# MKTG/STAT 4760/7760 Project 2:

Behind the Curtain: What Drives Adoption in Wicked's Ticket Sales?

April 9, 2025

**Abstract**

This paper models daily ticket sales for Wicked using continuous-time probability models to understand the timing and drivers of theatrical adoption. We compare Pareto II and Burr XII fits, ultimately selecting a Burr XII model with Weekend and Holiday covariates for its balance of accuracy, interpretability, and parsimony. This model achieves a cumulative in-sample MdAPE of 1.25% and out-of-sample MdAPE of 3.35%, capturing key behavioral patterns such as elevated weekend attendance and a second adoption wave during the holidays. These findings offer actionable insights for studios and theater operators, highlighting how timing and calendar effects can be leveraged to optimize theatrical performance.

## 1 Introduction

When *Wicked* opened in U.S. theaters on November 22, 2024, it entered a crowded holiday box office season with high expectations and an eager fanbase. This paper analyzes daily unit-level ticket sales for *Wicked* using continuous-time probability models to understand the timing and structure of adoption. Beyond fitting a curve, we aim to identify patterns that can inform decision-making: how weekends, holidays, and competitive releases affect purchase behavior, and how studios and theaters can better plan around these rhythms. Through this lens, we explore what really drove *Wicked's* performance and what lessons might apply the next time a big release hits the stage.

## 2 Fitting Baseline Models

### 2.1 Preliminary Hypothesis

An initial look at daily ticket sales for *Wicked* shows a sharp spike on November 22, 2024, the film's theatrical release in the United States, followed by increased sales throughout the opening weekend. After that, sales decline steadily, forming a classic adoption curve with rapid initial uptake and gradual tapering, likely driven by early anticipation, media attention, and fan enthusiasm.

Since ticket purchases can occur on any day, a continuous-time model is appropriate for these data. The observed pattern suggests that:

- The hazard of adoption is highest immediately after release, potentially due to marketing campaigns, celebrity promotions, and general "buzz" about the movie.
- The hazard declines with time, reflecting a negative duration dependence. That is, individuals who don't see *Wicked* early on become less likely to do so later.

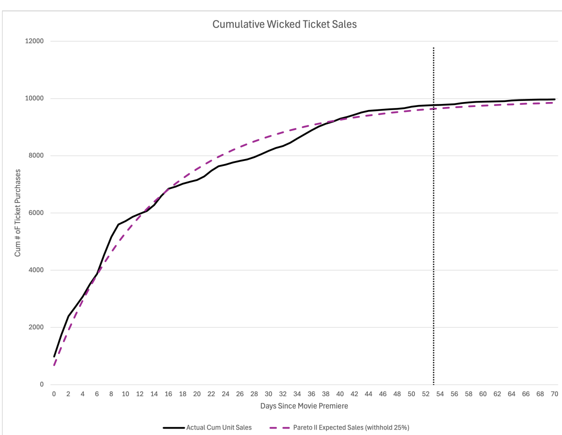
This decreasing hazard aligns with the Pareto II model, which captures this behavior while allowing for heterogeneity across individuals. We begin by fitting a baseline Pareto II model and then assess whether a more flexible alternative, such as the Burr XII model, offers a better fit.

## 2.2 Pareto II

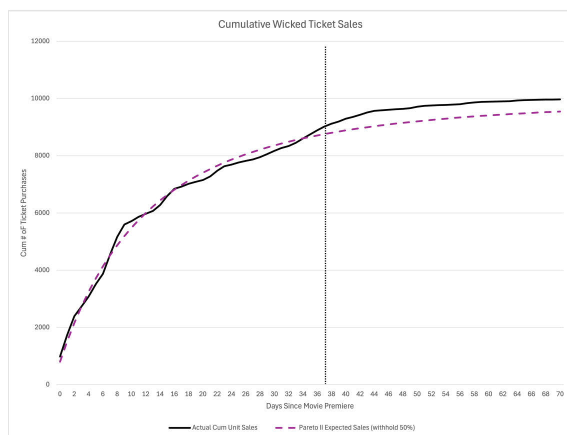
We estimate Pareto II model parameters using maximum likelihood and evaluate predictive accuracy using Median Absolute Percentage Error (MdAPE), both in-sample and out-of-sample. MdAPE is chosen over MAPE due to its robustness to outliers, which is relevant given the high initial sales spike and subsequent tapering. While cumulative MdAPE evaluates how well the model captures the overall shape of the adoption curve, incremental MdAPE focuses on day-to-day accuracy and allows us to evaluate how well the model predicts short-term fluctuations in sales. To assess generalizability and robustness, we evaluate the model under two train-test splits: withholding 25% and 50% of the data for testing.

