Time Series Analysis of Public Legal Interest: Dual Approach of Long-Term Patterns and Short-Term Events

Donghyun Ahn, Department of Data Science, Korea University

1. Introduction

As a student majoring in Public Governance and Data Science, I aimed to determine whether there are seasonal or time-related shifts in public interest within the legal domain. I am also interested in understanding how surges in event-related public interest return to baseline levels over the short term. By predicting both the long-term trends in legal interest and the short-term spikes associated with specific events, we could provide valuable insights for more effective legal policymaking and media strategies.

Legal interest—encapsulating the public's interest in justice, legislative activities, and high-profile trials—serves as a crucial barometer of the social and political climate. Understanding fluctuations in this sentiment is essential not only for gauging immediate public reactions to legal developments but also for forming a stable foundation for future-oriented, data-driven policymaking. This study addresses these needs by examining trends and influences on legal sentiment in South Korea using a time-series model, aiming to uncover patterns, periodicities, and external influences that shape public interest.

2. Datasets Description

In this study, the data is divided into two main analytical approaches: long-term time series analysis and outlier analysis for focused legal sentiment investigation.

1. **Long-Term Time Series Data Sources**: This dataset consists of Google Trends data from October 2019, focusing on four keywords: **Justice, Parliament, Economy,** and **Trial** within South Korea. Each term represents a facet of Korean society's collective concerns and reflects public interest in law, governance, economic matters, and legal proceedings over time. This structured dataset, with "Date" as the time index and the four keywords as separate columns, enables a unique view into the temporal patterns of public attention toward justice-related topics. Moreover, the reason why we include the term "Economy" is to figure out some specific events with new perspective that could influence huge fluctuation on Public Sentiment.
2. **Short-Term Time-Related Data Sources**: We aim to make DL model to handle out short term osciliation. Outlier analysis investigates sharp deviations from general trends, focusing on specific events (e.g., major legislation, specific crimes or trials, economic crises) to assess their impact on public legal sentiment. After finding the several events, we will collect time series data from other sources or more precise data. By analyzing whether searches for legal terms around certain events, this outlier analysis offers insights into how pivotal moments affect public interest in terms of time series.

3. Goals and Hypotheses

3.1. Goals

This study aims to develop a comprehensive understanding of public interest and semtiment trends toward legal topics, employing both long-term and short-term modeling approaches:

1. **Building a Stable Long-Term Model for the Legal Domain**: The primary goal is to construct a stable long-term model of legal sentiment. By identifying seasonal patterns and trends, we seek to offer a framework that supports ongoing forecasting within the legal domain. This model can serve as a predictive tool for understanding gradual long-term changes in public interest in justice, governance, and economic matters related to legal sentiment over time.
2. **Exploring and Analyzing Rapid Increases in Public Interest**: By examining significant spikes in search volumes, this study aims to uncover and explain public reactions to key legal and socio-political events. Such spikes might happen with notable trials, legislative changes, or economic disruptions, making it essential to understand these outliers for a complete view of public legal sentiment.
3. **Developing a Short-Term Model for Rapid Sentiment Shifts**: In addition to the long-term model, we aim to build a short-term model capable of identifying and responding to rapid shifts in sentiment. This model

will capture sudden increases in public interest, helping policymakers and stakeholders assess and address real-time public sentiment in response to specific events or emerging issues.

4. **Insights for Policy Adaptation**: By exploring both long-term trends and short-term spikes in legal area, this research ultimately seeks to provide insights that inform adaptive, responsive policy decisions. A robust understanding of public sentiment toward legal issues can offer guidance on how best to address public concerns, improve transparency, and build trust in legal institutions.

3.2. Hypotheses

- **Hypothesis 1**: Long-term trends will result in long-term time series patterns because legislation-related activities are cyclical.

- **Hypothesis 2**: Outliers or surges of interest in the legal field will likely follow the common model in the short term.

