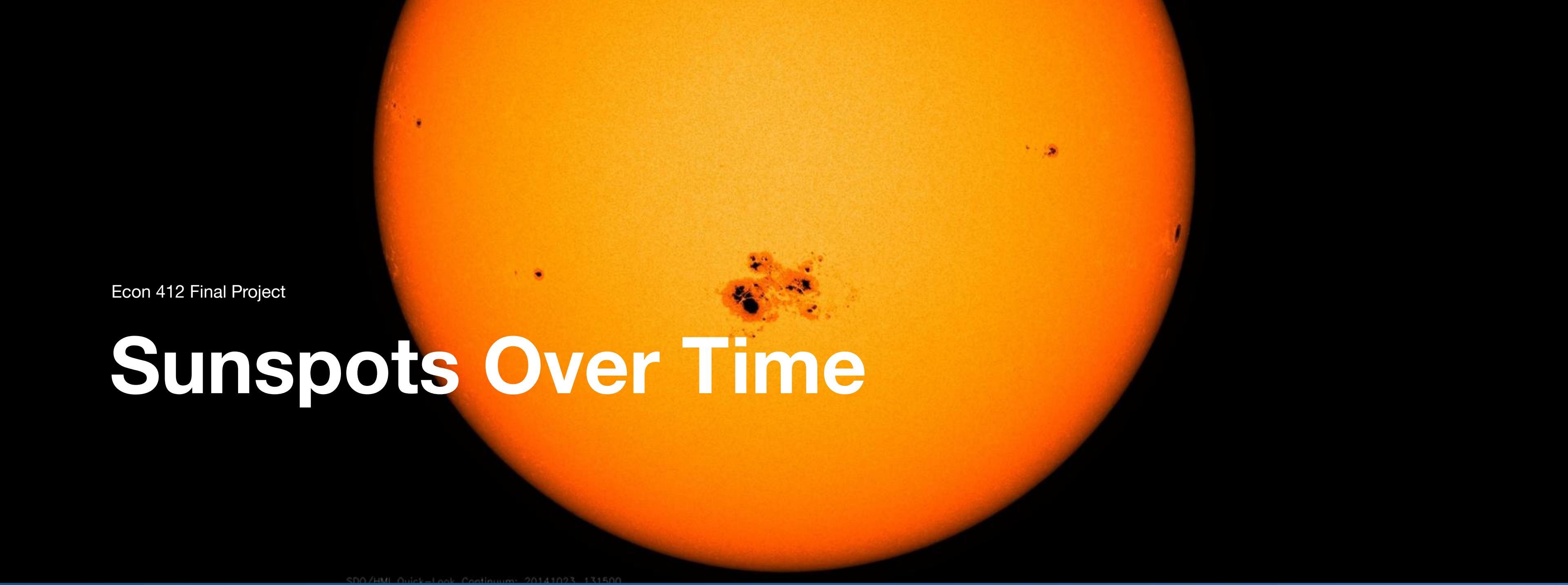
Econ 412 Final Project

Sunspots Over Time



Dylan Schmerer
Rhea Sethi
Talia Lieberman

Econ 412: Fundamentals of Big Data

Dr. Rojas

June 7, 2023

Motivation

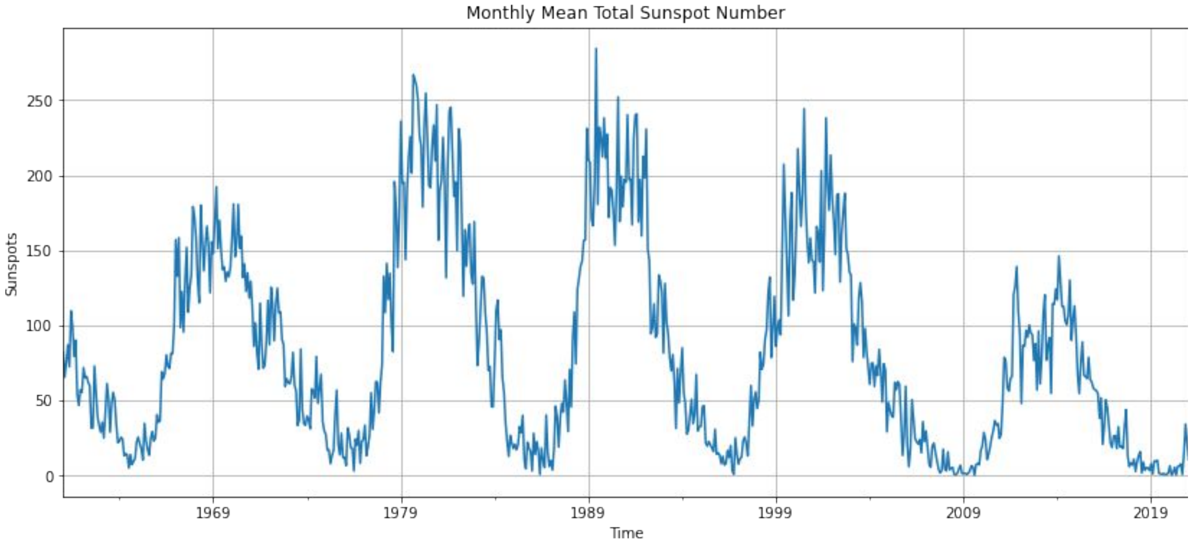
What are sunspots?