| **Metric** | **Model 1: Nov 22 – Jan 13 (75% train, 25% test)** | **Model 2: Nov 22 – Dec 27 (50% train, 50% test)** |
|---|---|---|
| **r** | 12.02 | 2.55 |
| $\boldsymbol{\alpha}$ | 169.61 | 30.10 |
| $\boldsymbol{\lambda}$ | 0.0709 | 0.0848 |
| **Max LL** | -36,323.79 | -32,760.13 |
| **Days to Adoption (1/$\lambda$)** | 14.11 | 11.79 |
| **Cumul. In-Sample MdAPE** | 3.04% | 2.55% |
| **Cumul. Out-of-Sample MdAPE** | 1.30% | 4.79% |
| **Incr. In-Sample MdAPE** | 42.38% | 42.30% |
| **Incr. Out-of-Sample MdAPE** | 54.53% | 59.33% |

Table 1: Comparison of Pareto II Model Parameters




(a) Model 1




(b) Model 2

Figure 1

*Evaluating Pareto II model performance*

Comparing the two fitted Pareto II models, we observe notable differences in their parameter estimates and performance. Model 1, trained on a larger portion of the data (75%), produces a higher shape parameter ($r = 12.02$) and scale ($\alpha = 169.61$), resulting in a smoother, more gradual decline in the hazard rate. Model 2, trained on the first 50% of the data, estimates lower values for both ($r = 2.55$,

$\alpha = 30.10$), yielding a steeper decay and a slightly shorter expected time to adoption (11.79 vs. 14.11 days).

While both models perform well overall, Model 2 achieves a lower cumulative in-sample MdAPE (2.55% vs. 3.04%), indicating a tighter fit to the training data. However, its cumulative out-of-sample error is notably higher (4.79% vs. 1.30%), suggesting reduced generalizability. When evaluating incremental MdAPE—which captures the model's accuracy in predicting day-to-day changes in ticket sales rather than total cumulative adoption—both models perform similarly in-sample (42%), but exhibit weaker performance out-of-sample (54.53% for Model 1, 59.33% for Model 2). This highlights the challenge of modeling short-term fluctuations with a smooth timing curve alone.

Taken together, Model 1 better captures the long-run trend and generalizes more cleanly, while Model 2 provides a closer fit to the early-stage adoption dynamics.

*Are there any Wicked haters?*

As an additional check, we also tested an Exponential-Gamma model with a spike at zero using a 50/50 test-train split, allowing for the possibility of hardcore never-triers, i.e. individuals who would never watch *Wicked* under any circumstances. We can see from Table 2 that the spike at zero was rejected by the model. The estimated value of p was nearly 1, suggesting that the model assigned no meaningful weight to the never-trier segment. The log-likelihood and parameter estimates were virtually identical to the baseline Pareto II model 2. This result is also consistent with the nature of the dataset given that the adoption behavior is drawn from a panel of self-described active moviegoers, a population inherently more inclined to engage with major theatrical releases like *Wicked*. As such, it is reasonable to assume that nearly all individuals in the sample had at least some nonzero propensity to adopt.

| **Parameter** | **Value** |
|---|---|
| $r$ | 2.55 |
| $\alpha$ | 30.09 |
| $p$ | 0.9999 |
| Max LL | -32,760.13 |

