PREZI

Stefania Colombo, Silvia D'Amicantonio and Federico Lancellotti
present

ON BOARD GAMES AND HOW TO BUILD THEM

A study on the next best board game

Nonparametric Statistics A.Y. 2023-2024 16th February 2024

INTRODUCTION

Emerging designers are developing new games every day, crowdfunding their projects on platforms such as **Kickstarter**.

Our goal is to characterize the **best board games** in terms of commercial success, to **optimize the efforts in the development and the investments**.

MOTIVATION

Business

This article is more than **1 month old**

Toymaker Hasbro lays off 1,100 staff as holiday season fails to lift sales

Manufacturer of Monopoly, Play-Doh and Transformer cuts workforce by almost one-fifth as 'last resort'

Callum Jones *in New York*

Mon 11 Dec 2023 23.41 CET

Most viewed


King Charles diagnosed with cancer, Buckingham Palace announces


Live King Charles: Prince Harry to travel to UK in coming days to see father after cancer diagnosis – latest news updates


Brentford 1-3 Manchester City: Premier League – as it happened

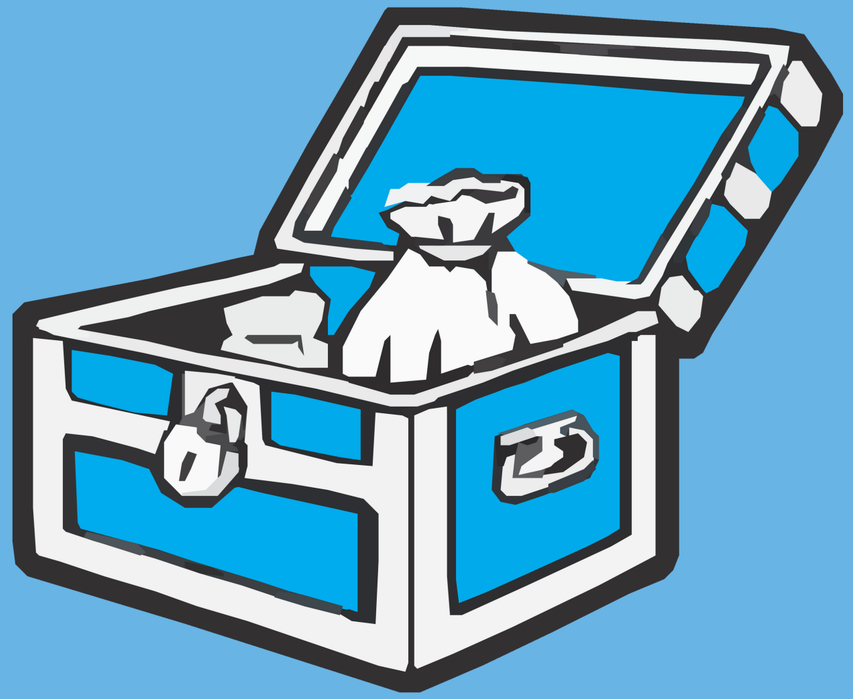
MOTIVATION

Emerging designers
and startups

Us

Established
Publishers

SNAPSHOT DATA



SNAPSHOT DATA

For approximately **21600 games**, we have **different features** such as: complexity, playtime, category, number of players, rank, average rating and other indicators.

Dataset comes from the **BoardGameGeek** database.

Data are collected for games released **between 1950 and 2023**.

Assumption: consider data as representative of the **real market**.

SNAPSHOT DATA

The column containing the categories and the publishers reports ***strings of characters***.
How do we deal with them?

Game ID	Category	Publisher
1	[Economic, Fantasy, War]	[Portal Games, Valley Games, …]
2	[Card, Fantasy]	[Wizards of the Coast, 999 Games, Hasbro, …]
3	[War]	[999 Games, Hans im Glück, Rio Grande Games]

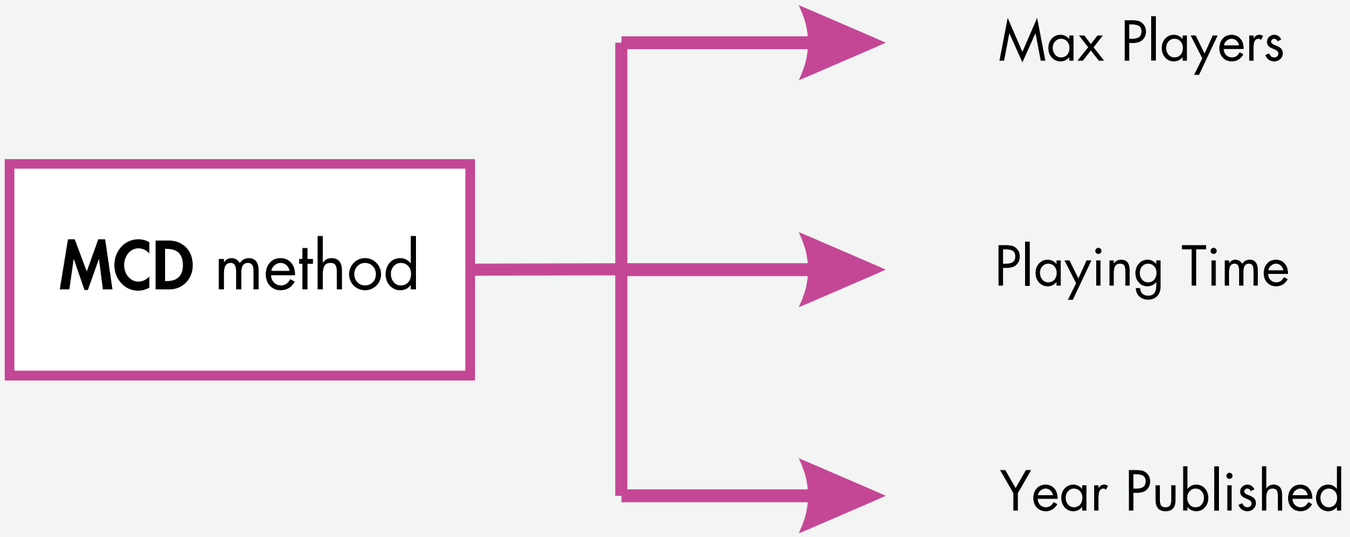
SNAPSHOT DATA

We decompose the strings ***distributing categories across columns*** and assigning ***binary values***, while for ***publishers*** we consider the ***number of published games***.
But are all ***84*** categories ***informative***?

Game ID	Card	Economic	Fantasy	War	...	Publisher
1	0	1	1	1	...	970
2	1	0	1	0	...	6144
3	0	0	0	1	...	37

SNAPSHOT DATA

We build a **ROBUST MODEL**,
removing some observations that might be **outliers**.



