POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

On board games and how to build them

A study on the next best board game

Nonparametric Statistics project

Master of Science in Mathematical Engineering

Academic Year: 2023-24

Students:

Stefania Colombo - stefania4.colombo@mail.polimi.it

Silvia D'Amicantonio - silvia.damicantonio@mail.polimi.it

Federico Lancellotti - federico.lancellotti@mail.polimi.it

Contents

Contents

1 | Introduction

In God we trust, all others bring data.

William Edwards Deming

1.1. Nonparametric statistics

Traditional statistical approaches typically depend on strict conditions about the data distribution characteristics, like normality, equal variances, and linear relationships. These assumptions might work in controlled laboratory settings, but they often don't accurately reflect the natural variability and diversity of real datasets. Nonparametric statistics offer a solution to this issue by providing distribution-free methods that do not rely on specific distributional assumptions. Instead of estimating population parameters directly, nonparametric techniques focus on ranking and ordering data, making them robust to deviations from parametric assumptions and resistant to outliers. Thanks to this flexibility nonparametric methods can be used to deal with a diverse range of data types, including ordinal, categorical, and skewed distributions, with minimal loss of efficiency and statistical power.

1.2. The study

The board game industry has experienced remarkable growth over the past few decades. Unlike other leisure activities that have suffered from economic crises, board games have maintained double-digit growth over the last 30 years. This phenomenon has gained popularity and reached a wide audience. Board game bars and cafes had been popping up and attendance at major games conventions is increasing. Social media is also playing a larger role in the space leading more and more game makers to consider the social media screen and length restraints when designing games.

Speaking of numbers, the global board game market has an estimated value between $11 billion and $13.4 billion and is projected to grow by about 7 to 11 percent within the next 5 years, according to market research companies Technavio and Imarc.
This increasing trend results also in more board game releases each year. The crowdfunding platform Kickstarter has made it easier than ever for unknown designers to release games. Over 3,000 new games are released each year (excluding expansion packs), according to the website and online forum BoardGameGeek [1]. However, this phenomenon has also intensified the competition among publishers, leading to a greater battle for profitable games. As a result, companies in the board game industry are demanding more from their employees and designers, and the pressure to produce successful titles has escalated. This competition has had significant repercussions, including recent layoffs at major companies in the last years [2].

In this growing market, we thought there is a little space for us. Our goal is to characterize the best board games out there in terms of commercial success, in order to optimize the efforts in the development and the investments. We want to serve as intermediaries between designers and publishers, assisting designers in making informed decisions about funding their new creative projects and advising publishers on selecting the most commercially promising games for their portfolios.

The starting point is trying to understand how to define the popularity of a game: indeed there is no popularity-meter we could use to measure it directly, hence the first task is to navigate the data we collected and find the best quantitative indicator to assess the popularity of board games. Once achieved the latter, we proceed by exploring data themselves and processing them according to the analyses and the needs we have. This part is also very challenging because of the great cardinality of the datasets involved.

Then we move to the actual analyses, the first of which is focused on understanding which game characteristics influence its popularity. Going on, we work on historical data to learn about the trends on the market during these last years. This part of the project involves two analyses: the first is again a study of the popular/unpopular categories of games, the second is a deeper study of the market life of games in the observed time window of our data. Both analyses aim at understanding the progression of the market and the evolution of buyers' tastes.

2 | Data and Methodology

2.1. Dataset description

The data we employ are derived from the website BoardGameGeek. The first dataset counts more than 21,000 board games with many related features such as their complexity, the number of players, the average time of a game, their category and so on: we'll refer to this dataset as the ***Snapshot Dataset*** since it contains a snapshot of these data collected in 2022. In particular we are using the data collected by TidyTuesday [3]. In addition, we have a second dataset about daily observations of a portion of the database since late 2016, that we will call ***Historical Dataset*** given that it gathers historical information in the form of time series. A deeper description of the data is available in the appendix.

Please note that we are aware of the fact that the BoardGameGeek database represents only a portion of the market, possibly the most "loyal". Therefore, in order to move on with our study, we need to make a strong assumption considering the data available as representative of the real market, ignoring the inevitable bias related to the "nerdish" nature of the source.

However our project is deliberately structured to propose a general approach for studies of this kind, adopting analytical instruments and techniques that offer both detailed insights into the particular dataset under examination and guarantee its adaptability with other collections of data. For instance, if we were to work with actual sales data or broader market information, it would be straightforward to carry on a similar study using the same tools and methods, thanks to the high level of scalability of our workflow.

2.2. Preprocessing data

2.2.1. Categories and publishers

A first issue to address is how to manipulate the columns related to the categories and the publishers, given their large dimensionality and their difficult format. Each game may be characterized by more than one category and more than one publisher. Moreover, for these two variables the data were collected as a single string for each observation, making even more complex to deal with them.

Therefore we introduce dummy variables, one for each of the unique categories, adding 84 columns to our datasets. For the publishers, we define their dimension as the total number of published games. For each observations, we consider the sum of the dimensions of the related publishers, in order to model what we think may be the marketing effect on that particular board game. So, in practice, for each publisher i we count the number of published games p_i and we fill a new column assigning to a game j the sum of all games published by its publishers, namely

$$\sum_{i \in \mathcal{C}_j} p_i$$

being $\mathcal{C}_j$ the set of publishers of game j. Indeed, for self-published games, the dimension is set to zero.

2.2.2. Historical data

The historical data can be found at [4] as .csv files associated with the date the daily snapshots of the BoardGameGeek database were scraped. As such, it is necessary to merge ∼2500 separate datasets. Moreover, given the partial nature of the snapshots, only a portion of the key information is immediately available, while, for instance, columns related to the categories are missing. Fortunately, the R library `bggAnalytics` [5] provides a function that is able to fetch data directly from the original online database through a series of get requests. While the process of accessing the online database is problematic, due to the large amount of data we need and the limits imposed to the traffic by BoardGameGeek, at the best of our knowledge this is the only way to obtain a complete dataset.

Once we have a dataframe complete of all the information we need, it has to be cleaned. In particular we expect a monotone behaviour from the number of ratings as a function of time; however the real data shows some inconsistencies, possibly due to disappearing user accounts and/or bad maintenance. Moreover, many games were published during the observation period, resulting in a bunch of NA values in the first part of the dataset. Our strategy consists in replacing NAs with 0 and adjusting each value to the one of the day before if it decreased.

While assuming the number of ratings in time as a measure of the popularity of the game (more on this later), it is natural to look at the derivative of such metric if we are interested in possible trends in the board games market. Due to the high level of irregularity in the data, we perform a monotone smoothing, using a spline basis of degree 4, with 10 basis functions and $\lambda = 1e - 2$ as penalising term on the third derivative. For some of the functions, the monotone smoothing process failed, possibly because of a jump discontinuity at the moment of publishing. For those instances we resorted to the classical unpenalised spline smoothing.

Finally, we consider the smoothed functions as well as their first derivatives. To ease the computational burden on the following analysis, we downsample in time, reducing the abscissa discretization from ∼2500 to 100 time instants.

3 | Tools and Analyses

3.1. Measure of popularity

Since our goal is to describe the commercial success of board games, we need a measure of their popularity. The natural choice would be to use sales data, but unfortunately they are not available, therefore we have to choose among three possibilities: the number of users who rated the game, the BGG rank or the average score in the BGG charts.

However, when considering the rank or the average score, the bias due to the specific pool of users that accesses and populates the database may be an issue. Indeed we noticed that some very notable games (i.e. commercially successful) are somehow treated poorly by the typical BoardGameGeek user, leading to low ranks and low average scores despite their popularity. In other terms, those two variables are skewed towards games that are known, appreciated and played mainly by the most loyal niche of the public.

While not completely solved, we found that the number of users who rated the game tempers the issue. Indeed we think that it better represents the number of people who played the game, which is conceptually close to the sales of such game, at least at the best of the available data.

Finally, we want to stress that this choice is made here once for the rest of the analysis. Up to some situational corrections, the number of rating will be the target variable for all the subsequent models and, in general, our measure of popularity.

3.2. A model for the Snapshot dataset

The *Snapshot dataset* is mainly used to build a generalized linear model with the goal of better understanding what characterise a popular game. To do so we fit a semiparametric model introducing the variables accounting for the categories as linear terms and the other numerical variables (gameplay features), such as *min players*, *weight*, *minage*, *dimpublisher*.., as smooth terms. For this purpose, after trying different splines fit, we choose to use thin plates regression splines of third order. Regarding the response variable, instead of considering directly the number of users rated, we fit the model using *log(users rated+1)*, with the choice motivated by a heavily skewed distribution.

After fitting the model, we proceed by doing variables selection to remove the categories with very little significance: we select all the significant covariates with respect to a permutational t-test at level $\alpha = 0.05$. Then, on the reduced model we introduce interactions between some of the numerical variables, smoothing them with cubic splines.

To address the issue of outliers, we build a robust model trimming away the most contaminated observations via Minimum Covariance Determinant method. For this purpose we set $\alpha = 0.05$ and we analyse only the most problematic numerical covariates, namely *max players*, *playing time*, *year published*.

Our final model is:

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

3.3. Analysing the Historical dataset

3.3.1. Promising game categories

To have a better comprehension of the evolution of the market, we focus also on studying how the popularity of games changed in the past years. Indeed, we perform in parallel two kinds of Analysis of Variance on both the curves and their derivative to study the commercial performance of the games across time with respect of their category.

In particular we are testing the hypothesis that the specific effects of each category are statistically null, namely:

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

Note that, in the case of the fANOVA (discussed below), the effects are functions of time $\tau_i(t)$. The p-values of the one-at-a-time tests are corrected via Bonferroni, considering a threshold of level $\alpha = 0.05$. The general idea is to identify a few promising categories associated with good commercial performances that would justify targeted investments and a renewed production effort.

Globalized Pointwise F-test

Since we have functional data, a natural approach would be to perform some sort of Analysis of Variance directly on the functions. However the usual MANOVA is not directly applicable, due to the high dimensionality of the data and the high correlation across time, that results in badly conditioned matrices and the failure of the numerical methods.

The literature proposes the Globalized Pointwise F-test [6], which consists in a series of "vanilla" ANOVA tests, performed pointwisely for each time instants:

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

where

- k is the number of groups of random functions defined over a given finite interval $\mathcal{T}$

- $n = \sum_{i=1}^{k} n_i$ is the total sample size
- $SSR_n(t)$ and $SSE_n(t)$ are respectively the pointwise between-subject and within-subject variations

In this framework, the F statistic is a function of time and it is globalized by integrating over time:

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

The test is extremely simple and also very computationally efficient. Moreover, it can be proven to be root-n consistent. Some of the details can be found in the Appendix A.2.

In R, the `fdANOVA` package [6] provides the function `fanova.tests` , that implements the one-way GPF test as well as a few (computationally expensive) alternatives. Considering the large amount of categories to test simultaneously and the fact that a game can belong to different groups at the same time, we eventually decided to implement our hard coded version of the multi-way test.

Moreover, the GPF test as proposed by Zhang and Liang assumes that the functional data are realizations of Gaussian processes as they draw a series of conclusions on the distribution of the F statistic (see Appendix A.2). Therefore we decided to follow a (nonparametric) permutational approach. Once obtained the functions $\tau(t)$, understood as the effects of the categories, we compute one-at-a-time confidence intervals at level $\alpha = 0.05$ in a bootstrap fashion.