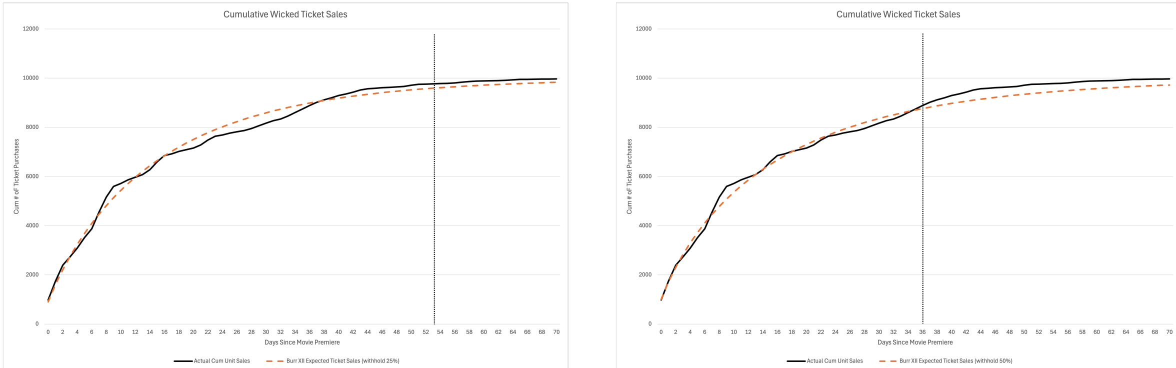
Table 2: Estimated Parameters for the EG "Never Triers" model

## 2.3 Burr XII

To assess whether a more flexible hazard structure improves model fit over the Pareto II, we also estimated a Burr XII model using the same 75/25 and 50/50 train-test splits as above.

| **Metric** | **Model 3: Nov 22 – Jan 13 (75% train, 25% test)** | **Model 4: Nov 22 – Dec 27 (50% train, 50% test)** |
|---|---|---|
| **r** | 9,712.30 | 516.87 |
| $\boldsymbol{\alpha}$ | 102,989.59 | 4,882.43 |
| $\boldsymbol{c}$ | 0.8835 | 0.8267 |
| $\boldsymbol{\lambda}$ | 0.0943 | 0.1059 |
| **Max LL** | -36,227.07 | -32,713.32 |
| **Days to Adoption ($1/\lambda$)** | 10.60 | 9.45 |
| **Cumul. In-Sample MdAPE** | 2.55% | 2.04% |
| **Cumul. Out-of-Sample MdAPE** | 1.62% | 3.23% |
| **Incr. In-Sample MdAPE** | 43.91% | 39.14% |
| **Incr. Out-of-Sample MdAPE** | 74.89% | 56.18% |

Table 3: Comparison of Burr XII Model Parameters




(a) Model 3

(b) Model 4

Figure 2

Overall, the Burr XII offers consistently better likelihood scores and improved cumulative in-sample fit, with especially meaningful gains in the 50/50 split. In both fitted models, we find that c is significantly less than 1, (0.88 in the 75/25 split and 0.82 in the 50/50 split). This provides further support for our earlier observation that individuals become less likely to purchase tickets as time passes. Additionally, the expected time to adoption under the Burr XII (9.45 to 10.60 days) is slightly shorter than under the Pareto II, reflecting the steeper early hazard implied by the more flexible model.

*How does the Burr XII performance compare with the Pareto II?*

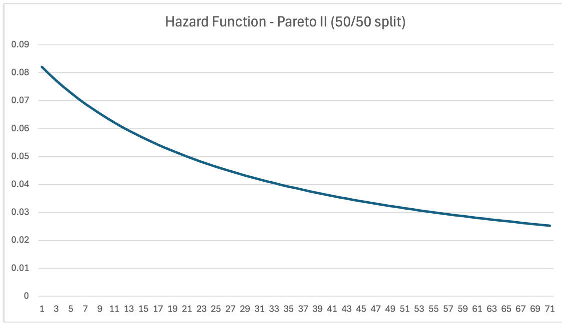
| **Model Comparison** | **LL (Pareto)** | **LL (Burr)** | **df** | **LRT Statistic** | **p-value** |
|---|---|---|---|---|---|
| Pareto II vs. Burr XII (25% test) | -36,323.79 | -36,227.07 | 1 | 193.45 | 0.0000 |
| Pareto II vs. Burr XII (50% test) | -32,760.13 | -32,713.22 | 1 | 93.62 | 0.0000 |