4. Data Exploratory Analysis

To understand the underlying characteristics of our dataset, we conducted an exploratory data analysis (EDA) on the search volume trends for each keyword. This EDA includes summary statistics, initial visualizations, and correlation analyses that provide insights into potential patterns and relationships within the data.

4.1. Basic EDA

- **Missing Value** There are no missing values in these dataset.

- **Interval Consistency** each date has same counts.

- **Outliers** There are some outliers, and we will conduct further test for them.

4.2. Basic Line Plots

To observe the general trends over time, we created line plots for each keyword: *Justice*, *Parliament*, *Economy*, and *Trial*. These plots, shown in Figure 1, reveal both the periodic patterns and fluctuations in search interest over time.

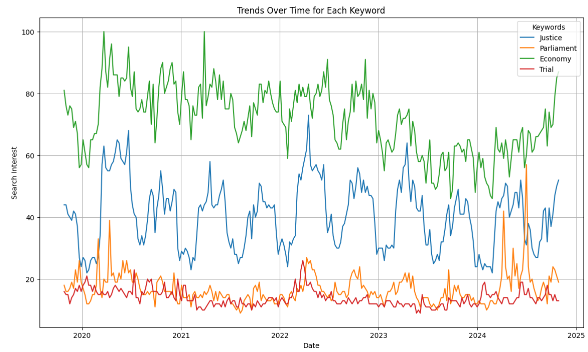

Figure 1. Time Series Plots of Search Interest for Each Keyword

Key observations from these plots include:

- **Justice and Economy** show a positive correlation.

- **Parliament and Justice** also exhibit a positive correlation, suggesting that legislative actions may influence interest in justice-related topics.

- **Trial** shows less correlation with other keywords, likely due to its event-driven nature and the influence of high-profile legal cases.

- Some outliers exists in these period. It seems that there are some specific events on South Korea.

4.3. Data Distribution

- **Histogram**: A histogram provides a visual representation of the distribution of values for each variable.

- **Q-Q Plot** : The Q-Q plot compares the distribution of the data to a theoretical normal distribution.

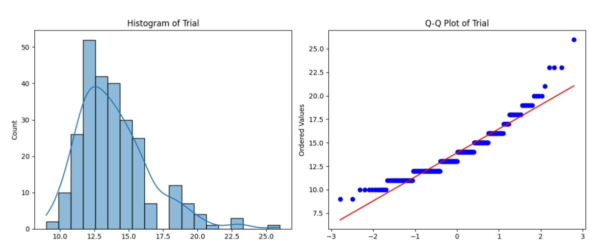

Figure 2. Justice Distribution

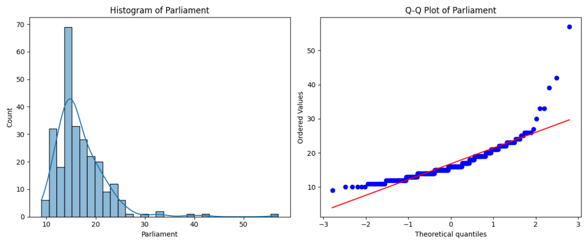

Figure 3. Parliament Distribution

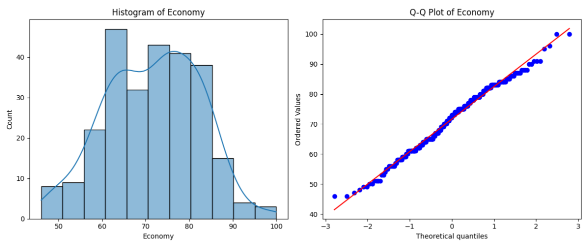

Figure 4. Economy Distribution

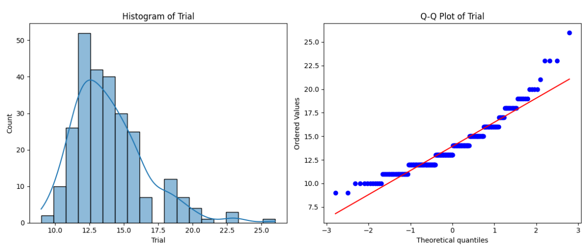

Figure 5. Trial Distribution

- **Trial and Parliament**: These variables show right-skewed distributions, confirmed by the deviations in their Q-Q plots, suggesting potential benefits from transformation before applying normality-dependent models.
- **Economy and Justice**: These distributions are closer to normal, as seen in both histograms and Q-Q plots.