ANOVA on the ranks

A second path we follow consists in performing the usual multi-way ANOVA not directly on the functional data, but on their rank. The general idea is to reduce the dimensionality, while preserving the relevant information about the amount of time a game or an entire category perform better than the others.

The natural choice for the target variable would be the Modified Band Depth, since it is the extension of the notion of rank for the one-dimensional functional data. However we are not exactly interested in a notion of "centrality", since it would associate well-performing and badly-performing games to similar small depths, but rather on a notion of "dominance" of a function with respect to the others. Therefore we resorted to the Modified Hypograph, that quantifies the amount of time the sample spends below a given curve and arranges the functions on a down-upward order.

This approach is highly flexible and computationally efficient, hence well suited for the problem at hand. For this reasons, it allows to explore some possible interactions between the categories. Also in this case, we performed a permutational version of the test. As we will see later, the results are fairly congruent with the ones from the GPF test.

Some alternatives

While we focus mainly on the previously discussed paths, we explored also some alternatives. The previously mentioned `fdANOVA` library provides some L^2-norm-based bootstrap tests and random projections based tests, which may be of use. The main reason we discarded them is their prohibitive computational cost, incompatible with the data and the machine available.

A second alternative we briefly explored is the Functional Regression Analysis, as presented by Ramsay and Silverman in [7]. Without going into too much detail, the general idea is to fit a model of the type $y_i(t) = \beta_1(t)z_i + e_i(t)$, where z_i is our constant dummy variable for the category of interest and $y_i(t)$ is the number of ratings. The R package `fda` [8] provides the function `fRegress`, which carries out this kind of analysis. The perspective of fitting this kind of model is fascinating, considering the high flexibility it provides. However, the computational cost is still an issue, especially if we want to follow a permutational or bootstrap approach.

Another possible path could be the classical MANOVA on the coefficients of the functional data in the spline basis, therefore reducing the dimensionality of the problem by a factor 10 in our case and possibly solving the problem of badly conditioned matrices. The general idea is that our data is strongly correlated in time and the coefficients obtained during the smoothing should be enough to characterize the data. While less trivial and more difficult to interpret, this approach could be interesting to be pursued.

3.3.2. Longevity and success in the market

At this stage of the analysis, we advance our investigation by employing a survival model to delve deeper into the commercial lifetime of a board game. Precisely, within our framework, our objective is to study the long-term efficiency of marketing and distribution strategies for a board game that has already been released and leverage this information for upcoming games that share similar characteristics with those already in existence.

Choice of the Time-to-Event

As is widely known and expected from a capitalist perspective, every board game in the market experiences a natural lifecycle shaped by players engagement, progressing from discovery to boredom and eventual migration to alternative entertainment forms. More specifically, it comprises five distinct phases: Introduction, Growth, Maturity, Decline, and Niche (i.e. the final demise of the genre in the mainstream market) [9].

Given that our original dataset lacks information about the survival time of games, after exploring the market behavior, we choose to experiment with different procedures to determine the accurate value for the Time-to-Event:

1. Month method;
2. Quantile method.

The first approach involves declaring the actual death of a game if, for a consecutive month, the number of ratings remains constant from one day to the next. To gain a more comprehensive understanding of the phenomenon, we explore additional variations, such as extending the evaluation period to include more months, ranging from two to twelve, and modifying the criteria for rating transitions to consider fewer than two new ratings per day or even five.
On the other side, the second approach consists of designating as the Time-to-Event the day on which a game first reaches a number of ratings equal to the 95^{th} percentile.

In the end, we decide to follow the first idea, considering a one-month timeframe with a constant number of ratings, as it appears to be the most logical and straightforward approach.
Additionally, we exclude a small number of games that were already discontinued at the beginning of the study and treat games that never ceased to exist in our time window as censored data.

Cox Model

Having both the survival time and the status for each game, we shift to the core aspect of this topic: the Proportional-Hazard Cox Model.

However, concerning continuous variables, we deviate from its basic version. Specifically, by employing splines [10], we relax one of the usual assumptions associated with the Cox model: the linearity in the covariates. This choice is practically motivated by the superior performance of splines in capturing complex patterns without imposing overly conservative beliefs on the shape of the relationship [11].
Initially, we consider various types of splines, including B-splines, Natural splines, Penalized splines (an extension of the well-known B-splines incorporating a penalty term to control model complexity and avoid overfitting) [12], and Restricted Cubic splines (again, a particular case of B-splines that forces the shape of the response to be linear before the first knot and after the last one)[11].
But, in the end, we opt for B-splines with 3 degrees of freedom since they perform fairly better thanks to their adjustable flexibility and direct interpretation.

Gameplay features As an initial step, we direct our attention to all the numerical variables that we deem meaningful in this specific context. These attributes include the *year of release*, *minimum age*, *playing time*, *maximum number of players*, *complexity*, and the *dimension of the publishers*.

Categories Then, we proceed by considering all the *categories* of games available. We have also computed the Kaplan-Meier estimator in this case. However, given that the results align with those obtained from the Cox model, we provide a detailed analysis solely for the latter.

Complete model Finally, to have a more comprehensive view of the impact of all the features together, we adopt a model that incorporates all the previously mentioned variables.

4 | Results and Achievements

There are three types of lies:
lies, damn lies, and statistics.

Benjamin Disraeli

4.1. On characterizing good board games

The first complete model we fit includes all the games categories and the seven gameplay features *min players, weight, log(playing time+1), min age, dim publisher, avg publisher* and *year published*. As mentioned before, we choose to introduce the former smoothed with thin plates regression splines and the latter as parametric linear terms. We find that $R^2_{adj} = 0.527$ and only 43 categories are significant at level $\alpha = 0.05$. However before removing the remaining 41 from the model, we use a permuational test to assess whether we can switch to a reduced model or not. Hence we test if all these variables' coefficients can be set to zero; we get a p-value of 0.44 strongly suggesting that we can move to the reduced model without further ado. Some of these uninformative categories are *Mythology, Space Exploration, Pirates, Music* and some more related to specific kinds of wargames. For what concerns the gameplay features, they are all very significant, with an estimated degree of freedom that ranges from 7 to 9.

The reduced model has a slightly better $R^2_{adj} = 0.584$; the numerical variables are again all significant while for the categories we still have a large number of variables with small p-values. In particular, among the significant ones we find *City Building, Farming, Science Fiction, Political* with a positive coefficient and *Children's game, Wargame, Memory, Educational, Napoleonic* with a negative coefficient.

The next step is building a robust model by removing some observations that might be outliers. For this purpose we focus on the gameplay features *max players, playing time, yearpublished* and, employing the Minimum Covariance Determinant method, we rule out 1062 observations. This robust model has an $R^2_{adj} = 0.513$ which grows to 0.542 when we perform variable selection by removing all the not significant categories at level $\alpha = 0.05$ and introduce the following interactions between the features using cubic splines: *maxplayers:log(playingtime+1)*, *weight:log(playingtime+1)*, *weight:dimpublisher*, *dimpublisher:log(playingtime+1)*.

In what follows we show the results of the robust models with interactions, plotting the variables' effects on the response.

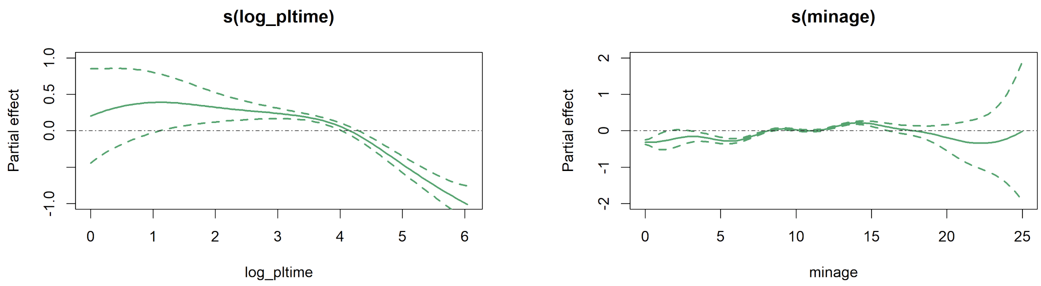


Figure 4.1: Left: a decreasing trend drives the variable behaviour, which assumes positive values for short to average playing time games and negative values for longer ones. Right: there seems to be a slight positive edge for games that require a higher minimum age. On the other hand, we expect a negative behaviour for games suitable for young children.

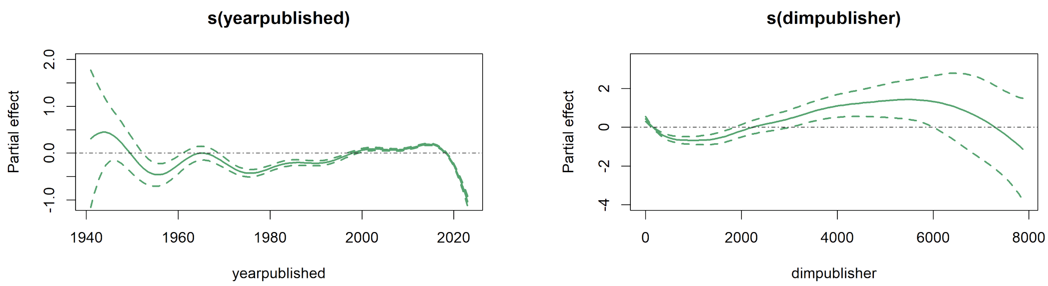


Figure 4.2: Left: it seems that old games have mostly terminated their commercial life, while recently published ones tend to be more popular in general. For the newest games, it may be too soon to draw conclusions. Right: self-published games tend to perform moderately well on average, the same cannot be said for board games published by small publisher, which may not have the resources to adequately market their products; on the contrary, large publishers have a positive effect on average, with the few largest ones probably affected by a sort of internal competition.

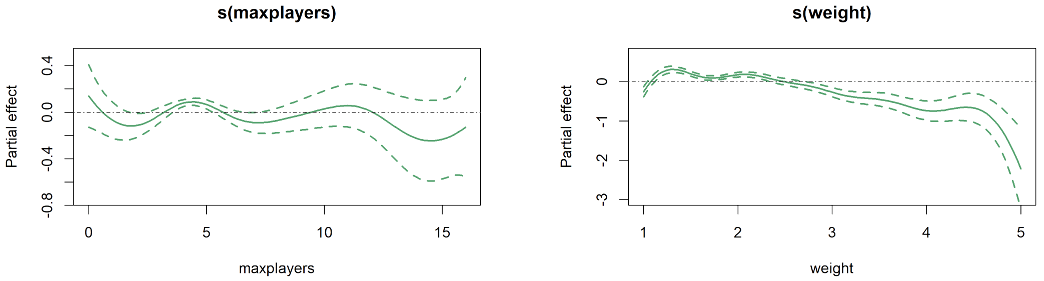


Figure 4.3: Left: the maximum number of players doesn't show a clear pattern, in general it has a positive effect on popularity when set around 5 and a not so relevant effect for all other values. Right: on average we see that the most complex games are also the less popular, with the public more attracted by simpler board games.