Table 4: Likelihood Ratio Tests Comparing Pareto II and Burr XII Model Fits

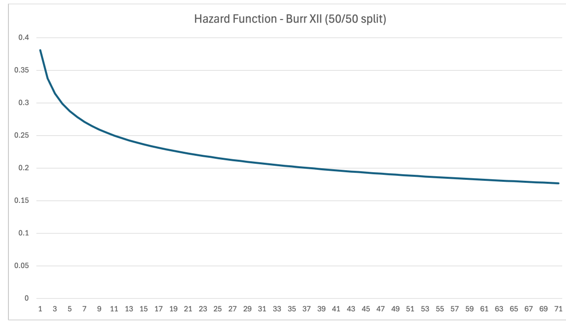
The Burr XII model 3 outperforms Pareto II model 1 in log-likelihood (–36,227.07 vs. –36,323.79) and cumulative in-sample MdAPE (2.55% vs. 3.04%). We also see small improvements for the Burr XII model 4 compared with the Pareto II model 2. The cumulative out-of-sample MdAPE for Burr XII model 4 (3.23%) is notably lower than Pareto II (4.79%), suggesting better generalization across time with a smaller training set. The likelihood ratio test also suggests that the Burr models offer a meaningful improvement over the Pareto models.

Cumulative MdAPE is especially important in this context because it reflects the model's ability to capture the overall shape and long-term trajectory of adoption, which is more relevant for forecasting total demand and informing operational decisions. In contrast, incremental MdAPE may be useful for measuring day-to-day predictive accuracy, but tends to be more sensitive to short-term volatility that smooth adoption models may not be designed to capture.

The fact that $c < 1$ in both Burr XII models reinforces our assumption of negative duration dependence, but also allows for a steeper decline in adoption than the Pareto II can capture (Figure 3). This added flexibility is especially relevant given the context: the streaming release on December 31, 2024, and the *Wicked* Oscars opening performance on March 2, 2025, both fall outside our data window but could plausibly drive secondary waves of adoption. Given this information, we will proceed with the Burr XII (Model 4) to evaluate covariates.




(a) Pareto Model 2 Hazard




(b) Burr Model 4 Hazard

Figure 3

# 3 Modeling with Covariates

## 3.1 Covariate Evaluation

To build on the baseline Burr XII model, additional covariates were introduced to enhance the model's explanatory power while maintaining narrative coherence with known patterns of movie-going behavior. The goal was to incorporate covariates that reflect real-world drivers of *Wicked* adoption, particularly those related to temporal variation in consumer availability and attention.

Although the dataset included scaled Google Trends for Ariana Grande and Broadway gross, these variables were excluded due to concerns about endogeneity. Both covariates may be more reflective of demand rather than predictive of it. For example, search interest in Ariana Grande could rise after major promotional events or spikes in ticket sales, and Broadway gross may simply scale with overall ticket volume. Instead, we evaluated covariates that align with established movie-watching patterns, including:

- **Weekend Indicator (Friday, Saturday, Sunday = 1; otherwise = 0):** Given that movie-going is likely to be highly concentrated on weekends, a binary weekend variable was constructed to reflect expected spikes in demand during these days. Following box office conventions, Friday was included as part of the weekend due to its significance in opening day traffic.
- **Holiday Indicator (November 28-29, December 24–January 1 = 1; otherwise = 0):** A binary variable was added to capture the well-documented year-end surge in movie attendance. According to a 2019 CNBC article.[1], the eight days from Christmas Eve through New Year's Eve can account for up to 5% of annual box office revenue, driven by widespread time off from school and work.
- **Smash Hit Indicator (If "smash hit" release on day = 1; otherwise = 0):** A binary "smash hit" competitor covariate was added to capture the impact of major competing film releases. This variable is set to 1 on and after the release of any film that grossed over $100 million at the domestic box office, which is a standard industry threshold for blockbuster status.

## 3.2 Fitting Burr XII with Covariates

We began by incorporating behaviorally motivated calendar effects, weekends and holidays, before adding an indicator for competitive theatrical pressure from other blockbuster releases. It was initially hypothesized that weekends and holidays would boost adoption, as people tend to have more availability and interest in attending movies during these times. Conversely, the release of smash hit

[1] https://www.cnbc.com/2019/12/23/this-is-the-most-important-week-of-the-year-for-movie-theater-owners.html

competitors was expected to detract from *Wicked's* adoption, by drawing attention and audiences toward other high-profile titles.