- A relatively cool area appearing on the surface of the Sun
- Sunspot activity follows a **cyclical pattern** known as a solar cycle which lasts approximately **11 years**.
  - Monthly sunspot counts recorded from **1749** to present by the Royal Observatory of Belgium

Why is modeling this series interesting?

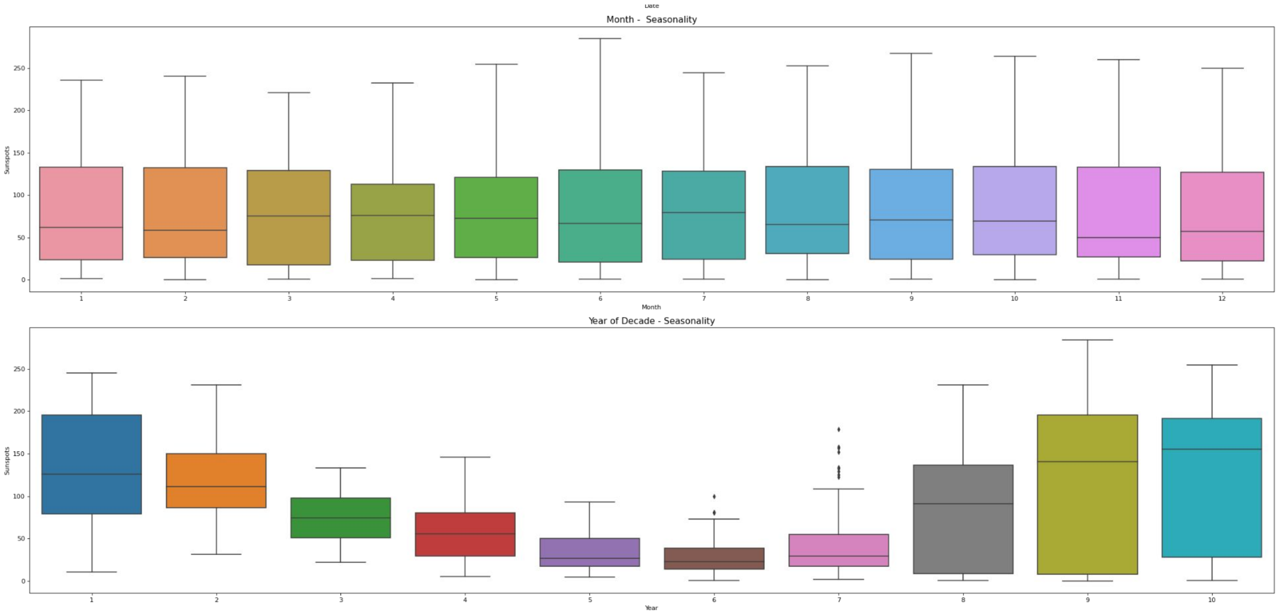
- Sunspots affect the Earth's **climate** and space weather events
- Can cause changes in **agricultural yields** and affect **communication systems**
- Physics based models vs. time series based

Data



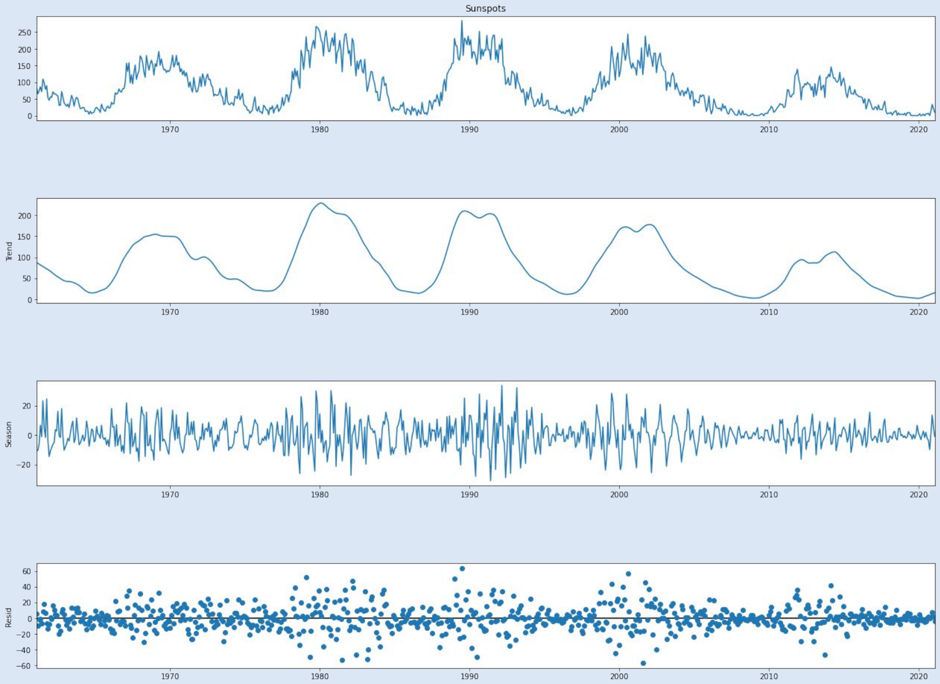
- **Variable of interest**: Mean number of sunspots per month
- **Frequency:** Monthly
- **Dates:** 1961 - 2021
- **Main challenge:** 11 year solar cycle (11yrs x 12mos = 132mo cycle)