4.4. Summary Tables

Table 1 presents summary statistics (mean, median, standard deviation) for each keyword, offering a quantitative view of the distribution and variability in search interest.

Keyword	Mean	Median	Standard Deviation
Justice	40.507634	41.500000	10.716744
Parliament	16.805344	16.000000	5.176232
Economy	71.622137	72.500000	10.798324
Trial	13.923664	13.000000	2.649711

Table 1. Summary Statistics for Each Keyword's Search Interest

4.5. Correlation Matrix

To evaluate the brief relationships between keywords, we generated a correlation matrix (Table 2).

	Justice	Parliament	Economy	Trial
Justice	1.000	0.428	0.613	0.162
Parliament	0.428	1.000	0.198	0.210
Economy	0.613	0.198	1.000	0.230
Trial	0.162	0.210	0.230	1.000

Table 2. Correlation Matrix for Search Interest Across Keywords

Correlation matrix shows:

1. **Justice and Economy** have a moderate positive correlation (0.61), suggesting economic conditions may influence public interest in justice topics.
2. **Justice and Parliament** show a weaker positive relationship (0.42), implying a minor impact of legislative activities on justice interest.
3. **Economy and Parliament** have a low correlation (0.20), indicating relative independence in public interest.
4. **Trial** shows low correlations with other keywords, reflecting its event-driven nature, with public interest in trials largely independent of broader justice, economy, or parliament trends.

But there could be some lagged relationship, seasonal trend and some outliers so we should watch these results just for brief result.

4.6. ADF Test

Looking at the basic line plot, it appeared that the time series data exhibited some degree of trend and seasonality so we decided to conduct further exploratory data analysis.

In the context of time series analysis, a stationary series has a constant mean and variance over time, which is important for making reliable predictive model. Non-stationary time series data, which often have trends or seasonality, can lead to inaccurate models if not properly handled. To figure out stationarity, we conducted an ADF test (Table 3)

The test operates under the following hypotheses:

- **Null Hypothesis (H0)**: The time series has a unit root, meaning it is non-stationary.
- **Alternative Hypothesis (H1)**: The time series does not have a unit root, meaning it is stationary.

If p-values are less than 0.05, we will reject the Null Hypothesis. And we conducted three tests: the first test(1st row) did not account for seasonality, the second test(2nd row) considered a 6-month cycle, and the third test(3rd row) used a 1-year cycle for validation.

	Justice	Parliament	Economy	Trial
p-value(1)	0.0031	0.0005	0.041	0.0095
p-value(2)	0.009	0.017	0.023	0.0014
p-value(3)	0.114	0.045	0.327	0.058

Table 3. ADF Test for Each Keword

All four series show p-values below 0.05 in test1 and test2, leading us to conclude that each time series is stationary. However, all four series in test3 have large p-values that can reject our Null hypothesis, asserting that there are some patterns(non-stationary). Hence, we have to figure out the model in seasonal approach.

4.7. ACF, PACF

- **ACF Plot**: Shows the correlation of a time series with its own lagged values. We can identify the overall structure of dependencies over different time lags, which is crucial for detecting seasonality and trend in the data.

- **PACF Plot**: Displays the correlation of the time series with its lagged values after removing the effect of intermediate lags. This is particularly useful for identifying the direct effect of each lag, aiding in determining the order of autoregressive (AR) terms in modeling.