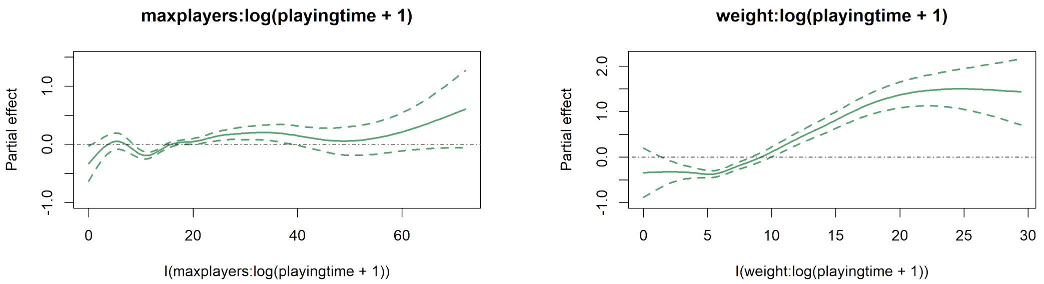


Figure 4.4: Left: the interaction between the *max players* and *playing time* mainly oscillates around the zero with a slightly positive edge for average of both. Right: high *weight* and *playing time* have a positive contribute to the popularity of a game, whereas games with both low values of these values tend to be dragged down.

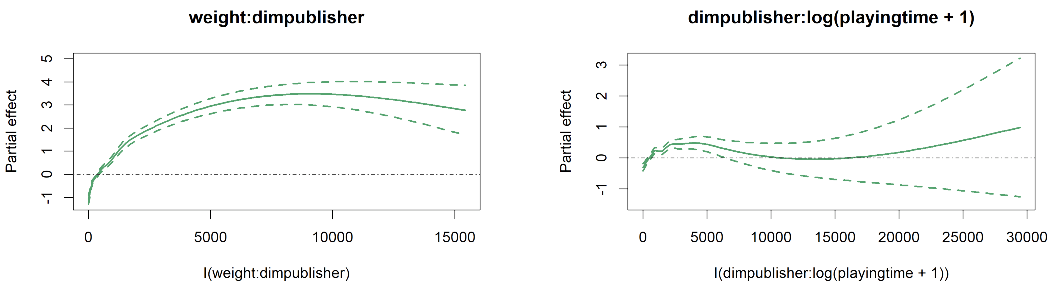


Figure 4.5: Left: there seems to be some form of trust towards the big publishers when they publish very complex games, with a positive interaction of high *weight* with large *dim publisher*. Right: large values of both *dim publisher* and the playing time do not seem to be particularly relevant; however, games of moderate length published by middle-sized publishers seem to have their share in the market.

4.2. On choosing the category

As we previously discussed in Chapter 3.3, we are able to select a handful of promising game categories on the basis of their performances in time since late 2016, in terms of both absolute popularity, studying the actual functions, and current trends, studying the derivatives.

Minor differences arise between the tests. Indeed, some categories are stronger in terms of trends (e.g. deduction games, which are significant in the tests on the derivatives, but they fail to stand out while considering the ranks of the cumulative popularity). On the other hand, some categories are popular in an absolute sense, but their current trends do not deviate too much from the benchmark of the market (e.g. bluffing games, not significant with respect to the fANOVA on the derivatives).

In general, we can say that the ANOVA tests on the Modified Hypographs are less conservative, possibly because of the heavy dimensionality reduction performed while considering only the rank.

4.2.1. The best categories

All the four tests we performed agree in identifying the following categories as significantly more popular, hence a safer choice while designing a game:

- economic;
- fantasy;
- medieval;
- territory building;
- civilization;
- city building;
- exploration;
- farming;
- science fiction;
- fighting;
- renaissance.

The following plots show the effects of each category $\tau_i(t)$, on the cumulative number of ratings (left) and on the current trends (right), as long as the behaviour of the market $\mu(t)$, that we can think of as the benchmark.

For some of these categories, the popularity has been considerably and consistently strong, as for *Economic* (Figure 4.6), *Territory* and *City building* games (Figures 4.9 and 4.11), with a specific effects that is even more impactful than the market.

Most of the categories observed some peaks, usually around the pandemic, highlighting a behaviour that is visible also for the market as a whole, which benefited from the last few years.

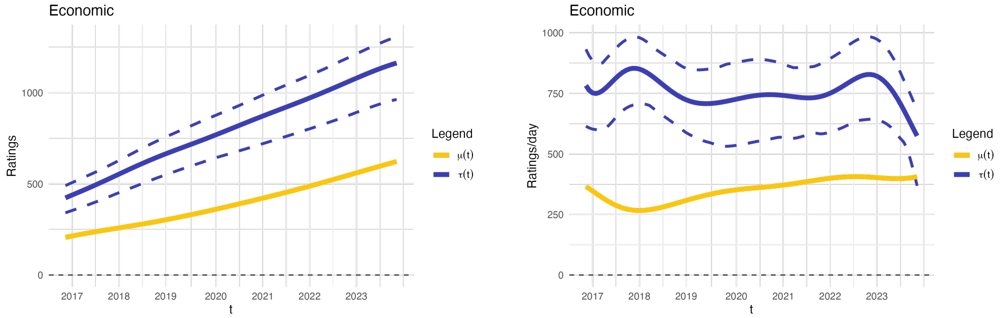


Figure 4.6: Economic games. The popularity is strong and growing, with a couple of peaks in velocity in 2018 and 2022. During the last few months the growth slowed.

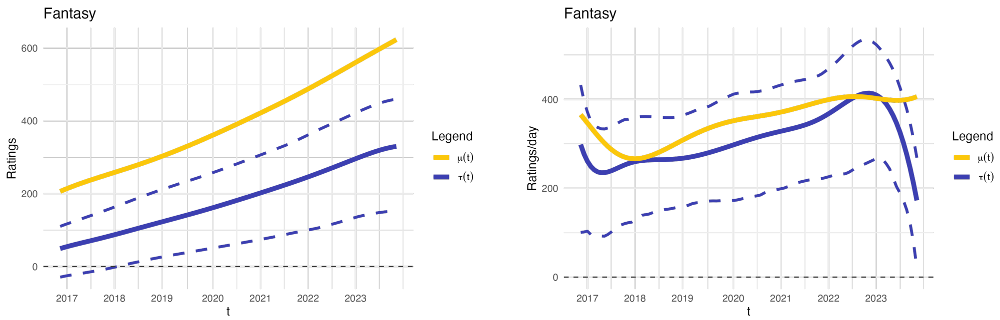


Figure 4.7: Fantasy games. Their market share expanded consistently, until 2023 when growth suffered a moderate contraction. A prelude of saturation, after the end of Game of Thrones?

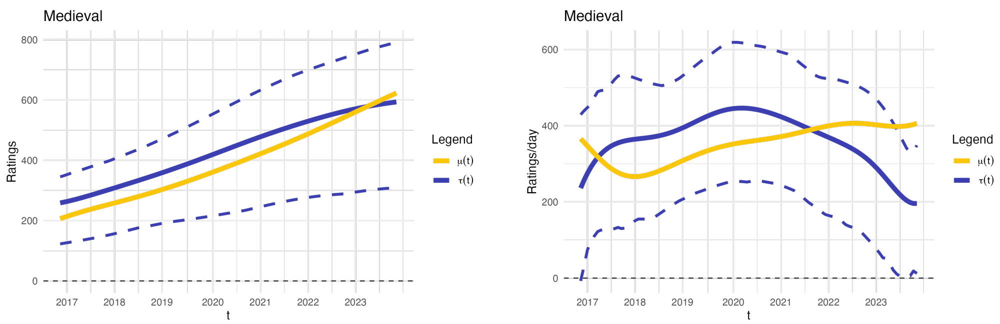


Figure 4.8: Medieval games. The growth in popularity reached a maximum in velocity around the pandemic, but they seem a solid category of games.

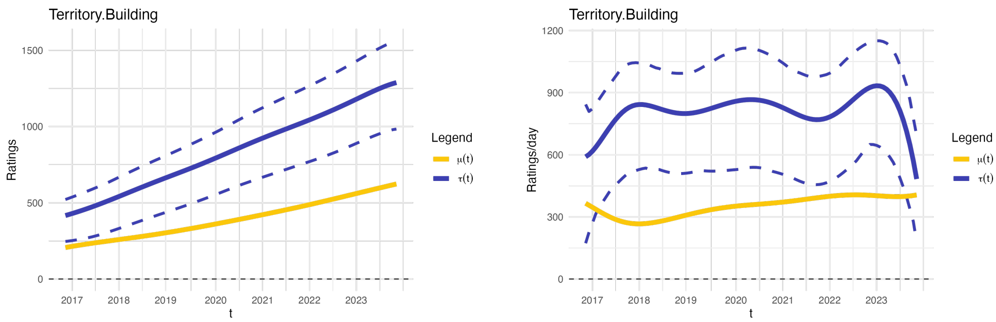


Figure 4.9: Territory building games. This category grew very fast during the last few years, until showing an initial slowdown in 2023. A solid category, nonetheless.

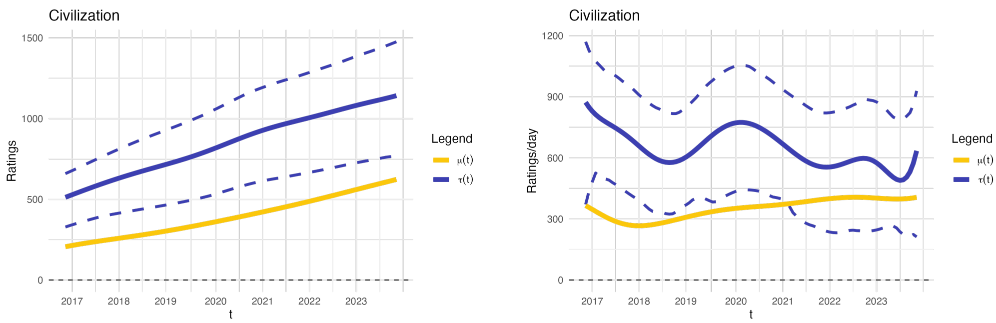


Figure 4.10: Civilization games. They show a few peaks in velocity and a hint of acceleration during the second half of 2023. We expect their market share to grow even more in the next months.

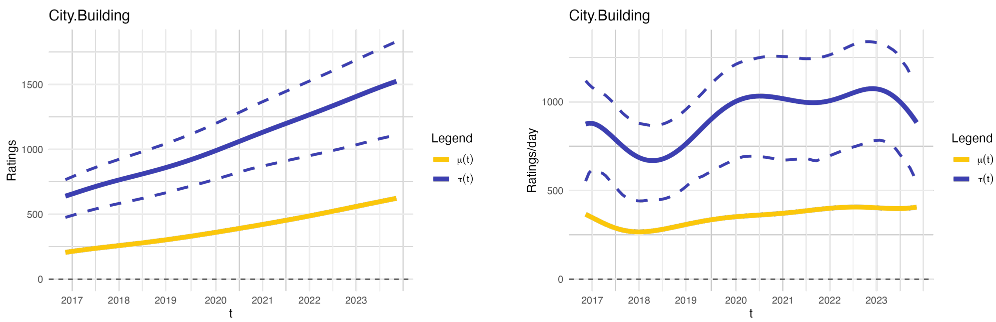


Figure 4.11: City building games. During the pandemic, their popularity hugely and consistently expanded.

