Time Series Analysis of *Wicked*

Shane Aaron Cu

PHY408H1: Time Series Analysis

Abstract

This experiment aims to understand the relationship of *Wicked* Broadway musical sales to the adapted movie sales. This analysis enables us to see how Broadway gross increase or decrease after a film is released. Using low-pass filter and cross-correlation, there is an increase on the Broadway sales around the time the movie was released. On the other hand, film sales continue to decrease as time progresses. The cross-correlation plot suggests that there is no cross correlation between the Broadway and film sales. This results concludes that no delayed increase or decreased was observed. Furthermore, there is a correlation in lag 10, implying a positive correlation 10 before the film was released. This phenomena happened possibly because of extensive advertisement campaigns. Thus, this experiment refuses the initial hypothesis of delayed increase on Broadway sales.

Contents

1 Introduction

Over the last decade, major motion pictures have continued to adapt Broadway musicals for people to enjoy. This adaptation enables people to watch and listen to their favorite musicals where a theater in their cities is inaccessible. However, it is not certain that a movie adaptation captures the essence of the original Broadway musical. For instance, fans of Andrew Lloyd Weber's *Cats* were greatly disappointed when the star-studded movie adaptation (2019) became one of the major flops due to poor high-tech animation and weak plot progression[1]. Unfortunately, the musical is not playing on Broadway to assess the consequences of this event.

Recently, Jon Chu's *Wicked (Part 1)* (2024) received much praise and attention when the plot and movie production nailed the essence of the original musical[5]. This experiment aims to analyze the relationship between the sales of the movie and the current Broadway show that is running in New York City. This analysis would enable movie producers, movie and musical enthusiasts to know the effect of producing a movie based on a musical. They would be able to see the impact of one variable on the other and clearly determine how flops like *Cats* may or may not harm the Broadway show.

This analysis is done by answering the hypothesis: **How does the release and sales of the Wicked movie compare to the sales of the Wicked Broadway show in New York City?**

I predict that the Broadway sales will have a delayed increase in the movie's initial release. Furthermore, even though the movie sales decreases over time, Broadway sales continue to increase due to mass media exposure and consumer's attention. Hence, this effect will give an positive trend suggesting a correlation between the data sets.

2 Methods

2.1 Chosen Analysis

To answer our question in this experiment, I will use low-pass filter and a correlation analysis, specifically lagged covariance and correlation.

Low-pass filter is appropriate because the raw data presents unusual spikes every new year, suggesting an unusual outlier. Increased holiday prices and special events cause these annual peculiar spikes (see Figure 1). This behavior has been consistent since the beginning of the play in Broadway. Thus, this method enables us to see a smoother trend where the high-frequency noise is removed and asses a unbiased dataset.

Lagged correlation (cross-correlation) is appropriate because this method would see the effects of the initial movie sales to the show sales in the following weeks. This technique could determine if there is a strong lagged relationship between the two datasets. A correlation calculation enables us to see a simplified explanation of the datasets' relationship.

In addition, de-trending the dataset is needed before applying these techniques. I will use `scipy.signal.detrend`.

2.2 Rejected Methods

Some of the rejected alternative methods are Fourier Methods and ARIMA models. Fourier methods and ARIMA models would help me smoothen out the noise of my dataset, however these alternatives would be inadequate to asses the anomalies present in the dataset. An ARIMA model is inadequate for non-linear datasets, while Fourier series fails to disregard the anomalies. These methods could falsify datasets and drive incoherent discussions.

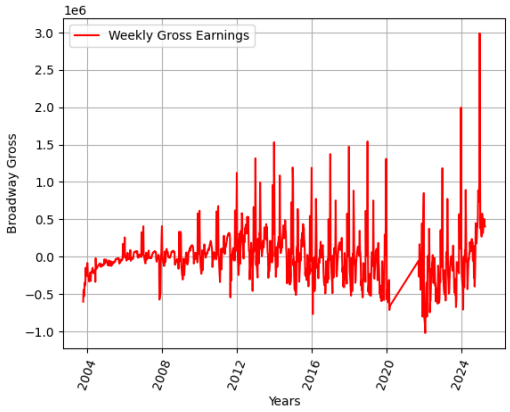

Figure 1: The detrended raw data set of Wicked Broadway sales from 2003 to present

3 Data

Description of the gathered data sets

3.1 Sources

The Wicked Broadway sales dataset was gathered from **BroadwayWorld**[3], the largest theater site that covers all forms of play including off-Broadway and Broadway plays and musicals in New York City, West End and many more.

https://www.broadwayworld.com/grosses/WICKED

The Wicked movie sales was gathered from BoxOffice Mojo from IMDb Pro[2]. This site contains box office receipts of different motion pictures around the world. They offer domestic and international gross earnings for each movie with different time-series.

