error	2.62 (df = 466)	0.01 (df = 466)	2.18 (df = 465)	0.72 (df = 454)
Note:				*p<0.1; **p<0.05; ***p<0.01

B References

ESRL global monitoring laboratory - mauna loa observatory n.d. Available at https://gml.noaa.gov/obop/mlo/

Global monitoring laboratory - carbon cycle greenhouse gases n.d. Available at https://gml.noaa.gov/ccgg/about/co2_measurements.html

Keeling, **C D** and **Whorf**, **T P** 1997 Scripps institution of oceanography (SIO), university of california, la jolla, california USA 92093-0220.

Keeling Curve 2024. *Wikipedia*. Available at https://en.wikipedia.org/w/index.php?title=Keeling_Curve&oldid=1223091306 [Last accessed 10 July 2024].