Exploratory Analysis



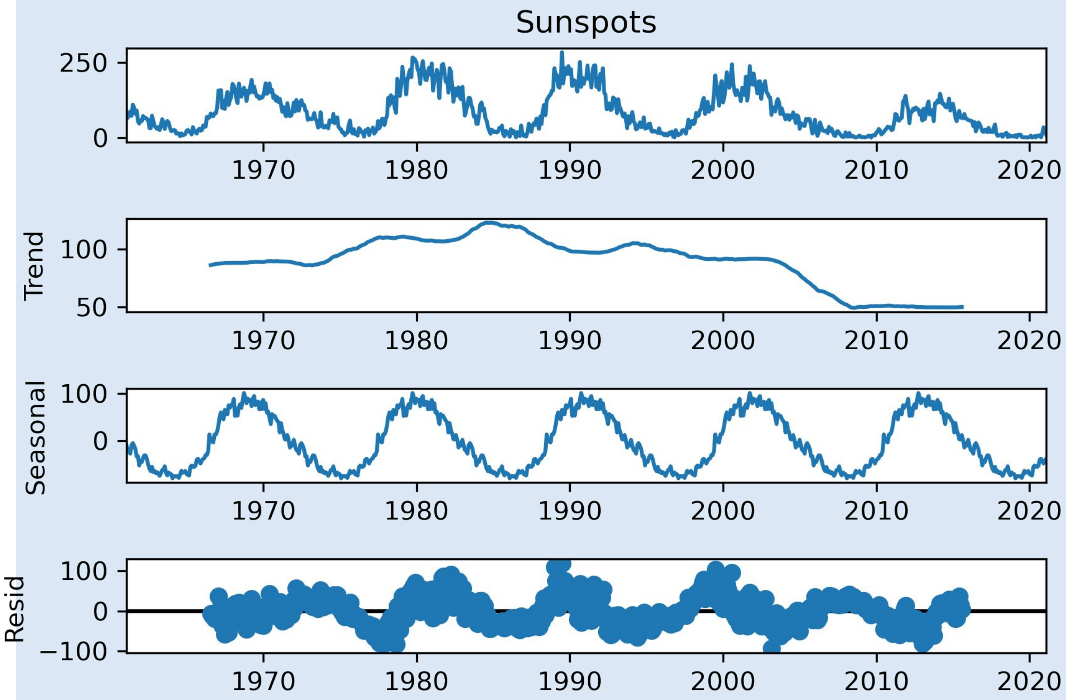
- **Seasonality** not present. Data does not vary month to month.
- **Cyclicity** present. Variation in years of a decade, highlighting 11 year solar cycle.

STL (12 month period)



- Using additive decomposition with 12 month period, the 11 year cycle is captured in the trend
- Seasonality exhibit a pattern and changing amplitude over time and changing
- Residuals also exhibit a clear pattern

Seasonal Decomposition (132 month period)



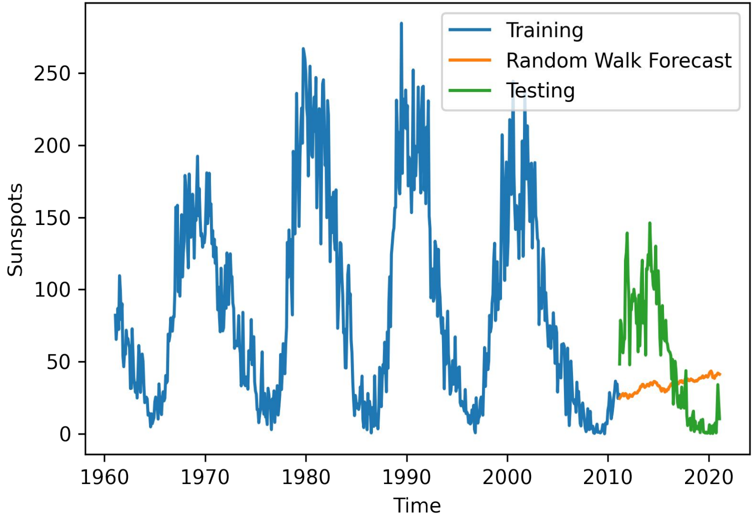
- For comparison, we try additive decomposition with a 32 month 'period' to see how this captures the 11 year cycle.
- The trend flattens out comparatively, and the 11 year cyclicity now appears in the seasonal component.
- Residuals still exhibit a clear pattern.