3.2 Raw Data

The raw data for Wicked Broadway sales will be spanning from the initial movie release to the final screening of the movie (November 2024 - February 2025), recorded on a weekly basis. This four-month starting point will help asses data behavior coinciding with the movie screen time.

It will also be helpful to separately asses the dataset starting from January 2024 to February 2025. This perspective enables us to see the behavior pre-release of the movie. Media attention and publicity could affect pre-release of the movie.

The raw data for the movie sales will be spanning from the initial movie release to the final screening of the movie (November 2024 - February 2025), also recorded on a weekly basis. This span describes the on-time screening of the film in the big screens domestically (USA).

4 Results and Analysis

4.1 Raw Data Analysis

I generated a plot that describes the sales for both Broadway and movie on the same time-span (i.e November 2024 to February 2025.) This time-span describes coinciding time frame of the movie release.

4.2 Low-pass filter Analysis

To filter out the data anomalies, I used Butterworth filter `scipy.signal.butter`. This method provides a smooth and simple filtered graph while maintaining the wanted signal frequencies. It properly filtered out the high-frequency anomalies while creating undistorted frequencies that generates a readable graph.

Since there are weekly samples in a year, the sampling frequency is 52 ($f_s = 52$) To filter out the annual high-frequency anomalies, I determined that the cut-off frequency is 4 because I want to suppress any frequency above 13-26+ -. This process eliminates the weekly noise while maintaining a clear monthly pattern (see Figure 4 and 5)

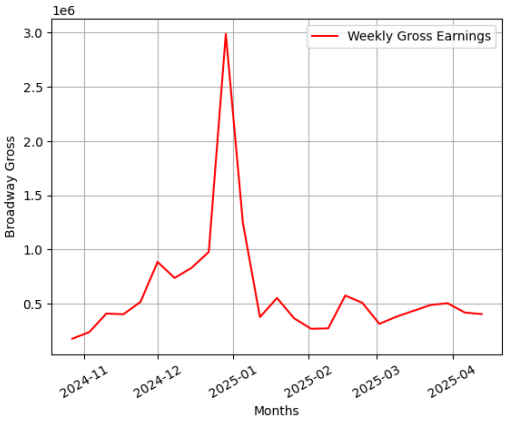

Figure 2: The detrended raw data of *Wicked* Broadway sales from November 2024 to February 2025

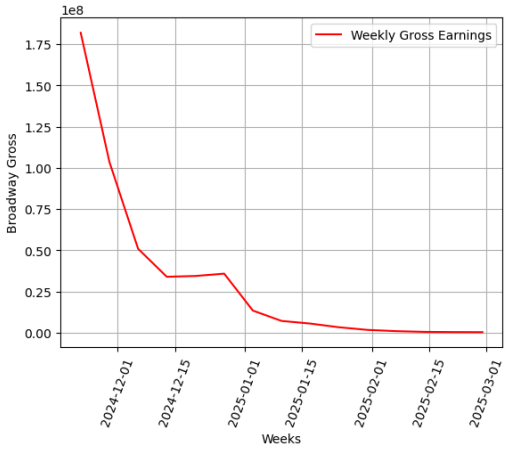

Figure 3: The raw data of *Wicked* movie sales from November 2024 to February 2025

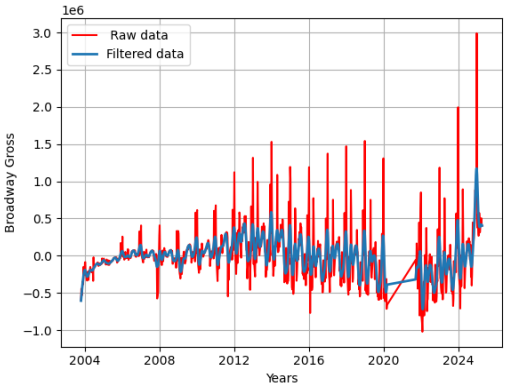

Figure 4: The filtered (low-pass) data of *Wicked* Broadway sales from 2003 to 2025 along with the raw data.