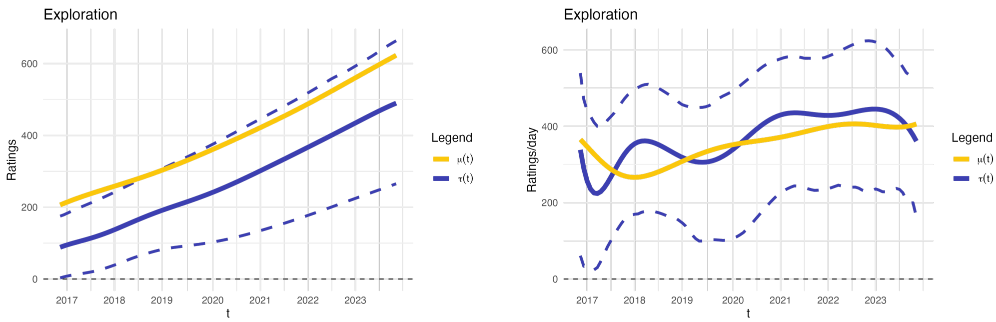


Figure 4.12: Medieval games. They show a moderate and consistent growth, especially since late 2020.

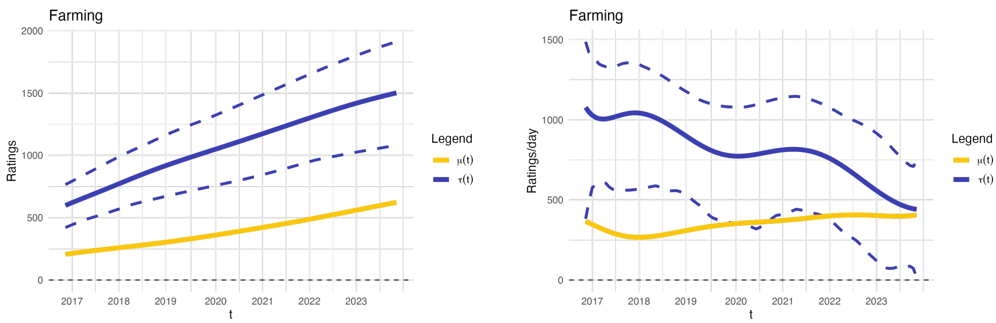


Figure 4.13: Farming games. While their popularity is solid, the growth is slowing. The next few months may be the last occasion to publish a farming game before reaching a level of saturation.

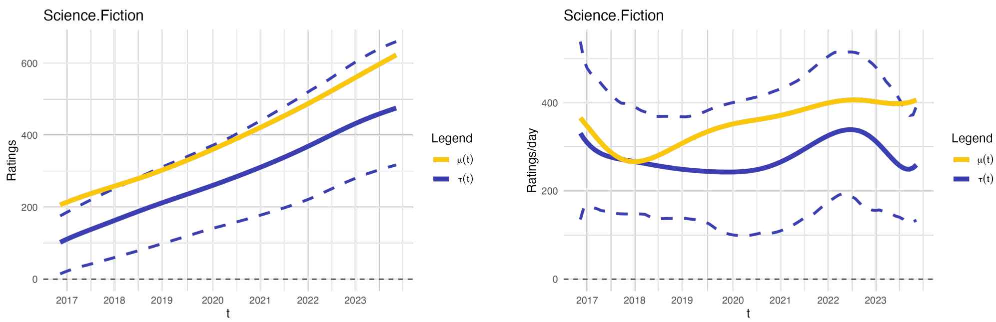


Figure 4.14: Science fiction games. Their popularity is moderately significant, with a mostly constant growth and a peak in 2022 (Dune effect?).

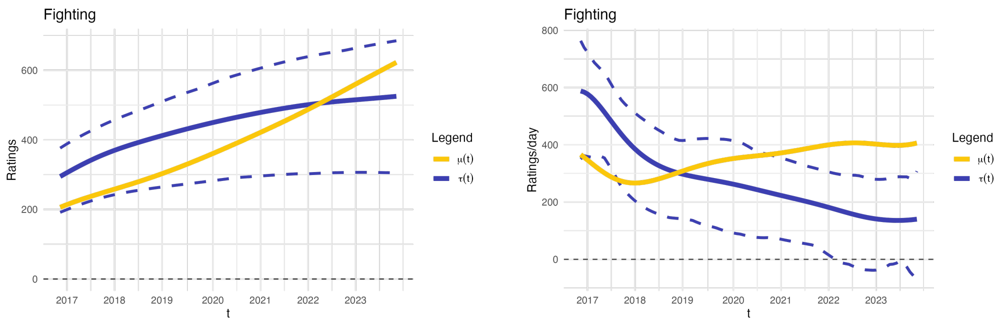


Figure 4.15: Fighting games. Since late 2022, the specific effect of this category is closer to be negligible with respect to the average of the market. While still a solid choice, we may have reached a point of saturation.

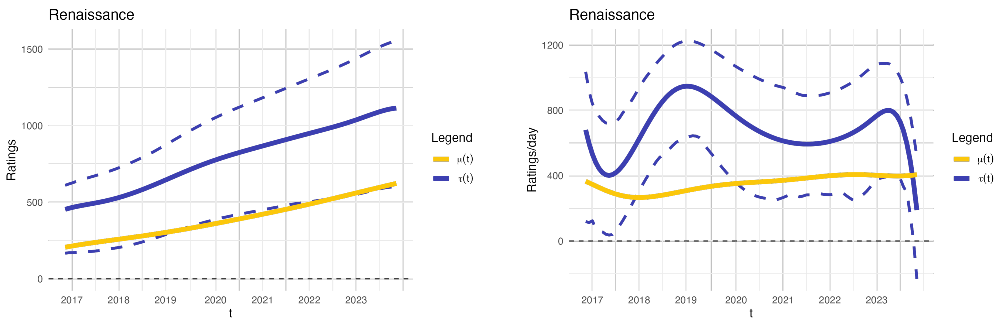


Figure 4.16: Renaissance games. They showed a consistent growth in popularity since late 2016, with a couple of peaks. However, the vertical fall of 2023 should be monitored.

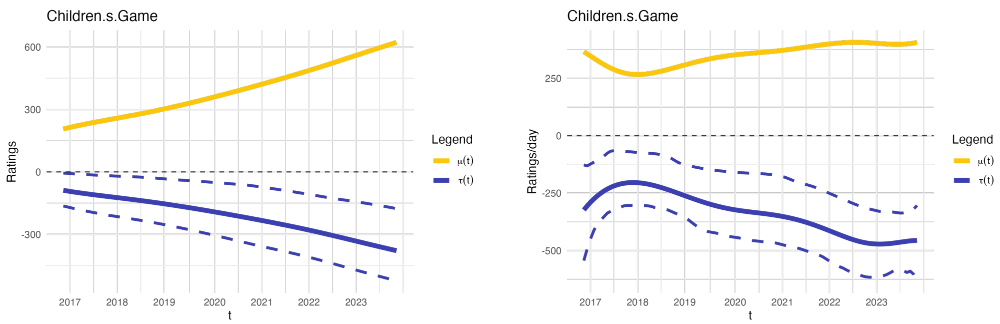


Figure 4.17: Children's games. The effect of this category is consistently negative, hinting that the market share is long saturated and it is advisable to reduce the production of this kind of games. However, especially in this case, we should take into account the bias of BGG.

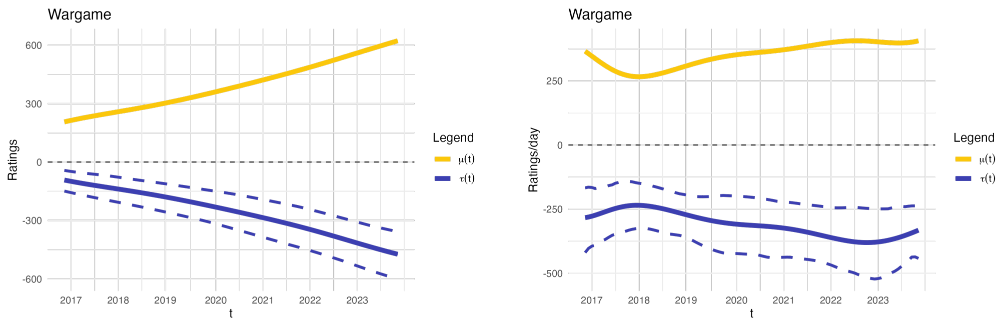


Figure 4.18: Wargames. The relative market share is significantly and continuously reducing, suggesting a saturation. Unless sudden changes of trend, it is advisable to avoid producing wargames.

4.2.2. What to expect from the near future

While purely indicative, the behaviour during 2023 could give us a hint on what to expect from the next months. Indeed, categories as *Economic* (Figure 4.6), *Fantasy* (Figure 4.7) and *Territory building* (Figure 4.9) observed a contraction in the velocity of their growth. While still well performing, they may have reached a peak in their growth. Expecting a similar level of success and a continuous departure from the benchmark may be unrealistic.

On the other hand, *Civilization* games weakly suggest an inversion of trend during late 2023, that could be followed by a strong 2024 (Figure 4.10). While the evergreen "*Twilight Imperium*" and "*Terra Mystica*" will lead the way, there should be room for some new gems as this market share expands.

In general, while most of these categories look promising, a few of them have slowly regressed toward the average of the market. This is the case for *Medieval* and *Fighting* games (Figures 4.8 and 4.15), and to some extent for *Farming* games (Figure 4.13). The vertical collapse of *Reinassance* games in the last months should be monitored as well (Figure 4.16).

4.2.3. Two categories to avoid

There are two categories that are consistently performing worse than the market at least since 2016 and they are *Children's games* and *Wargames*, as we can see in Figures 4.17 and 4.18. We suspect that these two categories long reached a point of saturation and we believe it would be unwise to pursue the design and the production of similar games.

4.2.4. A few words on interactions

A certain number of possible interactions between the significant categories have been explored with uncertain results. On one side it is clear that interactions are a direct measure of possible subcategories. Indeed, some of the significant ones include *Economic* with *Territory building*, *Medieval* with *Ancient*, *Medieval* with *Renaissance*, *Civilization* with *City building*, which are particular instances of the two broader categories.

However, for all the cases we examined, we observed negatives effects on the popularity suggesting that a too granular characterization could perform badly on average, since only exceptional games excel. While *Civilization* games in general are a safe choice, the specifics of the particular game will be more relevant and decisive in terms of commercial success than the choice of the subcategory.

If nothing else, these conclusions should highlight how complex is the design of a board game and how superficial its categorization could be.

4.3. On the longevity of a board game

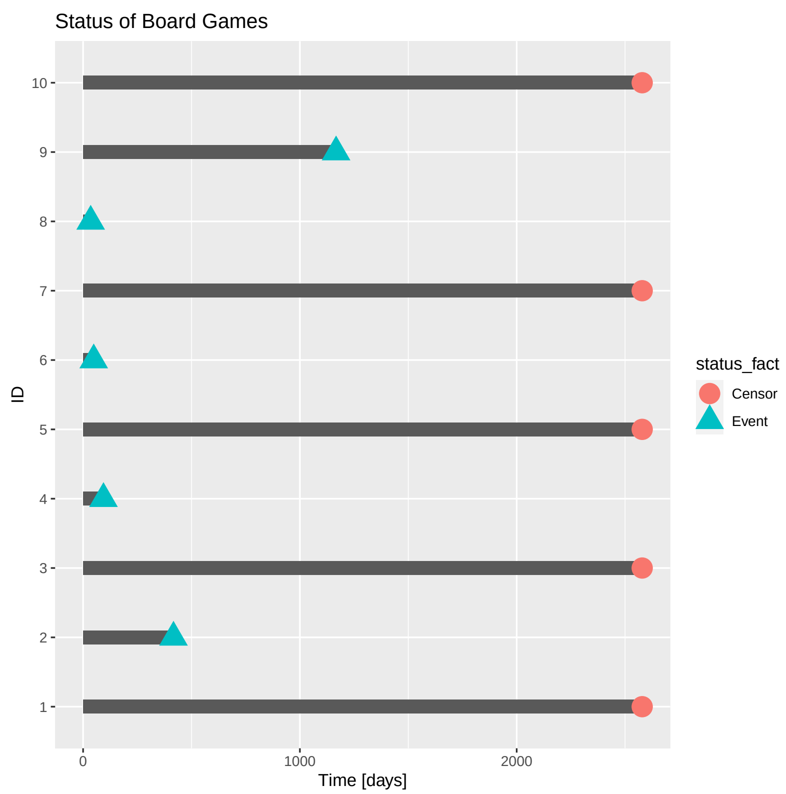


Figure 4.19: Status of some board games

4.3.1. Features impact

In this section, we move on to the results related to the *gameplay features*, crucial elements in our survival analysis. These variables provide future insight into how specific factors may influence the long-term duration and success of board games.