MODELS

We fit several models on our training set (1961-2011) and try to forecast the next 10 years (testing set). We use the out-of-sample Mean Absolute Percentage Error to evaluate our models' performance.

Note: While there are other metrics like RMSE, MAE etc. for model evaluation, we choose the MAPE owing to its interpretability

Baseline Model: Random Walk



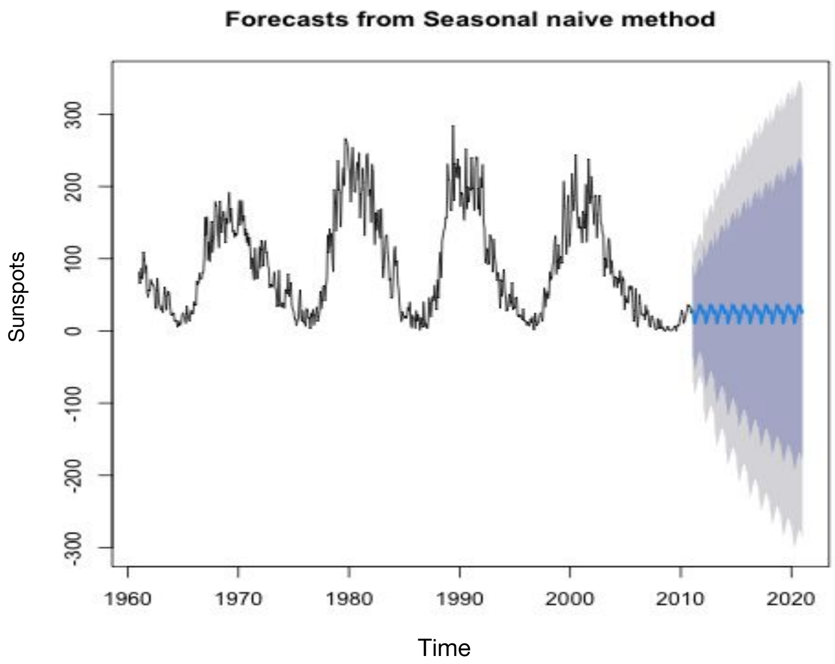
Model Description:

- Assumes future values will be the same as the most recent observed value
- Equation: $Y(t+1) = Y(t) + \varepsilon$
- We use this model as a point of comparison for the upcoming, more complicated models

Performance:

- **MAPE: 941.98%**
- As seen in the plot of the forecast as well as in the exorbitant MAPE, this model fails to capture the dynamics of our data.

Seasonal Naive



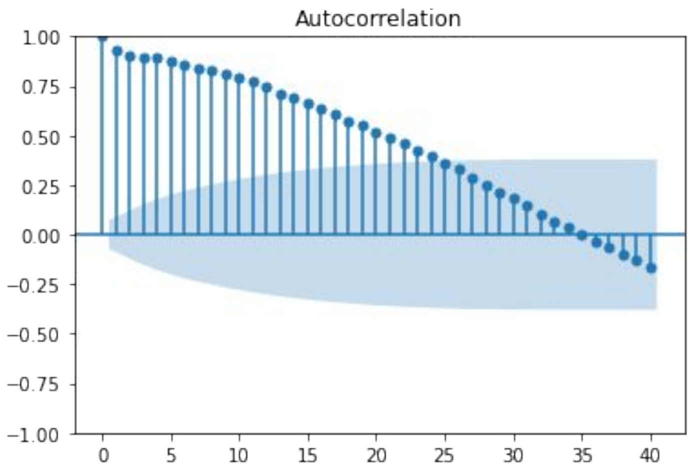
Model Description:

- Uses last season's value as this season's projection
- Builds on Random walk ; still a relatively simplistic model

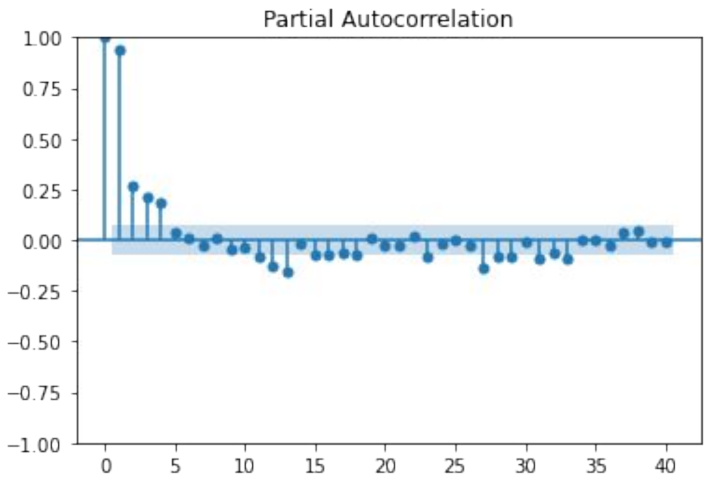
Performance:

- **MAPE: 628.21%**
- Improvement over the Random Walk Forecast but still doesn't capture the cyclical patterns

Series ACF and PACF



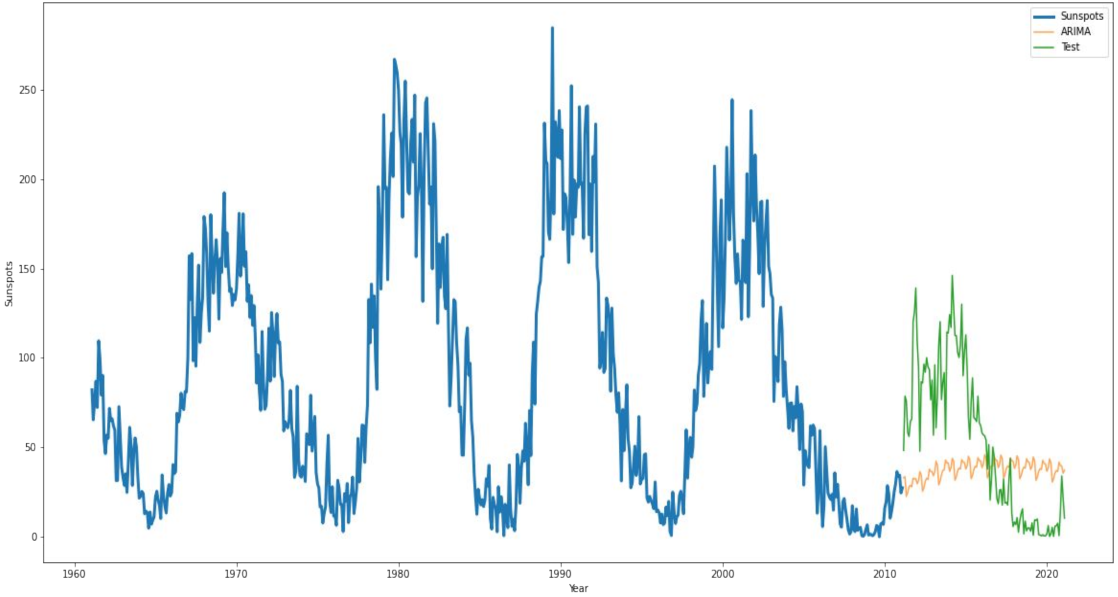
ACF: Decays to 0



PACF: 4 significant lags

Indicative of an AR (4) process with some additional underlying patterns (stemming from the 11-year cyclicity) which we try to model using seasonal terms

ARIMA (Period = 12 months)



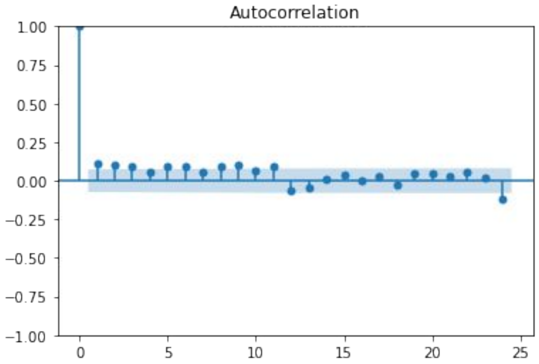
Model Description:

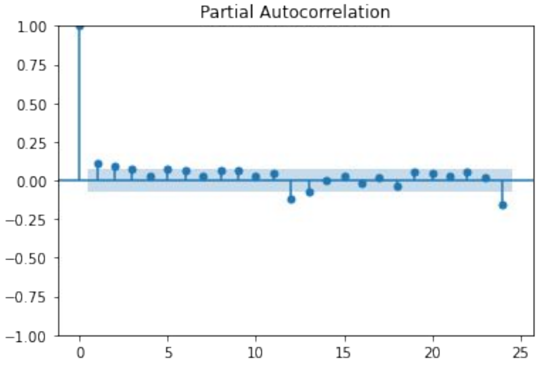
- ARIMA (4, 0, 0)(2, 1, 0)[12]
  - Autoregressive order 4
  - Seasonal autoregressive order 2
  - One seasonal differencing
  - No moving average terms
  - No seasonal moving averages
  - Seasonal period of 12