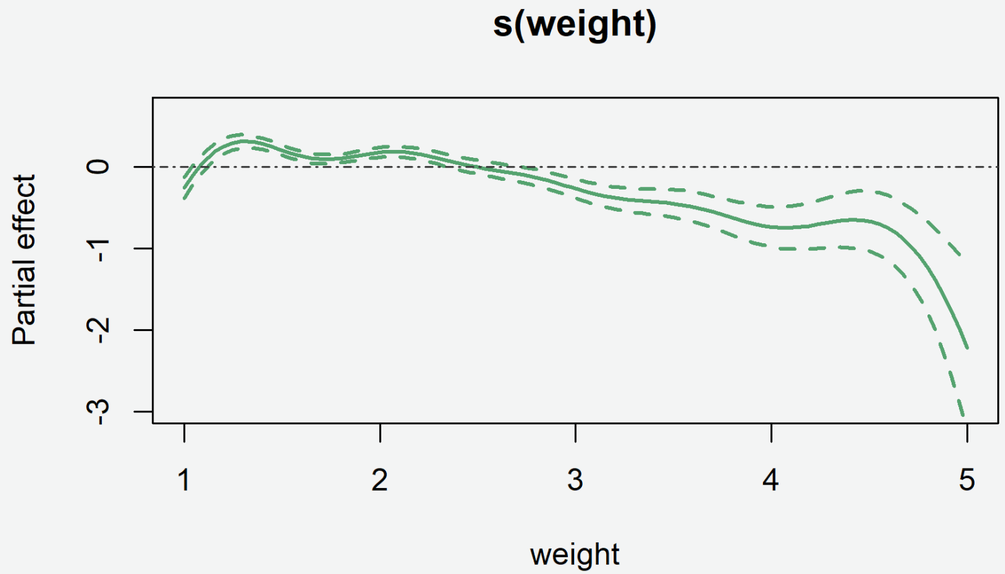
MODEL FOR SNAPSHOT DATA

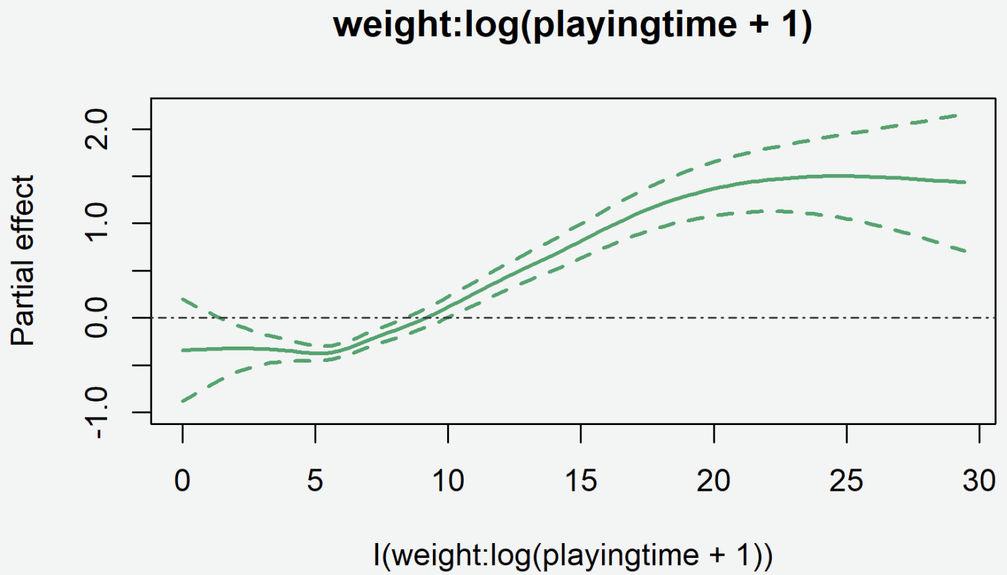
SNAPSHOT DATA

Robust model with interactions

$$
\begin{aligned}
log(users\ rated_i + 1) = {} & f(log(playing\ time_i + 1)) + f(min\ age_i) \\
& + f(year\ published_i) + f(dim\ publisher_i) \\
& + f(max\ players_i) + f(weight_i) \\
& + f(I(max\ players_i : log(playing\ time_i + 1))) \\
& + f(I(weight_i : log(playing\ time_i + 1))) \\
& + f(I(weight_i : dim\ publisher_i)) \\
& + f(I(dim\ publisher_i : log(playing\ time_i + 1))) \\
& + Economic_i + Negotiation_i + Political_i + \cdots + \epsilon_i
\end{aligned}
$$

SNAPSHOT DATA



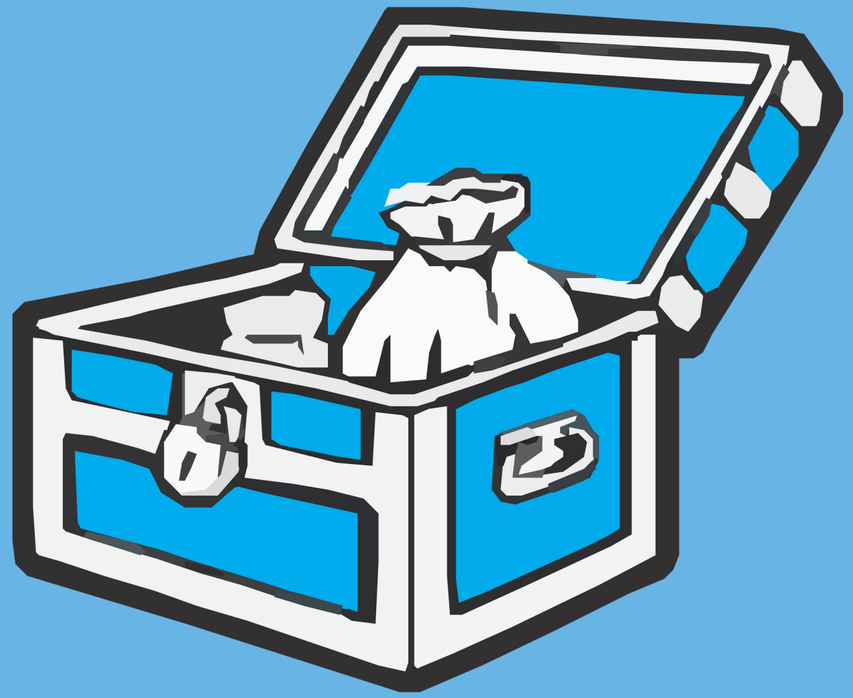


SNAPSHOT DATA

Significant Categories

+	−
• Exploration • Territory Building • City Building • Environmental • …	• Children's Game • War Game • Trivia • Book • …

HISTORICAL DATA



HISTORICAL DATA

We recover **~2500 daily snapshot** of the BoardGameGeek dataset, since late 2016, each containing the cumulative number of ratings of the games.