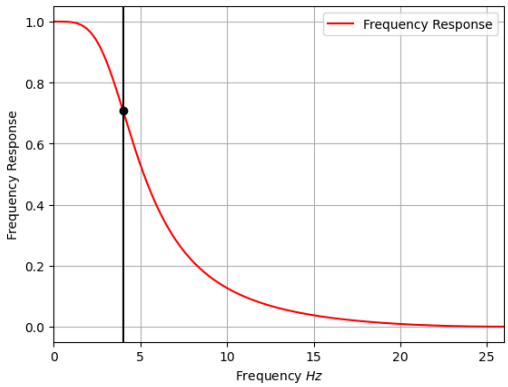

Figure 5: The Frequency response as a function of the frequency

4.3 Filtered Data Analysis

I plotted a the filtered data of Wicked Broadway and movie sales from November 2024 to Feb 2025. In reference to Figure 6, there is a general upward trend around November to December, and plateaus after New Year's As the same time, there is a downward trend on movie sales over time (see Figure 7). To further generalize the filtered data, I included a year overview of the data pre-release (see Figure 8). I also included the cross-correlation stem graph as a function of the lags. This normalized lagged covariance determines the relationship of the Broadway and movie sales by examining the lag regression for each time (see Figure 9)

$$Corr(X_t, Y_{t-\tau}) = \frac{Cov(X_t, Y_{t-\tau}}{\sqrt{Var(X_t)Var(Y_{t-\tau})}}$$

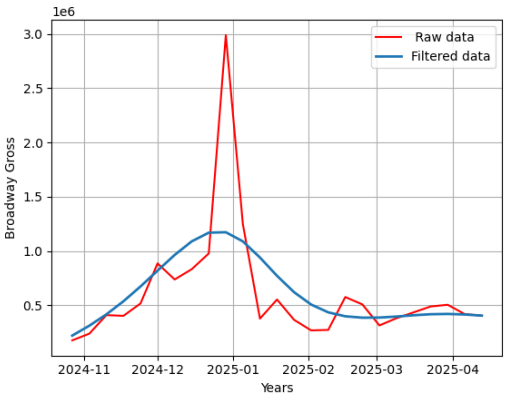

Figure 6: The filtered (low-pass) data of *Wicked* Broadway sales from November 2024 to February 2025 along with the raw data.

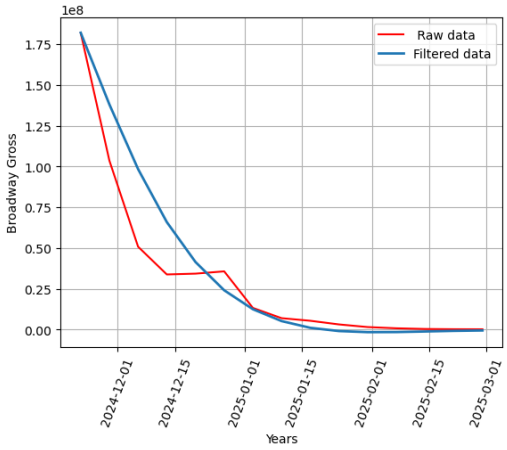

Figure 7: The filtered (low-pass) data of *Wicked* movie sales from November 2024 to February 2025 along with the raw data.

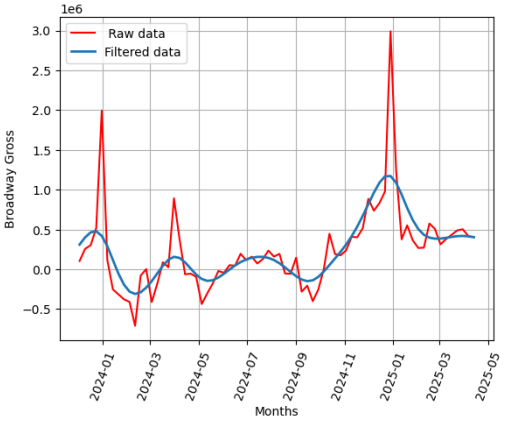

Figure 8: The filtered data from January 2024 to February 2025.

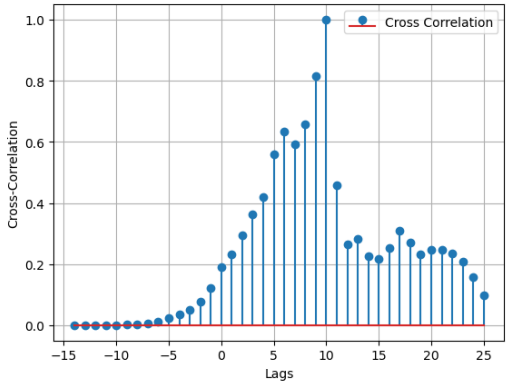

Figure 9: Cross-correlated function of Broadway and movie sales as a function of lagged time.