- **Seasonality**: Repeating peaks at regular intervals in the ACF suggest seasonal patterns.

- **trend**: A gradual decline in ACF values indicates trend data, which might require differencing to achieve stationarity.

- **AR and MA Components**: A quick drop-off in PACF or ACF values at certain lags can signal AR or MA components, which guide the choice of model parameters.

Justice and **Economy** series show signs of seasonality or trend, which might require seasonal adjustments or differencing in a time series model. The **Parliament** and **Trial** series suggest weaker dependencies, indicating simpler models and stationary. These findings help guide the selection of ARIMA or seasonal models for each series.

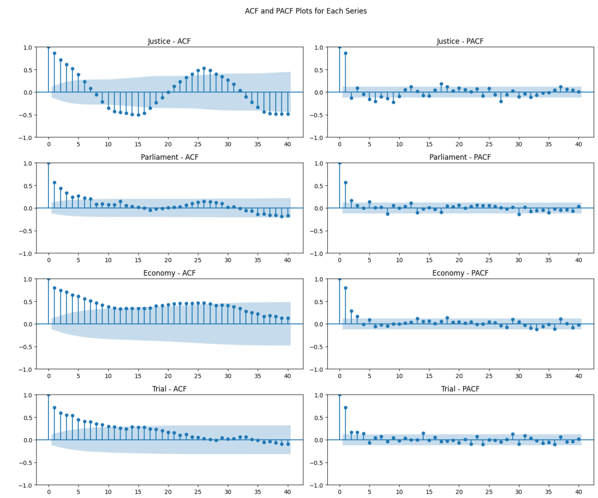

Figure 6. ACF and PACF Plots for Each Keyword

5. Long Term Modeling Process

5.1. Seasonal Decomposition and GARCH Test

We focus on analyzing the seasonal patterns and volatility of public sentiment data. We first use the stepwise SARIMAX method to decompose seasonality. Then, we perform the ARCH test using the residual square term to assess the presence of volatility clustering, offering a deeper understanding of the dynamic nature of sentiment variations.

1. **Justice**: The best model is SARIMAX (1,0,0) (1,0,0)(52). Its residual follows a stationary white noise process. The ARCH p-value is 0.2769, larger than 0.05. Therefore, we accept the null hypothesis: ARCH modeling is unnecessary.

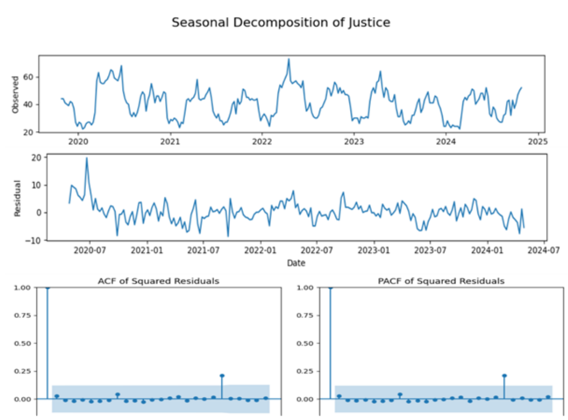

Figure 7. Residual and ARCH Analysis on Justice

2. **Parliament**: The best model is SARIMAX (1,0,1) (0,0,0)(52). The ARCH p-value is 0.02, smaller than 0.05. Therefore, we reject the null hypothesis: additional ARCH modeling is needed.