Fetch and **clean** the data from NAs and inconsistencies.

Monotone smoothing of the data as functions of time.

Derivatives and **downsampling** in time.

HISTORICAL DATA

We recover **~2500 daily snapshot** of the BoardGameGeek dataset, since late 2016, each containing the cumulative number of ratings of the games.

ID	Name	Published	2016-10-12	2016-10-13	...	2023-11-04
13	Catan	1995	62 048	62 048	...	122 477
1294	Clue	1949	11 629	11 629	...	20 013
1406	Monopoly	1933	18 075	18 075	...	34 655

THE BEST CATEGORIES

PROMISING GAME CATEGORIES

We want to test the statistical significance of specific effects of each category:

$$H_0 : \tau_{Economic} \equiv \tau_{CardGame} \equiv \tau_{Deduction} \equiv ... \equiv 0$$

fANOVA	ANOVA	Functional Regression Analysis
Permutational multi-way **Globalised Pointwise F-test**, on both the functions and the derivatives	Permutational multi-way ANOVA on the **Modified Hypograph**, of both the functions and the derivatives	$y_i(t) = \beta_1(t)z_i + e_i(t)$ (only explored)

PROMISING GAME CATEGORIES

A few words on the Globalised Pointwise F-test

1. Perform the classic pointwise ANOVA for each time instant:

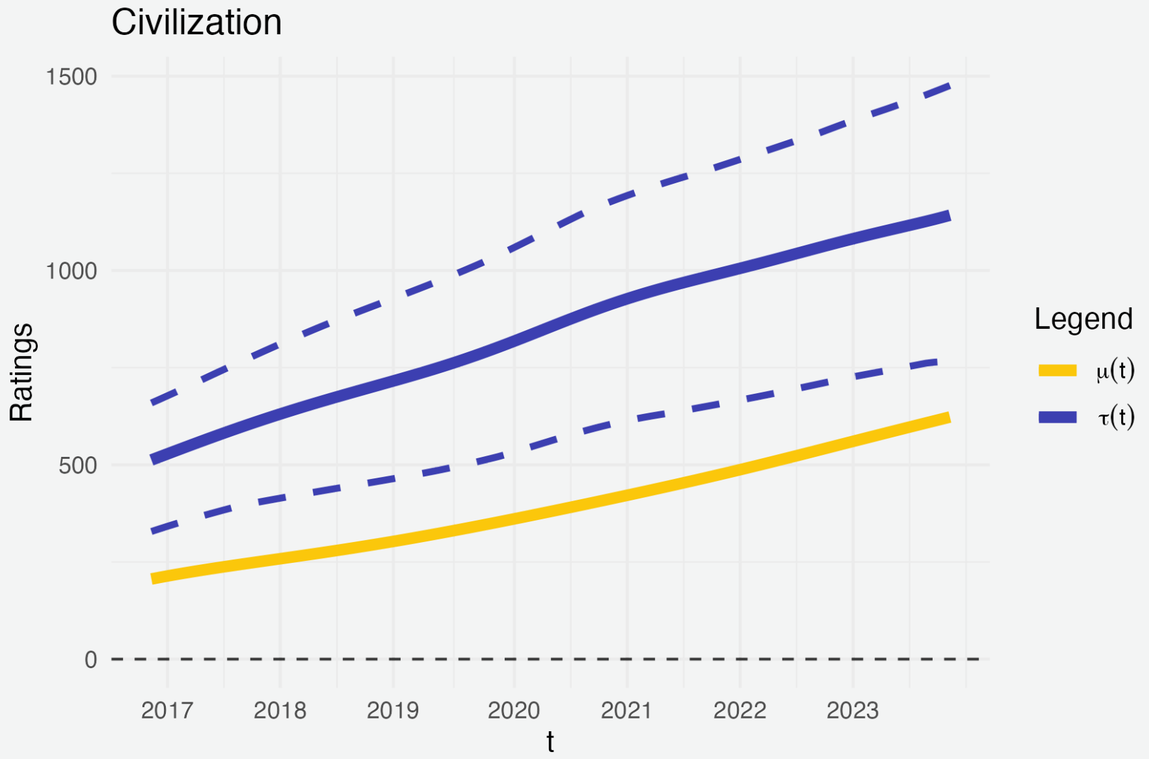
$$F_n(t) = \frac{SSR_n(t)/(k-1)}{SSE_n(t)/(n-k)}$$

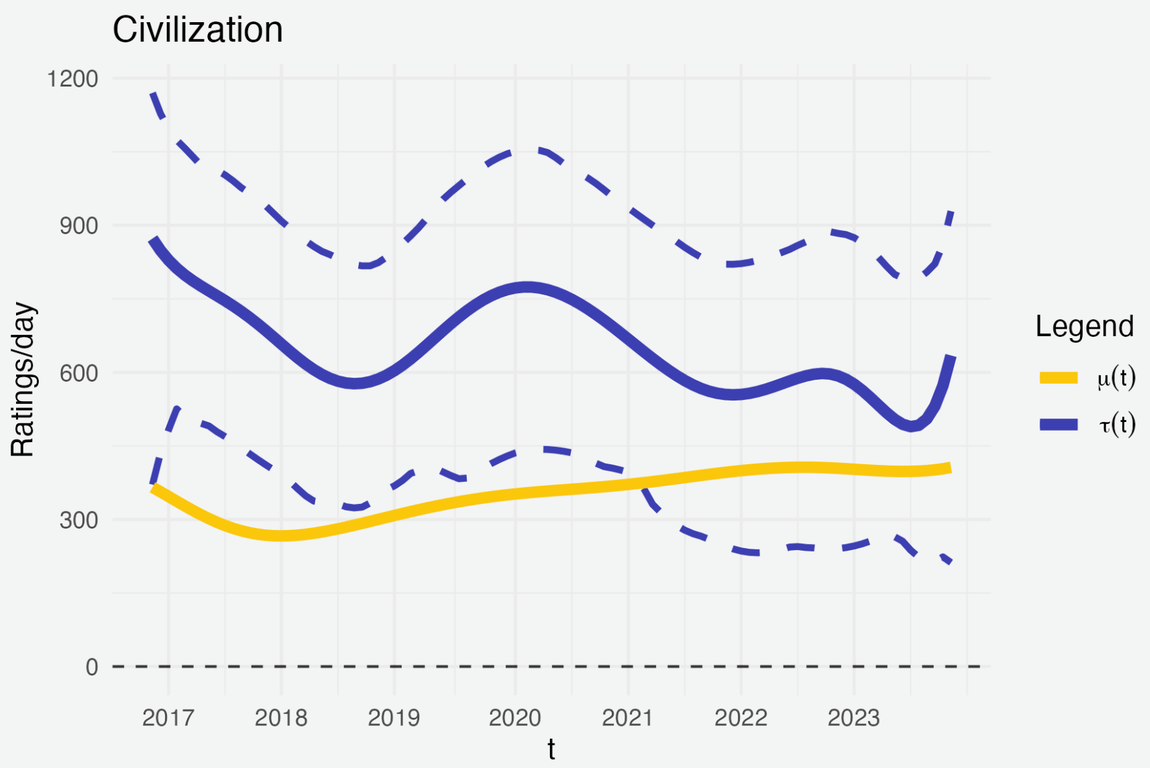
2. Globalise the test statistic by integrating in time:

$$T_n = \int_{\mathcal{T}} F_n(t)dt$$

The test is very efficient from a computational perspective and it can be proven to be root-n consistent.

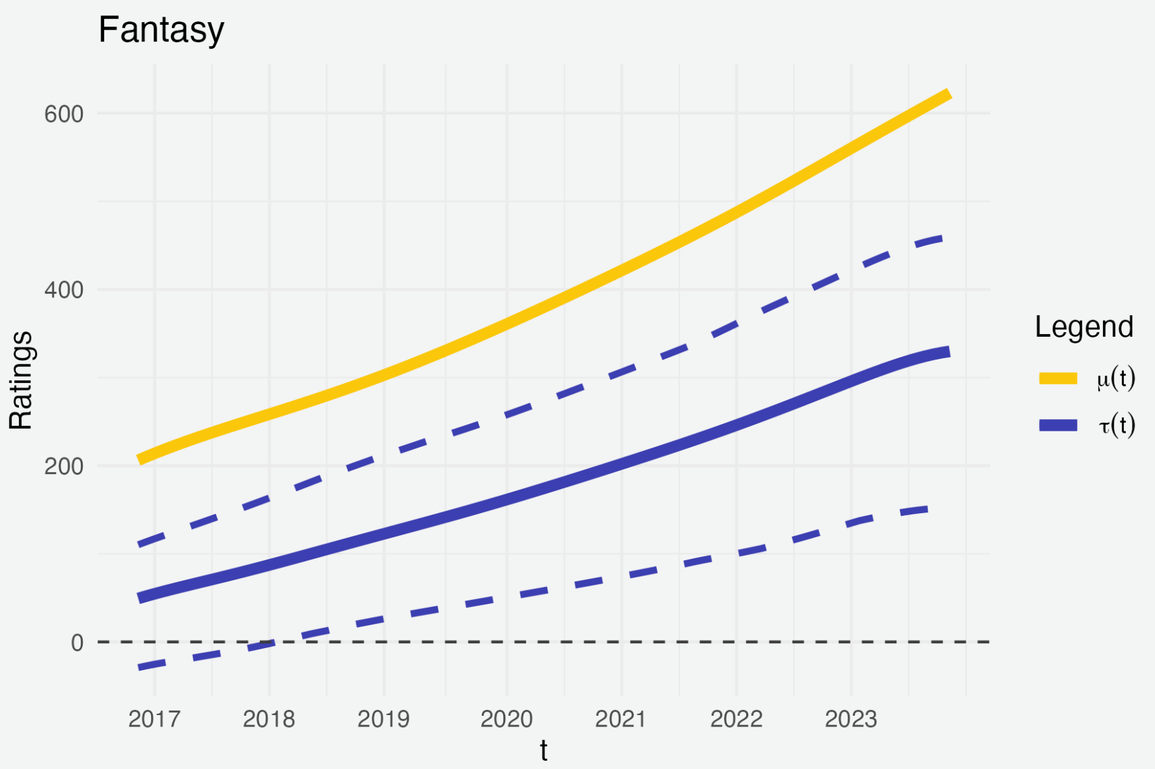
PROMISING GAME CATEGORIES

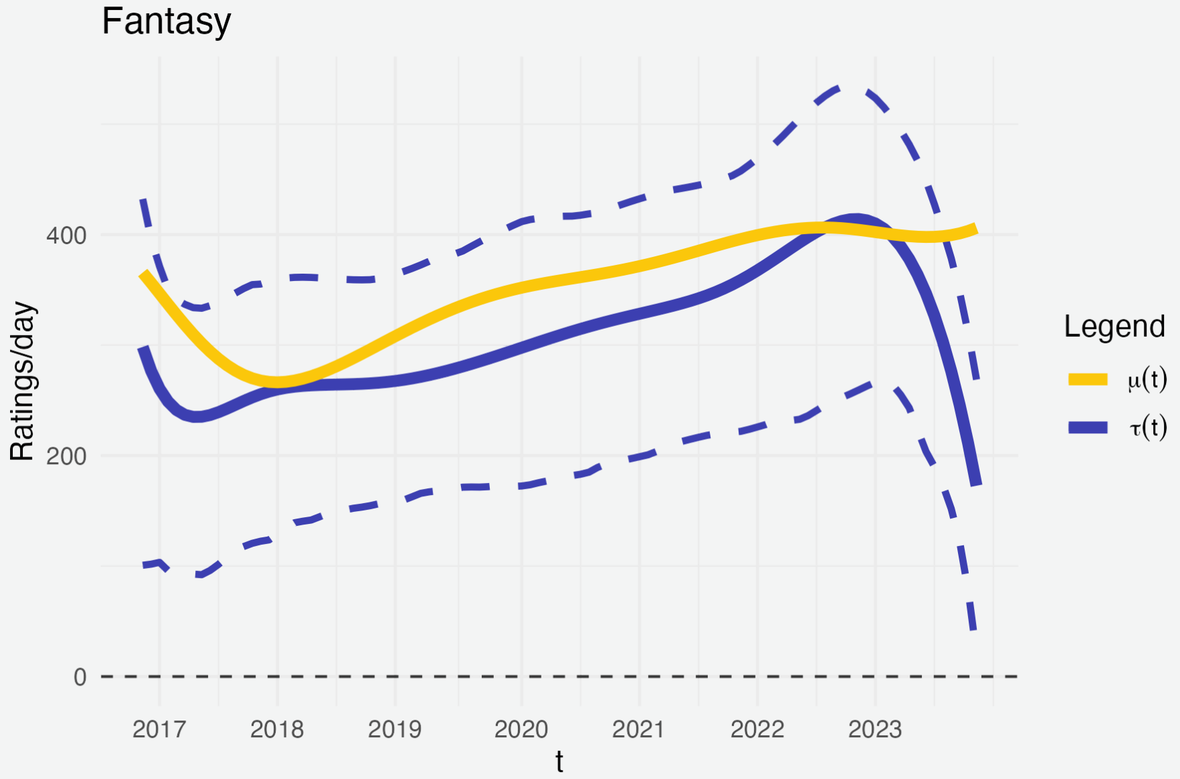




* the bands are one-at-a-time bootstrap CI at level 5%

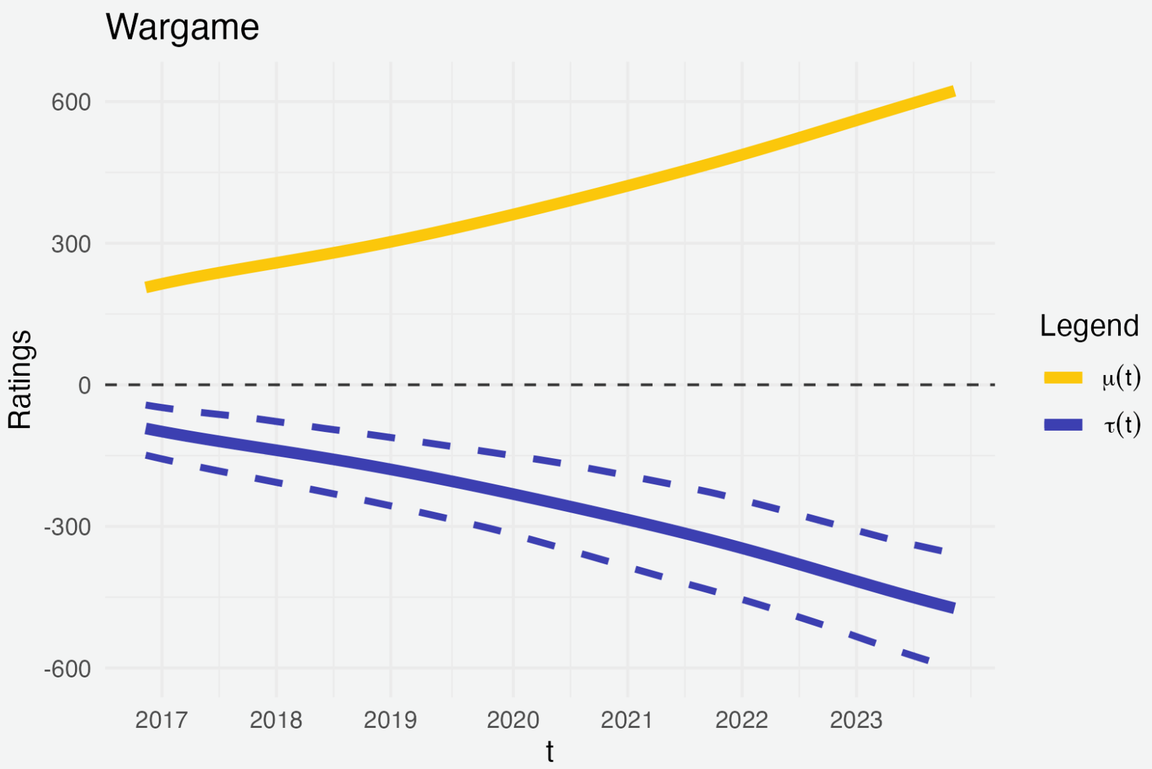
PROMISING GAME CATEGORIES

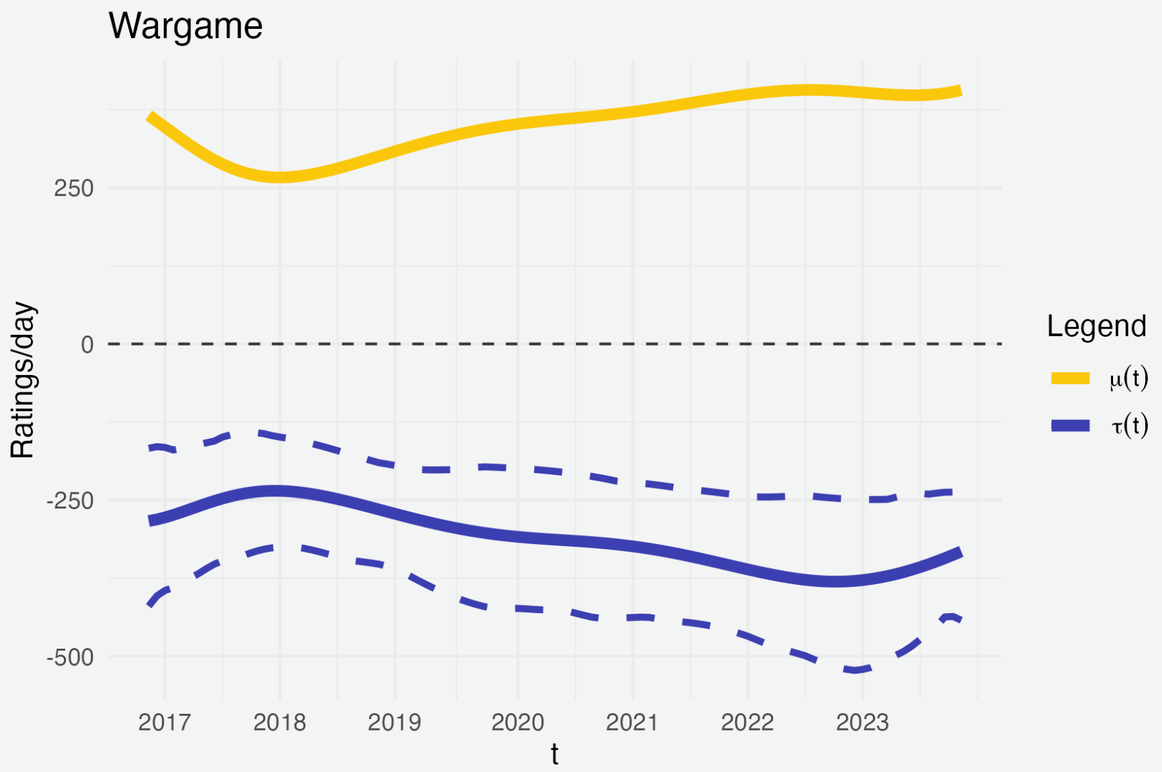




* the bands are one-at-a-time bootstrap CI at level 5%

PROMISING GAME CATEGORIES





* the bands are one-at-a-time bootstrap CI at level 5%