We explore the results of the Cox model (Figures 4.20, 4.21 and 4.22), aiming to capture the nuances contributing to the relevance of these variables within the context of our study.

Year of the release

We can first identify a bad change in the direction of survival in correspondence of the late fifties and early sixties, driven by the emergence of the very first video games, such as "Tennis for Two" (1958) and "Spacewar!" (1962). The peak is reached in the late nineties and early 2000s due to the advent of the internet. Fortunately, a new shift occurs thanks to the revival of old games and especially to the establishment of crowdfunding platforms like Kickstarter (2009). The situation has been steadily improving from the onset of the COVID-19 epidemic until the present due to the growing interest and necessity for social interaction during lockdowns. (Figure 4.20)

Minimum age

The age group between 9 and 18 years exhibits a positive impact on the survival of the game. This result can be interpreted by the habit of the gaming industry to prioritize the production and promotion of games targeting this specific age range and also to advance their variety and innovation. (Figure 4.20)

Playing time

Based on real-world scenarios, it is reasonable to assert that the time interval between 30 and 120 minutes is optimal for reaching a broader audience, leading to a decreased risk of discontinuation. Simultaneously, with the emergence of new figures in the gaming community, especially professional gamers [13] who dedicate increasing amounts of time to training, it is evident that more complex games with a playing time exceeding 3 days have a positive impact on survival. (Figure 4.21)

Maximum number of players

Also the results for the *maximum number of players* appear to be significant and consistent because of the wider audience reached by board games requiring a maximum number of players ranging from 3 to 9. (Figure 4.21)

Complexity of a game

As the complexity of a game increases, its effect on the survival of the game is positive. This might be influenced also by the origin of our data since they are linked to the nerdy component of the market, i.e., the one more interested in strategic and complex games. (Figure 4.22)

Marketing effect

We conclude this section with the last variable. As the size of a publisher increases, there is a noticeable positive impact on the game's survival. This observation is logical, as larger publishers tend to instill greater confidence in consumers compared to smaller ones. (Figure 4.22)

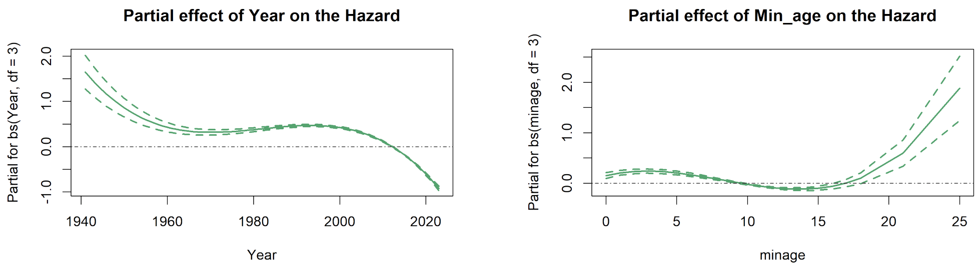


Figure 4.20: Left: partial effect of *Year of release* variable on the Hazard for the Cox model for gameplay features. Right: partial effect of *Minimum age* variable on the Hazard for the Cox model for gameplay features.

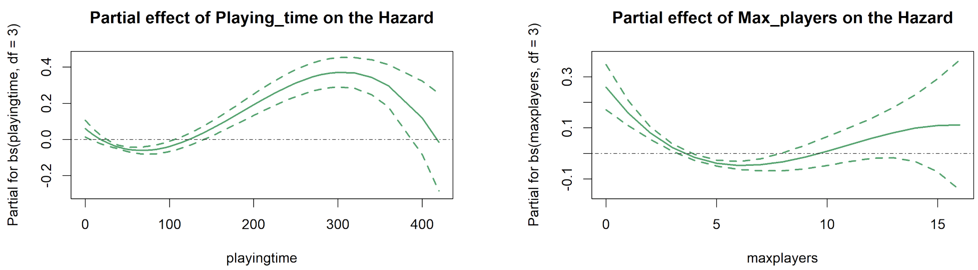


Figure 4.21: Left: partial effect of *Playing time* variable on the Hazard for the Cox model for gameplay features. Right: partial effect of *Max Players* variable on the Hazard for the Cox model for gameplay features.

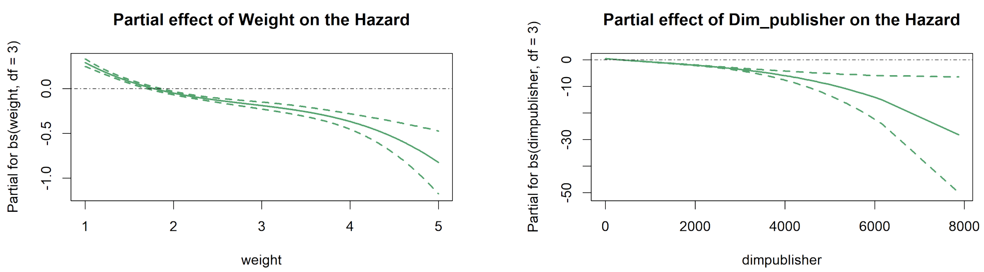
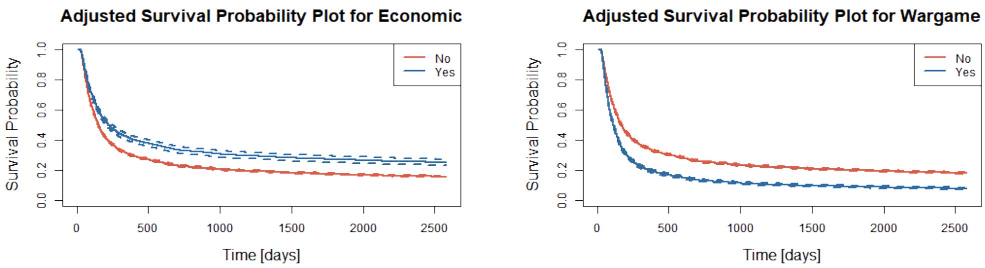


Figure 4.22: Left: partial effect of *Weight* variable on the Hazard for the Cox model for gameplay features. Right: partial effect of *Dim Publisher* variable on the Hazard for the Cox model for gameplay features.

4.3.2. Decoding diversity: game categories

As we delve into the heart of our analysis, a meticulous Cox model awaits, enabling us to explore the intricate dance of 84 distinctive board game categories. From the strategic realms of war games to the fantastical landscapes and the intrigues of political narratives, each category contributes a unique thread to the rich tapestry of board gaming.

Firstly, we analyze the survival dynamics of each category, stratifying into two groups based on the membership. Naturally, we encounter scenarios where being part of a specific category guarantees a reduced risk of discontinuation, while in other instances, the opposite is observed, as visible in Figure 4.23.



(a) Economic: a protective category (b) Wargame: a risk category

Figure 4.23: Behavior of different categories

But here comes the more interesting part. Since we need to provide an all-encompassing guide for producers navigating the intricate landscape of board game diversity, we provide a distinctive model that incorporates all these variable types, aiming to understand how their collective presence shapes the landscape of tabletop gaming.
Upon closer examination, a noteworthy pattern emerges. The majority of the categories seem to be significant at level $\alpha = 0.05$, except for 34 of them. Among the more risky ones, we recommend exercising caution with *Abstract Strategy, Children's Game, Collectible Components, Print&Play, Racing, War Game, Humor, Memory, Trivia, Sports, Game System, Book, Educational and Mature&Adult*, since in this way a decreasing impact on the risk of niche seems more plausible (Figure 4.24)

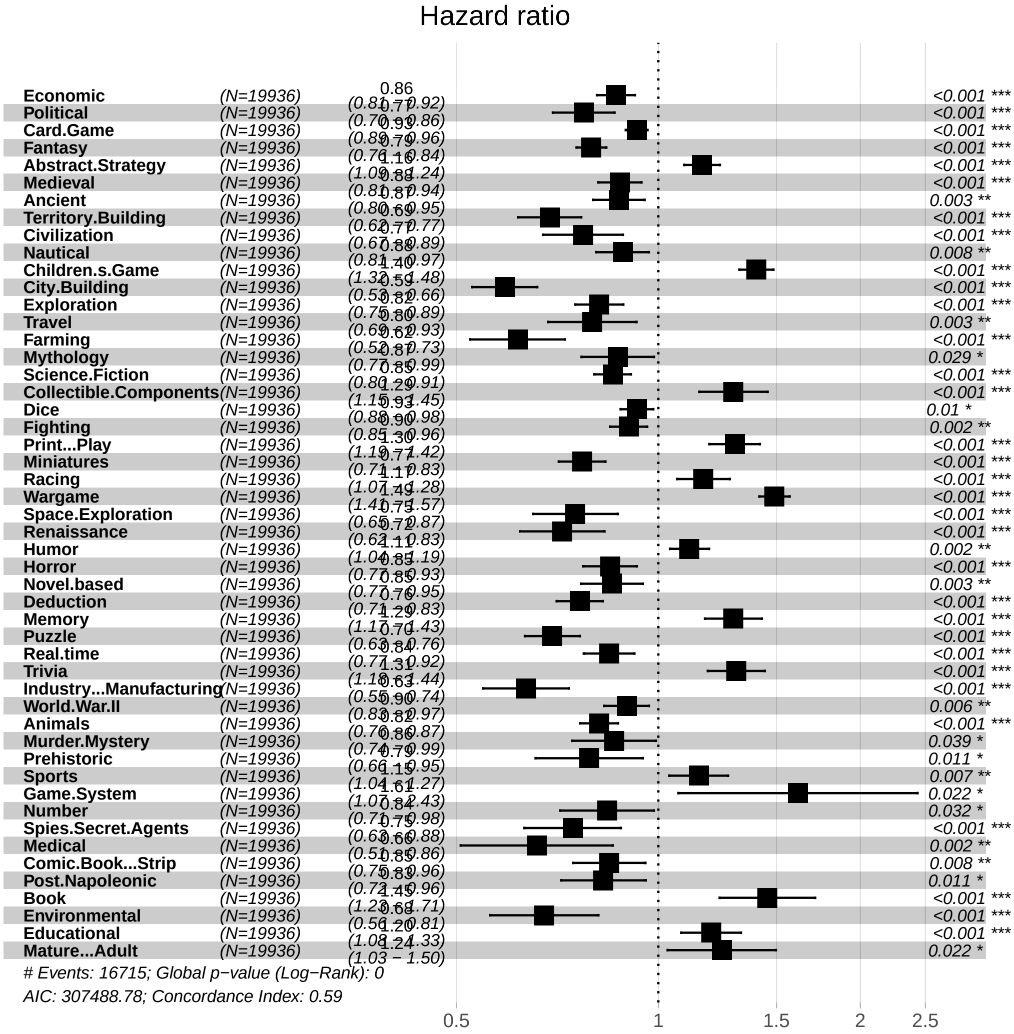


Figure 4.24: Hazard Ratio per category

4.3.3. Comprehensive board game dynamics

Relying on our comprehensive model, let's take a closer look at what makes board games stand the test of time, unraveling the factors that truly impact how long a game remains popular (Figure 4.25).
Upon an initial inspection, the baseline survival probability curve, derived from applying this model, enhances the traditionally brief lifespan of the majority of board games.