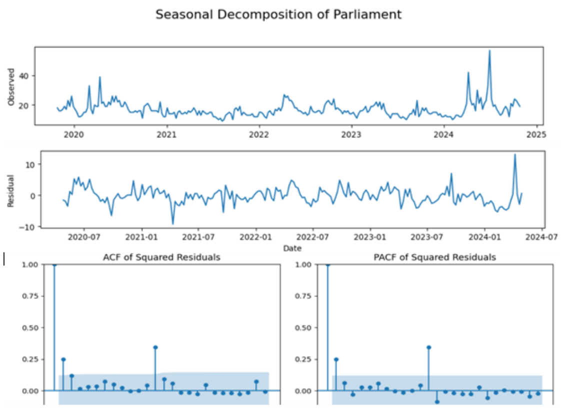

Figure 8. Residual and ARCH Analysis on Parliament

3. **Economy**: The best model is SARIMAX (0,1,1) (1,0,0)(52). Its residual follows a stationary white noise process. The ARCH p-value is 0.1, larger than 0.05. Therefore, we accept the null hypothesis: ARCH modeling is unnecessary.

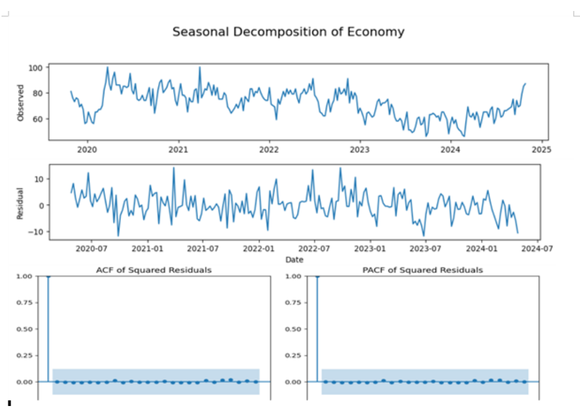

Figure 9. Residual and ARCH Analysis on Economy

4. **Trial**: The best model is SARIMAX (1,1,1) (0,0,1)(52). The ARCH p-value is almost zero, smaller than 0.05. Therefore, we reject the null hypothesis: ARCH modeling is necessary.

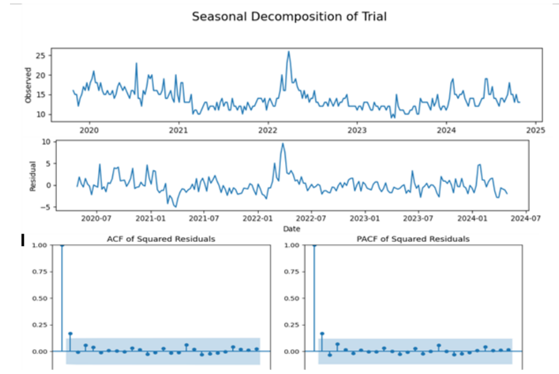

Figure 10. Residual and ARCH Analysis on Trial

We will check out outliers with one model in one data. Also, we use Google Colab T4-GPU resource. We tune the hyperparmeter with AIC values and MAPE values.

5.2. Justice Data Model fitting

We figured out best hyperparameter of SARIMAX model and the fact that we didn't need to handle out residual volatility. And we will compare some methods including basic SARIMAX model, Time Series Transformer, and pretrained-Prophet model.

1. **SARIMA Model**: We fitted data with SARIMAX (1, 0, 0) x (1, 0, 0, 52) model. It seems it well illustrates and predicts these data series.

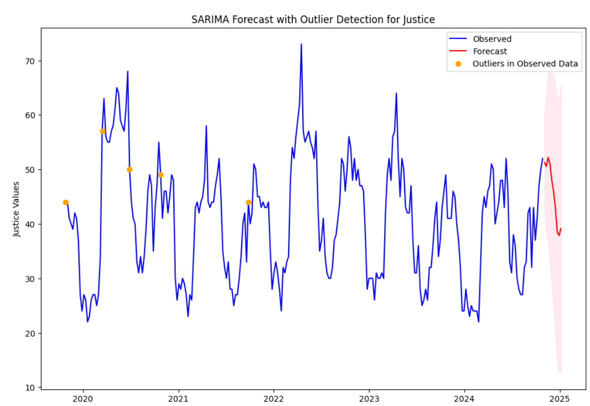

Figure 11. SARIMA Modeling on Justice

2. **Time Series Transformer**: We make transformer-based model to fitting the data. We slice our data with input windows and output windows to treat them like sequence like NLP. There are several Time Series

Transformer models in huggingface and github. We combine others code to make more suitable to our data.