Longevity in the market

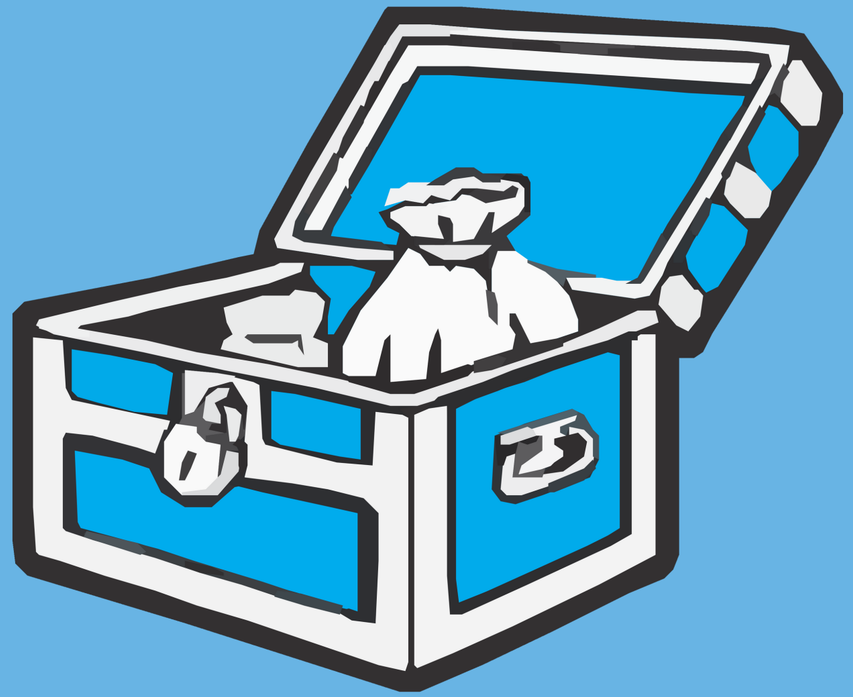
GAM + ANOVA

Absolute popularity
Growth on the market

SURVIVAL

Longevity
in the market

SURVIVAL DATA



LONGEVITY IN THE MARKET

Relying on both Snapshot and Historical datasets, we have to build the correct **Time-to-Event**

MONTH METHOD

Death after a consecutive month with constant number of ratings

QUANTILE METHOD

Death when the number of ratings is equal to the 95th percentile

LONGEVITY IN THE MARKET

Relying on both Snapshot and Historical datasets, we have to build the correct **Time-to-Event**