After fitting the model with these covariates, we find that while each addition improves statistical fit (Table 5), they also introduce greater complexity that requires careful interpretation. However, marginal gains in accuracy come at the cost of increased model complexity. Not all effects align with initial expectations, and the shifting shape of the baseline hazard suggests that the model becomes increasingly reliant on covariate structure to explain variation in adoption behavior.

| **Metric** | **Baseline Burr Model 4** | **+ Weekend** | **+ Weekend & Holiday** | **+ Weekend, Holiday & Smash Hit** |
|---|---|---|---|---|
| $r$ | 516.87 | 3.52 | 5.61 | 9014.15 |
| $\alpha$ | 4,882.43 | 70.33 | 105.57 | 169,228.41 |
| $c$ | 0.827 | 1.014 | 0.951 | 1.087 |
| $\beta_{\text{weekend}}$ | — | 0.702 | 0.704 | 0.702 |
| $\beta_{\text{holiday}}$ | — | — | 0.452 | 0.414 |
| $\beta_{\text{smash}}$ | — | — | — | 0.802 |
| Smash Hit Gamma | — | — | — | 0.900 |
| $\lambda$ | 0.106 | 0.0501 | 0.0531 | 0.0533 |
| Days to Purchase ($1/\lambda$) | 9.45 | 19.98 | 18.82 | 18.77 |
| Max Log Likelihood | -32,713.32 | -32,280.26 | -32,169.66 | -32,150.41 |
| Cumul. In-Sample MdAPE | 2.04% | 2.42% | 1.25% | 0.97% |
| Cumul. Out-of-Sample MdAPE | 3.23% | 4.00% | 3.35% | 2.64% |
| Incr. In-Sample MdAPE | 39.14% | 24.13% | 16.63% | 17.16% |
| Incr. Out-of-Sample MdAPE | 56.18% | 48.67% | 33.88% | 41.46% |

Table 5: Comparison of Burr XII Model Parameters with Increasing Covariate Complexity

| **Model Comparison** | **LL (Restricted)** | **LL (Full)** | **df** | **LRT Statistic** | **p-value** |
|---|---|---|---|---|---|
| Baseline → + Weekend | -32,713.32 | -32,280.26 | 1 | 866.12 | 0.0000 |
| + Weekend → + Weekend & Holiday | -32,280.26 | -32,169.66 | 1 | 221.20 | 0.0000 |
| + Weekend & Holiday → + Weekend, Holiday & Smash Hit | -32,169.66 | -32,150.41 | 1 | 38.50 | 0.0000 |

Table 6: Likelihood Ratio Test Comparing Nested Burr XII Models

Given that these covariates were binary in nature, exponentiating the coefficients provides a direct interpretation of their impact on the hazard of adoption:

- The **weekend effect** ($e^{0.702} = 2.02$) confirms that adoption is twice as likely on Fridays through Sundays compared to weekdays, which is aligned with our initial hypothesis.
- The **holiday effect** ($e^{0.452} = 1.57$) shows a 57% increase in hazard during public holidays (Thanksgiving and the period between Christmas and New Year's in the training data), supporting the idea of a second adoption wave driven by school and work breaks as parents are likely to take their children to go see movies.
- The **smash hit competitor** effect, however, yields a positive hazard ratio ($e^{0.802} = 2.23$), which runs counter to the initial hypothesis. Rather than detracting from *Wicked's* adoption, blockbuster releases may have boosted overall theater attendance, increasing the likelihood of movie-going in general. The estimated gamma parameter of 0.900 suggests a gradual diffusion effect where the impact of smash hits on *Wicked's* adoption builds over time rather than spiking immediately. However, without direct causal inference, this could also reflect noncompetitive audience segmentation, broader market momentum, or simply correlation without causation.