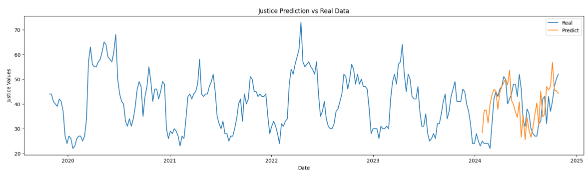

Figure 12. Transformer Modeling on Justice

3. **Prophet**: Lastly, we used pre-trained time series model named Prophet. It is powerful model that developed by Facebook. Unlike previous models, prophet has convenient illustration tools.

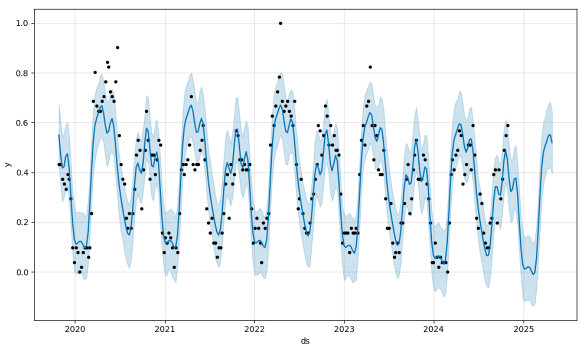

Figure 13. Prophet Modeling on Justice

5.3. Parliament Data Model fitting

We will first apply SARIMAX + GARCH Model. Then, we will fit our data with LSTM model and ETS model.

1. **SARIMA + GARCH Model**: We fitted data with SARIMAX (1, 0, 1) x (0, 0, 0, 52) model. Then we make GARCH(1,1) model with residuals term. Finally we combine those two model to predict our data.

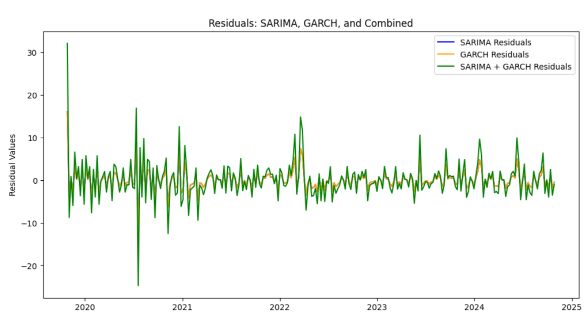

Figure 14. Model Decomposition

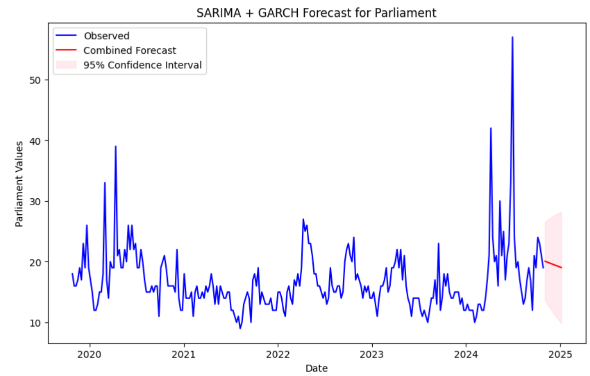

Figure 15. SARIMA + GARCH Modeling on Parliament

2. **LSTM**: We make LSTM-based model to fitting the data. We create sequence for LSTM model. It seems it didn't predict not very well, but when the data size gets more bigger, it could make more suitable prediction.