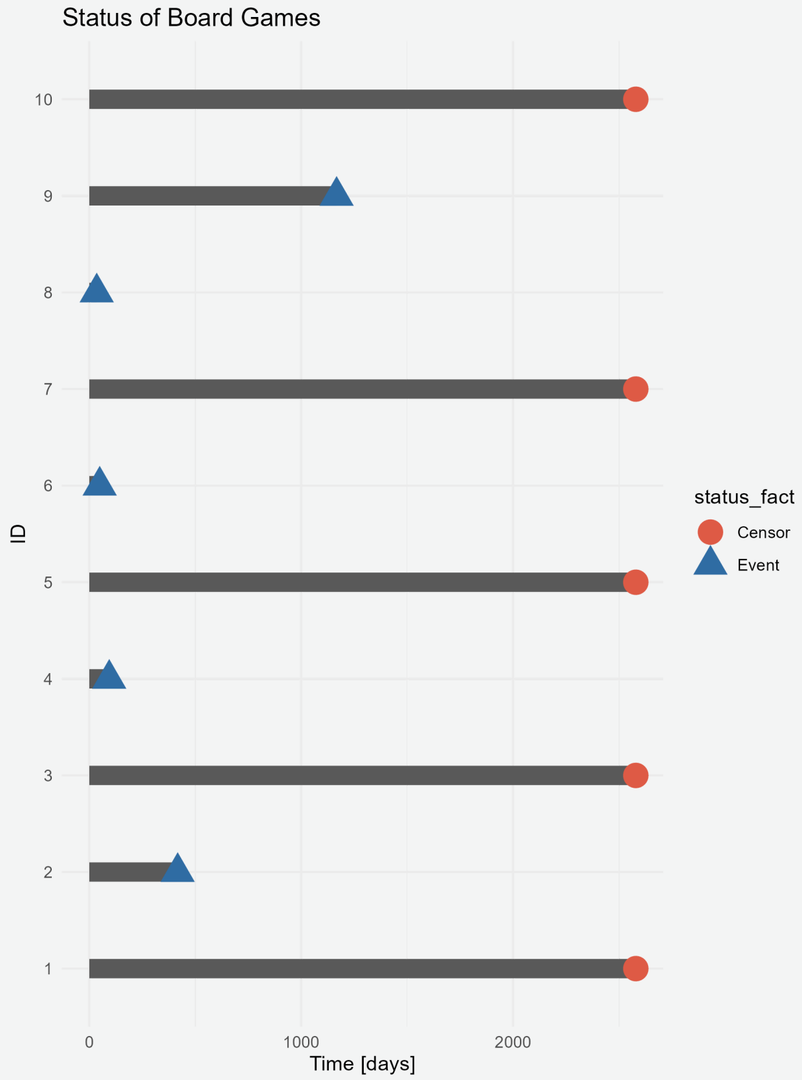
MONTH METHOD

Death after a consecutive month with constant number of ratings

QUANTILE METHOD

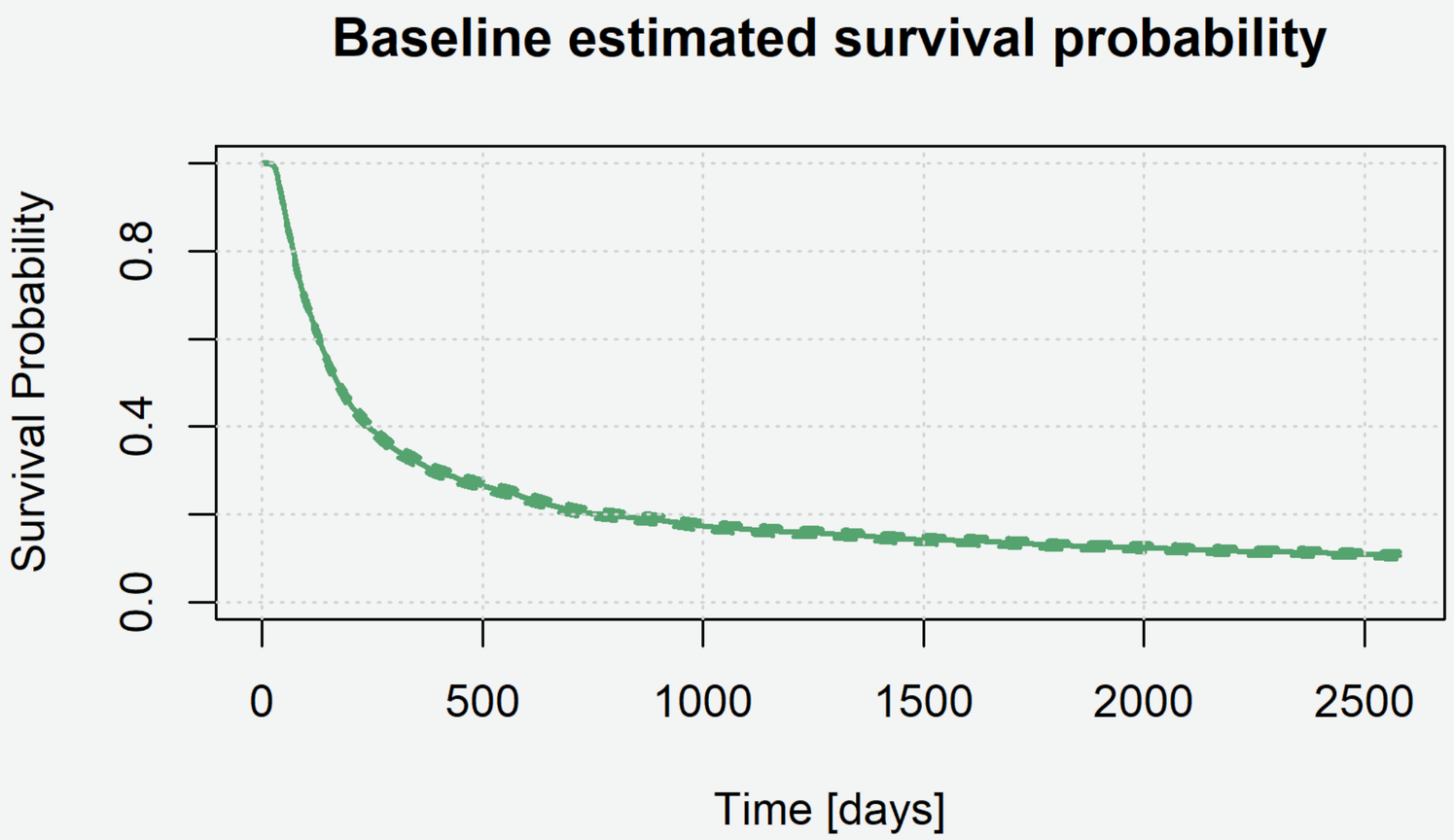
Death when the number of ratings is equal to the 95th percentile