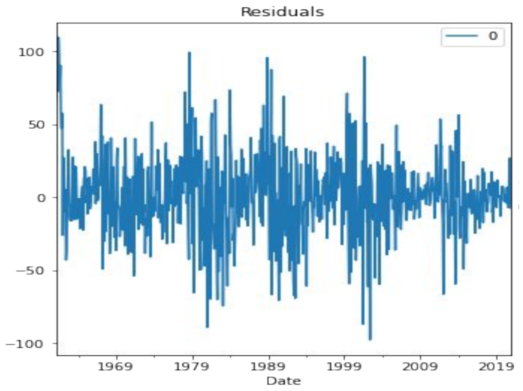
Performance:

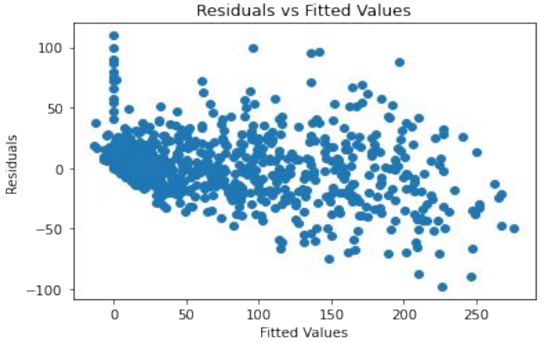
- **MAPE: 923.53%**
- Does not model 11 year cycle very well

ARIMA (Period = 12): Residuals



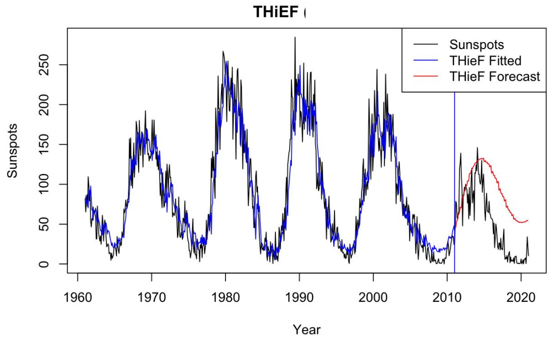






- While the ACF and PACF look clean overall, there are some significant spikes.
- The residuals exhibit non-constant volatility
- Therefore, there are still some dynamics left to model

Temporal Aggregation Methods

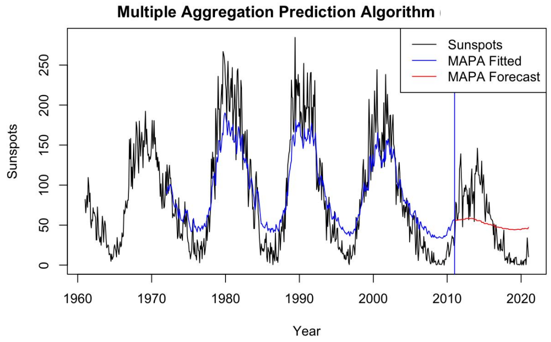


THiEF Model Description:

- Adapts forecast combination weights based on the series' performance on different levels- yearly, quarterly, monthly etc.
- THiEF period = 12

Performance:

- **MAPE: 1303%**



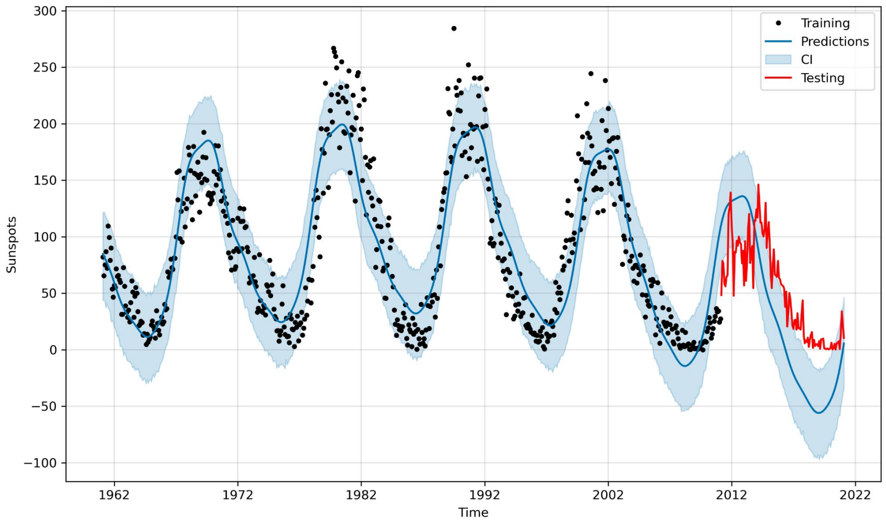
MAPA Model Description:

- Adapts model parameters to different time resolutions and determines the optimal level of aggregation for forecasting
- MAPA period = 12

Performance:

- **MAPE: 1042%**

Prophet



Model Description:

- Why is it useful for our data?
  - Highly customizable specification
  - We specify "decadely" seasonality to capture the 11-year cycles
  - We use a "multiplicative" mode since the amplitude changes over time
  - Fourier order "5" does best

Performance:

- **MAPE: 342.31%**
- This model performs significantly better than the others

Model Rankings

Models	Performance (MAPEs in %)
Prophet	342.31
Seasonal Naive	628.21
ARIMA	923.53
MAPA	1042.08
NNAR	1308.23
THIEF	1310.82
ETS	1607.58
Holt-Winters	3113.90

Combine the 3 best models (These beat the Random Walk)

- It is interesting to note that as MAPEs increase, the complexity of models decreases
- Including these suboptimal models can deteriorate performance of the combination

Combinations

I. OLS Combinations

OLS1- ***Prophet, ARIMA, Seasonal Naive***

- **MAPE: 461.38%**
- Seasonal Naive gets a negative weight of -0.09, we try dropping it

OLS2- ***Prophet, ARIMA***

- **MAPE: 406.59%**
- Performance of the combination improves when we drop Seasonal Naive

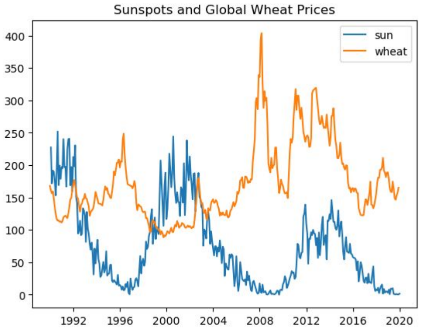
II. Combinations using Mean

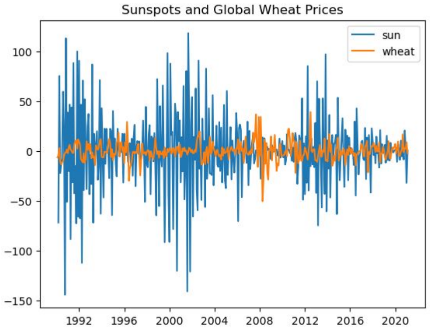
Gives equal weight to all 3 models

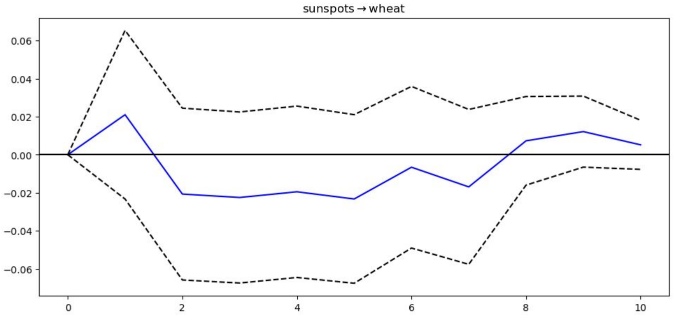
MAPE: 437.28%

- OLS2 > Combination using Mean > OLS1
- None of the combinations beat Prophet alone

VAR: Sunspot Effect on Global Wheat Price







- Research shows inconsistent findings on the effect of sunspots on wheat prices

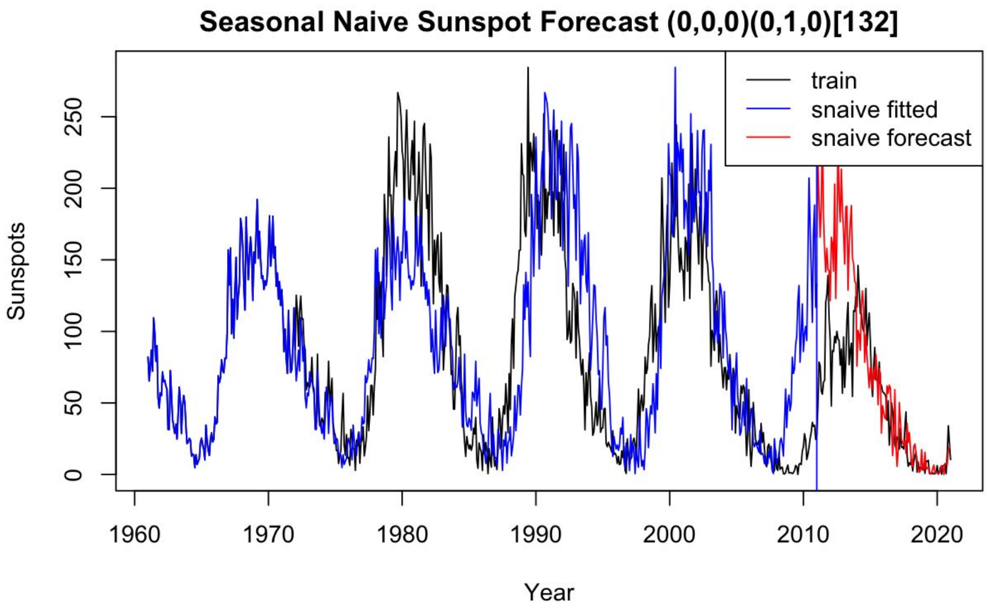
- Differenced to make stationary; deflated wheat prices with commodity ppi
- VAR(3) with significant coefficients on sunspot lags
- No granger causality
  - p-value = 0.58

- We find no evidence that sunspot activity affects wheat price

BONUS MODELS!