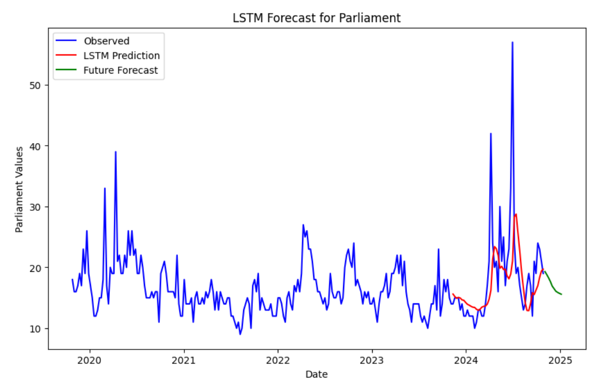

Figure 16. LSTM Modeling on Parliament

3. **ETS**: Lastly, we used pre-trained time series model ETS. It is powerful model that considered exponential impact on time series sequence.

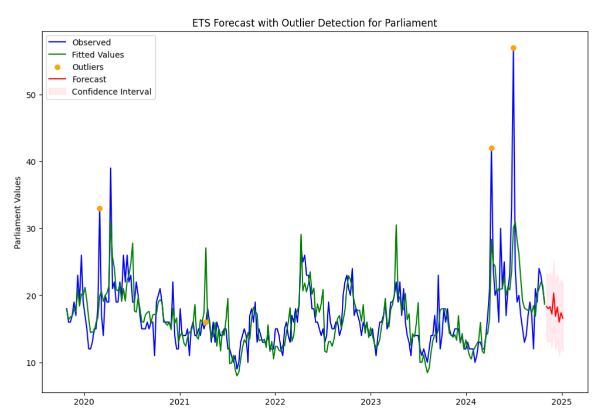

Figure 17. ETS Modeling on Parliament

5.4. Economy Data Model fitting

We will first apply basic SARIMA Model. Then, we will fit our data with ETS-based model.

1. **SARIMA Model**: We fitted data with SARIMAX (0, 1, 1) x (1, 0, 0, 52) model. It didn't well predicts the future term.

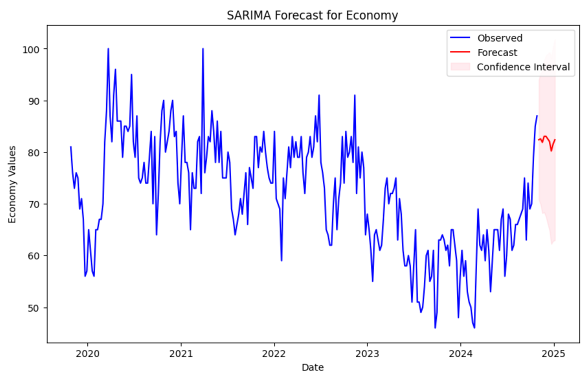

Figure 18. SARIMA Modeling on Economy

2. **ETS**: Lastly, we used pre-trained time series model ETS. It is powerful model that considered exponential impact on time series sequence. It seems well to predict.

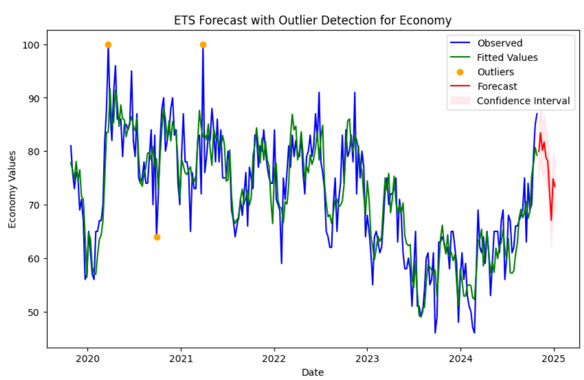

Figure 19. ETS Modeling on Ecnonomy

5.5. Trial Data Model fitting

We will apply SARIMA +GARCH model to handle residual volatility problem.