LONGEVITY IN THE MARKET

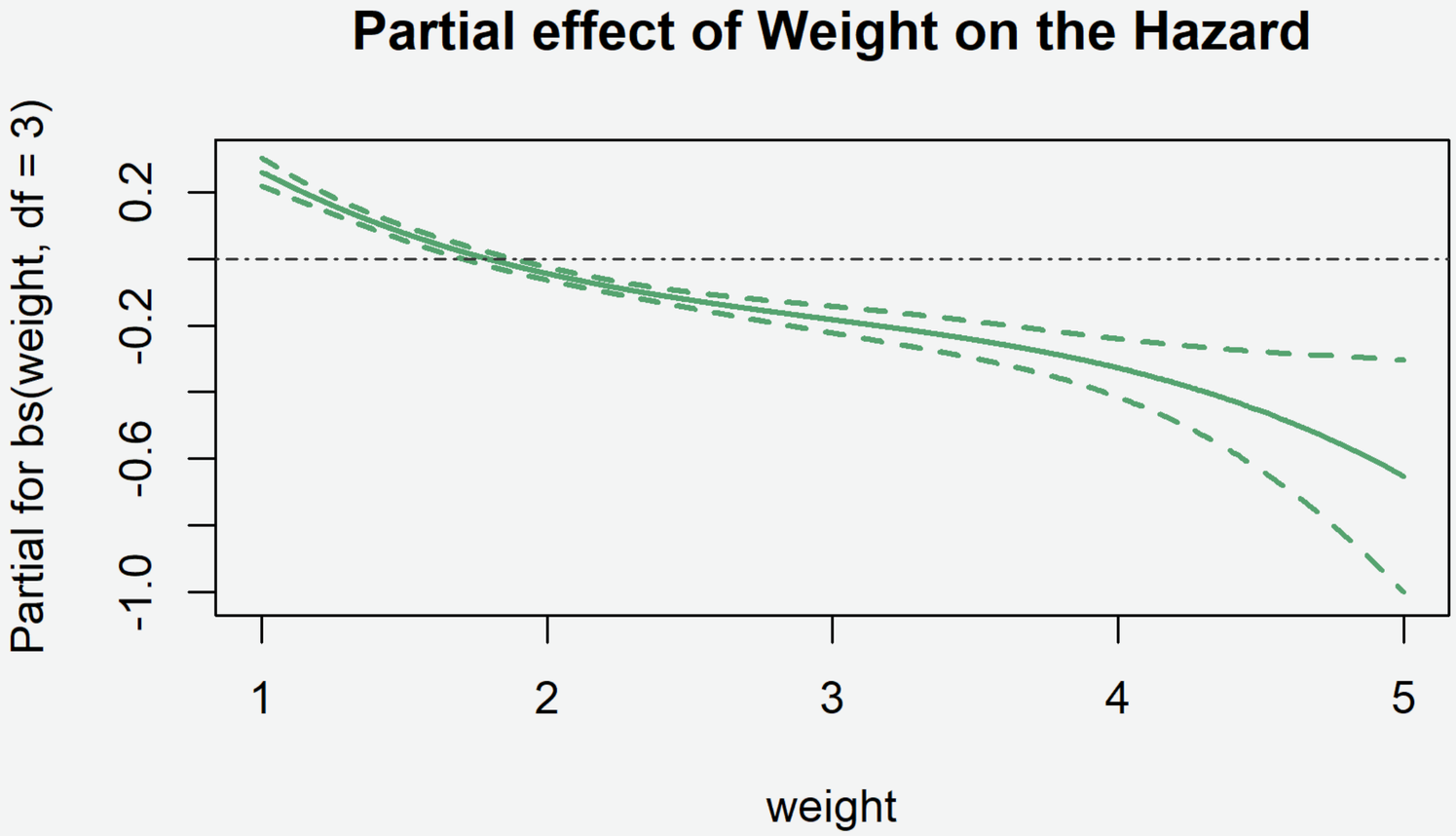


LONG-TERM SUCCESS

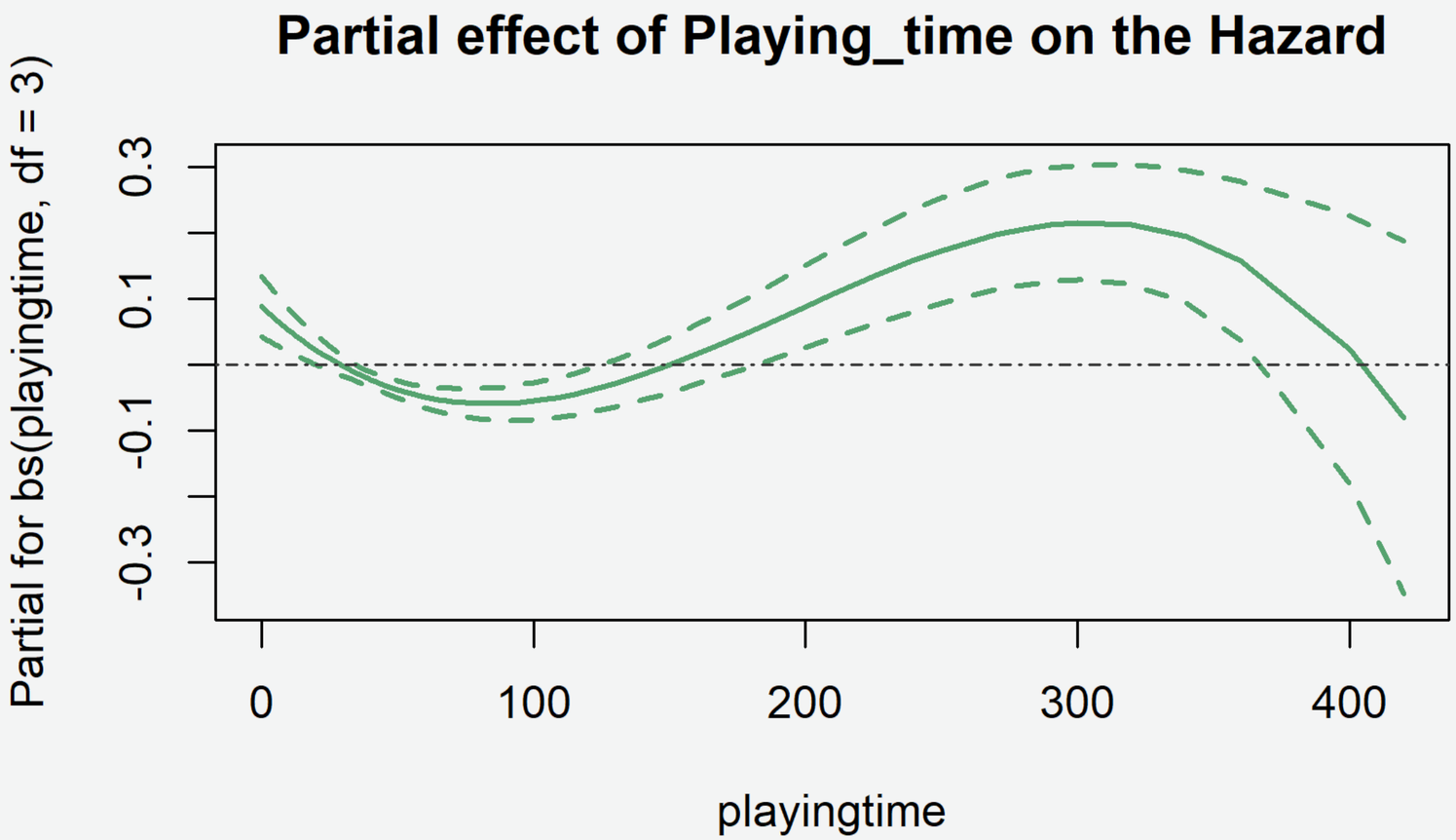
LONGEVITY IN THE MARKET



LONGEVITY IN THE MARKET



LONGEVITY IN THE MARKET



LONGEVITY IN THE MARKET

Significant Categories

+	−
• Exploration • Territory Building • City Building • Civilization • …	• Negotiation • Children’s Game • War Game • Memory • Book

CONCLUSIONS

DESIGNERS
and
PUBLISHERS

Categories: Exploration, Territory and City Building, Civilization

Features:
- Small weight & short playing time
- High weight & large playing time & large publisher dimension

PUBLISHERS

Small: improve promotion

Big: solve internal competition and saturation

FUTURE DEVELOPMENTS

ANOVA:
- Functional Regression

SURVIVAL:
- Reborn Games

PREDICTION:
- Time Series

THANK YOU

for your attention

References

- J. Peiser, “We’re in a golden age of board games. it might be here to stay,” The Washington post, 2022. https://www.washingtonpost.com/business/2022/12/24/board-game-popularity/ [Online; accessed 10-Feb-2024].
- rfordatascience, “tidyTuesday BGG.” https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-01-25/readme.md, 2022. [Online; accessed 10-Feb-2024].
- beefsack, “bgg-ranking-historicals.” https://github.com/beefsack/bgg-ranking-historicals, 2016-2024. [Online; accessed 10-Feb-2024].
