Time-Series Approach for Forecasting Twitch Viewership

C. Austin Brown Justin R. Sims
Department of Mathematics & Statistics, University of Tennessee at Martin

MARTIN

Introduction

- Twitch is a platform with a focus on live streaming video games and esports events.
- Accurate forecasting of viewership allows external companies to target their marketing toward games with high viewership.
- Mallari et al discuss the history and need for approachable analytics for video game streamers.
- **Motivation**: For a particular game, what is the optimal seasonal time series model for forecasting across a six month horizon?

Twitch Viewership Data

- Our data is obtained from the Twitch analytics site, SullyGnome. This site uses the Twitch API to collect data from Twitch creators with three or more viewers every 15 minutes. Data collection began August 2015 and continues to the present.
- Although the data is collected creator-to-creator, SullyGnome also combines the data for those creators currently streaming each game. From the list of most watched games on Twitch, we selected 15 games with at least three years worth of observations.
- For every 15 minutes a stream is active, a viewership count is recorded. These data are then aggregated across varying time scales (daily, monthly, etc.). We obtained **average monthly viewership counts** from the SullyGnome database.

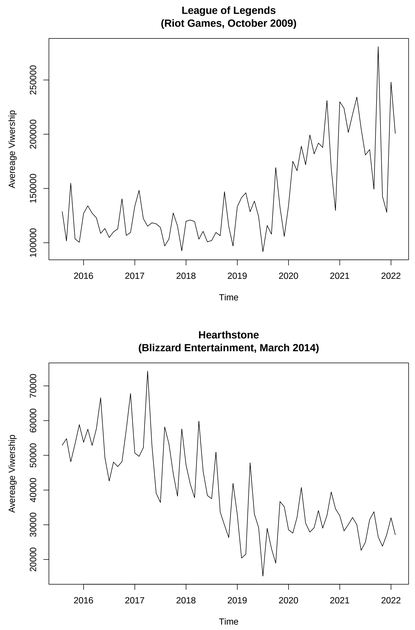

Figure: Time plots containing the average monthly Twitch viewership (August 2015 to February 2022) for League of Legends (top) and Hearthstone (bottom).

Seasonal ARIMA Model

- Define $\{X_t\}$ as a zero-mean univariate time series observed at T time points. An autoregressive integrated moving average model of orders p, d, q, P, D, and Q, ARIMA$(p,d,q)(P,D,Q)_S$, is given by

$$f_p(B)\, F_{S,P}(B)\,(1-B)^d \left(1-B^S\right)^D X_t = g_q(B)\, G_{S,Q}(B)\, \varepsilon_t,$$

for $t = 1, 2, \ldots, T$, where

  - B^k is the backshift operator which operates on a time series and has the effect of shifting the series back k time points;
  - $f_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the non-seasonal order-p autoregressive characteristic polynomial with coefficients $\phi_1, \ldots, \phi_p$;
  - $g_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ is the non-seasonal order-q moving average characteristic polynomial with coefficients $\theta_1, \ldots, \theta_q$;
  - S is the seasonal period defining the number of observations that make up a seasonal cycle;
  - $F_{S,P}(B) = 1 - \Phi_1 B^S - \cdots - \Phi_P B^{SP}$ is the seasonal order-P autoregressive characteristic polynomial with coefficients $\Phi_1, \ldots, \Phi_P$;
  - $G_{S,Q}(B) = 1 + \Theta_1 B^S + \cdots + \Theta_Q B^{SQ}$ is the seasonal order-Q moving average characteristic polynomial with coefficients $\Theta_1, \ldots, \Theta_Q$;
  - d is the non-seasonal differencing order and D is the seasonal differencing orders; and
  - $\{\varepsilon_t\}$ is zero-mean white noise with variance $\sigma^2 < \infty$.
- For example, an ARIMA$(1,1,0)(0,0,1)_{12}$ model is for monthly data and may be written as

$$X_t = X_{t-1} + \phi_1(X_{t-1} - X_{t-2}) + \varepsilon_t + \Theta_1 \varepsilon_{t-12}.$$

Forecasting the Seasonal ARIMA Model

- For each selected game, hold out the last six months of observations to test the model's predictive capability.
- For the remaining series, evaluate stationarity by visual inspection of the time plot and/or the Augmented Dicky-Fuller (ADF) test.
- If the series is not stationary, determine d and D by computing non-seasonal and seasonal differences and applying the ADF test to the differenced series.
- Once stationary determine candidate orders for p, q, P and Q by examining the autocorrelation functions.
- Fit the seasonal ARIMA model for each combination of candidate orders and check each for overfitting.
- For each fitted model obtain a six-month-ahead forecast. Compare the forecasted values to the held-out values by calculating the root mean square prediction error (RMSPE).

Results

Table: Optimal seasonal ARIMA model for various games with RMSPE of six-month-ahead forecast of average monthly viewers.

Game	Model	RMSPE
Fortnite	ARIMA$(0,1,1)$	19725.6
Hearthstone	ARMIA$(3,1,0)(1,0,0)_{12}$	3535.8
League of Legends	ARIMA$(0,1,1)(0,1,0)_{12}$	30703.6
Minecraft	ARIMA$(0,1,0)(1,0,0)_{12}$	9932.9
World of Warcraft	ARIMA$(1,0,1)(0,1,1)_{12}$	22563.4

League of Legends

$$X_t = X_{t-1} + (X_{t-12} - X_{t-13}) + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

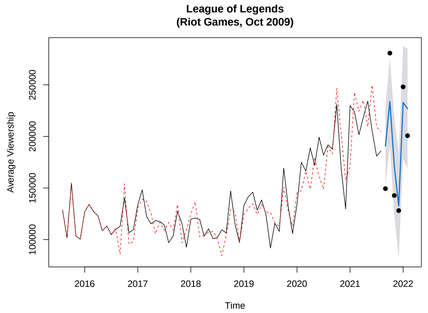

Figure: Time plot containing the modeled average monthly Twitch viewership in red for League of Legends with predicted average monthly viewership in blue.

Conclusion and Future Work

- Although there is not one set of optimal orders, we do note similarities in the optimal forecasting times scale across games.
- Furthermore, the interpretability of the seasonal ARIMA model allows us to explain our forecast more easily by using recent and long-term viewership counts.
- We plan to pursue an investigation of more creator-centric models (both parametric and semiparametric) and the use of multivariate models across similar genres.

Selected References

- Lessel, P., Mauderer, M., Wolff, C., and Krüger, A. (2017). "Let's Play My Way: Investigating Audience Influence in User-Generated Gaming Live-Streams," *2017 ACM Conference on Interactive Experiences for TV and Online Video*, June 2017, pp. 51-63.
- Mallari, K., Williams, S., Hsieh, G. (2021) "Understanding Analytics Needs of Video Game Streamers," *2021 CHI Conference on Human Factors in Computing Systems*, May 2021, pp. 1-12.

These results were part of research funded by the College of Engineering & Natural Sciences at UT Martin.

E-Mail: chaabrow@ut.utm.edu