Focusing on the *gameplay features*, we note a behavior largely consistent with the one detailed in section 4.3.1, except for the reduced significance at level $\alpha = 0.05$ of the variable *max players* (Figures 4.25, 4.26 and 4.27).

More intriguing findings emerge from the *category* variables. We can now exclude 64 categories from our analysis, allowing for a more focused examination of the remaining, more significant ones, at level $\alpha = 0.05$. Among these, only *Negotiation,* *Children's Game, War game, Memory and Book* exert a negative influence on the overall survival, while *Political, Card Game, Territory Building, City Building, Exploration, Farming, Dice, Fighting, Miniatures, Industry&Manufactoring, Animals, Transportation, Action&Dexterity, Spies Secret Agents and Environmental* are the categories we recommend considering.

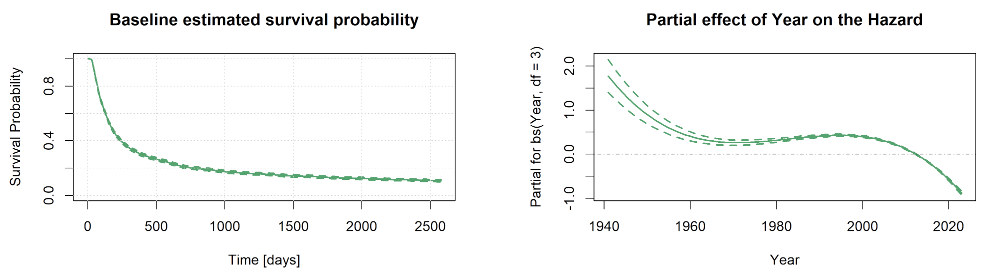


Figure 4.25: Left: Baseline estimated survival probability curve for the complete Cox model. Right: partial effect of *Year of release* variable on the Hazard for the complete Cox model.

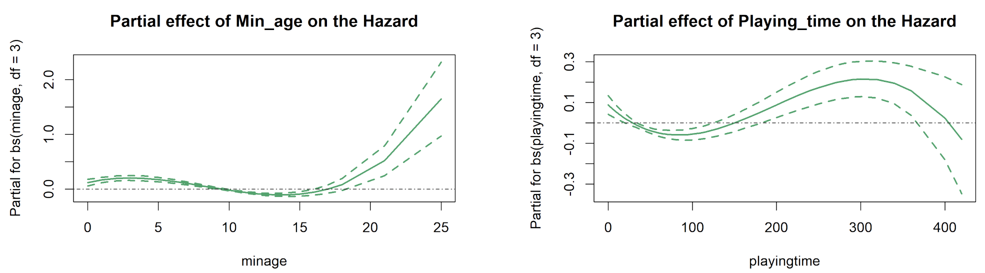


Figure 4.26: Left: partial effect of *Min age* variable on the Hazard for the complete Cox model. Right: partial effect of *Playing Time* variable on the Hazard for the complete Cox model.

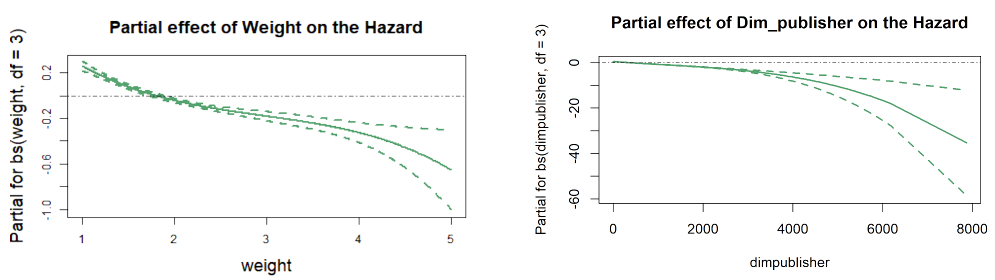


Figure 4.27: Left: partial effect of *Weight* variable on the Hazard for the complete Cox model. Right: partial effect of *Dim Publisher* variable on the Hazard for the complete Cox model.

5 | Conclusions

Let's roll some dice.

Me, right now

In this study we explored the features that characterise the best board games out there, both in terms of absolute commercial performance and longevity in the market. Our aim has been to draw a map for the designers, to aid them in their creative process and highlight which can be the safest choices with respect to the market demand. On the other hand, we focused on the specific needs of publishers. We want to give them the necessary elements to decide what will be their next production, which projects are worthy of their investments and to better assess the nature of the commercial performances of their internal departments.

The results of our analysis showed that *Civilization* games historically performed really well, both in terms of absolute popularity, growth and longevity on the market. Moreover, the last few months of 2023 suggests that we may expect their market share to enlarge once more. Some other categories that seems solid choices in this regards are *Exploration*, *Territory* and *City building* games, since they tend on average to be really appreciated by the public and to perform well for long periods of time.

From a technical perspective, we showed how on average less complex games with a moderate playing time are more dominant, both in absolute and in a long term perspective. However, also more complex games have their fair market share if combined with a more important time length of a game, especially if published by noted and trustworthy companies, which may guarantee support and expansions in time.

In terms of marketing, smaller publishers should focus their efforts on a proper promotion of their products, while larger companies may want to address possible cases of internal competition and saturation.

We believe this work propose a strong and solid characterisation, possibly the deepest publicly available, and we are excited to see what will be the next best board game.

Further developments

During this research, we frequently explored and discussed different paths that could follow this work.

First and foremost, while sales data could provide a more insightful knowledge in commercial terms, the BGG dataset still offers some features that may be of use. In particular, it may be interesting to perform an analysis on the game mechanics, similar to the one on the game categories but addressing the cumbersome issue of the high cardinality. Information about the designers and artists is also available and could be exploited with a bit of creativity. Moreover, daily snapshot of the database are still collected to this day at [4].

With respect to the historical data, the Globalised Pointwise F-test is very flexible and it may be interesting to integrate the test statistic only on a restricted time frame or possibly weight differently the integral, especially if working in a nonparametric framework. In Chapter 3.3, we discussed possible alternatives to the ANOVA tests we performed. We believe that the Functional Regression Analysis is the more promising and it could allow to include also the numerical feature in a single model.

Regarding the survival analysis part of this work, our measure of longevity has some flaws and we did not address the possibility of games with a "second life", namely that for some reasons they suddenly return to perform really well. To mitigate this concern, one possible strategy is to approach the situation from a reliability standpoint. This involves conceptualizing a game as, for instance, a circuit component, developing a procedure to examine its maintainability, and subsequently transitioning back to a survival framework.

Finally, it may be interesting to consider the historical data as time series, in order to perform some sort of forecasting. We briefly explored the path of adopting an ARMA model in a Conformal framework, but the possibilities are infinite.

Code

The code is available at the following GitHub page [14].

A | Appendix

A.1. Complete description of dataset

The *Snapshot Dataset* contains 14 variables:

ID	primary	published	min players	max players	playing time	min age
1294	Clue	1949	2	6	45	8
1406	Monopoly	1933	2	8	180	8
...	...	...	...	...	...	...

weight	category	mechanic
1.649	['Bluffing', 'Deduction','Movies / Radio / TV theme', 'Murder/Mystery']	['Deduction', 'Dice Rolling', 'Memory','Grid Movement', 'Roll / Spin and Move', 'Square Grid']
1.6278	['Economic', 'Negotiation']	['Auction/Bidding', 'Income', 'Loans', 'Stock Holding','Trading'..]
...	...	...

users rated	publishers	designers	artists
18 448	['Hasbro', 'John Waddington Ltd.', '(Self-Published)', '(Unknown)', 'Alga', 'Basic Fun, Inc.',...]	['Anthony E. Pratt']	['(Uncredited)', 'Peter Dobbin' Matt Groening', 'Rune Johansson',...]
30 803	['(Unknown)', 'Åhlén and Åkerlund', 'Alga', 'Altap', 'ASS Altenburger Spielkarten',...]	['(Uncredited)', 'Charles Darrow', 'Elizabeth J. Magie (Phillips)']	['Edison Girard']
...	...	...	...

After the preprocessing, the dataset we actually use is the following:

ID	primary	published	min players	max players	playing time	min age
1294	Clue	1949	2	6	45	8
1406	Monopoly	1933	2	8	180	8
...	...	...	...	...	...	...

Economic	***Bluffing***	***Deduction***	**...**	**dim publisher**
0	1	1	...	3 512
1	0	0	...	3 771
...	...	...	...	...

The *Historical dataset* is composed of partial daily snapshot of the BGG database. The following tables represent the dataframes related to 12/10/2016, 13/10/2016 and 04/11/2023, the first two days and the last one considered.

12th October 2016:

ID	**Name**	**Year**	**Rank**	**Average**	**Bayes average**	**Users rated**
13	Catan	1995	211	7.29	7.16	62 048
1294	Clue	1949	7 336	5.63	5.55	11 629
1406	Monopoly	1933	12 751	4.42	4.37	18 075
...	...	...	...	...	...	...

13th October 2016:

ID	**Name**	**Year**	**Rank**	**Average**	**Bayes average**	**Users rated**
13	Catan	1995	211	7.29	7.16	62 048
1294	Clue	1949	7 336	5.63	5.55	11 629
1406	Monopoly	1933	12 751	4.42	4.37	18 075
...	...	...	...	...	...	...

4th November 2023:

ID	**Name**	**Year**	**Rank**	**Average**	**Bayes average**	**Users rated**
13	Catan	1995	522	7.1	6.93	122 477
1294	Clue	1949	10 150	5.67	5.57	20 013
1406	Monopoly	1933	25 341	4.36	4.29	34 655
...	...	...	...	...	...	...

A.2. Proof of consistency of GPF

The test statistic proposed is root n consistent meaning that as the sample size grows larger, the probability of correctly rejecting the null hypothesis (if it is false) approaches 1 at a rate proportional to $\sqrt{n}$. This means that the test becomes increasingly reliable as more data is collected. To prove the latter statement we need to obtain first the asymptotic random expression of T_n under the null and alternative hypothesis. Note that in the following only a brief explanation is provided, if interested in the full dissertation please refer to Zhang and Liang[6].