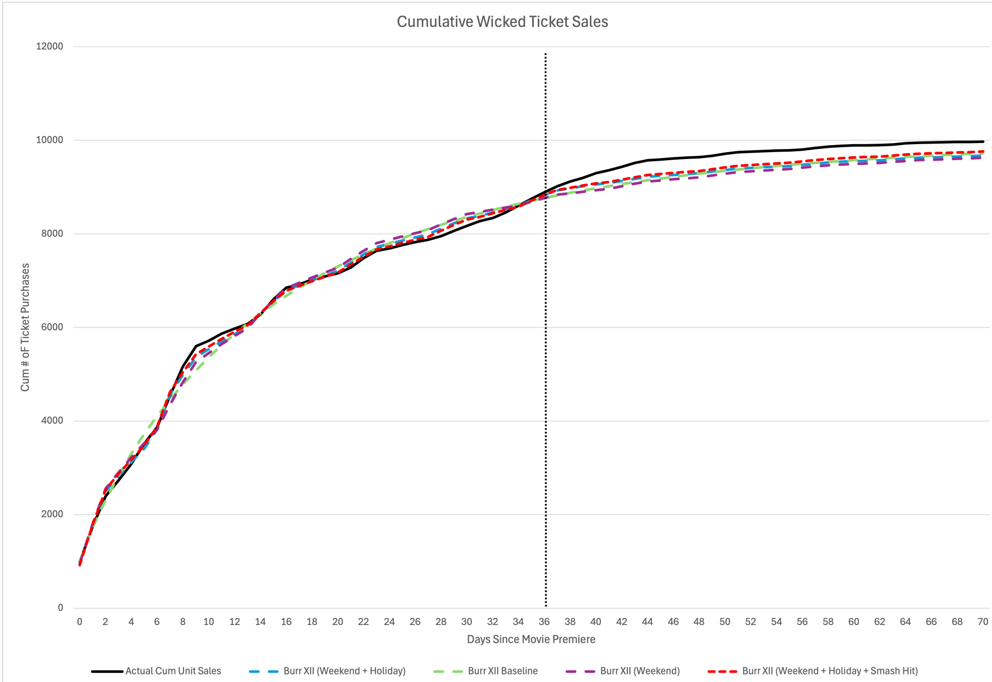



Figure 4: Observed and Predicted Adoption Curves Across Models (Cumulative)

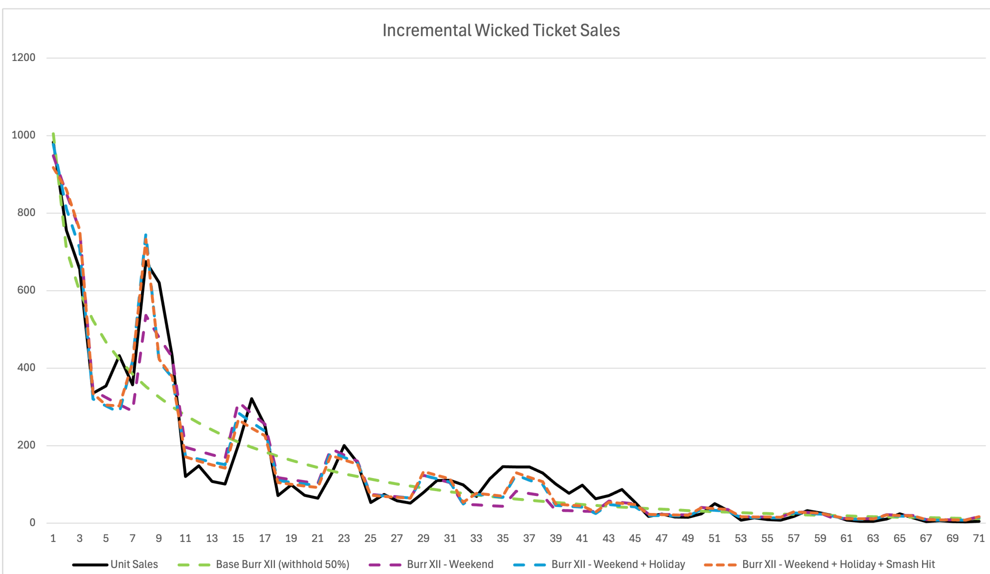
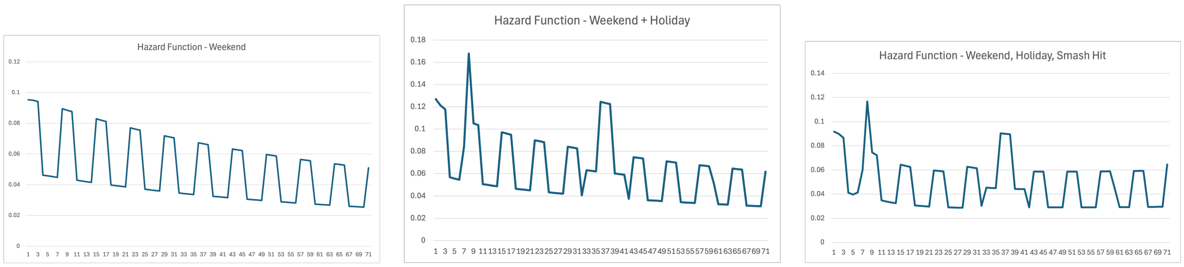