- J. Bujnowicz, bggAnalytics: BoardGameGeek’s Board Game Data Analysis Tools, 2022. R package version 0.2.1.
- J.-T. Zhang and X. Liang, “One-way anova for functional data via globalizing the pointwise f-test,” Scandinavian Journal of Statistics, vol. 41, 11 2013.
- J. O. Ramsay and B. W. Silverman, Functional Data Analysis. Springer, 2010
- J. Ramsay, fda: Functional Data Analysis, 2023. R package version 6.1.4.
- D. Cook, “The circle of life: An analysis of the game product lifecycle,” Game Developer, 2007. https://www.gamedeveloper.com/design/the-circle-of-life-an-analysis-of-the-game-product-lifecycle/ [Online; accessed 10-Feb-2024].
- T. Therneau, “Spline terms in a cox model,” MIT, 2017.
- D. Long, “Cox proportional hazard regression and splines: A cautionary tale for exposure-response assessment,” Department of Biostatistics, University of Alabama at Birmingham, 2019.
- P. H. C. Eilers, B. D. Marrx, and M. Durban, “Twenty years of p-splines,” SORT (Statistics and Operations Research Transaction), 2015.
- Z. Zegras, “The rise of the professional gamer: a new age career,” Zzeger, 2023. https://zzeger.com/EN/blog/rise-of-professional-gamer/.
- T. Gorecki and L. Smaga, fdANOVA: Analysis of Variance for Univariate and Multivariate Functional Data, 2018. R package version 0.1.2.