A.2.1. Asymptotic random expression under H_0

Let $y_{i1}(t), y_{i2}(t), \cdots, y_{in_i}(t)$ $i = 1, ..k$ denote the k groups of random functions defined over a given fine interval $t \in \mathcal{T} = [a, b]$. Let $SP(\mu, \gamma)$ denote a stochastic process with mean function $\mu(t), t \in \mathcal{T}$ and covariance function $\gamma(s, t), s, t \in \mathcal{T}$. Assuming that

$$y_{i1}(t), y_{i2}(t), \cdots, y_{in_i}(t) \overset{\text{iid}}{\sim} SP(\mu, \gamma), i = 1, \ldots, k$$

we want to test the equality of the k mean functions:

$$H_0 : \mu_1(t) \equiv \mu_2(t) \equiv \cdots \equiv \mu_k(t), t \in \mathcal{T}$$

For this one-way ANOVA problem, the k mean functions are often decomposed as $\mu_i(t) = \mu_0(t) + \alpha_i(t), i = 1, \ldots, k$ which enables to write the problem equivalently as

$$H_0 : \alpha_1(t) \equiv \alpha_2(t) \equiv \cdots \equiv \alpha_k(t) \equiv 0, t \in \mathcal{T}$$

For any $t \in \mathcal{T}$ the pointwise between-subject variation can be express as:

$$SSR_n(t) = [\mathbf{z}_n(t) + \boldsymbol{\mu}_n(t)]^T (\mathbf{I}_k - \mathbf{b}_n \mathbf{b}_n^T / n)[\mathbf{z}_n(t) + \boldsymbol{\mu}_n(t)]$$

being $\mathbf{I}_k$ the $k \times k$ identity matrix and

$$\mathbf{z}_n(t) = [\sqrt{n_1}[\bar{y}_{1.}(t) - \mu_1(t)], \sqrt{n_2}[\bar{y}_{2.}(t) - \mu_2(t)], \ldots, \sqrt{n_k}[\bar{y}_{k.}(t) - \mu_k(t)]]^T \quad \text{(A.1)}$$

$$\boldsymbol{\mu}_n(t) = [\sqrt{n_1}\mu_1(t), \sqrt{n_2}\mu_2(t), \ldots, \sqrt{n_k}\mu_k(t)], \quad \mathbf{b}_n = [\sqrt{n_1}, \sqrt{n_2}, \ldots, \sqrt{n_k}]^T \quad \text{(A.2)}$$

Proposition. *Under certain conditions and under the null hypothesis, as $n \to \infty$, $T_n \xrightarrow{d} T_0^*$ with*

$$T_0^* \overset{d}{=} (k-1)^{-1} \int_{\mathcal{T}} \boldsymbol{w}(t)^T (\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T) \boldsymbol{w}(t) dt \overset{d}{=} \frac{\sum_{r=1}^{\infty} \lambda_r A_r}{(k-1)}, \quad A_r \overset{iid}{\sim} \chi^2_{k-1} \quad \text{(A.3)}$$

where

- $\boldsymbol{w}(t) = [w_1(t), \ldots, w_k(t)]^T \sim GP_k(\boldsymbol{0}, \gamma_w \boldsymbol{I}_k)$ *and* $\gamma_w(s, t) = \gamma(s, t) / \sqrt{\gamma(s, s)\gamma(t, t)}$
- $\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T$ *is the limit of* $\boldsymbol{I}_k - \boldsymbol{b}_n \boldsymbol{b}_n^T$ *and* $\boldsymbol{b} = [\sqrt{\lim_{n\to\infty} n_1/n}, \ldots, \sqrt{\lim_{n\to\infty} n_k/n},]$
- $\lambda_r, r = 1, \ldots \infty$ *are the decreasing order eigenvalues of* $\gamma_w(s, t)$

This proposition shows tha the asympotical distribution of T_n under H_0 is the same as that of a central χ^2-type mixture.

A.2.2. Approximation of the null distribution

From the above Proposition we get that $w_i(t) \overset{\text{iid}}{\sim} GP(0, \gamma_w)$; moreover $\gamma(s,t)$ can be estimated as the pooled sample covariance function $\hat{\gamma}(s,t)$ which converges to $\gamma(s,t)$ uniformly over $\mathcal{T}^2$. Coherently we can approximate its egenvalues with $\hat{\lambda}_r, r = 1, 2, \dots$ eigenvalues of $\hat{\gamma}(s,t)$ and, since it is often sufficient to use only the $\hat{m}$ positive eigenvalues of $\hat{\gamma}(s,t)$, we can generate a sample from the distributon of T_n under the null hypothesis using

$$\hat{T}_0^* = (k-1)^{-1} \sum_r^{\hat{m}} \hat{\lambda}_r A_r, \quad A_r \overset{\text{iid}}{\sim} \chi^2_{k-1}$$

Although such procedure works well in general, to avoid estimating the eigenvalues and resample A_r from the chi squared distribution a large number of times, it's more convenient to use the *Welch-Satterthwaite χ^2-approximation*. The idea is to approximate the distribution of T_n under the null hypothesis by that of R_w a χ^2-random variable multiplied by a constant, namely $R_w \sim \beta_w \chi^2_{d_w}$. The parameters β_w and d_w are determined by matching the means and the variances of T_n and R_w.

$$E(R_w) = \beta_w d_w \qquad E(T_n) = E(T_0^*) + o(1) = (b-a) + o(1)$$

$$Var(R_w) = 2\beta_w^2 d_w \qquad Var(T_n) = Var(T_0^*) + o(1) = \frac{2tr(\gamma_w^{\otimes 2})}{k-1} + o(1)$$

where $\gamma_w^{\otimes 2}(s,t) = \int_{\mathcal{T}} \gamma_w(s,u)\gamma_w(u,t)du$. Hence matching the parameters and approximating the covariance function with its pooled sample version we get:

$$\hat{\beta}_w = \frac{tr(\hat{\gamma}_w^{\otimes 2})}{(k-1)(b-a)}, \quad \hat{d}_w = \frac{(k-1)(b-a)^2}{tr(\hat{\gamma}_w^{\otimes 2})}$$

Then, the proposed GPF test is conducted by computing th p-value using the following approximate distribution under the null hypothesis

$$T_n \sim \hat{\beta}_w \chi^2_{\hat{d}_w}$$

and for any given significance levelα the estimated critical value of T_n is specified as

$$\hat{T}_n(\alpha) = \hat{\beta}_w \chi^2_{\hat{d}_w}(\alpha) \quad \xrightarrow{n \to \infty} \quad T_0^*(\alpha) = \beta_w \chi^2_{d_w}(\alpha)$$

A.2.3. The asympotic power

To finally study the asymptotic power of the GFP test, let's specify the following local alternative

$$H_{1n} : \mu_i(t) = \mu_0(t) + \frac{d_i(t)}{\sqrt{n_i}}, \quad i = 1, 2, \dots, k$$

where $\mu_0(t)$ is the grand mean function and $d_i(t)$ are any real functions independent of n. With the local alternative we represent a scenario where there may be small, localized deviations from the null hypothesis. We can write $\boldsymbol{\mu}(t) = \mu_0(t)\mathbf{1}_k +$

$[\frac{d_1(t)}{\sqrt{n_1}}, \frac{d_2(t)}{\sqrt{n_2}}, \ldots, \frac{d_k(t)}{\sqrt{n_k}}]^T$. From expression (A.1) follows that $\boldsymbol{\mu}(t) = \mu_0(t)\mathbf{b}_n + \mathbf{d}_n(t)$. As n tends to ∞, the local alternative will tend to the null one with the root-n rate, however as long as the information provided by $\mathbf{d}(t)$ diverges to ∞ the GPF test can detect the local alternative (i.e. correctly reject the null hypothesis in favor of the local alternative hypothesis) with probability 1. In this sense we can call the test root-n consistent. To prove the latter consider first the following proposition.

Proposition. *Under certain conditions and under the local alternative hypothesis, as $n \to \infty$, $T_n \xrightarrow{d} T_1^*$ with*

$$T_1^* \stackrel{d}{=} (k-1)^{-1}_{\mathcal{T}} \left[\boldsymbol{w}(t) + \boldsymbol{h}(t)\right]^T \left(\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T\right) \left[\boldsymbol{w}(t) + \boldsymbol{h}(t)\right]^T dt \tag{A.4}$$

$$\stackrel{d}{=} (k-1)^{-1} \left[\sum_{r=1}^{m} \lambda_r B_r + (\delta^2 - \sum_{r=1}^{m} \delta_r^2)\right], \quad B_r \stackrel{iid}{\sim} \chi^2_{k-1}(\delta_r^2/\lambda_r) \tag{A.5}$$

where

- $\boldsymbol{w}(t) = [w_1(t), \ldots, w_k(t)]^T \sim GP_k(\boldsymbol{0}, \gamma_w \boldsymbol{I}_k)$
- $\boldsymbol{h}(t) = \boldsymbol{d}(t)/\sqrt{\gamma(t,t)}$
- $\delta_r^2 = \| \int_{\mathcal{T}} (\boldsymbol{I}_{k-1}, \boldsymbol{0})\, \boldsymbol{U}^T \boldsymbol{h}(t)\phi_r(t)dt\|^2, r = 1, 2, \ldots, m$
- $\delta^2 = \int_{\mathcal{T}} \boldsymbol{h}(t)^T (\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T)\boldsymbol{h}(t)dt$
- $\phi_1(t), \ldots, \phi_m(t)$ *are the eigenfunctions associated to* $\lambda_1, \ldots, \lambda_m$ *eigenvalues of* $\gamma_w(s,t)$
- $\boldsymbol{U}$ *is the matrix of the singular value decomposition of* $\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T$*:*

$$\boldsymbol{I}_k - \boldsymbol{b}\boldsymbol{b}^T = \boldsymbol{U} \begin{pmatrix} \boldsymbol{I}_{k-1} & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{pmatrix} \boldsymbol{U}^T$$

This proposition shows that the asympotical distribution of T_n under H_{1n} is the same as that of a non-central χ^2-mixture plus a constant.
Now let $\delta_\lambda^2 \leq \lambda_1 \sum_{r=1}^m \delta_r^2 \leq \lambda_1 \delta^2$, we finally state:

Proposition (The Asymptotic Power of GPF). *Let $Z \sim \mathcal{N}(0,1)$ then under certain conditions and under the local alternative, as $n \to \infty$, the power of T_n is given by*

$$P(T_n \geq \hat{T}_n(\alpha)) = P(T_0^* + 2(k-1)^{-1}\delta_\lambda Z \geq T_0^*(\alpha) - (k-1)^{-1}\delta^2) + o(1)$$

which will tend to 1 as $\delta \to \infty$.