Since the traditional models don't seem to perform very well given the complexity of our data, we fit some additional models to try and improve our forecast performance

Seasonal Naive (132 mo ‘Season’)



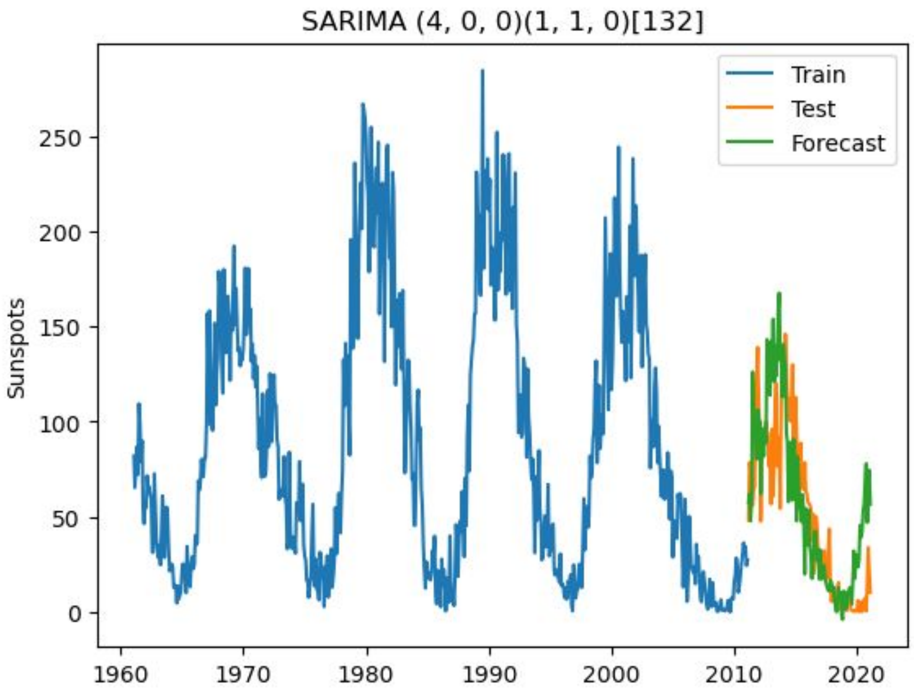
Model Description:

- Modified Periodicity of 132 months to capture the 11-year cyclicity

Performance:

- Using 132 month year seasonality, this model is a big improvement on the simple Seasonal Naive and also on the other models fit earlier
- **MAPE: 145%**

ARIMA (Period = 132 mos)



Model Description:

- ARIMA (4, 0, 0)(1, 1, 0)[132]
  - Autoregressive order 4
  - Seasonal autoregressive order 1
  - One seasonal differencing
  - No moving average terms
  - No seasonal moving averages
  - ‘Seasonal’ period of 132

Performance:

- **MAPE: 94.09%**
- This Modified ARIMA outperforms all the models fit earlier

LSTM

Why this model?

- Our data exhibit complex long-term dependencies
- LSTM can capture these in the memory cell which stores and propagates information to future time steps

Performance:

- **MAPE: 93.29%**
- Even without optimizing the hyperparameters or architecture of the model, it turns out to perform the best overall.
- Fine-tuning this model can lead to even better predictive performance

Limitations:

- Tradeoff between Accuracy and Interpretability (from Econ 430!)
- Though this model yields much better accuracy, it is less interpretable than models like ARIMA

Conclusion and Future Work

- The key takeaway from this analysis is that more advanced, state-of-the-art models like LSTM, Prophet work better for data as challenging as this one. We saw how much traditional models like ARIMA and Seasonal Naive need to be modified to get them to perform well.

- Some potential improvements for the future could be:
  - Smoothing the series- By using a 3-month rolling mean of the data (for example) or State Space Models that help to extract the signal from the noise
  - GARCH- Modelling the raw data as a GARCH process to predict volatility
  - Larger sample size with data from earlier periods to train our models

References

- Kaggle. (2021). Sunspots [Data set]. https://www.kaggle.com/datasets/robervalt/sunspots
- Kurnia, D. (2019). Time series Sunspots. https://rpubs.com/danielkurnia/timeseries_sunspots
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts. https://otexts.com/fpp3/
- Rojas, R. (2023). Econ 412, Course Material. Retrieved from BruinLearn
- Kunz, N. (2023). Econ 412, Course Material. Retrieved from BruinLearn

Thank You