1. **SARIMA + GARCH Model**: We fitted data with SARIMAX (1, 1, 1) x (0, 0, 1, 52) model. Then we make GARCH(1,1) model with residuals term. Finally we combine those two model to predict our data. And we thought that it seems be flattened because their near time series data possess small variance.

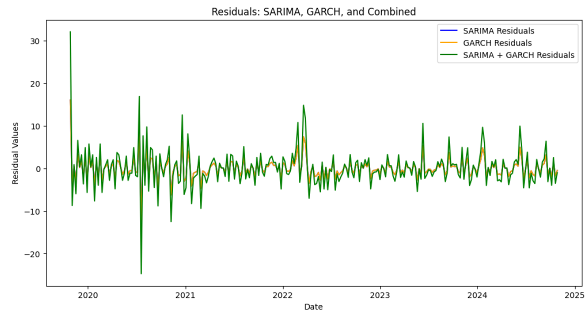

Figure 20. Model Decomposition

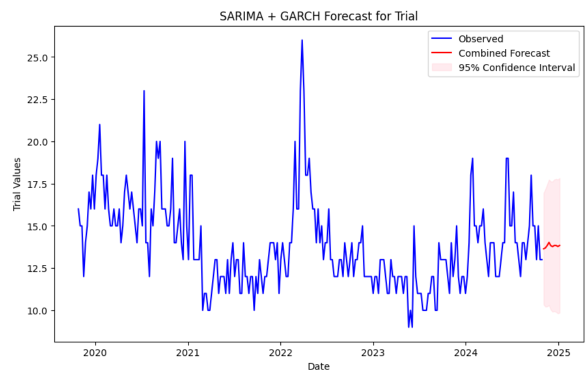

Figure 21. SARIMA + GARCH Modeling on Trial

6. Semi-Result

We successfully make long term models on these time series topic. We apply several models. Simple cyclic time series can be more well fitted in simple SARIMA model, but when the data become more complex and bigger, deep learning based model fits and predicts well. By these modeling, policy makers could consider interest of people. My dream is to be a policy and law makers in science and technology.

And we faced some significant obstacles in our research. Not only did we compare the testing data, but we also attempted to predict future values and planned to compare the results two weeks later. However, a major issue arose in South Korea, causing all four topics to exhibit abnormal outliers, which led to the failure of further validation.

7. Plan for Future Research

1. **Outlier Detection and Short-Term Model**: Through long-term modeling, we identified several outliers and aimed to proceed with short-term modeling. However, collecting data for past short-term events posed challenges, as past Google Trends data often has wider time intervals, such as monthly, making it difficult to obtain sufficient data points. So to do with these short term analysis project, we should gather nowadays political

trend data steadily.

For Example, we found common outliers, Sewolho accident and Itaewon accident. And these two events seems like similar trending graph. Moreover, we want to train the model with these data to successfully get short-term model.

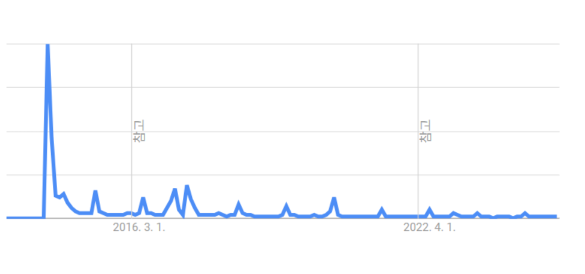

Figure 22. Sewol-ho dataset

Figure 23. Itaewon dataset

2. **Multivariate Modeling**: Finally, we want to sum all of these model into one model with ”social interest on public and law” term. To achieve effective model, we have to concern about correlation or multivariate modeling terms.

8. Related Works

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). *Dive into Deep Learning.* MIT Press. Available at: https://d2l.ai/

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*.

Busseti, E., Osband, I., & Wong, S. (2012). Deep learning for time series modeling. *Technical report, Stanford University*, 1-5.

Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. *IEEE Sensors Journal*, *21*(6), 7833-7848.