Proof. Thanks to (A.5) and (A.3), we can write

$$\begin{aligned} B_r &\stackrel{d}{=} z_{1r}^2 + \cdots + z_{(k-2)r}^2 + (z_{(k-1)r} + \delta_r/\sqrt{\lambda_r})^2 \\ &\stackrel{d}{=} Ar + 2z_{(k-2)r}\delta_r/\sqrt{\lambda_r} + \delta_r^2/\sqrt{\lambda_r} \end{aligned}$$

where $z_{ir} \overset{iid}{\sim} \mathcal{N}(0,1), i = 1, \ldots, k-1; r = 1, \ldots, m$. Thus

$$\begin{aligned} T_1^* &\overset{d}{=} (k-1)^{-1} \left[\sum_r^m \lambda_r A_r + 2 \sum_{r=1}^m \lambda_r^{1/2} \delta_r z_{(k-1)r} + \delta^2 \right] \\ &\overset{d}{=} T_0^* + 2(k-1)^{-1} \delta_\lambda Z + (k-1)^{-1} \delta^2 \end{aligned}$$

where $Z = 2 \sum_{r=1}^m \lambda_r^{1/2} \delta_r z_{(k-1)r} / \delta_\lambda \sim \mathcal{N}(0,1)$. Now, as n goes to infinity under H_{1n} we have $T_n \to T_1^*$ and $\hat{T}_n(\alpha) \to T_0^*(\alpha)$, so using the latter expression for T_1^* we get:

$$P(T_n \geq \hat{T}_n(\alpha)) = P(T_0^* + 2(k-1)^{-1} \delta_\lambda Z + (k-1)^{-1} \delta^2 \geq T_0^*(\alpha)) + o(1)$$

Finally it's immediate to see that the latter quantity goes to 1 as δ goes to ∞ since:

- if $\delta_r < \infty$ then it is obvious that the probability reaches the value 1
- if also $\delta_r \to \infty$ then $P(T_n \geq \hat{T}_n(\alpha)) = P(Z \geq -\delta^2/[2\delta_r^2]) + o(1) \geq P(Z \geq -\delta^2/[2\delta\lambda_1]) + o(1) = P(Z \geq -\delta/[2\lambda_1]) + o(1) \xrightarrow{\delta \to \infty} 1$

$\square$

A.3. Results

A.3.1. Robust Generalized Additive Model

Variable	Coeff	SE	t-value	p-value
(Intercept)	5.1283	0.0147	348.0010	0
Economic	0.0741	0.03109	2.3840	0.0171
Negotiation	-0.2434	0.0441	-5.5200	0
Political	0.2173	0.0483	4.4990	0
Card Game	0.0888	0.0175	5.0650	0
Fantasy	0.1270	0.0231	5.5050	0
Medieval	0.0902	0.0330	2.7320	0.0063
Ancient	0.1535	0.0394	3.9000	0
Territory Building	0.1961	0.0457	4.2890	0
Civilization	0.0896	0.0577	1.5510	0.1209
Nautical	0.1618	0.0412	3.9220	0
Children's Game	-0.1362	0.0377	-3.6150	0.0003
City Building	0.2495	0.0456	5.4730	0
Exploration	0.1502	0.0356	4.2240	0
Farming	0.2209	0.0655	3.3730	0.0007
Bluffing	0.0644	0.0318	2.0260	0.0428
Science.Fiction	0.1617	0.0286	5.6530	0
Dice	0.1523	0.0256	5.9480	0
Fighting	0.1189	0.0282	4.2240	0
Print Play	0.1439	0.0425	3.3860	0.0007
Miniatures	0.1524	0.0361	4.2190	0
Wargame	-0.2974	0.0298	-9.9680	0
Space Exploration	0.1224	0.0623	1.9660	0.0493
Renaissance	0.1604	0.0628	2.5550	0.0106
Electronic	-0.184	0.0741	-2.4840	0.0130
Horror	0.1655	0.0404	4.0980	0
Novel based	-0.1408	0.0464	-3.0340	0.0024
Aviation Flight	0.1861	0.0626	2.9730	0.0030
Movies TV Radio theme	-0.1923	0.0333	-5.7720	0
Memory	-0.2408	0.0464	-5.1930	0
Trivia	-0.2752	0.0481	-5.7210	0
Industry Manufacturing	0.1235	0.0606	2.0360	0.0418
Trains	0.1484	0.0602	2.4640	0.0137
Animals	0.1743	0.0300	5.8200	0
Murder.Mystery	-0.1221	0.0584	-2.0910	0.0365
Transportation	0.1335	0.0560	2.3820	0.0172
Prehistoric	0.1779	0.0800	2.2230	0.0262
Action Dexterity	0.1401	0.0349	4.0120	0
Spies Secret Agents	0.1539	0.0701	2.1950	0.0282
Book	-0.2253	0.0808	-2.7890	0.0053
Environmental	0.1796	0.0737	2.4370	0.0148

Variable	edf	Ref.df	F-value	p-value
s(log playingtime)	5.6200	6.4870	18.2870	0
s(minage)	9.4380	9.8670	27.0360	0
s(yearpublished)	9.8440	9.9850	66.4630	0
s(dimpublisher)	8.7980	9.5500	14.5970	0
s(maxplayers)	7.8350	8.6290	5.6070	0
s(weight)	9.7710	9.9760	29.2900	0
s(I(maxplayers:log(playingtime + 1)))	8.8670	8.9910	9.9950	0
s(I(weight:log(playingtime + 1)))	7.2580	8.2610	15.4160	0
s(I(weight:dimpublisher))	8.9480	8.9960	54.6350	0
s(I(dimpublisher:log(playingtime + 1)))	8.3580	8.7780	8.1860	0

A.3.2. ANOVA on the Modified Hypograph

Category	Test on the functions		Test on the derivatives	
	Coefficient	p-value	Coefficient	p-value
Economic	0.0860	0	0.0642	0
Fantasy	0.0119	0	0.0409	0
Medieval	0.0518	0	0.0260	0
Ancient	0.0810	0	0.0405	0
Territory Building	0.0597	0	0.0722	0
Civilization	0.0736	0	0.0778	0
City Building	0.0860	0	0.0991	0
Exploration	0.0368	0	0.0413	0
Farming	0.0679	0	0.0900	0
Bluffing	0.0574	0	0.0471	0
Science Fiction	0.0359	0	0.0213	0
Fighting	0.0457	0	0.0224	0
Renaissance	0.1128	0	0.0830	0
Horror	0.0427	0	0.0450	0
Novel based	0.0510	0	0.0455	0
Aviation Flight	0.0737	0	0.0463	0
Real time	0.0324	0	0.0699	0
Medieval:Ancient	-0.1545	0	-0.1301	0
Ancient:City Building	-0.1395	0	-0.0828	0.001
Territory Building:Exploration	-0.1243	0	-0.0831	0.006

A.3.3. Complete Cox Model

Variable	Coeff	Exp(Coeff)	SE(Coeff)	p-value	lower	upper
bs(weight, df = 3)1	-0.5967	0.5506	0.0872	0	0.4642	0.6533
bs(weight, df = 3)2	-0.2849	0.7521	0.1417	0.0443	0.5697	0.9927
bs(weight, df = 3)3	-0.9138	0.401	0.1819	0	0.2807	0.5728
bs(minage, df = 3)1	0.5129	1.6702	0.1278	1e-04	1.3	2.1457
bs(minage, df = 3)2	-1.5859	0.2048	0.2266	0	0.1313	0.3192
bs(minage, df = 3)3	1.528	4.6091	0.3442	0	2.3477	9.049
bs(dimpublisher, df = 3)1	-3.7771	0.0229	0.2254	0	0.0147	0.0356
bs(dimpublisher, df = 3)2	-1.7638	0.1714	1.9592	0.368	0.0037	7.9736
bs(dimpublisher, df = 3)3	-35.779	0	11.5887	0.002	0	0
bs(playingtime, df = 3)1	-0.5451	0.5798	0.1086	0	0.4686	0.7173
bs(playingtime, df = 3)2	0.6503	1.9161	0.1571	0	1.4083	2.607
bs(playingtime, df = 3)3	-0.1694	0.8442	0.1448	0.2421	0.6356	1.1212
bs(Year, df = 3)1	-3.3815	0.034	0.3228	0	0.0181	0.064
bs(Year, df = 3)2	0.4346	1.5443	0.1577	0.0058	1.1338	2.1035
bs(Year, df = 3)3	-2.6567	0.0702	0.197	0	0.0477	0.1032
Negotiation	0.1068	1.1127	0.0468	0.0225	1.0152	1.2196
Political	-0.2106	0.8101	0.055	1e-04	0.7273	0.9024
Card Game	-0.0923	0.9118	0.019	0	0.8785	0.9464
Territory Building	-0.2112	0.8096	0.0553	1e-04	0.7265	0.9022
Children's Game	0.1334	1.1427	0.0335	1e-04	1.0701	1.2203
City Building	-0.2333	0.7919	0.0571	0	0.7081	0.8856
Exploration	-0.1272	0.8805	0.0417	0.0023	0.8115	0.9555
Farming	-0.2369	0.7891	0.0834	0.0045	0.6701	0.9292
Dice	-0.1002	0.9047	0.0286	5e-04	0.8553	0.9569
Fighting	-0.1059	0.8995	0.0316	8e-04	0.8456	0.9569
Miniatures	-0.1117	0.8943	0.0402	0.0055	0.8265	0.9677
Wargame	0.2002	1.2217	0.0271	0	1.1584	1.2884
Memory	0.1535	1.1659	0.0494	0.0019	1.0582	1.2844
Industry Manufacturing	-0.2684	0.7646	0.0742	3e-04	0.6611	0.8843
Animals	-0.0812	0.922	0.0333	0.0148	0.8637	0.9843
Transportation	-0.1729	0.8413	0.0605	0.0043	0.7472	0.9472
Action Dexterity	-0.0877	0.9161	0.0373	0.0186	0.8515	0.9855
Spies Secret Agents	-0.2368	0.7892	0.0834	0.0045	0.6701	0.9293
Book	0.2246	1.2518	0.0842	0.0076	1.0614	1.4765
Environmental	-0.2367	0.7892	0.0939	0.0117	0.6566	0.9487

Test	Statistic	df	p-value
Likelihood ratio test	7365	35	0
Wald test	5594	35	0
Score (logrank) test	6310	35	0

Bibliography

[1] J. Peiser, "We're in a golden age of board games. it might be here to stay." https://www.washingtonpost.com/business/2022/12/24/board-game-popularity/, 2022. [Online; accessed 15-Feb-2024].

[2] The Guardian, "Toymaker hasbro lays off 1,100 staff as holiday season fails to lift sales." https://www.theguardian.com/business/2023/dec/11/toymaker-hasbro-layoffs, 2023. [Online; accessed 15-Feb-2024].

[3] rfordatascience, "tidyTuesday BGG." https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-01-25/readme.md, 2022. [Online; accessed 10-Feb-2024].

[4] beefsack, "bgg-ranking-historicals." https://github.com/beefsack/bgg-ranking-historicals, 2016-2024. [Online; accessed 10-Feb-2024].

[5] J. Bujnowicz, *bggAnalytics: BoardGameGeek's Board Game Data Analysis Tools*, 2022. R package version 0.2.1.

[6] 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.

[7] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. Springer, 2010.

[8] J. Ramsay, *fda: Functional Data Analysis*, 2023. R package version 6.1.4.

[9] D. Cook, "The circle of life: An analysis of the game product lifecycle." https://www.gamedeveloper.com/design/the-circle-of-life-an-analysis-of-the-game-product-lifecycle/, 2007. [Online; accessed 15-Feb-2024].

[10] T. Therneau, "Spline terms in a cox model," *MIT*, 2017.

[11] D. Long, "Cox proportional hazard regression and splines: A cautionary tale for exposure-response assessment," *Department of Biostatistics, University of Alabama at Birmingham*, 2019.

[12] P. H. C. Eilers, B. D. Marrx, and M. Durban, "Twenty years of p-splines," *SORT (Statistics and Operations Research Transaction)*, 2015.

[13] Z. Zegras, "The rise of the professional gamer: a new age career." https://zzeger.com/EN/blog/rise-of-professional-gamer/, 2023. [Online; accessed 15-Feb-2024].

[14] Colombo, D'Amicantonio, and Lancellotti, "boardgames-nps-project." https://github.com/federico-lancellotti/boardgames-nps-project.git, 2023-2024. [Online; accessed 14-Feb-2024].

[15] T. Gorecki and L. Smaga, *fdANOVA: Analysis of Variance for Univariate and Multivariate Functional Data*, 2018. R package version 0.1.2.