5 Discusion

5.1 Interpretation of lagged correlation

In reference to the filtered Broadway and film data (see Figure 6 and 7) , there is an upward increase in Broadway sales right after the release of the movie. On the other hand, there is a logarithmic decrease on film sales. This phenomena is normal because as time progresses, high number of consumers have watched the film right on the release date. One can suggest that this phenomena is due to the high-grossing weekend of the film during that time.

To examine further correlation, we can see there is a significant positive correlation on lag 10. The significance of this correlation suggest that there is a an increase of Broadway sales before 10 weeks. This event implies that before the release, Broadway sales continue to increase. On the other hand, lags -15 to 0 has no correlation effect whatsoever. This result demonstrate that after the release of the film, there is no relationship between the Broadway and movie sales. We can see this phenomena in Figure 6 as Broadway sales plateaus as time progress.

5.2 Significance of Lag 10

Though Broadway sales does not correlate after the release of the film, as mentioned above, we can see some correlation before the release. In reference to Figure 8, we can see the an annual sales progression of the Broadway musical. Starting from the month of September, sales continue to increase until New Years. This phenomena may be caused by media attention and huge marketing campaign around the film. Wicked casts (*Cynthis Erivo, Ariana Grande, Jonathan Bailey, etc.*) continue to promote their film around the world using press and public tours, brand partnerships, and immersive experiences[4]. Seeing this campaign, consumers may decide to watch the show beforehand, thus increasing Broadway sales.

However, no data nor any quantity was gathered to solidify this argument. This hypothesis might only one of the explanation for the increase in sales.

5.3 Data Errors

Conducting this data analysis involved inevitable experimental errors. One of this errors is the collected data. The data collected to compare both time-series is short. . Even though the original Broadway musical is composed of extensive data points, we need to adjust the time-frame to match the comparison of the film and the show. This dataset solely depends on the longevity of the film in the big screens, Even though Wicked ran for almost 5 months, making it one of the longest film in theaters, this data set is not enough to have a full comparison and analysis on the time-series. Thus, leading us to drastically reduce dataset to approximately 30.

Another error could be the lack of other resources. It is seen that there is an increase on Broadway sales around the time the file was released. Film gross is solely not the main contributor to this growth, Social media trends and streaming platforms are some of the quantitative dataset that could possibly use to universalize the analysis on the growth of the Broadway sales.

6 Conclusions

The result and analysis in this experiment rejects the initial hypothesis on the correlation of *Wicked* Broadway sales with film sales. The results denies the delayed increase on Broadway sales. However, various conclusions was detected in the course of the experiment. First, low-pass filters successfully eliminated the all unwanted high-frequency rendering the dataset smooth and palatable. Second. an increase on lag 10 shows other possible reasons like extensive marketing strategies for the increase on Broadway sales around the release of the film.

Other ways to improve this experiment could include more quantitative data. Data from social media trends like *TikTok, Instagram, X, etc.* could see how people respond to the film and the original Broadway show. We should also include streaming data services to asses how many consumers at home brought subscriptions or pay-per-view once the film was released. Another way to improve is to generalize the experimental approach. Instead of focusing on one show, we should include every musical that was adapted into a movie (e.g. *Dear Evan Hansen, Phantom of the Opera, West Side Story, etc.*). This approach could see the overall trends on how a movie affects the original Broadway musical. It would enable us to detect a general correlation once a movie was released.

References

[1] Armstrong, P. (2020). How the +50 million 'cats' flop is a lesson to all startups in 2020. https://www.forbes.com/sites/paularmstrongtech/2020/01/02/how-the-50-million-cats-flop-is-a-lesson-to-all-startups-in-2020/.

[2] BoxOfficeMojo (2025). Wicked box office mojo. https://www.boxofficemojo.com/ release/rl1199474177/.

[3] BroadwayWorld (2025). Wicked gross. https://www.broadwayworld.com/grosses/wicked.

[4] Hardcastle, K. (2024). Defying the rulebook: How wicked is soaring beyond traditional marketing. https://www.forbes.com/sites/katehardcastle/2024/11/15/defying-the-rulebook-how-wicked-is-soaring-beyond-traditional-marketing/.

[5] Kain, E. (2024). 'wicked' review: A spectacular cinematic triumph. https://www.forbes.com/sites/erikkain/2024/11/25/wicked-review-a-spectacular-cinematic-triumph/.