Figure 5: Observed and Predicted Adoption Curves Across Models (Incremental)

In the baseline Burr XII model, the shape parameter $c < 1$ reflects high initial adoption that fades over time. *Wicked's* fanbase and early marketing likely drove front-loaded adoption, particularly among superfans eager to attend opening weekend. When covariates like weekend, holiday, and smash hit competitors are introduced, $c$ flips to greater than 1 in some cases. This shift indicates that the baseline hazard no longer needs to explain early spikes, which are now captured directly by the covariates. The model reallocates explanatory power, resulting in a flatter or slightly increasing baseline hazard that accounts for residual variation.

The Weekend-only model shows recurring spikes every Friday through Sunday, matching expected moviegoing patterns, with overall hazard tapering over time (Figure 5). The Holiday covariate introduces a sharp rise around days 33–40, aligning with the high-traffic December 24–31 window, which captures a second wave of adoption. The full model, including Smash Hit competitors, reveals additional bumps around the releases of *Gladiator II*, *Moana 2*, and *Mufasa*, suggesting these blockbusters may have boosted overall theater attendance, indirectly benefiting *Wicked*.




(a) Hazard - Weekend Only

(b) Hazard - Weekend + Holiday

(c) Hazard - Weekend + Holiday + Smash Hit

Figure 6: Comparison of Hazard Curves Across Burr XII Models

# 4 Model Fit and Managerial Implications

In selecting a final model, we considered the story behind the model, parsimony, robustness, and both in-sample and out-of-sample fit. The Weekend + Holiday model offers the strongest balance across these factors, which we viewed as the most useful for managerial decision-making. In addition to its strong cumulative performance, the Weekend + Holiday model also reduces incremental in-sample MdAPE by over 22 percentage points compared to the baseline Burr XII model (16.63% vs. 39.14%). This suggests that it captures not just the long-run trend, but also short-term adoption spikes, which is especially useful for programming and marketing coordination. The added holiday indicator captures the well-documented end-of-year adoption boost in the movie industry without overcomplicating the model. In contrast, the baseline model relies entirely on the shape of the hazard to explain behavior, leading to a more rigid and less nuanced fit. Although the full model with all covariates achieves slightly better metrics, it also introduces additional complexity and risks overfitting to a small number of calendar events.

The Burr XII model with weekend and holiday covariates highlights how much timing matters in theatrical adoption. For studios, the findings reinforce that weekends and the holiday window are critical windows for driving attendance. Marketing efforts should lean heavily into these periods, i.e. trailer drops, cast interviews, or social media pushes tied to weekends and late December, to capitalize on when people are most available and in the mindset to go to the movies. For theater owners, with demand peaking on weekends and around the holidays, it makes sense to expand show times, staff up, and even get creative with special programming or themed promotions to get people into seats. A “Wicked Weekend” bundle with special merchandise, a holiday sing-along showing, or tie-ins with local events could help theaters make the most of these natural adoption spikes.

# 5 Limitations

While the model offers useful insights into Wicked’s adoption patterns, there are a few key limitations. The data is based on a scaled panel of 10,000 self-identified active moviegoers, which may not fully represent the broader U.S. population. Moviegoing behavior also varies by age, income, location, and lifestyle. Urban audiences, for example, may behave differently than rural ones, and families likely adopt on a different rhythm than young adults. Without demographic or behavioral segmentation, the model can’t capture these nuances. It also assumes equal access to theaters, overlooking regional differences in availability, pricing, or screen count that could influence adoption timing.

The model further assumes one-time adoption within a fixed window, but theatrical runs often have long tails. In *Wicked's* case, a second wave of adoption could emerge with the release of Part 2 on November 21, 2025, as fans revisit Part 1 in theaters or on digital platforms. Lastly, although covariates improved both cumulative and day-to-day fit, incremental MdAPE remains high, highlighting the challenge of capturing short-term volatility with smooth adoption models.

Future work could explore: How do adoption patterns differ across audience segments? How does theater access affect regional behavior? And when Part 2 hits, will *Wicked* rise again for an encore?